Some Observations on the Interaction of Recombination and Self-Adaptation in Evolution Strategies

Lothar Gruenz
Universität Dortmund
LS für Fertigungsvorbereitung
Leonhard-Euler-Str. 5
D-44227 Dortmund, Germany
email:gruenz@lfv.mb.uni-dortmund.de

Hans-Georg Beyer
Universität Dortmund
Informatik XI (Systemanalyse)
Joseph-von-Fraunhofer-Str. 20
D-44227 Dortmund, Germany
email:beyer@zappa.cs.uni-dortmund.de

Abstract- The performance of the multirecombinant $(\mu/\mu, \lambda)$-evolution strategy (ES) with σ-self-adaptation (σSA) is investigated on the sphere model. The investigation includes the computation of the maximal performance of an ES using recombination can reach when the mutation strength is optimally adjusted during the whole evolution. The comparison between the strategies with and without σSA shows that SA (self-adaptation) is not always able to drive the ES in its optimal working regime, although it still guarantees linear convergence order. The static and dynamic aspects of SA are discussed and it is shown that the learning parameter has a sensible influence on the progress rate.

1 Introduction

Evolution strategies (ES) are known as powerful optimization tools. The search process is based on mutation, recombination, and selection applied to a population of individuals. An individual $\vec{a} := (\vec{y}, \vec{s}, F(\vec{y}))$ consists of a set of object parameters $\vec{y} = (y_1, \ldots, y_N)^T$ to be optimized and a set of endogenous strategy parameters $\vec{s} = (\sigma_1, \ldots, \sigma_{N_s})$. The object parameters determine the fitness $F(\vec{y})$ of the individual. The strategy parameters are intended to control the dynamics of the ES in such a way that the algorithm exhibits maximal optimization performance. For example, using isotropic Gaussian mutations efforts to control the mutation strength σ (i.e. the standard deviation of the components of the mutation vector).

The convergence velocity depends on the mutation strength σ. Since its optimal value depends on the fitness landscape, it is necessary to adjust σ according to the local topology of the landscape to ensure convergence and maximal performance of the strategy. The strategy parameter set $\vec{s}$ is adapted by a second level of evolution called σ-self-adaptation (σSA). Each individual has its own strategy parameter set. By selecting the fittest individuals, well-adapted mutation strengths are chosen indirectly.

Experiments have shown that the SA yields the desired performance, but only a few theoretical investigations and detailed simulations exist.

The article investigates the SA on the sphere model with isotropic Gaussian mutation on the strategy parameter. First, the progress rate of ESs without SA is computed to determine the maximal performance of ESs using recombination. Next, the static and dynamic aspects of SA are discussed and the progress rate is compared with the strategy without SA. Further investigations concerning the learning parameter are performed to determine the maximal performance of a self-adaptive ES. In the final section the results are summarized and a short outlook is given.

2 Experiments

This section presents simulation results of the $(\mu/\mu, \lambda)$-ES on the sphere model. First the experiments on the standard ES with optimally adjusted mutation strength are discussed. The maximal performance is calculated using the normalized progress rate $\hat{\varphi}^*$. In the second part these results are compared with the progress rate of ESs with SA. Further investigation on the average mutation strength is performed to explain the observed performance loss. The influence of the learning parameter on the progress rate is analyzed.

Two different kinds of recombination are used: the intermediate recombination μ/μ_I and the dominant recombination[1] μ/μ_D. Using intermediate recombination the recombinant $\vec{y}^{(g+1)}$ is obtained by calculating the center of mass of the μ parental individuals

$$\vec{y}^{(g+1)} := \frac{1}{\mu} \sum_{m=1}^{\mu} \vec{y}^{(g)}_{m;\lambda}. \tag{1}$$

The notation $(.)_{m;\lambda}$ specifies the m-th best individual out of a population consisting of λ individuals. The dominant recombination produces the descendant by randomly choosing components from the μ individuals in the parent population

$$\forall l = 1, \ldots, \lambda : \vec{y}_l^{(g+1)} := \sum_{i=1}^{N} \left(\vec{e}_i^{\,T} \vec{y}^{(g)}_{m_i;\lambda} \right) \vec{e}_i, \qquad m_i = \mathtt{Random}\{1, \ldots, \mu\}, \tag{2}$$

based on the representation of the $\vec{y}$-vector by the orthogonal basis $\{\vec{e}_i\}$.

[1] Dominant recombination is also known as global discrete recombination [5].

