The Steady State Behavior of $(\mu/\mu_I, \lambda)$-ES on Ellipsoidal Fitness Models Disturbed by Noise*

Hans-Georg Beyer and Dirk V. Arnold

Department of Computer Science XI,
University of Dortmund, D-44221 Dortmund, Germany
{hans-georg.beyer, dirk.arnold}@cs.uni-dortmund.de

Abstract. The method of differential-geometry is applied for deriving steady state conditions for the $(\mu/\mu_I, \lambda)$-ES on the general quadratic test function disturbed by fitness noise of constant strength. A new approach for estimating the expected final fitness deviation observed under such conditions is presented. The theoretical results obtained are compared with real ES runs showing a surprisingly excellent agreement.

1 Introduction

Understanding the impact of noise on the optimization behavior of evolutionary algorithms (EAs) is of great interest: There is a certain beliefe that EAs are especially good at coping with noisy information due to the use of a population of candidate solutions. There is empirical evidence as well as some theoretical support for this beliefe [3]. Furthermore, noise models on the level of the control parameters to be optimized, also called *actuator noise models* in [11], are of interest in the context of *robust optimization* [17,18,12].

While there is a need for a deeper understanding of the behavior of EAs on such noisy problems, a theoretical analysis is still at its beginning. Up to now, only the behavior of evolution strategies (ES) on the sphere model has been analyzed [1]. Performing similar analyses on other test functions still remain to be done. However, such analyses starting from scratch are expensive. Therefore, it would be desirable to use results obtained from the sphere theory as a starting point for deriving statements on the behavior of ES on other test functions.

This article is exactly in that spirit by taking up the thread from [8] where $(1, \lambda)$-ES has been considered. First, it applies the differential-geometrical model [7] in order to derive the condition for the zero progress rate in *recombinant* ES on general quadratic models disturbed by fitness noise of constant strength. Second, it provides a new and simple but surprisingly accurate method for estimating the expected final fitness deviation observed under such conditions.

The paper is organized as follows. After introducing the general quadratic test function disturbed by fitness noise we will determine the steady state condition

* This work was supported by the Deutsche Forschungsgemeinschaft (DFG), grant Be1578/6-3, and by the Collaborative Research Center (SFB) 531.

E. Cantú-Paz et al. (Eds.): GECCO 2003, LNCS 2723, pp. 525–536, 2003.

starting from the standard noisy sphere model. Then we will provide the new approach for determining the expected final fitness deviation. The predictions of this model will be compared with $(\mu/\mu_I, \lambda)$-ES runs. In the concluding section an outlook will be given emphasizing the potential of the methods presented.

2 The Steady State Condition of $(\mu/\mu_I, \lambda)$-ES on Noisy Quadratic Functions

2.1 The General Quadratic Fitness Noise Model

We consider the general quadratic fitness model based on the quality function

$$Q_{\mathrm{g}}(\mathbf{y}) := \mathbf{b}^{\mathrm{T}}\mathbf{y} - \mathbf{y}^{\mathrm{T}}\mathbf{Q}\mathbf{y} \tag{1}$$

where $\mathbf{b}$ and $\mathbf{y}$ are N-dimensional real-valued vectors and $\mathbf{Q}$ is a symmetric, (w.l.o.g.) positive definite matrix. Given an object vector $\mathbf{y}$ the actually observed objective value, i.e. fitness F_{ng}, is disturbed by Gaussian noise of strength σ_δ

$$F_{\mathrm{ng}}(\mathbf{y}) := Q_{\mathrm{g}}(\mathbf{y}) + \mathcal{N}(0, \sigma_\delta^2). \tag{2}$$

It is assumed that σ_δ is constant for each single generation. That is, all offspring within the same generation experience the same noise strength.

2.2 Determining the Steady State Condition

It is a common phenomenon that EA optimizing fitness functions disturbed by noise of constant strength exhibit some kind of steady state behavior (after a certain transient time of approaching the optimum) which is – on average – away from the optimal solution [6]. If this steady state regime has been reached, the expected fitness improvement will be zero. In order to determine the steady state condition of this behavior, we will reconsider the standard noisy sphere model and apply the differential-geometrical model [7] to it.

Results from the Sphere Model Theory. The qualitative properties of an ES can be characterized by evolution criteria [7, p. 90] which describe the appoach toward the optimum in terms of inequalities in the space of the endogenous strategy parameters such as the mutation strength and the noise strength. This concept has been developed for the $(1, \lambda)$-ES on the noisy sphere model

$$F_{\mathrm{nsp}}(\mathbf{y}) := f(\|\mathbf{y}\|) + \mathcal{N}(0, \sigma_\delta^2), \qquad f = f(r) \text{ monotonic function,} \tag{3}$$

in [5] and recently extended for the $(\mu/\mu_I, \lambda)$-ES in [2]. The asymptotically correct $(N \to \infty)$ evolution criterion reads

