On the Asymptotic Behavior of Multirecombinant Evolution Strategies

Hans-Georg Beyer*

University of Dortmund, Department of Computer Science,
Systems Analysis Research Group,
D-44221 Dortmund, Germany

Abstract. The performance of $(\mu/\mu, \lambda)$-ESs (Evolution Strategies) in the asymptotic limit for $N \to \infty$ and $\lambda \to \infty$ is investigated. The conjecture made by Schwefel that the maximum performance of such strategies scales like $\mu \ln(\lambda/\mu)$ will be proved. Furthermore, it will be shown that an optimally tuned $(\mu/\mu, \lambda)$-ES performs exactly λ times faster than an optimally tuned $(1+1)$-ES, if the hyper-sphere is taken as the fitness model (using the number of generations as the performance measure). The notion of fitness efficiency will be introduced and will be used to derive the ES time complexity. The results are compared to the non-recombinant (μ, λ)-ES.

1 Introduction

The performance of Evolution Strategies (ES) can be investigated by experiments and by theoretical analysis. Both approaches have their advantages and disadvantages. Usually it is much easier to perform ES runs and present the results by graphs. This method can be applied to each fitness function $F(\mathbf{x})$. However, there remains always an uncertainty as to the significance of such results with respect to its generality. The question is what can we learn from those very special ES results as to the ES performance in general? For example, what would we expect if the size N (dimension of the parameter vector $\mathbf{x}$ to be optimized) of the optimization problem is increased?

Indeed, it is really difficult to estimate the computational complexity of the ES algorithms by experiments. The same holds for the influence of the strategy parameters like the mutation strength σ, the number of parents μ and offspring λ used, etc. Even for the simplest fitness functions, as e. g. the inclined hyperplane or the sphere model, the ES time complexity G (for the definition, see Sec. 3 below) becomes a multidimensional function of these exogenous parameters (N, μ, $\lambda, \ldots$). Deriving (better: estimating) this function from empirical data and by ES simulations is almost excluded if there is not any clue from the theory. Actually, *the theory determines what we will measure* (after Einstein).

The second approach to the ES performance analysis is by theory. Such analysis is necessarily restricted to simple fitness models and special ES variants.

* e-mail: beyer@LS11.informatik.uni-dortmund.de

However, one hopes that the results obtained can be applied as approximations to real world fitness functions. That is, if one derives results from the sphere model, for example, one expects that "similar" fitness landscapes exhibit a similar behavior. Thus, the result of the sphere model analysis can be extended to more complicated fitness functions. The predictions made from such extensions should be the starting point for ES experiments (and not the other way around). From this point of view ES experiments should serve first of all as *tests* to verify the approximations made and to investigate the application limits of the model(s) used.

This paper deals mainly with the asymptotic performance behavior of multirecombinant $(\mu/\mu, \lambda)$-ESs, i. e., with the question how the ES does work for $N \to \infty$, $\lambda \to \infty$, and $\mu \to \infty$. It is the goal to derive formulae that describe the ES performance in a rather simple fashion than the more accurate results provided in [3]. Note, this is a good example for which the experimental approach is almost excluded.

The paper is divided in two parts dealing with the microscopic and the macroscopic aspects, respectively, of the ES performance. The microscopic aspect of the ES performance can be measured by the progress rate φ. In the first part we will investigate the asymptotic behavior of the *normalized* progress rate $\varphi^\star$ on the sphere model. This comprises the analysis of the progress rate formula for $N \to \infty$ and the investigation of the *progress coefficients* $c_{\mu/\mu,\lambda}$ for $\lambda \to \infty$. This result will be used to prove Schwefel's $\varphi^* \approx \mu \ln(\lambda/\mu)$-conjecture [10]. Furthermore, the *fitness efficiency* η will be introduced and it will be shown that in the asymptotic limit the $(\mu/\mu, \lambda)$-ES performs exactly λ-times faster than the simple $(1+1)$-ES.

The shorter second part is devoted to the ES dynamics and the (local) ES time complexity, again for the asymptotic case.

It is assumed that the reader is familiar with the recent results of the ES theory. Because of the page limitation, the readers are to refer to the article of Beyer [3] where the $(\mu/\mu, \lambda)$-algorithms are defined, the basic theory is developed, and the benefit of recombination is explained by the *genetic repair principle*. Here, we can give only a very rough overview, aiming first of all as a reminder.

The $(\mu/\mu, \lambda)$-ES. The mutations used are in accordance with those from [3]. The mutation strength is denoted by σ, the progress rate by φ. The normalizations $\varphi^\star := \varphi\ N/R$ and $\sigma^\star := \sigma\ N/R$ use N as the dimension of the object parameter vector $\mathbf{x}$ to be optimized (with respect to a fitness function $F(\mathbf{x})$) and R is the distance of the *center of mass parent* to the center of the spherical model at generation (g).

