Sizing the Population with Respect to the Local Progress in $(1,\lambda)$-Evolution Strategies — A Theoretical Analysis

Nikolaus Hansen, Andreas Gawelczyk and Andreas Ostermeier

Technische Universität Berlin, Sekr. ACK 1

Ackerstr. 71–76, 13355 Berlin, Germany

E-mail: {hansen,gawelczyk,ostermeier}@fb10.tu-berlin.de

ABSTRACT

This paper presents an analysis of the local serial rate of progress with respect to the number of offspring λ for the $(1,\lambda)$-evolution strategy. It is shown that local serial progress is maximized when the expected progress of the second best offspring is zero. The theoretical results lead to a simple but efficient adaptation rule for λ, which needs no extra fitness function evaluations and only small computational expense. Simulations of the λ-adaptation on simple test functions are shown.

Keywords: evolutionary algorithms, evolution strategy; population size, offspring number; adaptation, strategy parameters; progress, serial progress

1. Introduction

In most applications of evolution strategies (ESs), the number of offspring λ is held fixed throughout optimization. Based on results yielded for a spherical objective function in [7], its value can be chosen such that the serial rate of progress (e.g. the expected fitness gain per function evaluation) is optimal with respect to the hyper-sphere problem.

One idea of λ-adaptation, sketched out by Schwefel [7] and successfully applied by Herdy [2], proposes a mutative adaptation in a two-level ES, where on the upper level, called *population level*, λ is treated as a variable to be optimized while on the lower level, called *individual level*, the object parameters are optimized. After isolating the populations for a given number of function evaluations, a selection between populations takes place. Thus, the best value of λ with regard to the quality gain per function evaluation is detected and inherited to the next set of populations. The drawback of this algorithm is that the computational expense rises by the factor *number of populations*.

Optimizing internal strategy parameters, like the number of offspring, by mutation and selection on the *individual level* turns out to be difficult, because within one generation the relevance for selection due to internal strategy parameters may be quite low, and stochastic fluctuations of these parameters may deteriorate the whole optimization. In principle, this problem can be overcome by producing high variations between the strategy parameter realizations within one population (to rise the selection relevance) and small variations of the strat-egy parameters between succeeding populations (to reduce stochastic fluctuations), as pointed out in [3]. When adapting parameters of the mutation distribution (e.g. step sizes), this is often achieved by using large populations and recombination techniques, but can also be done independent of the population size by the so called *derandomization* of the adaptation process [3].

In this paper, we do not deal with *mutative* parameter control. Instead, we will introduce a *deterministic* adaptation scheme for the number of offspring based on theoretical considerations on the relation between serial rates of progress for the actual number of offspring λ, for $\lambda - 1$ and for the optimal number of offspring.

2. Theoretical Analysis

2.1. Motivation

To optimize the local serial progress (i.e. progress per fitness function evaluation) in a $(1,\lambda)$-ES with respect to the number of offspring λ, it would be ideal to know the progress-rate as a function of λ.[1] In practical applications it may be sufficient to estimate the "gradient" of this progress function at the current λ, to get the right direction for λ-adaptation. The fitness gain of the fittest offspring can be used to estimate the progress for current λ. To estimate the progress for $\lambda - 1$, we can

---

[1]Progress-rate means the expected progress of the selected (i.e. fittest) offspring with respect to the current parent. For example, the expected difference between the fitness values of the fittest offspring and the parent *can* be considered as progress-rate.

remove one offspring and use the fitness gain of the remaining fittest. The choice of the offspring to be removed has to be independent of its fitness. Repeating the estimating process with different offspring will raise the estimation quality. It can be expected that removing every offspring exactly once and calculating the mean of the fitness gains is the best estimator. This basic idea (pursued in Lemma 1) lays the foundation of the theoretical results, which are now briefly summarized:

1. The serial progress-rate as a function of λ is either a function with exact one (local and global) maximum or a strictly monotonically increasing function.
2. The serial progress-rate of a $(1,\lambda)$-ES increases with the reduction of λ, exactly when the progress-rate of the second fittest offspring is greater than zero.
3. The serial progress-rate of a $(1,\lambda)$-ES decreases with the reduction of λ, exactly when the progress-rate of the second fittest offspring is less than zero.

