UNIVERSITY OF DORTMUND

REIHE COMPUTATIONAL INTELLIGENCE

COLLABORATIVE RESEARCH CENTER 531

Design and Management of Complex Technical Processes and Systems by means of Computational Intelligence Methods

Predicting the Solution Quality in Noisy Optimization

Hans-Georg Beyer and Silja Meyer-Nieberg

No. CI-160/04

Technical Report ISSN 1433-3325 September 2004

Secretary of the SFB 531 · University of Dortmund · Dept. of Computer Science/XI
44221 Dortmund · Germany

This work is a product of the Collaborative Research Center 531, "Computational Intelligence," at the University of Dortmund and was printed with financial support of the Deutsche Forschungsgemeinschaft.

Predicting the Solution Quality in Noisy Optimization

Hans-Georg Beyer and Silja Meyer-Nieberg

Department of Computer Science XI,
University of Dortmund, D-44221 Dortmund, Germany
{beyer,meyer}@Ls11.cs.uni-dortmund.de

Abstract. Noise often occurs in real-world optimization tasks. Although evolution strategies and other evolutionary algorithms are thought to be robust against the effects of noise, even their performance is degraded. Generally, one observes a reduction of the convergence velocity and a deterioration of the final solution quality.

There are two local progress measures describing the performance of evolution strategies from one generation to the next. The first – the progress rate – is defined directly on the object parameter space, whereas the second, called the quality gain, operates in the space of the fitness values instead. Although both are local performance measures, they can be utilized to derive evolution criteria and steady state conditions. The latter can be used to predict the final solution quality.

We will determine the quality gain for $(1, \lambda)$-ES in the presence of noise. After the derivation of the steady state conditions, we will develop an equation describing the final fitness error for some test functions. The results will be extended to $(\mu/\mu_I, \lambda)$-ES and compared with ES runs using σ-self-adaptation as well as CSA.

Table of Contents

Chapter 1

Introduction

Noise is a common phenomenon in optimization tasks dealing with real-world applications. Typically, one is confronted with measurement errors, actuator vibrations, production tolerances or/and is searching for robust (i.e. insensitive) solutions.

Evolution strategies are nature-inspired search algorithms that move through the search space by means of variation and selection. It is widely believed that they and other evolutionary algorithms are especially good at coping with noisy information due to the use of a population of candidate solutions (see for example [15], [11], [10], or [4]).

However, since noise does deceive the information obtained from the objective function and thus influences the selection process, even population based algorithms are degraded in their performance. Basically, there are two negative effects that can be observed. The convergence velocity is reduced and a final steady state is reached which deviates from the real optimum.

There are two local performance measures that can be used to describe the progress of the ES from one generation to the next. The first – called the progress rate – is defined on the object parameter space whereas the quality gain operates on the fitness values, instead.

The local performance of the $(1 \overset{+}{,} \lambda)$-ES (compare [3]) and the (μ, λ)-ES (compare [4] and [2]) has been studied in noisy spherically symmetric and linear fitness environments. The case of an $(\mu/\mu_I, \lambda)$-ES was examined in [7] applying the progress rate. Using it to derive a steady state condition, the authors were able to determine the solution quality of the steady state, i.e. the final fitness error, in the case of quadratic objective functions. The crucial point of the derivation was the so-called *equipartition assumption*, stating that once the steady state is reached, the weighted distributions of the components to the final fitness error can be supposed to be the same. We will use similar assumptions in order to derive the final fitness error for some selected fitness functions in the case of an $(1, \lambda)$-ES.

In this paper, we will focus on the quality gain of $(1, \lambda)$-ES in the presence of noise. After the development of the equation, we will derive evolution criteria allowing for the prediction of the steady state behavior of an $(1, \lambda)$-ES. The predictions of the quality gain and the final fitness error will then be compared with the results of real ES-runs. Furthermore, we will extend the latter to $(\mu/\mu_I, \lambda)$ strategies and compare their prediction power with real ES runs.

In the concluding section, an outlook will be given presenting a road map for future research.

1.1 Evolution Strategies

Evolution strategies try to optimize a goal or fitness function F by applying variation and selection to a population of candidate solutions mimicking the natural evolution process. The variation operator comprises the processes recombination and mutation. In the case of an $(\mu/\mu_I, \lambda)$-ES, the recombination is performed by computing the centroid of all μ individuals of the parental population. A mutation vector is then added to the centroid for all λ candidate solutions of the offspring population. Usually, these mutations are assumed to obey a normal distribution with a constant standard deviation σ which is called the mutation strength. After creating λ new individuals, the μ best of them are selected to form the next parental population. In the case of an $(1, \lambda)$-ES only one offspring is chosen to be the parent of the next generation.

1.2 Fitness Environments

In this paper, we consider several classes of fitness functions which will be described below. The equation describing the quality gain will be tested on all function types, while cubic functions are not considered for the final fitness error.

The first function class to be considered are quadratic functions given by the general form

$$F_1(\mathbf{y}) := \mathbf{b}^{\mathrm{T}}\mathbf{y} - \mathbf{y}^{\mathrm{T}}\mathbf{Q}\mathbf{y}, \tag{1.1}$$

where $\mathbf{y}$ and $\mathbf{b}$ are N-dimensional real-valued vectors and $\mathbf{Q}$ is a positive definite matrix. They can represent real-world goal functions in the vicinity of an optimum.

In order to test the predictive validity of the quality gain on the class of polynomial fitness functions, we consider two further functions of that type, cubic functions of the form

$$F_2(\mathbf{y}) := \sum_{i=1}^{N} a_i y_i - c_i y_i^3 \tag{1.2}$$

and biquadratic functions, i.e.

$$F_3(\mathbf{y}) := \sum_{i=1}^{N} a_i y_i - c_i y_i^4. \tag{1.3}$$

In contrast to the other test functions, F_2 has two local optima, a minimum at $-\sqrt{a_i/(3c_i)}$ and a maximum at $\sqrt{a_i/(3c_i)}$. For $y_i \to -\infty$, it approaches infinity whereas it goes to $-\infty$ for $y_i \to \infty$.

To test the range of the applicability of the equation describing the quality gain, the bit-counting function OneMax is also considered. It is given by

$$F_4(\mathbf{y}) := \sum_{i=1}^{N} y_i, \tag{1.4}$$

where $y_i \in \{0, 1\}$.

Finally we introduce the function

$$F_5(\mathbf{y}) = c - \sum_{i=1}^{N} |y_i|, \tag{1.5}$$

i.e. the negative sum of the absolute values or – in other words – the negative L_1-norm of $\mathbf{y}$. Its isometric plot forms a rotated N-dimensional hypercube.

Chapter 2

The $(1, \lambda)$-Quality Gain and Steady-State Behavior

2.1 The $(1, \lambda)$-Quality Gain

Let us consider an individual at the position $\mathbf{y}$ in an N-dimensional object parameter space. The position of an offspring is obtained by adding a mutation vector $\mathbf{x}$. The components x_i are assumed to be normally distributed with standard deviation or *mutation strength* σ. The associated change of the fitness function F is

$$Q(\mathbf{x}) := F(\mathbf{y} + \mathbf{x}) - F(\mathbf{y}). \tag{2.1}$$

We consider an $(1, \lambda)$-ES that has to cope with noisy function evaluations and apply the standard fitness noise model assuming that the noise term is added to the fitness function. Therefore, only the perceived local quality change

$$\tilde{Q}(\mathbf{x}) = Q(\mathbf{x}) + \epsilon \tag{2.2}$$

can be observed, where the noise ϵ is modeled by a normally distributed random variable with an expected value of zero. The standard deviation σ_ϵ is called the *noise strength* and is assumed to be constant. Since the measurement of F is disturbed by noise, the seemingly best offspring chosen to be the parent of the next generation is not necessarily the actually optimal candidate. We are interested in the expected value of the local quality gain of this individual which is called the quality gain $\overline{\Delta Q}_{1,\lambda}$ and is defined as

$$\overline{\Delta Q}_{1,\lambda} := \int_{-\infty}^{\hat{Q}} Q p_{1,\lambda}(Q) \, \mathrm{d}Q, \tag{2.3}$$

where $\hat{Q} = \max Q(\mathbf{x})$. In order to continue, we have to develop an expression for $p_{1,\lambda}(Q)$ which denotes the density function of the seemingly best candidate of the offspring population. Applying the concept of noisy order statistics [1], that density is given as

$$p_{1,\lambda}(Q) = p_Q(Q) \lambda \int_{-\infty}^{\infty} p_\epsilon(\tilde{Q}|Q) P(\tilde{Q})^{\lambda-1} \, \mathrm{d}\tilde{Q}, \tag{2.4}$$

where $p_Q(Q)$ is the density function of Q for a single individual and $p_\epsilon(\tilde{Q}|Q)$ is the conditional density function of $\tilde{Q}$ given the undisturbed Q, i.e. $p_\epsilon(\tilde{Q}|Q) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon} \mathrm{e}^{-\frac{(\tilde{Q}-Q)^2}{2\sigma_\epsilon^2}}$. Inserting Eq. (2.4) into Eq. (2.3), we get

$$\overline{\Delta Q}_{1,\lambda} = \lambda \int_{-\infty}^{\hat{Q}} \int_{-\infty}^{\infty} Q p_Q(Q) p_\epsilon(\tilde{Q}|Q) P(\tilde{Q})^{\lambda-1} \, \mathrm{d}\tilde{Q} \, \mathrm{d}Q. \tag{2.5}$$

The upper integration limit of the outer integral can formally be extended to ∞. When using an approximation for $p_Q(Q)$, this will cause an additional approximation error. Therefore, one has to require that the approximate density function drops sufficiently fast for $Q > \hat{Q}$. Under this condition, the resulting error can be expected to be rather small.

In order to proceed, we center Q with its expected value M_Q and its standard deviation S_Q, that is, we transform to the standardized variable $z = (Q - M_Q)/S_Q$ and $p_z(z)\,\mathrm{d}z := p_Q(Q)\,\mathrm{d}Q$. The quality gain is now given as

$$\begin{aligned}\overline{\Delta Q}_{1,\lambda} &= \lambda \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (S_Q z + M_Q) p_z(z) p_\epsilon(\tilde{Q}|S_Q z + M_Q) P(\tilde{Q})^{\lambda-1}\,\mathrm{d}\tilde{Q}\,\mathrm{d}z \\ &= \underbrace{\lambda S_Q \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z p_z(z) p_\epsilon(\tilde{Q}|S_Q z + M_Q) P(\tilde{Q})^{\lambda-1}\,\mathrm{d}\tilde{Q}\,\mathrm{d}z}_{I_1} + \\ &\quad \underbrace{\lambda M_Q \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_z(z) p_\epsilon(\tilde{Q}|S_Q z + M_Q) P(\tilde{Q})^{\lambda-1}\,\mathrm{d}\tilde{Q}\,\mathrm{d}z}_{I_2}. \end{aligned} \tag{2.6}$$

We will treat the two integrals separately. The solution of I_2 is easily obtained. Changing the order of integration and taking into account that $p(\tilde{Q}) = \int_{-\infty}^{\infty} p_z(z) p_\epsilon(\tilde{Q}|S_Q z + M_Q)\,\mathrm{d}z$, we get

$$I_2 = \lambda M_Q \int_{-\infty}^{\infty} p(\tilde{Q}) P(\tilde{Q})^{\lambda-1}\,\mathrm{d}\tilde{Q} = M_Q. \tag{2.7}$$

In order to calculate I_1, we expand $p_z(z)$ into a Gram-Charlier series

$$p_z(z) = \frac{\mathrm{e}^{-\frac{z^2}{2}}}{\sqrt{2\pi}}\left(1 + \frac{\kappa_3}{3!}\mathrm{He}_3(z) + \frac{\kappa_4}{4!}\mathrm{He}_4(z) + \frac{\kappa_5}{5!}\mathrm{He}_5(z) + \frac{\kappa_6 + 10\kappa_3^2}{6!}\mathrm{He}_6(z) + \ldots\right), \tag{2.8}$$

where κ_i are the cumulants and $\mathrm{He}_k(z)$ is the Hermite polynomial of order k, i.e. $\mathrm{He}_k(z) = (-1)^k \mathrm{e}^{z^2/2} \frac{d^k}{dz^k} \mathrm{e}^{-z^2/2}$.

We will first consider the limit case $N \to \infty$. Let us assume that the cumulants κ_i ($i > 2$) vanish in the case of an infinite-dimensional search space. The density function of z simplifies to $p_z(z) \simeq \mathrm{e}^{-\frac{z^2}{2}}/\sqrt{2\pi}$ enabling us to derive a closed expression for the probability of $\tilde{Q}$

$$P(\tilde{Q}) \simeq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\tilde{Q}}\int_{-\infty}^{\infty} p_\epsilon(\tilde{q}|S_Q z + M_Q) \mathrm{e}^{-\frac{z^2}{2}}\,\mathrm{d}z\,\mathrm{d}\tilde{q}. \tag{2.9}$$

Changing the order of integration and performing the integration over $\tilde{Q}$ (see [6, p.329f]) leads to

$$P(\tilde{Q}) \simeq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \Phi\left(\frac{\tilde{Q} - S_Q z - M_Q}{\sigma_\epsilon}\right) \mathrm{e}^{-\frac{z^2}{2}}\,\mathrm{d}z = \Phi\left(\frac{\tilde{Q} - M_Q}{\sigma_\epsilon\sqrt{1 + (\frac{S_Q}{\sigma_\epsilon})^2}}\right). \tag{2.10}$$

Inserting $P(\tilde{Q})$ into I_1, it can be expressed as

$$I_1 \simeq \frac{\lambda S_Q}{2\pi\sigma_\epsilon} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z \mathrm{e}^{-\frac{z^2}{2}} \mathrm{e}^{-\frac{1}{2}(\frac{\tilde{Q} - S_Q z - M_Q}{\sigma_\epsilon})^2} \Phi\left(\frac{\tilde{Q} - M_Q}{\sigma_\epsilon\sqrt{1 + (\frac{S_Q}{\sigma_\epsilon})^2}}\right)^{\lambda-1} \mathrm{d}\tilde{Q}\,\mathrm{d}z. \tag{2.11}$$

We now change the order of integration and transform $\tilde{Q}$ into $u = (\tilde{Q} - M_Q)/(\sigma_\epsilon\sqrt{1+(\frac{S_Q}{\sigma_\epsilon})^2})$. It follows that $\mathrm{d}\tilde{Q} = \sigma_\epsilon\sqrt{1+(\frac{S_Q}{\sigma_\epsilon})^2}\,\mathrm{d}u$. Inserting u into (2.11), I_1 is given as

$$I_1 \simeq \frac{\lambda S_Q}{\sqrt{2\pi}}\sqrt{1+(\frac{S_Q}{\sigma_\epsilon})^2} * \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\frac{z\mathrm{e}^{-\frac{z^2}{2}}}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(\sqrt{1+(\frac{S_Q}{\sigma_\epsilon})^2}u-\frac{S_Q}{\sigma_\epsilon}z)^2}\,\mathrm{d}z\right)\Phi^{\lambda-1}(u)\,\mathrm{d}u. \tag{2.12}$$

The solution of the inner integral can be found in [6, p.330]. Recalling the definition of the progress coefficient [6, p.72] $c_{1,\lambda} := \frac{\lambda}{\sqrt{2\pi}}\int_{-\infty}^{\infty} u\mathrm{e}^{-\frac{u^2}{2}}\Phi^{\lambda-1}(u)\,\mathrm{d}u$, we finally obtain

$$\begin{aligned}\overline{\Delta Q}_{1,\lambda} &\simeq \frac{\lambda S_Q}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{\sqrt{1+(\frac{S_Q}{\sigma_\epsilon})^2}}{\left(1+(\frac{S_Q}{\sigma_\epsilon})^2\right)}\left(\frac{S_Q}{\sigma_\epsilon}\right)u\mathrm{e}^{-\frac{u^2}{2}}\Phi^{\lambda-1}(u)\,\mathrm{d}u + M_Q \\ &= \frac{S_Q^2}{\sqrt{S_Q^2+\sigma_\epsilon^2}}c_{1,\lambda} + M_Q.\end{aligned} \tag{2.13}$$

Equation (2.13) was developed for the case of an infinite-dimensional search space assuming that the density function of the standardized variable z is given as the standard normal density function. It will be used as an approximate formula in finite-dimensional spaces presuming that the approximation for $p_z(z)$ is also valid there and the error that results from extending the upper integration limit of (2.5) when using the approximation is small. The predictive quality of (2.13) will be assessed by experiments in Section 3.1.