There are two variants of "μ/μ"-recombination which have been investigated by theoretical analysis: the *intermediate* version, denoted by $(\mu/\mu_I, \lambda)$, and the *dominant* version, often called *global discrete recombination* [9], denoted by $(\mu/\mu_D, \lambda)$.

In $(\mu/\mu_I, \lambda)$-strategies the λ offspring $\mathbf{y}_l$ are generated from the μ parents $\mathbf{x}_m$ by mutation of the center of mass parent $\langle \mathbf{x} \rangle_\mu := \frac{1}{\mu} \sum_{m=1}^{\mu} \mathbf{x}_m$, i. e. $\mathbf{y}_l := \langle \mathbf{x} \rangle_\mu + \mathbf{z}$, where $\mathbf{z}$ is a random vector with N iid normal variates $z_i = \mathcal{N}(0, \sigma^2)$.

The dominant recombination $(\mu/\mu_D,\lambda)$ produces each vector component $(\mathbf{y}_l)_i$ of the offspring $\mathbf{y}_l$ by randomly choosing *one* of the ith components from the μ parents $\mathbf{x}_m$ and subsequent addition of a $\mathcal{N}(0,\sigma^2)$ normally distributed random number. As in the $(\mu/\mu_I,\lambda)$ case as well as for (μ,λ)-strategies the new parents are produced by (μ,λ)-selection, sometimes called *truncation selection*, i. e. the μ best offspring are chosen.

2 The Asymptotic Progress Law of the $(\mu/\mu,\lambda)$-ES

2.1 The Limit $N\to\infty$ Progress Rate Formulae

The progress rate φ measures - roughly speaking - the expected distance change (in the parameter space) of the parent's center of mass from generation (g) to $(g+1)$. For the $(\mu/\mu_I,\lambda)$-ES on the hyper-sphere one finds [3, p. 90] the normalized progress rate

$$\varphi^\star_{\mu/\mu_I,\lambda}(\sigma^\star) = N\left(1-\sqrt{1+\frac{\sigma^{\star 2}}{\mu N}}\right) + \sigma^\star\, c_{\mu/\mu,\lambda}\,\frac{1+\frac{\sigma^{\star 2}}{2\mu N}}{\sqrt{1+\frac{\sigma^{\star 2}}{2N}}\sqrt{1+\frac{\sigma^{\star 2}}{\mu N}}} + \ldots \quad (1)$$

and for the $(\mu/\mu_D,\lambda)$ dominant case [3, p. 101]

$$\varphi^\star_{\mu/\mu_D,\lambda} = N\left(1-\sqrt{1+\frac{\sigma^{\star 2}}{N}}\right) + \frac{\sqrt{\mu}\, c_{\mu/\mu,\lambda}\,\sigma^\star}{\sqrt{1+\frac{\sigma^{\star 2}}{N}}} + \ldots, \quad (2)$$

with the $c_{\mu/\mu,\lambda}$-progress coefficient having an integral representation [3]

$$c_{\mu/\mu,\lambda} = \frac{\lambda-\mu}{2\pi}\binom{\lambda}{\mu}\int_{-\infty}^{\infty} \mathrm{e}^{-t^2}\left(\Phi(t)\right)^{\lambda-\mu-1}\left(1-\Phi(t)\right)^{\mu-1} dt \quad (3)$$

where Φ is the cdf (cumulative distribution function) of the standard normal variate

$$\Phi(t) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z=t} \mathrm{e}^{-\frac{1}{2}z^2}\, dz. \quad (4)$$

From the N-dependent progress rate (1) and (2) the asymptotic $N\to\infty$ formulae are easily obtained by Taylor expansion using $\sqrt{1+x} = 1+\frac{x}{2}+\mathcal{O}(x^2)$ and $(1+x)^{-\frac{1}{2}} = 1-\frac{x}{2}+\mathcal{O}(x^2)$. Thus one gets $\varphi^\star$ formulae which are *independent* of N (for $N\to\infty$; also obtained by Rechenberg [7])

$$\varphi^\star_{\mu/\mu_I,\lambda}(\sigma^\star) = c_{\mu/\mu,\lambda}\sigma^\star - \frac{\sigma^{\star 2}}{2\mu} \quad (5)$$

and

$$\varphi^\star_{\mu/\mu_D,\lambda}(\sigma^\star) = \sqrt{\mu}\, c_{\mu/\mu,\lambda}\sigma^\star - \frac{\sigma^{\star 2}}{2}. \quad (6)$$