In the following, first we will prove a theorem, which gives us 2. and 3. The Corollary 1 then proves 1. Important notations and definitions given in the following section are summarized in Appendix A.

2.2. Theoretical Results

Let $x^g = x \in \mathbb{R}^n$ be the object variable vector of the parent in generation g. Offspring are produced by adding (multi-)normally distributed random numbers on x and can be interpreted as λ independent and identically distributed random vectors $X_1, \ldots, X_\lambda$.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be the fitness function to be maximized and $F_i := f(X_i) \in \mathbb{R}$ the fitness value of offspring i, which is a random variable as well. Then $F_1, \ldots, F_\lambda$ rearranged in *descending* order of magnitude will be written as $F_{1:\lambda}, \ldots, F_{\lambda:\lambda}$, denoting the order statistics of F_i. Furthermore, we define $X_{i:\lambda}$ to be the ith fittest offspring, so that $f(X_{i:\lambda}) = F_{i:\lambda}$ holds (in case of equal fitness the offspring may be ordered arbitrarily).

The progress function $\Phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $(x, X) \mapsto \Phi(x, X) =: \Phi(X)$ defines the progress of offspring X with respect to parent x. The progress of the fittest individual implies the actual progress of the ES and will be denoted by $\Phi_{1:\lambda} := \Phi(X_{1:\lambda})$. For the moment, there will be no more requirements on the definition of $\Phi_{1:\lambda}$. Especially we do not assume that, in general, the fittest individual will have the largest progress, that is $\Phi_{1:\lambda} \geq \ldots \geq \Phi_{\lambda:\lambda}$, where $\Phi_{i:\lambda} := \Phi(X_{i:\lambda})$, may not hold. The only connection between f and $\Phi_{1:\lambda}$ is, that the selection criterion f determines the argument of Φ for $\Phi_{1:\lambda}$. The reason for not assuming any direct relation be-

tween f and Φ is to keep generality, e.g. permitting noisy fitness evaluation.

The progress-rate $\varphi_{1,\lambda}$ of the $(1,\lambda)$-ES will now be defined by the expected progress of the fittest offspring, while the progress-rate of the ith fittest offspring will be denoted by $\varphi_{i:\lambda} := \mathsf{E}[\Phi(X_{i:\lambda})]$, $i = 1, \ldots, \lambda$. Now we can turn to the first lemma, a recurrence relation between $\varphi_{1:\lambda}$ and $\varphi_{1:\lambda-1}$, which was stated before by Rechenberg [6] in the special case of linear fitness functions.

Lemma 1 *For $\lambda \geq 2$ holds the relation:*

$$\lambda \, \varphi_{1:\lambda-1} = \varphi_{2:\lambda} + (\lambda - 1) \, \varphi_{1:\lambda}$$

Proof Let $\Phi_{1:\lambda}^{\{i\}}$ be the progress of the $(1,\lambda)$-ES, if the ith offspring[2] is removed before selection, and $\Phi_{1:\lambda}^{[i]}$ the progress of the $(1,\lambda)$-ES, if the ith *fittest* offspring is removed before selection. Obviously, removing the ith offspring is the same as producing only $\lambda - 1$ offspring:

$$\Phi_{1:\lambda}^{\{i\}} = \Phi_{1:\lambda-1} \qquad \forall i = 1, \ldots, \lambda. \qquad (1)$$

