Analysis of convergence of an evolutionary algorithm with self-adaptation using a stochastic Lyapunov function

Mikhail A. Semenov* and **Dmitri A. Terkel** *mikhail.semenov@bbsrc.ac.uk
Rothamsted Research, Harpenden, Herts, AL5 2JQ, United Kingdom

Abstract

This paper analyses the convergence of evolutionary algorithms using a technique which is based on a stochastic Lyapunov function and developed within the martingale theory. This technique is used to investigate the convergence of a simple evolutionary algorithm with self-adaptation, which contains two types of parameters: fitness parameters, belonging to the domain of the objective function; and control parameters, responsible for the variation of fitness parameters. Although both parameters mutate randomly and independently, they converge to the "optimum" due to the direct (for fitness parameters) and indirect (for control parameters) selection. We show that the convergence velocity of the evolutionary algorithm with self-adaptation is asymptotically exponential, similar to the velocity of the optimal deterministic algorithm on the class of unimodal functions. Although some martingale inequalities have not be proved analytically, they have been numerically validated with 0.999 confidence using Monte-Carlo simulations.

Keywords

Self-adaptation, Stochastic Lyapunov function, Convergence, Martingale.

1 Evolutionary algorithm with self-adaptation

Evolutionary algorithms with self-adaptation represent a relatively new development in evolutionary computation (Beyer, 1995; Schwefel, 1981; 1995; Schwefel and Rudolph, 1995; Semenov and Terkel, 1984; 1985). A recent review of self-adaptation can be found in (Back, 1998; Eiben et al., 1999). Both reviews indicate that there have been few attempts to theoretically analyze the convergence of evolutionary algorithms with self-adaptation. In this paper we develop a technique based on the stochastic Lyapunov function for the analysis of convergence and convergence velocity of evolutionary algorithms.

For simplicity we consider a class of unimodal functions $K(\mathcal{R}) = \{f : \mathcal{R} \to \mathcal{R}\}$. We use an evolutionary algorithm with self-adaptation to solve the maximization problem for any $f \in K(\mathcal{R})$.

$$\max_{x \in \mathcal{R}} f(x)$$

Let us describe a simple evolutionary algorithm with self-adaptation. There is a finite population of N individuals $\left\{\left(x_{t,n}, x^*_{t,n}\right), n = 1, ..N\right\}$. Each individual at discrete time t is described by a pair of real numbers $\left(x_{t,n}, x^*_{t,n}\right)$, where $x_{t,n}$determines an individual fitness $f(x_{t,n})$, and $x^*_{t,n}$ controls the variability of $x_{t,n}$. Each step of the evolution includes the replication of individuals with random and independent mutations for both parameters $\left(x_{t,n}, x^*_{t,n}\right)$ and the selection of a new population from generated rivals according to the fitness function $f\left(x\right)$.

REPLICATION. At a discrete time t each individual from population $\left\{\left(x_{t,n}, x^*_{t,n}\right), n = 1, ..N\right\}$ produces M offspring. Each offspring $\left(x_{t,n,m}, x^*_{t,n,m}\right)$ originates from a parent according to the following rules:

$$\begin{cases} x_{t,n,m} = x_{t,n} + x^*_{t,n}\mathcal{N}\left(0,1\right)_{t,n,m} \\ x^*_{t,n,m} = x^*_{t,n}\exp(\vartheta_{t,n,m}) \end{cases}$$

where $\mathcal{N}\left(0,1\right)_{t,n,m}$ are independent normally distributed random variables with zero mean and a standard deviation of 1. $\vartheta_{t,n,m}$ are independent and uniformly distributed on the interval $[-a, a]$ random variable, where a is a constant.

SELECTION. From the $N \times M$ generated rivals $\left\{\left(x_{t,n,m}, x^*_{t,n,m}\right)\right\}$ N individuals are selected with the highest values of the fitness function $f\left(x_{t,n,m}\right)$. The selected individuals form a new population $\left\{\left(x_{t+1,n}, x^*_{t+1,n}\right), n = 1, ..N\right\}$ for the next time step $t+1$.

The algorithm described above can be considered as the $(N, N+M) - \sigma$SA - ES, σ-*self-adapting evolution strategy* algorithm (Beyer, 1995; Schwefel and Rudolph, 1995). The distinctive characteristic of self-adaptation is that the control parameters x^* are incorporated as a part of the evolving objects and they are allowed to evolve along with the main parameters x.

A self-adaptation mechanism, as described above, has a direct analogy in natural evolution. It has been assumed for a long time that the accuracy of gene replication is not perfect and that mutations, or abrupt hereditary changes, arise randomly through copy errors. Detailed studies of spontaneous mutations have shown that the real situation is more complex. It is generally accepted now that the mutation process is controlled by the genotype itself (Auerbach, 1976; Grant, 1977) and a high level of gene stability cannot be explained by the resistance of genes to changes alone, but by the presence in the organism of an active system which controls the process of mutation. The first example of this phenomenon, so-called mutator genes, was described in 1937 by Demerec (1937). Mutator genes control the mutation rate of the other genes, increasing or decreasing it by 1000-fold. There are numerous speculations on the possible role of the mutation mechanism in evolution (Karlin and McGregor, 1972; Kimura, 1967; Semenov and Terkel, 1985; Travis, 1992). Mutator genes do not contribute directly to the organism fitness, and they are not subject to direct selection. Thus it is important to find out whether the evolution of the mutator genes might be convergent or represented by a random walk (Grant, 1977). The most natural explanation of the evolution of mutators is indirect selection. An organism with a successfully altered mutator will have an evolutionary benefit in a subsequent generation by supplying a larger variety of genes, which directly affects individual fitness. Thus, co-evolution of both types of genes can be expected in the direction of increasing population adaptability to the environment and increasing the rate of adaptation. The evolution of the *Esherichia coli* mutator gene *mutT1* was examined experimentally (Gibson, 1970). Results showed that *mutT1* populations consistently outgrew *mut* $^+$ populations when the two were grown together in the same chemostat. A likely explanation was that the *mutT1* allele increased the variation in individual fitness more than the *mut*$^+$ allele because of a higher mutation rate in the *mutT1* populations, which supplied a larger variety of phenotypes for testing in the environment. It has been shown through an experiment (Sniegowski et al., 1997) and simulations (Taddei et al., 1997) that mutators can predominate in finite populations due to indirect selection.

