Actuator Noise in Recombinant Evolution Strategies on General Quadratic Fitness Models*

Hans-Georg Beyer

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

Abstract. This paper addresses the influence of actuator noise on the steady state behavior of multirecombinant evolution strategies (ES) on general quadratic fitness functions. Actuator noise degrades the ES's ability to locate the global optimizer. After a certain transient time the ES approaches a steady state behavior characterized by an expected fitness deviation from the global optimum. This expected value is calculated and the predictions are compared with ES runs on quadratic test functions.

1 Introduction

Actuator noise is a phenomenon widely observed in practice when trying to control the behavior of a device or machine by a set of control parameters which cannot be tuned exactly. While the control parameters can be prescribed exactly, its actual realization on the machine is disturbed by random perturbations such as vibrations (ground motion, turbulence effects, etc.) or other sources of noise (e.g., resistor noise, recombination noise, burst noise in electronic devices like resistors or transistors). If one wants to optimize the performance of such devices or machines, taking the actuator noise into account, one has to deal with goal functions which are intrinsically random functions. That is, optimizing such functions by neglecting its randomness can lead to a false optimal object parameter set.

Another problem domain with similar implications concerns *robust design* and optimization. Here one seeks to find optimal solutions which are robust with respect to random perturbation of design parameters [4,13,14]. The main application area is in the field of coping with production tolerances. There is only a limited degree of accuracy by which devices can be produced. Optimizing the design of a device has to serve two goals: On the one hand the device's performance, production costs, etc. should be maximal/minimal, on the other hand it must be "producable", i.e., the production process must allow for production tolerances. This can be achieved at the level of product design by superimposing random perturbations on the design variable modeling the impact of the production tolerances. Therefore, the aim of the design optimization process is

* This work was supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the Collaborative Research Center (SFB) 531.

K. Deb et al. (Eds.): GECCO 2004, LNCS 3102, pp. 654–665, 2004.

basically not finding an optimum solution represented by a sharp peak in the fitness landscape, but rather optimal solutions which are less sensitive to small changes of the design parameters.

It is widely believed that evolutionary algorithms (EA) are good at such design tasks. These algorithms seem rather suited for finding large optimum attractors than finding the "needle in the haystack" peak. Up to now, there are only a few references addressing the question whether this "folklore" belief can be substantiated by hard and provable facts. Most investigations done on this topic are mainly of empirical nature [4,13] or consider special one-dimensional cases [14] without analyzing the EA's behavior on the test functions proposed. Only recently an attempt has been made to understand the behavior of evolution strategies (ES) on simple N-dimensional test functions disturbed by actuator noise [3,12]. These investigations revealed interesting behaviors such as (actuator) noise-induced bistabilities on a unimodal fitness landscape [12] and the appearance of an optimum localization error on a sphere model with actuator noise [3]. In this paper we apply a technique proposed in [3] to investigate the behavior of $(\mu/\mu_I, \lambda)$-ES[1] on *general* quadratic fitness functions disturbed by actuator noise. The results to be presented here extend the findings obtained for the simple (i.e. symmetrical) sphere model to a more realistic situation of a general quadratic fitness model. Such models can be regarded as local attractor models of real-world objective functions.

The rest of the paper is organized as follows. First, we will introduce the actuator noise model. Second, the steady state condition for the $(\mu/\mu_I, \lambda)$-ES on this fitness model will be derived. In Section 4 we compare the theoretical predictions with real ES runs. Finally, in the concluding section a short summary will be given including an outlook to future research.

2 The Actuator Noise Model

The actuator noise model was introduced in [3] to account for object parameter fluctuations like actuator jittering which are beyond the control of the user and the optimization algorithm, respectively. The model considered was the quadratic sphere. This paper investigates an *arbitrary* N-dimensional quadratic function $Q(\mathbf{y})$ (to be maximized)

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

with the N-dimensional real-valued vectors $\mathbf{b}$ and $\mathbf{y}$ and the symmetric (positive definite) matrix $\mathbf{Q}$. Given an object (or actuator) vector $\mathbf{y}$, the actually observed objective value, i.e. the fitness, is defined by the *actuator noise model*

$$F_{\mathrm{a}}(\mathbf{y}) := Q(\mathbf{y} + \mathbf{z}), \qquad \text{where} \qquad \mathbf{z} \sim \boldsymbol{\mathcal{N}}(\mathbf{0}, \varepsilon^2 \mathbf{1}). \tag{2}$$

That is, each object parameter component is disturbed by independent normally distributed random events z_i with the same standard deviation ε.

[1] For the definitions of the evolution strategies used, see Appendix A.

