Fitness Noise and Localization Errors of the Optimum in General Quadratic Fitness Models

Hans-Georg Beyer
Department of Computer Science XI
University of Dortmund
44221 Dortmund, Germany
beyer@zappa.cs.uni-dortmund.de

Dirk V. Arnold
Department of Computer Science XI
University of Dortmund
44221 Dortmund, Germany
arnold@ls11.cs.uni-dortmund.de

Abstract

Evolutionary algorithms are generally believed to perform well in the presence of noise. Thus, they are often used for optimization in noisy environments. It comes as a surprise that hardly more than a handful of studies has dealt with the question of just how well they are doing and what can be done to improve their performance. The present paper presents empirical results regarding the behavior of genetic algorithms and evolution strategies in the presence of fitness noise for a range of objective functions. Bounds for the mean residual location error and the stationary fitness error for optimization of general quadratic fitness models with evolution strategies are derived and compared with measurements.

1 INTRODUCTION

Noise is a common phenomenon in almost all real world optimization problems. It can stem from sources as different as possibly systematic measurement errors in experiments, the need to sample from spaces too large to be treated in their entirety, the use of stochastic simulation procedures, or from interaction with users. Moreover, it has been claimed that voluntarily adding noise can improve convergence reliability in multimodal optimization.

Evolutionary algorithms (EAs) are optimization procedures that are generally believed to work well in noisy environments. This hope is fostered by the conviction that Darwinian evolution can be regarded as an optimization process and that evolution in Nature appears to be highly disturbed, and by the observation that in many applications EAs do seem to be able to cope with noise. However, there is hardly more than a handful of studies dealing explicitly with effects of noise on the convergence behavior of EAs, and most of those that do exist are empirical and it is questionable whether their results can be generalized.

At least two effects of the presence of fitness noise have been recognized. One is reduced convergence velocity, the other one the occurrence of a residual location error, meaning that even after having converged, in the mean the EA will remain at a certain distance on the average from the optimizer. These effects can be observed for genetic algorithms (GAs) as well as for evolution strategies (ESs) and evolutionary programming (EP).

The goal of this paper is to investigate the influence of different fitness noise levels on the residual location error. Section 2 provides short a survey of previous noise-related work in the field of evolutionary optimization. Section 3 introduces the fitness and noise models used in this paper, and Section 4 empirically investigates the behavior of a simple GA applied to the bit-counting function. Section 5 deals with $(1, \lambda)$-ESs. It contains a theoretical analysis of the residual localization error of the optimum and of the stationary fitness error in general quadratic fitness models and compares the results with empirical measurements. Section 6 concludes.

2 PREVIOUS WORK

In spite of the importance of the subject, there are few serious investigations of the effects of noise on the convergence behavior of EAs. A more extensive survey than the following can be found in Beyer (1998b).

In possibly the first attempt to study EAs in noisy environments, Rechenberg (1973) succeeded in calculating the progress rate of a $(1 + 1)$-ES in the presence of fitness noise for an N-dimensional linear function with certain constraints, the so-called corridor model. Beyer (1993) obtained analogous results for a $(1, \lambda)$-ES on the N-dimensional sphere. In particular, he developed a progress rate theory and established a lower bound for the residual location error. Empirical research by Hammel and Bäck (1994) confirmed Beyer's predictions. The progress rate theory was extended to $(1, \lambda)$-ESs with rescaled mutations in Beyer (1998a). In a similar spirit, Miller and Goldberg (1997) presented performance analyses of GAs in the presence of noise. Their investigations mainly concentrate on the fitness dynamics of the bit-counting function.

Resampling, i.e. averaging over a number of independent fitness measurements at one location, and increasing the population size are two techniques known to improve the performance of EAs in noisy environments at the price of increased computational costs. Research into the question of which technique is preferable was initiated by Fitzpatrick and Grefenstette (1988) who empirically investigated fitness noise in GAs. That their results cannot easily be generalized was shown in research by Beyer (1993) and Hammel and Bäck (1994) who investigated similar questions in the realm of ESs, but obtained differing results. Recently, Stagge (1998) readdressed the problem of resampling and suggested to use simple statistical tests to minimize the number of times the objective function is resampled. His experiments explored the trade-off between the number of function evaluations and the quality of the solution that can be obtained. Nissen and Propach (1998) empirically investigated the robustness of population-based versus point-based optimization methods for a number of parameter optimization problems in noisy environments and reported that their experiments favor population-based methods.

3 FITNESS AND NOISE MODELS

The objective functions used in the present study extend the realm of spherical functions used by Beyer (1998b) in several ways.