2.1 Simulation conditions

Before discussing the experimental results a description of the simulation conditions is given. The fitness function to be minimized is the quadratic sphere model:

$$F(\vec{x}) = \sum_{i=1}^{N} (\hat{x_i} - x_i)^2, \tag{3}$$

where $\hat{x}$ is the position of the optimum usually chosen as $\hat{x_i} = (0, \dots, 0)^T$. The progress rate φ is defined as the expected value $\varphi := E\{R - \tilde{R}\}$, where R is the distance of the population's center $\langle \vec{y} \rangle^{(g)}$ to the optimum and $\tilde{R}$ that of $\langle \vec{y} \rangle^{(g+1)}$. $\langle \vec{y} \rangle^{(g)}$ is defined by

$$\langle \vec{y} \rangle^{(g)} := \frac{1}{\mu} \sum_{m=1}^{\mu} \vec{y}_m^{(g)}. \tag{4}$$

Using normalized quantities[2] φ^* and σ^* the normalized progress rate for the generations g_{start} to g_{end} can be calculated during the simulation

$$\varphi^* = \frac{1}{g_{end} - g_{start}} \sum_{g=g_{start}}^{g_{end}} \varphi_{local}^{*(g)} \tag{5}$$
$$\text{and} \quad \varphi_{local}^* = (R - \tilde{R}) \frac{N}{R}.$$

The population is initialized in two different ways. The first places the individuals uniformly distributed on the surface of a hypersphere with radius $R^{(0)}$. The second shifts the population by R on one coordinate: Each initial individual is generated from an N-dimensional random vector $\vec{z} = (z_1, \dots, z_N)^T$ where the z_i are normally distributed random numbers with a fixed standard deviation σ and zero mean. The starting populations are defined by

$$\vec{y} = (\frac{R}{\|\vec{z}\|} z_1, \dots, \frac{R}{\|\vec{z}\|} z_N)^T \tag{6}$$
$$\text{and} \quad \vec{y} = (z_1 + R, z_2, \dots, z_N)^T,$$

respectively. Although both initializations are totally different, they lead to the same results. If one allows for a sufficiently large g_{start}, the observed progress rates φ^* are almost identical.

2.2 Experiments without Self-Adaptation

This part is devoted to the progress rate analysis of recombinant ESs *without* SA on the sphere model. We are aiming at the determination of $\varphi^*(\sigma^*)$. In order to get $\varphi^*(\sigma^*)$ from evolution runs, Eq. (6) must be computed keeping σ^* constant. This is accomplished by tuning σ according to $\sigma^{(g)} = \sigma^* \cdot R^{(g)}/N$ (see Footnote 2) during the evolution run. As a result we obtain table functions of $\varphi^*(\sigma^*)$ (see Figure 1, below). They will be calculated for an interval $I_{\mu/\mu,\lambda}$ of σ^* values defined by $I_{\mu/\mu,\lambda} := [0 \dots \sigma_0^*]$ for which $\varphi^*(\sigma^*) \geq 0$ is fulfilled. That is, $I_{\mu/\mu,\lambda}$ defines the σ^* region where (linear) convergence of the ES is observed.

The φ^* function depends also on the exogenouos strategy parameters μ (parental population size) and λ (offspring size) as well as the parameter space dimension N.

Since SA aims to drive the ES into its optimal working regime, we are especially interested in the optimal μ and σ^* values, denoted by $\hat{\mu}$ and $\hat{\sigma}^*$ that provide maximal progress rate $\hat{\varphi}^*$ given λ and N

$$\hat{\varphi}^*(\lambda, N) := \max_{\sigma^*} \left[\varphi_{\hat{\mu}/\hat{\mu},\lambda}^*(\sigma^*, N) \right], \tag{7}$$
$$\hat{\sigma}^*(\lambda, N) := \arg\max_{\sigma^*} \left[\varphi_{\hat{\mu}/\hat{\mu},\lambda}^*(\sigma^*, N) \right].$$

We will compare the simulation results with the theoretical predictions [2].

2.2.1 Comparison of intermediate and dominant recombination using the same exogenous parameters

First, we will compare simulations with the same selection pressure and mutation strength for both dominant and intermediate recombinative ESs. The mutation strength σ^*, the size of the parent population μ, and the offspring population λ are chosen according to the theoretical results of Beyer [2, p. 97, Table 2]. Using the theoretically optimal mutation strength $\hat{\sigma}_I^*$ (Subscript I denotes intermediate recombination, whereas D refers to dominant recombination.) from the intermediate recombination for the dominant recombination an adjustment by $\sigma_D^* = \hat{\sigma}_I^* / \sqrt{\mu}$ is necessary.[3] Table 1 shows the simulation results.