3 Determination of the Steady State of the ES

The analysis of the steady state behavior follows the decomposition technique proposed in [3]. The basic idea is to transform the random function $F_\mathrm{a}(\mathbf{y})$ in such a way that it appears as a sum of two parts, one carrying the stochastics and the other being deterministic. That is, the transformed problem appears as a fitness function with *additive* fitness noise. This transformation is admissible because the ES acts as a black-box algorithm which only uses the fitness information but not structural information from the fitness function. If we were able to make the transformation in such a way that we obtain a fitness noise model already analyzed then we are done. Therefore, the aim of the next section is to derive such a (approximative) model. As a result we will obtain the general quadratic model with (approximately) normally distributed fitness noise. Treating this model in Section 3.2 using techniques from [2] will yield the desired expected steady state fitness deviation from the global optimum.

3.1 Reducing the Actuator Noise Model to the General Quadratic Noisy Fitness Model

In order to obtain an approximative fitness noise model one has to decompose (2) into a deterministic part in terms of (1) and a normally distributed additive noise term. Since the $(\mu/\mu_I, \lambda)$-ES with isotropic mutations is considered, it is reasonable to express the fitness model and its decomposition in the eigensystem of the matrix $\mathbf{Q}$. Let q_i be the eigenvalues of $\mathbf{Q}$ and $\mathbf{e}_i$ the corresponding eigenvectors of length 1, i.e. $q_i\mathbf{e}_i = \mathbf{Q}\mathbf{e}_i$, the entire actuator noise model (1), (2) can be rewritten by a principal axes transformation as

$$F_\mathrm{a}(\mathbf{y}) = \sum_{i=1}^{N}[b_i(y_i + z_i) - q_i(y_i + z_i)^2]$$

$$F_\mathrm{a}(\mathbf{y}) = \sum_{i=1}^{N}[b_i y_i - q_i y_i^2] + \sum_{i=1}^{N}[(b_i - 2q_i y_i)z_i - q_i z_i^2], \tag{3}$$

where $z_i \sim \mathcal{N}(0, \varepsilon^2)$ and $b_i = \mathbf{e}_i^\mathrm{T}\mathbf{b}$, $y_i = \mathbf{e}_i^\mathrm{T}\mathbf{y}$. Since it is the aim to decompose $F_\mathrm{a}(\mathbf{y})$ in such a manner that

$$F_\mathrm{a}(\mathbf{y}) \simeq E[F_\mathrm{a}|\mathbf{y}] + \mathcal{N}(0, \mathrm{Var}[F_\mathrm{a}|\mathbf{y}]) + \ldots, \tag{4}$$

one has to calculate $E[F_\mathrm{a}|\mathbf{y}]$ and $\mathrm{Var}[F_\mathrm{a}|\mathbf{y}]$. For the first conditional moment we easily obtain (recall $\mathrm{E}[z_i] = 0$, $\mathrm{E}[z_i^2] = \varepsilon^2$, $\sum q_i = \mathrm{Tr}[\mathbf{Q}]$)

$$E[F_\mathrm{a}|\mathbf{y}] = \sum_{i=1}^{N}[b_i y_i - q_i y_i^2] - \sum_{i=1}^{N} q_i \overline{z_i^2} = Q(\mathbf{y}) - \varepsilon^2 \mathrm{Tr}[\mathbf{Q}]. \tag{5}$$

For $\mathrm{Var}[F_\mathrm{a}|\mathbf{y}]$ one obtains (recall that $\mathrm{E}[z_i^3] = 0$, $\mathrm{E}[z_i^4] = 3\varepsilon^4$, $\sum q_i^2 = \mathrm{Tr}[\mathbf{Q}^2]$)

$$\mathrm{Var}[F_\mathrm{a}|\mathbf{y}] = \sum_{i=1}^{N} \mathrm{Var}[(b_i - 2q_i y_i)z_i - q_i z_i^2]$$

$$\begin{aligned}
\mathrm{Var}[F_{\mathrm{a}}|\mathbf{y}] &= \sum_{i=1}^{N} \left[\mathrm{E}[((b_i - 2q_i y_i)z_i - q_i z_i^2)^2] - (\mathrm{E}[(b_i - 2q_i y_i)z_i - q_i z_i^2])^2\right] \\
&= \sum_{i=1}^{N} \left[(b_i - 2q_i y_i)^2 \varepsilon^2 + 2q_i^2 \varepsilon^4\right] \\
&= \varepsilon^2 \sum_{i=1}^{N} (b_i - 2q_i y_i)^2 + 2\varepsilon^4 \mathrm{Tr}[\mathbf{Q}^2]. \\
&= 4\varepsilon^2 \sum_{i=1}^{N} q_i^2 \left(y_i - \frac{b_i}{2q_i}\right)^2 + 2\varepsilon^4 \mathrm{Tr}[\mathbf{Q}^2].
\end{aligned} \tag{6}$$