To evaluate GA performance, the bit counting function

$$F(\mathbf{y}) = \frac{1}{N}\sum_{i=1}^{N} y_i, \tag{1}$$

where $y_i \in \{0, 1\}$, is used as fitness function. In that case, the residual distance $\langle R \rangle$ from the optimizer is naturally defined as the proportion of bits of value zero.

The behavior of $(1, \lambda)$-ESs is theoretically investigated for a general quadratic fitness model of the form

$$F(\mathbf{y}) = \mathbf{b}^T\mathbf{y} - \mathbf{y}^T\mathbf{Q}\mathbf{y}, \tag{2}$$

where $\mathbf{b}$ and $\mathbf{y} = (y_1, \ldots, y_N)^T$ are N-dimensional real-valued vectors and $\mathbf{Q} = \mathbf{Q}^T$ is a positive definite $N \times N$ matrix. The maximum $\hat{F}$ of F is attained at $\hat{\mathbf{y}} = \text{argmax}_{\mathbf{y}} F(\mathbf{y})$ and can be computed by solving

$$\mathbf{b} = 2\mathbf{Q}\hat{\mathbf{y}} \tag{3}$$

for $\hat{\mathbf{y}}$. Special cases of Equation (2) selected for empirical investigation in Section 5.2 are the quadratic sphere model

$$F(\mathbf{y}) = -\sum_{i=1}^{N} y_i^2 \tag{4}$$

with $\mathbf{b} = \mathbf{0}$ and $\mathbf{Q} = \mathbf{1}$, two coordinate axes-aligned elliptical functions

$$F(\mathbf{y}) = -\sum_{i=1}^{N} i y_i^2 \tag{5}$$

and

$$F(\mathbf{y}) = -\sum_{i=1}^{N} (i y_i)^2 \tag{6}$$

with $\mathbf{b} = \mathbf{0}$ and diagonal matrices $\mathbf{Q}$ with diagonal entries $1, 2, \ldots, N$ and $1, 4, \ldots, N^2$, respectively, and Schwefel's function

$$F(\mathbf{y}) = -\sum_{i=1}^{N} \left(\sum_{j=1}^{i} y_j \right)^2 \tag{7}$$

with $\mathbf{b} = \mathbf{0}$ and matrix $\mathbf{Q}$ consisting of entries $q_{ij} = \min\{N-i+1, N-j+1\}$. In all four cases the residual distance $\langle R \rangle$ is simply the average Euclidean distance between the location of the parent individual and the optimizer $\hat{\mathbf{y}} = \mathbf{0}$.

Fitness noise is modeled by adding a normally distributed noise term

$$\epsilon(F(\mathbf{y})) = \sigma_\epsilon(F(\mathbf{y}))\mathcal{N},$$

where $\mathcal{N}$ is a standard normally distributed random variable, with mean 0 and standard deviation $\sigma_\epsilon(F(\mathbf{y}))$ to $F(\mathbf{y})$. Choosing $\sigma_\epsilon(F(\mathbf{y}))$ proportional to $F(\mathbf{y})$ allows for the modeling of relative measuring errors. In the following, however, σ_ϵ is assumed to be constant. It is the disturbed function

$$F_{\sigma_\epsilon}(\mathbf{y}) = F(\mathbf{y}) + \sigma_\epsilon\mathcal{N}$$

which is accessible to the EA for fitness computations for selection purposes.

4 GA PERFORMANCE

Figure 1 illustrates how a simple GA with uniform crossover with probability one and without mutations behaves when optimizing the noisy version of the bit-counting function introduced in Equation (1) with bit strings of length $N = 100$. Both proportionate selection and tournament selection with tournament size 2 have been investigated for population sizes 60 and 120.

Obviously, the residual distance $\langle R \rangle$, defined as the proportion of bits of value 0, approaches a steady state value R_∞ which monotonically increases with the noise strength. For the selected range of noise levels, the dependence of this steady state value on the noise strength σ_ϵ is less severe for proportionate selection than for tournament selection. For low noise levels, tournament selection outperforms proportionate selection. For high noise levels, proportionate selection seems to have slight advantages over tournament selection. Increasing the population size leads to better performance in either case. Simulations not presented here have shown that introducing mutations prevents premature convergence at the cost of increasing R_∞. These observations are in accordance with the results of Beyer (1998b) who investigated GA performance on a sphere model. A theoretical analysis of these effects in the realm of GAs is still lacking.

