# Convergence in Evolutionary Programs with Self-Adaptation

**Garrison W. Greenwood**
Department of Electrical and Computer Engineering
Portland State University
Portland, OR 97207 USA
greenwd@ee.pdx.edu

**Qiji J. Zhu**
Department of Mathematics and Statistics
Western Michigan University
Kalamazoo, MI 49008 USA
zhu@math-stat.wmich.edu

**Abstract**

Evolutionary programs are capable of finding good solutions to difficult optimization problems. Previous analysis of their convergence properties has normally assumed the strategy parameters are kept constant, although in practice these parameters are dynamically altered. In this paper we propose a modified version of the 1/5-success rule for self-adaptation in evolution strategies (ES). Formal proofs of the long-term behavior produced by our self-adaptation method are included. Both elitist and non-elitist ES variants are analyzed. Preliminary tests indicate an ES with our modified self-adaptation method compares favorably with both a non-adapted ES and a 1/5-success rule adapted ES.

**Keywords**

Convergence, evolutionary programming, evolution strategies, self-adaptation

## 1 Introduction

Evolutionary programs are probabilistic algorithms that use the principles of population genetics to search for problem solutions. They are capable of finding good solutions to a wide variety of optimization problems, including NP-hard combinatorial optimization problems (Bäck, 1996). The current implementations of evolutionary programs are closely related but independently developed approaches: *evolution strategies* (ES) and *evolutionary programming* (EP). Although most of the work in this paper discusses ES, the concepts are equally applicable to EP.

It is easy to formulate an ES as a finite state Markov chain. Under some rather mild assumptions, Fogel (1994) proved that this Markov chain asymptotically converges to an absorbing state (the globally optimal solution). In practice, however, most ES implementations involve self-adaptations. These self-adaptations lead to non-stationary Markov chains—thus limiting the significance of previous convergence studies. In fact, recently Rudolph (1999) constructed a non-convex function for which the ES with 1/5-success rule does not converge to the global optimum with probability one.

In this paper we propose a modification to the 1/5-success rule. We prove our new self-adapted (elitist) ES algorithm converges with probability one for a wide range of (non-convex) functions. Our proof uses measure theory rather than Markov chain theory. Measure theory possesses a number of powerful tools for characterizing sets and spaces; it has been previously used to analyze evolutionary algorithms (Rudolph, 1997). We also prove a corresponding non-elitist ES algorithm with self-adaptation locates the global minimum with probability one. Numerical results show that our new

 

ES algorithm with self-adaptation compares favorably with both the 1/5-success rule and an ES algorithm without self-adaptation. We begin with an overview of ES.

## 2 Background

Each optimization problem solution can be considered as a distinct point in a finite space of all possible solutions. Associated with each point in this space is a fitness value which is a measure of goodness or quality. The ES begins with an initial population of $\mu > 0$ "parents" which are randomly chosen. Each parent encodes a set of parameters that completely describes a solution to the problem of interest. In the $(\mu + \lambda)$-ES, these $\mu$ parents (candidate solutions) produce $\lambda$ offspring (new solutions) by mutating one or more problem parameters. Parents and offspring compete equally for survival; only the $\mu$ best (i.e., those with the highest fitness) will survive to reproduce in the next generation. Done properly, the population will evolve towards increasingly better regions of the search space by means of reproduction and survival of the fittest. Its canonical form is given in Figure 1. In the $(\mu, \lambda)$-ES version, $\mu$ parents produce $\lambda > \mu$ offspring, but only the $\mu$ best offspring survive. Thus individuals live for only a single generation regardless of their fitness level. This approach may result in short periods of recession, but it does avoid long periods of stagnation. There are other variants of the ES and the interested reader should see Bäck (1996) for further information. Both the $(\mu + \lambda)$-ES and $(\mu, \lambda)$-ES versions are investigated in this paper.