Taking into account that the optimal state $\hat{\mathbf{y}}$ of $Q(\mathbf{y})$ is easily obtained from (5)

$$\hat{y}_i = \frac{b_i}{2q_i} \qquad \text{and} \qquad \hat{Q} := \max[Q] = \sum_{i=1}^{N} \frac{b_i^2}{4q_i}, \tag{7}$$

one gets $\mathrm{Var}[F_{\mathrm{a}}|\mathbf{y}] = 4\varepsilon^2 \sum_{i=1}^{N} q_i^2 (y_i - \hat{y}_i)^2 + 2\varepsilon^4 \mathrm{Tr}[\mathbf{Q}^2]$. This can be written in vector notation

$$\mathrm{Var}[F_{\mathrm{a}}|\mathbf{y}] = 4\varepsilon^2 \|\mathbf{Q}(\hat{\mathbf{y}} - \mathbf{y})\|^2 + 2\varepsilon^4 \mathrm{Tr}[\mathbf{Q}^2]. \tag{8}$$

Inserting (5) and (8) into (4) yields finally

$$F_{\mathrm{a}}(\mathbf{y}) \simeq Q(\mathbf{y}) - \varepsilon^2 \mathrm{Tr}[\mathbf{Q}] + \underbrace{\varepsilon\sqrt{4\|\mathbf{Q}(\hat{\mathbf{y}} - \mathbf{y})\|^2 + 2\varepsilon^2 \mathrm{Tr}[\mathbf{Q}^2]}}_{=\sigma_\delta} \mathcal{N}(0,1). \tag{9}$$

Note, the constant term $-\varepsilon^2 \mathrm{Tr}[\mathbf{Q}]$ (w.r.t. $\mathbf{y}$) in (9) is without relevance for the derivation of the evolution criterion because this term does not depend on the location in the object parameter space. While this is true for the derivations to be presented below, the effect of this term with respect to the attainable objective function values is of considerable importance because it degrades the maximal fitness independent of the ES used. Even if one were able to determine the optimal object parameter vector $\hat{\mathbf{y}}$, the expected value of the maximal fitness (7) $\hat{Q}$ will still be reduced by the term $\varepsilon^2 \mathrm{Tr}[\mathbf{Q}]$.

3.2 Deriving the Evolution Criterion

Due to the (approximate) decomposition (9) we have reduced our problem to a case already known: Equation (14) in [2] characterizes the steady state behavior of the $(\mu/\mu_I, \lambda)$-ES on an arbitrary ellipsoidal function $Q(\mathbf{y})$ with Gaussian *fitness* noise of strength σ_δ

$$\|\mathbf{Q}(\hat{\mathbf{y}} - \mathbf{y})\|^2 \geq \frac{\sigma_\delta \mathrm{Tr}[\mathbf{Q}]}{4\mu c_{\mu/\mu,\lambda}} \tag{10}$$

where $c_{\mu/\mu,\lambda}$ is the progress coefficient (for its definition, see Appendix B). Taking the expectation in (10) one first obtains

$$\mathrm{E}[\|\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})\|^2] \geq \frac{\mathrm{E}[\sigma_\delta]\mathrm{Tr}[\mathbf{Q}]}{4\mu c_{\mu/\mu,\lambda}}. \tag{11}$$

Since the expected value expression $\mathrm{E}[\sigma_\delta]$ cannot be calculated analytically from (9), the approximation

$$\mathrm{E}[\sigma_\delta] \simeq \varepsilon\sqrt{4\mathrm{E}[\|\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})\|^2] + 2\varepsilon^2\mathrm{Tr}[\mathbf{Q}^2]} \tag{12}$$

must be used leading to an implicit evolution criterion

$$\mathrm{E}[\|\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})\|^2] \geq \frac{\varepsilon\mathrm{Tr}[\mathbf{Q}]}{4\mu c_{\mu/\mu,\lambda}}\sqrt{4\mathrm{E}[\|\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})\|^2] + 2\varepsilon^2\mathrm{Tr}[\mathbf{Q}^2]}. \tag{13}$$

This criterion can be resolved for $\mathrm{E}[\|\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})\|^2]$. After squaring (13), and solving the quadratic inequality one gets

$$\mathrm{E}[\|\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})\|^2] \geq \frac{\varepsilon^2\mathrm{Tr}[\mathbf{Q}]^2}{8\mu^2 c_{\mu/\mu,\lambda}^2}\left(1+\sqrt{1+\frac{8\mu^2 c_{\mu/\mu,\lambda}^2\mathrm{Tr}[\mathbf{Q}^2]}{\mathrm{Tr}[\mathbf{Q}]^2}}\right). \tag{14}$$