$$\sigma_\delta^{*2} + \sigma^{*2} \leq (2\mu c_{\mu/\mu,\lambda})^2 \tag{4}$$

where

$$\sigma^* = \sigma \frac{N}{R} \qquad \text{and} \qquad \sigma^*_\delta = \sigma_\delta \frac{N}{R|f'|} \quad \text{with} \quad f' = \left.\frac{\mathrm{d}f}{\mathrm{d}r}\right|_{r=R} \tag{5}$$

are the normalized mutation strength (isotropic mutations assumed) and the normalized noise strength, respectively. $R := \|\mathbf{y}_\mathrm{p}\|$ is the distance of the parental centroid (center of mass of the μ parents $\mathbf{y}_m$) to the optimum, and $c_{\mu/\mu,\lambda}$ is the progress coefficient (see e.g. [7, p. 247]).

Criterion (4) characterizes those endogenous strategy parameter states which guarantee convergence toward the optimum. That is, when the "<" relation is fulfilled in (4), the expected R value decreases from one generation to the next. Criterion (4) necessarily implies

$$\sigma^*_\delta \leq 2\mu c_{\mu/\mu,\lambda}. \tag{6}$$

Using (6) and the σ^*_δ definition in (5), the evolution criterion becomes

$$R|f'| \geq \frac{\sigma_\delta N}{2\mu c_{\mu/\mu,\lambda}}. \tag{7}$$

If the equal sign holds in (4) and (6, 7), respectively, the expected R value remains constant (from one generation to the next) and the evolution stagnates (on average). The latter case is of particular interest because it appears as the steady state of the ES working in a fitness environment with constant noise strength $\sigma_\delta = \text{const}$. Considering the special case

$$f(r) = \beta r^\alpha, \tag{8}$$

(7) can be solved for R yielding for the steady state

$$R \geq \sqrt[\alpha]{\frac{\sigma_\delta N}{2\alpha|\beta|\mu c_{\mu/\mu,\lambda}}} =: R_\infty. \tag{9}$$

It appears that R_∞ in (9) provides a good approximation for the expected steady state value of R in ES runs. This is so because the standard mutation strength adaptation techniques yield comparatively small normalized mutation strengths σ^*. This holds for the mutative σ self-adaptation ES (σSA-ES) as well as for the cumulative step-size adaptation ES (CSA-ES).

Distorting the Sphere – the General Quadratic Model. In order to transfer the idea behind the evolution criterion (7) to the general quadratic case (2), we have to introduce the generalized quantities for R and f'. This step is by analogy. A mathematically rigorous proof for the correctness of this step is still pending, however, there is experimental evidence for its accuaracy (see Section 2.4).

The analogue of the absolute value of the first derivative f' in the $\mathbb{R}^N$ space is the length of the gradient. Similarly, the radius R must be replaced by the differential-geometrical mean radius $\overline{R}$. Calculating the gradient in (1) yields

$$\nabla Q_\mathrm{g} = \mathbf{b} - 2\mathbf{Q}\mathbf{y}. \tag{10}$$

Since the optimal state $\hat{\mathbf{y}}$ is given by $\nabla Q_{\mathrm{g}} = \mathbf{0}$, we obtain from (10) $\mathbf{b} = 2\mathbf{Q}\hat{\mathbf{y}}$ and therefore we get

$$\nabla Q_{\mathrm{g}} = 2\mathbf{Q}(\hat{\mathbf{y}} - \mathbf{y}) =: \mathbf{a}. \tag{11}$$

Using the mean radius formula from [7, p.46], one gets

$$\overline{R} = \frac{N-1}{2} \frac{\|\mathbf{a}\|}{\left|\mathrm{Tr}[\mathbf{Q}] - \frac{\mathbf{a}^{\mathrm{T}}\mathbf{Q}\,\mathbf{a}}{\|\mathbf{a}\|^2}\right|}. \tag{12}$$

In order to obtain the necessary evolution criterion we now substitute $\|\nabla Q_{\mathrm{g}}\|$ for $|f'|$ and $\overline{R}$ for R in (7). After rearranging terms and considering $N \to \infty$, the evolution criterion reads

$$\frac{\|\mathbf{a}\|^2}{\left|\mathrm{Tr}[\mathbf{Q}] - \frac{\mathbf{a}^{\mathrm{T}}\mathbf{Q}\,\mathbf{a}}{\|\mathbf{a}\|^2}\right|} \geq \frac{\sigma_\delta}{\mu c_{\mu/\mu,\lambda}}. \tag{13}$$

Provided that the Rayleigh quotient $\mathbf{a}^{\mathrm{T}}\mathbf{Q}\mathbf{a}/\|\mathbf{a}\|^2$ can be neglected, this expression can be further simplified using (11). One obtains

$$\|\mathbf{Q}(\hat{\mathbf{y}} - \mathbf{y})\|^2 \geq \frac{\sigma_\delta \mathrm{Tr}[\mathbf{Q}]}{4\mu c_{\mu/\mu,\lambda}}. \tag{14}$$

Unlike the sphere model where the steady state is characterized by a constant (expected) residual distance to the optimal state, the general case is characterized by $\mathbf{y}$-states located on an ellipsoidal hypersurface (equal sign in (14)). It is important to realize that these ellipsoidal hypersurfaces are *not* geometrically similar to the ellipsoid defined by $Q_{\mathrm{g}}(\mathbf{y}) = \mathrm{const.}$ (in contrast to the sphere).