5 ES PERFORMANCE

In this section, the behavior of a $(1, \lambda)$-ES is analyzed theoretically for objective functions of the form of Equation (2). Then the results of the theory are verified experimentally using the objective functions previously introduced, and the limitations of the theoretical approach are discussed.

A $(1, \lambda)$-ES uses a population consisting of a single individual which spawns λ offspring individuals in every generation by adding normally distributed vectors $\mathbf{z}$ with mean 0 and variances σ^2 to the location $\mathbf{y}$ of the parent individual. The best of the offspring individuals survives and serves as the parent individual for the next generation. The step size σ can be controlled by self-adaptation, but is regulated externally in the experiments below to single out the effects of noise.

5.1 RESIDUAL DISTANCE THEORY

Using Equations (2) and (3), the fitness value of an offspring individual generated by a mutation $\mathbf{z}$ is

$$\begin{aligned} F_{\sigma_\epsilon}(\mathbf{y} + \mathbf{z}) &= F_{\sigma_\epsilon}(\mathbf{y}) + (\mathbf{b}^T - 2\mathbf{y}^T\mathbf{Q})\mathbf{z} - \mathbf{z}^T\mathbf{Q}\mathbf{z} \\ &= F_{\sigma_\epsilon}(\mathbf{y}) + 2(\hat{\mathbf{y}}^T - \mathbf{y}^T)\mathbf{Q}\mathbf{z} - \mathbf{z}^T\mathbf{Q}\mathbf{z}. \end{aligned}$$

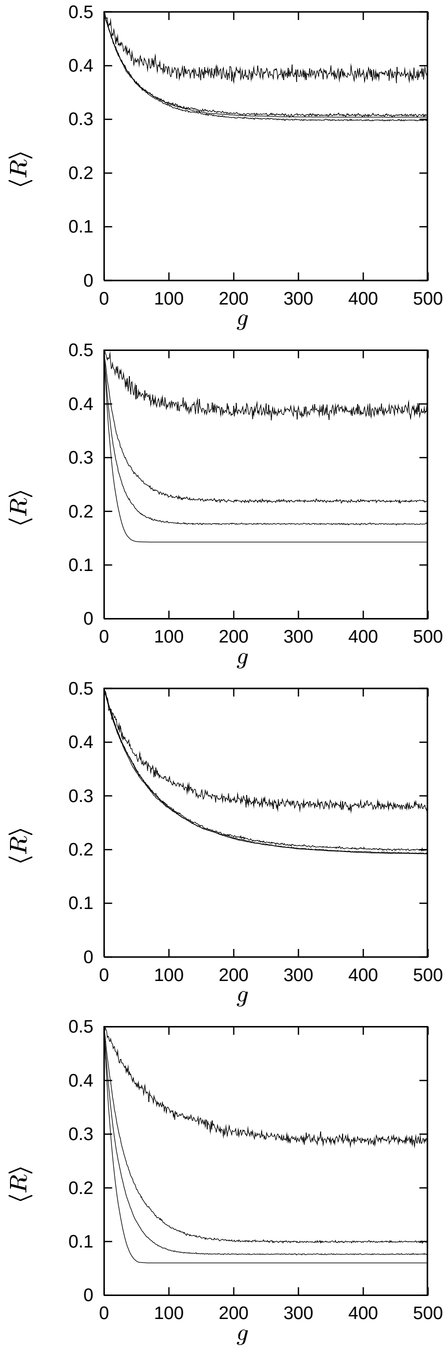

Figure 1: The mean residual distance $\langle R \rangle$ as a function of the generation number g averaged over 100 independent runs. The graphs display, from top to bottom, the results for proportionate selection and population size 60, tournament selection with tournament size 2 and population size 60, proportionate selection and population size 120, and tournament selection with tournament size 2 and population size 120. Each graph displays the curves for noise levels σ_ϵ of 0.00, 0.05, 0.10, and 0.50, with the lower curves corresponding to lower noise levels.

Defining the local quality function $Q_{\mathbf{y}}(\mathbf{z}) := F_{\sigma_\epsilon}(\mathbf{y}+\mathbf{z}) - F_{\sigma_\epsilon}(\mathbf{y})$ we have[1]

$$Q_{\mathbf{y}}(\mathbf{z}) = \mathbf{a}^T\mathbf{z} - \mathbf{z}^T\mathbf{Q}\mathbf{z}, \tag{8}$$

where $\mathbf{a} = 2\mathbf{Q}(\hat{\mathbf{y}} - \mathbf{y})$.