2.2 Steady-State Behavior

The quality gain (2.13) can be used to derive conditions that ensure progress towards the optimum or characterize the steady state. Evolution strategies with (μ,λ)-selection that are disturbed by noise of constant variance cannot approach the optimum arbitrarily close. Instead, they finally end in a steady state that deviates from the actual optimum. Once the steady state is reached, the average quality gain becomes zero. As long as the ES still progresses towards the optimum, the quality gain is strictly positive. Therefore, the condition $\overline{\Delta Q}_{1,\lambda} \geq 0$ is also called the sufficient evolution criterion.

Let us assume in the following that $M_Q \leq 0$ which is a typical characteristic of many test functions. Considering the sufficient evolution criterion, we see that $\overline{\Delta Q}_{1,\lambda} \geq 0$ is satisfied for $S_Q^2/\sqrt{S_Q^2+\sigma_\epsilon^2} \geq |M_Q|/c_{1,\lambda}$. Solving the inequality for S_Q^2, one obtains

$$\begin{aligned}(S_Q^\infty)^2 &\geq \frac{|M_Q^\infty|}{c_{1,\lambda}}\left(\frac{|M_Q^\infty|}{2c_{1,\lambda}} + \sqrt{\left(\frac{M_Q^\infty}{2c_{1,\lambda}}\right)^2 + \sigma_\epsilon^2}\right) \\ &= \frac{(M_Q^\infty)^2}{2c_{1,\lambda}^2}\left(1+\sqrt{1+\left(\frac{2c_{1,\lambda}\sigma_\epsilon}{M_Q^\infty}\right)^2}\right).\end{aligned} \tag{2.14}$$

Since the terms below the root-sign are all positive, we can write

$$(S_Q^\infty)^2 \geq \frac{(M_Q^\infty)^2}{2c_{1,\lambda}^2}\left(1+\sqrt{1+\left(\frac{2\sigma_\epsilon c_{1,\lambda}}{M_Q^\infty}\right)^2}\right) > \frac{(M_Q^\infty)^2}{2c_{1,\lambda}^2}\sqrt{1+\left(\frac{2\sigma_\epsilon c_{1,\lambda}}{M_Q^\infty}\right)^2}$$

$$> \frac{(M_Q^\infty)^2}{2c_{1,\lambda}^2}\sqrt{\left(\frac{2\sigma_\epsilon c_{1,\lambda}}{M_Q^\infty}\right)^2} = \frac{|M_Q^\infty|}{c_{1,\lambda}}\sigma_\epsilon \tag{2.15}$$

and thus arrive at a lower bound for the mutation strength (since M_Q and S_Q depend on the mutation strength) constituting a necessary evolution criterion.

On the other hand, the case "=" in (2.15) leads to the steady state condition $\overline{\Delta Q}_{1,\lambda} = 0$ provided that $S_Q^2 \ll \sigma_\epsilon^2$. Therefore, the noise strength has to be sufficiently large or the variance S_Q^2 sufficiently small. Since S_Q^2 generally depends strongly on the mutation strength σ, we require $\sigma \ll \sigma_\epsilon$ – as it is often the case at the steady state. The steady state condition can therefore be formulated as

$$(S_Q^\infty)^2 \simeq \frac{|M_Q^\infty|}{c_{1,\lambda}}\sigma_\epsilon. \tag{2.16}$$

This equation can be used to derive the final fitness error $\mathrm{E}[\Delta F]$ which is defined as the deviation of the expected value of the fitness function from the real optimal value. If $\mathbf{y}_p$ is a parental state chosen in the steady state regime, we set $\Delta F := \hat{F} - F(\mathbf{y}_p)$, where $\hat{F} = \max F$. Thus, the expected value of the deviation is given as $\mathrm{E}[\Delta F] = \hat{F} - \mathrm{E}[F(\mathbf{y}_p)]$.

In the following, we will derive the final fitness error for quadratic test functions $F_1(\mathbf{y}) = \mathbf{b}^\mathrm{T}\mathbf{y} - \mathbf{y}^\mathrm{T}\mathbf{Q}\mathbf{y}$, a class of biquadratic fitness functions $F_3(\mathbf{y}) = -\sum_{i=1}^N c_i y_i^4$, $F_5(\mathbf{y}) = c - \sum_{i=1}^N |y_i|$, and the function OneMax $F_4(\mathbf{y}) = \sum_{i=1}^N y_i$.

Quadratic fitness functions The local quality is given by $Q(\mathbf{x}) = F(\mathbf{y}+\mathbf{x}) - F(\mathbf{y}) = \mathbf{b}^\mathrm{T}\mathbf{x} - 2\mathbf{y}^\mathrm{T}\mathbf{Q}\mathbf{x} - \mathbf{x}^\mathrm{T}\mathbf{Q}\mathbf{x}$ and can be simplified to $Q(\mathbf{x}) = (\mathbf{b} - \mathbf{2Qy})^\mathrm{T}\mathbf{x} - \mathbf{x}^\mathrm{T}\mathbf{Q}\mathbf{x}$. Since $\mathbf{b} = 2\mathbf{Q}\hat{\mathbf{y}}$, with $\hat{\mathbf{y}} = \arg\max F(\mathbf{y})$, $Q(\mathbf{x})$ can be written as

$$Q(\mathbf{x}) = (2\mathbf{Q}(\hat{\mathbf{y}} - \mathbf{y}))^\mathrm{T}\mathbf{x} - \mathbf{x}^\mathrm{T}\mathbf{Q}\mathbf{x}. \tag{2.17}$$

The expected value and variance of Q have already been obtained in [6, p.122f] as $M_Q = -\sigma^2\mathrm{Tr}[\mathbf{Q}]$ and $S_Q^2 = 4\sigma^2||\mathbf{Q}(\hat{\mathbf{y}} - \mathbf{y})||^2 + 2\sigma^4\mathrm{Tr}[\mathbf{Q}^2]$. Inserting these expressions into (2.16), we obtain

$$||\mathbf{Q}(\hat{\mathbf{y}} - \mathbf{y})||^2 \simeq \frac{\mathrm{Tr}[\mathbf{Q}]}{4c_{1,\lambda}}\sigma_\epsilon - \frac{\sigma^2\mathrm{Tr}[\mathbf{Q}^2]}{2}. \tag{2.18}$$

Assuming smallness of the mutation strength σ at the steady state, i.e. $\sigma^2 \to 0$, Equation (2.18) can be simplified to

$$||\mathbf{Q}(\hat{\mathbf{y}} - \mathbf{y})||^2 \simeq \frac{\mathrm{Tr}[\mathbf{Q}]}{4c_{1,\lambda}}\sigma_\epsilon. \tag{2.19}$$

This agrees with the result found in [7]. Following the same assumptions used there, we can give an approximation for the expected final fitness error $\mathrm{E}[\Delta F]$ without repeating the deviations presented in [7]

$$\mathrm{E}[\Delta F] \approx \frac{\sigma_\epsilon N}{4c_{1,\lambda}}. \tag{2.20}$$

The validity of this equation was already tested in [7], yielding good results. One of the most important assumptions made in that paper was the so-called equipartition assumption stating that the weighted contribution of all components to the final fitness error can be supposed to be the same.

Biquadratic fitness functions We will make use of a similar assumption in order to calculate the fitness error in the case of the test function (1.3). We obtain for $Q(\mathbf{x}){=}F(\mathbf{y}+\mathbf{x})-F(\mathbf{y})$

$$Q(\mathbf{x}) = \sum_{i=1}^{N}(a_i - 4c_i y_i^3)x_i - 6c_i y_i^2 x_i^2 - 4c_i y_i x_i^3 - c_i x_i^4. \tag{2.21}$$

Considering the special case $a_i = 0$, we have

$$Q(\mathbf{x}) = \sum_{i=1}^{N}(-4c_i)y_i^3 x_i - 6c_i y_i^2 x_i^2 - 4c_i y_i x_i^3 - c_i x_i^4. \tag{2.22}$$

Since the optimal point $\hat{y}_i = \sqrt[3]{\frac{a_i}{4c_i}}$ also becomes zero, $\mathit{\Delta} F = \hat{F} - F(\mathbf{y})$ can be written as $\mathit{\Delta} F = \sum_{i=1}^{N} c_i y_i^4$ and the final fitness error $\mathrm{E}[\mathit{\Delta} F]$ is given by

$$\mathrm{E}[\mathit{\Delta} F] = \sum_{i=1}^{N} c_i \mathrm{E}[y_i^4]. \tag{2.23}$$

To continue, we need the expected value M_Q and the variance S_Q^2 of (2.21). To simplify the calculations, we set $Q_i(x_i) := (a_i - 4c_i y_i^3)x_i - 6c_i y_i^2 x_i^2 - 4c_i y_i x_i^3 - c_i x_i^4$. The expected value and the variance can therefore be written as

$$M_Q = \sum_{i=1}^{N} \mathrm{E}[Q_i(x_i)] = -3\sigma^2 \sum_{i=1}^{N} c_i(2y_i^2 + \sigma^2) \tag{2.24}$$

and

$$\begin{aligned} S_Q^2 &= \sum_{i=1}^{N} \mathrm{Var}[Q_i(x_i)] = \sum_{i=1}^{N} \mathrm{E}[Q_i^2] - E[Q_i]^2 \\ &= \sum_{i=1}^{N} \sigma^2(a_i - 4c_i y_i^3)^2 - 24 a_i c_i y_i \sigma^4 + 168 c_i^2 y_i^4 \sigma^4 + 384 c_i^2 y_i^2 \sigma^6 + 96 c_i^2 \sigma^8. \end{aligned} \tag{2.25}$$

As before, we will consider small mutation strengths. Therefore, the higher order terms of σ can be neglected and (2.25) simplifies to $S_Q^2 \simeq \sigma^2 \sum_{i=1}^{N}(a_i - 4c_i y_i^3)^2$. Inserting the expressions for M_Q and S_Q^2 into (2.16), we obtain the following steady state condition (note, $a_i = 0$)

$$\sum_{i=1}^{N}(4c_i y_i^3)^2 = \sum_{i=1}^{N} 16 c_i^2 y_i^6 \simeq \frac{3\sigma_\epsilon}{c_{1,\lambda}} \sum_{i=1}^{N} c_i(2y_i^2 + \sigma^2). \tag{2.26}$$

Considering the limit case $\sigma^2 \to 0$, the influence of σ^2 can be neglected in comparison to $2y_i^2$ and the equation above can be simplified. In order to eliminate the factors c_i in $\mathit{\Delta} F = \sum_{i=1}^{N} c_i y_i^4$, we introduce a new variable $z_i = \sqrt[4]{c_i} y_i$. Taking the expected value of z_i on both sides of (2.26), we get

$$\sum_{i=1}^{N} 16\sqrt{c_i}\, \overline{z_i^6} \simeq \frac{3\sigma_\epsilon}{c_{1,\lambda}} \sum_{i=1}^{N} 2\sqrt{c_i}\, \overline{z_i^2}. \tag{2.27}$$

Since the steady state is reached, it can be assumed that the single components show a very similar behavior, so that the relation for the sum is also valid for a single term

$$\overline{z_i^6} \approx \frac{3\sigma_\epsilon}{8c_{1,\lambda}} \overline{z_i^2}. \tag{2.28}$$

Since our approach does not allow for a determination of the z_i density from first principles, we assume the z_i to be $\mathcal{N}(0,\sigma_z^2)$-distributed exhibiting approximately the same variance σ_z^2. Using this rather crude approximation, the required sixth and the second central moments are $\overline{z_i^6} = 15\sigma_z^6$ and $\overline{z_i^2} = \sigma_z^2$, respectively, and we obtain

$$\sigma_z^4 \approx \frac{\sigma_\epsilon}{40 c_{1,\lambda}}. \tag{2.29}$$

Now we can calculate $\mathrm{E}[\varDelta F]$, i.e. (2.23). Taking $y_i = \frac{z_i}{\sqrt[4]{c_i}}$ and (2.29) into account, $\mathrm{E}[\varDelta F]$ simplifies to $\mathrm{E}[\varDelta F] = 3N\sigma_z^4$, so that the final fitness error can be estimated as

$$\mathrm{E}[\varDelta F] \approx \frac{3N\sigma_\epsilon}{40 c_{1,\lambda}}. \tag{2.30}$$

This equation was obtained by assuming that once the steady state is reached, the average behavior of each of the weighted components is approximately the same. This assumption agrees with the postulates in [7]. We derived (2.30) under the condition that σ is sufficiently small so that the higher-order terms of the variance (2.25) can be neglected. Since the mutation strength tends to be small in the steady state, this condition can be supposed to be justified. Finally, we assumed that the normalized variables can be approximated roughly with normally distributed variables. The validity of this approach will be examined in Section 3.2.1.

It is a rather surprising result that the equation does *not* depend on any terms of function (2.21), that is, the fitness error should be roughly the same regardless of the c_i-values chosen. We will see in Section 3.2.1 that this simple equation predicts the final fitness error surprisingly well – even when extended to the $(\mu/\mu_I,\lambda)$-variants by substituting $c_{1,\lambda}$ with $\mu c_{\mu/\mu_I,\lambda}$.

Negative L_1-Norm The function $F_5(\mathbf{y}) = c - \sum_{i=1}^N |y_i|$ has its maximum at $\mathbf{y} = \mathbf{0}$. The fitness error of a vector $\mathbf{y}$ and the final fitness error are therefore given as $\varDelta F := \sum_{i=1}^N |y_i|$ and $\mathrm{E}[\varDelta F] = \sum_{i=1}^N \mathrm{E}[|y_i|]$. In order to derive an equation describing the final deviation, we use the simplified steady state condition (2.16), $|M_Q^\infty| \frac{\sigma_\epsilon}{c_{1,\lambda}} \simeq {S_Q^\infty}^2$.