1. Randomly create an initial population of $\mu$ individuals.

2. From the current population generate $\lambda$ offspring by applying a reproduction operator (described below).

3. Determine the fitness of each individual.

4. Select the $\mu$ fittest individuals for survival. Discard the other individuals.

5. Proceed to step 2 unless $\Gamma$ generations have been processed.

Figure 1: The canonical form of the $(\mu + \lambda)$-ES algorithm.

The ES is readily formulated as a Markov chain. Denote by $i \mapsto j$ the transition from state $i$ to state $j$ in a Markov chain operating in a finite space $S$. The Markov chain is described by a sequence of transition matrices $\{P_k\}_{k=1}^{\infty}$ where at the $k$-th step $p_{ij}^{(k)} \in P_k$ is the probability of $i \mapsto j$. In a stationary Markov chain $p_{ij}^{(k)}$ is independent of $k$.

Now consider a (1+1)-ES—a single parent produces a single offspring via stochastic mutation and the one with the highest fitness becomes the parent in the next generation. Let state $i$ and $j$ be the parent and offspring, respectively. The elitist property permits only the fittest to survive; in the Markov chain $i \mapsto j$ means the the offspring replaces the parent and $i \mapsto j \mapsto i$ means the offspring is discarded.

To fix ideas, consider a simple (1+1)-ES used to find $x \in R^N$ which minimizes a highly multi-modal, objective function $f(x)$. The state space is finite since any computer implementation of real numbers is discrete. Offspring are produced by mutating the parent with some $\delta x$ term

$$x' = x + \delta x \tag{1}$$

Here $\delta x$ is a normally distributed, random variable with zero mean and standard deviation $\sigma$. If state $i$ is the parent and state $j$ is the offspring, then the interpretation of $p_{ij}^{(k)}$ becomes the probability that at generation $k$, $\delta x$ is of sufficient magnitude to take the ES from parent $x$ (state $i$) to offspring $x'$ (state $j$).

Although $\sigma$ may be kept constant, usually it is periodically modified—a procedure called self-adaptation—to improve the search efficiency. For example, Rechenberg (1973) proposed the *1/5-success rule* to self-adapt (1+1)-ES algorithms. First, the percentage of successful mutations ($p_s$) is recorded over intervals of around $10 \cdot n$ trials, where $n$ is the number of parameters undergoing evolution[1]. Then, with $c = 0.85$, $\sigma$ is modified as follows:

$$\sigma = \begin{cases} \sigma/c & \text{if } p_s > 1/5 \\ \sigma \cdot c & \text{if } p_s < 1/5 \\ \sigma & \text{if } p_s = 1/5 \end{cases} \tag{2}$$

Asymptotic convergence in evolutionary programs with a fixed $\sigma$ has previously been proven by Fogel (1994). A fixed $\sigma$ naturally makes the corresponding Markov chain model stationary. Unfortunately, most ES implementations allow self-adaptation of $\sigma$—making most ES Markov chain models non-stationary. To see this, let $S$ be a finite space and suppose at step $k$ the Markov chain is at state $i \in S$. For each potential offspring state $j \in S$ there is a $p_{ij}^{(k)} > 0$. Assume the produced offspring has a lower fitness so that it is discarded. Then $i \mapsto j \mapsto i$, and at step $k+1$ the Markov chain will again be in state $i$. If $\sigma$ is a constant, $p_{ij}^{(k+1)} = p_{ij}^{(k)}$ and the Markov chain is stationary. But, if $k$ is some integer multiple of $n$, Eq. (2) says it is entirely possible that $p_{ij}^{(k+1)} \neq p_{ij}^{(k)}$, which makes the Markov chain non-stationary. Since, in practice, ES implementations operate in this manner, Fogel's proof cannot be used to describe their behavior in the limit.