The quantity $\mathrm{E}[\|\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})\|^2]$ is not directly observable, therefore, we consider now the expected fitness deviation from the optimum state $\hat{\mathbf{y}}$. Defining ΔF as

$$\Delta F := F_\mathrm{a}(\hat{\mathbf{y}}) - F_\mathrm{a}(\mathbf{y}), \tag{15}$$

one obtains using (2) and (3)

$$\begin{aligned}\Delta F &= Q(\hat{\mathbf{y}}+\mathbf{z}) - Q(\mathbf{y}+\mathbf{z}) \\ &= Q(\hat{\mathbf{y}}) - Q(\mathbf{y}) + 2\sum_{i=1}^{N} q_i(y_i-\hat{y}_i)z_i.\end{aligned} \tag{16}$$

Taking the expected value, one ends up with

$$\mathrm{E}[\Delta F] = \mathrm{E}[\underbrace{Q(\hat{\mathbf{y}}) - Q(\mathbf{y})}_{=:\Delta Q}]. \tag{17}$$

That is, we have reduced the calculation of $\mathrm{E}[\Delta F]$ to that of $\mathrm{E}[\Delta Q]$, i.e, to the expected fitness deviation of the general ellipsoidal model with Gaussian fitness noise.

In order to calculate $\mathrm{E}[\Delta Q]$ we rewrite $\mathrm{E}[\|\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})\|^2]$ using the principal axes transformation

$$\mathrm{E}[\|\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})\|^2] = \sum_{i=1}^{N} q_i^2\mathrm{E}[(y_i-\hat{y}_i)^2]. \tag{18}$$

Now using the equipartition assumption taken from [2]

$$\mathrm{E}[(y_i - \hat{y}_i)^2] = \frac{\mathrm{E}[\Delta Q]}{N q_i} \tag{19}$$

that holds at the steady state, Eq. (18) becomes

$$\mathrm{E}[\|\mathbf{Q}(\hat{\mathbf{y}} - \mathbf{y})\|^2] = \sum_{i=1}^{N} q_i^2 \frac{\mathrm{E}[\Delta Q]}{N q_i} = \frac{\mathrm{Tr}[\mathbf{Q}]}{N} \mathrm{E}[\Delta Q]. \tag{20}$$

Inserting this in (14), we finally obtain with (16) the expected fitness deviation at the steady state

$$\mathrm{E}[\Delta F] \geq \frac{N\varepsilon^2 \mathrm{Tr}[\mathbf{Q}]}{8\mu^2 c_{\mu/\mu,\lambda}^2} \left(1 + \sqrt{1 + \frac{8\mu^2 c_{\mu/\mu,\lambda}^2 \mathrm{Tr}[\mathbf{Q}^2]}{\mathrm{Tr}[\mathbf{Q}]^2}.} \right). \tag{21}$$

Equation (19) provides also an estimate for the steady state variance of an y_i component. Considering (19), one sees that

$$\mathrm{Var}[y_i] = \frac{\mathrm{E}[\Delta F]}{N q_i}, \tag{22}$$

therefore, using (19) we are able to estimate the variances of the parents at the steady state measured in principal axis directions.

While $\mathrm{E}[\Delta F]$ describes the expected deviation of F_a from the optimum state *with* actuator noise, testing the correctness and approximation quality of the equal sign in (21) can be easier performed by measuring the average of the difference $\tilde{\Delta} F$ defined by

$$\tilde{\Delta} F := Q(\hat{\mathbf{y}}) - F_\mathrm{a}(\mathbf{y}). \tag{23}$$

Here, $\hat{\mathbf{y}}$ is given by (7). Using (2) and the first line in (16), one gets

$$\tilde{\Delta} F = Q(\hat{\mathbf{y}}) - Q(\hat{\mathbf{y}} + \mathbf{z}) + \Delta F. \tag{24}$$

Since

$$Q(\hat{\mathbf{y}} + \mathbf{z}) = \underbrace{\mathbf{b}^\mathrm{T}\hat{\mathbf{y}} - \hat{\mathbf{y}}^\mathrm{T}\mathbf{Q}\hat{\mathbf{y}}}_{=Q(\hat{\mathbf{y}})} + \underbrace{(\mathbf{b}^\mathrm{T} - 2\hat{\mathbf{y}}^\mathrm{T}\mathbf{Q})}_{=\mathbf{0}}\,\mathbf{z} - \mathbf{z}^\mathrm{T}\mathbf{Q}\mathbf{z} \tag{25}$$

(24) becomes

$$\tilde{\Delta} F = \mathbf{z}^\mathrm{T}\mathbf{Q}\mathbf{z} + \Delta F. \tag{26}$$