The parameter M_Q and S_Q can be given after a short calculation. The local quality is defined as

$$Q(\mathbf{x}) := \sum_{i=1}^N |y_i| - |y_i + x_i|. \tag{2.31}$$

We assume again the x_i to be normally distributed with variance σ^2. Therefore, M_Q is

$$\begin{aligned} M_Q &= \sum_{i=1}^N |y_i| - \int_{-\infty}^{\infty} |y_i + x_i| \phi_{0,\sigma^2}(x_i)\,\mathrm{d}x_i \\ &= \sum_{i=1}^N |y_i| + \int_{-\infty}^{-y_i} (y_i + x_i)\phi_{0,\sigma^2}(x_i)\,\mathrm{d}x_i \\ &\qquad - \int_{-y_i}^{\infty} (y_i + x_i)\phi_{0,\sigma^2}(x_i)\,\mathrm{d}x_i \\ &= \sum_{i=1}^N |y_i| - y_i(2\varPhi_{0,\sigma^2}(y_i) - 1) - 2\sigma^2\phi_{0,\sigma^2}(y_i), \end{aligned} \tag{2.32}$$

where $\phi_{0,\sigma^2}(x_i) := \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{x_i^2}{2\sigma^2}}$ and Φ_{0,σ^2} is the corresponding distribution function.

Defining $Q(\mathbf{x}) := \sum_{i=1}^N q_i = \sum_{i=1}^N |y_i| - |y_i + x_i|$, the variance S_Q^2 is given as $S_Q^2 = \sum_{i=1}^N \mathrm{Var}[q_i] = \sum_{i=1}^N \mathrm{E}[q_i^2] - \mathrm{E}[q_i]^2$. The expected values are obtained as

$$\mathrm{E}[q_i^2] = 2y_i^2 + \sigma^2 - 2|y_i|y_i(2\Phi_{0,\sigma^2}(y_i) - 1) - 4\sigma^2|y_i|\phi_{0,\sigma^2}(y_i) \tag{2.33}$$

and

$$\begin{aligned}\mathrm{E}[q_i]^2 = 2y_i^2 + 4y_i^2\Phi_{0,\sigma^2}^2(y_i) - 4y_i^2\Phi_{0,\sigma^2}(y_i) - 4\sigma^2\phi_{0,\sigma^2}(y_i)\left[y_i(1 - 2\Phi_{0,\sigma^2}(y_i)) - \sigma^2\phi_{0,\sigma^2}(y_i)\right] \\ +2|y_i|y_i[1 - 2\Phi_{0,\sigma^2}(y_i)] - 4\sigma^2|y_i|\phi_{0,\sigma^2}(y_i).\end{aligned} \tag{2.34}$$

Therefore, the variance is given by

$$\begin{aligned}S_Q^2 = N\sigma^2 + 4\sum_{i=1}^N y_i^2\Phi_{0,\sigma^2}(y_i)[1 - \Phi_{0,\sigma^2}(y_i)] \\ -4\sigma^2\sum_{i=1}^N \phi_{0,\sigma^2}(y_i)y_i[2\Phi_{0,\sigma^2}(y_i) - 1] - 4\sigma^4\sum_{i=1}^N \phi_{0,\sigma^2}^2(y_i).\end{aligned} \tag{2.35}$$

First, we note that $S_Q^2 \leq N\sigma^2 - 4\sigma^4\sum_{i=1}^N \phi_{0,\sigma^2}^2(y_i)$ and that $|M_Q| = \sum_{i=1}^N y_i(2\Phi_{0,\sigma^2}(y_i) - 1) + 2\sigma^2\phi_{0,\sigma^2}(y_i) - |y_i|$. A proof of both equations can be found in Appendix 5.1. Considering (2.16), i.e. $|M_Q^\infty|\sigma_\epsilon/c_{1,\lambda} \simeq S_Q^{\infty 2}$, we thus obtain

$$N\sigma^2 - \frac{2\sigma^2}{\pi}\sum_{i=1}^N \mathrm{e}^{-\frac{y_i^2}{\sigma^2}} \geq (\sum_{i=1}^N y_i(2\Phi_{0,\sigma^2}(y_i) - 1) + 2\frac{\sigma}{\sqrt{2\pi}}\mathrm{e}^{-\frac{y_i^2}{2\sigma^2}} - |y_i|)\frac{\sigma_\epsilon}{c_{1,\lambda}}. \tag{2.36}$$

Setting $y_i = \sigma z_i$, the inequalities become

$$\begin{aligned}N\sigma^2 - \frac{2\sigma^2}{\pi}\sum_{i=1}^N \mathrm{e}^{-z_i^2} &\geq (\sigma\sum_{i=1}^N z_i(2\Phi(z_i) - 1) + 2\frac{\sigma}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z_i^2}{2}} - \sigma|z_i|)\frac{\sigma_\epsilon}{c_{1,\lambda}} \\ \Rightarrow N\sigma - \frac{2\sigma}{\pi}\sum_{i=1}^N \mathrm{e}^{-z_i^2} &\geq (\sum_{i=1}^N z_i(2\Phi(z_i) - 1) + \frac{2}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z_i^2}{2}} - |z_i|)\frac{\sigma_\epsilon}{c_{1,\lambda}}.\end{aligned} \tag{2.37}$$

Applying the equipartition assumption to $F_5(\mathbf{y}) = c - \sum_{i=1}^N |y_i|$ (1.5), it becomes clear that each component y_i in (1.5) should obey the same behavior (on average in the steady state regime). Therefore, the same holds for $z_i = y_i/\sigma$ and we can assume $\mathrm{E}[h(z_i)] = \mathrm{E}[h(z_k)]$, $k \neq i$, $h(x)$ one of the functions in (2.37), thus introducing again an equipartition assumption. Furthermore, we approximate z_i by a $\mathcal{N}(0, \sigma_z^2)$ distributed variable. Thus, the expected values of the terms in (2.37) become

$$\begin{aligned}\mathrm{E}[|z_i|] &= \sqrt{\frac{2}{\pi}}\sigma_z, \\ \mathrm{E}[z_i(2\Phi(z_i) - 1)] &= \sqrt{\frac{2}{\pi}}\frac{\sigma_z^2}{\sqrt{\sigma_z^2 + 1}}, \\ \frac{2}{\sqrt{2\pi}}\mathrm{E}[\mathrm{e}^{-\frac{z_i^2}{2}}] &= \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{\sigma_z^2 + 1}}, \text{ and} \\ \frac{2\sigma}{\pi}\mathrm{E}[\mathrm{e}^{-z_i^2}] &= \frac{2}{\pi\sqrt{2\sigma_z^2 + 1}}\sigma.\end{aligned} \tag{2.38}$$

The expected value of (2.37) is therefore

$$N\sigma(1-\frac{2}{\pi\sqrt{2\sigma_z^2+1}}) \geq \frac{\sigma_\epsilon}{c_{1,\lambda}} N\left(\sqrt{\frac{2}{\pi}}\sqrt{\sigma_z^2+1}-\sigma_z\sqrt{\frac{2}{\pi}}\right) \tag{2.39}$$

leading to a relationship between the mutation strength in the steady state and the variance of z_i

$$\sigma \geq \frac{\sigma_\epsilon}{c_{1,\lambda}}\sqrt{\frac{2}{\pi}}\left(\frac{\sqrt{\sigma_z^2+1}-\sigma_z}{(1-\frac{2}{\pi\sqrt{2\sigma_z^2+1}})}\right). \tag{2.40}$$

Note, that the left hand side of (2.40) is a monotonically decreasing function of σ. Once the steady state is reached, the mutation strength is generally small. Therefore, (2.40) can only be expected to be valid, if σ_z is relatively high, especially $\sigma_z \geq 1$. Thus, we assume that the standard deviation of the y_i is higher than the mutation strength.

We can use (2.40) to derive a lower bound for the final fitness error of F_5, i.e. $\mathrm{E}[\Delta F] = \sum_{i=1}^N \mathrm{E}[|y_i|]$. Applying the assumptions of the distribution of y_i, the final fitness error is given by $\mathrm{E}[\Delta F] = \sigma\sum_{i=1}^N \mathrm{E}[|z_i|] = N\sqrt{2/\pi}\sigma\sigma_z$. Considering the inequality (2.40) for σ, we get

$$\mathrm{E}[\Delta F] = N\sqrt{\frac{2}{\pi}}\sigma_z\sigma \geq N\sqrt{\frac{2}{\pi}}\sigma_z\frac{\sigma_\epsilon}{c_{1,\lambda}}\sqrt{\frac{2}{\pi}}\left(\frac{\sqrt{\sigma_z^2+1}-\sigma_z}{(1-\frac{2}{\pi\sqrt{2\sigma_z^2+1}})}\right). \tag{2.41}$$

In order to continue, we need to find a lower bound for $h(x) := (\sqrt{x^4+x^2}-x^2)/(1-\frac{2}{\pi\sqrt{2x^2+1}})$, $x \geq 1$.

Since we do not know an upper bound for σ_z, we assume it to be unbounded above. As the denominator $1-2/(\pi\sqrt{2x^2+1})$ approaches 1 for $x \to \infty$ and the numerator is given as

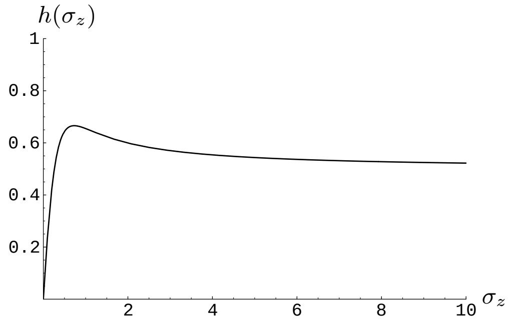
$$\begin{aligned}\sqrt{x^4+x^2}-x^2 &= \frac{(\sqrt{x^4+x^2}-x^2)(\sqrt{x^4+x^2}+x^2)}{\sqrt{x^4+x^2}+x^2} = \frac{x^2}{\sqrt{x^4+x^2}+x^2}\\ &= \frac{1}{1+\sqrt{1+\frac{1}{x^2}}},\end{aligned} \tag{2.42}$$

$h(x) = \frac{\sqrt{x^4+x^2}-x^2}{1-\frac{2}{\pi\sqrt{2x^2+1}}}$ goes to 0.5. This is a lower bound for $h(x)$, $x \geq 1$, as we will show in Appendix 5.1. Therefore, we can give a lower bound for the final fitness error as

$$\mathrm{E}[\Delta F] \approx N\sqrt{\frac{2}{\pi}}\sigma\sigma_z \geq N\sqrt{\frac{2}{\pi}}\left(\sqrt{\frac{2}{\pi}}\frac{\sigma_\epsilon}{c_{1,\lambda}}h(\sigma_z)\right) \geq \frac{\sigma_\epsilon}{\pi c_{1,\lambda}}N. \tag{2.43}$$

To derive (2.43), we used again a variant of the equipartition assumption, first introduced in [7]. Furthermore, we had to assume that the standardized variables $z_i = y_i/\sigma$ obey a normal distribution. Under this assumption, we can derive (2.40) which gives a function of the noise strength and the standard deviation of z as a lower bound for the mutation strength. Equation (2.40) can further be used to derive an estimate for the final inequality (2.43). Assuming that $\sigma_z \geq 1$, $h(x)$ is always higher than its limit value which can then be used to derive the lower bound.

Fig. 2.1. The function $h(\sigma_z)$ obtained in the derivation of the lower bound for function F_5 using the normal distribution.



OneMax The local quality of OneMax, i.e. $F_4(\mathbf{y}) = \sum_{i=1}^{N} y_i$, is given by $Q(\mathbf{y}) = F(\mathbf{x}) - F(\mathbf{y})$. For strings of length N, the statistical parameters are $M_Q = -p_m(2F_0 - N)$ and $S_Q = \sqrt{N}\sqrt{p_m(1-p_m)}$ (see [6, p.128f]), where p_m is the mutation rate and $F_0 = \sum_{i=1}^{N} y_i$ the parental fitness value. Inserting these values into (2.16), $(S_Q^{\infty})^2 \simeq |M_Q^{\infty}|\sigma_\epsilon/c_{1,\lambda}$, gives

$$N p_m(1-p_m) \simeq \frac{p_m(2F_0 - N)}{c_{1,\lambda}}\sigma_\epsilon \tag{2.44}$$

leading to the final fitness error

$$\mathrm{E}[\mathit{\Delta F}] = N - \sum_{i=1}^{N}\mathrm{E}[y_i] \simeq \frac{N}{2}\left(1 - \frac{c_{1,\lambda}(1-p_m)}{\sigma_\epsilon}\right). \tag{2.45}$$

Note that this equation is only valid for sufficiently high noise strengths. A more appropriate formula is obtained by inserting the expected value and standard deviation into (2.13), $\overline{\mathit{\Delta Q}}_{1,\lambda} \simeq S_Q^2/\sqrt{S_Q^2 + \sigma_\epsilon^2}\,c_{1,\lambda} + M_Q$, and setting the quality gain equal to zero. After rearranging the terms and taking the expectation of $F_0 = \sum_{i=1}^{N} y_i$, one immediately gets

$$\sum_{i=1}^{N}\mathrm{E}[y_i] = \frac{N}{2}\left(1 + \frac{1-p_m}{\sqrt{Np_m(1-p_m) + \sigma_\epsilon^2}}c_{1,\lambda}\right). \tag{2.46}$$

The final fitness error is then given as

$$\mathrm{E}[\mathit{\Delta F}] = N - \sum_{i=1}^{N}\mathrm{E}[y_i] \simeq \frac{N}{2}\left(1 - \frac{1-p_m}{\sqrt{Np_m(1-p_m) + \sigma_\epsilon^2}}c_{1,\lambda}\right). \tag{2.47}$$

Considering the mutation rate $p_m = 1/N$, usually recommended in literature, and assuming N to be sufficiently large, the equations (2.45) and (2.47) simplify to

$$\mathrm{E}[\mathit{\Delta F}] \approx \frac{N}{2}\left(1 - \frac{c_{1,\lambda}}{\sigma_\epsilon}\right) \tag{2.48}$$

and

$$\mathrm{E}[\mathit{\Delta F}] \approx \frac{N}{2}\left(1 - \frac{c_{1,\lambda}}{\sqrt{1+\sigma_\epsilon^2}}\right). \tag{2.49}$$

Since $\mathrm{E}[\mathit{\Delta F}] \geq 0$, (2.49) is only valid if the inner fraction is smaller than 1, i.e. $c_{1,\lambda} \leq \sqrt{1+\sigma_\epsilon^2}$. The predictive quality of the final fitness error obtained will be assessed in Section 3.2.1.

Chapter 3

Comparison with Experiments

In the following, Equation (2.13) describing the quality gain will be applied to several classes of fitness functions which were introduced in Section 1.2. Afterwards, we will examine the predictive quality of the final fitness error for some biquadratic functions, the negative L_1-norm, and OneMax extending it to $(\mu/\mu_I, \lambda)$-ES for the first two function classes. Since the equation for quadratic functions equals the one developed in [7], where its validity was already shown, it will not be considered further.