Self-adaptation can improve a search algorithm's performance, but its use does not always guarantee convergence to the global optimum. For example, Rudolph (1999) constructed a function with one local minimum and one global minimum for which the 1/5-success rule did not lead to convergence of an ES. Intuitively, the 1/5-success rule may cause an ES to become stuck at a local minimum point as follows. When the algorithm happens to search near a local minimum, a percentage of successful mutations will drop below 1/5. Now, the 1/5-success rule calls for a reduction in $\sigma$, which restricts the search to an even smaller neighborhood of the local minimum. Such the search will eventually become stuck on a local minimum and miss the global minimum.

## 3 Modified 1/5-success Rule and its Convergence

In this section we propose a modification to the 1/5-success rule for adapting the ES strategy parameter $\sigma$. The elitist version (a (1+1)-ES) is discussed first, followed by a non-elitist version (a (1,$\lambda$)-ES). In both cases a theorem on the long term behavior is provided.

[1] A mutation is "successful" if the offspring has a higher fitness level than its parent.

### 3.1 The Elitist (1+1)-ES

We consider the following general optimization problem:

*Find the global minimum of a function $f$ over a bounded space $E \subset R^N$.*

subject to the conditions

($i$) the search space $E$ is Lebesgue measurable

($ii$) $f : E \to R$ is a measurable function[2]

($iii$) a global minimum for $f$ exists in $E$

($iv$) if $m(S)$ denotes the Lebesgue measure of a set $S$, then for any positive number $a > 0$,

$$m(\{x|f(x) \leq \min_{y \in E} f(y) + a\}) > 0$$

Several remarks below explain these conditions:

1. We observe first that in general stochastic search methods cannot find isolated discontinuous minimum. This is illustrated by the following pathological example. Let $E = [-1, 1]$ with $f(x) = 1$ if $x \in E \backslash \{0\}$ and $f(0) = 0$. Obviously $x = 0$ is the global minimum of $f$. However, starting from any $x \neq 0$, the probability of $x + \delta x$ hitting 0 is zero. Therefore, we cannot expect an ES to converge to this global optimum.

2. Conditions ($i$) and ($ii$) are both mild and reasonable. Intuitively, a stochastic search algorithm works only if, for any positive number $a$, the probability of the search intersecting the set $\{x|f(x) \leq \min_{y \in E} f(y) + a\}$ is greater than zero. This establishes the need for some concept of "size", or *measure* with respect to a subset. A very convenient framework is the Lebesgue measure, which provides a natural extension to the familiar notions of area or volume (Reddy, 1998). It is also worth mentioning that a Lebesgue measurable space includes both open and closed sets as well as their unions and intersections—it may even be disjoint. Measurable functions therefore encompass a broad family of objective functions. It is for these reasons we decided to work on the class of Lebesgue measurable functions on Lebesgue measurable sets.

3. Conditions ($i$)-($iv$) are satisfied as long as $E$ is a bounded region and $f$ is continuous, although the pair $(f, E)$ that satisfies these conditions is much broader. Specifically, if $E$ is a bounded region and $f$ is a piecewise continuous function in the general sense, i.e., there exists a finite partition $E_i, i = 1, 2, \ldots, I$ of $E$ where each $E_i$ is a subregion and $f$ is continuous when restricted to $E_i$, then $(f, E)$ also satisfies conditions ($i$)-($iv$).

4. Condition ($iv$) is the key. It precludes $f$ from having an isolated discontinuous optimum point. This condition cannot be dispensed with as illustrated by the example in item 1 above. The main focus in this section is to show that this condition is sufficient to ensure our modified 1/5-success rule for ES converges to the global minimum with probability one.

[2] A measurable function is the limit a.e. of a sequence of step functions.