Furthermore it holds

$$\sum_{i=1}^{\lambda} \Phi_{1:\lambda}^{\{i\}} = \sum_{i=1}^{\lambda} \Phi_{1:\lambda}^{[i]}, \qquad (2)$$

because the LHS is a rearrangement of the RHS. With linearity of the expectation value we yield

$$
\begin{aligned}
\lambda \, \varphi_{1:\lambda-1} &= \mathsf{E}\left[\lambda \, \Phi_{1:\lambda-1}\right] \\
&\overset{\text{eq. (1)}}{=} \mathsf{E}\left[\sum_{i=1}^{\lambda} \Phi_{1:\lambda}^{\{i\}}\right] \\
&\overset{\text{eq. (2)}}{=} \mathsf{E}\left[\sum_{i=1}^{\lambda} \Phi_{1:\lambda}^{[i]}\right] \\
&= \sum_{i=1}^{\lambda} \mathsf{E}\left[\Phi_{1:\lambda}^{[i]}\right] \\
&= \varphi_{2:\lambda} + (\lambda - 1) \, \varphi_{1:\lambda}
\end{aligned}
$$

$\square$

This kind of recurrence relation is well known in order statistics and of course is true for the relation between $\mathsf{E}[F_{1:\lambda-1}]$ and $\mathsf{E}[F_{1:\lambda}]$ (see e.g. [1]), but the equation holds also for any function from the domain of the order statistics $F_{i:m}$, as shown in Lemma 1. Now we can prove a theorem on the relation of the *serial* progress-rates $\varphi_{1:\lambda}/\lambda$ and $\varphi_{1:\lambda-1}/(\lambda - 1)$.[3]

---

[2]This is the offspring X_i, which — for example — may always be *produced* as the ith.

[3]If $\varphi_{1:\lambda}$ is less than zero, this definition of the serial progress-rate — which is to be maximized — intuitively makes no sense. Nevertheless, the theoretical results even hold in that case and furthermore show that this fact is irrelevant.

Theorem 1 *The serial progress-rate for $\lambda - 1$ offspring is less/equal/greater than the serial progress-rate for λ offspring, exactly when the progress-rate of the second fittest offspring is less/equal/greater than zero.*

Proof

$$\frac{\varphi_{1:\lambda-1}}{\lambda-1} \gtreqless \frac{\varphi_{1:\lambda}}{\lambda}$$

$$\Longleftrightarrow \quad \lambda\, \varphi_{1:\lambda-1} \gtreqless (\lambda - 1)\, \varphi_{1:\lambda}$$

$$\overset{\text{Lemma 1}}{\Longleftrightarrow} \quad \varphi_{2:\lambda} + (\lambda - 1)\, \varphi_{1:\lambda} \gtreqless (\lambda - 1)\, \varphi_{1:\lambda}$$

$$\Longleftrightarrow \quad \varphi_{2:\lambda} \gtreqless 0$$

where $\gtreqless$ means "is greater than or equal or less than, respectively" and $\lambda \geq 2$. □

Exactly when the progress-rate of the second fittest offspring is greater than zero, it is better to produce only $\lambda - 1$ than λ offspring. Progress of the second best offspring can be seen as wasted progress.

Now we want to characterize the local, serial progress-rate $\varphi_{1:\lambda}^{\text{serial}} := \varphi_{1:\lambda}/\lambda$ as a function of λ. We will prove that $\varphi_{1:\lambda}^{\text{serial}}(\lambda)$ is a function without *local* maxima and, apart from the global optimum, strictly monotonic. To show this, we have to make a slightly stronger assumption on Φ: The more offspring are produced, the higher should be the expected progress of the ES, i.e. it holds for all $\lambda \in \mathbb{N}$: $\mathsf{E}\left[\Phi_{1:\lambda}\right] \leq \mathsf{E}\left[\Phi_{1:\lambda+1}\right]$, and especially $\mathsf{E}\left[\Phi_{2:\lambda}\right] \leq \mathsf{E}\left[\Phi_{2:\lambda+1}\right]$ as used in the proof of Corollary 1. If we assume a given optimization problem and a corresponding progress function, then maintaining this inequation is a necessary requirement on the fitness function for doing reasonable optimization with any selection scheme.