3.1 The Quality Gain

3.1.1 Quadratic Test Functions

The local quality for quadratic functions is given as $Q(\mathbf{x}){=}2[\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})]^{\mathrm{T}}\mathbf{x}-\mathbf{x}^{\mathrm{T}}\mathbf{Q}\mathbf{x}$ (cp. Sec. 2.2). As expected value and variance of Q, $M_Q=-\sigma^2\mathrm{Tr}[\mathbf{Q}]$ and $S_Q^2=4\sigma^2||\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})||^2+2\sigma^4\mathrm{Tr}[\mathbf{Q}^2]$ were obtained. Inserting these expressions into (2.13), the quality gain for quadratic functions is given as

$$\begin{aligned}\overline{\Delta Q}_{1,\lambda} &\simeq \frac{S_Q^2}{\sqrt{S_Q^2+\sigma_\epsilon^2}}c_{1,\lambda}+M_Q\\ &= \frac{4\sigma^2||\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})||^2+2\sigma^4\mathrm{Tr}[\mathbf{Q}^2]}{\sqrt{4\sigma^2||\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})||^2+2\sigma^4\mathrm{Tr}[\mathbf{Q}^2]+\sigma_\epsilon^2}}c_{1,\lambda}-\sigma^2\mathrm{Tr}[\mathbf{Q}]\end{aligned} \tag{3.1}$$

or assuming that $2\sigma^4\mathrm{Tr}[\mathbf{Q}^2]$ is small compared with $4\sigma^2||\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})||^2$

$$\overline{\Delta Q}_{1,\lambda} \simeq \frac{4\sigma^2||\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})||^2}{\sqrt{4\sigma^2||\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})||^2+\sigma_\epsilon^2}}c_{1,\lambda}-\sigma^2\mathrm{Tr}[\mathbf{Q}]. \tag{3.2}$$

To compare (3.2) with the result obtained in [6] for the noise-free case, we introduce the following normalizations [6, p.132] $\sigma^*=\frac{\mathrm{Tr}[\mathbf{Q}]}{||\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})||}\sigma$, $\sigma_\epsilon^*=\frac{\mathrm{Tr}[\mathbf{Q}]}{2||\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})||^2}\sigma_\epsilon$, and $\overline{\Delta Q}^*_{1,\lambda}=\frac{\mathrm{Tr}[\mathbf{Q}]}{2||\mathbf{Q}(\hat{\mathbf{y}}-\mathbf{y})||^2}\overline{\Delta Q}_{1,\lambda}$ and get

$$\overline{\Delta Q}^*_{1,\lambda} \simeq c_{1,\lambda}\sigma^*\frac{1}{\sqrt{1+(\frac{\sigma_\epsilon^*}{\sigma^*})^2}}-\frac{\sigma^{*2}}{2}. \tag{3.3}$$

As in the noise-free case with $\overline{\Delta Q}^*_{1,\lambda}\simeq c_{1,\lambda}\sigma^*-\frac{\sigma^{*2}}{2}$, the equation can be decomposed into a gain and a loss term. The noise only influences the gain part of the equation, an observation already made in [6] for the progress rate on the noisy sphere.

The noisy quality gain $\overline{\Delta Q}^*_{1,\lambda}$ has a maximum whose value depends on the noise strength σ^*_ϵ. If the noise is too large, it will reduce the linear gain part such that the loss term will be the dominating factor leading to a monotonically decreasing function.

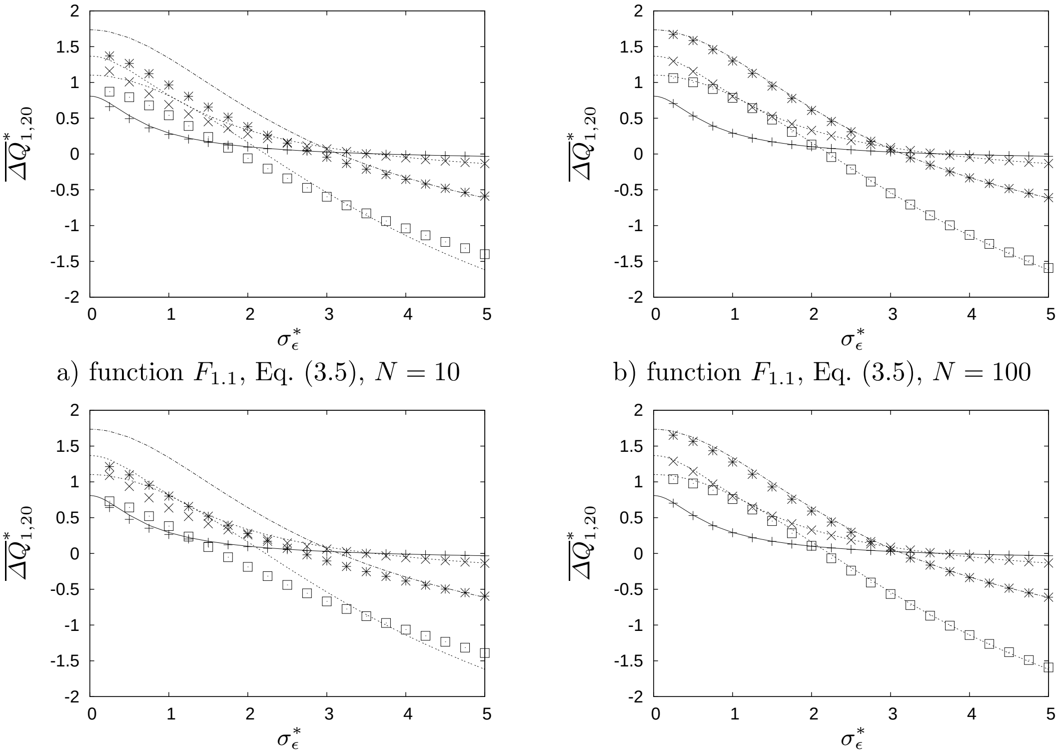
As representatives for the quadratic test functions,

$$F_{1.1}(\mathbf{x}) = -\sum_{i=1}^{N} i x_i^2 \text{ and} \tag{3.4}$$

$$F_{1.2}(\mathbf{x}) = -\sum_{i=1}^{N} i^2 x_i^2 \tag{3.5}$$

were chosen. Figure 3.1 shows the dependence of the normalized quality gain on the noise

Fig. 3.1. Dependence of $\overline{\Delta Q}^*_{1,20}$ on the noise strength σ^*_ϵ for quadratic test functions (Equations (3.5) and (3.5)) for several choices of σ^*. Depicted are from top to bottom the curves for $\sigma^* = 2$, 1, 3, and 0.5.



a) function $F_{1.1}$, Eq. (3.5), $N = 10$

b) function $F_{1.1}$, Eq. (3.5), $N = 100$

c) function $F_{1.2}$, Eq. (3.5), $N = 10$

d) function $F_{1.2}$, Eq. (3.5), $N = 100$

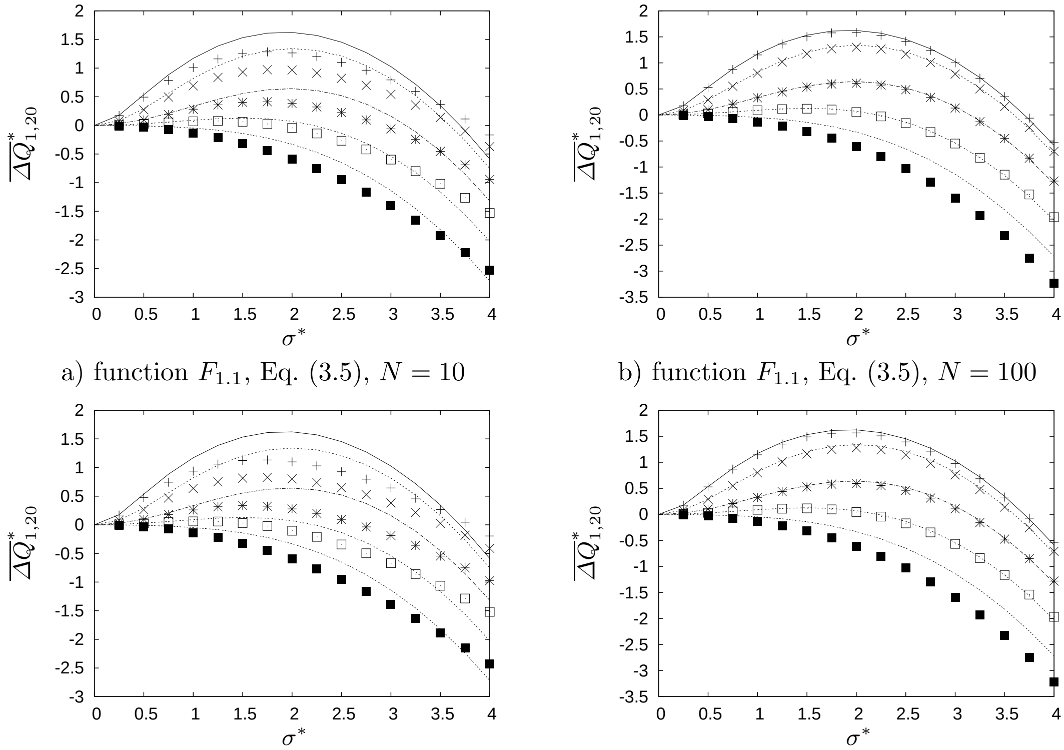
strength σ^*_ϵ in the 10- and 100-dimensional object parameter space. The experimental values were obtained by performing 1,000,000 one-generation experiments per data point using $\mathbf{y} = \mathbf{2}$ as starting vector.

As σ^*_ϵ increases, the quality gain decreases approaching $-\frac{\sigma^{*2}}{2}$. For $N = 100$, the agreement between equation and experiment is very good, while greater deviations can be observed for the lower dimensional case. Here, good results for all choices of σ^*_ϵ can only be obtained for small normalized mutation strengths, e.g. $\sigma^* = 0.5$ – the lowest mutation strength examined – where experimental results and predicted values show an excellent agreement.

Figure 3.2 shows the dependency of $\overline{\Delta Q}^*_{1,20}$ on σ^*, depicting the quality gain (3.3) and the experimentally obtained values for several choices of σ^*_ϵ. For σ^* sufficiently small, i.e.

$\sigma^* < 1$, the agreement between (3.3) and measurement is quite good even in the lower dimensional search space. That range increases decidedly for $N = 100$. The only exception is the curve for $\sigma_\epsilon^* = 5$ where (3.3) leads to an overestimation of the experimental values.

Fig. 3.2. Dependence of $\overline{\Delta Q}^*_{1,20}$ on the mutation strength σ^* for quadratic test functions (Equations (3.5) and (3.5)). Depicted are from top to bottom the curves for the noise strengths 0.5, 1, 2, 3, and 5.



a) function $F_{1.1}$, Eq. (3.5), $N = 10$

b) function $F_{1.1}$, Eq. (3.5), $N = 100$

c) function $F_{1.1}$, Eq. (3.5), $N = 10$

d) function $F_{1.1}$, Eq. (3.5), $N = 100$

3.1.2 Cubic Functions

In this subsection, we will consider the quality gain in the case of cubic functions, i.e. functions of the form $F_2(\mathbf{y}) = \sum_{i=1}^N a_i y_i - c_i y_i^3$, for which the local quality is given as

$$\begin{aligned} Q(\mathbf{x}) &= \sum_{i=1}^{N} a_i y_i + a_i x_i - c_i (y_i + x_i)^3 - a_i y_i + c_i y_i^3 \\ &= \sum_{i=1}^{N} (a_i - 3c_i y_i^2) x_i - 3c_i y_i x_i^2 - c_i x_i^3. \end{aligned} \tag{3.6}$$

To obtain the specific quality gain, we need the expected value and variance of (3.6). The expected value is given by

$$M_Q = \sum_{i=1}^{N} \mathrm{E}[q_i(x_i)] = \sum_{i=1}^{N} \mathrm{E}[(a_i - 3c_i y_i^2) x_i - 3c_i y_i x_i^2 - c_i x_i^3] = -3\sigma^2 \sum_{i=1}^{N} c_i y_i. \tag{3.7}$$

In order to obtain the variance, we consider

$$S_Q^2 = \sum_{i=1}^{N} \mathrm{Var}[q_i] = \sum_{i=1}^{N} \mathrm{E}[q_i^2] - \mathrm{E}[q_i]^2. \tag{3.8}$$

Since $\mathrm{E}[q_i]^2$ equals $\mathrm{E}[q_i]^2 = 9\sigma^4 c_i^2 y_i^2$ and the expected value of q_i^2 is given by $\mathrm{E}[q_i^2] = \mathrm{E}[(a_i - 3c_i y_i^2)^2 x_i^2 - 6c_i y_i (a_i - 3c_i y_i^2) x_i^3 - 2c_i(a_i - 3c_i y_i^2) x_i^4 + 9c_i^2 y_i^2 x_i^4 + 6c_i^2 y_i x_i^5 + c_i^2 x_i^6] = 15c_i^2\sigma^6 + 45c_i^2 y_i^2 \sigma^4 - 6c_i a_i \sigma^4 + (a_i - 3c_i y_i^2)^2\sigma^2$, the variance is obtained as

$$S_Q^2 = \sum_{i=1}^{N} 15c_i^2\sigma^6 + 6c_i(6c_i y_i^2 - a_i)\sigma^4 + (a_i - 3c_i y_i^2)^2\sigma^2. \tag{3.9}$$

By inserting these values into (2.13) we get

$$\overline{\mathit{\Delta} Q}_{1,\lambda} \simeq \frac{15\sigma^6 \sum_{i=1}^N c_i^2 + 6\sigma^4 \sum_{i=1}^N 6c_i^2 y_i^2 - c_i a_i + \sigma^2 \sum_{i=1}^N (a_i - 3c_i y_i^2)^2}{\sqrt{15\sigma^6 \sum_{i=1}^N c_i^2 + 6\sigma^4 \sum_{i=1}^N 6c_i^2 y_i^2 - c_i a_i + \sigma^2 \sum_{i=1}^N (a_i - 3c_i y_i^2)^2 + \sigma_\epsilon^2}} c_{1,\lambda} - 3\sigma^2 \sum_{i=1}^{N} c_i y_i. \tag{3.10}$$

The quality gain consists of a gain term and a quadratic loss term. The gain part of (3.10) is approximately a polynomial of degree three that is influenced by the noise although that influence is rather weak. Eventually, if σ is increased far enough, the quality gain will be a monotonically increasing function of the mutation strength.

For $\sigma \ll 1$, the variance simplifies to $S_Q^2 \simeq 9\sigma^2 \sum_{i=1}^N (a_i - 3c_i y_i^2)^2$, leading to the following normalized quality gain

$$\overline{\mathit{\Delta} Q^*}_{1,\lambda} \simeq \sigma^* \frac{1}{\sqrt{1 + (\frac{\sigma_\epsilon^*}{\sigma^*})^2}} c_{1,\lambda} - \frac{(\sigma^*)^2}{2}, \tag{3.11}$$

where $\sigma^* = \frac{6\sum_{i=1}^N c_i y_i}{\sqrt{\sum_{i=1}^N (a_i - 3c_i y_i^2)^2}}\sigma$, $\sigma_\epsilon^* = \frac{6\sum_{i=1}^N c_i y_i}{\sum_{i=1}^N (a_i - 3c_i y_i^2)^2}\sigma_\epsilon$, and $\overline{\mathit{\Delta} Q^*}_{1,\lambda} = \frac{6\sum_{i=1}^N c_i y_i}{\sum_{i=1}^N (a_i - 3c_i y_i^2)^2}\overline{\mathit{\Delta} Q}_{1,\lambda}$. Equation (3.10) equals the normalized quality gain of quadratic fitness functions (3.3).