These expressions can be used as approximations for the case $N < \infty$, if the conditions

$$\sigma^{\star 2} \ll N \qquad \text{and} \qquad \mu^2 \ll N \tag{7}$$

are fulfilled. Furthermore, they allow for an analytical calculation of the maximal achievable (normalized) progress rate $\hat{\varphi}^\star = \varphi^\star(\hat{\sigma}^\star) = \max_{\sigma^\star}[\varphi^\star(\sigma^\star)]$. By maximizing (5) and (6) one finds

$$\hat{\varphi}^\star_{\mu/\mu_I,\lambda} = \mu \, \frac{c^2_{\mu/\mu,\lambda}}{2} \qquad \text{at} \qquad \hat{\sigma}^\star_{\mu/\mu_I,\lambda} = \mu c_{\mu/\mu,\lambda} \tag{8}$$

and

$$\hat{\varphi}^\star_{\mu/\mu_D,\lambda} = \mu \, \frac{c^2_{\mu/\mu,\lambda}}{2} \qquad \text{at} \qquad \hat{\sigma}^\star_{\mu/\mu_D,\lambda} = \sqrt{\mu}\, c_{\mu/\mu,\lambda}. \tag{9}$$

As can be seen, in the asymptotic $N \to \infty$ limit both strategies have the same maximal achievable progress rate, however, the corresponding mutation strengths are different.

2.2 The $\lambda \to \infty$ Asymptotic of the $c_{\mu/\mu,\lambda}$ Progress Coefficient

Though the formulae (5), (6) and (8), (9) are rather simple, they contain the $c_{\mu/\mu,\lambda}$ coefficient which is the complicated integral (3). To see the influence of λ and μ on the ES performance the $\lambda \to \infty$ is investigated. In [3] an asymptotically *exact* $c_{\mu/\mu,\lambda}$ expression has been derived

$$c_{\mu/\mu,\lambda} \simeq \frac{1}{\vartheta} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left[\Phi^{-1}(1-\vartheta) \right]^2 \right\} \qquad \text{with} \qquad \vartheta := \frac{\mu}{\lambda} \tag{10}$$

which depends on the truncation ratio $\vartheta := \mu/\lambda$ only.[2] In this paper we will even go further and ask for the λ, μ order of (10). To derive this (new) order expression we resolve (10) for ϑ in $\Phi^{-1}(1-\vartheta)$

$$\vartheta \simeq 1 - \Phi\left(\sqrt{-\ln\left(2\pi\vartheta^2 c^2_{\mu/\mu,\lambda}\right)} \right). \tag{11}$$

For small ϑ ($\vartheta \to 0$) $\Phi(\cdots) \to 1$ does hold, therefore one can apply the asymptotic expansion of $\Phi(t)$ (see, e. g. [1, p. 85]) $\Phi(t) \simeq 1 - \exp(-t^2/2)/\sqrt{2\pi}\, t$ with the result

$$\begin{aligned} c_{\mu/\mu,\lambda} &\simeq \sqrt{-\ln\left(2\pi\vartheta^2 c^2_{\mu/\mu,\lambda}\right)} \\ c^2_{\mu/\mu,\lambda} &\simeq -\ln(2\pi) - 2\ln\vartheta - 2\ln c_{\mu/\mu,\lambda}. \end{aligned} \tag{12}$$

[2] N. B., $\Phi^{-1}(y)$ is the quantile of the standard normal variate, i. e., $\Phi^{-1}(y)$ is the inverse to $\Phi(t)$, equation (4).

Thus, we obtain for sufficiently large $c_{\mu/\mu,\lambda}$

$$c_{\mu/\mu,\lambda} \sim \sqrt{-2\ln\vartheta}, \qquad \text{that is,} \qquad c_{\mu/\mu,\lambda} = \mathcal{O}\left(\sqrt{\ln\frac{\lambda}{\mu}}\right). \tag{13}$$

This result is in accordance with the order relation of the $(1,\lambda)$-ES from [2, p. 171] $c_{1,\lambda} = \mathcal{O}(\sqrt{\ln\lambda})$, if $\mu = 1$ is chosen.

2.3 The Asymptotic Progress Law

Now we are able to ask for the scaling law of the $(\mu/\mu,\lambda)$-ES working in the vicinity of its performance optimum.[3] This question has gained some attention. Schwefel [10] conjectures:

> Based upon the observation that in the linear theory the convergence rate mainly depends on the ratio λ/μ if the population size is not too small, one might speculate about putting together what we know so far to the rather simple formula[4]
>
> $$\varphi^* \approx \mu\ln\frac{\lambda}{\mu}$$
>
> A proof is still missing, however.

