Lyapunov Design of a Simple Step-Size Adaptation Strategy Based on Success

Claudia R. Correa[1], Elizabeth F. Wanner[2], and Carlos M. Fonseca[3]

[1] Post-Graduate Program in Mathematical and Computational Modeling
CEFET-MG, Brazil
`claudinharmc@yahoo.com.br`
[2] Department of Computer Engineering, CEFET-MG, Brazil
`efwannerr@decom.cefetmg.br`
[3] CISUC, Department of Informatics Engineering
University of Coimbra, Portugal
`cmfonsec@dei.uc.pt`

Abstract. A simple success-based step-size adaptation rule for single-parent Evolution Strategies is formulated, and the setting of the corresponding parameters is considered. Theoretical convergence on the class of strictly unimodal functions of one variable that are symmetric around the optimum is investigated using a stochastic Lyapunov function method developed by Semenov and Terkel (2003) in the context of martingale theory. General expressions for the conditional expectations of the next values of step size and distance to the optimum under $(1 \overset{+}{,} \lambda)$-selection are analytically derived, and an appropriate Lyapunov function is constructed. Convergence rate upper bounds, as well as adaptation parameter values, are obtained through numerical optimization for increasing values of λ. By selecting the number of offspring that minimizes the bound on the convergence rate with respect to the number of function evaluations, all strategy parameter values result from the analysis.

Keywords: Step-size Adaptation, Evolution Strategy, Lyapunov Function Theory, Convergence Rate

1 Introduction

Evolution strategies (ESs) are a particular class of Evolutionary Algorithms (EAs) that have attracted a significant amount of attention in the last decades. ESs traditionally emphasize the use of selection and mutation as search operators, where the mutation operator consists in creating an offspring by adding a random vector to the current solution, or individual. Adaptive methods that dynamically rescale mutation step-length parameters have proved to be rather effective [1–3].

Convergence analysis of ESs with step-size adaptation has deserved due attention in the research community, but it has proved to be a difficult task. The present work extends the methodology proposed in [4] to the analysis of both

$(1, \lambda)$-ESs and $(1+\lambda)$-ESs with an arbitrary number of offspring, λ, and a success-based step-size adaptation rule. Based on the theoretical results developed in [5, 4], general expressions for the expectation of the next individual's step size and distance to the optimum are analytically derived for both types of selection and, using a Lyapunov function, upper bounds on the convergence rate on the class of strictly unimodal functions of one variable that are symmetric around the optimum are determined. Moreover, the number of offspring that minimizes the bound on the convergence rate with respect to the number of function evaluations can be easily determined as a by-product, providing a useful guideline for the selection of this remaining strategy parameter.

2   Literature Review

Convergence analysis is a major topic of research in evolutionary algorithms, which has been addressed mostly separately depending on whether the optimization problems of interest are continuous or discrete. Drift analysis is a state-of-art technique for the study of the expected hitting time of randomized search heuristics on discrete problems. The use of drift analysis combined with Markov chain theory in order to obtain lower and upper bounds on the expected hitting time on discrete search spaces is presented, for example, in [6–10]. Proofs of convergence of the $(1+1)$-EA applied to pseudo-Boolean linear functions can be found in [6, 9, 10]. In [6], the author combines drift analysis and Markov-chains for the first time to state bounds on the expected optimization time. The upper bound on the expected runtime of the algorithm was improved in [9], using multiplicative drift analysis. Later, in [10], the author improves the upper bound further, also using multiplicative drift analysis.

For continuous optimization problems, works dealing with convergence analysis of self-adaptive $(1, \lambda)$-ES applied to sphere functions using martingale theory include [11, 5, 4]. In [5], a stochastic Lyapunov function method is developed in the context of martingale theory, but Monte Carlo simulations are used to verify the convergence of a mutative self-adaptive $(1, \lambda)$-ES. Based on [5], a Lyapunov synthesis procedure for the adaptation parameters of a simple derandomized self-adaptive $(1, 2)$-ES is proposed in [4]. The methodology is based on particular candidate functions which become stochastic Lyapunov functions through suitable choices of the algorithm adaptation parameters. Considering the class of strictly unimodal functions of one variable which are symmetric around the optimum, and through the appropriate setting of the algorithm parameters, it is proved that both the decision variable and the mutation step-size converge almost surely to the optimum and to zero, respectively. Moreover, an upper bound on the rate of convergence is derived, and suitable values for the ES parameters are determined numerically.