The quality gain (3.10) and the normalized quality gain (3.11) on cubic functions were tested using the fitness functions

$$F_{2.1}(\mathbf{y}) = -\sum_{i=1}^{N} i y_i^3 \quad \text{and} \tag{3.12}$$

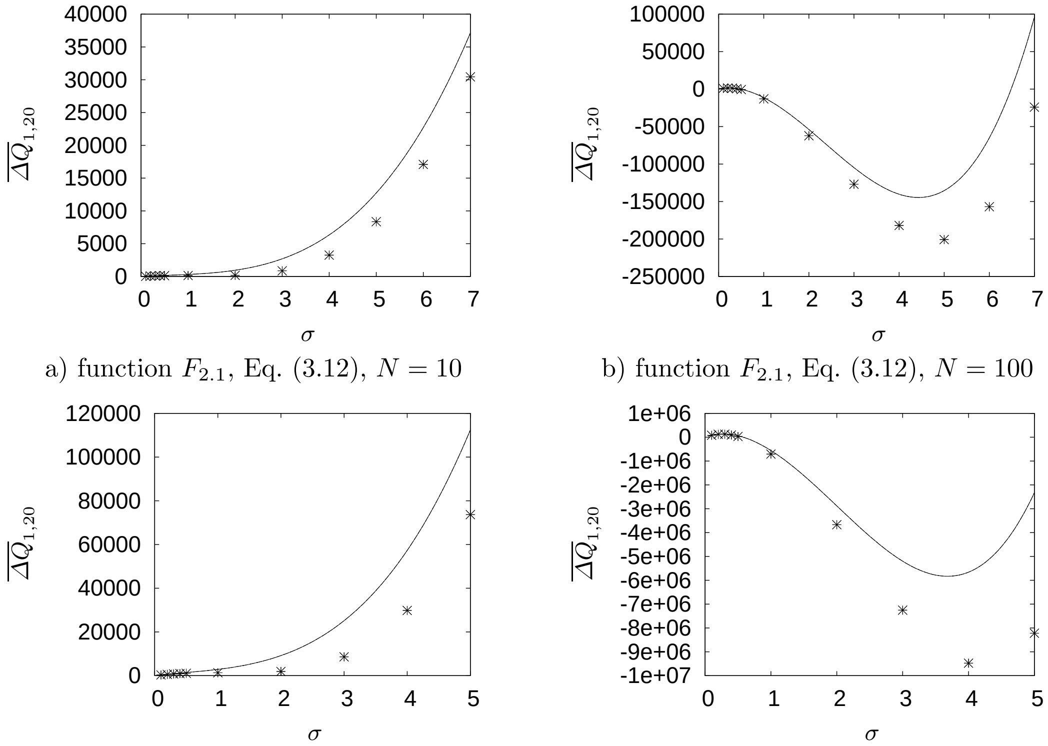
$$F_{2.2}(\mathbf{y}) = -\sum_{i=1}^{N} i^2 y_i^3. \tag{3.13}$$

Figures 3.3 to 3.5 show the dependency of the respective quality gain on the noise strength and mutation strength. The values were obtained by performing 1,000,000 one-generations experiments per data point. The vector $\mathbf{y} = \mathbf{2}$ was chosen as starting vector.

Figure 3.3 depicts the quality gain (3.10) as a function of the mutation strength. For $N = 10$, the quality gain is a monotonically increasing function, for which experimental values and (3.10) start to deviate rather soon, i.e. for mutation strengths close to one. For $N = 100$, the predictive quality is acceptable for a wider range of σ.

Considering the normalized noise strength and Fig. 3.5, (3.11) agrees well with the experimental results as long as σ^* is sufficiently small, i.e. $\sigma^* \leq 1.5$ for $N = 10$. That range can be increased to $\sigma^* \leq 3$ in the case of the 100-dimensional space. As one can see in Fig. 3.3, for $N = 10$ (3.10) is a monotonically increasing function for the fitness functions considered. Therefore, (3.11) can only be expected to agree well with the experiments as long as the gain part outweighs the loss term.

Fig. 3.3. Dependence of $\overline{\Delta Q}_{1,20}$ on the mutation strength σ for cubic functions (Equations (3.12) and (3.13)). Depicted is the curve for the noise strength $\sigma_\epsilon = 1$.



a) function $F_{2.1}$, Eq. (3.12), $N = 10$

b) function $F_{2.1}$, Eq. (3.12), $N = 100$

c) function $F_{2.2}$, Eq. (3.13), $N = 10$

d) function $F_{2.2}$, Eq. (3.13), $N = 100$

3.1.3 Biquadratic Functions

We will now consider functions of the form $F_3(\mathbf{y}) = \sum_{i=1}^{N}(a_i y_i - c_i y_i^4)$ (1.3) with the local quality function

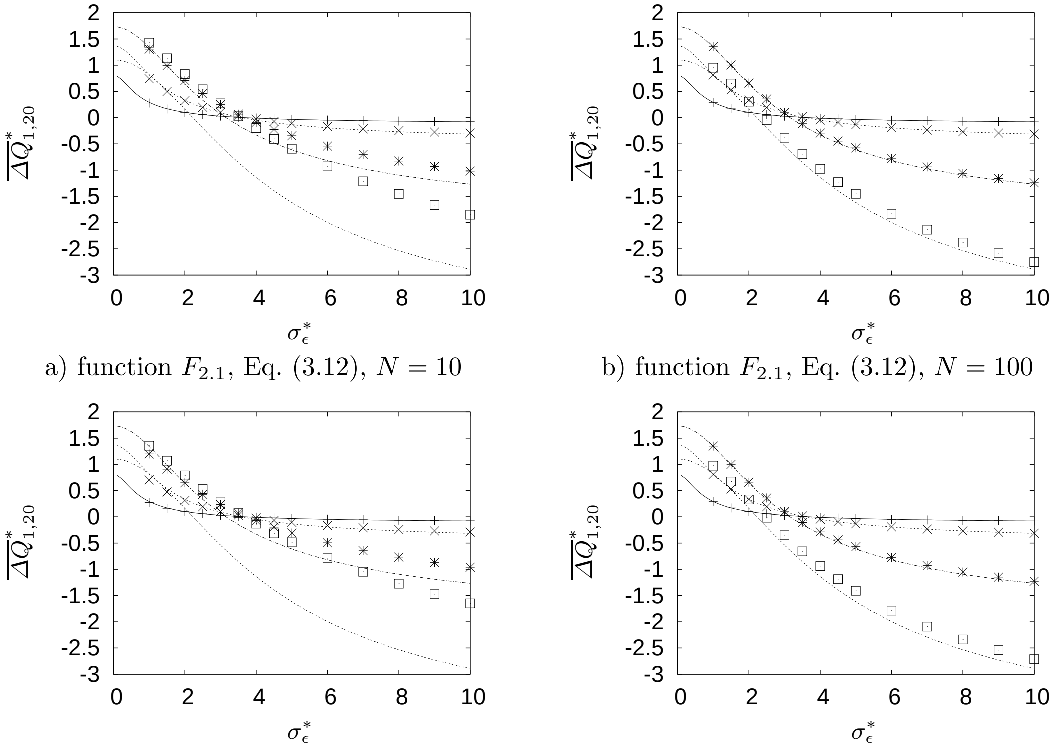
$$Q(\mathbf{x}) = \sum_{i=1}^{N}(a_i - 4c_i y_i^3)x_i - 6c_i y_i^2 x_i^2 - 4c_i y_i x_i^3 - c_i x_i^4. \tag{3.14}$$

In order to determine the quality gain (2.13), $\overline{\Delta Q}_{1,\lambda} \simeq \frac{S_Q^2}{\sqrt{S_Q^2+\sigma_\epsilon^2}}c_{1,\lambda} + M_Q$, we need the expressions for M_Q and S_Q which were already developed in Sec. 2.2 as

$$M_Q = -3\sigma^2 \sum_{i=1}^{N} c_i(2y_i^2 + \sigma^2) \text{ and}$$

$$S_Q^2 = \sum_{i=1}^{N} \sigma^2(a_i - 4c_i y_i^3)^2 - 24a_i c_i y_i \sigma^4 + 168c_i^2 y_i^4 \sigma^4 + 384c_i^2 y_i^2 \sigma^6 + 96c_i^2 \sigma^8.$$

Fig. 3.4. Dependence of $\overline{\Delta Q}^*_{1,20}$ on the noise strength σ^*_ϵ for cubic functions (Equations (3.12) and (3.13)) for several choices of σ^*. Depicted are from top to bottom the curves for $\sigma^* = 2$, $\sigma^* = 1$, $\sigma^* = 3$, and $\sigma^* = 0.5$.



a) function $F_{2.1}$, Eq. (3.12), $N = 10$

b) function $F_{2.1}$, Eq. (3.12), $N = 100$

c) function $F_{2.2}$, Eq. (3.13), $N = 10$

d) function $F_{2.2}$, Eq. (3.13), $N = 100$

Assuming σ to be sufficiently small, S_Q^2 can be simplified to $S_Q^2 \simeq \sigma^2 \sum_{i=1}^N (a_i - 4c_i y_i^3)^2$. The expressions for S_Q and M_Q are then inserted into (2.13) leading to

$$\overline{\Delta Q}_{1,\lambda} \simeq c_{1,\lambda}\sigma^2 \frac{\sum_{i=1}^N (a_i - 4c_i y_i^3)^2}{\sqrt{\sigma^2 \sum_{i=1}^N (a_i - 4c_i y_i^3)^2 + \sigma_\epsilon^2}} - 3\sigma^2 \sum_{i=1}^N c_i(2y_i^2 + \sigma^2). \tag{3.15}$$

Unfortunately, function (1.3) generally does not allow for a normalization of (3.15). An exception to this is the case $\mathbf{y} = \mathbf{0}$. The expressions above for the variance and the expected value simplify then to $S_Q^2 \simeq \sigma^2 \sum_{i=1}^N a_i^2 = \sigma^2 ||\mathbf{a}||^2$ and $M_Q = -3\sigma^4 \sum_{i=1}^N c_i$. Setting $c = 3\sum_{i=1}^N c_i$ and using the normalizations $\sigma^* = \sqrt[3]{c/||\mathbf{a}||}\sigma$, $\sigma^*_\epsilon = \sqrt[3]{c/||\mathbf{a}||^4}\sigma_\epsilon$, and $\overline{\Delta Q}^*_{1,\lambda} = \sqrt[3]{c/||\mathbf{a}||^4}\,\overline{\Delta Q}_{1,\lambda}$ (see [6, p.134]), we obtain

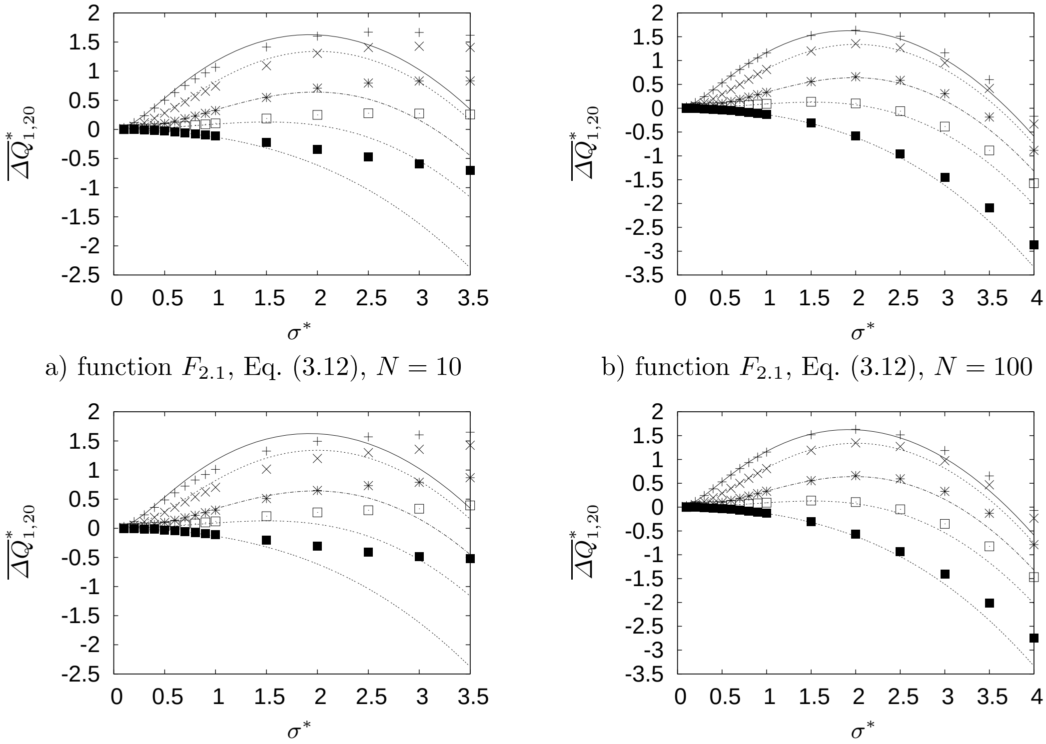
$$\overline{\Delta Q}^*_{1,\lambda} \simeq c_{1,\lambda}\sigma^* \frac{1}{\sqrt{1 + (\frac{\sigma^*_\epsilon}{\sigma^*})^2}} - \sigma^{*4}. \tag{3.16}$$

Again the equation can be decomposed into a linear gain part being influenced by the noise and a loss term which is biquadratic in this case. If σ^* is kept constant, $\overline{\Delta Q}^*_{1,\lambda}$ approaches $-\sigma^{*4}$ if the noise strength σ^*_ϵ is increased.

The predictive quality of (3.16) was examined using the fitness functions

$$F_{3.1}(\mathbf{y}) = \sum_{i=1}^N y_i - y_i^4, \tag{3.17}$$

Fig. 3.5. Dependence of $\overline{\Delta Q}^*_{1,20}$ on the mutation strength σ^* for cubic functions (Equations (3.12) and (3.13)). Depicted are from top to bottom the curves for the noise strengths $\sigma^*_\epsilon = 0.5$, $\sigma^*_\epsilon = 1$, $\sigma^*_\epsilon = 2$, $\sigma^*_\epsilon = 3$, and $\sigma^*_\epsilon = 5$.



a) function $F_{2.1}$, Eq. (3.12), $N = 10$

b) function $F_{2.1}$, Eq. (3.12), $N = 100$

c) function $F_{2.2}$, Eq. (3.13), $N = 10$

d) function $F_{2.2}$, Eq. (3.13), $N = 100$

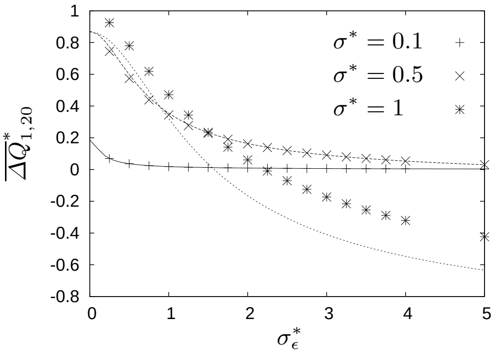
$$F_{3.2}(\mathbf{y}) = \sum_{i=1}^{N} y_i - i y_i^4, \quad \text{and} \tag{3.18}$$

$$F_{3.3}(\mathbf{y}) = \sum_{i=1}^{N} y_i - i^2 y_i^4. \tag{3.19}$$

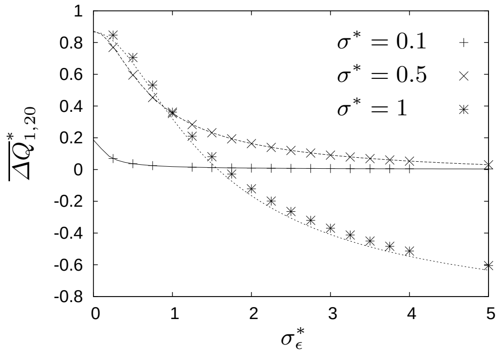
Figure 3.6 shows the dependency of $\overline{\Delta Q^*}_{1,\lambda}$ on σ^*_ϵ for several choices of σ^*. As in the cubic case, the experimental values were obtained by averaging over 1,000,000 one-generation trials using $\mathbf{y} = \mathbf{0}$ as parental state. The behavior we observe is similar to that of the quadratic functions. An increase of the noise strength decreases the linear gain part of (3.10) so that the maximum becomes less pronounced and approaches zero for $\sigma^*_\epsilon \to \infty$. Even for $N = 10$, a good agreement exists between the values obtained by (3.10) and those of the experiments as long as σ^* is small. As it approaches one, the predictive quality deteriorates.