Let $\lambda_{\text{optmin}} < \ldots < \lambda_{\text{optmax}}$ be all the offspring number of one global maximum of $\varphi_{1:\lambda}(\lambda)$ (in case of existence). If the maximum is isolated, let $\lambda_{\text{optmin}} = \lambda_{\text{optmax}}$ by definition. For notational convenience we assume $\lambda_{\text{optmin}} \geq 2$ and permit $\lambda_{\text{optmax}} = \infty$. Due to this definition it holds $\varphi_{1:\lambda_{\text{optmin}}-1}^{\text{serial}} < \varphi_{1:\lambda_{\text{optmin}}}^{\text{serial}} = \ldots = \varphi_{1:\lambda_{\text{optmax}}}^{\text{serial}}$ and if $\lambda_{\text{optmax}} < \infty$ also $\varphi_{1:\lambda_{\text{optmax}}}^{\text{serial}} > \varphi_{1:\lambda_{\text{optmax}}+1}^{\text{serial}}$.

Corollary 1
If an optimal number of offspring exists, it holds

$$\varphi_{1:1}^{\text{serial}} < \ldots < \varphi_{1:\lambda_{\text{optmin}}}^{\text{serial}} \tag{3}$$

$$\text{and if} \quad \lambda_{\text{optmax}} < \infty$$

$$\varphi_{1:\lambda_{\text{optmax}}}^{\text{serial}} > \varphi_{1:\lambda_{\text{optmax}}+1}^{\text{serial}} > \ldots \tag{4}$$

If no optimal number of offspring exists, it holds

$$\varphi_{1:1}^{\text{serial}} < \varphi_{1:2}^{\text{serial}} < \ldots \tag{5}$$

Proof

(3): For all $k = 0, \ldots, \lambda_{\text{optmin}} - 2$:

$$\varphi_{1:\lambda_{\text{optmin}}-1}^{\text{serial}} < \varphi_{1:\lambda_{\text{optmin}}}^{\text{serial}}$$

$$\overset{\text{Theorem 1}}{\Longleftrightarrow} \quad \varphi_{2:\lambda_{\text{optmin}}} < 0$$

$$\Longrightarrow \quad \varphi_{2:\lambda_{\text{optmin}}-k} < 0$$

$$\overset{\text{Theorem 1}}{\Longleftrightarrow} \quad \varphi_{\lambda_{\text{optmin}}-k-1}^{\text{serial}} < \varphi_{\lambda_{\text{optmin}}-k}^{\text{serial}}$$

$$\Longleftrightarrow \quad (3)$$

(4): For all $k \in \mathbb{N}_0$:

$$\varphi_{1:\lambda_{\text{optmax}}}^{\text{serial}} > \varphi_{1:\lambda_{\text{optmax}}+1}^{\text{serial}}$$

$$\overset{\text{Theorem 1}}{\Longleftrightarrow} \quad \varphi_{2:\lambda_{\text{optmax}}+1} > 0$$

$$\Longrightarrow \quad \varphi_{2:\lambda_{\text{optmax}}+1+k} > 0$$

$$\overset{\text{Theorem 1}}{\Longleftrightarrow} \quad \varphi_{\lambda_{\text{optmax}}+k}^{\text{serial}} > \varphi_{\lambda_{\text{optmax}}+k+1}^{\text{serial}}$$

$$\Longleftrightarrow \quad (4)$$

(5): Assume the existence of a λ_0 with $\varphi_{1:\lambda_0}^{\text{serial}} \geq \varphi_{1:\lambda_0+1}^{\text{serial}}$. It follows $\varphi_{1:\lambda_0+k}^{\text{serial}} \geq \varphi_{1:\lambda_0+k+1}^{\text{serial}}$ for all $k \in \mathbb{N}$ as in the proof of (4), in contradiction to the non-existence of an optimal number of offspring. □

In summary, we can state the following relations between a given $\lambda \geq 2$, the progress-rate of the second best offspring $\varphi_{2:\lambda}$ and an optimal number of offspring $\lambda_{\text{opt}} \leq \infty$:

Corollary 2

$$\varphi_{2:\lambda} > 0 \quad \Longleftrightarrow \quad \lambda > \lambda_{\text{opt}}$$

$$\varphi_{2:\lambda} < 0 \quad \Longrightarrow \quad \lambda \leq \lambda_{\text{opt}}$$

$$\varphi_{2:\lambda} = 0 \quad \Longleftrightarrow \quad \lambda \text{ and } \lambda-1 \text{ produce maximal serial progress}$$

Proof Follows from applying first Theorem 1 and then Corollary 1. □

Due to Corollary 1 and 2, $\varphi_{2:\lambda}$ becomes a useful source of information for the adaptation of λ, which will be utilized in the next section. In conclusion of this section, we give a **summary of assumptions** for getting the given theoretical results:

1. Offspring in one generation are independent and identically distributed. The number of offspring is greater than one.
2. The fitness function f determines the selection and therefore the argument $\boldsymbol{X}_{1:\lambda}$ for the progress function Φ, which then defines the local progress of the ES.
3. Progress-rate $\varphi_{1:\lambda}$ is defined as *expectation of the local progress* of the fittest offspring, $\mathsf{E}\left[\Phi(\boldsymbol{X}_{1:\lambda})\right]$.
4. Progress-rate increases with increasing λ, due to a corresponding statistical relation between f and Φ.

3

Confined to these assumptions, the results can be applied to *any* reproduction–$(1,\lambda)$-selection scheme. Further investigations aim at generalization of the results for (μ,λ)-ESs with $\mu > 1$.

3. An Adaptation Scheme for λ

As shown in the last section, we can get evidence for the relation between the offspring number λ and the optimal number of offspring with respect to the serial progress-rate by evaluating the progress-rate of the second best offspring $\varphi_{2:\lambda}$ (cf. Corollary 2). To estimate $\varphi_{2:\lambda}$, we will use the fitness difference between the second best offspring and the parent, denoted by $\Delta f_{(2)}$. Introducing the parameter $\lambda^g \in \mathbb{R}$ at generation g, the adaptation scheme is:

$$\lambda^{g+1} = \lambda^g \exp\left(\beta_\lambda \frac{-\Delta f_{(2)}}{\sigma_\lambda}\right) \qquad (6)$$

where

$\Delta f_{(2)}$ fitness difference between second fittest offspring and parent, in case of minimization multiplied by -1.

$\beta_\lambda \geq 0$ parameter which determines adaptation speed. Large values lead to fast adaptation and large stochastic fluctuations of λ^g. If $\beta_\lambda = 0$ no adaptation takes place. Sensible values are clearly less than 1 (see Figure 1 below).

$\sigma_\lambda = \sqrt{\sum_{i=1}^{\lambda} \left(\Delta f_{(i)}\right)^2 / (\lambda - 1)}$,   where $\Delta f_{(i)}$ is the fitness difference between i^{th} fittest offspring and parent. σ_λ normalizes $\Delta f_{(2)}$ in (6). This makes the adaptation invariant for fitness function scaling.

The adaptation mechanism increases λ^g if $\Delta f_{(2)} < 0$, which tends to be the case if $\varphi_{2:\lambda} < 0$, where the current offspring number is less than or equal to λ_{opt} (cf. Corollary 2). It decreases λ^g, if $\Delta f_{(2)} > 0$, which tends to be the case if $\varphi_{2:\lambda} > 0$, where the current offspring number is greater than λ_{opt}. λ^g should be restricted to values ≥ 2, and the number of offspring in generation g is calculated by truncation of λ^g. Due to the normalization, $\Delta f_{(2)}/\sigma_\lambda$ is restricted to values between -1 and $\sqrt{(\lambda - 1)/2}$. We want to emphasize that the adaptation rule is independent of translation and scaling of the fitness function.