5.1.1 Deriving the Stationarity Condition

As shown in Beyer (1993), for the noisy $(1,\lambda)$-ES sphere model defined in Equation (4) with sufficiently large parameter space dimension N, the expected normalized distance change $\varphi^* = \varphi N/R$ toward the optimum is given by

$$\varphi^* = \frac{c_{1,\lambda}\sigma^{*2}}{\sqrt{\sigma^{*2}+\sigma_\epsilon^{*2}}} - \frac{\sigma^{*2}}{2}, \tag{9}$$

where $c_{1,\lambda}$ denotes the $(1,\lambda)$-progress coefficient, $\sigma^* = \sigma N/R$ is the normalized mutation strength, and the normalized noise strength is given by

$$\sigma_\epsilon^* = \sigma_\epsilon \frac{N}{RQ'}, \quad Q' = \frac{dQ}{dR}, \tag{10}$$

where R is the parental distance to the optimum. Progress towards the optimum is expected for $\varphi^* > 0$ and stationarity is reached for $\varphi^* \to 0$. With $\varphi^* = 0$ we obtain from Equation (9) $2c_{1,\lambda} = \sqrt{\sigma^{*2}+\sigma_\epsilon^{*2}} \geq \sigma_\epsilon^*$, and with Equation (10) it follows

$$\frac{\sigma_\epsilon N}{2RQ'} \leq c_{1,\lambda}. \tag{11}$$

Thus, for $\sigma^{*2} \ll \sigma_\epsilon^{*2}$ the equal sign in Equation (11) defines the stationary residual distance $R = R_\infty$ to the optimum for the $(1,\lambda)$-ES sphere model.

The idea pursued in the following is to generalize this result to the general quadratic fitness model in Equation (8) by substituting the mean radius $R_{\langle\varkappa\rangle}$ as defined in Beyer (1995) for R and the length of the gradient of $Q_{\mathbf{y}}(\mathbf{z})$ at the parental state $\mathbf{y}$ for Q'. The latter is given by

$$\|\nabla_{\mathbf{z}} Q|_{\mathbf{z}=0}\| = \|2\mathbf{Q}(\hat{\mathbf{y}} - \mathbf{y})\| = \|\mathbf{a}\|, \tag{12}$$

the former by Equation (89) from Beyer (1995, p.108) which reads

$$R_{\langle\varkappa\rangle} = \frac{N-1}{2}\frac{\|\mathbf{a}\|}{\mathrm{Tr}[\mathbf{Q}] - \mathbf{a}^T\mathbf{Q}\mathbf{a}/\|\mathbf{a}\|^2}.$$

In order to simplify further calculations, sufficiently large N is assumed. Furthermore, the Rayleigh quotient $\mathbf{a}^T\mathbf{Q}\mathbf{a}/\|\mathbf{a}\|^2$, which is bounded by the largest and the smallest eigenvalues of $\mathbf{Q}$, is assumed to be small compared to $\mathrm{Tr}[\mathbf{Q}]$. This is admissible if N is sufficiently large and the eigenvalues of $\mathbf{Q}$ are concentrated around $\mathrm{Tr}[\mathbf{Q}]/N$. In that case $R_{\langle\varkappa\rangle}$ can be simplified to

$$R_{\langle\varkappa\rangle} = \frac{N}{2}\frac{\|\mathbf{a}\|}{\mathrm{Tr}[\mathbf{Q}]}. \tag{13}$$

With Equations (12) and (13) in Equation (11) we obtain the stationarity condition

$$\|\mathbf{Q}(\hat{\mathbf{y}} - \mathbf{y})\| \geq \frac{1}{2}\sqrt{\frac{\sigma_\epsilon \mathrm{Tr}[\mathbf{Q}]}{c_{1,\lambda}}} =: \sqrt{K} \tag{14}$$

for general quadratic fitness models. Equation (14) demonstrates that for $\sigma_\epsilon > 0$ there always is a residual distance $R_\infty = \|\hat{\mathbf{y}} - \mathbf{y}\|$. This residual distance can be reduced by increasing the number of offspring as $c_{1,\lambda}$ is monotonically increasing with λ. Note that Inequality (14) contains the residual distance R_∞ of the sphere model derived in Beyer (1993) as a special case. It will be shown in Section 5.2 that in experiments Inequality (14) yields good results as long as $\mathbf{a}^T\mathbf{Q}\mathbf{a}/\|\mathbf{a}\|^2 \ll \mathrm{Tr}[\mathbf{Q}]$.

