Convergence velocity of an evolutionary algorithm with self-adaptation

Mikhail A. Semenov

IACR-Rothamsted
Harpenden, Hertfordshire
AL5 2JQ, United Kingdom

Abstract

A stochastic Lyapunov function was used to assess the convergence velocity of a simple evolutionary algorithm with self-adaptation, which searches for a maximum of a "fitness" function. This algorithm uses two types of parameters: "fitness" parameters belonging to the domain of the function, and strategy parameters, which control changes of fitness parameters. It was shown that the convergence velocity of the evolutionary algorithm with self-adaptation is exponential, similar to the convergence velocity of the optimal deterministic algorithm, the Fibonacci search, on the class of unimodal functions.

1 OPTIMAL DETERMINISTIC SEARCH ALGORITHM

Let $K[a,b]$ be a class of unimodal functions $f:[a,b]\to\mathbb{R}$. Let P^n be a set of *n*-point sequential deterministic algorithms $\{p_n\}$. A *n*-point algorithm p_n searches for maximum of a function $f\in K[a,b]$ by sequentially selecting x_k, based on calculation of values $f(x_1),\dots,f(x_{k-1})$, where $k\le n$. For any $f\in K[a,b]$ and $\forall\, p_n$ let the error of an algorithm p_n on a function f be defined as

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

where x_f is a value where the function f has maximum $f(x_f)=\max_{y\in[a,b]} f(y)$. 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

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

We denote $\gamma(n)=\inf_{p_n\in P^n}\Lambda(p_n)$

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

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

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.

2 CONVERGENCE VELOCITY OF SUPERMARTINGALES

We use the stochastic Lyapunov function to assess the convergence velocity of the evolutionary algorithm with self-adaptation (Kushner, 1967; Semenov and Terkel, 1984; Semenov, 2001). If X_t is a stochastic process with values from an arbitrary state space X, then its stochastic Lyapunov function is a numerical function $V(X_t)$ decreasing on average along the trajectories of the process X_t, i.e. it is a super-martingale (Doob, 1990; Neveu, 1964; Williams, 2000). The convergence velocity of a super-martingale allows the assessment of the convergence velocity of the stochastic Lyapunov function, which in turn can be used to assess a convergence velocity of a stochastic process itself. The aim of this section is the Proposition 2.1 , which estimates the convergence velocity of a super-martingale

Let us fix a probability space (Ω, A, P), where Ω denotes the sample space, A is a σ-algebra of measurable sets and P is a probability measure, provided with an increasing family of σ-algebras A_t (t is the discrete time). Let (V_t) be a supermartingale adapted to the family (A_t). In Proposition 2.1 we analyse the asymptotic behaviour of (V_t) under the following restrictions: (1) V_t decreases on average each time by a fixed constant $a > 0$, (2) the variation of V_t does not exceed on average $b > 0$.

PROPOSITION 2.1. *Let (V_t) be a supermartingale, $V_0 = 0$. If the following conditions hold*

1. $E^{A_t}(V_{t+1}) \le V_t - a$
2. $E^{A_t}\left(\left(V_{t+1} - E^{A_t}(V_{t+1})\right)^2\right) \le b$

where $a>0$, $b>0$, then $\forall \varepsilon>0$ the inequality holds asymptotically almost everywhere

$$V_t \le -at + \overline{o}(t^{1/2+\varepsilon})$$

where $E^{A_t}(V_{t+1})$ is a conditional expectation with respect to a σ-algebra A_t.

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

(2.1) $\quad V_t = Y_t - H_t$

where

$$Y_0 = V_0,\ Y_{t+1} - Y_t = V_{t+1} - E^{A_t}(V_{t+1})$$
$$H_0 = 0,\ H_{t+1} - H_t = V_t - E^{A_t}(V_{t+1})$$

Using inequality 1 from Proposition 2.1, H_t can be assessed as

(2.2) $\quad H_t = \sum_{i=1}^{t}(H_i - H_{i-1}) + H_0 \ge at$

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

$$Y_t = \sum_{i=1}^{t}\left(V_i - E^{A_{i-1}}(V_i)\right) + V_0$$

and according to Inequality 2

$$E\left(\left(V_{t+1} - E^{A_t}(V_{t+1})\right)^2\right) = E\left(E^{A_t}\left(\left(V_{t+1} - E^{A_t}(V_{t+1})\right)^2\right)\right) \le b$$

