On the “Explorative Power” of ES/EP-like Algorithms

Hans-Georg Beyer*

University of Dortmund, Department of Computer Science XI,
D-44221 Dortmund, Germany

Abstract. This paper discusses the question how ES/EP-like algorithms perform the evolutionary search in real-valued N-dimensional parameter spaces. It will be shown that the sometimes invoked model of a perturbed gradient search does not seem to give an appropriate picture of the search process. Instead, the search behavior is described as the antagonism of exploitation and exploration, where exploitation works in one dimension, whereas the exploration is a random walk on a $(N-1)$-dimensional manifold in the search space. As an example the exploration dynamics on the sphere model will be investigated.

1 Introduction

ES/EP-like evolutionary algorithms (EAs) have been proved to provide optimal or at least nearly optimal solutions to real-valued multi-dimensional optimization problems (for a comprehensive bibliography, see [1]). In contrast to their great success in real-world applications, their working mechanisms are analyzed and understood only partially. The aims and key issues of an analysis are (list is by no means exhaustive):

- How does the EA work at all?
- What is the advantage of a population concept (is it just for parallelization or is there more)?
- What is the influence/importance of the selection and the “genetic” operators (mutation and recombination)?

As far as the *local* behavior of (μ, λ)- and $(\mu/\mu, \lambda)$-ESs is considered, there was some progress in the last four years [10, 3, 2]. This progress concerns mainly the question of local performance measures, i.e. the expected generational change of the population toward an optimum. Such performance measures like the progress rate φ (measuring the expected distance change in the object parameter space), the quality gain $\overline{Q}$ (measuring the expected fitness change, also known as response to selection [9]), and the self-adaptation response [4] can be derived for simple fitness functions, mainly for the spherical model class.

* e-mail: beyer@ludo.cs.uni-dortmund.de

One side effect of these derivations is that they also provide a deeper insight into how EAs work locally, giving rise to some *basic EA working principles* recently published in [6]. These principles are the

1. evolutionary progress principle (EPP)
2. genetic repair hypothesis (GR)
3. mutation-induced speciation by recombination hypothesis (MISR).

The first principle applies to all EA-like strategies; we will come back to it in the next section. The second and the third apply to those strategies that use recombination. It is important to notice that these principles explain the effect of the operators with respect to the generational change of suitably defined progress measures. They do not explain how the EA is (or is not) able to locate a global optimum. This is so because the local progress measures (φ, $\overline{Q}$, etc.) are defined to quantify the "exploitative power" of the EA. The explorative behavior and its quantification, however, has received only little interest up to now. Such questions as

- Can the population escape from a local optimum attractor?
- Can we quantify the EA's ability to leave such an attractor?

are of real importance. A first attempt to evaluate the effect of different selection techniques ("hard" and "soft" selection) was made in [7]. That work puts emphasis on the ability to escape from an attractor, leaving open the question of how the population may get trapped in a better attractor. Though this is one of the most important problems, we will not be able to give a definite answer here. Instead, an explanation will be given how ES/EP-like strategies *explore* the search space. The intention is to give a feeling how the population or the parents are placed in the fitness landscape. For this purpose we will shortly re-examine the EPP (evolutionary progress principle) emphasizing the relation between exploitation and exploration. This will be done in Section 2. In Section 3 we will analyze the explorative behavior of the ES using the N-dimensional sphere as the model.

2 The EPP and the Exploitation/Exploration Behavior of EAs

Our current understanding of how an EA explores the search space is unsatisfactory. To get a feeling on how the optimization process proceeds in the search space, one can observe the sampling points $\mathbf{y}^{(g)}$ during the evolution (g is the generation counter). For example, if we consider a real-valued N-dimensional space and a population of μ parents and λ offspring, it becomes clear that there is no simple way to get a realistic "picture" of the evolution process. Due to our restriction to think and live in three dimensions, we are forced to project down from the $N \times (\mu|\lambda)$ dimensions onto two or three dimensions. This projection, however, if wrongly applied, can produce misleading pictures of the search process. One of these misleading pictures sometimes invoked to explain the working of the ES is similar to the one of a *gradient-path-following population* [10]. There are two clues that seem to support that picture:

1. The ES on the sphere model exhibits linear convergence order just like a gradient strategy. Its time complexity is $\mathcal{O}(N)$.
2. If one projects the N-dimensional individual $\mathbf{y} := (y_1, \ldots, y_N)^{\mathrm{T}}$ into a (x_1, x_2)-plane by [10]

$$x_1 := \sqrt{y_1^2 + \cdots + y_{(N/2)}^2}\,, \quad x_2 := \sqrt{y_{(N/2)+1}^2 + \cdots + y_N^2}, \quad (N \text{ even}) \quad (1)$$

   one actually observes a gradient like behavior, as displayed in Fig. 1.