2 Convergence Theory

The convergence of evolutionary algorithms, including algorithms with self-adaptation, has been investigated numerically in the past with few attempts to perform a theoretical analysis (Beyer, 1995; Rudolph, 1997a; b). One natural approach to investigating the convergence of stochastic processes, which describes the behavior of evolutionary algorithms, is an analogue of the method of the stochastic Lyapunov function in the theory of stability of stochastic processes (Kushner, 1967; 1984). If z_t is a stochastic process with values from an arbitrary state space $\mathcal{Z}$, then its Lyapunov function is a numerical function $V : \mathcal{Z} \to \mathcal{R}$, such that $V(z_t)$ decreases on average along the trajectories of the process, i.e. $V(z_t)$ is a supermartingale (Doob, 1990; Neveu, 1975; Williams, 2000). The convergence theorem for supermartingales allows one to conclude the convergence of the stochastic Lyapunov function, which in turn can be used to prove, under some additional assumptions, the convergence of the process z_t to an optimal state z_0. The aim of this section is to formulate a theorem on the convergence of the supermartingale $V_t \to \Gamma$, where Γ is a subset of $\mathcal{R}$ real numbers. Along with the main process (V_t) we will also consider the random process (V_t^*) characterizing the variation of the process (V_t) (Definition 2.1). Although the results can be formulated in terms of arbitrary processes, the special assumptions imposed on them are determined by the fact that (V_t) is the stochastic Lyapunov function of the evolutionary process.

In all that follows we denote by $(\Omega, \mathcal{A}, P)$ a probability space, where Ω is a sample space, $\mathcal{A}$ is a σ-algebra of measurable sets and P is a probability measure, and by $(\mathcal{A}_t, t \in \mathcal{N})$ an increasing family of sub-σ-algebras $\mathcal{A}_t \subset \mathcal{A}$. Let (V_t) and (V_t^*) be two stochastic processes adapted to the family $(\mathcal{A}_t)$, such that V_t takes values from the closed set $U \subset \mathrm{R}$. In the following definition we describe a condition that characterizes (V_t^*) as a process controlling the variation of a stochastic process (V_t).

DEFINITION 2.1 *The pair* (V_t, V_t^*) *of stochastic processes satisfies the A-condition on the set* $G \subset U$, *if* $\forall V \in G,\ V^* \in \mathcal{R}\ \exists \delta > 0$ *such that*

$$\inf_t \underset{\omega \in D}{ess \inf} P_\omega^{At} \{ \bigcup_{k=1}^{\infty} \{ \omega \in \Omega : | V_{t+k}(\omega) - V_t(\omega) | > \delta \} \} > 0$$

where $D = \{\omega \in \Omega : |V_t(\omega) - V| < \delta\} \cap \{\omega \in \Omega : V_t^*(\omega) > V^*\}$ *and* $P_\omega^{\mathrm{A}_t}$ *denotes the conditional probability with respect to the* σ*-algebra* $\mathcal{A}_t$ *at a point* $\omega \in \Omega$.

The *A-condition* means that V_t^*controls the variation of $V_t$in the following sense. If the value of V_t^* is larger than any arbitrary number V^*, then the conditional probability of the process V_t to leave a δ-neighborhood of the point $V \in G$ is separated from zero. The following proposition states that under the *A-condition* the convergence of V_t to the elements of the set G implies the convergence $V_t^* \to -\infty$.

PROPOSITION 2.1. *If the pair* (V_t, V_t^*) *of stochastic processes satisfies the A-condition on the set G, then*

$$\lim_t V_t^* \to -\infty$$

almost surely (a.s.) on $\left\{ \omega \in \Omega : \exists \lim_t V_t(\omega) \in G \right\}$

PROOF: Let us prove

$$P\left\{\lim_t V_t \in G\right\} = P\left\{\lim_t V_t \in G \cap \lim_t V_t^* = -\infty\right\}$$

or

$$P\left\{\lim_t V_t \in G \cap \lim_t V_t^* \neq -\infty\right\} = 0 \tag{2.1}$$

The event $\left\{\lim_t V_t^* \neq -\infty\right\}$is an enumerable union of events

$$\left\{\lim_t V_t^* \neq -\infty\right\} = \bigcup_{V^* \in \mathrm{Q}} \bigcap_{s \in \mathcal{N}} \bigcup_{t \geq s} \{V_t^* \geq V^*\}$$

To prove (2.1) it is sufficient to show that $\forall\, V^* \in \mathcal{Q}$ the equality (2.2) holds

$$P\left\{\lim_t V_t \in G \cap \left\{\bigcap_{s \in \mathcal{N}} \bigcup_{t \geq s} \{V_t^* \geq V^*\}\right\}\right\} = 0 \tag{2.2}$$

For any $V \in G$ and $V^* \in \mathcal{Q}$ we select $\delta = \delta(V, V^*)$ according to the *A-condition*. From a set of intervals $\left\{I_{V,\frac{\delta}{2}},\ V \in G\right\}$, where $I_{V,\frac{\delta}{2}} = \left\{x \in \mathcal{R} : |x - V| < \frac{\delta}{2}\right\}$, we select an enumerable set $\bigcup_{i \in \mathcal{N}} I_{V_i, \frac{\delta_i}{2}} \supset G$. The probability P is countably additive; therefore to proof (2.2) it is sufficient to demonstrate that $\forall \quad i \in \mathcal{N}$

$$P_i = P\left\{\left\{\lim_t V_t \in G \cap I_{V_i, \frac{\delta_i}{2}}\right\} \cap \left\{\bigcap_{k \in \mathcal{N}} \bigcup_{t \geq k} \{V_t^* \geq V^*\}\right\}\right\} = 0 \tag{2.3}$$

It is true that $\left\{\lim_t V_t \in G \cap I_{V_i, \frac{\delta_i}{2}}\right\} \subset \bigcup_{s \in \mathcal{N}} \bigcap_{t > s} \{V_t \in I_{V_i, \delta_i}\}$, therefore P_i can be estimated as

$$P_i \leq \lim_s P\left\{\bigcap_{t > s} \{V_t \in I_{V_i, \delta_i}\} \cap \left\{\bigcap_{k \in \mathcal{N}} \bigcup_{t \geq k} \{V_t^* \geq V^*\}\right\}\right\}.$$

Let us show that if the *A-condition* is valid on G, then for any $s \in \mathcal{N}$ the probability

$$P\left\{\bigcap_{t > s} \{V_t \in I_{V_i, \delta_i}\} \cap \left\{\bigcap_{k \in \mathrm{N}} \bigcup_{t \geq k} \{V_t^* \geq V^*\}\right\}\right\} = 0 \tag{2.4}$$

Let us choose $\delta > 0$ and $p > 0$ based on the Definition 2.1 such that

$$P^{\mathcal{A}_t}\left\{\bigcup_{k \in \mathcal{N}} \{|V_{t+k} - V_t| > \delta\}\right\} \geq p\chi\left(D_{t,V,V^*,\delta}\right)$$

where χ is a characteristic function of $D_{t,V,V^*,\delta} = \{|V_t - V| < \delta\} \cap \{V_t^* > V^*\}$. A similar inequality is true for any finite stopping time τ, e.g.

$$P^{\mathcal{A}_t}\left\{\bigcup_{k \in \mathcal{N}} \{|V_{\tau+k} - V_\tau| > \delta\}\right\} \geq p\chi\left(D_{t,V,V^*,\delta}\right)$$