μ	λ	σ^*	φ_I^*	φ_D^*
3	10	3.03	1.64	1.54
5	20	5.17	3.27	2.86
7	30	6.76	4.55	3.72
8	40	7.70	5.57	4.46
9	50	8.47	6.41	5.07
14	100	11.2	9.26	6.63
17	150	12.6	11.1	6.96
19	200	13.5	12.3	7.38
21	250	14.2	13.3	7.58

Table 1: Progress rates for the intermediate recombination with mutation strength σ^* and dominant recombination with mutation strength $\sigma^*/\sqrt{\mu}$ depending on population size in a 100-dimensional parameter space ($N = 100$). The φ_I^* values obtained are very close to the theoretical predictions in [2, p. 97].

Comparing the progress rates φ_I^* and φ_D^* a performance surplus for the intermediate recombination of the object parameters is observable. From this point of view, intermediate

[2] $\varphi^* := \varphi \frac{N}{R}$ and $\sigma^* := \sigma \frac{N}{R}$

[3] The relation between σ_D^* and σ_I^* arises from the surrogate mutation model in flat fitness landscapes and may be regarded as an upper bound on the surrogate mutation strength [2].

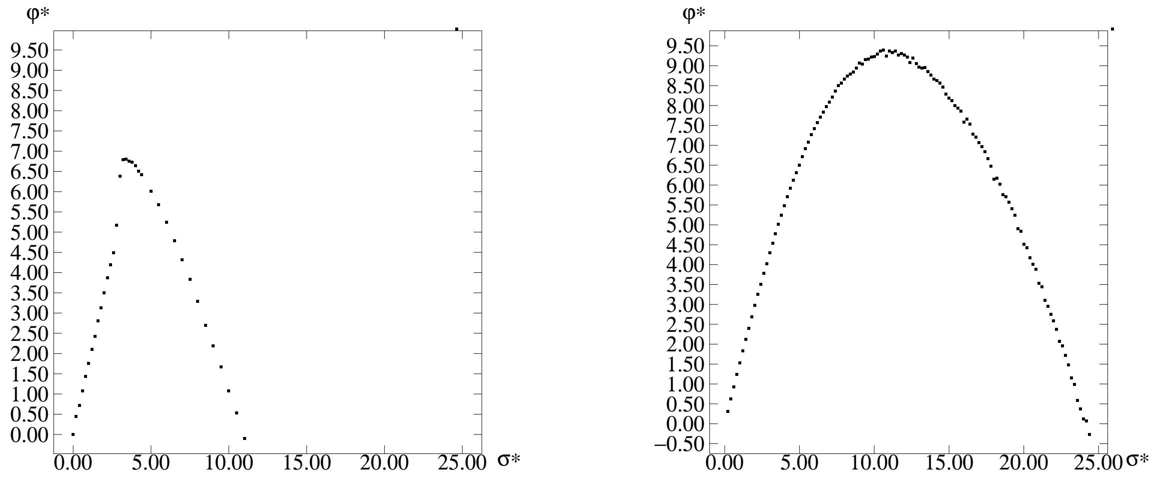


Figure 1: Comparison of the progress rates on the example of a $(14/14, 100)$-ES for a 100-dimensional parameter space. Left plot: dominant recombination, right plot: intermediate recombination

recombination should be preferred. The lack in performance can be due to an inappropriate mutation strength setting or an incorrect selection pressure, or both. Further simulations show that the parameters driving the intermediate recombinative ESs in its optimum working regime are not optimal for the dominant strategy. But, even if we know the optimal σ^*_D, it turns out that for $N < \infty$ the intermediate version exhibits better performance (see below).

2.2.2 The influence of the mutation strength on the dominant recombination

Varying the mutation strength σ^* in an interval $I_{\mu/\mu,\lambda}$ yields plots showing the relation between the progress rate and σ^*. As an example, the $\varphi^*(\sigma^*)$ functions of the $(14/14, 100)$-ES are plotted for both dominant and intermediate recombination in Figure 1.