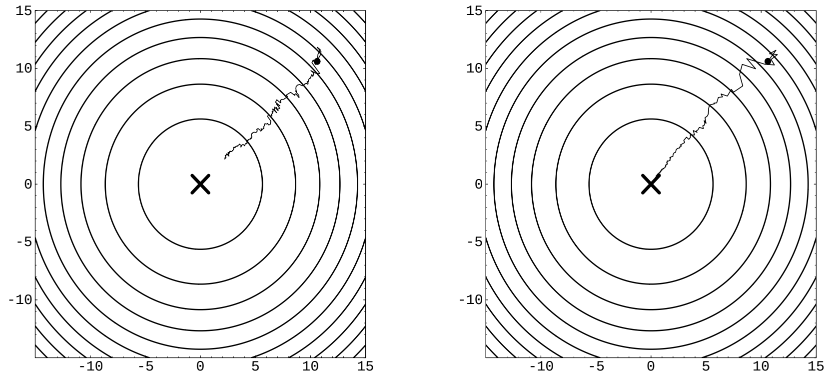


Fig. 1. Path of the *best* individual in a spherical model produced by a non-recombinant $(4, 20)$-ES (left picture) and a multirecombinant $(4/4, 20)$-ES (right picture) projected from the $N = 100$ dimensional parameter space onto two dimensions using Eq. (1). "$\bullet$" is the starting point, "$\times$" is the optimum. The horizontal axis displays x_1 and the vertical x_2. The evolution runs over $g = 200$ generations.

Figure 1 seems to strongly motivate the idea that the ES traces the local gradient with some minor random perturbations. If this picture were correct then there were only a very small probability to escape from the local attractor: the gradient field would define the stream lines guiding the individuals to the local fixed point. As a consequence, one could discard such strategies totally and go back to gradient strategies using the multi-start approach [11] of global optimization.

In contrast to the projection (1) the actual path of two of the N coordinates from the best parent over the generations (phase diagram) is displayed in Fig. 2. As one can see, the actual trajectories of the ESs are by no means gradient driven but resemble much more a random walk. However, this random walk is

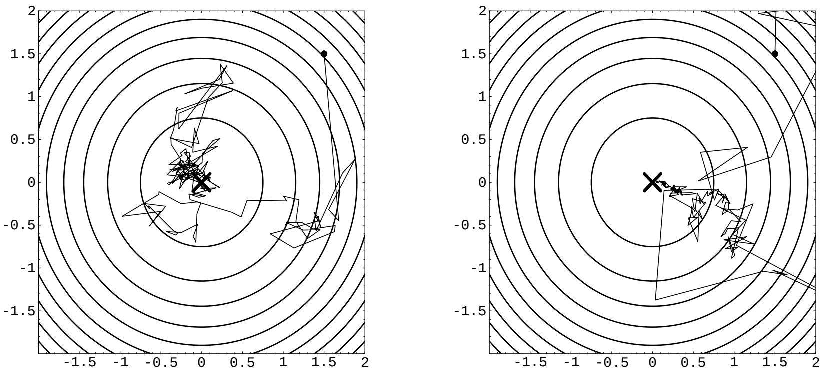


Fig. 2. The $x_1 := y_1$ and $x_2 := y_2$ coordinates (x_1 horizontal axis, x_2 vertical axis) of the path of the *best* individual from the ES-runs of Fig. 1 are displayed. The $(4, 20)$-ES (left picture) as well as the $(4/4, 20)$-ES (right picture) produce strong erratic trails in the search space. Note, instead of $x_1 := y_1$, $x_2 := y_2$ one could as well arbitrarily choose, e.g., $x_1 := y_{10}$ and $x_2 := y_3$ to obtain "similar" pictures.

restricted due to the selection and the "genetic" operators applied to the population. The usual way to characterize the effect of these operators is to measure the population's change toward the optimum from generation g to generation $g + 1$, known as the progress rate φ.