This can also be seen in Fig. 3.7 showing the dependency of $\overline{\Delta Q}^*_{1,20}$ on the mutation strength σ^* for the noise strengths 0.5, 1, 2, 3, and 5. For $N = 10$, (3.16) predicts the experimental values up to a mutation strength of one reasonably well. This can be extended to 1.25 in the higher-dimensional parameter space.

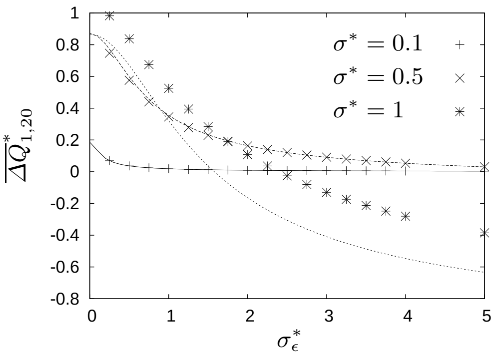
Fig. 3.6. Dependence of $\overline{\Delta Q}^*_{1,20}$ on the noise strength σ^*_ϵ for biquadratic functions. Depicted are from top to bottom the results for $\sigma^* = 1$, $\sigma^* = 0.5$, and $\sigma^* = 0.1$.



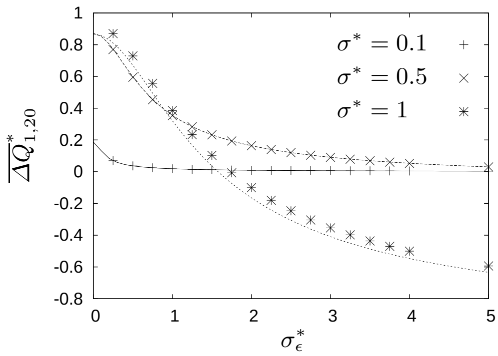
a) function $F_{3.1}$, Eq. (3.17), $N = 10$



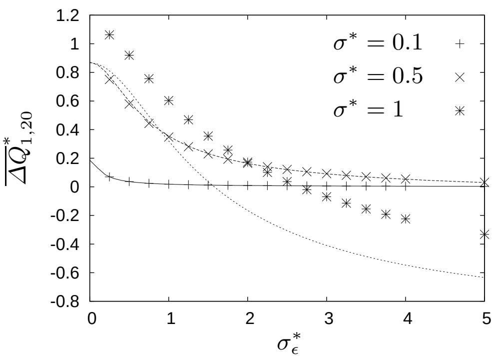
b) function $F_{3.1}$, Eq. (3.17), $N = 100$



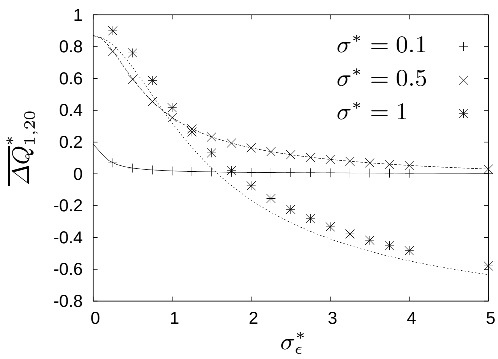
c) function $F_{3.2}$, Eq. (3.18), $N = 10$



d) function $F_{3.2}$, Eq. (3.18), $N = 100$



e) function $F_{3.3}$, Eq. (3.19), $N = 10$



f) function $F_{3.3}$, Eq. (3.19), $N = 100$

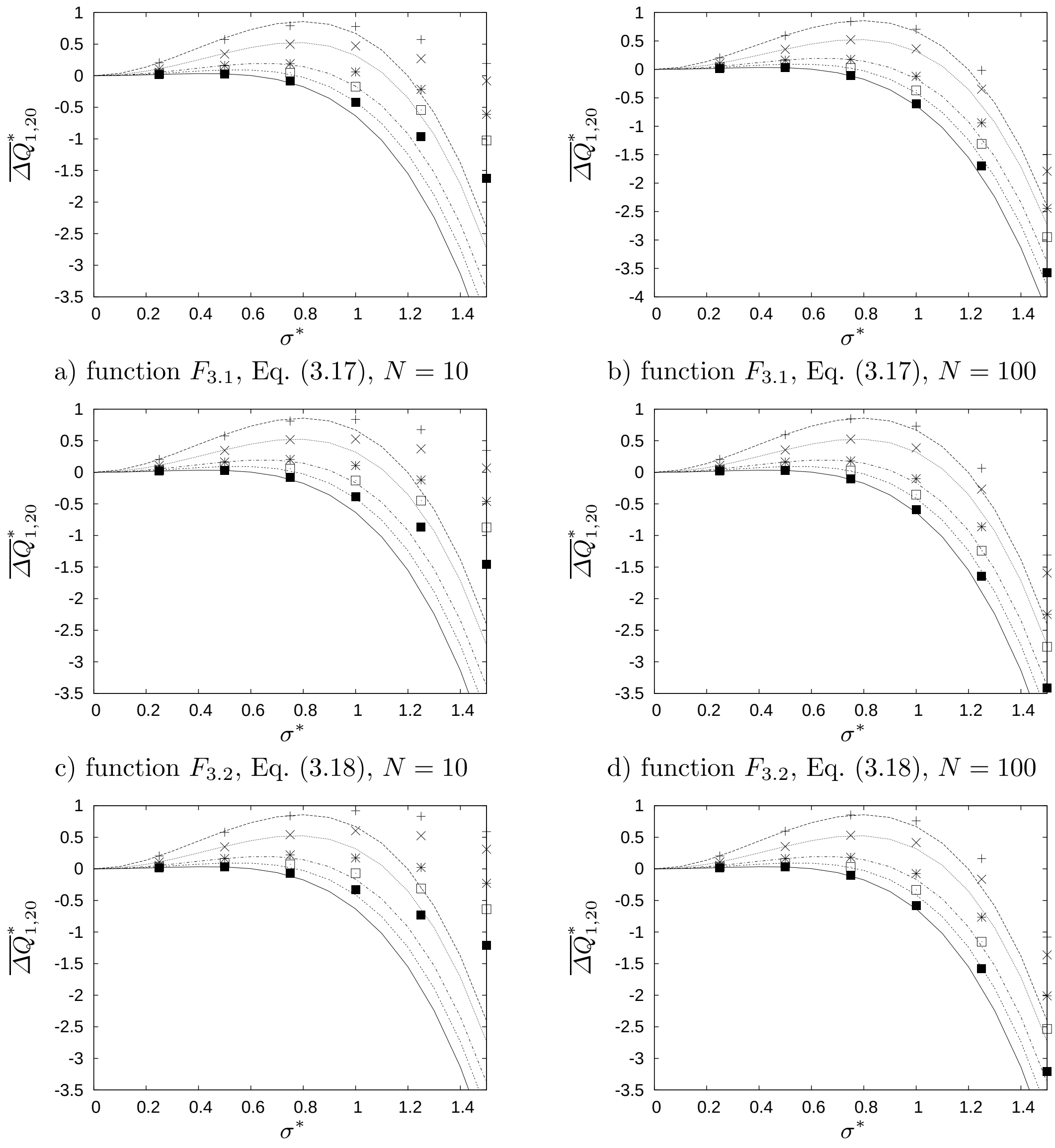
3.1.4 Negative L_1-Norm

We will now consider functions of the general form $F_5(\mathbf{y}) = c - \sum_{i=1}^N |y_i|$, for which the statistical parameters are given as (see (2.32) and (2.35))

$$M_Q = \sum_{i=1}^N |y_i| - y_i(2\Phi_{0,\sigma^2}(y_i) - 1) - 2\sigma^2\phi_{0,\sigma^2}(y_i) \text{ and}$$

$$S_Q^2 = N\sigma^2 + 4\sum_{i=1}^N y_i^2\Phi_{0,\sigma^2}(y_i)[1 - \Phi_{0,\sigma^2}(y_i)]$$
$$-4\sigma^2\sum_{i=1}^N \phi_{0,\sigma^2}(y_i)y_i[2\Phi_{0,\sigma^2}(y_i) - 1] - 4\sigma^4\sum_{i=1}^N \phi^2_{0,\sigma^2}(y_i),$$

Fig. 3.7. Dependence of $\overline{\Delta Q^*}_{1,20}$ on the mutation strength σ^* for biquadratic functions. Depicted are from bottom to top the results for $\sigma^*_\epsilon = 5$,$\sigma^*_\epsilon = 3$,$\sigma^*_\epsilon = 2$, $\sigma^*_\epsilon = 1$, and $\sigma^*_\epsilon = 0.5$.



a) function $F_{3.1}$, Eq. (3.17), $N = 10$ b) function $F_{3.1}$, Eq. (3.17), $N = 100$

c) function $F_{3.2}$, Eq. (3.18), $N = 10$ d) function $F_{3.2}$, Eq. (3.18), $N = 100$

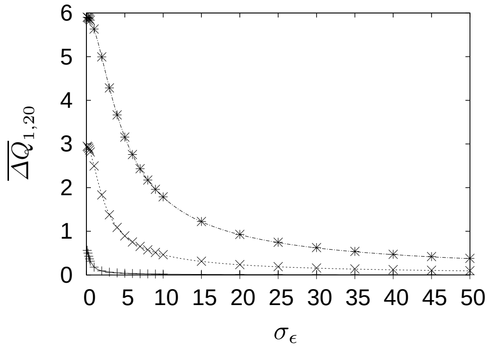
e) function $F_{3.3}$, Eq. (3.19), $N = 10$ f) function $F_{3.3}$, Eq. (3.19), $N = 100$

where $\phi_{0,\sigma^2}(y_i) := \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{y_i^2}{2\sigma^2}}$ and Φ_{0,σ^2} is the normal distribution function. By inserting M_Q and S_Q^2 into (2.13), $\overline{\Delta Q}_{1,\lambda} \simeq S_Q^2/\sqrt{S_Q^2 + \sigma_\epsilon^2}c_{1,\lambda} + M_Q$, we get the quality gain of F_5.

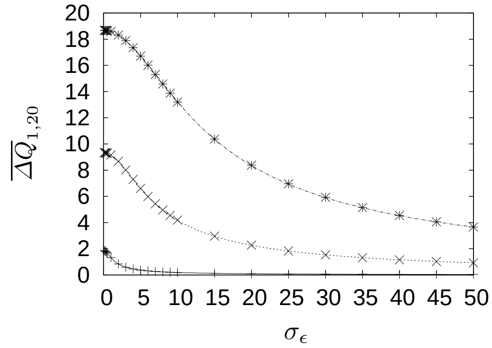
Considering the mean value $M_Q = \sum_{i=1}^N |y_i| - y_i(2\Phi_{0,\sigma^2}(y_i) - 1) - 2\sigma^2\phi_{0,\sigma^2}(y_i)$, we note that $|y_i| - y_i(2\Phi_{0,\sigma^2}(y_i) - 1)$ is given as $2y_i(1 - \Phi_{0,\sigma^2}(y_i))$ if $y_i \geq 0$, whereas it simplifies to $-2y_i\Phi_{0,\sigma^2}(y_i)$ if $y_i < 0$. For small σ-values and $y_i \neq 0$ therefore, $M_Q = \sum_{i=1}^N |y_i| - y_i(2\Phi_{0,\sigma^2}(y_i) - 1) - 2\sigma^2\phi_{0,\sigma^2}(y_i)$ tends to zero.

Similarily, the variance can be approximated with $N\sigma^2$ if the mutation strength is sufficiently small. As a consequence, the quality gain can be assumed to be positive and approximately a linear function of σ as long as the mutation strength is sufficiently small.

Fig. 3.8. Dependence of $\overline{\Delta Q}_{1,20}$ on the noise strength σ_ϵ for function $F_{5.1}$ (Equation (3.20)). Depicted are the results for $\sigma = 1$, $\sigma = 0.5$, and $\sigma = 0.1$ from top to bottom.



a) function $F_{5.1}$, Eq. (3.20), $N = 10$



b) function $F_{5.1}$, Eq. (3.20), $N = 100$

The validity of the quality gain (2.13) was assessed by experiments using

$$F_{5.1}(\mathbf{y}) = -\sum_{i=1}^{N} |y_i| \tag{3.20}$$

as test function. The values were obtained by averaging over $500,000$ one-generation experiments with $\mathbf{y} = \mathbf{5}$ as starting vector. As can be seen in Figs. 3.8 and 3.9, the predictive quality is generally acceptable, which is even the case in the 10-dimensional space.