Let us fix $\varepsilon > 0$. For a stopping time τ let us define a stopping time $[\tau]_\varepsilon^+$ such as

$$[\tau]_\varepsilon^+|_{\{\tau=t\}} = k(t)$$

where $k(t) > t$ is such that the inequality

$$P^{\mathcal{A}_t}\left\{\bigcup\nolimits_{t\le j\le k(t)}\{|V_j - V_t| > \delta\}\right\} \ge \frac{p}{2}\chi\left(D\backslash\Omega_t\right)$$

holds and $P\left(\Omega_t\right) \le \varepsilon/2^t$. It follows from the definition of $[\tau]_\varepsilon^+$ that

$$P^{\mathcal{A}_t}\left\{\bigcup\nolimits_{\tau\le j\le\tau^+}\{|V_j - V_\tau| > \delta\}\right\} \ge \frac{p}{2}\chi\left(D\backslash\Omega_\tau\right)\chi\left\{\tau < \infty\right\}$$

where $\Omega_\tau = \bigcup\limits_t \Omega_t$ and $P\left(\Omega_\tau\right) < \varepsilon$.

Let us fix a sequence of $\varepsilon_i, i \in \mathcal{N}$. Let us generate a sequence of stopping times τ_i , $i \in$ N

$$\begin{aligned}&\tau_1\left(\omega\right) = \min\left\{t > n\ :\ \ V_t^*\left(\omega\right) > V^*\right\}\\ &\tau_i\left(\omega\right) = \min\left\{t > [\tau_{i-1}]_{\varepsilon_i}^+\left(\omega\right) :\ \ V_t^*\left(\omega\right) > V^*\right\}, \qquad i > 1\end{aligned}$$

Let us estimate the probability

$$P\left\{\bigcap_{t>n}\left\{V_t \in I_{V,\frac{\delta}{2}}\right\} \cap \left\{\bigcap_{k\in\mathcal{N}}\bigcup_{t\ge k}\{V_t^* > V^*\}\right\}\right\}$$

Note that $\bigcap\limits_{i\in\mathcal{N}}\{\tau_i < \infty\} = \left\{\bigcap\limits_{k\in\mathcal{N}}\bigcup\limits_{t\ge k}\{V_t^* > V^*\}\right\}$, therefore the following inequality holds

$$\begin{aligned}&P\left\{\bigcap_{n<t}\left\{V_t \in I_{V,\frac{\delta}{2}}\right\} \cap \left\{\bigcap_{k\in\mathcal{N}}\bigcup_{t\ge k}\{V_t^* > V^*\}\right\}\right\} \le\\ &\lim_i P\left\{\bigcap_{n\le t\le\tau_i}\left\{V_t \in I_{V,\frac{\delta}{2}}\right\}\bigcap\left\{\bigcap_i\{\tau_i < \infty\}\right\}\right\}\end{aligned}$$

Let $C = \bigcap\limits_{n<t\le\tau_i}\left\{|V_t - V| \le \frac{\delta}{2}\right\}\bigcap\left\{\bigcap\limits_i\{\tau_i < \infty\}\right\}$, using the definition of stopping times τ_i and $[\tau]_i^+$ we can infer that

$$\begin{aligned}&P\left\{\bigcap_{n<t\le\tau_i}\left\{V_t \in I_{V,\frac{\delta}{2}}\right\}\bigcap\left\{\bigcap_i\{\tau_i < \infty\}\right\}\right\} \le\\ &\int_C\left(1 - P^{\mathcal{A}_t}\left\{|V_t - V_{\tau_i}| > \delta,\, \tau_i \le t \le [\tau]_i^+\right\}\right)dP =\\ &\left[\int_{C\backslash\Omega_{\varepsilon_i}} dP + \int_{\Omega_{\varepsilon_i}} dP\right]\left(1 - P^{\mathcal{A}_t}\left\{|V_t - V_{\tau_i}| > \delta,\, \tau_i \le t \le [\tau]_i^+\right\}\right) \le\\ &\left(1 - \frac{p}{2}\right)P\left\{\bigcap_{n<t\le\tau_{i-1}}\left\{V_t \in I_{V,\frac{\delta}{2}}\right\}\bigcap\left\{\bigcap_i\{\tau_i < \infty\}\right\}\right\} + \varepsilon_i \le \left(1 - \frac{p}{2}\right)^i + \sum_{j=1}^i \varepsilon_j\end{aligned}$$

or finally

$$P\left\{\bigcap_{n\le t}\left\{V_t\in I_{V,\frac{\delta}{2}}\right\}\bigcap\left\{\bigcap_{i\in\mathcal{N}}\{\tau_i<\infty\}\right\}\right\}\le\sum_{j=1}^{\infty}\varepsilon_j$$

We can always choose $\sum\limits_{j=1}^{\infty}\varepsilon_j$ to be arbitrarily small, therefore (2.4) is true, which proves the proposition.

□

Under the second condition imposed on the pair (V_t, V_t^*), the random process (V_t^*) possesses the submartingale property, if (V_t) deviates insignificantly from a state $V \in G$ and the value of (V_t^*) is sufficiently small (inequality (1) in Definition 2.2).

DEFINITION 2.2 *The pair of stochastic processes* (V_t, V_t^*) *satisfies the B-condition on the set* $G \subset U$, *if* $\forall V \in G\, \exists\, \delta > 0,\, V^* \in \mathcal{R}$ *and* $c \in \mathrm{R}$ *such that for all t and for almost all* $\omega \in \{\omega \in \Omega : |V_t(\omega) - V| < \delta\} \cap \{\omega \in \Omega : V_t^*(\omega) < V^*\}$ *the inequalities*

$$(1)\; E^{\mathcal{A}_t}(V_{t+1}^*)(\omega) > V_t^*$$
$$(2)\; E^{\mathcal{A}_t}\left(V_{t+1}^*\chi\left\{V_{t+1}^*\ge V^*\right\}\right)(\omega)\le cP_\omega^{\mathcal{A}_t}\left\{V_{t+1}^*\ge V^*\right\}$$

hold.