4. Simulations

Tests of the adaptation scheme were done on the hyper-plane $(\max_{\vec{x}} \{\sum_{i=1}^{n} x_i\})$ and hyper-sphere $(\min_{\vec{x}} \{\sum_{i=1}^{n} x_i^2\})$ with dimension $n = 100$. The optimal number of offspring is two *and* three for the former, five for the latter and large n (see [7] and [2]). The first number can easily be achieved
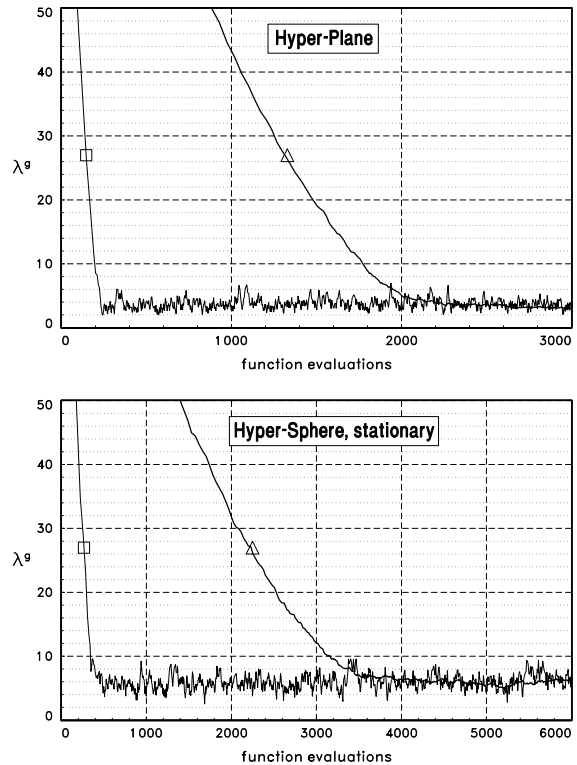

Figure 1: Evaluation of sensible values for β_λ. $\square$: $\beta_\lambda = 0.3$, $\triangle$: $\beta_\lambda = 0.03$. Simulations with fixed — in the case of the hyper-sphere optimal — global step size, starting with $\lambda^1 = 100$.

from our theoretical results: Obviously the expected progress of the second out of three offspring is zero on the hyper-plane.[4] Due to Corollary 2 two and three are optimal offspring numbers then. Considering more than three offspring, the expected progress of the second best is obviously greater than zero and therefore, due to Corollary 2, offspring numbers greater three are not optimal.

To show the adaptation *process*, no averaged data are presented, but typical runs out of at least ten. We found different runs with identical parameter settings looking very similar.

Simulations for different values of β_λ are shown in **Figure 1**. Step-sizes are fixed. On the hyper-sphere, the best offspring is always projected back onto the unit sphere, to keep on mutating with optimal step size. Fast adaptation and large stochastic fluctuations are mutually dependent. The faster adaptation on the hyper-plane is due to the normalization process: Offspring on the hyper-sphere produce more often fitness values which are clearly worse than the parent's fitness, increasing σ_λ remarkably, compared to the hyper-plane. In further simulations we found no evidence that sensible values for β_λ depend on the dimension of the problem.

Figure 2 shows the adaptation process for different, fixed step sizes on the hyper-sphere, which

---

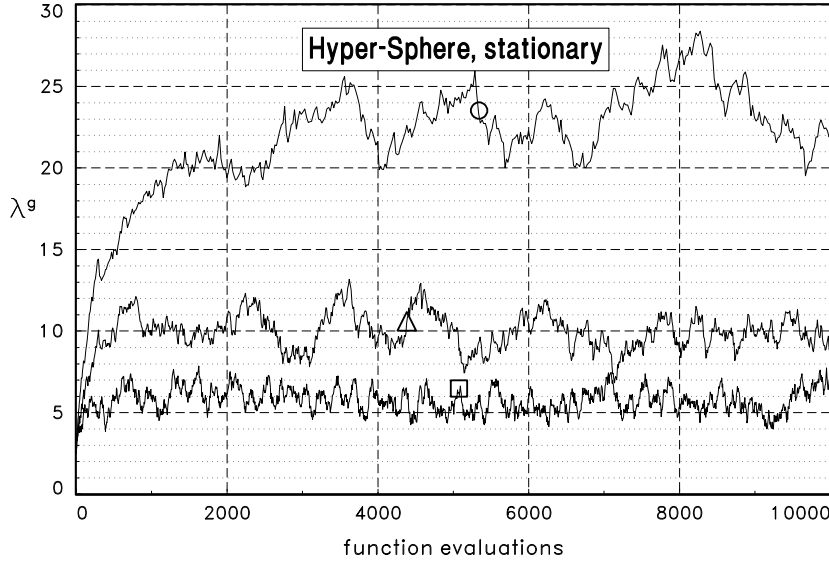[4] Assuming a reasonable progress definition.