His conjecture can be easily proved by the use of (13) and (8), (9). If we insert the asymptotic $c_{\mu/\mu,\lambda}$ (the left formula from (13)) into (8), (9), then we obtain

$$\boxed{\hat{\varphi}^\star \sim \mu\,\ln\frac{\lambda}{\mu},} \qquad \text{that is,} \qquad \boxed{\hat{\varphi}^\star = \mathcal{O}\left(\mu\,\ln\frac{\lambda}{\mu}\right).} \tag{14}$$

As can be seen, the progress rate increases logarithmically with λ whereas the μ influence can be decomposed in a linearly increasing part $\mu\ln\lambda$ and a nonlinear loss part $-\mu\ln\mu$. For small μ (let $\lambda = \text{const.}$) the linear part dominates. Therefore, for small μ the function $\hat{\varphi}^\star(\mu,\lambda)$ increases with μ, however, for $\mu \to \lambda$ it approaches zero. There must be an optimal $\hat{\vartheta} = \mu/\lambda$ ratio for which the $(\mu/\mu,\lambda)$-ES works with maximal efficiency. This is the question on the most efficient $(\mu/\mu,\lambda)$-ES. They cannot be answered by maximizing (14) directly, because (14) was obtained by neglecting terms in (12), i. e., for $\mu \ll \lambda$. If one ignores this fact and maximizes the left formula in (14) with respect to μ, then one finds $\lambda/\mu = \mathrm{e} \approx 2.718$, or for the truncation ratio $\hat{\vartheta} \approx 0.368$. This value is not so far from the exact result $\hat{\vartheta} \approx 0.270$ that will be derived in the next section.

3 This requires the control of σ in such a way that the optimal $\hat{\sigma}^\star$, given by (8) and (9), respectively, is roughly realized. Usually this is attained by σ-self-adaptation developed by Schwefel [9] or statistical inference methods (see Ostermeier et al. [6]).

4 N. B., Schwefel uses the φ^* to indicate the maximum of the normalized φ, whereas the author uses $(\cdot)^\star$ to label normalized quantities. Therefore $\hat{\varphi}^\star = \varphi^*$ holds.

2.4 The Fitness Efficiency η and the Optimal ϑ-Choice

Motivation and Definition of η. Besides the main goal of maximal progress towards the optimum there remains the question how efficient an ES algorithm performs the generational change. In order to have a "fair" measure, the progress rate $\varphi^\star$ is to be normalized with respect to the number of fitness evaluations within one generation. Since there are λ offspring there are λ fitness computations, thus the *progress rate per offspring* $\varphi^\star/\lambda$ is that "fair" performance measure.

As an example the $\varphi^\star/\lambda = f_\lambda(\sigma^\star)$ curves of the $(1,\lambda)$- and $(1+\lambda)$-ES (see Beyer [2]) are displayed in Figure 1. It is obvious that there is always an optimal

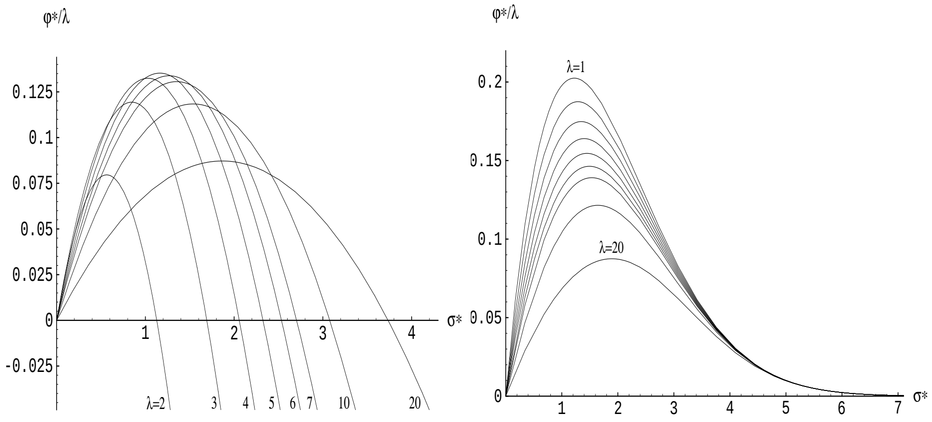

Fig. 1. The normalized progress rate per offspring for selected ES-variants. Left picture: the $(1,\lambda)$-ES. Right picture: the $(1+\lambda)$-ES ($\lambda = 1$, 2, 3, 4, 5, 6, 7, 10, and 20).

strategy [8]. In the case of the $(1,\lambda)$-ES the $\hat{\lambda} = 5$ variant has the highest maximum, whereas for the $(1+\lambda)$-ES $\hat{\lambda} = 1$ provides the largest progress per offspring.

In order to have a measure that quantifies this observation for each algorithm, the *fitness efficiency* η is defined

$$\text{FITNESS EFFICIENCY:} \qquad \eta := \frac{\max_{\sigma^\star}[\varphi^\star(\sigma^\star)]}{\lambda} = \frac{\hat{\varphi^\star}}{\lambda}. \tag{15}$$

In the case of the $(1,\lambda)$-ES, for example, one easily finds $\eta_{1,\lambda} = \eta(\lambda) = c_{1,\lambda}^2/2\lambda = c_{1/1,\lambda}^2/2\lambda$. This function is plotted in Figure 2. As can be seen, the most efficient strategy is obtained for $\hat{\lambda} \approx 5.02$.