THEOREM 2.1. *Let* V_t *be a supermartingale bounded from below and taking values from the closed set* $U \subset \mathcal{R}$ *and let* V_t^* *denote a random process adapted to the family of* σ*-algebras* $(\mathcal{A}_t)$. *If the pair* (V_t, V_t^*) *satisfies the A- and B-conditions on the set* $U\backslash\Gamma$, *where* $\Gamma \subset U$, *then* $\lim\limits_t V_t(\omega) \in \Gamma$ *exists a.s. If the A-condition is satisfied on the set* U, *then* $\lim\limits_t V_t^*(\omega) = -\infty\;\, a.s.$

PROOF: Let us analyze the properties of (V_t, V_t^*), resulting from the *B-condition* on G. For $V \in G$ let us select $\delta > 0, V^* \in \mathcal{R}$ according to the definition 2. Let us fix V, V^*, δ and $n \in \mathcal{N}$ and define a stopping time $T(\omega)$ as

$$T(\omega)\,=\,T_{V,V^*,\delta,n}(\omega)\,=\,\min\left\{i\in N:\,i\ge n\cap\{|V_i(\omega)-V|\,>\delta\cup V_i^*(\omega)>V^*\}\right\}$$

Let us show that the stopped process $V_{t\wedge T}^*$ is a submartingale for $t > n$, and $\sup\limits_t E(V_{t\wedge T}^{*+}) < \infty$.

$$E^{\mathcal{A}_t}(V_{t+1\wedge T}^*)\,=\,E^{\mathcal{A}_t}(V_{t+1\wedge T}^*\chi\,(T>t))\,+\,E^{\mathcal{A}_t}(V_{t+1\wedge T}^*\chi\,(T\le t))$$

The random variable $V_{t+1\wedge T}^*\chi\,(T \le t) = V_{t\wedge T}^*\chi\,(T \le t)$ is $\mathcal{A}_t$-measurable, therefore

$$E^{\mathcal{A}_t}(V_{t+1\wedge T}^*\chi\,(T\le t))\,=\,V_{t\wedge T}^*\chi\,(T\le t)\,.$$

To assess $E^{\mathcal{A}_t}(V_{t+1\wedge T}^*\chi\,(T > t))$ we use the submartingale inequality (1) from Definition 2.2, which is applicable on the set $\{T > t\}$

$$E^{\mathcal{A}_t}(V_{t+1}^*\chi\,(T>t))\,=\,E^{\mathcal{A}_t}(V_{t+1}^*)\chi\,(T>t)\ge V_t^*\chi\,(T>t)$$

Finally

$$E^{\mathcal{A}_t}(V^*_{t+1\wedge T}) \geq V^*_{t\wedge T}$$

Using inequality (2) from definition 2.2 the following upper bound can be obtained:

$$E^{\mathcal{A}_t}(V^+_{t\wedge T}) \leq \max(c, V^*) \tag{2.5}$$

According to Theorem 2.1 (V_t) is a supermartingale bounded from below, hence $\exists \lim_t V_t = V_\infty$ and $P(V_\infty(\omega) \in U) = 1$. To prove the first statement of Theorem 2.1, i.e. $\exists \lim_t V_t \in \Gamma$ a.s., it is sufficient to prove that $P(V_\infty(\omega) \in U\backslash\Gamma) = 0$. The second statement of the Theorem, i.e. $\lim_t V^*_t(\omega) = -\infty$, is a direct conclusion from Proposition 2.1.
Suppose the opposite is true, i.e.

$$P(V_\infty(\omega) \in U\backslash\Gamma) > 0 \tag{2.6}$$

We will prove below that this would result in $\sup_t E(V^{*+}_{t\wedge T}) = \infty$, which contradicts inequality (2.5), hence proving that $P(V_\infty(\omega) \in U\backslash\Gamma) = 0$.

The *A-condition* is valid on $U\backslash\Gamma$. If inequality (2.6) is true then using proposition 2.1 we conclude

$$P\left(\{V_\infty(\omega) \in U\backslash\Gamma\} \cap \left\{\lim_t V^*_t = -\infty\right\}\right) > 0 \tag{2.7}$$

The *B-condition* for (V_t, V^*_t) is true on $U\backslash\Gamma$, therefore $\forall \quad V \in U\backslash\Gamma$ we can select $\delta > 0$ according to definition 2. From a set of intervals $\left\{I_{V,\frac{\delta}{2}},\ V \in U\backslash\Gamma\right\} \supset U\backslash\Gamma$, let us select an enumerable sub-set $\left\{I_{V_i,\frac{\delta_i}{2}},\ i \in \mathcal{N}\right\} \supset$ U\Γ. P is countably additive, then

$$0 < P(V_\infty \in U\backslash\Gamma) \leq \sum_{i\in\mathcal{N}} P\left(V_\infty \in U\backslash\Gamma \cap I_{V_i,\frac{\delta_i}{2}}\right)$$

and we can select $m \in \mathcal{N}$ such that

$$0 < P\left(V_\infty \in U\backslash\Gamma \cap I_{V_m,\frac{\delta_m}{2}}\right)$$

or using inequality (2.7)

$$0 < P\left(\left\{V_\infty \in U\backslash\Gamma \cap I_{V_m,\frac{\delta_m}{2}}\right\} \cap \left\{\lim_t V^*_t = -\infty\right\}\right) \tag{2.8}$$

If $V_\infty = \lim_t V_t(\omega) \in U\backslash\Gamma \cap I_{V_m,\frac{\delta_m}{2}}$, then we can select *k*$(\omega)$ such that $V_t(\omega) \in I_{V_m,\delta_m} \forall \quad t > k(\omega)$ or

$$V_\infty = \lim_t V_t(\omega) \in U\backslash\Gamma \cap I_{V_m,\frac{\delta_m}{2}} \subset \bigcap_{t>k(\omega)} V_t(\omega) \in I_{V_m,\delta_m} \tag{2.9}$$

Using (2.8) and (2.9) we can select *k*∈ N such that

$$0 < P\left(\bigcap_{t>k} \{V_t(\omega) \in I_{V_m,\delta_m}\} \bigcap \left\{\lim_t V^*_t(\omega) = -\infty\right\}\right)$$

Using the *B-condition*, we can choose V^* for previously selected values of V_m and δ_m . Considering that

$$\left\{\lim_t V_t^* = -\infty\right\} = \bigcap_{V^* \in \mathcal{Q}} \bigcup_n \bigcap_{t>n} \{V_t^* < V^*\}$$

we can select $V_m^* < V^*$, where $V_m^* \in \mathcal{Q}$, and $n_m \in \mathcal{N}$ such that

$$0 < P\left(\bigcap_{t>k} \{V_t(\omega) \in I_{V_m,\delta_m}\} \bigcap_{t>n_m} \{V_t^* < V_m^*\} \bigcap \left\{\lim_t V_t^*(\omega) = -\infty\right\}\right) \tag{2.10}$$

Let

$$T(\omega) = T_{V_m, V_m^*, \delta_m, n_m}(\omega)$$

be a stopping time defined at the beginning of the proof. We show that

$$\sup_t E(V_{t\wedge T}^{*+}) = \infty$$

It is true that

$$E\left(V_{t\wedge T}^{*-}\right) \geq \int_{\{T=\infty \cap \lim V_t^* = -\infty\}} V_t^{*-} dP$$

Using inequality (2.10) and the definition of the stopping time T, we conclude that $\{T = \infty \cap \lim V_t^* = -\infty\} > 0$, and consequently

$$\int_{\{T=\infty \cap \lim V_t^* = -\infty\}} V_t^{*-} dP \to \infty$$

because $\lim_t V_{t\wedge T}^{*-} = \infty$ a.s. on $\{T = \infty \cap \lim V_t^* = -\infty\} > 0$. $V_{t\wedge T}^*$ is a submartingale, therefore $\sup_t E(V_{t\wedge T}^{*+}) = \infty$.