There is a more or less comprehensive progress rate theory for the sphere model and there are also first results for the parabolic ridge function (to be published). From those results the *evolutionary progress principle* (EPP) has been inferred [6]:

$$\boxed{\begin{array}{c}\text{EVOLUTIONARY}\\\text{PROGRESS}\end{array} = \begin{array}{c}\text{PROGRESS}\\\text{GAIN}\end{array} - \begin{array}{c}\text{PROGRESS}\\\text{LOSS}\end{array}}$$

That is, progress is the result of two opposite forces producing gain and loss, respectively. The gain and loss parts for the sphere model are displayed in Table 1. A thorough analysis of the derivations (which can be found in [2, 3]) reveals that the gain is mainly influenced by the selection operator, whereas the loss can be reduced by the recombination operator.[1] Both parts depend on the mutation strength that must be chosen appropriately to maximize the local evolutionary progress. But most importantly, it is possible to relate the gain/loss parts in the φ formulae to certain components of the mutation vectors $\mathbf{z}$ producing the λ

[1] It is important to note that the loss reduction effect has been proved for the so-called intermediate and the dominant *multirecombination* in real-valued parameter spaces. Whether there is a similar effect for one-point crossover, often used in GAs, is still an open question.

	(μ,λ)	$(\mu/\mu_I,\lambda)$	$(\mu/\mu_D,\lambda)$
gain	$c_{\mu,\lambda}\sigma^*$	$c_{\mu/\mu,\lambda}\sigma^*$	$\sqrt{\mu}c_{\mu/\mu,\lambda}\sigma^*$
loss	${\sigma^*}^2/2$	${\sigma^*}^2/(2\mu)$	${\sigma^*}^2/2$

Table 1. Gain and loss parts of the progress rate formulae in the asymptotic case ($N \to \infty$): σ^* – normalized mutation strength, $c_{\mu,\lambda}$ and $c_{\mu/\mu,\lambda}$ – progress coefficients, $_I$ – intermediate recombination, $_D$ – dominant (discrete) recombination (for definitions, see [2, 3]).

offspring from the μ parents. As have been shown [6], a mutation $\mathbf{z}$ can always be decomposed into a component in progress direction[2] and in a perpendicular $\mathbf{h}$. If the progress is defined as the reduction of the Euclidean distance R to the global optimum $\hat{\mathbf{y}}$, see Fig. 3, then the decomposition reads

$$\mathbf{z} = -x\mathbf{e}_R + \mathbf{h} \tag{2}$$

and the gain part in the EPP is related to x, whereas the loss part is mainly determined by the length of the $\mathbf{h}$-vector.

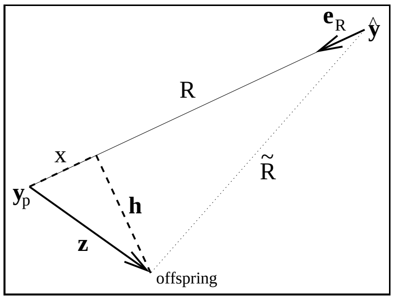
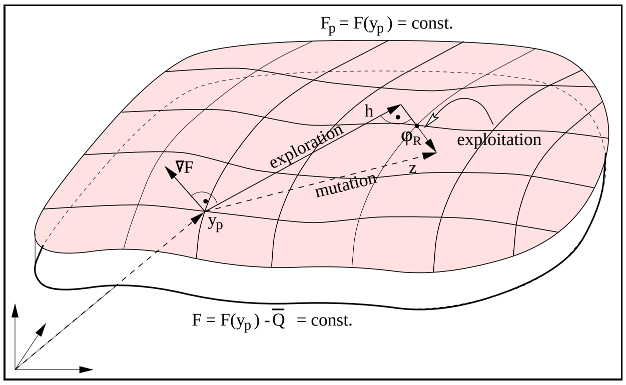


Fig. 3. Decomposition of the $\mathbf{z}$ mutation in a part x in optimum direction and a perpendicular $\mathbf{h}$-vector. R is the parental distance $\|\mathbf{y}_p - \hat{\mathbf{y}}\|$, and $\tilde{R}$ measures the distance of the offspring to the optimum $\hat{\mathbf{y}}$.