The parallel between the drift analysis used in [9] and the method used in [5] and [4] is worth noting. With both methods, the behavior of an evolutionary algorithm is analyzed through an auxiliary function, which must be chosen in such a way that convergence of the algorithm on the true objective function

can be proved by verifying conditions on the auxiliary function only. In the case of drift analysis, the auxiliary function, also known as a potential function, is used to derive bounds on the expected runtime of the algorithm with respect to problem size. On the other hand, in [5] and [4], continuous optimization is considered and bounds on the convergence rate are provided instead.

3   The Proposed ES

The aim of the $(1 \overset{+}{,} \lambda)$-ES analyzed here is to minimize a real-valued function $f : \mathbb{R} \to \mathbb{R}$. A vector (x_t, d_t) describes an individual, the fitness of which depends on the value of the decision variable $x_t \in \mathbb{R}$. The step size which controls the variation is given by $d_t \in \mathbb{R}_+ \backslash \{0\}$. The evolution cycle consists of two steps: the mutation step, which creates λ offspring, $x_{i,t}$, $i = 1, \dots, \lambda$, and the selection step, which selects the next parent, x_{t+1}, from the offspring (and possibly the current parent, x_t, depending on the selection scheme), and determines the associated step size, d_{t+1}.

Consider the following $(1, \lambda)$-ES, where $\mu_{i,t}$, $i = 1, \dots, \lambda$, are random variables uniformly distributed in $[-1, 1]$, $0 < \alpha_f \le 1$, and $\alpha_s \ge 1$:

$$
\begin{aligned}
x_{t+1} &= F(x_t, d_t, \mu_{1,t}, \dots, \mu_{\lambda,t}) \\
&= \underset{x \in \{x_{i,t} = x_t + \mu_{i,t} d_t, \ i=1,\dots,\lambda\}}{\arg \min} f(x) \\[2mm]
d_{t+1} &= G(x_t, d_t, \mu_{1,t}, \dots, \mu_{\lambda,t}) \\
&= \begin{cases} \alpha_f \cdot d_t \ \text{if} \ f(x_{t+1}) > f(x_t) \\ \alpha_s \cdot d_t \ \text{if} \ f(x_{t+1}) \le f(x_t) \end{cases}
\end{aligned}
\tag{1}
$$

Note that, if f is any strictly unimodal function of one variable that is symmetric around the minimum, x_{t+1} is selected as the offspring $x_{i,t}$, $i = 1, \dots, \lambda$, that is closest to the minimum point of f, even if that minimum point is not known. In fact, the selection process is translation invariant, meaning that the minimum of f can be considered to be located at zero after an appropriate translation, without loss of generality. Therefore, any even function with such properties could be chosen. Moreover, observe that d_{t+1} depends indirectly on $(\mu_{1,t}, \mu_{2,t}, \dots, \mu_{\lambda,t})^T$ through a direct dependence on x_{t+1}. The search space $\mathbb{R}$ is equipped with the Borel σ-algebra.

The mutation step-size adaptation process can be described as follows: (i) if any offspring is at least as good as the parent, then the step size is increased, and (ii) if all offspring are worse than the parent, then the step-size is decreased. The only difference between this $(1, \lambda)$-ES and the corresponding $(1 + \lambda)$-ES is the selection scheme: in the latter, the best among the λ offspring *and* the parent itself, x_t, is selected to become x_{t+1}. As a consequence, the parent represents the all-time best individual at each iteration.