□

3 Convergence Velocity of Supermartingales

In this section we estimate the convergence velocity of supermartingales, which enable the use of a stochastic Lyapunov function in the assessment of an evolutionary process convergence velocity. Let (V_t) be a supermartingale. In proposition 3.1 we analyze the asymptotic behavior of (V_t) under the following restrictions: (1) V_t decreases on average each time by a fixed constant $a > 0$, (2) variation of V_t does not exceed on average $b > 0$.

PROPOSITION 3.1. *Let (V_t) be a supermartingale, $V_0 = 0$. If the following conditions hold*

$$\begin{aligned} &(1)\ \ E^{\mathcal{A}_t}\left(V_{t+1}\right) \leq V_t - a \\ &(2)\ \ E^{\mathcal{A}_t}\left(\left(V_{t+1} - E^{\mathcal{A}_t}\left(V_{t+1}\right)\right)^2\right) \leq b \end{aligned}$$

where $a > 0, b > 0$, then $\forall \varepsilon > 0$ the following inequality holds a.s.

$$V_t \leq -at + o\left(t^{0.5+\varepsilon}\right)$$

PROOF: Let us present a supermartingale V_t as a sum of martingale Y_t and an increasing process H_t

$$V_t = Y_t - H_t \tag{3.1}$$

where

$$\begin{array}{l} Y_0 = V_0,\ Y_{t+1} - Y_t = V_{t+1} - E^{\mathcal{A}_t}\left(V_{t+1}\right) \\ H_0 = 0,\ H_{t+1} - H_t = V_t - E^{\mathcal{A}_t}\left(V_{t+1}\right) \end{array}$$

Using inequality 1 from proposition 3.1 H_t can be estimated as

$$H_t = \sum_{i=1}^{t}\left(H_i - H_{i-1}\right) + H_0 \geq at \tag{3.2}$$

Let us show that a martingale $Y_t \in L^2$. Indeed

$$Y_t = \sum_{i=1}^{t-1}\left(V_{i+1} - E^{\mathcal{A}_t}\left(V_{i+1}\right)\right) + V_0$$

and according to inequality 2

$$E\left(\left(V_{t+1} - E^{\mathcal{A}_t}\left(V_{t+1}\right)\right)^2\right) = E\left(E^{\mathcal{A}_t}\left(\left(V_{t+1} - E^{\mathcal{A}_t}\left(V_{t+1}\right)\right)^2\right)\right) \leq b$$

Let K_t be an increasing process in the Doob decomposition of submartingale Y_t^2 (Doob, 1990). Let us evaluate K_t using the inequality 2

$$K_{t+1} - K_t = E^{\mathcal{A}_t}\left(Y_t^2\right) - Y_t^2 = E^{\mathcal{A}_t}\left(Y_{t+1} - Y_t\right)^2 = E^{\mathcal{A}_t}\left(V_{t+1} - E^{\mathcal{A}_t}\left(V_{t+1}\right)\right)^2 \leq b$$

therefore

$$K_{t+1} \leq \sum_{i=1}^{t}\left(K_i - K_{i-1}\right) + K_0 \leq bt \tag{3.3}$$

Using Proposition VII –**2**-4 from (Neveu, 1975) and inequality (3.3), we conclude that

$$Y_t = o\left(K_t^{0.5+\varepsilon}\right) = o\left(t^{0.5+\varepsilon}\right),\quad \forall\varepsilon > 0 \tag{3.4}$$

Replacing Y_t and H_t in (3.1) by (3.2) and (3.4) we obtain

$$V_t \leq -at + o\left(t^{0.5+\varepsilon}\right)$$

□

4 Convergence of the Evolutionary Algorithm with Self-Adaptation

In this section, we illustrate how to apply the developed techniques to analyze the convergence of the evolutionary algorithm with self-adaptation and evaluate its convergence velocity. To verify some martingale inequalities, required by Theorem 2.1 and Proposition 3.1, we used numerical Monte Carlo simulations. Although this approach cannot be considered as a rigorous mathematical proof, it has an important practical attractiveness. Typically evolutionary algorithms are too complex to be analyzed analytically. However most of them are implemented as computer programs, which allows a straightforward application of Monte Carlo simulations to verify martingale inequalities. By using theoretical results from Section 2 and 3, one can then deduce the convergence of the algorithm.

We apply theorem 2.1 to analyze the convergence of a modified version of the evolutionary algorithm with self-adaptation described in Section 1. The population consists of only one individual N = 1, which produces M offspring according to the following formulas:

$$\begin{aligned} x^*_{t,m} &= x^*_t \exp\left(\vartheta_{t,m}\right) \\ x_{t,m} &= x_t + x^*_{t,m}\xi_{t,m} \end{aligned} \tag{4.1}$$

where the random variables $\vartheta_{t,i}$ are independent and uniformly distributed on the interval [-2,2] and $\xi_{t,i}$ are independent and uniformly distributed on [-1,1]. From the M generated rivals $\{(x_{t.m}, x^*_{t.m}), m = 1, ..M\}$ only one is selected, which has maximum of $f(x_{t,m})$ and becomes the next state $\left(x_{t+1}, x^*_{t+1}\right)$, where $f(x) = -|x|$.

For $M \geq 3$ the convergence $x_t \to 0$ and $x^*_t \to 0$ takes place a.s. The following stochastic Lyapunov function was used

$$V_t = V(x_t, x^*_t) = \left[-\int f(x_{t+1})dP_{x_t,x^*_t}\right]^\alpha$$

where $0 < \alpha < 1$ and dP_{x,x^*} denotes the transition probability of the Markov process (x_t, x^*_t) . The process (V^*_t) is defined as $V^*_t = \ln(x^*_t)$, and we set $\Gamma = \{0\}$.