It is quite clear that such a decomposition of $\mathbf{z}$ is always possible (even in the case of non-spherical fitness landscapes); however, it is not the only possibility. One might as well decompose $\mathbf{z}$ into a component in direction of the local gradient ∇F ($F(\mathbf{y})$ is the function to be optimized) and a perpendicular $\mathbf{h}$. Such a decomposition is of special interest if we want to quantify the relation between the local "exploitation" and the "exploration." If we interpret "exploitation" as the ability to step into the local gradient direction and "exploration" as the ability to leave the gradient path then we might have a way to understand the evolutionary search behavior. This is depicted in Fig. 4. This decomposition sheds a new light on the question how the EAs in real-valued spaces perform the evolutionary search. The search is *not* simply path oriented. Given a (fixed) parent vector $\mathbf{y}_p$ with fitness $F_p := F(\mathbf{y}_p)$, the hypersurface $F(\mathbf{y}) = F_p = \text{const.}$

[2] This direction is not necessarily the globally optimal direction, it depends on the definition of the progress.

Fig. 4. On the hypersurface of the fitness landscape $F(\mathbf{y}) = F_p$ a mutation $\mathbf{z}$ is performed from the parental point $\mathbf{y}_p$. The mutation can be decomposed locally into a part parallel to the local gradient $\nabla F(\mathbf{y}_p)$ and a part lying in the tangential hyperplane.



forms a $(N-1)$-dimensional manifold in which the $\mathbf{z}$ components are randomly placed. That is, the EA performs a random walk in that manifold. Each component on this hypersurface is selectively neutral. Selection is only in one dimension which is normal to the hypersurface. Exploitation is thus of the magnitude φ, whereas exploration is of magnitude $\|\mathbf{h}\|$. One way of relating exploitation and exploration is to determine its ratio. By simple calculations one finds for isotropic Gaussian mutations

$$\frac{\text{exploitation}}{\text{exploration}} = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) . \qquad (3)$$

As one can see for isotropic Gaussian mutations, the strategies spend $\sqrt{N}$ times more effort to explore the search space than to follow the gradient path.

3 The Exploration Dynamics on the Sphere Model

The ratio (3) gives us a vague feeling how exploration and exploitation are related to each other. To get a deeper insight and also for sake of simplicity, the exploration dynamics on the sphere model is investigated. Due to space limitations, the derivations must be omitted here (see [5]), only the ideas and the results for the (μ, λ)- and the $(\mu/\mu_I, \lambda)$-ES will be presented.

3.1 The Angular Dynamics and the (μ, λ)-ES

Since the exploration dynamics evolves on the $(N-1)$-dimensional hypersphere surface, we are faced again with the problem of finding a suitable projection from the $(N-1)$ dimensions onto one or two dimensions assessable by human beings. Our goal is to quantify how fast the population members spread over the surface starting from an initial parent $\mathbf{R}^{(0)} = \mathbf{R}_\star = (0, 0, \ldots, R_0)^{\mathrm{T}}$. (For sake of simplicity, the coordinate system is located in the center of the sphere.) Since the surface is selectively neutral, it suffices to trace the evolution of one individual $\mathbf{R}^{(g)}$. For a fixed g, $\mathbf{R}_\star$ and $\mathbf{R}^{(g)}$ span a two-dimensional plane and one can ask about the angle $\alpha^{(g)}$ between them (see Fig. 5).

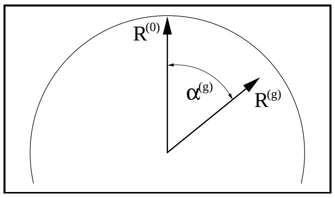


Fig. 5. On the definition of the angular dynamics.

Using the definition of the scalar product one obtains

$$\alpha^{(g)} = \arccos\left[\mathbf{R}_\star^{\mathrm{T}}\mathbf{R}^{(g)}/(\|\mathbf{R}_\star\|\,\|\mathbf{R}^{(g)}\|)\right]. \tag{4}$$

Though this formula can be easily applied to get the $\alpha^{(g)}$-dynamics from ES-experiments, an analytical calculation of $\alpha^{(g)}$ is a rather difficult task. It comprises mainly four steps.