For the intermediate strategy the plot is almost symmetric to the optimum $\hat{\sigma}^*$. Dominant recombination, however, produces $\varphi^*(\sigma^*)$ curves without this symmetry and a smaller interval $I_{\mu/\mu,\lambda}$ is observed. Although the latter can be expected from the surrogate mutation model mentioned before, the shape of the φ^*-curve comes as a surprise. In the first part of the interval the progress rate depends linearly on the mutation strength. Then, a *break* in the curve can be noticed with a very sharp increase of the progress rate up to the optimum. From the optimum to the zero of the progress rate one observes a nearly quadratic descent similar to the case of intermediate recombination.

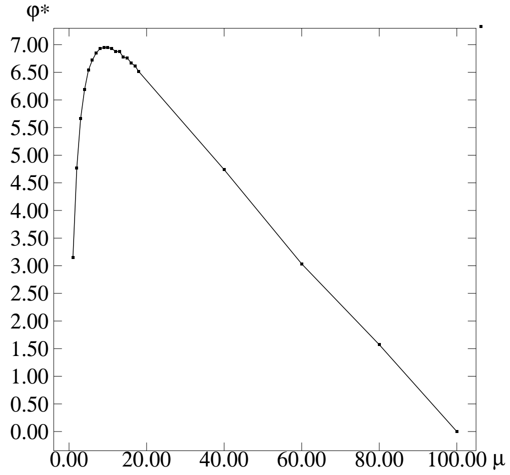
Further results for $\hat{\varphi}^*_D$ and $\hat{\sigma}^*_D$ on different population sizes are given in Table 2. These results can be compared with those in Table 1: If dominant recombination is used, on the one hand the relation between progress rate and mutation strength is different, on the other hand the performance does not reach the level of the intermediate recombination. The optimal mutation strength is smaller than the value given by the surrogate mutation model. That is, tuning the mutation strength, a small performance surplus is observed compared to the dominant strategy using the mutation strength $\sigma^*/\sqrt{\mu}$.

μ	λ	σ^*_D	φ^*_D
3	10	1.6	1.55
5	20	2.2	2.84
7	30	2.4	3.72
8	40	2.6	4.44
9	50	2.8	5.03
14	100	3.4	6.80
17	150	3.4	7.86
19	200	3.8	8.58
21	250	4.0	9.13

Table 2: Maximal progress rates and optimal mutation strength for dominant recombination depending on the population size in a 100-dimensional parameter space using the optimal selection pressure from the intermediate recombination.

2.2.3 The optimal exogenous parameter setting for the $(\mu/\mu_D, 100)$-ES

Next the influence of the selection pressure is investigated. The previous results are obtained under the assumption that the $\hat{\mu}$-value of the intermediate recombination was also well suited for the dominant recombinative ES. On the example of the $(\mu/\mu_D, 100)$-ES the optimal size of the parent population and the values of $\hat{\varphi}^*_D$ and $\hat{\sigma}^*_D$ (see Eq. (8)) are determined. Varying the size of the parent population yields plots shown in Figure 2. For each population size μ the optimal mutation strength is calculated and the maximal progress rate is plotted. The $(10/10_D, 100)$-ES achieves the maximal performance $\hat{\varphi}^*_D = 6.95$ with mutation strength $\hat{\sigma}^*_D = 3.4$.



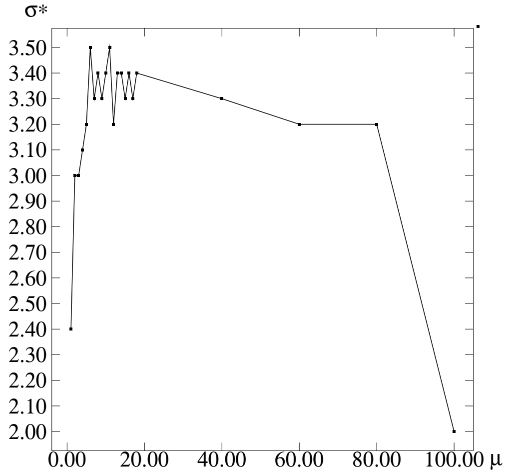


Figure 2: The maximal progress rate for the $(\mu/\mu_D, 100)$-ES with dominant recombination for $N = 100$. Left: progress rate with optimal mutation strength, right: optimal mutation strength

Although the optimal parental population size $\hat{\mu}_D = 10$ is smaller than $\hat{\mu}_I = 14$ (see Point 2.2.1), the relation between selection pressure and progress rate/mutation strength is similar to the intermediate recombinative ESs.