Now taking the expectation, one finally obtains

$$\mathrm{E}[\tilde{\Delta} F] = \varepsilon^2 \mathrm{Tr}[\mathbf{Q}] + \mathrm{E}[\Delta F]. \tag{27}$$

As one can see, the average deviation from the optimum without actuator noise comprises two terms: the constant term $\varepsilon^2\mathrm{Tr}[\mathbf{Q}]$ independent of the ES used and a strategy specific part $\mathrm{E}[\Delta F]$ given by the equal sign in (21). $\mathrm{E}[\tilde{\Delta F}]$ can be easily tested in ES runs: Since $Q(\hat{\mathbf{y}})$ is known for the models considered, it suffices to calculate the mean fitness over *all* offspring generated after reaching the vicinity of the steady state.[2]

4 Comparision with Experiments

The behavior of the $(\mu/\mu_I, \lambda)$-ES on the actuator noise function class (1), (2) have been tested on three ellipsoidal test functions given in Table 1 for dimensionality $N = 30$ and $N = 100$. Q_1 and Q_2 are axes-parallel ellipsoids. Q_3 has

Table 1. Definitions and properties of the actuator noise test functions.

	Q_1	Q_2	Q_3
$Q(\mathbf{y}) :=$	$-\sum_{i=1}^N iy_i^2$	$-\sum_{i=1}^N i^2y_i^2$	$-\sum_{j=1}^N \left(\sum_{i=1}^j y_i\right)^2$
$(\mathbf{Q})_{i,k} =$	$i\delta_{ij}$	$i^2\delta_{ij}$	$\min[N-i+1, N-j+1]$
$\mathrm{Tr}[\mathbf{Q}]_{N=30} =$	465	9455	465
$\mathrm{Tr}[\mathbf{Q}^2]_{N=30} =$	9455	5273999	144305
$\mathrm{Tr}[\mathbf{Q}]_{N=100} =$	5050	338350	5050
$\mathrm{Tr}[\mathbf{Q}^2]_{N=100} =$	338350	2050333330	17003350

a certain non-parallel orientation. Since we are using $(\mu/\mu_I, \lambda)$-ES (see Appendix A for its definition) with isotropic mutations, the orientation of the ellipsoid does not influence the performance of the strategy. However, Q_3 possesses a dominating eigenvalue, such that the shape of this ellipsoid resembles a distorted discus.

Similar to observations made on the behavior of ES on ellipsoidal test functions with fitness noise in [2], the σ control rule based on cumulative step-length adaptation (CSA) [6,7,8,9] does not work well on the test functions when the non-sphericity gets too large. This is shown in Fig. 1. The CSA-ES is not able to get close to the steady state but exhibits premature convergence: The mutation strength σ quickly reaches values too small for further object parameter evolution. There is a remedy to prevent this behavior by keeping σ above a certain limit σ_0. However, choosing σ_0 is a nontrivial task. Clearly, one should consider the covariance matrix adaptation (CMA-ES, [8]) instead, however, this is beyond the scope of this paper.

Figure 2 compares the predictive quality of (27) using (21) with ES runs. The data points (displayed as dots) have been obtained by recording the fitness values

[2] This assumes that the mutation strenght σ is sufficiently small. If this is not fulfilled, the fitness of the parental centroid must be evaluated at each generation in order to obtain $\tilde{\Delta F}$.