4 Theoretical Background

Consider a deterministic discrete dynamic system represented by $x_{t+1} = F(x_t)$ where $t \in \mathbb{N}$ is the time index, $x_t \in \mathbb{R}^n$ is the system state vector at time t, n is the number of system state variables, and $F : \mathbb{R}^n \to \mathbb{R}^n$ is a function in class C^p, with $p \geq 1$. An equilibrium point, x^*, of such a system is a point such that $F(x^*) = x^*$. A discrete version of the direct method of Lyapunov [12] essentially states that, if there is a function $V \in \mathcal{C}^1$ such that $V(x^*) = 0$ and $V(x) > 0$ $\forall x \neq x^*$, and, in addition, $V(x_{t+1}) < V(x_t)$ $\forall x_t \neq x^*$, then the equilibrium is uniformly asymptotically stable [13]. Function V is known as a Lyapunov function.[4]

Since the proposed $(1 \overset{+}{,} \lambda)$-ESs are stochastic processes, the above result is not directly applicable, but convergence may still be studied through the conditional expectation $E^{A_t}[(V(x_{t+1}, d_{t+1})] \triangleq E[V(x_{t+1}, d_{t+1})|x_1, \ldots, x_t, d_1, \ldots, d_t]$ of a suitable Lyapunov function $V(x_t, d_t)$ of the stochastic process (x_t, d_t). Theoretical results presented in [5] allow the convergence rate to be studied as well.

Consider that the inequality

$$|x_t| \leq \exp(-at), \tag{2}$$

holds asymptotically almost surely[5] for some scalar $a > 0$. Then, the convergence rate of x_t to the point of equilibrium $x^* = 0$ is $e^{-\bar{a}}$, where $\bar{a}$ is the supremum of the set of values of a for which (2) holds asymptotically almost surely. If only a lower bound a on the exponential decay constant $\bar{a}$ can be determined, with $0 < a \leq \bar{a}$, then $|x_t| \leq \exp(-\bar{a}t) \leq \exp(-at)$.

In [5], the asymptotic behavior of a supermartingle[6] V_t is analyzed under the following conditions: at each time step, V_t decreases on average by at least a constant $a > 0$, and the conditional variance of V_t is at most $b > 0$. The following result is proved therein:

Proposition 1. *Let V_t be a supermartingale and $V_0 = 0$. If the following conditions hold*

$$E^{A_t}(V_{t+1}) \leq V_t - a \tag{3}$$

$$E^{A_t}([V_{t+1} - E^{A_t}(V_{t+1})]^2) \leq b \tag{4}$$

where $a > 0$, $b > 0$, then $\forall \epsilon > 0$ the following inequality holds almost surely:

$$V_t \leq -at + o(t^{0.5+\epsilon}). \tag{5}$$

Proposition 1 is employed in the following result to establish the exponential convergence of the proposed $(1, \lambda)$-ES algorithm.

---

[4] Equivalently to $V(x^*) = 0$ and $V(x) > 0$ $\forall x \neq x^*$, one may require that $V(x) \to -\infty$ only when $x \to x^*$.

[5] An event holds asymptotically almost surely if it holds with probability $1 - o(1)$, i.e. the probability of success goes to 1 in the limit as $n \to \infty$ [14].

[6] A stochastic process V_t is said to be a *supermartingale* if $E^{A_t}(V_{t+1}) \leq V_t$.

Proposition 2. *Consider the stochastic process (x_t, d_t) defined in Eq. (1). If f is a strictly unimodal function of one variable that is symmetric around its minimum, then this process converges to $(0,0)$ almost surely, and the following inequalities*

$$|x_t| \leq \exp(-at) \quad and \quad d_t \leq \exp(-at) \tag{6}$$

hold asymptotically almost surely for some α_f, α_s, and $a > 0$.

Convergence of the corresponding $(1 + \lambda)$-ES can be stated (and proved) in the same way.

5   Convergence Rate Analysis

The following Lyapunov function is used in this work:

$$V_t = V(x_t, d_t) = \ln(|x_t| + w d_t) - k \ln(d_t) \tag{7}$$

where $w, k \in \mathbb{R}$, $w > 0$, and $0 \leq k < 1$. Observe that $V_t \to -\infty$ only when $x_t \to 0$ and $d_t \to 0$.