2.3 Estimating the Expected Stationary Fitness Error

As we have seen in Section 2.2, noisy fitness information implies a localization error of the optimizer in the $\mathbf{y}$ (object) parameter space. Therefore, choosing a parental state $\mathbf{y}_{\mathrm{p}}$ produced by the ES after reaching the vicinity of the steady state regime results in a $\mathbf{y}_{\mathrm{p}} \neq \hat{\mathbf{y}}$. Thus, the actually obtained *undisturbed* objective function value $Q = Q_{\mathrm{g}}(\mathbf{y}_{\mathrm{p}})$ will also deviate from the optimum value $\hat{Q} = Q_{\mathrm{g}}(\hat{\mathbf{y}})$. Since $\Delta Q := \hat{Q} - Q_{\mathrm{g}}(\mathbf{y}_{\mathrm{p}})$ is a random variate one can ask for its expected value and its variance.

A first attempt for estimating ΔQ by neglecting its random character has been presented in [8]. The idea was to use the respective stationarity condition (similar to Eq. (14) with $\mu = 1$) as a constraint on the optimization of Q_{g} given by (1). While this approach yielded a first rough lower bound on ΔQ, the predicted strong dependency of the Q deviation on the largest eigenvalue of $\mathbf{Q}$ and $\mathrm{Tr}[\mathbf{Q}]$ was not observed in experiments. Astonishingly, one observed ΔQ values the expected values of which were almost independent of the $\mathbf{Q}$ matrix.

We will now present an approach for estimating the expected value of ΔQ in accordance with the experimental observations mentioned above. In order to simplify the formulae, we will first switch to an appropriate coordinate system by performing a principal axes transformation. Let $\mathbf{e}_i$ $(i = 1, \ldots, N)$ be the normalized eigenvectors and q_i the corresponding eigenvalues of $\mathbf{Q}$. Eq. (1) can be transformed using the completeness condition $\mathbf{1} = \sum_{i=1}^{N} \mathbf{e}_i \mathbf{e}_i^{\mathrm{T}}$ into

$$Q_{\mathrm{g}}(\mathbf{y}) := \sum_{i=1}^{N} (b_i y_i - q_i y_i^2) \tag{15}$$

with $b_i = \mathbf{e}_i^{\mathrm{T}} \mathbf{b}$ and $y_i = \mathbf{e}_i^{\mathrm{T}} \mathbf{y}$. Performing quadratic completion in (15) yields

$$Q_{\mathrm{g}}(\mathbf{y}) := \sum_{i=1}^{N} \frac{b_i^2}{4q_i} - \sum_{i=1}^{N} q_i \left(y_i - \frac{b_i}{2q_i} \right)^2 . \tag{16}$$

One can easily prove that the optimum state of (15) is given by

$$\hat{y}_i = \frac{b_i}{2q_i} \qquad \text{and} \qquad \hat{Q} = \max[Q_{\mathrm{g}}] = \sum_{i=1}^{N} \frac{b_i^2}{4q_i}. \tag{17}$$

Thus, using Eq. (16) ΔQ can be expressed in terms of

$$\Delta Q = \hat{Q} - Q_{\mathrm{g}}(\mathbf{y}) = \sum_{i=1}^{N} q_i (y_i - \hat{y}_i)^2. \tag{18}$$

As one can see, the principal axes transformation decomposes ΔQ in a sum of N independent fitness contributions f_i

$$\Delta Q = \sum_{i=1}^{N} f_i \qquad \text{where} \qquad f_i = q_i (y_i - \hat{y}_i)^2. \tag{19}$$

Using the same transformation, the evolution criterion (14) can be expressed similarly. Since $\|\mathbf{Q}(\hat{\mathbf{y}} - \mathbf{y})\|^2 = (\hat{\mathbf{y}} - \mathbf{y})^{\mathrm{T}} \mathbf{Q}^2 (\hat{\mathbf{y}} - \mathbf{y})$ we immediately obtain

$$\|\mathbf{Q}(\hat{\mathbf{y}} - \mathbf{y})\|^2 = \sum_{i=1}^{N} q_i^2 (y_i - \hat{y}_i)^2 \tag{20}$$

and therefore with (14)

$$\sum_{i=1}^{N} q_i^2 (y_i - \hat{y}_i)^2 \geq \frac{\sigma_\delta \mathrm{Tr}[\mathbf{Q}]}{4\mu c_{\mu/\mu,\lambda}}. \tag{21}$$