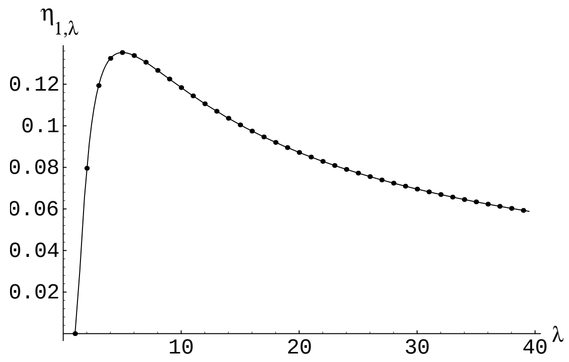

Fig. 2. The fitness efficiency of the $(1, \lambda)$-ES. One observes a maximum at $\lambda \approx 5$.

It is interesting to notice that the $\eta_{1+\lambda}$ of the $(1+\lambda)$-ES does *not* exhibit an maximum for $\lambda > 1$. That is, enlarging the number of offspring λ improves the progress rate, however, the fitness efficiency is reduced (as can be inferred from the right picture of Figure 1). The question arises whether there are strategies which do *not* exhibit such a strong efficiency degradation. Actually, they do exist – at least in the asymptotic limit – the $(\mu/\mu, \lambda)$-strategies.

The Fitness Efficiency of the $(\mu/\mu, \lambda)$-ES. In order to investigate η of the $(\mu/\mu, \lambda)$-ES in the asymptotic $\lambda \to \infty$ case, we insert (10) in (8) and (9). Thus $\eta_{\mu/\mu,\lambda}(\vartheta)$ becomes

$$\eta_{\mu/\mu,\lambda}(\vartheta) = \frac{1}{4\pi}\,\frac{1}{\vartheta}\,\exp\Big\{-\big[\Phi^{-1}(1-\vartheta)\big]^2\Big\}. \tag{16}$$

Figure 3 shows this function. As one can see, the maximal $\hat{\eta} = \eta(\hat{\vartheta}) = \max_\vartheta[\eta(\vartheta)]$ is achieved for $\hat{\vartheta} \approx 0.270$. This value should sound quite familiar to readers well acquainted with the $(1+1)$-ES theory (this is the optimal success probability, see [2] and below). Looking at the $\hat{\eta}$-value gives $\hat{\eta} \approx 0.202$ which is the maximum of $\varphi^\star_{1+1}$ (cf. Figure 1, right picture). This is *not* merely an incidence, but a deep connection that can be formally proved.

In order to uncover the connection, we first derive the equation for $\hat{\vartheta}$ from (16) by solving $\frac{d\eta}{d\vartheta}\Big|_{\vartheta=\hat{\vartheta}} = 0$ for $\hat{\vartheta}$

$$\frac{d\eta}{d\vartheta} \stackrel{!}{=} 0 \qquad \Rightarrow \qquad 1 = 2\hat{\vartheta}\sqrt{2\pi}\,\Phi^{-1}(1-\hat{\vartheta})\exp\left\{\frac{1}{2}\left[\Phi^{-1}(1-\hat{\vartheta})\right]^2\right\}. \tag{17}$$

If one substitutes $x := \Phi^{-1}(1-\hat{\vartheta})$, i. e.,

$$\hat{\vartheta} = 1 - \Phi(x) \tag{18}$$

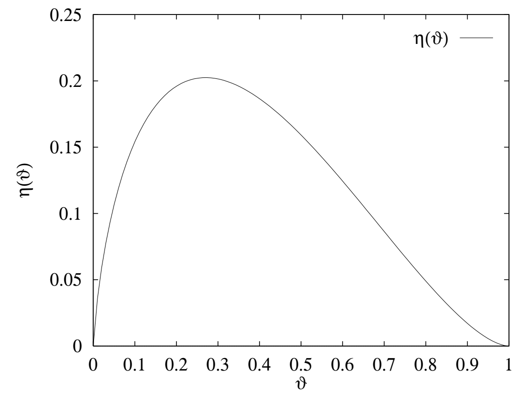

Fig. 3. The fitness efficiency $\eta_{\mu/\mu,\lambda}(\vartheta)$ in the asymptotic limit case as a function of the truncation ratio $\vartheta = \mu/\lambda$.

in (17), then one obtains after a simple re-arrangement the nonlinear equation

$$0 = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) - 2x\left[1 - \Phi(x)\right]. \tag{19}$$