2.3 Experiments with σ-Self-Adaptation

The simulations in Section 2.2 were made under the assumption that the position of the optimum is known. So it was possible to use an optimal mutation strength during the whole optimization. The relation between the mutation strength σ^* and the progress rate φ^* for dominant and intermediate recombination on the object parameters was analyzed. Now the σ-self-adaptation is introduced and the effects of learning the mutation strength during the optimization are discussed. First, simulations on the static aspect of the adaptation process are performed. To explain the lack in performance further investigations on the dynamic aspect of SA are necessary. One aspect considered here concerns the learning parameter τ. It is analysed in detail yielding surprizing results on the relation between τ and φ^*.

2.3.1 Basics of σ-self-adaptation

There are four combinations of using recombination on the object and the strategy parameters. Early investigations by Schwefel [6] suggested to use dominant recombination on the object parameters and intermediate recombination on the strategy values intermediate. In this section both the intermediate-intermediate (II) [4] and the dominant-intermediate (DI) variant of σSA are considered. Dominant recombination on the strategy parameters is not considered here because it has provend to be less successful as to the attainable performance.

[4]This notation means that both the object and the strategy parameters are treated by intermediate recombination

Only one strategy parameter is used to determine the mutation strength on all axes of the R^N (isotropic mutations). The general notation of an individual reduces to $\vec{a} := (\vec{y}, \sigma, F(\vec{y}))$. The $(\mu/\mu, \lambda)$-ES with SA is given by Algorithm 1.

```
repeat
 recombine strategy:
  ∀l = 1...λ : a_l(σ)^(g+1) = a_m(R_s(σ))^(g), m = 1...μ
 recombine object:
  ∀l = 1...λ : a_l(y)^(g+1) = a_m(R(y))^(g), m = 1...μ
 mutate strategy:
  ∀l = 1...λ : a_l(σ)^(g+1) = a_l(M_s(σ))^(g+1)
 mutate object:
  ∀l = 1...λ : a_l^(g+1) = M(a_l^(g+1))
 select:
  ∀m = 1...μ : a_m^(g+1) = S(a_l^(g+1)), l = 1...λ
until stop criterion
```

Algorithm 1: The evolution loop of the SA algorithm. The notation $\vec{a}(.)$ specifies the component of the individual. R_s is the recombination operator on the strategy parameters, R works on the object parameters. The M_s function mutates the strategy parameter and M operates on the object parameters. S is the selection operator.

First the mutation operator on the strategy parameters is defined. We are using the *log-normal* operator which generates a continuous distribution of random numbers. It is based on the exponential transformation of a $\mathcal{N}(0, \tau^2)$ normally distributed random number

$$\tilde{\sigma} := \xi\sigma, \quad \xi := \exp(\tau\mathcal{N}(0,1)), \tag{8}$$

σ is the parental strategy parameter multiplied by a random number ξ achieving the value for the offspring. An exogenous

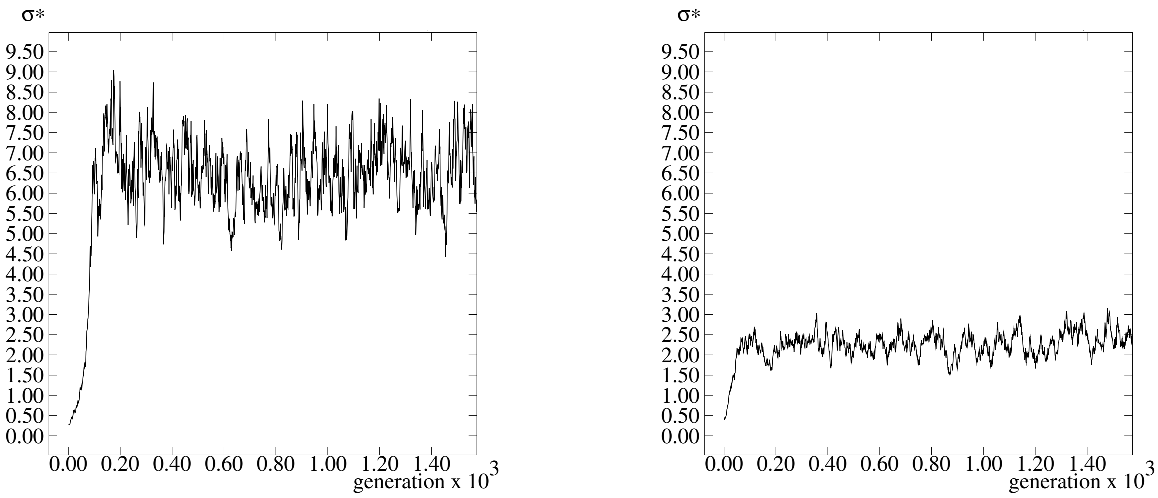


Figure 3: The mutation strength $\tilde{\sigma}^{*(g)}_{local}$ of a $(14/14, 100)$-ES. Left plot: dominant recombination of the object parameters, right plot: intermediate recombination.