We are now at that position where we can formulate the core ideas for the derivation of the expected Q deviation. At first we note that condition (21) does hold for all states y_i. Since y_i are random variates, we can take the expected value in (21) leading to

$$\sum_{i=1}^{N} q_i^2 \mathrm{E}[(y_i - \hat{y}_i)^2] \geq \frac{\sigma_\delta \mathrm{Tr}[\mathbf{Q}]}{4\mu c_{\mu/\mu,\lambda}}. \tag{22}$$

Provided that $\hat{y}_i = \mathrm{E}[y_i]$, the $\mathrm{E}[(y_i - \hat{y}_i)^2]$ expressions can be interpreted as the variances of the respective y_i variates. This is admissible for the steady state because of symmetry of the y_i states in (21). In other words, after reaching the vicinity of the steady state, the y_i fluctuate around the optimizer state $\hat{y}_i$.

Consider the expected value of ΔQ. Using (19) we have

$$\mathrm{E}[\Delta Q] = \sum_{i=1}^{N} \mathrm{E}[f_i] \qquad \text{where} \qquad \mathrm{E}[f_i] = q_i \mathrm{E}[(y_i - \hat{y}_i)^2]. \tag{23}$$

Now comes the crucial assumption which is quite similar to the *equipartition theorem* in statistical thermodynamics: Each degree of freedom in (23) contributes on average the same effect to the whole system. That is,

$$\text{Equipartition Assumption:} \quad \forall i, j : \ \mathrm{E}[f_i] = \mathrm{E}[f_j]. \tag{24}$$

It is quite clear that this assumption can only hold under equilibrium conditions, i.e. after reaching the steady state. However, considering the steady state, it should also be clear that the $\mathrm{E}[f_i] = \mathrm{E}[f_j]$ assumption is a natural one: First, the mutations generating new y_i states are not directed. Second, selection only "sees" the whole fitness. Therefore it cannot prefer a specific f_i degree and the f_i degrees fluctuate independently of each other. If a specific f_i degree dominated the others (i.e. having had a much larger $\mathrm{E}[f_i]$) then this would mean that its f_i contributions to the actual Q values were much higher. Such states, however, are likely to go extinct because selection prefers $\mathbf{y}$ realizations with lower Q values. Third, since the y_i states fluctuate independently around the optimum state $\hat{y}_i$, selection does not prefer any specific $\mathbf{y}$ direction (if this were not the case we would not be at the steady state but – on average – still move through the search space in a specific direction).

Accepting the validity of (24) we obtain with (23) $\mathrm{E}[f_i] = \mathrm{E}[\Delta Q]/N$ and therefore

$$\mathrm{E}[(y_i - \hat{y}_i)^2] = \frac{\mathrm{E}[\Delta Q]}{N q_i}. \tag{25}$$

Taking into account that at the steady state $\mathrm{E}[y_i] = \hat{y}_i$ does hold (recall the discussion above), (25) is also the variance of y_i

$$\mathrm{Var}[y_i] = \frac{\mathrm{E}[\Delta Q]}{N q_i}. \tag{26}$$

After insertion of (25) into the evolution criterion (22) we end up with a surprisingly simple condition (recall that $\sum_i q_i = \mathrm{Tr}[\mathbf{Q}]$)

$$\mathrm{E}[\Delta Q] \geq \frac{\sigma_\delta N}{4\mu c_{\mu/\mu,\lambda}}. \tag{27}$$

As can be checked by experiments (see Section 2.4), the equal sign in (27) predicts the average steady state ΔQ well as long as the mutation strength of the ES is controlled appropriately. The most astonishing message from (27) is the independence of $\mathrm{E}[\Delta Q]$ on the $\mathbf{Q}$ matrix. This was already observed in $(1, \lambda)$-ES runs in [8]. However, the reason for this interesting behavior remained obscure. Now we are able to explain this behavior by the equipartition effect which decomposes the (arbitrarily oriented) ellipsoid into its principal fitness components.

Inserting (27) into (26) yields an estimate for the object parameter fluctuations at the steady state. Since the steady state is characterized by the equal sign in (27), we obtain under steady state conditions

$$\mathrm{Var}[y_i] = \frac{\sigma_\delta}{4\mu c_{\mu/\mu,\lambda} q_i}. \tag{28}$$

This result is also in accordance with experiments. The main conclusion that can be drawn from (28) is that y_i parameter fluctuations decrease with the increase of the corresponding eigenvalue q_i. This is reasonable: A large eigenvalue q_i (compared to a smaller q_j) results in a higher sensitivity of the fitness on the particular parameter space direction $\mathbf{e}_i$. That is, it produces (on average) a larger deviation from the optimal fitness value. Such deviations, however, are singled out by the (μ, λ) selection, only small deviations will survive. On the other hand, small eigenvalues reduce the influence of the y_i fluctuations on the fitness. Therefore, y fluctuations in such $\mathbf{e}$-directions will be larger.