The proof of Proposition 2 consists of three steps:

1. Determining values for α_s, α_f, w, k and a such that inequality (3) is verified.
2. Proving that the conditional variance of V_t is bounded (inequality (4)).
3. Proving that, under Proposition 1, inequalities (6) hold asymptotically almost surely.

Since steps 2 and 3 depend only on the structure of the Lyapunov function and/or on the type of mutation and form of step-size adaptation considered, but not on whether step-size adaptation is based on success or on the length of the selected step, the corresponding proofs are identical to those presented in [4], where the same Lyapunov function and type of mutation are used together with a different two-point adaptation rule. Moreover, they apply to both $(1, \lambda)$ and $(1 + \lambda)$ selection.

In particular, step 2 can be accomplished by showing that V_{t+1} has support of bounded length for all x_t and d_t. By noting that the next value of the step size must satisfy $d_t \alpha_f \leq d_{t+1} \leq d_t \alpha_s$, is possible to show that both $\ln(d_{t+1})$ and $\ln(|x_{t+1}| + w d_{t+1})$ have supports of bounded length, and so does V_{t+1}.

Having this in mind, only step 1 is considered here for the proposed $(1 \stackrel{+}{,} \lambda)$-ESs. In order to derive (an upper bound on) $E^{A_t}(V_{t+1})$ for an arbitrary number of offspring, λ, expectations $E^{A_t}(|x_{t+1}|)$ and $E^{A_t}(d_{t+1})$ are calculated first for each type of selection. Actually, since the step-size adaptation scheme is exactly the same for $(1, \lambda)$ and $(1 + \lambda)$ selection, $E^{A_t}(d_{t+1})$ is the same in both cases. However, depending of the type of selection, $E^{A_t}(|x_{t+1}|)$ will result in different expressions.

The $(1, \lambda)$-ES is considered first. In determining $E^{A_t}(|x_{t+1}|)$, two cases must be considered: $0 \leq \frac{x_t}{d_t} < 1$, meaning that x_t is close to the origin, and $\frac{x_t}{d_t} \geq 1$, meaning that x_t is far from the origin.

In the case where x_t is close to the origin,

$$
\begin{aligned}
E^{A_t}(|x_{t+1}|) &= \tfrac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{0} -z(2z+2d_t)^{\lambda-1}\mathrm{d}z\right]\\
&+ \tfrac{\lambda}{(2d_t)^\lambda}\left[\int_{0}^{-x_t+d_t} z(2d_t-2z)^{\lambda-1}\mathrm{d}z\right]\\
&+ \tfrac{\lambda}{(2d_t)^\lambda}\left[\int_{-x_t+d_t}^{x_t+d_t} z(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right]\\
&= \left(\tfrac{1}{\lambda+1}\right)\left[\tfrac{x_t^{\lambda+1}}{d_t^{\lambda+1}}+1\right]d_t
\end{aligned}
\tag{8}
$$

and in the case where x_t is far from the origin,

$$
\begin{aligned}
E^{A_t}(|x_{t+1}|) &= \tfrac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{x_t+d_t} z(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right]\\
&= \left[\tfrac{x_t}{d_t}-\left(\tfrac{\lambda-1}{\lambda+1}\right)\right]d_t.
\end{aligned}
\tag{9}
$$

Considering $(1+\lambda)$-ES, three cases arise: $0 \leq \tfrac{x_t}{d_t} < \tfrac{1}{2}$, $\tfrac{1}{2} \leq \tfrac{x_t}{d_t} < 1$ and $\tfrac{x_t}{d_t} \geq 1$. In the case where $0 \leq \tfrac{x_t}{d_t} < \tfrac{1}{2}$,