parameter τ is needed to determine the mutability. Schwefel [4] suggested to set this *learning parameter* $\tau \propto 1/\sqrt{N}$. Beyer [3] has shown for the $(1, \lambda)$-strategy with sufficiently large λ that $\tau \simeq c_{1,\lambda}/\sqrt{N}$ with the progress coefficient $c_{1,\lambda}$. Based on this result it seems reasonable to choose the learning parameter for the multirecombination strategy as

$$\tau = \frac{c_{\mu/\mu,\lambda}}{\sqrt{N}}. \tag{9}$$

However, as we will see, this choice might be wrong.

2.3.2 Static aspects of the adaptation of the mutation strength

In the simulations the progress rate and the mutation strength are evaluated. The average mutation strength σ^*_{sa} is defined analogous to the progress rate

$$\sigma^*_{sa} := \frac{1}{g_{end} - g_{start}} \sum_{g=g_{start}}^{g_{end}} \tilde{\sigma}^{*(g)}_{local} \tag{10}$$

$$\text{and} \quad \tilde{\sigma}^{*(g)}_{local} := \frac{N}{\mu R^{(g)}} \sum_{m=1}^{\mu} \tilde{\sigma}^{(g)}_{m}.$$

The locally normalized mutation strength $\tilde{\sigma}^{*(g)}_{local}$ is defined as the average of the whole parental population. σ^*_{sa} is the expectation of the probability distribution of $\tilde{\sigma}^{*(g)}_{local}$ for sufficiently large g (steady state case). The initial value $\sigma = 1$ was chosen for generation $g = 0$.

The Table 3 shows the simulation results. Intermediate recombination is applied to the strategy parameters, the object parameters are recombined dominantly and intermediatly, respectively. For sufficiently large λ a performance surplus of the DI (dominant-intermediate) strategy can be observed. This is in agreement with the recommendations given in [1].

μ	λ	$c_{\mu/\mu,\lambda}$	$\sigma^*_{sa_I}$	$\sigma^*_{sa_D}$	φ^*_I	φ^*_D
3	10	1.07	2.03	2.16	1.33	1.13
5	20	1.21	2.35	3.64	2.08	1.93
7	30	1.29	2.37	4.39	2.47	2.45
8	40	1.38	2.20	4.77	2.57	3.06
9	50	1.44	2.21	5.16	2.74	3.47
14	100	1.59	2.22	6.58	3.14	4.65
17	150	1.70	2.26	7.32	3.46	5.53
19	200	1.77	2.30	7.72	3.71	6.22
21	250	1.81	2.35	8.18	3.93	6.69

Table 3: Progress rate and adapted mutation strength for N = 100 dimensions with learning parameter $\tau = \frac{c_{\mu/\mu,\lambda}}{\sqrt{N}}$. The progress coefficients are taken from [2].

Section 2.2.2 has shown that the progress rate depends strongly on the mutation strength (see Figure 1). Looking at Table 3 the mutation strength for the intermediate recombination of the object parameters is almost constant even when the number of individuals is raised. Only a slight increase of this value can be observed that is much too small. On the other hand for the dominant recombining ESs σ^*_{sa} increases with growing λ and the progress rate increases, too.