2.4 ES-Dynamics and Comparison with Experiments

The dynamical behavior of the ES maximizing the noisy function class (1), (2) has been investigated on three ellipsoidal test functions

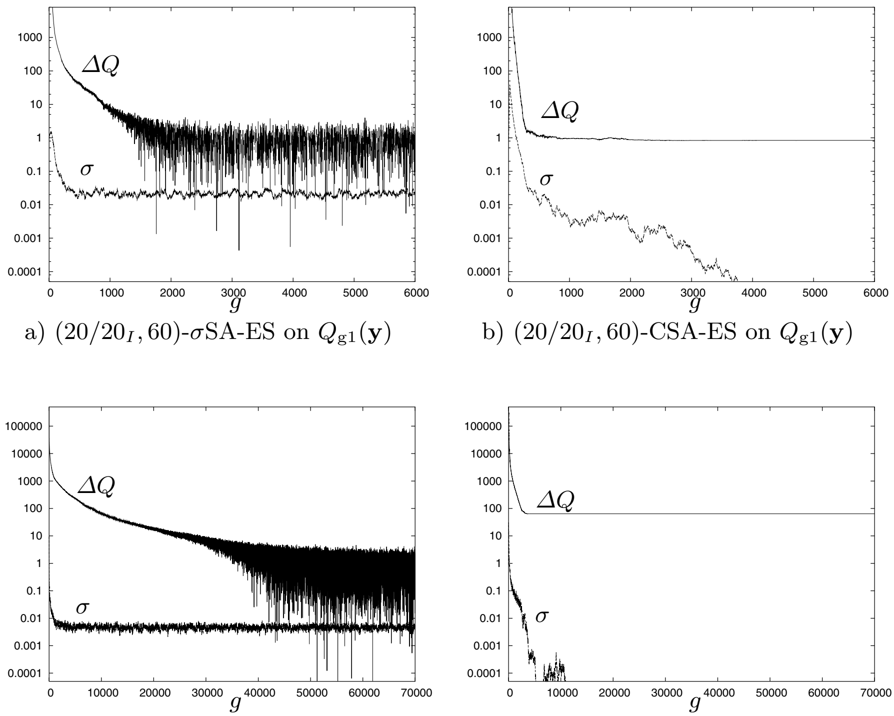
$$Q_{\mathrm{g1}}(\mathbf{y}) = -\sum_{i=1}^{N} i y_i^2, \qquad (\mathbf{Q})_{ij} = i\delta_{ij}, \quad (q_i = i), \tag{1.29a}$$

$$Q_{\mathrm{g2}}(\mathbf{y}) = -\sum_{i=1}^{N} i^2 y_i^2, \qquad (\mathbf{Q})_{ij} = i^2\delta_{ij}, \quad (q_i = i^2), \tag{1.29b}$$

$$Q_{\mathrm{g3}}(\mathbf{y}) = -\sum_{j=1}^{N} \left(\sum_{i=1}^{j} y_i\right)^2, \qquad (\mathbf{Q})_{ij} = \min[N-i+1, N-j+1], \tag{1.29c}$$

for dimensionalities $N = 30$ and 100. While $Q_{\mathrm{g1}}(\mathbf{y})$ and $Q_{\mathrm{g2}}(\mathbf{y})$ define axis-parallel ellipsoids, the third function, also known as "Schwefel's" function, has a certain non-parallel orientation. Since we are using isotropic mutations (sphere-symmetrical mutations) the orientation of the ellipsoids does not affect the performance of the ES on these test functions. However, $Q_{\mathrm{g3}}(\mathbf{y})$ has the peculiarity that the eigenvalue spectrum of $\mathbf{Q}$ possesses a dominating eigenvalue. Thus the shape of this ellipsoid resembles a distorted discus. This might influence the dynamical behavior, however, in the expriments performed using the noise model (2) no peculiarities concerning the steady-state behavior have been observed.

Differences are observed, however, concerning the dynamic behavior of the different mutation strength σ control rules used. We have tested the standard mutative σ self-adaptation (σSA, see e.g. [4]) and the cumulative step-length adaptation (CSA) proposed by Gawelczyk, Hansen, and Ostermeier [14,15,16] without covariance matrix adaptation, i.e. using isotropic mutations. Figures 1a–d show the typical behaviors. Both ES versions end up with a steady state behavior where the fitness values are in expectation away from the global optimum $\Delta Q = 0$, i.e. $\mathrm{E}[\Delta Q] > 0$. This is the typical behavior when noise is involved in the fitness evaluations. Clearly, the main aim is to have $\mathrm{E}[\Delta Q]$ as small as possible. When comparing the resulting steady-state $\mathrm{E}[\Delta Q]$ in Fig. 1c,d one notices that the CSA-ES yields a much larger $\mathrm{E}[\Delta Q]$ than the σSA-ES. From this point of view, the σSA-ES should be preferred. However, as one can see this is brought at the expense of a slower approach to the steady state. Comparing the steady state behavior of the two ES types on the two test functions (1.29a) and (1.29b) one also sees that, using CSA-ES, the effect of larger $\mathrm{E}[\Delta Q]$ gets larger with



a) $(20/20_I, 60)$-σSA-ES on $Q_{\mathrm{g1}}(\mathbf{y})$ b) $(20/20_I, 60)$-CSA-ES on $Q_{\mathrm{g1}}(\mathbf{y})$

c) $(20/20_I, 60)$-σSA-ES on $Q_{\mathrm{g2}}(\mathbf{y})$ d) $(20/20_I, 60)$-CSA-ES on $Q_{\mathrm{g2}}(\mathbf{y})$