This equation could be solved numerically for x yielding $\hat{\vartheta}$ by (18). Provided that x and therefore $\hat{\vartheta}$ are known, then $\hat{\eta}$ can be calculated. One can get an alternative $\hat{\eta}$ formula (needed below) if the exponential in (16) is expressed by (17)

$$\hat{\eta}_{\mu/\mu,\lambda} = 2\left[\Phi^{-1}(1-\hat{\vartheta})\right]^2 \hat{\vartheta}. \tag{20}$$

In a second step we maximize the $(1+1)$-ES progress rate formula [2, 7]

$$\varphi^\star_{1+1}(\sigma^\star) = \frac{\sigma^\star}{\sqrt{2\pi}} \exp\left(-\frac{1}{8}\sigma^{\star 2}\right) - \frac{\sigma^{\star 2}}{2}\left[1 - \Phi\left(\frac{\sigma^\star}{2}\right)\right]. \tag{21}$$

By differentiation one obtains

$$\frac{d\varphi^\star}{d\sigma^\star} \stackrel{!}{=} 0 \qquad \Rightarrow \qquad 0 = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{8}\hat{\sigma}^{\star 2}\right) - \hat{\sigma}^\star\left[1 - \Phi\left(\frac{\hat{\sigma}^\star}{2}\right)\right]. \tag{22}$$

Here $\hat{\sigma}^\star$ is the mutation strength which maximizes $\varphi^\star_{1+1}$. Now, let us compare (22) with (19). Both equations become equivalent, if one substitutes $x = \frac{\hat{\sigma}^\star}{2}$ in (22). For (18) one gets

$$\hat{\vartheta} = 1 - \Phi\left(\frac{\hat{\sigma}^\star}{2}\right). \tag{23}$$

Given the $\hat{\sigma}^\star$ one can compute $\hat{\varphi}^\star_{1+1}$ from (21). An alternative expression is obtained, if one substitutes the exponential term in (21) by means of (22). This gives $\hat{\varphi}^\star_{1+1} = \frac{1}{2}\hat{\sigma}^{\star 2}\,[1 - \Phi(\hat{\sigma}^\star/2)]$. Taking (23) into account yields finally[5]

$$\hat{\varphi}^\star_{1+1} = 2\left[\Phi^{-1}(1-\hat{\vartheta})\right]^2 \hat{\vartheta}. \tag{24}$$

Now we can compare (24) with (20) and obtain the remarkable result

$$\boxed{\hat{\eta}_{\mu/\mu,\lambda} = \hat{\varphi}^\star_{1+1}.} \tag{25}$$

That is, for the asymptotic case ($N \to \infty$, $\lambda \to \infty$) the fitness efficiency of the $(\mu/\mu, \lambda)$ multirecombinant ES becomes equal to the efficiency of the $(1+1)$-ES.

Besides this more or less formal relation between $(\mu/\mu, \lambda)$-ESs and the $(1+1)$-ES there is another remarkable connection. In the theory of the $(1+1)$-ES the *success probability*[6] $P_s(\sigma^\star)$ is given by the expression $P_s(\sigma^\star) = 1 - \Phi(\sigma^\star/2)$. Therefore, an optimally performing $(1+1)$-ES, i. e. $\sigma^\star = \hat{\sigma}^\star$, exhibits an optimal success probability $\hat{P}_s = P_s(\hat{\sigma}^\star)$

$$\boxed{\hat{P}_{s1+1} = 1 - \Phi\left(\frac{\hat{\sigma}^\star}{2}\right) = \hat{\vartheta},} \qquad \hat{P}_{s1+1} \approx 0.27027, \tag{26}$$

which is equal to the optimal truncation ratio $\hat{\vartheta} = \mu/\lambda$ of the $(\mu/\mu, \lambda)$-ES. This holds because of equation (23). Thus, the $\hat{\vartheta}$ value has an interesting interpretation. To obtain optimal performance for the $(1+1)$-ES the mutation strength σ has to be tuned in such a way that in average each $1/0.27 \approx 3.7$th offspring replaces the parent (this holds exactly for the sphere model). That is, *one* parent has to produce 3.7 offspring in order to get optimally replaced. In $(\mu/\mu, \lambda)$ strategies μ parents have to produce $\hat{\lambda} = \mu/\hat{\vartheta} \approx 3.7\mu$ offspring. It is as if the $(1+1)$-ES has get its "natural continuation" in the realm of multi-membered strategies. Note, the $(1+\lambda)$-ES cannot do the continuation (without performance degradation), because $\eta_{1+\lambda}$ is a decreasing function of λ. If we switch back to the maximum progress rate $\hat{\varphi}^\star$ using (15) we will find $\hat{\varphi}^\star_{1+\lambda}(\lambda)$ as a sub-linearly increasing function of λ, however, as to the multirecombinant strategies one finds

$$\hat{\varphi}^\star_{\mu/\mu,\lambda} = \lambda\hat{\eta}_{\mu/\mu,\lambda} \approx \lambda\cdot 0.20245. \tag{27}$$

3 The Dynamics of the Evolution Process and the ES Time Complexity

A prime goal of theoretical computer science concerns the computational complexity of the algorithms. Since this measure depends also on the complexity of

[5] N. B., this $\hat{\varphi}^\star_{1+1}$-formula must not be applied for values $\vartheta \neq \hat{\vartheta}$. The same holds for (20).