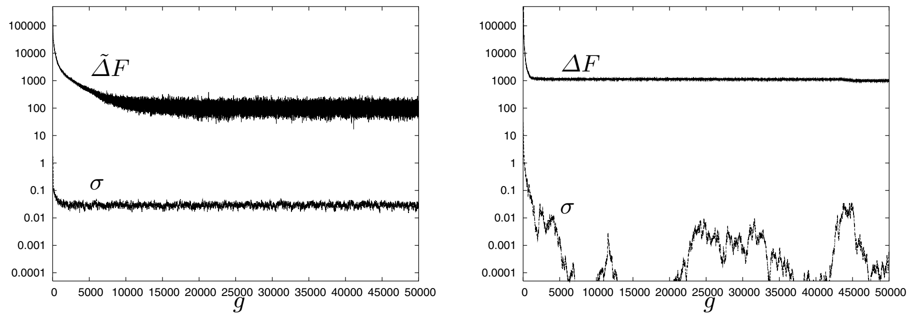
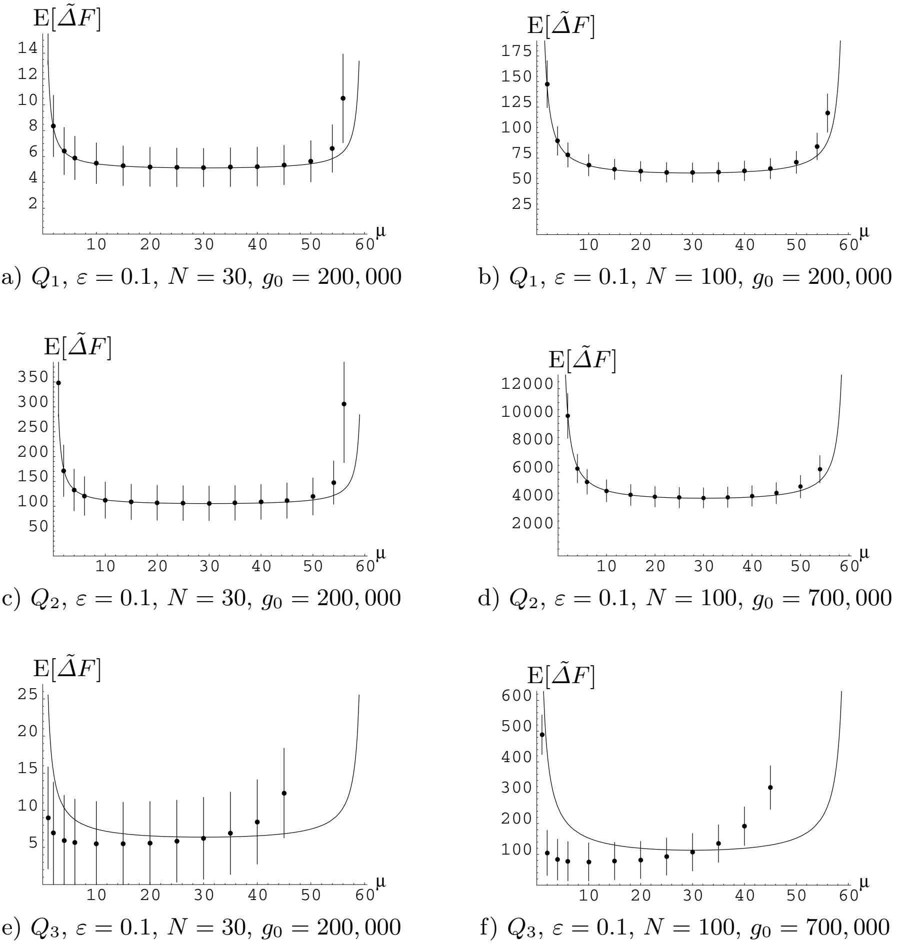


Fig. 1. Evolution dynamics of the $(20/20_I, 60)$-ES on the test function Q_2, $N = 30$, with actuator noise strength $\varepsilon = 0.1$. Adaptation of mutation strength σ is by σSA-ES (left-hand side) and by CSA (right-hand side). One observes the typical behavior of EAs on noisy problems: The fitness values reach a certain steady state distribution the expected value of which deviates from the optimum.

of the parental centroid states over a number of 200,000 generations starting after a number of generations g_0 (transient time for reaching the vicinity of the steady state). Since CSA-ES can exhibit premature convergence, the σSA-ES has been used. As one can see, the theory predicts the steady state behavior of the ES on Q_1 and Q_2 well (leaving aside the cases $\mu = 1$ and $\mu/\lambda \approx 1$). Unfortunately this does not hold for Q_3. In [2] the same test functions have been investigated, however, disturbed by fitness noise. There the authors found a good predictive quality on Q_3. Therefore, the reason for the deviations observed must be in the approximative decomposition (4): It has been assumed that the stochastics can be well approximated by a normal distribution. While this is indeed correct for Q_1 and Q_2 (actually, both functions reach normality exactly for $N \to \infty$) this is not the case for Q_3. As have already been mentioned, Q_3 has an eigenvalue spectrum where the ratio of the largest eigenvalue q_1 to the second largest eigenvalue q_2 approaches 9 (from below) as $N \to \infty$. This is in contrast to Q_1 and Q_2 where this ratio goes to 1. Even worse, considering the ratio $q_1/\sum_{i=2}^{N} q_i$ one finds (numerically) that it approaches ≈ 4.279. In other words, the isolated large eigenvalue q_1 prevents Q_3 from reaching normality for $N \to \infty$ by violating the Lindeberg condition (see, e.g., [5]) and the central limit theorem of statistics does not apply. That is why, we do not observe an improved prediction quality for the $N = 100$ case compared to $N = 30$. The fitness noise produced by Q_3 has a high degree of skewness. The corresponding theory for non-Gaussian noise remains still to be developed.