Fig. 1. Evolution dynamics by $(20/20_I, 60)$-ES on (1.29a, b) ($N = 30$) using mutative self-adaptation (σSA, Figs. a and c) and cumulative step length adaptation (CSA, Figs. b and d). The CSA exhibits premature convergence on test function $Q_{\mathrm{g2}}(\mathbf{y})$ (Fig. d). As noise strength $\sigma_\delta = 1$ has been chosen.

increasing non-sphericity.[1] The reason for this undesirable behavior can be explained when considering the mutation strengths σ actually realized during the evolution. While the σSA-ES produces a quasi-constant steady state mutation strength, the CSA-ES produces an almost random walk like σ behavior on the logarithmic scale with very small σ values. That is, the CSA σ control rule produces a nearly premature convergence behavior and the ES is not able to further evolve towards the optimizer state. The reason for this – at first glance – astonishing behavior can be traced back to the optimality condition the CSA control rule is based upon [13]: Consecutive changes of the parental centroids should be – on average - perpendicular to each other in order to have maximal progress on the sphere model. The analysis in [9] shows, however, that this assumption leads to a wrong adaptation behavior when fitness information is disturbed by noise. As a result, σ is decreased even though it should be kept nearly constant (for an in-depth discussion on the sphere model, see [9]).

There is a remedy for the undesired σ decrease in CSA-ES: Simply keep the mutation strength σ above a certain (but small) limit σ_0. Figure 2 shows the effect of this remedy. The CSA-ES is prevented from premature convergence.

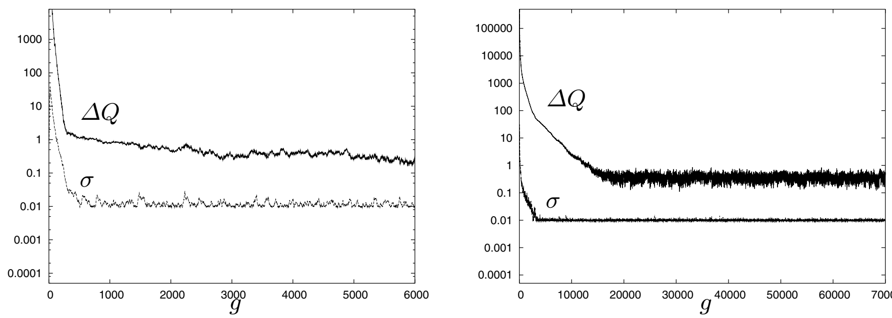


Fig. 2. Evolution dynamics of the $(20/20_I, 60)$-CSA-ES keeping σ explicitly above $\sigma_0 = 0.01$. Left: test function (1.29a), $N = 30$; right: test function (1.29b), $N = 30$. As noise strength $\sigma_\delta = 1$ has been chosen.

The problem is, however, that fixing σ_0 is a difficult task and also that the approach to the steady state is slowed down. Therefore, this method cannot be recommended as a clever strategy. In the following we will not consider the CSA-ES further because in this article we are mainly interested in the expected steady state ΔQ. Therefore, our simulations will be performed using the old σSA-ES.

Figure 3 compares the predictive quality of the equal sign in (27) as an estimate for the expected steady state ΔQ. The $(\mu/\mu_I, 60)$-σSA-ES has been used for the simulations. ΔQ was recorded at each generation after a number of transient generations g_0 by evaluating the (noisy) fitness (1), (2) of the parental centroid

[1] This might be an argument for using the covariance matrix adaptation (CMA) [15], however, this is not the focus of this paper.

using the test functions (1.29a,b,c). The number of generations used for averaging ΔQ is 200,000. The noise strength used is $\sigma_\delta = 1$. There is a good agreement between experiments and the lower bound of $\mathrm{E}[\Delta Q]$ given by the curve obtained from (27). Recall that the lower bound corresponds to vanishing normalized mutation strength in the original evolution criterion (7). Considering the actually realized σ values (see, e.g., the figures on the left-hand sides of Figs. 1 and 2) one realizes that the σSA-ES exhibits a behavior where σ is obviously that small at the steady state such that the equal sign in (27) is roughly fulfilled. That is