5.1.2 Bounds on the Stationary Fitness Error

The residual optimum location error necessarily implies a stationary fitness error, i.e. an expected residual deviation of the fitness without noise F_∞ from the optimum value $\hat{F} = F(\hat{\mathbf{y}})$. It is the goal of this section to derive a bound on this deviation. Note that the fitness without noise F is not directly accessible to the EA; however, it can be obtained as the average over an infinite number of samples.

Writing Inequality (14) as

$$(\hat{\mathbf{y}}^T - \mathbf{y}^T)\mathbf{Q}^2(\hat{\mathbf{y}} - \mathbf{y}) \geq K \tag{15}$$

it becomes clear that the stationarity condition for the equal sign defines an $(N-1)$-dimensional elliptical manifold which, except for the spherical case, is not similar to the equifitness manifolds $F(\mathbf{y}) = \mathit{const.}$ This is so because the eigenvalues of $\mathbf{Q}$ and $\mathbf{Q}^2$ in general cannot be transformed into each other by a linear transformation.

[1] This definition can serve as the basis for a quality gain analysis of the noisy $(1,\lambda)$-ES along the lines of the theory for the unperturbed $(1,\lambda)$-ES developed in Beyer (1994). The $(1,\lambda)$-quality gain $\overline{Q}_{1,\lambda}$ is defined as $\mathrm{E}[Q_{\mathbf{y}}(\mathbf{z}_{1;\lambda})]$, the expected quality gain for the best offspring, denoted by $\mathbf{z}_{1;\lambda}$. If this quantity was known the residual distance could be computed from the stationarity condition $\overline{Q}_{1,\lambda}(\mathbf{y}) = 0$.

However, instead of extending the derivation in Beyer (1994) to the noisy case, we will rather show how the concept of mean radius $R_{\langle\varkappa\rangle}$ developed in Beyer (1995) can be applied to the noisy sphere theory outlined in Beyer (1993) in order to derive a stationarity condition.

The idea for computing F_∞ is to consider Inequality (15) as a constraint on the optimization process performed by the $(1,\lambda)$-ES. In the following, we solve this constrained optimization problem by Lagrange's method. In order to simplify the calculations, we first change the coordinate basis to the eigensystem of $\mathbf{Q}$. Let $\mathbf{Q}$ have normalized eigenvectors $\mathbf{u}_i$ with respective eigenvalues q_i. Then we have $\mathbf{u}_i^T\mathbf{u}_j = \delta_{ij}$ and

$$\mathbf{Q} = \sum_{i=1}^{N} q_i \mathbf{u}_i \mathbf{u}_i^T \quad \text{and} \quad \mathbf{1} = \sum_{i=1}^{N} \mathbf{u}_i \mathbf{u}_i^T .$$

With this basis, Equation (2) can be written as

$$F(\mathbf{y}) = \sum_{i=1}^{N} b_i x_i - \sum_{i=1}^{N} q_i x_i^2 , \tag{16}$$

where $x_i := \mathbf{u}_i^T \mathbf{y}$ and $b_i := \mathbf{b}^T \mathbf{u}_i$. The components of the unconstrained maximizer $\hat{\mathbf{x}}$ from Equation (3) can easily be calculated from Equation (16) as

$$\hat{x}_i = \frac{b_i}{2q_i}, \tag{17}$$

and therefore

$$\hat{F} = F(\hat{\mathbf{x}}) = \frac{1}{4}\sum_{i=1}^{N} \frac{b_i^2}{q_i}. \tag{18}$$

Inequality (15) is transformed into

$$K \le \sum_{i=1}^{N} (\hat{x}_i - x_i)^2 q_i^2 . \tag{19}$$

Considering the equal sign in Inequality (19) only and inserting $\hat{x}_i$ from Equation (17), the constraint reads

$$K = \sum_{i=1}^{N} (\frac{1}{4} b_i^2 - b_i q_i x_i + x_i^2 q_i^2). \tag{20}$$

Combining Equations (16) and (20) the Lagrange function can be written as

$$L(x_1,\ldots,x_N,\vartheta) = \sum_{i=1}^{N} (b_i x_i - q_i x_i^2) + \vartheta \left[\sum_{i=1}^{N} (\frac{1}{4} b_i^2 - b_i q_i x_i + x_i^2 q_i^2) - K \right].$$

Computing the derivatives one obtains

$$\frac{\partial L}{\partial x_i} = b_i - 2q_i x_i - b_i q_i \vartheta + 2 x_i q_i^2 \vartheta \tag{21}$$

$$\frac{\partial L}{\partial \vartheta} = \sum_{i=1}^{N} (\frac{1}{4} b_i^2 - b_i q_i x_i + x_i^2 q_i^2) - K. \tag{22}$$