2.3.3 Dynamic aspects of the adaptation process

In the previous section only the static aspects of the self-adaptation were analyzed and the average of the mutation strength was considered based on the whole evolution. It is necessary to look at the mutation strength in detail. Figure 3 shows the mutation strength $\tilde{\sigma}^{*(g)}_{local}$ of the $(14/14, 100)$-ES during the evolution. A steady state adjustment of σ^* can be observed. After a certain transient time the mutation strength reaches a steady state domain and fluctuates around $\tilde{\sigma}^*_D \approx 6.5$ for the dominant recombination and $\tilde{\sigma}^*_I \approx 2.1$ for the inter-

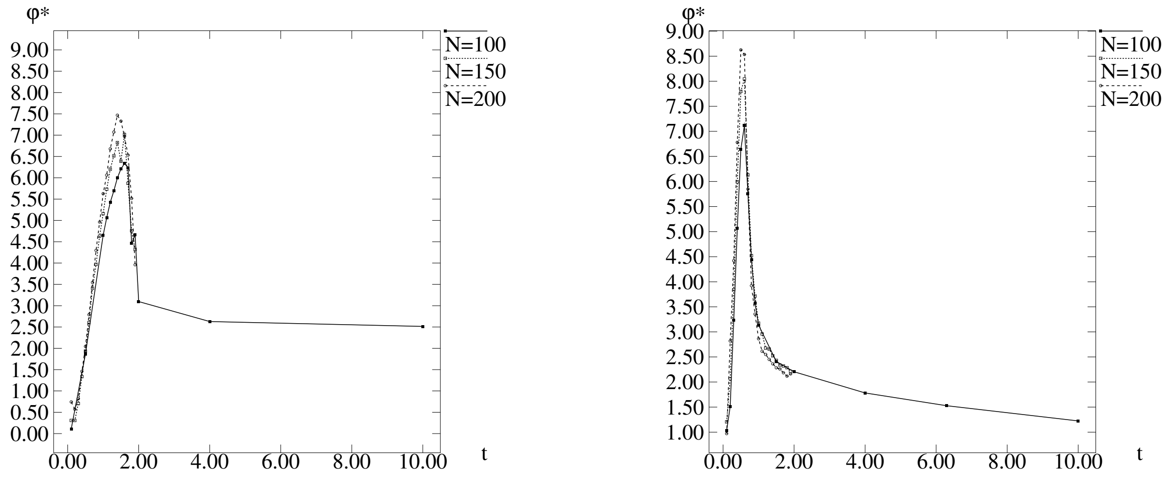


Figure 4: The maximal progress rate φ^* is sensitive to the parameter t. The example of a $(14/14, 100)$-ES with parameter space dimensions $N = 100$, $N = 150$, and $N = 200$ is depicted. Left plot: dominant recombination of the object parameters, right plot: intermediate recombination.

mediate one. $\tilde{\sigma}^*_D$ varies in an interval from 4.5 to 9 and $\tilde{\sigma}^*_I$ in an interval from 1.5 to 3. Considering Figure 1, $\tilde{\sigma}^*_D$ lies at the further end of $I_{\mu/\mu_D,\lambda}$ [5]. On the other hand $\tilde{\sigma}^*_I$ is placed at the beginning of $I_{\mu/\mu_I,\lambda}$.

In both cases the SA does not learn the optimal mutation strength. But, it is important to realize that both variants perform self-adaptation, because the $\tilde{\sigma}^*$ values are within the respective $I_{\mu/\mu_I,\lambda}$ intervals.

As in the case of the $(1, \lambda)$-σSA-ES [3] one observes large σ^* fluctuations. These fluctuations are mainly responsible for the performance loss in $(1, \lambda)$ strategies. Investigations on the influence of fluctuations of $\tilde{\sigma}^{*(g)}_{local}$ are necessary to make sure that the variation of the mutation strength during the evolution cannot be accounted for the performance degradation in the case considered here. To this end, the distribution of the observed (normalized) mutation strengths is taken as input for a table function $\varphi^*(\sigma^*)$ (displayed in Figure 1) to compute the theoretical average progress rate $\bar{\varphi}^*$ for the fluctuating $\tilde{\sigma}^{*(g)}$. The theoretical average progress rate is defined by:

$$\bar{\varphi}^* = \frac{1}{g_{end} - g_{start}} \sum_{g=g_{start}}^{g_{end}} \texttt{progress}(\tilde{\sigma}^{*(g)}), \qquad (11)$$

where `progress` is the table function that calculates the progress rate for a given σ^*. Comparing $\bar{\varphi}^*$ with $\varphi^*(\sigma^*_{sa})$ confirms that the fluctuation of the mutation strength has no serious influence on the progress rate. The lack in performance is due to the badly adjusted σ^*_{sa}.

[5] $I_{\mu/\mu_D,\lambda}$ is the interval of the mutation strength for ESs with dominant recombination, $I_{\mu/\mu_I,\lambda}$ is the interval for ESs with intermediate recombination (see also the definition of $I_{\mu/\mu,\lambda}$ at the beginning of Section 2.2).