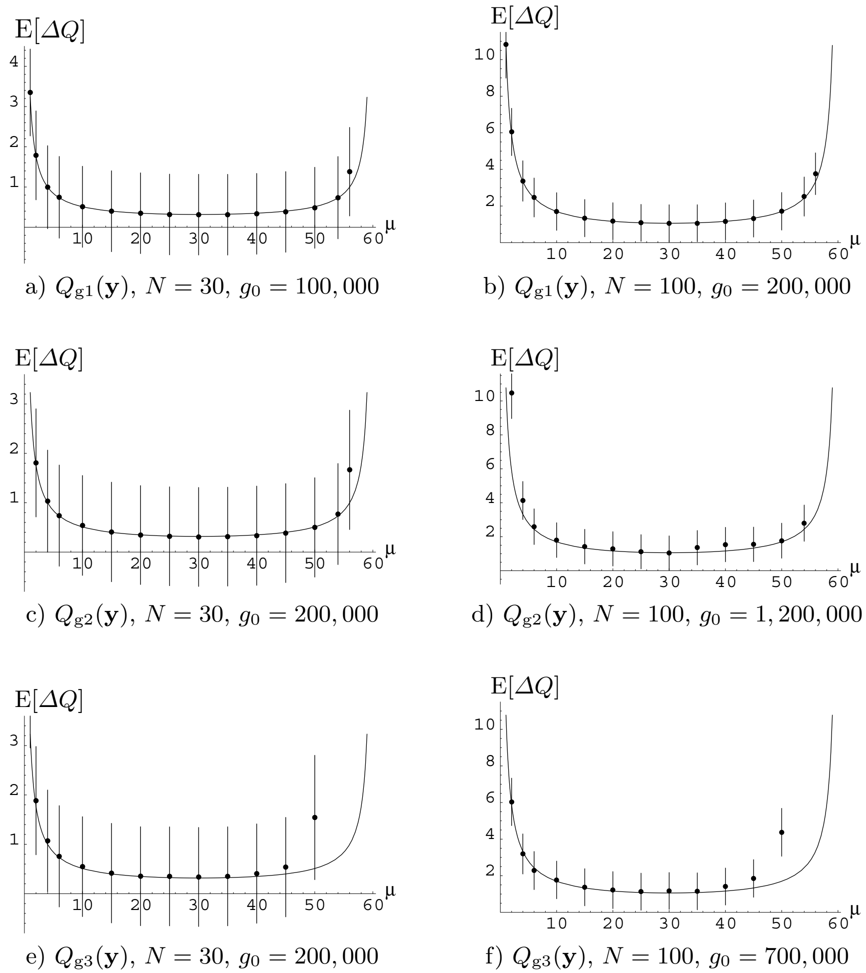


Fig. 3. Dependence of the expected steady state fitness error $\mathrm{E}[\Delta Q]$ on the parent numbers $\mu = 1, 2, 4, 6, 10, 15, 20, 25, 30, 35, 40, 45, 50, 54, 56, 58, 59$ given fixed offspring number $\lambda = 60$. The vertical bars indicate the measured $\pm$ standard deviation of ΔQ. Note, some data points are missing, see explanation in the text.

why we observe such a good agreement between theory and experiments. On the other hand the mutation strength is large enough to ensure convergence to the vicinity of the steady state described by (27). This is in contrast to the CSA-ES where the mutation strength goes down very rapidly when reaching the vicinity of the steady state. Violating the smallness assumption of σ, however, will result in a similar behavior: The ES cannot approach states which are described by the equal sign in (27). This can be observed in σSA-ES with μ/λ near 1, i.e. in strategies with low selective pressure, and can also be seen in the plots: Depending on the test function and the dimensionality there are some data points missing (usually $\mu = 59$ and $\mu = 58$, sometimes even for smaller μ) due to divergence. The behavior of the σSA-ES is diametrically opposite to the CSA-ES under this condition. Having a very small selection pressure results in an almost random selection behavior. As has been shown in [10], random selection results in an exponential increase of the mutation strength of the σSA-ES. Therefore, one observes a continuously increasing mutation strength if $\lambda - \mu$ is chosen too small. This effect starts gradually with increasing μ (keeping λ constant) and can be observed in the experiments presented.

3 Conclusions and Outlook

Using the equipartition assumption we were able to derive a simple formula which predicts the final expected fitness deviation surprisingly well. While the σSA-ES reaches the predicted fitness deviation, the CSA-ES exhibits premature convergence on ellipsoidal test function with a high degree of non-sphericity.

Formula (27) can be used for population sizing. In order to get to the optimizer as closely as possible $\mu/\lambda = 0.5$ should be chosen. Getting to the steady state as fast as possible, however, requires $\mu/\lambda \approx 0.27$ (sphere model assumption and $N \to \infty$, not considered in this paper). Considering the plots in Fig. 3, $\mu/\lambda = 0.3$ seems to be a good compromise.