Figure 2: Simulations with different, fixed step sizes δ. Offspring number is always calculated by truncation of λ^g.

$\square$: $\delta = 1.163$ (optimal step size),

$\triangle$: $\delta = 2.0$,

$\bigcirc$: $\delta = 3.0$.

The optimal offspring numbers are 5, 9 and 23, respectively (for large n). Parameter $\beta_\lambda = 0.1$, start with $\lambda^1 = 2$.
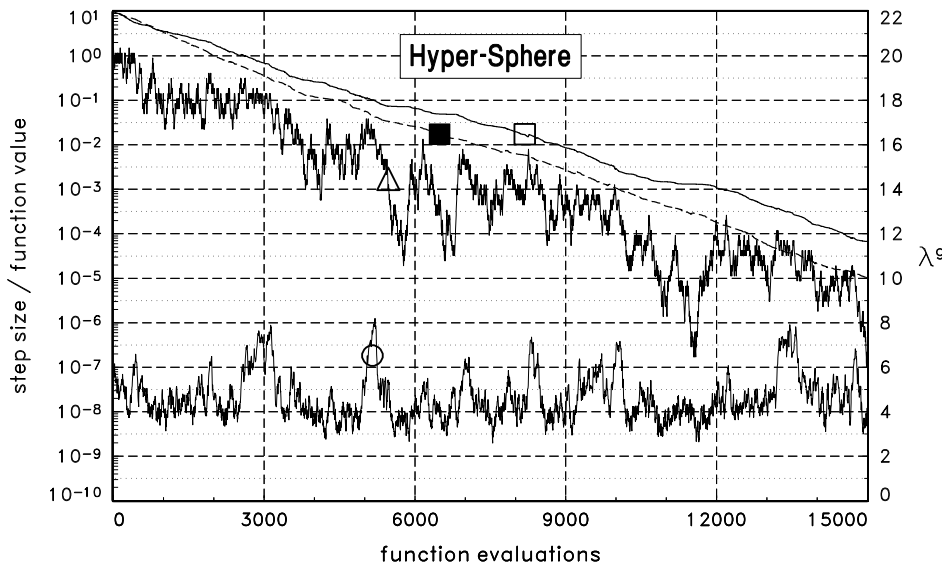

Figure 3: Simulations with mutative step size control.

$\blacksquare$: Fitness with $\lambda = 5$ fixed,

$\square$: Fitness with λ-adaptation,

$\triangle$: Step size for $\square$,

$\bigcirc$: λ^g for $\square$ (scale on the right).

$\beta_\lambda = 0.1$, start with $\lambda^1 = 5$.

lead to different optimal offspring numbers as given in the caption (cf. [2]). $\lambda_{\rm opt}$ increases with increasing step size here. Apart from stochastic effects, the adaptation produces optimal offspring number or optimal offspring number plus one, as expected from the theory. This can be observed better using smaller values of β_λ (not shown here).

Figure 3 shows simulations with global mutative step size control at work, where step size mutation is carried out by multiplying the parent's step size by 1.3 or 1/1.3, both with probability 0.5 ([5], [6]). In general, mutative step size control adapts the global step size to achieve maximal probability to produce the fittest offspring with the adapted step size, rather than to achieve maximal fitness gain with the fittest offspring. This results in a step size smaller than optimal. In consequence, the optimal number of offspring decreases as discussed before, resulting in a further reduction of the step size. The mean offspring number becomes four then. This behavior leads to a reduction of convergence speed

by about 15% compared to the simulation with theoretically optimal offspring number. It can be avoided using better step size adaptation methods, for example cumulative step size adaptation [4], which adapts the optimal step size exactly and reduces the stochastic fluctuations of the step size as well.