$$
\begin{aligned}
E^{A_t}(|x_{t+1}|) &= \tfrac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{-x_t} x_t(2z+2d_t)^{\lambda-1}\mathrm{d}z\right]\\
&+ \tfrac{\lambda}{(2d_t)^\lambda}\left[\int_{-x_t}^{0} -z(2d_t+2z)^{\lambda-1}\mathrm{d}z\right]\\
&+ \tfrac{\lambda}{(2d_t)^\lambda}\left[\int_{0}^{x_t} z(2d_t-2z)^{\lambda-1}\mathrm{d}z\right]\\
&+ \tfrac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t}^{-x_t+d_t} x_t(2d_t-2z)^{\lambda-1}\mathrm{d}z\right]\\
&+ \tfrac{\lambda}{(2d_t)^\lambda}\left[\int_{-x_t+d_t}^{x_t+d_t} x_t(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right]\\
&= \tfrac{d_t}{\lambda+1}-\left(1-\tfrac{x_t}{d_t}\right)^\lambda\left(\tfrac{d_t-x_t}{\lambda+1}\right).
\end{aligned}
\tag{10}
$$

For $\tfrac{1}{2} \leq \tfrac{x_t}{d_t} < 1$,

$$
\begin{aligned}
E^{A_t}(|x_{t+1}|) &= \tfrac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{0} -z(2z+2d_t)^{\lambda-1}\mathrm{d}z\right]\\
&+ \tfrac{\lambda}{(2d_t)^\lambda}\left[\int_{0}^{-x_t+d_t} z(2d_t-2z)^{\lambda-1}\mathrm{d}z\right]\\
&+ \tfrac{\lambda}{(2d_t)^\lambda}\left[\int_{-x_t+d_t}^{x_t} z(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right]\\
&+ \tfrac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t}^{x_t+d_t} x_t(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right]\\
&= \left(\tfrac{x_t}{\lambda+1}\right)\left(\tfrac{x_t}{d_t}\right)^\lambda+\tfrac{d_t}{\lambda+1}-\tfrac{d_t}{2^\lambda(\lambda+1)}.
\end{aligned}
\tag{11}
$$

Finally, for $\tfrac{x_t}{d_t} \geq 1$,

$$
\begin{aligned}
E^{A_t}(|x_{t+1}|) &= \tfrac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{x_t} z(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right]\\
&+ \tfrac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t}^{x_t+d_t} x_t(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right]\\
&= x_t+d_t\left(\tfrac{1-\lambda}{\lambda+1}\right)-\tfrac{d_t}{2^\lambda(\lambda+1)}.
\end{aligned}
\tag{12}
$$

Regarding the expected step size, $E^{A_t}(d_{t+1}) = d_t\left[\alpha_f P_F + \alpha_s P_S\right]$, and all that is required is to derive expressions for the failure and success probabilities,

P_F and $P_S = 1 - P_F$, respectively. The probability of failure is simply $P_F = [P(|x_{i,t}| > |x_t|)]^\lambda = [P(x_{i,t} > x_t) + P(x_{i,t} < -x_t)]^\lambda$, leading to two distinct cases.

If $0 \leq \frac{x_t}{d_t} < \frac{1}{2}$, then $-x_t > x_t - d_t$, and thus $P(x_{i,t} < -x_t) \neq 0$. Therefore, $P_F = \left(1 - \frac{x_t}{d_t}\right)^\lambda$ and

$$E^{A_t}(d_{t+1}) = d_t \left[\left(1 - \tfrac{x_t}{d_t}\right)^\lambda (\alpha_f - \alpha_s) + \alpha_s\right]. \tag{13}$$

If $\frac{x_t}{d_t} \geq \frac{1}{2}$, then $-x_t \leq x_t - d_t$ and $P(x_{i,t} < -x_t) = 0$. Thus, $P_F = \frac{1}{2^\lambda}$ and

$$E^{A_t}(d_{t+1}) = d_t \left[\frac{1}{2^\lambda}(\alpha_f - \alpha_s) + \alpha_s\right]. \tag{14}$$

To complete step 1, inequality (3) can be rewritten as:

$$E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})] - kE^{A_t}[\ln(d_{t+1})] - \ln(|x_t| + wd_t) + k\ln(d_t) \leq -a. \tag{15}$$