Equating the derivatives in Equation (21) to zero yields the existence of an eigenvalue q_j such that

$$\vartheta = \frac{1}{q_j}$$

and

$$\tilde{x}_i = \frac{b_i}{2q_i} \quad \text{for } i \neq j. \tag{23}$$

With Equation (22) one finds for $\partial L/\partial\vartheta = 0$ using Equation (23)

$$\sum_{i\neq j} \underbrace{(\frac{1}{4} b_i^2 - b_i q_i \tilde{x}_i + \tilde{x}_i^2 q_i^2)}_{0} + \frac{1}{4} b_j^2 - b_j q_j \tilde{x}_j + q_j^2 \tilde{x}_j^2 = K,$$

and therefore

$$b_j \tilde{x}_j - q_j \tilde{x}_j^2 = \frac{1}{q_j}\left(\frac{b_j^2}{4} - K\right). \tag{24}$$

This quadratic equation can be solved for $\tilde{x}_j$. However, to compute $F_\infty = F(\tilde{x}_1,\ldots,\tilde{x}_N)$ this is not necessary. Using Equations (23) and (24) and subsequently Equation (18) in Equation (16) we get

$$\begin{aligned} F_\infty &= \sum_{i\neq j} (b_i \tilde{x}_i - q_i \tilde{x}_i^2) + b_j \tilde{x}_j - q_j \tilde{x}_j^2 & (25)\\ &= \frac{1}{4}\sum_{i\neq j} \frac{b_i^2}{q_i} + \frac{1}{q_j}\frac{b_j^2}{4} - \frac{K}{q_j} & (26)\\ &= \hat{F} - \frac{K}{q_j}. & (27) \end{aligned}$$

As the task at hand is maximization q_j must be the largest eigenvalue of $\mathbf{Q}$ and we get with the definition of K in Equation (14) a lower bound

$$\hat{F} - F_\infty \ge \frac{\sigma_\epsilon \mathrm{Tr}[\mathbf{Q}]}{4 c_{1,\lambda} q_j} \tag{28}$$

for the stationary fitness error.

5.2 SIMULATION RESULTS

Figures 2 and 3 show the residual location error and the deviation from the optimal fitness over the number of generations for a $(1,12)$-ES for fitness functions (4), (5), (6), and (7) with parameter space dimension $N = 40$. In all cases the population was randomly initialized and consisted of a single individual with a single strategy parameter σ. So as to not introduce effects brought about by self-adaptation, the step size was set externally to $\sigma_g = \sigma_0 \exp(-kg)$, where g denotes the generation number and constants σ_0 and k were determined by trial and error. Values of $\sigma_0 = 0.4$ and $k = 2 \cdot 10^{-4}$ for functions (4), (5), and (6) and $\sigma_0 = 0.001$ and $k = 5 \cdot 10^{-5}$ for function (4) proved to be useful. Many simulations with this and also with other control schemes were run to ensure that the observed convergence is not the result of ill-adjusted step sizes. All graphs are averaged over 20 independent runs.

Little surprisingly, not only the bound in Equation (14), but also that in (28) prove to be very good approximations for the simple sphere model. The small deviations present may be attributed to the parameter space dimension of $N = 40$ while $N \to \infty$ was assumed in the theoretical derivation of the bounds. More remarkably, Equation (14) turns out to be very useful for the first two elliptical fitness functions as well. The quality of the bounds provided by Equation (28) is no longer as good, but still within an order of magnitude. One reason for the reduced accuracy can be found in the static treatment of the constraint in Equation (15). A better approximation could be obtained if the fluctuations of the population around the static solution $\tilde{\mathbf{x}}$ were taken into account. With fluctuations $\delta\mathbf{x}$ it follows

$$\begin{aligned}\tilde{F}_\infty :&= F(\tilde{\mathbf{x}} + \delta\mathbf{x}) \\ &= F(\tilde{\mathbf{x}}) + (\mathbf{b}^T - 2\tilde{\mathbf{x}}\mathbf{Q})\delta\mathbf{x} - \delta\mathbf{x}^T\mathbf{Q}\delta\mathbf{x},\end{aligned}$$

and as $F(\tilde{\mathbf{x}}) = F_\infty$ and $\mathrm{E}[\delta\mathbf{x}] = \mathbf{0}$ due to the stationarity of $\tilde{\mathbf{x}}$ the mean

$$\mathrm{E}[\tilde{F}_\infty] = F_\infty - \mathrm{E}[\delta\mathbf{x}^T\mathbf{Q}\delta\mathbf{x}] < F_\infty$$

proves to be a better bound for the values observed in Figure 3 than the static approximation F_∞.