PROPOSITION 4.1. *The stochastic process*$(V_t,\, V^*_t)$ *defined for the evolutionary algorithm with self-adaptation* $(x_t,\, x^*_t)$ *as* $V_t = V(x_t, x^*_t) = \left[-\int F(x_{t+1})dP_{x_t,x^*_t}\right]^\alpha$ and $V^*_t = \ln(x^*_t)$ satisfies the *A-condition* on $\mathcal{R}$ and the *B-condition* on $\mathcal{R}\backslash\{0\}$. *Moreover*, (V_t) *is a supermartingale.*

PROOF: (based on Monte Carlo simulations)
1). To prove that the *A-condition* is satisfied a.s. on

$$D = \{(x, x^*)\ :\ |V(x, x^*)\ - V|\ <\ \delta \cap\ V^*(\,x, x^*)\ >\ V^*\,\}$$

we show that $\forall V \in G, V^* \in \mathcal{R} \exists \quad \delta > 0$ such that

$$\inf_{(x,x^*)\in D} P_{\,x,x^*}\ \{\,(x, x^*)\ :\ \left|\,V(x_+, x^*_+) -\, V(x, x^*)\,\right|\ > \delta\,\} > 0$$

where $\left(x_+,\, x^*_+\right)$ is the state of the process followed after $(x,\, x^*)$. Using the definition of the evolutionary process (4.1) we obtain

$$P_{x,x^*}\left\{x_+^* - x^* \geq x^*\right\} \geq P\left\{\vartheta \geq \ln 2\right\}^M \geq p \tag{4.2}$$

$$P_{x,x^*}\left\{(x_+ - x)\, sign\,(x) \geq {x^*}/{2}\right\} \geq \left(\frac{1}{8}\right)^M \geq p$$

where $p = \left(\frac{1}{8}\right)^{2M} > 0$. For fixed δ and V let us define a function $\rho :\ \mathrm{A}\,(\delta, V) \to \mathcal{R}$

$$\rho\,(x, x^*) = \min\left\{\max_{\alpha:(x,x^*+\alpha)\in \mathrm{A}(\delta,V)} \alpha,\ \max_{\beta:(x+\beta,x^*)\in \mathrm{A}(\delta,V)} \beta\right\}$$

where $\mathrm{A}\,(\delta, V) = \{|V\,(x, x^*) - V| \leq \delta\}$. Let us denote

$$\gamma\,(\delta, V) = \sup_{(x,x^*)\in \mathrm{A}(\delta,V)} \rho\,(x, x^*)\,.$$

Note that $\gamma\,(\delta, V) \to 0$ when $\delta \to 0$. For any $V \in \mathcal{R}\backslash\{0\}, V^* \in \mathrm{R}$ let us select δ such that $\gamma\,(2\delta, V) < {x^*}/{2} = \exp\left({V^*}/{2}\right)$. Then

$$\begin{aligned}
&\inf_{(x,x^*)\in \mathrm{A}(\delta,V)} P_{\,x,x^*}\ \{\,(x, x^*)\ :\ \left|\,V(x_+, x_+^*) - V(x, x^*)\,\right|\ > \delta\,\} \geq \\
&\inf_{(x,x^*)\in \mathrm{A}(\delta,V)} P_{\,x,x^*}\ \{\,|\,x_+ - x\,| > \gamma\,(2\delta, V) \cap \left|x_+^* - x^*\right| > \gamma\,(2\delta, V)\,\} \geq \\
&\inf_{(x,x^*)\in \mathrm{A}(\delta,V)} P_{\,x,x^*}\ \{\,|\,x_+ - x\,| > {x^*}/{2} \cap \left|x_+^* - x^*\right| > x^*\,\} \geq p > 0
\end{aligned}$$

2). To prove that the *B-condition* is satisfied on $\mathcal{R}\backslash\{0\}$, we have to show that

$$E_{x,x^*}\left(\ln\left(x_+^*\right)\right) \geq \ln\,(x^*) \tag{4.3}$$

and

$$E_{x,x^*}\left(\ln\left(x_+^*\right)\chi\left(\ln\left(x_+^*\right) \geq V^*\right)\right) \leq cP_{x,x^*}\left\{\ln\left(x_+^*\right) \geq V^*\right\} \tag{4.4}$$

The inequality (4.3) is true on $\left\{|x| \geq x^* e^2\right\}$for $M \geq 3$. To prove inequality (4.4) it is sufficient to choose $c = V^* + 2$, because $\ln\left(x_+^*\right) \leq \ln\,(x^*) + 2 \leq V^* + 2$.

3). The verification that $V_t = V\,(x_t, x_t^*)$ is a supermartingale was based on the numerical Monte Carlo simulations for $f(x_t) = -|x_t|$, α =0.5 and $M = 3$. It was shown numerically that $E_{x,x^*}\left(V\left(x_+, x_+^*\right)\right) \leq V\,(x, x^*)$ (Appendix 1).

□

COROLLARY 4.1 *If* (V_t, V_t^*) *is a stochastic process as defined in Proposition 4.1, then*

$$\lim_{t\to\infty} V_t = 0 \text{ and } \lim_{t\to\infty} V_t^* = -\infty\ a.s.$$

COROLLARY 4.2 *If* (x_t, x_t^*) *is an evolutionary process with self-adaptation as defined in Section 4, then*

$$\lim_{t\to\infty} x_t = 0 \text{ and } \lim_{t\to\infty} x_t^* = 0\ a.s.$$

Let us estimate the convergence velocity of the evolutionary algorithm with self-adaptation (x_t, x_t^*) described at the beginning of this section. We show that the evolutionary process with self-adaptation has an exponential convergence velocity. This result does not use theorem 2.1 and could be considered as an independent proof of convergence $(x_t, x_t^*) \to (0,0)$. The evaluation of the convergence velocity will be based on proposition 3.1. We selected a more complex stochastic Lyapunov function in this case

$$V(x, x^*) = \ln\left(E_{x,x^*}\left(|x_+|\right)\right) - k\ln\left(x^*\right) \tag{4.5}$$

PROPOSITION 4.2 *A stochastic process* $V_t = V(x_t, x_t^*)$*, defined by (4.5), is a supermartingale and* $\exists\, a > 0\,,\ b > 0$ *such that inequalities*

$$E_{x,x^*}(V) \le V(x, x^*) - a$$
$$E_{x,x^*}\left(\left[V\left(x_+, x_+^*\right) - E_{x,x^*}(V)\right]^2\right) \le b$$

hold.

PROOF: (based on Monte Carlo simulations)
1). It can be shown that

$$E_{x,x^*}\left(V\left(x_+, x_+^*\right)\right) - V(x, x^*) = E_{x/x^*,1}\left(V\left(x_+/x_+^*, 1\right)\right) - V(x/x^*, 1)$$

The function $s(y) = E_{y,1}(V(y_+, 1)) - V(y, 1)$ was investigated numerically using the Monte Carlo simulations (see Appendix 2). It was shown that $\exists\, a > 0\ :\ \forall y\ s(y) \le -a$.

2). The second inequality of the proposition 4.2 could be directly infer from the fact that both $E_{x,x^*}\left(\left(\ln\left(E_{x_+,x_+^*}|x|\right) - E_{x,x^*}\ln\left(E_{x_+,x_+^*}|x|\right)\right)^2\right)$, and $E_{x,x^*}\left(\left(\ln\left(x_+^*\right) - E_{x,x^*}\ln\left(x_+^*\right)\right)^2\right)$ are bounded, and the Cauchy-Schwarz integral inequality $\left[E(fg)\right]^2 \le \left[E\left(f^2\right)\right]\left[E\left(g^2\right)\right]$ holds.

□

COROLLARY 4.3 *The following inequality for the process* $V_t = V(x_t, x_t^*)$*, defined by (4.5)*

$$V_t \le -at + o\left(t^{0.5+\varepsilon}\right)$$

holds asymptotically a.s.