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

$$K_{t+1} - K_t = E^{A_t}(Y_{t+1}^2) - Y_t^2 = E^{A_t}(Y_{t+1} - Y_t)^2 =$$
$$E^{A_t}\left(V_{t+1} - E^{A_t}(V_{t+1})\right)^2 \le b$$

therefore

(2.3) $\quad K_t \le \sum_{i=1}^{t}(K_i - K_{i-1}) + K_0 \le bt$

According to (Neveu, 1964)

(2.4) $\quad Y_t = \overline{o}\left(K_t^{1/2+\varepsilon}\right)$ a.e. on $\{K_\infty = \infty\}$

where $K_\infty = \lim_{t\to\infty} K_t$ and by definition

$g(t) = \overline{o}(h(t))$, if $\lim_{t\to\infty}\frac{g(t)}{h(t)} \to 0$. Using inequality (2.3) and (2.4) $\forall \varepsilon > 0$

(2.5) $\quad Y_t = \overline{o}\left(t^{1/2+\varepsilon}\right)$

Replacing Y_t and H_t in (2.1) by (2.2) and (2.5) we obtain

$$V_t \le -at + \overline{o}\left(t^{1/2+\varepsilon}\right)$$

□

3 CONVERGENCE VELOCITY OF EVOLUTIONARY ALGORITHM WITH SELF-ADAPTATION

We now apply Proposition 2.1 to prove the convergence of a simple evolutionary algorithm with self-adaptation (Semenov, 2001; Semenov and Terkel, 1985) (Back et al., 2000b; Beyer, 1995). The population consists of only one individual (x_t, x_t^*) at time t, which produces M offspring according to the following formulae:

(3.1)
$$\begin{cases} x_{t,i}^* = x_t^* \exp(\vartheta_{t,i}) \\ x_{t,i} = x_t + x_{t,i}^* \xi_{t,i} \end{cases}$$

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 $\{(x_{t,i}, x_{t,i}^*), i = 1,..M\}$ only one is selected, which has a maximum $f(x_{t,i})$ and it becomes the next state (x_{t+1}, x_{t+1}^*), where $f(x) = -|x|$.

Let us estimate the convergence velocity of the evolutionary algorithm with self-adaptation $\left(x_t, x_t^*\right)$. Corollary 3.4 shows that the evolutionary algorithm with self-adaptation has an exponential convergence velocity. The convergence velocity will be estimated using Proposition 3.1. Let us construct a stochastic Lyapunov function for the process $\left(x_t, x_t^*\right)$ as

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

where $\left(x_+, x_+^*\right)$ is the state generated from $\left(x, x^*\right)$.

PROPOSITION 3.1 *A stochastic process* $V_t = V\left(x_t, x_t^*\right)$*, defined by (3.2), is a supermartingale and* $\exists a > 0,\ b > 0$ *such that inequalities*

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

hold. $E_{x,x^*}\left(V\left(x_+, x_+^*\right)\right)$ *is a conditional expectation.*

PROOF: 1). It can be shown that

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

The function $G(x) = E_{x,1}\left(V\left(x_+, x_+^*\right)\right) - V\left(x, 1\right)$ was investigated numerically (see Appendix) and it was shown that $\exists a > 0 : \forall x \in \mathbb{R}\ G(x) \le -a$.

2). The second inequality is a direct deduction from the fact that both

$$E_{x,x^*}\left(\left[\ln E_{x_+,x_+^*}\left(\left|x_{++}\right|\right) - E_{x,x^*}\left(\ln E_{x_+,x_+^*}\left(\left|x_+\right|\right)\right)\right]^2\right)$$

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

$$\left[E\left(fg\right)\right]^2 \le \left[E\left(f^2\right)\right]\left[E\left(g^2\right)\right].$$

□

COROLLARY 3.3 *The following inequality for the process* $V_t = V\left(x_t, x_t^*\right)$*, defined by (3.2)*

$$V_t \le -at + \overline{o}\left(t^{1/2+\varepsilon}\right)$$

holds asymptotically almost everywhere.

COROLLARY 3.4 *Evolutionary algorithm with self-adaptation* $\left(x_t, x_t^*\right)$ *converges to* $(0,0)$ *almost everywhere; moreover the following inequalities*

$$\left|x_t\right| \le \exp\left(-at\right), \qquad x_t^* \le \exp\left(-at\right)$$

hold asymptotically almost everywhere.

PROOF: Using formula (3.2) and Corollary 3.3 we can conclude that the following inequality

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

holds asymptotically almost everywhere. Taking into account that $\exists\gamma>0$ $E_{x_t,x_t^*}\left(\left|x_{t+1}\right|\right) = x_t^* E_{x_t/x_t^*,1}\left(\left|x_{t+1}\right|\right) \ge \gamma x_t^*$, we can transform (3.6) to

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

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

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

holds asymptotically almost everywhere. The second inequality can be proved similarly.