2.3.4 The learning parameter τ

In the next step the influence of the learning parameter τ is analysed. Based on $\tau \simeq c_{1,\lambda}/\sqrt{N}$ a learning parameter formula is introduced

$$\tau = t\frac{c_{\mu/\mu,\lambda}}{\sqrt{N}}. \qquad (12)$$

The parameter t is introduced to tune the learning parameter. Although it has been shown in [3] that the performance of the $(1, \lambda)$-ES does not sensitively depend on t, we will see that this does not hold for recombinant strategies. The following plots show the steady state progress rate φ^* versus t for parameter space dimensions $N = 100$, $N = 150$, and $N = 200$ on the $(14/14, 100)$-ES.

Considering Figure 4 the newly introduced parameter t seems almost independent of the parameter space dimension: Using dominant recombination for the object parameters, $t_D = 1.6$ leads to the maximal progress rate $\varphi^*_{t_D=1.6} = 6.33$; for intermediate recombination of the object parameters, $t_I = 0.6$ is optimal achieving a progress rate $\varphi^*_{t_I=0.6} = 7.11$. Comparing the progress rates to the results of the previous simulations with $t = 1$ (see also Table 3) a performance surplus of 36.1% for dominant strategy and surplus of 126% for the intermediate strategy can be noticed. Furthermore using intermediate recombination on both object and strategy parameter yields better results.

In contrast to the $(1, \lambda)$-ES, the learning parameter τ has a strong influence on the progress rate in multirecombinant ESs. That is, only when τ is adjusted correctly the strategy can reach its maximal performance.

3 Summary and Outlook

The multirecombinant ES with σSA has been investigated in this article on the sphere model. For a standard ES with constant normalized mutation strength, the performance was computed in order to determine the maximal progress rate. Using these results, the standard SA was investigated and compared with the strategies using the optimal mutation strength during the whole evolution.

With respect to the $(\mu/\mu, \lambda)$-ES considered, the simulations have shown that self-adaptation has considerable problems to drive the ES into its optimal working regime. In contrast to SA on $(1, \lambda)$-ESs, the learning parameter τ influences the the steady state progress rate φ^* sensitively. The adjustment of the mutation strength is optimal only in a small τ interval and the strategy can nearly reach its theoretically maximal performance. Outside this interval, the steady state σ^*_{sa} appears far away from the optimal $\hat{\sigma}^*$ and performance losses are inevitable. However, under these conditions the fluctuations of σ^* have no serious influence on the progress rate, because σ^*_{sa} lies in the linear interval of the progress curve. This may be the reason for the stable working of SA-ESs in practical applications.

At this point it is important to recall that we have considered *optimal* SA performance. The failure of not reaching optimal performance does *not* mean that the SA does not work. As long as $\varphi^* > 0$, the SA works and guarantees on average linear convergence order in general.

This paper has presented some unexpected results on the behavior of multi-recombinant ES with self-adaptation. Due to the lack of a theory, getting a deeper understanding of these observations remains as a goal for future research. Furthermore, other recombination operators and SA with more than one strategy parameter are to be investigated in order to see whether the observations made are a common rule or rather an exception.

4 Acknowledgements

The authors would like to thank Kalyanmoy Deb, David B. Fogel, and A. Irfan Oyman for helpful comments. The second author (HGB) acknowledges support as Heisenberg Fellow of the Deutsche Forschungsgemeinschaft (DFG), grant Be 1578/4-1.

Bibliography

[1] T. Bäck and H.-P. Schwefel. An Overview of Evolutionary Algorithms for Parameter Optimization. *Evolutionary Computation*, 1(1):1–23, 1993.

[2] 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.

[3] H.-G. Beyer. Toward a Theory of Evolution Strategies: Self-Adaptation. *Evolutionary Computation*, 3(3):311–347, 1996.

[4] H.-P. Schwefel. *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie*. Interdisciplinary systems research; 26. Birkhäuser, Basel, 1977.

[5] 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.

[6] Hans-Paul Schwefel. Collective phenomena in evolutionary systems. In P. Checkland and I. Kiss, editors, *Problems of Constancy and Change — the Complementarity of Systems Approaches to Complexity, Papers presented at the 31st Annual Meeting of the Int'l Soc. for General System Research*, volume 2, pages 1025–1033, Budapest, 1.–5. Juni 1987. Int'l Soc. for General System Research.