First one has to introduce a spherical coordinate system (r, γ_i) [8] in which the Cartesian reads

$$y_j = r\left(\prod_{k=1}^{j-1}\sin\gamma_k\right)\cos\gamma_j, \qquad y_N = r\left(\prod_{k=1}^{N-2}\sin\gamma_k\right)\sin\gamma_{N-1} \tag{5}$$

(NB, $1 \le j \le N-1$) with

$$0 \le \gamma_k < \pi, \quad \text{for} \quad 1 \le k \le N-2, \qquad \text{and} \qquad 0 \le \gamma_{N-1} < 2\pi. \tag{6}$$

Thus, the evolution of the random walk on the hypersphere with (fixed) radius R (i.e., $r = R$) is mapped onto the angular space

$$[0,\pi)^{N-2} \times [0, 2\pi). \tag{7}$$

In the second step, one has to transform the $\mathcal{N}(0,\sigma^2)$ normally distributed components of the mutation vector $\mathbf{z}$ into mutations in the angular space. Due to the statistical independence of the mutation components one obtains by Taylor expansion for sufficiently small σ

$$\tilde{\gamma}_i \approx \mathcal{N}\left(\gamma_i, \frac{\sigma^2}{R^2}\right) = \mathcal{N}\left(\gamma_i, \frac{\sigma^{*2}}{N^2}\right), \quad \text{with} \quad \sigma^* := \sigma\frac{N}{R}. \tag{8}$$

That is, the ith angle having the parental value γ_i is mutated by a Gaussian with standard deviation σ/R and zero mean.

In the third step, we investigate the random walk of a single γ_i coordinate. For a small number of generations this is simply a Gaussian diffusion process obeying a $\sqrt{g}$ law, i.e., the angular distribution disperses over the angular space $[0, \pi)$ (and $[0, 2\pi)$, respectively) with a normal distribution and standard deviation $s^{(g)}$

$$s^{(g)} = \sqrt{g}\,\frac{\sigma^*}{N}. \tag{9}$$

It is important to realize that this law has only restricted applicability. Due to the periodicity of the angular space, values outside of $[0, \pi)$ are mapped back modulo π (for $i = N-1$: modulo 2π). The result is that for $g \to \infty$ the space $[0, \pi)$ is uniformly sampled and the Gaussian distribution has changed into a uniform one. For this latter case one can ask for the expected angular difference $\overline{\Delta\gamma} := \mathrm{E}\{|\gamma - \gamma'|\}$ of two randomly chosen $\mathbf{R}$-vectors. This yields

$$g \to \infty: \qquad \overline{\Delta\gamma_i} = \frac{\pi}{3} \quad (i \le N-2) \qquad \text{and} \quad \overline{\Delta\gamma_{N-1}} = \frac{2\pi}{3}. \tag{10}$$

In a fourth step we are able to estimate now the $\alpha^{(g)}$-dynamics for both the small g case and for $g \to \infty$. For small g we have to consider the joint density of the $\gamma_i^{(g)}$. Let $\gamma_i^{(0)} = 0$, then the density at generation g reads

$$p\Big(\gamma_1^{(g)}, \gamma_2^{(g)}, \ldots, \gamma_{N-1}^{(g)}\Big) = \frac{1}{\big(\sqrt{2\pi}\; s^{(g)}\big)^{N-1}} \exp\left[-\frac{1}{2}\frac{\sum_{i=1}^{N-1}\Big(\gamma_i^{(g)}\Big)^2}{\big(s^{(g)}\big)^2}\right]. \tag{11}$$

These γ_i build up the $\alpha^{(g)}$ angle defined by Eq. (4) in a nonlinear fashion. Again, we consider small γ. In such a case, the $\alpha^{(g)}$ is simply the geometric sum of the γ_i. This follows because both the Cartesian and the spherical system have orthogonal coordinate lines, i.e., their metric tensors are diagonal. Therefore, we obtain for small γ_i

$$\alpha^{(g)} \approx \sqrt{\sum_{i=1}^{N-1}\Big(\gamma_i^{(g)}\Big)^2}. \tag{12}$$

Since the $\gamma_i^{(g)}$ are normally distributed, the right hand side of Eq. (12) is χ-distributed. For large N ($N \to \infty$) the expected value of the χ-distribution is asymptotically equal to "standard deviation time $\sqrt{N}$." Thus, the expected value $\overline{\alpha}^{(g)} = \mathrm{E}\{\alpha^{(g)}\} \approx \sqrt{N} s^{(g)}$ and with (9) it follows

$$\boxed{\overline{\alpha}^{(g)} \approx \sqrt{g}\,\frac{\sigma^*}{\sqrt{N}}} \qquad \text{(for } g \text{ sufficiently small).} \tag{13}$$