For Schwefel's function, the situation is quite different. The eigenvalue spectrum of its $\mathbf{Q}$ matrix is dominated by a single large value, with the next largest one following at some distance. In that case, neglecting the Rayleigh quotient is impractical, and especially the bounds provided by Inequality (28) more substantially deviate from the observed values.

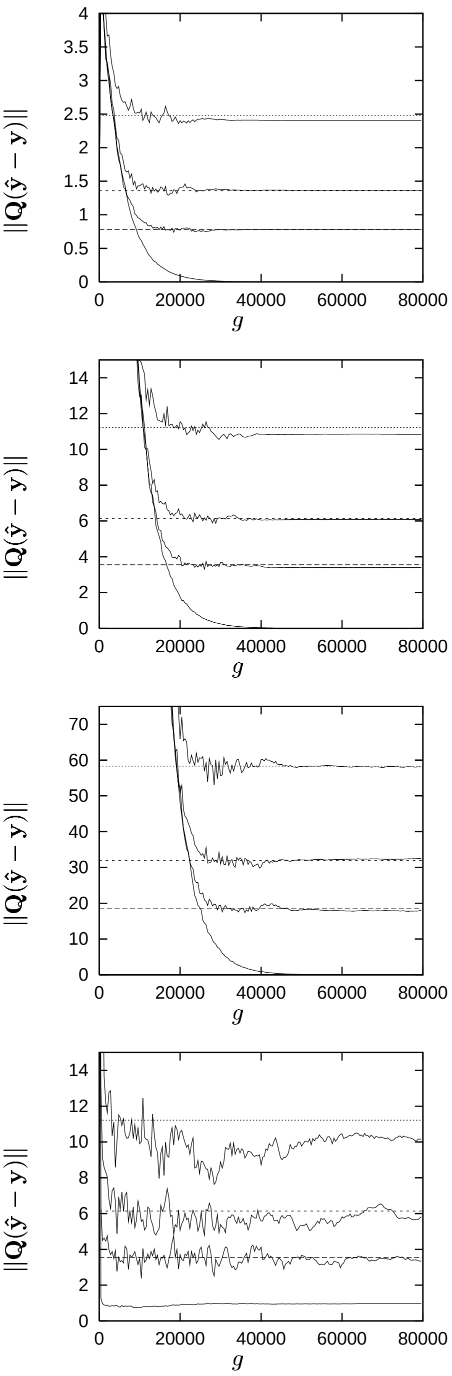

Figure 2: The transformed residual location error $\|\mathbf{Q}(\hat{\mathbf{y}} - \mathbf{y})\|$ over the generation number g for a $(1, 12)$-ES and, from top to bottom, fitness functions (4), (5), (6), and (7). Each graph displays the curves for noise levels σ_ϵ of 0.0, 0.1, 0.3, and 1.0 as well as the bounds $\sqrt{K}$ predicted by Equation (14).

6 CONCLUSION

To summarize, reduced convergence velocity and the occurrence of a residual location error of the optimum are observable in both GAs and ESs in noisy environments. This paper has provided estimates of a lower bound for the residual location error and the stationary fitness error for $(1, \lambda)$-ESs in general quadratic fitness models with fitness noise present. The results