It is also of interest to look at the quality gain, if the parent vector is of the form $c\mathbf{e}^k$, where $\mathbf{e}^k$ is the kth unit vector – that is, we consider the quality gain starting in one of the edges of the rotated hypercube. There, the success probability is quite small which is reflected in the quality gain. The mean value of Q is given in this case by

$$M_Q = |c| - c(2\Phi_{0,\sigma^2}(c) - 1) - 2\sigma^2\phi_{0,\sigma^2}(c) - \frac{2}{\sqrt{2\pi}}(N-1)\sigma, \tag{3.21}$$

whereas the variance reduces to

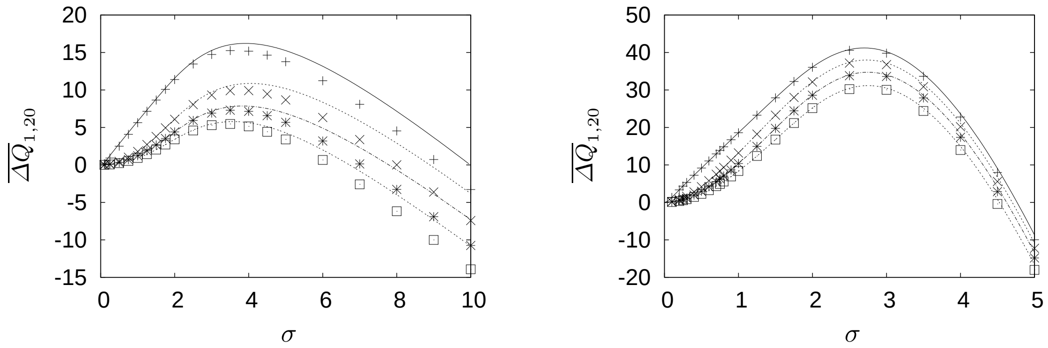
$$\begin{aligned} S_Q^2 &= \sigma^2[N - \frac{2}{\pi}(N-1)] + 4c^2\Phi_{0,\sigma^2}(c)[1 - \Phi_{0,\sigma^2}(c)] \\ &\quad -4\sigma^2 c\phi_{0,\sigma^2}(c)[\Phi_{0,\sigma^2}(c) - 1] - 4\sigma^4\phi^2_{0,\sigma^2}(c). \end{aligned} \tag{3.22}$$

For $N \gg 1$ and/or σ sufficiently small in relation to c, the statistical parameters can be approximated with $M_Q \approx -\frac{2}{\sqrt{2\pi}}(N-1)\sigma$ and $S_Q^2 \approx [N - \frac{2}{\pi}(N-1)]\sigma^2$ leading to the quality gain

$$\overline{\Delta Q}_{1,\lambda} \approx \sigma \frac{[N - \frac{2}{\pi}(N-1)]}{\sqrt{[N - \frac{2}{\pi}(N-1)] + (\frac{\sigma_\epsilon}{\sigma})^2}} c_{1,\lambda} - \frac{2}{\sqrt{2\pi}}(N-1)\sigma. \tag{3.23}$$

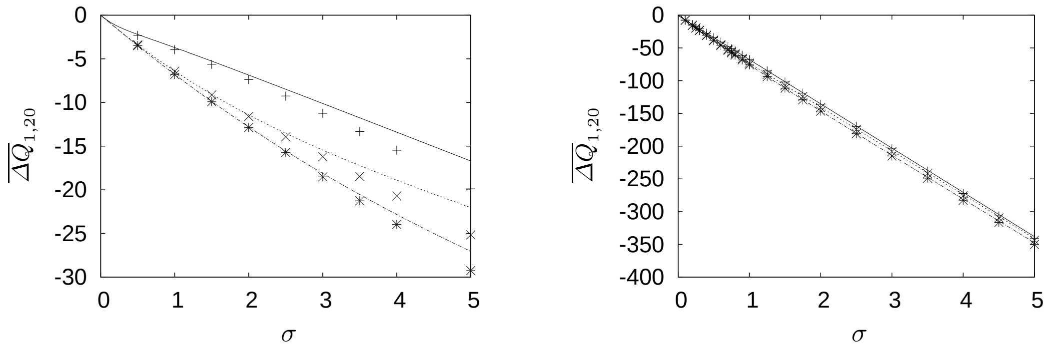
The quality gain now degrades into a approximately linear function of σ only assuming positive values if the offspring population size λ is rather high and the dimension N and the noise strength sufficiently small. Figure 3.10 shows the quality gain as a function of the mutation strength for $N = 10$ and $N = 100$. As starting vector $\mathbf{y} = 5\mathbf{e}^1$ was chosen.

Fig. 3.9. Dependence of $\overline{\Delta Q}_{1,20}$ on the mutation strength σ for function $F_{5.1}$ (3.20). Depicted are from top to bottom the results for $\sigma_\epsilon = 1$, $\sigma_\epsilon = 10$, $\sigma_\epsilon = 15$, and $\sigma_\epsilon = 20$.



a) function $F_{5.1}$, Eq. (3.20), $N = 10$ b) function $F_{5.1}$, Eq. (3.20), $N = 100$

Fig. 3.10. Dependence of $\overline{\Delta Q}_{1,20}$ on the mutation strength σ for function $F_{5.1}$, (3.20), starting in $\mathbf{y} = 5\mathbf{e}^1$. Depicted are from top to bottom the results for $\sigma_\epsilon = 1$, $\sigma_\epsilon = 10$, and $\sigma_\epsilon = 20$.



a) function $F_{5.1}$, Eq. (3.20), $N = 10$ b) function $F_{5.1}$, Eq. (3.20), $N = 100$

3.1.5 OneMax

Equation (2.13), $\overline{\Delta Q}_{1,\lambda} \simeq S_Q^2 / \sqrt{S_Q^2 + \sigma_\epsilon^2} c_{1,\lambda} + M_Q$, was developed for continuous distributions. As already mentioned, the discrete bit-counting function OneMax $F_5(\mathbf{y}) = \sum_{i=1}^N y_i$, therefore represents an extreme test case for the applicability. For OneMax, the local quality is defined as $Q(\mathbf{x}) = F(\mathbf{x}) - F(\mathbf{y})$. Considering strings of length N, the statistical parameters are given as (see [6, p. 128f])

$$M_Q = -p_m(2F_0 - N) \tag{3.24}$$

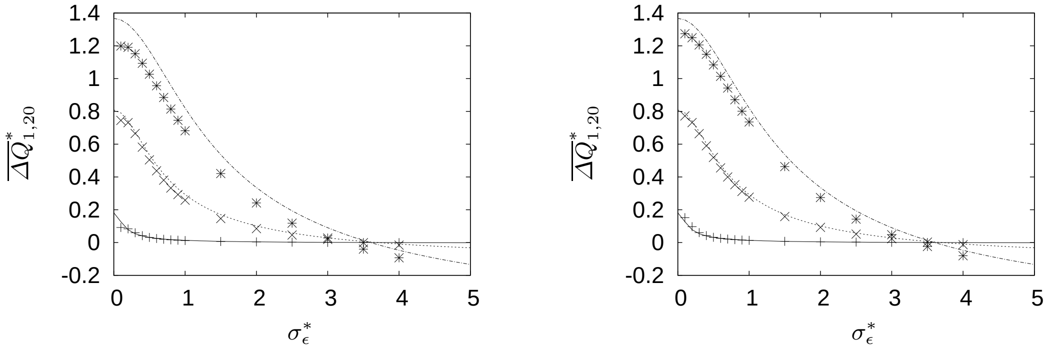
and

$$S_Q = \sqrt{N}\sqrt{p_m(1 - p_m)}, \tag{3.25}$$

where $F_0 := \sum_{i=1}^N y_i$ is the fitness value of the parental vector and p_m is the mutation probability. If we insert these values into (2.13), we obtain

$$\overline{\Delta Q}_{1,\lambda} \simeq \frac{N p_m(1 - p_m)}{\sqrt{N p_m(1 - p_m) + \sigma_\epsilon^2}} c_{1,\lambda} - p_m(2F_0 - N). \tag{3.26}$$

Fig. 3.11. Dependence of $\overline{\Delta Q^*}_{1,20}$ on the noise strength σ^*_ϵ for OneMax. Depicted are from bottom to top the results for $\sigma^* = 0.1$,$\sigma^* = 0.5$, and $\sigma^* = 1$.



a) function OneMax, Eq. (1.5), $N = 30$ b) function OneMax, Eq. (1.5), $N = 100$

If $F_0 < N/2$, the resulting quality gain will be positive since less than half of the bit positions are occupied with ones.

Equation (3.26) can be further simplified and normalized. Provided that $p_m \ll 1$, the influence of $(1 - p_m)$ can be neglected and Np_m can be formally identified with the variance σ^2. If the parental value F_0 is greater than $N/2$, we can introduce the following normalizations $\sigma^* = 2(2F_0/N - 1)\sigma$,
$\sigma^*_\epsilon = 2(2F_0/N - 1)\sigma_\epsilon$, and $\overline{\Delta Q^*}_{1,\lambda} = 2(2F_0/N - 1)\overline{\Delta Q}_{1,\lambda}$ (see [6, p.130]), leading finally to

$$\overline{\Delta Q^*}_{1,\lambda} \simeq \sigma^* \frac{1}{\sqrt{1 + (\frac{\sigma^*_\epsilon}{\sigma})^2}} c_{1,\lambda} - \frac{(\sigma^*)^2}{2}. \tag{3.27}$$

The predictive quality of (3.27) was examined for a 30- and 100-dimensional search space where $F_0 = 20$ and $F_0 = 60$, respectively have been chosen. The values were obtained by averaging over $500,000$ one-generation experiments. As can be seen in Fig. 3.12, (3.27) predicts the behavior up to a mutation strength of $\sigma^* = 1.5$ reasonably well, which holds even for the case $N = 30$. The mutation strength 1.5 corresponds to a mutation probability of $p_m = 0.141$ for $N = 100$ and $p_m = 0.169$ for $N = 30$. In both cases, the mutation probabilities are considerably higher than the mutation rate $p_m = 1/N$, which is often recommended.

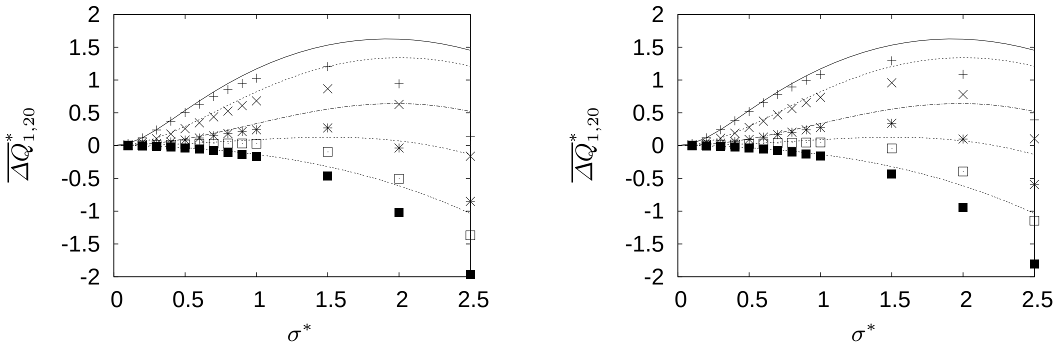
3.2 Steady State Behavior

This section is concerned with the predictive quality of the equations or inequality that describe the final fitness error in the case of the L_1-norm. The equations are derived for $(1, \lambda)$-ES but will be extended using a simple substitution to $(\mu/\mu_I, \lambda)$-ES. The only exception is the function OneMax which was only examined for $(1, \lambda)$ strategies.

3.2.1 Biquadratic Functions

This subsection is concerned with the predictive quality of $\mathrm{E}[\Delta F] \approx 3N\sigma_\epsilon/(40c_{1,\lambda})$, (2.30), describing the final fitness error of biquadratic functions of type $F(\mathbf{y}) = -\sum_{i=1}^{N} c_i y_i^4$. Its applicability was tested not only for the $(1, \lambda)$-ES but also for the $(\mu/\mu_I, \lambda)$ variant,

Fig. 3.12. Dependence of $\overline{\Delta Q^*}_{1,20}$ on the mutation strength σ^* for OneMax. Depicted are from top to bottom the results for $\sigma^*_\epsilon = 0.5$, $\sigma^*_\epsilon = 1$, $\sigma^*_\epsilon = 2$, $\sigma^*_\epsilon = 3$, and $\sigma^*_\epsilon = 5$.



a) function OneMax, Eq. (1.5), $N = 30$ b) function OneMax, Eq. (1.5), $N = 100$

substituting $c_{1,\lambda}$ with $\mu c_{\mu/\mu,\lambda}$. As test functions, variants of the functions (3.17) - (3.19)

$$F_{3.4}(\mathbf{y}) = -\sum_{i=1}^{N} y_i^4, \tag{3.28}$$

$$F_{3.5}(\mathbf{y}) = -\sum_{i=1}^{N} i y_i^4, \quad \text{and} \tag{3.29}$$

$$F_{3.6}(\mathbf{y}) = -\sum_{i=1}^{N} i^2 y_i^4. \tag{3.30}$$

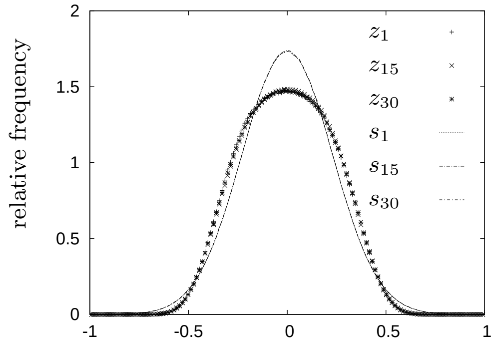
were chosen.

Validity of the Normal Distribution Approximation Figure 3.13 shows relative frequencies for some $(\mu/\mu_I, 60)$-ES runs on the biquadratic test functions, i.e. (3.28)-(3.30). The values were obtained by several trials – sampling the normalized z_i values for a total of $7,200,000$ generations in the steady state region and aggregating them in 109 intervals ranging from -1.5 to 1.5. Also included in the figure are relative frequencies drawn from corresponding normal distributions with the experimentally obtained standard deviations.

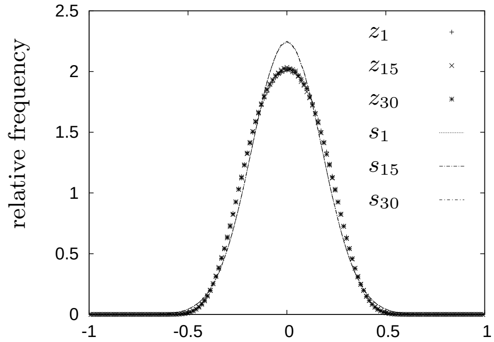
As can be seen in the plots, especially the lower-order components show deviations from the normal distribution. This seems to be a result of the weak selective pressure that acts on these coordinate values in the case of (3.29) and (3.30). In order to enable a comparison with the normal distribution, the skewness and the kurtosis were also calculated. The skewness generally assumes the smallest values in the case of (3.28) ranging from 0.00014 to 0.01. In the case of (3.29) and (3.30), several lower-order components tend to show higher deviations from zero, although sometimes the skewness of (3.29) is even lower than that of (3.28). The kurtosis assumes values between -0.5 and -0.8 for all functions which does not support the assumption of normally distributed variables. For higher μ-values, the components generally agree better with the assumption of a normal distribution than for smaller.

Predictive Quality of Equation (2.30) The plots in Figure 3.14 show the dependency of the final fitness error $\mathrm{E}[\Delta F]$ on the number of parents μ for some $(\mu/\mu_I, 60)$-ES

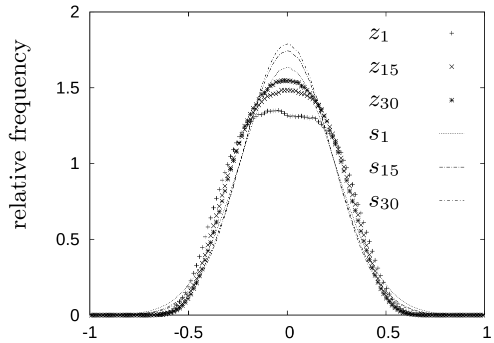
Fig. 3.13. Relative frequencies for several $(\mu/\mu_I, 60)$-ES using σSA ($N = 30$). Shown are the values for the first, 15th, and 30th normalized coordinate. Also included are the corresponding relatives frequencies s_i drawn from normal distributions using the experimentally found standard deviations.