Since $\ln(\cdot)$ is a concave function, using Jensen's inequality,

$$E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})] \leq \ln[E^{A_t}(|x_{t+1}| + wd_{t+1})] \tag{16}$$

and it is sufficient to prove that there exist α_f, α_s, w, k and a such that the following inequality holds:

$$\ln[E^{A_t}(|x_{t+1}|) + wE^{A_t}(d_{t+1})] - kE^{A_t}[\ln(d_{t+1})] - \ln(|x_t| + wd_t) + k\ln(d_t) \leq -a. \tag{17}$$

Due to the particular form of its left-hand side, inequality (17) may be rewritten as:

$$\Psi\left(\frac{|x_t|}{d_t}\right) = \ln\left[\frac{E^{A_t}(|x_{t+1}|) + wE^{A_t}(d_{t+1})}{|x_t| + wd_t}\right] - k(E^{A_t}[\ln(d_{t+1})] - \ln(d_t)) \leq -a. \tag{18}$$

Analytical expressions for function $\Psi(r)$, $r = |x_t|/d_t$, can be obtained analytically by considering the intervals (A) $0 \leq \frac{x_t}{d_t} < \frac{1}{2}$, (B) $\frac{1}{2} \leq \frac{x_t}{d_t} < 1$, and (C) $\frac{x_t}{d_t} \geq 1$, and combining expressions (8) to (14) as appropriate for each interval and each type of selection.

Then, it is necessary to show that, for all $r \geq 0$, $\Psi(r) \leq -a$ for some $a > 0$, particularly at the ends of each interval ($r = 0, 1/2, 1$ and $r \rightarrow +\infty$) and at any critical points inside those intervals ($r = r^*$ such that $\Psi'(r^*) = 0$).

Note that, regardless of the selection scheme, $\Psi(r)$ is represented in each interval by a sum of logarithms of rational fractions, and that $\Psi'(r)$ is represented by a rational fraction, in variable r. Therefore, the critical points within each interval, and the corresponding values of Ψ, can be determined numerically for given values of α_f, α_s, w and k, and the following constrained nonlinear

optimization problem can be formulated:

$$(\alpha_s^*, \alpha_f^*, w^*, k^*, a^*) = \underset{\alpha_s, \alpha_f, w, k, a}{\arg\max} \; a$$

$$\text{subject to:} \begin{cases} \alpha_s \geq 1 \\ 0 < \alpha_f \leq 1 \\ w > 0 \\ 0 \leq k < 1 \\ \Psi(r) + a \leq 0, \;\; r = 0, 1/2, 1, +\infty \\ \Psi(r^*) + a \leq 0, \, r^* : \Psi'(r^*) = 0 \end{cases} \tag{19}$$

Problem (19) can be solved numerically for either $(1 \overset{+}{,} \lambda)$-ES and any selected value of λ. To prove this statement, the principle of finite induction is used. Firstly, it is shown that Problem (19) can be solved both for the $(1+1)$-ES and for the $(1, 2)$-ES (see Table 1). Then, it must also be shown that, if it can be solved for a generic λ, then it can also be solved for $\lambda + 1$.

In order to highlight the dependence of function Ψ on the number of offspring λ, the following notation is introduced:

$$\Psi_\lambda(r) = \ln\left[\frac{E_\lambda^{A_t}(|x_{t+1}|) + w E_\lambda^{A_t}(d_{t+1})}{|x_t| + w d_t} \right] - k\{E_\lambda^{A_t}[\ln(d_{t+1})] - \ln(d_t)\} \tag{20}$$

Clearly, $\forall \lambda \in \mathbb{N}$,

$$-k\{E_{\lambda+1}^{A_t}[\ln(d_{t+1})] - \ln(d_t)\} \leq -k\{E_\lambda^{A_t}[\ln(d_{t+1})] - \ln(d_t)\}, \tag{21}$$

so it is sufficient to show that $\forall \lambda \in \mathbb{N}, \exists \, w' > 0 :$

$$E_{\lambda+1}^{A_t}(|x_{t+1}|) + w' E_{\lambda+1}^{A_t}(d_{t+1}) \leq E_\lambda^{A_t}(|x_{t+1}|) + w E_\lambda^{A_t}(d_{t+1}), \tag{22}$$

where $E_{\lambda+1}^{A_t}(|x_{t+1}|) \leq E_\lambda^{A_t}(|x_{t+1}|)$ and $E_{\lambda+1}^{A_t}(d_{t+1}) \geq E_\lambda^{A_t}(d_{t+1})$.