5 Conclusions and Outlook

In this paper the impact of actuator noise on the steady state behavior of $(\mu/\mu_I, \lambda)$-ES optimizing general quadratic fitness functions has been analyzed. It has been shown that the decomposition method of [3] together with the equipar-



a) Q_1, $\varepsilon = 0.1$, $N = 30$, $g_0 = 200,000$
b) Q_1, $\varepsilon = 0.1$, $N = 100$, $g_0 = 200,000$
c) Q_2, $\varepsilon = 0.1$, $N = 30$, $g_0 = 200,000$
d) Q_2, $\varepsilon = 0.1$, $N = 100$, $g_0 = 700,000$
e) Q_3, $\varepsilon = 0.1$, $N = 30$, $g_0 = 200,000$
f) Q_3, $\varepsilon = 0.1$, $N = 100$, $g_0 = 700,000$

Fig. 2. Dependence of the expected fitness error $\mathrm{E}[\tilde{\Delta}F]$ 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 $\tilde{\Delta}F$. Missing data points are due to divergence (for μ/λ near 1) and premature convergence (for $\mu = 1$), respectively. The curves are the predictions made by (27) using the equal sign in (21).

tition assumption of [2] can be used to predict accurately the final fitness error (provided that the fitness noise induced is approximately normally distributed).

From the results obtained one can derive recommendations concerning the population sizing in order to get a minimal steady state fitness error $\mathrm{E}[\tilde{\Delta}F]$: Looking at Fig. 2 one sees that – assuming normality of the actuator induced fitness noise (i.e., skipping Q_3) – $\mu/\lambda = 1/2$ yields minimal $\mathrm{E}[\tilde{\Delta}F]$. On the other hand, $\mu/\lambda = 1/2$ is not the optimal population ratio for maximal progress toward the steady state. From sphere model theory we know that for $N \to \infty$ the ratio

$\mu/\lambda \approx 0.27$ should be preferred. Taking the behavior on Q_3 into account, a good population ratio compromise seems to be in the interval $0.2 \ldots 0.3$.

The results obtained might have more far-reaching implications. Consider the general quadratic fitness model as a local attractor model of real-world objective functions under actuator noise. The long term behavior of an ES at the end of an evolution process might be well described by such a fitness model and the steady state predictions of the theory might be valid for more complicated objective functions. Therefore, additional investigations are needed to determine the limitations of the model analysis presented.

As has been mentioned, the CSA-ES using isotropic mutations is not good at these ellipsoidal test functions with noise. It is reasonable to use non-isotropic mutations instead. This leads to the problem of adapting a full covariance matrix describing the distribution of the mutations. This is usually done by the CMA method [8]. Investigating the behavior of the CMA-ES should be one of the next steps in future research.

References

1. H.-G. Beyer. *The Theory of Evolution Strategies.* Natural Computing Series. Springer, Heidelberg, 2001.
2. H.-G. Beyer and D. V. Arnold. The Steady State Behavior of $(\mu/\mu_I, \lambda)$-ES on Ellipsoidal Fitness Models Disturbed by Noise. In E. Cantú-Paz et al., editors, *GECCO-2003: Proceedings of the Genetic and Evolutionary Computation Conference*, pages 525–536, Berlin, Germany, 2003. Springer.
3. 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*, pages 307–328, San Francisco, CA, 2003. Morgan Kaufmann.
4. J. Branke. *Evolutionary Optimization in Dynamic Environments.* Kluwer Academic Publishers, Dordrecht, 2001.
5. M. Fisz. *Wahrscheinlichkeitsrechnung und mathematische Statistik.* VEB Deutscher Verlag der Wissenschaften, Berlin, 1971.
6. 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.
7. 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.
8. N. Hansen and A. Ostermeier. Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation*, 9(2):159–195, 2001.
9. A. Ostermeier, A. Gawelczyk, and N. Hansen. A Derandomized Approach to Self-Adaptation of Evolution Strategies. *Evolutionary Computation*, 2(4):369–380, 1995.
10. I. Rechenberg. *Evolutionsstrategie '94.* Frommann-Holzboog Verlag, Stuttgart, 1994.

11. H.-P. Schwefel. *Numerical Optimization of Computer Models.* Wiley, Chichester, 1981.
12. B. Sendhoff, H.-G. Beyer, and M. Olhofer. On Noise Induced Multi-Modality in Evolutionary Algorithms. In L. Wang, K.C. Tan, T. Furuhashi, J.-H. Kim, and F. Sattar, editors, *Proceedings of the 4th Asia-Pacific Conference on Simulated Evolution and Learning – SEAL*, volume 1, pages 219–224, 2002.
13. S. Tsutsui and A. Ghosh. Genetic Algorithms with a Robust Solution Searching Scheme. *IEEE Transactions on Evolutionary Computation*, 1(3):201–208, 1997.
14. 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.