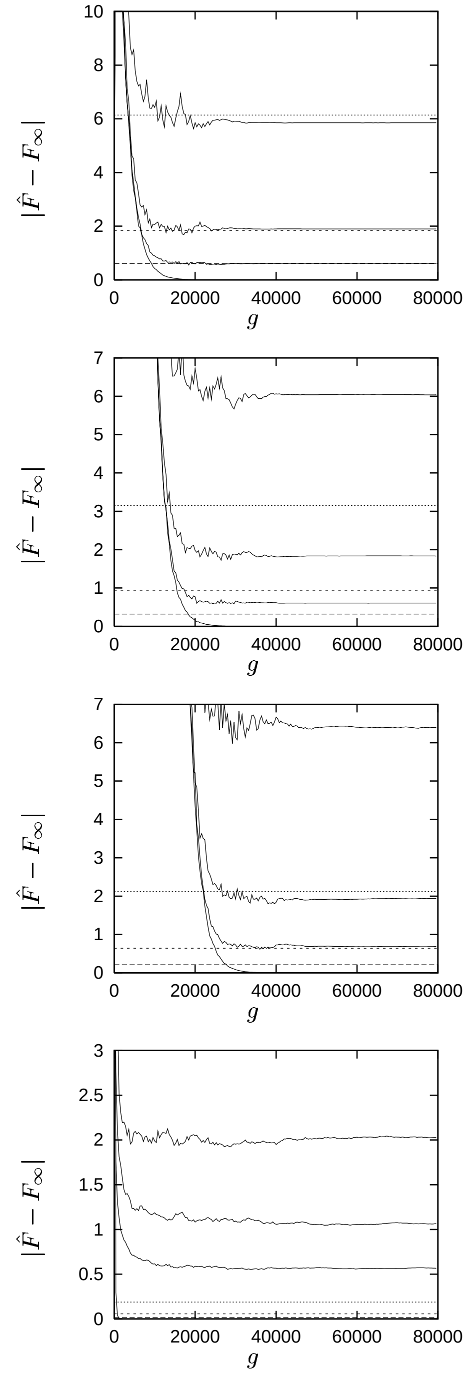

Figure 3: The deviation from the optimal fitness value $|\hat{F} - F_\infty|$ over the generation number g for a $(1,12)$-ES and, from top to bottom, fitness functions (4), (5), (6), and (7). Each graph displays the curves for noise levels σ_ϵ of 0.0, 0.1, 0.3, and 1.0 as well as the bounds predicted by Equation (28).

have been tested for several instances of quadratic fitness models, and have proven to be useful under certain conditions. Notably, the use of the mean curvature as an approximation to the local geometric properties of the fitness function has proven to be quite exact for some functions substantially different from the simple sphere model.

Clearly, the investigation of the effects of fitness noise in quadratic fitness models holds research potential for the future. An extension of the theory developed in Beyer (1994) to the noisy case instead of the ad hoc use of the mean curvature promises to yield better bounds on the residual distance. Finally, extending the theory to the multimembered (μ, λ)-ES and to the $(\mu/\mu, \lambda)$-ES remains as a challenge for the future.

Acknowledgements

This paper is a result of research supported by the Deutsche Forschungsgemeinschaft (DFG). The first author is supported as a Heisenberg fellow under grant Be 1578/4-1, the second author is supported under grant Be 1578/6-1.

References

H.-G. Beyer, "Toward a Theory of Evolution Strategies: Some Asymptotical Results from the $(1\overset{+}{,}\lambda)$-Theory", *Evolutionary Computation*, 1(2), 165-188, (1993).

H.-G. Beyer, "Towards a Theory of 'Evolution Strategies': Progress Rates and Quality Gain for $(1\overset{+}{,}\lambda)$-Strategies on (Nearly) Arbitrary Fitness Functions", in Y. Davidor, R. Männer, and H.-P. Schwefel (eds.), *Parallel Problem Solving from Nature*, 3, 58-67, (Heidelberg, Springer, 1994).

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).

H.-G. Beyer, "Mutate Large, But Inherit Small! On the Analysis of Rescaled Mutations in $(\tilde{1}, \tilde{\lambda})$-ES With Noisy Fitness Data", in A.E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel (eds.), *Parallel Problem Solving from Nature*, 5, 109-118, (Berlin, Springer, 1998a).

H.-G. Beyer, "Evolutionary Algorithms in Noisy Environments: Theoretical Issues and Guidelines for Practice", *Computer Methods in Applied Mechanics and Engineering*, in print, (1998b).

J.M. Fitzpatrick and J.J. Grefenstette, "Genetic Algorithms in Noisy Environments", in P. Langley (ed.), *Machine Learning: Special Issue on Genetic Algo-*

rithms, 3, 101-120, (Dordrecht, Kluwer, 1988).

U. Hammel and T. Bäck, "Evolution Strategies on Noisy Functions. How to Improve Convergence Properties", in Y. Davidor, R. Männer, and H.-P. Schwefel (eds.), *Parallel Problem Solving from Nature*, 3, 159-168, (Heidelberg, Springer, 1994).

B.L. Miller and D.E. Goldberg, "Genetic Algorithms, Selection Schemes, and the Varying Effects of Noise", *Evolutionary Computation*, 4(2), 113-131, (1997).

V. Nissen and J. Propach, "Optimization with Noisy Function Evaluations", in A.E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel (eds.), *Parallel Problem Solving from Nature*, 5, 159-168, (Berlin, Springer, 1998).

I. Rechenberg, *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*, (Stuttgart, Frommann-Holzboog, 1973).

P. Stagge, "Averaging Efficiently in the Presence of Noise", in A.E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel (eds.), *Parallel Problem Solving from Nature*, 5, 188-197, (Berlin, Springer, 1998).