Suppose that there exist $0 < \alpha_f \leq 1$, $\alpha_s \geq 1$, $w > 0$, $0 \leq k < 1$, and $a > 0$ such that $\Psi_\lambda(r) \leq -a$ for all $r \geq 0$, and let $w_1 > 0$ and $w_2 > 0$ be defined as follows:

$$w_1 = w \frac{2^{\lambda+1} \alpha_f}{\alpha_f + \alpha_s(2^{\lambda+1} - 1)} < w \frac{(1-r)^\lambda(\alpha_f - \alpha_s) + \alpha_s}{(1-r)^{\lambda+1}(\alpha_f - \alpha_s) + \alpha_s} = w \frac{E_\lambda^{A_t}(d_{t+1})}{E_{\lambda+1}^{A_t}(d_{t+1})} \tag{23}$$

where $0 \leq \frac{x_t}{d_t} < \frac{1}{2}$, and

$$w_2 = 2w \frac{\alpha_f + \alpha_s(2^\lambda - 1)}{\alpha_f + \alpha_s(2^{\lambda+1} - 1)} = w \frac{E_\lambda^{A_t}(d_{t+1})}{E_{\lambda+1}^{A_t}(d_{t+1})} \quad \text{where} \quad \frac{x_t}{d_t} \geq \frac{1}{2}. \tag{24}$$

Letting $w' = \min\{w_1, w_2\}$, inequality (22) is shown to hold true, and there are indeed $0 < \alpha_f \leq 1$, $\alpha_s \geq 1$, $w = w' > 0$, $0 \leq k < 1$ and $a > 0$ such that $\Psi_{\lambda+1}(r) \leq -a$ for all $r \geq 0$, concluding the induction step.

$(1,\lambda)$-ES		$\lambda=2$	$\lambda=3$	$\lambda=4$	$\lambda=5$	$\lambda=6$	$\lambda=7$
α_s		1.15180	1.19847	1.19591	1.18729	1.18036	1.17567
α_f		0.72873	0.52779	0.42236	0.36531	0.33489	0.31895
w		2.51320	1.44610	1.09380	0.90415	0.77973	0.68860
k		0.31395	0.30306	0.29601	0.29158	0.28901	0.28761
a		0.00843	0.02379	0.03370	0.03932	0.04223	0.04362
a/λ		0.00422	0.00793	0.00842	0.00786	0.00704	0.00623
$(1+\lambda)$-ES	$\lambda=1$	$\lambda=2$	$\lambda=3$	$\lambda=4$	$\lambda=5$	$\lambda=6$	$\lambda=7$
α_s	1.89274	1.73948	1.64704	1.59411	1.56531	1.55027	1.54260
α_f	0.75675	0.60518	0.51496	0.46382	0.43603	0.42133	0.41370
w	0.11219	0.10567	0.09626	0.08560	0.07521	0.06598	0.05819
k	0.23988	0.24304	0.24488	0.24591	0.24648	0.24681	0.24701
a	0.04309	0.07039	0.08660	0.09569	0.10060	0.10318	0.10453
a/λ	0.04309	0.03518	0.02887	0.02392	0.02012	0.01720	0.01493

Table 1. Values for α_s, α_f, w, k, a and a/λ obtained by solving optimization problem (19) for $(1,\lambda)$ and $(1+\lambda)$-ESs and different numbers of offspring (λ).

By solving Problem (19) numerically for each type of selection and number of offspring, suitable step-size adaptation parameters α_f and α_s are obtained, and inequality (3) is shown to hold true, completing step 1 of the proof of Proposition 2. An upper bound e^{-a} on the convergence rate of such an ES on any strictly unimodal, symmetric function of one variable is also obtained as a by-product.