A Description of the ESs Used

For the simulation of the dynamic behavior of $(\mu/\mu_I, \lambda)$-ES the ES must control the endogenous strategy parameter σ. We used the two standard approaches to this control problem: the σ self-adaptation [11,10] and alternatively the cumulative step size adaptation (CSA) [6,8].

The σ self-adaptation technique is based on the coupled inheritance of object and strategy parameters. Using the notation

$$\langle \mathbf{a} \rangle^{(g)} := \frac{1}{\mu} \sum_{m=1}^{\mu} \mathbf{a}_{m;\lambda}^{(g)} \tag{28}$$

for intermediate recombination (centroid calculation, i.e., averaging over the **a** parameters of the μ best offspring individuals), the $(\mu/\mu_I, \lambda)$-σSA-ES can be expressed in "offspring notation"

$$\forall l = 1, \dots, \lambda : \begin{cases} \sigma_l^{(g+1)} := \langle \sigma \rangle^{(g)} \mathrm{e}^{\tau \mathcal{N}_l(0,1)} \\ \mathbf{y}_l^{(g+1)} := \langle \mathbf{y} \rangle^{(g)} + \sigma_l^{(g+1)} \boldsymbol{\mathcal{N}}_l(\mathbf{0}, \mathbf{1}). \end{cases} \tag{29}$$

That is, each offspring individual (indexed by l) gets its own mutation strength σ. And this mutation strength is used as mutation parameter for producing the offspring's object parameter. In (29) the log-normal update rule for mutating the mutation strength has been used. As learning parameter $\tau = 1/\sqrt{N}$ has been chosen in the simulations.

While in evolutionary self-adaptive ES each individual get its own set of endogenous strategy parameters, cumulative step-size adaptation (CSA) uses a single mutation strength parameter σ per generation to produce all the offspring. This σ is updated by a deterministic rule which is controlled by certain statistics gathered over the course of generations. The statistics used is the so-called (normalized) cumulative path-length $\mathbf{s}$. If $\|\mathbf{s}\|$ is greater than the expected length of a random path, σ is increased. In the opposite situation, σ is decreased. The update rule reads

$$\left.\begin{array}{l} \forall l = 1, \ldots, \lambda : \mathbf{y}_l^{(g+1)} := \langle \mathbf{y} \rangle^{(g)} + \sigma^{(g)} \boldsymbol{\mathcal{N}}_l(\mathbf{0}, \mathbf{1}) \\ \mathbf{s}^{(g+1)} := (1-c)\mathbf{s}^{(g)} + \sqrt{(2-c)c}\frac{\sqrt{\mu}}{\sigma^{(g)}} \left(\langle \mathbf{y} \rangle^{(g+1)} - \langle \mathbf{y} \rangle^{(g)} \right) \\ \sigma^{(g+1)} := \sigma^{(g)} \exp\left(\frac{\|\mathbf{s}^{(g+1)}\| - \overline{\chi}_N}{D \overline{\chi}_N} \right) \end{array} \right\}, \qquad (30)$$

where $\mathbf{s}^{(0)} = \mathbf{0}$ is chosen initially. The recommended standard settings for the cumulation parameter c and the damping constant D are used, i.e., $c = 1/\sqrt{N}$ and $D = \sqrt{N}$. For the expected length of a random vector comprising N standard normal components, the approximation $\overline{\chi}_N = \sqrt{N}(1 - 1/4N + 1/21N^2)$ was used.

B The Progress Coefficient $c_{\mu/\mu,\lambda}$

The progress coefficient $c_{\mu/\mu,\lambda}$ is defined as the expectation of the average over the μ largest samples out of a population of λ random samples from the standard normal distribution. According to [1, p. 247], $c_{\mu/\mu,\lambda}$ can be expressed by a single integral

$$c_{\mu/\mu,\lambda} = \frac{\lambda - \mu}{2\pi} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} \mathrm{e}^{-t^2} \left(\Phi(t)\right)^{\lambda-\mu-1} \left(1 - \Phi(t)\right)^{\mu-1} \mathrm{d}\,t, \qquad (31)$$

where $\Phi(t)$ is the cumulative distribution function of the standard normal variate. The special $c_{\mu/\mu,\lambda}$ values used in this paper are given in the table below.

μ	1	2	4	6	10	15	20
$c_{\mu/\mu,60}$	2.31928	2.12722	1.88199	1.71349	1.47183	1.25171	1.07569
μ	25	30	35	40	45	50	56
$c_{\mu/\mu,60}$	0.924168	0.787546	0.66012	0.537847	0.417235	0.294366	0.134428
μ	58	59	60				
$c_{\mu/\mu,60}$	0.0733524	0.0393098	0				