The opposite case, $g \to \infty$, yields an $\alpha_\infty := \lim_{g\to\infty} \overline{\alpha}^{(g)}$ that can be determined by estimating the expectation of the formula

$$\alpha^{(g)} = \arccos\left(\prod_{k=1}^{N-1} \sin\gamma_k^{(g)}\right) \tag{14}$$

which is obtained by substitution of $\mathbf{R}_\star$ and $\mathbf{R}^{(g)}$ in Eq. (4) using the transformation (5). If Taylor expansion is applied to Eq. (14) neglecting higher-order terms then one finds that α_∞ is expected to be in the vicinity of 90^o

$$\alpha_\infty \approx \frac{\pi}{2}. \tag{15}$$

The predictions of Eq. (13) and (15) can be verified by ES experiments. Figure 6 shows the $\alpha^{(g)}$- and the $R(g)$-dynamics of an *optimally tuned* (i.e. $\sigma^* = c_{\mu,\lambda}$), $N = 100$, $(5, 30)$-ES starting at $\mathbf{y}^{(0)} = \mathbf{R}_\star = (0, 0, \ldots, R_0)^{\mathrm{T}}$ with $R_0 = 100$. The plots have been obtained by averaging over 76 independent ES runs.

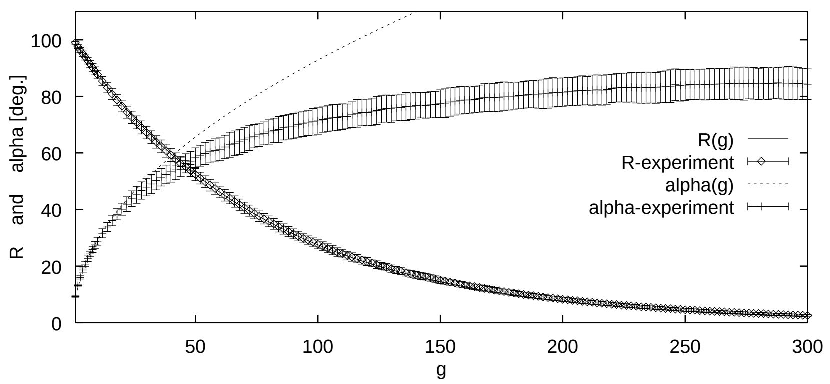


Fig. 6. The α- and R-dynamics of an *optimally tuned* $(5, 30)$-ES. The experimental data are displayed by error bars indicating the standard deviation over 76 independent ES runs. As one can see, the theoretical $R(g)$ curve, Eq. (16), is covered by the error bars of the experiments. As to the $\alpha(g)$ (dashed curve), only a few generations can be predicted by Eq. (13).

As to the $\alpha^{(g)}$ one can see that the theoretical formula (13), displayed as the dashed curve, predicts the dynamics satisfactorily up to angles of $\approx 45^o$. For large g, however, the estimate (15) does hold roughly.

The $R(g)$-dynamics, i.e. the exploitation toward the optimum, is also displayed in Fig. 6. It can be theoretically calculated as

$$R(g) = R_0 \exp\left[-\frac{\varphi^*(\sigma^*)}{N} g\right], \qquad \hat{\varphi}^*(\sigma^*) = \frac{c_{\mu,\lambda}^2}{2} \tag{16}$$

with $\varphi^* = \hat{\varphi}^*_{(5,30)} \approx 1.29$. We will come back to this in Section 3.3.