Table 1 shows the parameter and auxiliary values, as well as the convergence bounds, obtained for $(1,\lambda)$ and $(1+\lambda)$ selection and several values of λ. The results were obtained using the sequential quadratic programming solver in the GNU Octave numerical package. From the table, it is possible to select the number of offspring λ which minimizes the convergence rate bound for each $(1 \overset{+}{,} \lambda)$-ES with respect to the number of function evaluations. For the $(1,\lambda)$-ES, $\lambda = 4$ can be seen to lead to the highest value of a/λ, whereas for the $(1+\lambda)$-ES, the best number of offspring appears to be $\lambda = 1$.

6 Conclusions

In this paper, a simple success-based $(1 \overset{+}{,} \lambda)$-ES with uniformly-distributed mutations for functions of one variable was proposed. Following the theoretical approach proposed in [5], the convergence of the ES was studied on the class of unimodal functions symmetric around the optimum using stochastic Lyapunov function and martingale theory. General expressions for the expectation of x_{t+1} and d_{t_1} were derived considering both $(1 \overset{+}{,} \lambda)$ selection. Using an appropriate Lyapunov function, upper bounds on the convergence rate and specific ES parameter values were obtained via numerical optimization, for growing values of λ. The number of offspring that minimizes the bound on the convergence rate with respect to the number of function evaluations was also determined in this

way. Future work includes an experimental study of the actual convergence rates achieved with the proposed parameter settings as well as extending the theoretical results to other function classes, other mutation distributions, and multiple decision variables.

Acknowledgement

References

1. N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159 –195, 2001.
2. B. Doerr and C. Doerr. Optimal parameter choices through self-adjustment: Applying the 1/5-th rule in discrete settings. In *Proceedings of the 2015 ACM-GECCO Genetic and Evolutionary Computation Conference*, pages 1335 –1342. ACM, 2015.
3. B. Doerr and C. Doerr. A tight runtime analysis of the $(1+(\lambda, \lambda))$ genetic algorithm on onemax. In *Proceedings of the 2015 ACM-GECCO Genetic and Evolutionary Computation Conference*, pages 1423 –1430. ACM, 2015.
4. E. F. Wanner, C. M. Fonseca, R. T. N. Cardoso, and R. H. C. Takahashi. Lyapunov stability analysis and adaptation law synthesis of a derandomized self-adaptive $(1, 2)$-ES. Under review.
5. M. A. Semenov and D. A. Terkel. Analysis of convergence of an evolutionary algorithm with self-adaptation using a stochastic Lyapunov function. *Evolutionary Computation*, 11(4):363–379, 2003.
6. J. Jägersküpper. A blend of Markov-chain and drift analysis. In *Proceedings of 10th International Conference on Parallel Problem Solving from Nature*, pages 41–51. Springer, 2008.
7. J. He and X. Yao. Drift analysis and average time complexity of evolutionary algorithms. *Artificial Intelligence*, 127:57–85, 2001.
8. J. He and X. Yao. A study of drift analysis for estimating computational time of evolutionary algorithms. *Natural Computing*, 3:21–35, 2004.
9. B. Doerr, D. Johannsen, and C. Winzen. Multiplicative drift analysis. *Algorithmica*, 64(4):673–697, 2011.
10. C. Witt. Tight bounds on the optimization time of a randomized search heuristic on linear function. *Combinatorics Probability and Computing*, 22(02):294 –318, 2013.
11. W. E. Hart. Rethinking the design of real-coded evolutionary algorithms: Making discrete choices in continuous search domains. *Soft Computing Journal*, 9:225–235, 2002.
12. A. M. Lyapunov. The general problem of stability of motion (reprint of the original paper of 1892). *International Journal of Control*, 55(3):531–773, 1992.
13. W. Hahn. *Stability of Motion*. Springer-Verlag, 1967.
14. S. Janson, T. Luczak, and A. Rucinski. *Random Graphs*. Wiley, 2000.