We modify the 1/5-success rule as follows. Let $p_s$ denote the percentage of successful mutations over $n$ ($n \geq 5$) trials. Then, with $c = 0.85$, $\sigma$ is modified every $n$ steps as follows:

$$\sigma = \begin{cases} \min(\sigma/c, D) & \text{if } p_s > 1/5 \\ \sigma & \text{if } p_s = 1/5 \\ \sigma \cdot c & \text{if } 1/20 \leq p_s < 1/5 \\ \min(2\sigma, D) & \text{if } p_s < 1/20 \end{cases} \tag{3}$$

Where $D$ is the diameter of the search region $E$. Notice that the search is conducted over a bounded space. Therefore, $\sigma$ needs never exceed $D$. It is therefore reasonable to choose $D$ as an upper bound of $\sigma$, i.e., whenever $\sigma$ reaches $D$ we stop increasing it. The main modification we made here is when $p_s < 1/20$ we increase $\sigma$, which helps to escape any local minima. We can now investigate the convergence property our modified 1/5-success rule imparts to a (1+1)-ES.

A point $x_0 \in E$ is a solution to the minimization problem if

$$f(x_0) \leq \min_{y \in E} f(y) + \varepsilon.$$

where $x \in E$ and $\varepsilon$ is the error bound, which reflects the accuracy needed in the application at hand. When we say a search algorithm converges with probability one, we mean that, for any error bound $\varepsilon > 0$, the algorithm converges to a solution corresponding to this error bound with probability one. The convergence of a (1+1)-ES with our modified 1/5-success rule is expressed in the following theorem:

**Theorem 1** *Let f be an objective function defined over a search space E. If the pair (f,E) satisfy conditions (i)-(iv) above, then a (1+1)-ES with self-adaptation as given in equation* (3) *will converge to the globally minimum value of f with probability one.*

*Proof.* Let us assume that the error bound is $\varepsilon$. In other words, $x_0$ is a solution if

$$f(x_0) \leq \min_{y \in E} f(y) + \varepsilon$$

During each iteration of the ES, a parent with value $x$ will produce an offspring with value $x + \delta x$ where $\delta x$ is a normally distributed random variable. The parent and its offspring will compete equally for survival; we update (replace) the parent with its offspring if

$$f(x + \delta x) < f(x) - \varepsilon$$

Each update improves the fitness (decreases $f$) by more than $\varepsilon$. Hence, the globally optimal solution can always be reached—from any initial point $x \in E$—after a finite number of updates. To complete the proof, we need only show that if $x$ is *not* a solution, i.e.,

$$f(x) > \min_{y \in E} f(y) + \varepsilon$$

then with probability one, $x$ is updated within a finite number of steps. We define

$$\alpha = \frac{1}{2}(f(x) - \min_{y \in E} f(y) - \varepsilon) > 0.$$

Notice that

$$\begin{aligned} f(x) - \varepsilon &= \min_{y \in E} f(y) + 2\alpha \\ &> \min_{y \in E} f(y) + \alpha \end{aligned} \tag{4}$$

This allows us to define the following set

$$A = \{z \in E | f(z) \leq \min_{y \in E} f(y) + \alpha\}.$$

By condition $(iv)$, $m(A) > 0$.

We proceed by way of contradiction. Suppose that $x$ is not updated in a finite number of steps. Then, in any $n$ steps $p_s = 0$ and the self-adaptation mechanism will continually double $\sigma$ until it reaches an upper bound of $\sigma = D$ in finite number of steps, which promotes ergodicity. With $\delta x$ normally distributed and $m(A) > 0$, we have

$$p(x) = \text{Prob}\{x + \delta x \in A\} > 0 \tag{5}$$

The probability that no update occurs at the $i$-th step is

$$\begin{aligned} q(i) &= \text{Prob}\{f(x + \delta x) \geq f(x) - \varepsilon\} \\ &\leq \text{Prob}\{f(x + \delta x) > \min_{y \in E} f(y) + \alpha\} \end{aligned} \tag{6}$$

where the latter inequality results from Eq. (4). It then follows that

$$\begin{aligned} q(i) &\leq \text{Prob}\{x + \delta x \in E \backslash A\} \\ &= 1 - \text{Prob}\{x + \delta x \in A\} \\ &= 1 - p(x) \\ &< 1 \end{aligned}$$