3.2 The Angular Dynamics of the $(\mu/\mu_I, \lambda)$-ES

The calculations for the intermediate $(\mu/\mu, \lambda)$-ES are similar to those of the (μ, λ)-ES except that we now have to consider the center of mass $\langle \mathbf{y} \rangle^{(g)}$ of the parents. That is, the $\mathbf{R}$-vectors in Eq. (4) are to be replaced by the arithmetic mean of the μ parents $\mathbf{y}_m$, $\langle \mathbf{y} \rangle^{(g)} = \frac{1}{\mu} \sum_{m=1}^{\mu} \mathbf{y}_m^{(g)}$. Due to this averaging procedure, the variance of the center of mass parent (which is produced by the selected mutations) is reduced by a factor of μ. Thus, the effective normalized mutation strength σ^* reduces by a factor of $\sqrt{\mu}$ (since it is the standard deviation). Therefore the σ^* in Eq. (9) must be replaced by $\sigma^*/\sqrt{\mu}$. This gives with Eq. (13)

$$\boxed{\overline{\alpha}^{(g)} \approx \sqrt{g}\,\frac{\sigma^*}{\sqrt{\mu\, N}}} \qquad \text{for} \quad \overline{\alpha}^{(g)} \lessapprox \frac{\pi}{4}. \tag{17}$$

The predictions of Eq. (17) will be compared with ES experiments. In Fig. 7 the α- and R-dynamics of the $(5/5_I, 30)$-ES are depicted. The graph of Eq. (17) is displayed as a dashed curve. Additionally $R(g) = \|\langle \mathbf{y} \rangle^{(g)}\|$ is displayed. Initially

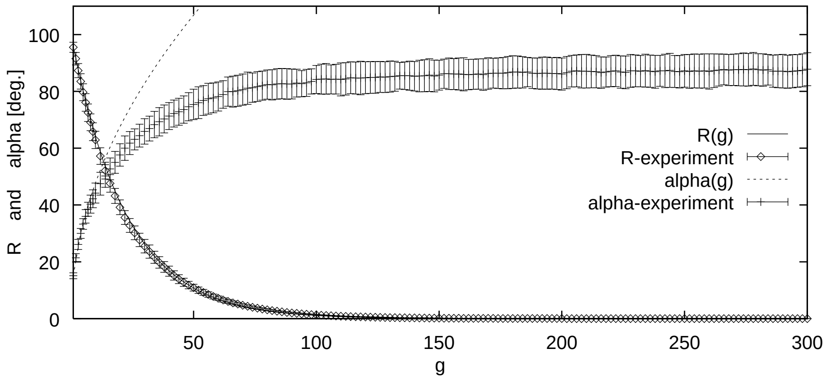


Fig. 7. The α- and R-dynamics of an *optimally tuned* $(5/5_I, 30)$-ES. Theoretical curves are compared with real ES runs; see also Fig. 6.

the parents were located at $\mathbf{R}_\star = (0, 0, \ldots, R_0)^{\mathrm{T}}$. The R-dynamics reads

$$R(g) = R_0 \exp\left[-\frac{\varphi^*(\sigma^*)}{N} g\right], \qquad \hat{\varphi}^*(\sigma^*) = \mu \frac{c_{\mu/\mu,\lambda}^2}{2} \tag{18}$$

with $\varphi^* = \hat{\varphi}^*_{(5/5,30)} \approx 4.46$, obtained by a normalized mutation strength $\hat{\sigma}^* = \mu c_{\mu/\mu,\lambda} \approx 5.9$, and $N = 100$, $R_0 = 100$.

As one can see, the angle of the center of mass parent moves away very fast from the initial starting point. At first glance it seems that the recombinant strategy has a higher explorative power than the (μ, λ)-ES. However, we have to relate this to the R-dynamics to be discussed in the next section.

3.3 Comparison of the (μ, λ)- and the $(\mu/\mu_I, \lambda)$-ES on the Sphere

If one compares the simulation results of Fig. 6 and 7, then one might conclude that the $(\mu/\mu, \lambda)$-ES is somehow more explorative than the (μ, λ)-ES, because the offspring are driven away faster from the gradient direction, as indicated by the α-dynamics. However, at the same time, the distance to the local optimum is decreased, too. To this end, let us assume that both strategies are tuned for optimal performance (maximal progress rate φ), i.e. $\hat{\sigma}^*_{\mu,\lambda} = c_{\mu,\lambda}$ and $\hat{\sigma}^*_{\mu/\mu_I,\lambda} = \mu c_{\mu/\mu,\lambda}$. Both strategies are compared in Table 2.