Since both CSA-ES and σSA-ES use isotropic mutations, in a next step ES with nonisotropic mutations should be investigated. One might expect an improved ES behavior using covariance matrix adaptation (CMA) [15]. While the CMA-ES may yield better results than the CSA-ES, theoretically, CMA can *not* significantly improve the *steady state results* of the σSA-ES (basically, CMA-ES transforms $\mathbf{Q}$ into another $\tilde{\mathbf{Q}}$, but (27) does *not* depend on $\mathbf{Q}$ or $\tilde{\mathbf{Q}}$ at all). However, we can expect an improved transient behavior (decreasing g_0) of the CMA-ES compared to the ES with isotropic mutations. This remains to be investigated in the future.

References

1. D. V. Arnold. *Noisy Optimization with Evolution Strategies.* Kluwer Academic Publishers, Dordrecht, 2002.
2. D. V. Arnold and H.-G. Beyer. Performance Analysis of Evolution Strategies with Multi-Recombination in High-Dimensional $\mathbb{R}^N$-Search Spaces Disturbed by Noise. *Theoretical Computer Science*, 289:629–647, 2002.

3. D. V. Arnold and H.-G. Beyer. A Comparison of Evolution Strategies with Other Direct Search Methods in the Presence of Noise. *Computational Optimization and Applications*, 24:135–159, 2003.
4. T. Bäck, U. Hammel, and H.-P Schwefel. Evolutionary computation: comments on the history and current state. *IEEE Transactions on Evolutionary Computation*, 1(1):3–17, 1997.
5. H.-G. Beyer. Toward a Theory of Evolution Strategies: Some Asymptotical Results from the $(1,^+ \lambda)$-Theory. *Evolutionary Computation*, 1(2):165–188, 1993.
6. H.-G. Beyer. Evolutionary Algorithms in Noisy Environments: Theoretical Issues and Guidelines for Practice. *Computer Methods in Applied Mechanics and Engineering*, 186(2–4):239–267, 2000.
7. H.-G. Beyer. *The Theory of Evolution Strategies*. Natural Computing Series. Springer, Heidelberg, 2001.
8. H.-G. Beyer and D. V. Arnold. Fitness Noise and Localization Errors of the Optimum in General Quadratic Fitness Models. In W. Banzhaf, J. Daida, A.E. Eiben, M.H. Garzon, V. Honavar, M. Jakiela, and R.E. Smith, editors, *GECCO-99: Proceedings of the Genetic and Evolutionary Computation Conference*, pages 817–824, San Francisco, CA, 1999. Morgan Kaufmann.
9. H.-G. Beyer and D.V Arnold. Qualms Regarding the Optimality of Cumulative Path Length Control in CSA/CMA-Evolution Strategies. *Evolutionary Computation*, 11(1):19–28, 2003.
10. H.-G. Beyer and K. Deb. On Self-Adaptive Features in Real-Parameter Evolutionary Algorithms. *IEEE Transactions on Evolutionary Computation*, 5(3):250–270, 2001.
11. H.-G. Beyer, M. Olhofer, and B. Sendhoff. On the Behavior of $(\mu/\mu_I, \lambda)$-ES Optimizing Functions Disturbed by Generalized Noise. In K. De Jong, R. Poli, and J. Rowe, editors, *Foundations of Genetic Algorithms, 7*, San Francisco, CA, 2003. Morgan Kaufmann. in print.
12. J. Branke. *Evolutionary Optimization in Dynamic Environments*. Kluwer Academic Publishers, Dordrecht, 2001.
13. N. Hansen and A. Ostermeier. Adapting Arbitrary Normal Mutation Distributions in Evolution Strategies: The Covariance Matrix Adaptation. In *Proceedings of 1996 IEEE Int'l Conf. on Evolutionary Computation (ICEC '96)*, pages 312–317. IEEE Press, NY, 1996.
14. N. Hansen and A. Ostermeier. Convergence Properties of Evolution Strategies with the Derandomized Covariance Matrix Adaptation: The $(\mu/\mu_I, \lambda)$-CMA-ES. In H.-J. Zimmermann, editor, *5th European Congress on Intelligent Techniques and Soft Computing (EUFIT'97)*, pages 650–654, Aachen, Germany, 1997. Verlag Mainz.
15. N. Hansen and A. Ostermeier. Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation*, 9(2):159–195, 2001.
16. A. Ostermeier, A. Gawelczyk, and N. Hansen. A Derandomized Approach to Self-Adaptation of Evolution Strategies. *Evolutionary Computation*, 2(4):369–380, 1995.
17. S. Tsutsui and A. Ghosh. Genetic Algorithms with a Robust Solution Searching Scheme. *IEEE Transactions on Evolutionary Computation*, 1(3):201–208, 1997.
18. D. Wiesmann, U. Hammel, and T. Bäck. Robust Design of Multilayer Optical Coatings by Means of Evolutionary Algorithms. *IEEE Transactions on Evolutionary Computation*, 2(4):162–167, 1998.