Since we assume that $x$ is not updated, $p(x)$ remains unchanged in the whole process. Since each iteration is independent, the probability of no update occurring in $n$ consecutive iterations equals $q(1)q(2)\cdots q(n) = \Pi_{i=1}^{n} q(i) \leq (1 - p(x))^n$. Thus,

$$\lim_{n \to \infty} \Pi_{i=1}^{n} q(i) = 0 \tag{7}$$

Hence, if $x$ is not a solution, then within a finite number of steps an update is guaranteed to occur. □

It is important to emphasize several key points. First, if $x$ is a global minimum then the probability that no update to $x$ occurs is $\text{Prob}\{f(x+\delta x) \leq f(x)-\varepsilon\} = 1$, because of the elitist strategy in the ES. Thus, the algorithm will not discard the globally optimal solution. Second, the proof does not depend on a normally distributed $\delta x$. The only requirement is we must have $m(A) > 0$, which means $p(x) = \text{Prob}(x + \delta x \in A) > 0$. *Any* distribution of $\delta x$ with a density function that is nowhere zero over $E$ satisfies this requirement. The normal distribution has this property, but so do many others. This has strong implications. Some researchers suggest Cauchy distributed mutations, with strategy parameter $t$, can sometimes enhance the search (Yao et al., 1997). The above proof also holds for these versions of ES.

We don't differentiate between two function values that differ less than the positive error bound $\varepsilon$. In fact, the validity of the above proof requires that an update occurs if and only if the fitness has increased by more than the error bound $\varepsilon$. To see why this is necessary, suppose the current state is not the global solution and assume no $\delta x$ improves fitness (decreases $f(\cdot)$) by more than $\gamma < \varepsilon$. That is, $\forall\, \delta x$

$$\begin{aligned} f(x + \delta x) &\geq f(x) - \gamma \\ &> f(x) - \varepsilon \end{aligned}$$

Then by Eq. (6), $q(i) = 1 \; \forall \; i$. This means the limit in Eq. (7) no longer holds, the parent is never replaced and convergence becomes impossible. Conversely, $\gamma > \varepsilon$ makes $q(i) \leq 1 - p(x) < 1$ and the zero product limit is attained.

Restricting the updates only to instances where a fitness improvement of more than $\varepsilon$ occurs should *not* be considered a limitation. In practice $\varepsilon$ is either the machine epsilon (i.e., machine precision) or the desired accuracy—no practical benefit is therefore gained from fitness improvements less than the error bound $\varepsilon$. In fact, Theorem 1 is invalid for $\varepsilon = 0$ because the probability of a stochastic search algorithm exactly hitting the global optimum is zero.

While Theorem 1 asserts an ES converges with probability one, it does not provide information on the rate of the convergence. It is possible to compute convergence rates for very specific objective functions (Bäck, 1996), but our assumptions about $f$ are quite general. The significance of this convergence theorem is as follows. Arguably the best known evolutionary search algorithm is the genetic algorithm. Fogel (1994) showed genetic algorithms can converge to local optima if they don't use some mutation in producing offspring. Our results here show that with the modified 1/5-success rule we can avoid stagnation on local minimum points. This means, in principle, one can always run an evolutionary program longer if the current solution is unsatisfactory—something which may prove to be futile with genetic algorithms. Some numerical results comparing our modified self-adaptation ES with other methods are given in Section 4.

### 3.2 The Non-Elitist (1,λ)-ES

We now investigate long term behavior in the non-elitist ES algorithm with self-adaptation and $\lambda > 1$. Because individuals live for only a single generation, it will be necessary to track the best fit solution off-line. More precisely, let $\{x_1, x_2, \ldots, x_l\}$ be a sequence of length $l$, where each $x_j$ is the single parent in generation $j$. Additionally, define the off-line *tracking sequence* $v_1, v_2, \ldots, v_l$ where $v_1 = f(x_1)$ and

$$v_j \;=\; \min(f(x_j), v_{j-1}) \;;\quad j = 2, 3, \ldots, l-1, l.$$

Note that for any integer $n$,

$$v_n \;=\; \min\{f(x_i)\} \;;\quad i = 1, 2, \ldots, n.$$

We say that a sequence $\{x_1, \ldots, x_n, \ldots\}$ *locates* the global minimum provided that $v_n$ converges to the global minimum of $f(\cdot)$.