□

4 CONCLUDING REMARK

In Section 3 by assessing a convergence velocity of the evolutionary algorithm with self-adaptation, we automatically proved its convergence. This proof is independent from the technique described in (Semenov, 2001).

Evolutionary algorithms with self-adaptation can be considered as universal methods for optimum search and can be used for solving optimisation problems of high complexity where heuristic deterministic procedures are difficult to develop (Back et al., 2000a). Universality of evolutionary algorithms with self-adaptation is achieved by allowing control parameters evolve along with the "fitness" parameters by an indirect effect of selection (Semenov and Terkel, 1985). Although stochastic search algorithms in general are less effective than deterministic ones, an evolutionary algorithm with self-adaptation has an exponential convergence velocity, which results from the co-evolution of control and "fitness" parameters. The Fibonacci search, the optimal deterministic search algorithm (Section 1), has an exponential convergence velocity $\left| x_n - x_f \right| < \beta e^{-\alpha n}$ with $\alpha \approx 0.48$. The evolutionary algorithm with self-adaptation has $\alpha \approx 0.09$ (note that n is a number of times f was calculated, therefore, $n = t * M$ for the evolutionary algorithm).

Acknowledgements

I am eternally indebted to D.A. Terkel for his contribution. IACR receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom.

APPENDIX

Let us show that $\exists a > 0 : G(x) \le -a$. Because of apparent difficulties in dealing with this function analytically , we assess it numerically using the Monte-Carlo method. Let us decompose the function

$$G(x) = E_{x,1}\left(V\left(x_+, x_+^*\right)\right) - V(x,1) = V_1(x) + V_2(x)$$

where $V_1(x) = E_{x,1}\left(\ln\left(E_{x_+, x_t^*}\left|x_{++}\right|\right)\right) - \ln\left(E_{x,1}\left|x_+\right|\right)$

and $V_2(x) = -k\, E_{x,1} \ln\left(x_+^*\right)$. Note, that

$E_{x,x^*}\left|x_+\right| = x^* E_{x/x^*,1}\left|y_+\right|$ and

$E_{x,1}\left|x_+\right| = E_{e^2,1}\left|x_+\right| + x - e^2 \quad \forall x \ge e^2$

Also note, that $E_{x,1} \ln\left(x_+^*\right) = const > 0 \quad \forall x \ge e^2$.

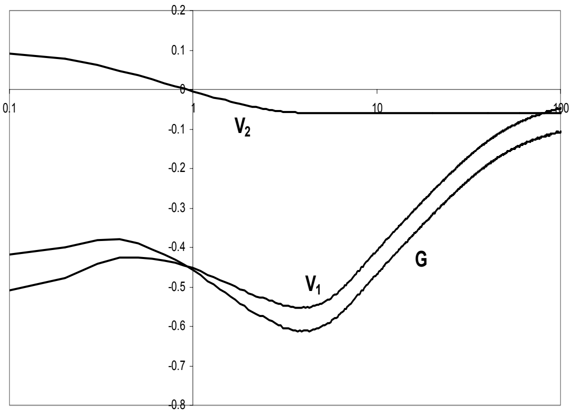

Figure 1. Functions V_1, V_2 and G were calculated by the Monte-Carlo methods for the number of offspring M=5. A logarithmic scale is used for the x-axis.

Results of Monte-Carlo calculations for the number of offspring M=5 are presented on Figure 1, $k = 0.1$

References

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

T. Back, D. B. Fogel, and Z. Michalewicz. (2000b). *Evolutionary Computation 2. Advanced Algorithms and Operators*, IOP Publishing Ltd, Bristol.

H. G. Beyer (1995). Toward a theory of evolution strategies: self-adaptation. *Evolutionary Computation*, **3**(3):311-348.

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

J. H. Kushner. (1967). *Stochastic stability and control*, Academic Press.

J. Neveu. (1964). *Bases mathematiques du calcul des probabilities*, Masson, Paris.

M. A. Semenov (2001). Analysis of evolutionary search with mutators using a stochastic Lyapunov function. *GECCO 2001*, San Francisco, 372-376.

M. A. Semenov, and D. A. Terkel (1984). On the convergence set of stochastic Lyapunov function. *Am.Math.Society: Sov.Math.Dokl*, **30**(2):376-378.

M. A. Semenov, and D. A. Terkel (1985). On the evolution of hereditary variability mechanisms by means of indirect effect of selection. *General Biology*, **XLVI**(2):271-278.

F. P. Vasilev. (1980). *Computational methods for optimisation problems*, Nauka, Moscow.

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