[6] The P_s is defined as the probability by which the parent is replaced by the offspring.

the fitness function (as far as optimization is considered), it is not well suited for comparing optimization algorithms with respect to their performance. A better measure for EAs (Evolutionary Algorithms) is usually the number of generations G needed to obtain a certain improvement under the assumption of an optimally working EA. The author will call this G the *EA* or *ES time complexity*.[7] Alternatively one can consider the number of fitness evaluation ν^F given by

$$\nu^F = \lambda G. \tag{28}$$

In order to compute G it is necessary to investigate the *macroscopic* behavior of the ES, i. e., the dynamics of the R change over the generations g. For large N this evolution is governed by the differential equation [2, 5] $\frac{dR(g)}{dg} = -R(g)\,\varphi^\star(g)/N$ which must be solved for $R(g)$. If we assume that there is a σ-control mechanism (see also footnote 3) which drives the ES into its optimal working regime,[8] then one can expect that $\varphi^\star(g) \approx \text{const.} \approx \hat{\varphi}^\star$ roughly holds. For this case the differential equation has the simple solution

$$R(g) = R(0)\,\exp\left(-\frac{\hat{\varphi}^\star}{N}\,g\right), \tag{29}$$

with the distance R to the optimum at generation g starting from an initial distance $R(0)$. If one asks for a certain relative improvement $R(g)/R(0)$, then one can determine the number of generations expected to reach this objective. From (29) we obtain with (15)

$$g = \frac{\lg(R(0)/R(g))}{\lg(\mathrm{e})}\,\frac{N}{\hat{\varphi}^\star} = \frac{\lg(R(0)/R(g))}{\lg(\mathrm{e})}\,\frac{N}{\lambda\eta}. \tag{30}$$

Applying this to the optimal $(\mu/\mu,\lambda)$-ES yields with (27)

$$G_{\mu/\mu,\lambda} = \hat{g}_{\mu/\mu,\lambda} = \frac{\lg(R(0)/R(g))}{\lg(\mathrm{e})\,\hat{\eta}_{\mu/\mu,\lambda}}\,\frac{N}{\lambda}. \tag{31}$$

Thus one obtains for the

$$\boxed{(\mu/\mu,\lambda)\text{-ES Time Complexity:} \qquad G_{\mu/\mu,\lambda} = \mathcal{O}\left(\frac{N}{\lambda}\right).} \tag{32}$$

That is, in multirecombinant ESs the (average) time complexity is of order N and inversely proportional to the population size λ. Equation (32) is a strong

[7] One might argue that such a measure neglects those parts of the ES-algorithm which are due to the selection, mutation, and the recombination operators. Furthermore, if parallel computers are under investigation, the communication overhead is to be taken into account. However, if the fitness calculation time is sufficiently large, then the other parts of the algorithm can be neglected. Developing more realistic performance models remains as a future task.

[8] This has been proved for the $(1,\lambda)$-σ-self-adaptation in [5] and it is observed in multi-parent strategies, too.

argument for parallelizing $(\mu/\mu, \lambda)$-ESs provided that the fitness calculation time does sufficiently exceed the communication time in the multiprocessor system under consideration.

The number of fitness evaluations ν^F becomes by virtue of (28) $\nu^F = \mathcal{O}(N)$. That is, in the asymptotic limit ($N \to \infty$, $\lambda \to \infty$, and assuming the validity of the sphere approximation) ν^F *cannot* be reduced by the multi-recombination, it is of the same order as for the $(1+1)$-ES.

It is interesting to compare these results with non-recombinant (μ, λ)-strategies. For the progress rate one finds [7, 4] $\varphi^\star = c_{\mu,\lambda}\sigma^\star - \sigma^{\star 2}/2$. The asymptotic behavior of the progress coefficient $c_{\mu,\lambda}$ can be obtained from the solution of an integral equation constituting some kind of nonlinear eigenvalue problem for the offspring distribution. Due to the space limitations the derivation must be omitted here, they will be published elsewhere. One finds for $\lambda \to \infty$ $c_{\mu,\lambda} \sim \sqrt{2\ln(\lambda/\mu)}$. Thus one obtains for the expected number of generations (30)

$$g = \frac{\lg(R(0)/R(g))}{\lg(\mathrm{e})} \frac{N}{\ln(\lambda/\mu)}. \tag{33}$$