Offspring in the (1,λ)-ES are produced as before (see Eq. (1)) but the self-adaptation of parameter $\sigma$ is determined as follows: We track the percentage rate $p_s$ of updates of $v_i$ in $n$ steps and adapt $\sigma$ every $n$ steps according to Eq. (3). Note that here we use the percentage update rate of $v_i$ instead of $f(x_i)$ because the non-elitist nature of the algorithm makes $f(x_i)$ no longer an indicator of how much progress has been made. Our objective is now to prove the following theorem:

**Theorem 2** *Let f be an objective function defined over a search space E. If the pair (f,E) satisfies conditions (i)-(iv), then a (1,λ)-ES, with self-adaptation defined by Eq. (3), will locate the global minimum of f with probability one.*

*Proof.* Let us assume that the error bound is $\varepsilon$. Observe that each update of $v_n$ improves the tracking value by at least $\varepsilon$. Therefore, a global minimum will be reached by $v_n$

after a finite number of updates. Thus, we need only show that, for any $n$, if $v_n$ has not yet reached the global minimum, i.e.

$$v_n \; > \; \min_{y \in E} f(y) + \varepsilon$$

then with probability one $v_n$ is updated within a finite number of steps. Define

$$\alpha \; = \; \frac{1}{2}(v_n - \min_{y \in E} f(y) - \varepsilon) > 0$$

and

$$A \; = \; \{z \in E | f(z) \leq \min_{y \in E} f(y) + \alpha\}.$$

Then

$$v_n - \varepsilon \; > \; \min_{y \in E} f(y) + \alpha \tag{8}$$

Now, by condition $(iv)$, $m(A) > 0$. If $v_n$ is not updated in a finite number of steps, i.e., $v_{n+1} = v_{n+2} = \ldots = v_{n+k}$, for any $k = 1, 2, \ldots$, then the self-adaptation mechanism will continually increase $\sigma$ until it reaches an upper bound of $\sigma = D$, which promotes ergodicity. With $\delta x$ normally distributed and $m(A) > 0$, we can find a positive constant $\beta$ such that for any $x \in E$

$$p(x) \; = \; \text{Prob}\{x + \delta x \in A\} \geq \beta > 0. \tag{9}$$

With a normally distributed $\delta x$, a value for $\beta$ is easily computed. For any $x \in E$,

$$\text{Prob}\{x + \delta x \in A\} = \frac{1}{\sqrt{(2\pi)^N \cdot D^2}} \int_A e^{-(\frac{\|y-x\|^2}{2D^2})} dy.$$

Since $\|y - x\| \leq D$ for any pair $x, y \in E$, we have

$$\frac{\|y-x\|^2}{2D^2} \leq \frac{1}{2}.$$

This means that

$$e^{-(\frac{\|y-x\|^2}{2D^2})} \geq e^{-(\frac{1}{2})} = \frac{1}{\sqrt{e}}.$$

Therefore,

$$\text{Prob}\{x + \delta x \in A\} \geq \frac{m(A)}{\sqrt{(2\pi)^N \cdot D^2 \cdot e}} = \beta > 0.$$