COROLLARY 4.4 The *evolutionary process with self-adaptation* (x_t, x_t^*) *converges to* $(0,0)$ *a.s.; moreover the following inequalities*

$$|x_t| \le \exp(-at)\,, \quad x_t^* \le \exp(-at)$$

hold asymptotically a.s.

PROOF: Using formula (4.5) and Corollary 4.3 we can conclude that the following inequality

$$E_{x_t,x_t^*}\left(|x_{t+1}|\right)/x_t^{*k} \leq \exp\left(-at + o\left(t^{1/2+\varepsilon}\right)\right) \tag{4.6}$$

holds asymptotically a.s. Taking into account that $\exists\gamma > 0 \; E_{x_t,x_t^*}\left(|x_{t+1}|\right) = x_t^* E_{x_t/x_t^*,1}\left(|x_{t+1}|\right) \geq \gamma$, we can transform (4.6) to

$$x_t^* \leq \exp\left(-at\right)\left[\gamma^{-1/1-k}\exp\left(-\frac{a}{1-k}t + \bar{o}\left(t^{1/2+\varepsilon}\right)\right)\right]$$

The expression in the square brackets is less than 1 for large t, therefore the inequality

$$x_t^* \leq \exp\left(-at\right)$$

holds asymptotically a.s. The second inequality can be proved similarly.

□

5 Concluding Remarks

1. In its early stage, evolutionary algorithms did not include control parameters as a part of the evolving object, but considered them as external fixed parameters. Very soon it was realized that in order to achieve optimal convergence these parameters should be altered in the process of evolution (Schwefel, 1995; Eiben et al., 1999). The control parameters were adjusted with time by using heuristic rules, which take into account information about the progress achieved. However, heuristic rules, which might be optimal for one optimization problem, might be inefficient or even fail to guarantee convergence for another problem. A logical step in the development of evolutionary algorithms was to include control parameters into the evolving objects and allow them to evolve along with the main parameters. The proof of convergence of evolutionary algorithms with self-adaptation is difficult, because control parameters are changed randomly and the selection does not affect their evolution directly.

2. The technique based on the stochastic Lyapunov function has been developed and can be used to analyze a wide range of evolutionary algorithms. Theorem 2.1 gives sufficient conditions for convergence, which can be verified for many existing evolutionary algorithms. Although a universal method for constructing Lyapunov functions for arbitrary stochastic processes is unknown, properties of the evolutionary processes suggest an almost regular way to construct such a function as a mathematical expectation of the objective function.

3. Evolutionary algorithms with self-adaptation can be considered as universal methods for optimum search and can be used for solving optimization problems of high complexity, where heuristic deterministic procedures are difficult to develop (Back et al., 2000). Universality of evolutionary algorithms with self-adaptation is achieved by allowing control parameters to evolve along with the main parameters (Semenov and Terkel, 1985). Although stochastic search algorithms in general are less efficient than deterministic algorithms, an evolutionary algorithm with self-adaptation has an exponential convergence velocity. The Fibonacci search, the optimal deterministic search algorithm (Appendix 3), has an exponential convergence velocity $|x_n - x_f| < \beta e^{-\alpha n}$ with $\alpha \approx 0.48$. The evolutionary algorithm with self-adaptation has $\alpha \approx 0.09$ for $N = 1$ and $M = 5$ (note that n is a number of times the function f was evaluated, therefore, $n = t \times N \times M$ for the evolutionary algorithm).

Acknowledgements

We are grateful to reviewers and the editor for their helpful comments. Rothamsted Research receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom.

Appendix 1

Let us show that

$$V_t = V(x_t, x_t^*) = \left[- \int f(x_{t+1}) dP_{x_t, x_t^*} \right]^\alpha$$

is supermartingale for $M = 3$ and α=0.5. Note that $E_{x,x^*}(V) = (x^*)^\alpha E_{x/x^*,1}(V)$, hence we need to verify a supermartingale inequality $E_{x,x^*}(V) - V(x,x^*) \leq 0$ only for $\{(x,1) : x \in \mathrm{R}\}$. Values of $g(x) = E_{x,1}(V) - V(x,1)$ were estimated numerically by Monte Carlo simulations $\tilde{g}(x) = \frac{1}{K} \sum_{i=1}^{K} V(x_i, x_i^*) - V(x,1)$. The sampling value K was chosen large enough to provide that the standard error σ_g of $g(x)$ satisfies $\tilde{\sigma_g}(x) \leq 0.025 |\tilde{g}(x)|$, which guarantees that the martingale inequality is valid with probability of greater than 0.999. The results of calculation of $\tilde{g}(x) = E_{x,1}(V) - V(x,1)$ are presented in Figure 1. Note, that $g(x) \sim -x^{\alpha-1}$. The convergence of the evolutionary algorithm for $M > 3$ is a direct corollary of the case for $M = 3$.

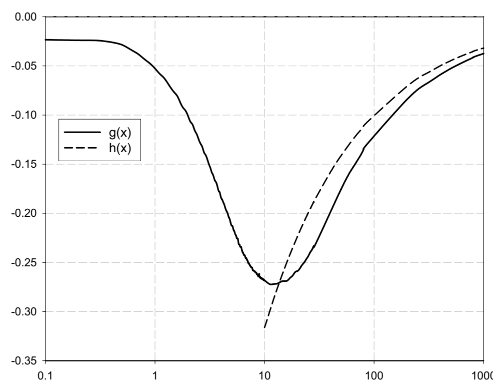

Figure 1. Functions $g(x) = E_{x,1}(V) - V(x,1)$ and $h(x) = -x^{\alpha-1}$, where $V(x,x^*) = \left[- \int f(x_+) dP_{x,x^*} \right]^\alpha$, were calculated using the Monte Carlo simulations for $M = 3$ and $\alpha = 0.5$ A logarithmic scale is used for the x-axis.

Appendix 2

Let us show that $\exists\, a > 0 : s(x) \leq -a$. Because of apparent difficulties in evaluating this function analytically, we evaluate it numerically, using Monte Carlo simulations. A similar procedure (as in appendix 1) was used for selecting parameters of Monte Carlo simulations , which guarantee with at least 0.999 probability that the inequality is true. Let us decompose the function $s(x) = E_{x,1}\left(V\left(x_+, x_+^*\right)\right) - V(x,1) = v_1(x) + v_2(x)$

where $v_1(x) = E_{x,1}\left(\ln\left(E_{x_+,x_+^*}|x_{++}|\right)\right) - \ln\left(E_{x,1}|x_+|\right)$ and $v_2(x) = -k\,E_{x,1}\ln\left(x_+^*\right)$. Note, that $E_{x,x^*}|x_+| = x^* E_{x/x^*,1}|x_+|$ and $E_{x,1}|x_+| = E_{e^2,1}|x_+| + x - e^2 \quad \forall x \geq e^2$ Also note, that $E_{x,1}\ln\left(x_+^*\right) = const > 0 \;\; \forall\, x \geq e^2$.