As one can see, maximal performance (i. e. minimal g) is achieved for $\mu = 1$, therefore we get the

$$\boxed{(\mu, \lambda)\text{-ES Time Complexity:} \qquad G_{\mu,\lambda} = \mathcal{O}\left(\frac{N}{\ln \lambda}\right).} \tag{34}$$

Again the complexity scales linearly with the problem size N, however, increasing λ decreases $G_{\mu,\lambda}$ only logarithmically. Furthermore, due to (33) the $\mu = 1$ strategy is the fastest (μ, λ) variant,[9] its ν^F scales sub-linearly with the population size $\nu^F = \mathcal{O}\big(\frac{\lambda N}{\ln \lambda}\big)$.

4 Conclusions and Outlook

This paper has dealt with the asymptotic properties of $(\mu/\mu, \lambda)$-strategies. The advantage of such an analysis is that it yields simple analytical formulae which can be much better interpreted than the bulky results from the N-dependent theory presented in [3]. Of course, this is bought at the price of lower accuracy, if one tries to use these results as *approximations* for the real world case $N < \infty$, $\lambda < \infty$. For example, the optimal truncation ratio $\hat{\vartheta} \approx 0.27$ becomes a function $\hat{\vartheta} = f(N, \mu, \lambda)$, however, this function can only be computed numerically (a table for the optimal μ-choice has been presented in [3, p. 97]).

Having a theory on the ES time complexity, we are able to determine what to measure (cf. the introduction). Now it is the ES-experimenter's turn to look at their algorithms for the scaling laws described by (30) – (34). The question is,

[9] Note, this statement concerns the local performance of the ES. It does *nothing* say about the ES's ability to locate the global optimum in multi-modal fitness landscapes.

whether one can observe the scaling laws in (continuous) *real world optimization problems*, too.

In ES-experiments it is common practice to display the evolutionary progress by $\lg(R)$ vs. g plots. This allows for the comparison with (29). If (29) is fulfilled, i. e., if there is linear convergence the $\lg(R)$ vs. g plot should exhibit nearly straight lines. Indeed, this is often observed. However, the scaling behavior with respect to N and λ as indicated by (32) and (34) are up until now *not* common subjects of investigation. This will hopefully change with this article.

The analysis presented can be extended to other strategies. Results for the (μ, λ)-ES have been already obtained. They are to be published in a forthcoming paper. Furthermore, one might think about the behavior of $(\mu+\lambda)$- and $(\mu/\mu+\lambda)$-ESs. This is a task for future research.

5 Acknowledgement

The author is grateful to the anonymous referee # 13 for his helpful comments. This work was funded by the DFG, grant Be 1578/1-2.

References

1. M. Abramowitz and I. A. Stegun. *Pocketbook of Mathematical Functions*. Verlag Harri Deutsch, Thun, 1984.
2. H.-G. Beyer. Toward a Theory of Evolution Strategies: Some Asymptotical Results from the $(1,^{+}\lambda)$-Theory. *Evolutionary Computation*, 1(2):165–188, 1993.
3. H.-G. Beyer. Toward a Theory of Evolution Strategies: On the Benefit of Sex - the $(\mu/\mu, \lambda)$-Theory. *Evolutionary Computation*, 3(1):81–111, 1995.
4. H.-G. Beyer. Toward a Theory of Evolution Strategies: The (μ, λ)-Theory. *Evolutionary Computation*, 2(4):381–407, 1995.
5. H.-G. Beyer. Toward a Theory of Evolution Strategies: Self-Adaptation. *Evolutionary Computation*, 3(3):311–347, 1996.
6. A. Ostermeier, A. Gawelczyk, and N. Hansen. A Derandomized Approach to Self-Adaptation of Evolution Strategies. *Evolutionary Computation*, 2(4):369–380, 1995.
7. I. Rechenberg. *Evolutionsstrategie '94*. Frommann–Holzboog Verlag, Stuttgart, 1994.
8. H.-P. Schwefel. Adaptive Mechanismen in der biologischen Evolution und ihr Einfluß auf die Evolutionsgeschwindigkeit. Technical report, Technical University of Berlin, 1974. Abschlußbericht zum DFG-Vorhaben Re 215/2.
9. H.-P. Schwefel. *Evolution and Optimum Seeking*. Wiley, New York, NY, 1995.
10. H.-P. Schwefel and G. Rudolph. Contemporary Evolution Strategies. In F. Morana, A. Moreno, J. J. Merelo, and P. Chacon, editors, *Advances in Artificial Life. Third ECAL Proceedings*, pages 893–907, Berlin, 1995. Springer-Verlag.

This article was processed using the LaTeX macro package with LLNCS style