Now, in each iteration $n + k$, since $v_{n+k-1} = v_n$, the probability that no update occurs is

$$\begin{aligned} q(k) \; &= \; \Pi_{j=1}^{\lambda}[\text{Prob}\{f(x_{n+k-1} + \delta x_j) \geq v_{n+k-1} - \varepsilon\}] \\ &\leq \; \Pi_{j=1}^{\lambda}[\text{Prob}\{f(x_{n+k-1} + \delta x_j) > \min_{y \in E} f(y) + \alpha\}] \end{aligned}$$

where, for each $j$, $\delta x_j$ has distribution $N(0, D)$. It then follows that

$$\begin{aligned} q(k) \; &\leq \; \Pi_{j=1}^{\lambda}[\text{Prob}\{x_{n+k-1} + \delta x_j \in E \backslash A\}] \\ &= \; \Pi_{j=1}^{\lambda}[1 - \text{Prob}\{x_{n+k-1} + \delta x_j \in A\}] \\ &= \; [1 - p(x_{n+k-1})]^{\lambda} \\ &\leq \; (1 - \beta)^{\lambda} \end{aligned}$$

Since each iteration is independent, the probability of no update for $v_n$ after $k$ consecutive iterations equals $\Pi_{i=1}^{k} q(i)$. Now

$$\Pi_{i=1}^{k} q(i) \leq (1-\beta)^{k\lambda}$$

and

$$\lim_{k\to\infty} (1-\beta)^{k\lambda} = 0,$$

which implies that

$$\lim_{k\to\infty} \Pi_{i=1}^{k} q(i) = 0.$$

Hence, if $v_n$ has not reached the global minimum, then within a finite number of steps an update of $v_n$ is guaranteed to occur. □

Two points are worth observing. First, modifications to the self-adaptation parameter $\sigma$ are tied to tracking sequence updates rather than to the number of successful mutations directly. This may appear counterintuitive vis-a-vis the original 1/5-success rule because the tracking sequence plays no direct role in reproduction. However, the number of updates to the tracking sequence are proportional to the number of successful mutations so a philosophical consistency actually is maintained. Secondly, observe that Theorem 2 says locate—it does <u>not</u> say converge. This is a non-elitist algorithm and so the globally optimum individual would be discarded in the next generation. Nevertheless, the tracking sequence $\{v_1, v_2 \ldots\}$ would converge to the globally optimum value.

## 4 Numerical Results

In this section we compare our modified 1/5-success rule with the original 1/5-success rule and an ES with no self-adaptation. Both elitist and non-elitist versions are investigated. A number of objective functions are used as test cases (Yao et al., 1997). The global optimum is $f^* = 0.0$ in all cases.

1. Six-Hump Camel-Back Function

$$f_1(\vec{x}) = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4 + 1.0316285$$

   with a search region $E = [-5, 5]^2$.

2. Sphere Model

$$f_2(\vec{x}) = \sum_{i=1}^{12} x_i^2$$

   with a search region $E = [-100, 100]^{12}$.

3. Generalized Rastrigin's Function

$$f_3(\vec{x}) = \sum_{i=1}^{12} [x_i^2 - 10\cos(2\pi x_i) + 10]$$

   with a search region $E = [-5.12, 5.12]^{12}$.

4. Generalized Griewank Function

$$f_4(\vec{x}) = \frac{1}{4000}\sum_{i=1}^{12} x_i^2 - \prod_{i=1}^{12}\cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$$

with a search region $E = [-600, 600]^{12}$.

All three adaptation rules were tried with the elitist ES version, but the original 1/5-success rule was not tried with the non-elitist ES version because it was never intended to work with it. All trials were run for 2500 generations with the single strategy parameter initialized to 0.5—a value that appears to provide reasonable performance in the no adaptation method over all test functions. The percentage of successful mutations were recorded over intervals of 50 trials. In the non-elitist version, the single parent produced seven offspring. The means ($\langle f \rangle$) and standard deviations ($\sigma_f$) are averaged over 50 runs. An initial population was randomly generated and stored for use in all subsequent runs. The results are summarized in the following tables:

**Elitist Version**:

| | Without Adaptation | | 1/5-Success Rule | | Modified 1/5-Success Rule | |
|---|---|---|---|---|---|---|
| Functions | $\langle f \rangle$ | $\sigma_f$ | $\langle f \rangle$ | $\sigma_f$ | $\langle f \rangle$ | $\sigma_f$ |
| $f_1$ | $1.19 \times 10^{-3}$ | $1.42 \times 10^{-4}$ | $5.64 \times 10^{-1}$ | $1.47 \times 10^{-1}$ | $6.29 \times 10^{-3}$ | $7.46 \times 10^{-4}$ |
| $f_2$ | $6.65 \times 10^{-1}$ | $2.10 \times 10^{-2}$ | $4.53 \times 10^{3}$ | $2.90 \times 10^{2}$ | 9.14 | $3.11 \times 10^{-1}$ |
| $f_3$ | $9.78 \times 10^{1}$ | 2.20 | $9.79 \times 10^{1}$ | 3.35 | $8.48 \times 10^{1}$ | 3.05 |
| $f_4$ | $1.37 \times 10^{2}$ | 5.85 | $4.67 \times 10^{1}$ | 2.36 | $8.62 \times 10^{-1}$ | $4.69 \times 10^{-2}$ |

**Non-Elitist Version**:

| | Without Adaptation | | Modified 1/5-Success Rule | |
|---|---|---|---|---|
| Functions | $\langle f \rangle$ | $\sigma_f$ | $\langle f \rangle$ | $\sigma_f$ |
| $f_1$ | $1.55\times10^{-3}$ | $2.46\times10^{-4}$ | $2.17\times10^{-3}$ | $4.68\times10^{-4}$ |
| $f_2$ | 4.13 | $1.35\times10^{-1}$ | $6.07\times10^{1}$ | 2.38 |
| $f_3$ | $8.45\times10^{1}$ | 1.17 | $8.55\times10^{1}$ | 1.02 |
| $f_4$ | $6.01 \times 10^{1}$ | 4.96 | $8.66\times10^{-1}$ | $1.16 \times 10^{-2}$ |

In functions $f_1, f_3$ and $f_4$ the elitist version of the modified 1/5-success rule compares favorably with both the original 1/5-success rule and stochastic search without adaptation. All these functions has multiple local optimal points. On the other hand, as expected, stochastic search method with variable $\sigma$ does not do as well for convex functions with only a single minimum. This is reflected in the test results for $f_2$. Conversely, the more rugged the fitness landscape, the better the modified 1/5-success rule performs compared to others. In particular, both the elitist and non-elitist versions with the modified 1/5-success rule outperformed the other search algorithms in function $f_4$, which is multi-modal with many local optima.

## 5 Conclusion

We have proposed an ES algorithm with self-adaptation that uses a modified 1/5-success rule. Formal proofs have been included to show that our adaptation method permits an ES to converge to the optimum in the elitist case with probability one; it locates the globally optimum solution with probability one in the non-elitist case. For a number of benchmark functions, numerical results show that our modified 1/5-success rule compares favorably with both a non-adapted ES and an ES adapted by the original 1/5-success rule.

## Acknowledgment

The authors thank Dr. Günter Rudolph for his helpful comments and review of the proofs. Q. J. Zhu was supported in part by National Science Foundation grant DMS-9704203.

## References

Bäck, T. (1996). *Evolutionary Algorithms in Theory and Practice*, Oxford University Press, NY.

Fogel, D.B. (1994). Asymptotic convergence properties of genetic algorithms and evolutionary programming: analysis and experiments. *Cyber. & Sys.*, **25**:389-407.

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

Rudolph, G. (1999). Self-adaptation and global convergence: a counter-example. *Proc. of Cong. on Evol. Comp. '99*, 646-651.

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

Reddy, B. (1998). *Introductory Functional Analysis*, Springer, New York.

Yao, X., Liu, Y. and Lin, G. (1997). Evolutionary Programming Made Faster. *IEEE Trans. Evol. Comput.*, **3**: 82-102