The results of Monte Carlo calculations are presented in Figure 2 for $M = 3$ and $k = 0.1$ The exponential convergence velocity of the evolutionary algorithm for $M > 3$ is a direct corollary of the case for $M = 3$.

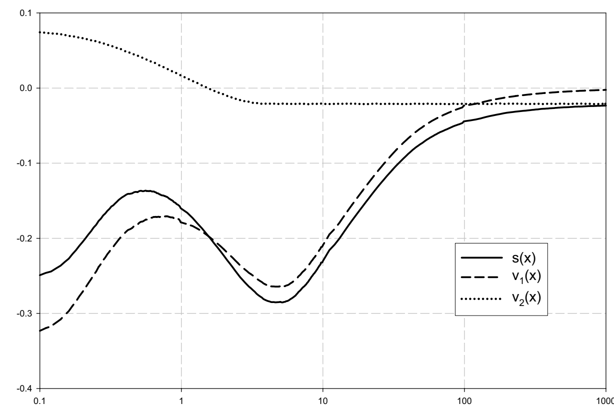

Figure 2. Functions $s(x), v_1(x)$ and $v_2(x)$ were calculated using Monte Carlo simulations for $M = 3$ and $k = 0.1$. A logarithmic scale is used for the x-axis.

Appendix 3

Optimal Deterministic Search Algorithm

Let $K[a,b]$ be a class of unimodal functions $f : [a,b] \to \mathcal{R}$. Let P^n be a set of n-point sequential deterministic algorithms $\{p_n\}$. An n-point algorithm p_n searches for the maximum of a function $f \in K[a,b]$ by sequentially selecting x_k, based on previous calculation of the function $f(x_1), \ldots, f(x_{k-1})$, where $k \leq n$. An algorithm p_n searches for the maximum trying to minimize the number of evaluations of the function f. For any $f \in K[a,b]$ and any p_n let the error of the algorithm p_n on a function f be defined as

$$\delta(p_n, f) = |x_n - x_f|$$

where x_f is a value where the function f has maximum $f(x_f) = \max_{x\in[a,b]} f(x)$. A guaranteed error of the algorithm p_n on the class of functions $K[a,b]$ is defined as

$$\Lambda(p_n) = \sup_{f\in K[a,b]} \delta(p_n, f)$$

An algorithm p_{opt} is an optimal n-point sequential deterministic algorithm, if it has the minimum guaranteed error compared with all other algorithms from P^n

$$\Lambda\left(p_{opt}\right)=\inf_{p_n\in P^n}\Lambda\left(p_n\right)$$

We denote $\gamma\left(n\right)=\inf\limits_{p_n\in P^n}\Lambda\left(p_n\right)$

THEOREM (Vasilev, 1980) *Fibonacci search* Φ_n *is the only optimal algorithm on the class* $K\left[a,b\right]$ *with*

$$\gamma\left(n\right)=\Lambda\left(\Phi_n\right)=\frac{\left(b-a\right)}{F_{n+2}}$$

where $F_n=\left[\left(\left(1+\sqrt{5}\right)/2\right)^n-\left(\left(1-\sqrt{5}\right)/2\right)^n\right]\Big/\sqrt{5}$ *is a Fibonacci number.*

References

Auerbach, C. (1976). *Mutation Research*, Chapman and Hall, London.

Bäck, Th. (1998). An Overview of Parameter Control Methods by Self-Adaptation in Evolutionary Algorithms. *Fundamenta Informaticae*, **35**:51-66.

Bäck, Th., Fogel, D. B., and Michalewicz, Z. (2000). *Evolutionary Computation 1. Basic Algorithms and Operators*, IOP Publishing Ltd, Bristol.

Beyer, H. G. (1995). Toward a Theory of Evolution Strategies: Self-adaptation. *Evolutionary Computation*, **3**(3):311-348.

Demerec, M. (1937). Frequency of Spontaneous Mutation in Certain Stocks of Drosophila Melanogaster. *Genetics*, **22**:469-478.

Doob, J. L. (1990). *Stochastic Processes*, John Wiley & Sons Inc, London.

Eiben, A. E., Hinterding, R., and Michalewicz, Z. (1999). Parameter Control in Evolutionary Algorithms. *IEEE Transaction on Evolutionary Computation*, **3**(2):124-141.

Gibson, T. C., Scheppe, M.L. & Cox, E.C. (1970). Fitness of an *Escherichia coli* Mutator Gene. *Science*, **169**:686-688.

Grant, V. (1977). *Organismic Evolution*, W.H. Freeman and Company, San Francisco.

Karlin, S., and McGregor, J. (1972). The Evolutionary Development of Modifier Genes. *Proc. Nat. Acad. Sci. USA*, **69**(12):3611-3614.

Kimura, M. (1967). Evolutionary Adjustment of Spontaneous Mutation Rates. *Genetical Research*, **9**(1):25-34.

Kushner, H. J. (1967). *Stochastic Stability and Control*, Academic Press.

Kushner, H. J. (1984). *Approximation and Weak Convergence Methods for Random Processes, with Application to Stochastic System Theory*, The MIT Press, Massachusetts.

Neveu, J. (1975). *Discrete-Parameter Martingales*, Noth-Holland Publishing Company, Amsterdam.

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

Rudolph, G. (1997b). Convergence Rates of Evolutionary Algorithms for a Class of Convex Objective Functions. *Control and Cybernetics*, **26**(3):375-390.

Schwefel, H. P. (1981, 1995). *Evolution and Optimum Seeking*, Willey, New York.

Schwefel, H.-P., and Rudolph, G. (1995). Contemporary Evolution Strategies. Advances in Artificial Life, F. M. e. al, ed., Springer, Berlin, 893-907.

Semenov, M. A., and Terkel, D. A. (1984). On the Convergence Set of Stochastic Lyapunov Function. *Am.Math.Society: Sov.Math.Dokl*, **30**(2):376-378.

Semenov, M. A., and Terkel, D. A. (1985). On the Evolution of Hereditary Variability Mechanisms by Means of Indirect Effect of Selection. *General Biology*, **XLVI**(2):271-278.

Sniegowski, P. D., Gerrish, P. J., and Lenski, R. E. (1997). Evolution of High Mutation Rates in Experimental Populations of *E.coli*. *Nature*, **387**:703-705.

Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P. H., and Godelle, B. (1997). Role of Mutator Appeps in Adaptive Evolution. *Nature*, **387**:700-702.

Travis, J. (1992). Possible Evolutionary Role Explored for ”Jumping Genes”. *Science*, **257**:884-885.

Vasilev, F. P. (1980). *Computational Methods for Optimisation Problems*, Nauka, Moscow.

Williams, D. (2000). *Probability with Martingales*, Cambridge University Press, Cambridge.