(μ, λ)-ES	$(\mu/\mu, \lambda)$-ES
$\hat{\overline{\alpha}}^{(g)}_{\mu,\lambda} \approx \sqrt{g}\, \frac{c_{\mu,\lambda}}{\sqrt{N}}$	$\hat{\overline{\alpha}}^{(g)}_{\mu/\mu,\lambda} \approx \sqrt{\mu\, g}\, \frac{c_{\mu/\mu,\lambda}}{\sqrt{N}}$
$\hat{R}(g) = R_0 \exp\left[-\frac{c^2_{\mu,\lambda}}{2N}\, g\right]$	$\hat{R}(g) = R_0 \exp\left[-\frac{c^2_{\mu/\mu,\lambda}}{2N}\, \mu g\right]$

Table 2. The α- and R-dynamics of optimally tuned ESs on the sphere model.

Now let us "stretch" the time in the second column of Table 2 by introducing a new time scale $\tilde{g}$

$$g =: \tilde{g}\, \frac{1}{\mu} \frac{c^2_{\mu,\lambda}}{c^2_{\mu/\mu,\lambda}}. \tag{19}$$

As is easily seen, the dynamics of the $(\mu/\mu, \lambda)$-ES is transformed into the dynamics of the (μ, λ)-ES. That is, the only difference between both dynamics is their time scale. The $(\mu/\mu, \lambda)$-ES is by a factor of $\mu c^2_{\mu/\mu,\lambda}/c^2_{\mu,\lambda}$ faster than the (μ, λ)-ES.

4 Summary

In order to evaluate the "explorative power" of ES/EP-like algorithms, the relation of exploitation and exploration has been discussed.

Exploitation is defined in this context as the ability of an algorithm to step into the direction of desired improvement. This can be the local gradient, but it might also be another desired direction, such as the global optimum direction. The progress in that direction can be quantified by the progress rate φ. It is thus a measure for the exploitative power of the algorithm.

Exploration is the process that drives the offspring away from the desired (local) exploitation direction. Unlike the exploitation which takes place in an (local) one-dimensional manifold of the search space, the exploration works in an $(N-1)$-dimensional hyperspace which is locally perpendicular (i.e. linearly independent) to the exploitation direction.

The actual "trail" of the population through the search space is by no means a perturbed gradient descent/ascent following the steepest hill direction. The "search path" resembles much more the *serpentine trails* to be found at hillsides.

One might speculate that the ability of the EA to locate a global optimal solution (surrounded by local optima) depends on its ability to find the right relation between exploitation and exploration. Up to now there are no models that allow for a deeper evaluation of this speculation. The approach introduced in this paper, however, might be a first step toward such investigations.

5 Acknowledgment

The author is grateful to David B. Fogel, A. Irfan Oyman, and Ralf Salomon for helpful comments. This work was funded by the DFG, grant Be 1578/4-1.

References

1. J. T. Alander. Indexed Bibliography of Genetic Algorithms. Report, University of Vaasa, Department of Information Technology and Production Economics, 1996. available via `ftp.uwasa.fi:/cs/report94-1/ga96bib.ps.Z`.
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: The (μ, λ)-Theory. *Evolutionary Computation*, 2(4):381–407, 1995.
4. H.-G. Beyer. Toward a Theory of Evolution Strategies: Self-Adaptation. *Evolutionary Computation*, 3(3):311–347, 1996.
5. H.-G. Beyer. *Zur Analyse der Evolutionsstrategien*. Habilitationsschrift, University of Dortmund, 1996.
6. H.-G. Beyer. An Alternative Explanation for the Manner in which Genetic Algorithms Operate. *BioSystems*, 41:1–15, 1997.
7. H.-G. Beyer and D. B. Fogel. A Note on the Escape Probabilities for Two Alternative Methods of Selection under Gaussian Mutation. In P.J. Angeline, R.G. Reynolds, J.R McDonnell, and R. Eberhart, editors, *Evolutionary Programming VI: Proceedings of the Sixth Annual Conference on Evolutionary Programming*, pages 265–274, Heidelberg, 1997. Springer-Verlag.
8. K. S. Miller. *Multidimensional Gaussian Distributions*. J. Wiley, New York, 1964.
9. H. Mühlenbein and D. Schlierkamp-Voosen. Predictive models for the Breeder Genetic Algorithm. *Evolutionary Computation*, 1(1):25–49, 1993.
10. I. Rechenberg. *Evolutionsstrategie '94*. Frommann–Holzboog Verlag, Stuttgart, 1994.
11. A. Törn and A. Zilinskas. *Global Optimization*. Springer-Verlag, Berlin, 1987.