The fluctuations of λ^g are correlated with stochastic fluctuations of the step size and can mainly be interpreted as an adaptation process due to the current step size. They can be significantly reduced by reducing β_λ. In either case, due to the mutative step size control, λ^g should be restricted to values ≥ 4, and in practical applications β_λ should be chosen clearly less than 0.1.

To reveal the practicability of the adaptation scheme — especially with regard to the interactions with mutative step size control — more simulations with a wide range of different fitness functions have to be done.

Acknowledgements

References

[1] David, H.A. (1970). *Order Statistics*. New York: John Wiley & Sons.

[2] Herdy, M. (1993). The Number of Offspring as Strategy Parameter in Hierarchically Organized Evolution Strategies. *SIGBIO Newsletter*, 13(2): 2–7.

[3] Ostermeier, A., Gawelczyk, A. & Hansen, N. (1994). A Derandomized Approach to Self-Adaptation of Evolution Strategies. *Evolutionary Computation 2(4)*: 369–380.

[4] Ostermeier, A., Gawelczyk, A. & Hansen, N. (1994). Step-size Adaptation Based on Non-local Use of Selection Information. In Davidor, Y., Schwefel, H.-P. & Männer, R. (eds.), *Parallel Problem Solving from Nature – PPSN III, Proceedings*: 189–198. Berlin: Springer.

[5] Rechenberg, I. (1989). Artificial evolution and artificial intelligence. In Forsyth, R. (ed.), *Machine Learning: Principles and Techniques*: 83–103. London: Chapman and Hall.

[6] Rechenberg, I. (1994). *Evolutionsstrategie '94*. Stuttgart: fromman-holzboog.

[7] Schwefel, H.-P. (1981). *Numerical optimization of computer models*. Chichester: Wiley.

A. Notations and Definitions

$x \in \mathbb{R}^n$ object variable vector of the parent.

$X_i \in \mathbb{R}^n$ object variable vector of offspring i. The X_i are independent and identically distributed random vectors.

$X_{i:m}$ the ith fittest of $X_1, \ldots, X_m$, i.e. $f(X_{1:m}) \geq \ldots \geq f(X_{m:m})$.

λ number of offspring.

λ_{opt} an optimal offspring number with respect to the serial rate of progress.

$\left.\begin{array}{c}\lambda_{\text{optmin}} \\ \lambda_{\text{optmax}}\end{array}\right\}$ smallest/largest optimal offspring number. If $\lambda_{\text{optmin}} \neq \lambda_{\text{optmax}}$ it holds $\varphi^{\text{serial}}_{1:\lambda_{\text{optmin}}} = \varphi^{\text{serial}}_{1:\lambda_{\text{optmin}}+1} = \ldots = \varphi^{\text{serial}}_{1:\lambda_{\text{optmax}}-1} = \varphi^{\text{serial}}_{1:\lambda_{\text{optmax}}}$.

$f : \mathbb{R}^n \to \mathbb{R}, X \mapsto f(X)$ objective or fitness function (to be maximized).

$F_i := f(X_i)$ fitness value of the ith offspring.

$F_{i:m} := f(X_{i:m})$ fitness value of the ith fittest out of m offspring $X_{i:m}$.

$\Phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, (x, X) \mapsto \Phi(x, X) =: \Phi(X)$ progress function which defines the progress of offspring X with respect to parent x.

$\Phi_{i:m} := \Phi(x, X_{i:m})$ progress of the ith fittest out of m individuals.

$\varphi_{i:m} := \mathsf{E}\left[\Phi_{i:m}\right]$ progress-rate (expectation value of the progress) of the ith fittest out of m individuals.

$\varphi^{\text{serial}}_{1:\lambda} := \varphi_{1:\lambda}/\lambda$ serial progress-rate, i.e. progress-rate per fitness function evaluation of $(1,\lambda)$-ES.