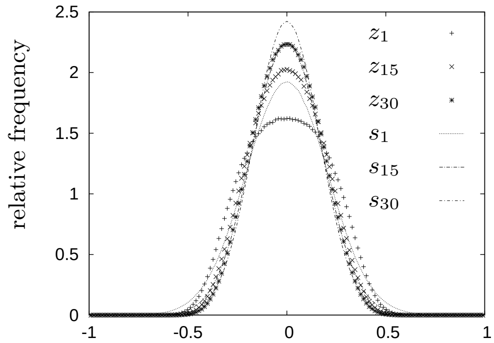
a) $(10/10_I, 60)$-σSA-ES, function $F_{3.4}$ (3.28)



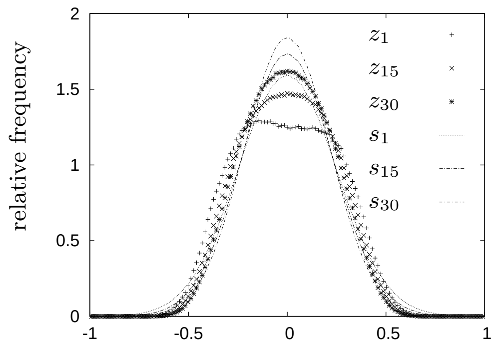
b) $(30/30_I, 60)$-σSA-ES, function $F_{3.4}$ (3.28)



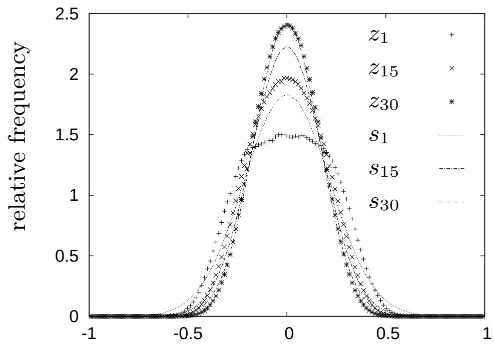
c) $(10/10_I, 60)$-σSA-ES, function $F_{3.5}$ (3.29)



d) $(30/30_I, 60)$-σSA-ES, function $F_{3.5}$ (3.29)



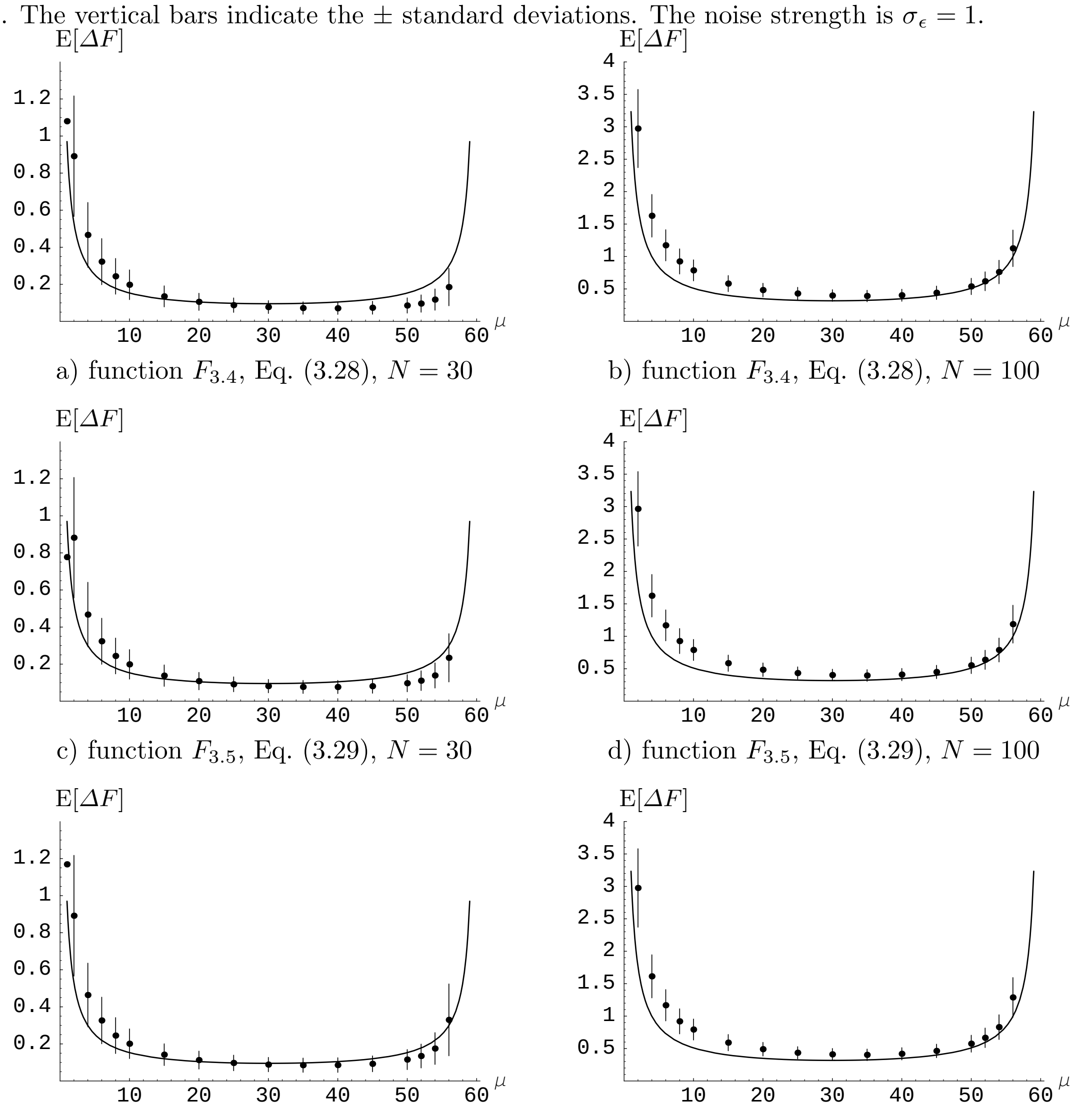
e) $(10/10_I, 60)$-σSA-ES, function $F_{3.6}$ (3.30)



f) $(30/30_I, 60)$-σSA-ES, function $F_{3.6}$ (3.30)

runs. Depicted are the values as predicted by (2.30) and the actually measured ones. The experimental values were obtained by averaging over 900,000 generations in the steady state region. The values for $\mu = 58$ and $\mu = 59$ are not shown because of convergence problems for these ES. The standard mutative σ self-adaptation (σSA, cp. [5]) was used as σ-control rule in all experiments. The overall agreement between equation and experiment is rather good although some deviations can be observed. The experimentally found final fitness error of the $(1, 60)$ strategy is significantly higher than the theoretically obtained lower bound. A similar case was already observed in [7] in the case of quadratic test functions. $(1, \lambda)$ strategies using σSA tend to premature stagnation which is the consequence of a very fast reduction of the mutation strength. This can also be observed in Fig. 3.15

Fig. 3.14. Final fitness error for biquadratic functions. Shown are the calculated and experimental values for E[$\varDelta F$] of $(\mu/\mu_I, 60)$-ES with μ =1, 2, 4, 6, 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 52, 54, 56. The vertical bars indicate the ± standard deviations. The noise strength is $\sigma_\epsilon = 1$.



a) function $F_{3.4}$, Eq. (3.28), $N = 30$ b) function $F_{3.4}$, Eq. (3.28), $N = 100$

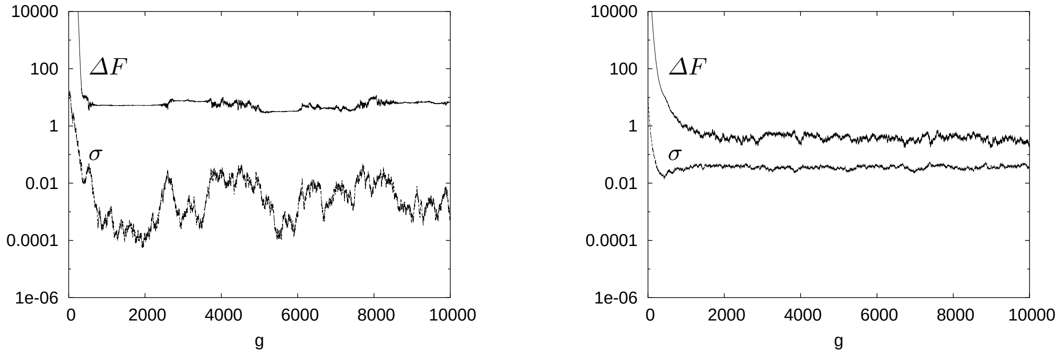
c) function $F_{3.5}$, Eq. (3.29), $N = 30$ d) function $F_{3.5}$, Eq. (3.29), $N = 100$

e) function $F_{3.6}$, Eq. (3.30), $N = 30$ f) function $F_{3.6}$, Eq. (3.30), $N = 100$

showing exemplary runs of an $(1, 60)$-ES and of a $(30/30_I, 60)$-ES optimizing test function $F_{3.5}$, (3.29). In the case of the second ES, σ is reduced to a value around 0.03, where it starts to fluctuate. As a consequence, $\varDelta F$ also approaches its steady state a short time later. In contrast to this, the mutation strength is reduced faster and further in the case of the $(1, 60)$-ES reaching values that are much closer to zero. Therefore, $\varDelta F$ stagnates very soon. For shorter periods, σ can attain higher values though being unable to stabilize at this level. The reasons for this behavior are not fully understood. Some deviations between the predicted curve and the experimental data can be found. The equation leads to a symmetric curve for the fitness error that is centered around $\mu = 30$. The experiments do not always show exactly the same behavior sometimes exhibiting a minimal value for $\mu = 35$.

Apart from this, the measured values tend to be similar. Their standard deviations decrease when μ approaches 30 and increase when μ is close to one or 60.

Fig. 3.15. ΔF and σ-dynamics of an $(1, 60)$ and a $(30/30_I, 60)$ ES using σSA. The noise strength is $\sigma_\epsilon = 1$.



a) $(1, 60)$-ES on function $F_{3.5}$, Eq. (3.29) b) $(30/30_I, 60)$-ES on function $F_{3.5}$, Eq. (3.29)

3.2.2 Negative L_1-Norm

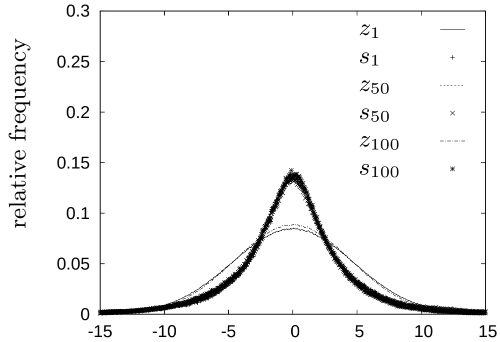
In the case of the negative L_1-norm, i.e. function F_5, we were only able to derive a lower bound for the final fitness error given by $\mathrm{E}[\Delta F] \geq N\sigma_\epsilon/(\pi c_{1,\lambda})$ (2.43), resp. $\mathrm{E}[\Delta F] \geq N\sigma_\epsilon/(\pi\mu c_{\mu/\mu_I,\lambda})$, making several assumptions the validity of which will be checked in the next subparagraph. The experiments were conducted using $(\mu/\mu_I, 60)$-ES with σSA and CSA(cumulative step length adaptation) [12, 14, 1] as σ control rules. In the case of the latter, a σ_{min}-value was introduced in order to prevent premature convergence. The noise strength used was $\sigma_\epsilon = 1$.

Validity of the Assumptions The first assumption we introduced was the equipartition assumption postulating that the differences between the distributions of the variables z_i can be neglected once the ES has reached the steady state. In the following, we also assumed that the distribution of the z_i can be approximated by a normal distribution. Using these assumptions, we could derive the lower bound (2.40) for the mutation strength σ, i.e. $\sigma \geq \sigma_\epsilon/c_{1,\lambda}\sqrt{2/\pi}(\sqrt{\sigma_z^2+1} - \sigma_z)/(1 - \frac{2}{\pi\sqrt{2\sigma_z^2+1}})$ which depends on the standard deviation σ_z of the z_i, on the noise strength σ_ϵ, and on the progress coefficient. Finally, we assumed $\sigma_z \geq 1$ to derive the lower bound for $\mathrm{E}[\Delta F]$.

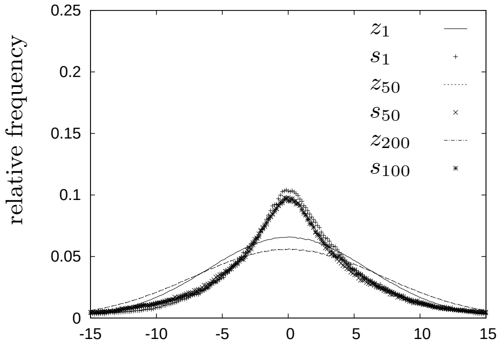
We will first examine the validity of the equipartition assumption and the approximation with the normal distribution. Figure 3.16 shows some relative frequency plots for $(10/10_I, 60)$-, $(30/30_I, 60)$-, and $(40/40_I, 60)$-ES runs using σSA as the control rule for the mutation. The values were obtained by sampling over a total of $800,000$ generations in the steady state with eight restarts grouping the values in 509 ($N = 100$) and 209 ($N = 200$) intervals ranging from -15.5 to 15.5. As test function function $F_{5.1}$, Eq. (3.20), was used. Also shown are the frequency plots of normal distribution functions with the experimentally obtained σ_z-values. As one can see in the figure, the standard deviations of the selected z_i are different. For $N = 100$ for example, the standard deviations range from 1.672 to 1.951 (average value 1.823) in the case of the $(40/40_I, 60)$-ES, from 1.998 to 2.424 (2.172) for the $(30/30_I, 60)$-ES, and from 4.215 to 5.058 with an mean value of 4.581 for the $(10/10_I, 60)$-ES.

Also the assumption of a normal distribution seems to be only a rough approximation since the actual distribution function is generally steeper and narrower. This is also the case for the higher dimensional search space.

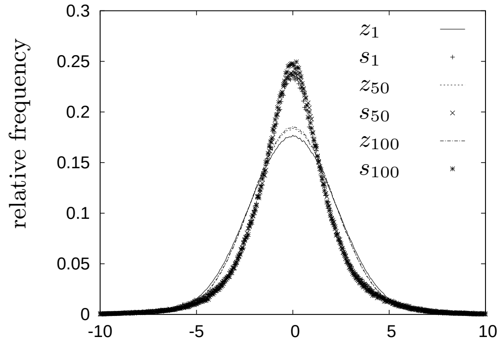
Fig. 3.16. Relative frequency plots for function $F_{5.1}$ (3.20) for some $(\mu/\mu_I, 60)$-ES using σSA. Also included are the corresponding relatives frequencies s_i drawn from normal distributions using the experimentally found standard deviations.



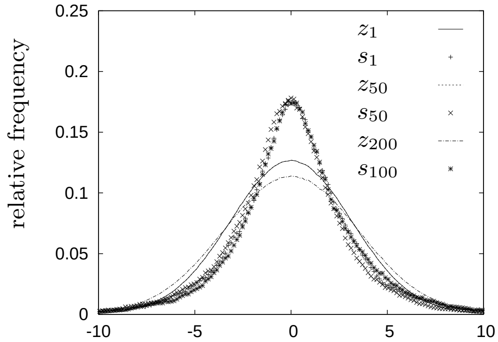
a) $F_{5.1}$, $N = 100$, $\sigma_\epsilon = 1$, $\mu = 10$



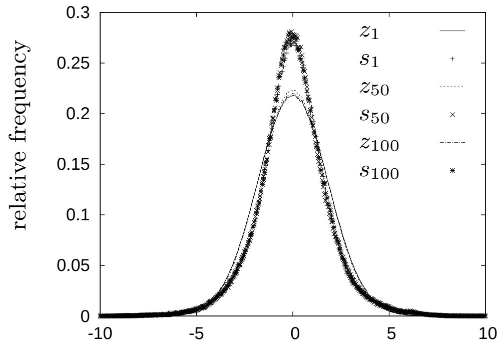
b) $F_{5.1}$, $N = 200$, $\sigma_\epsilon = 1$, $\mu = 10$



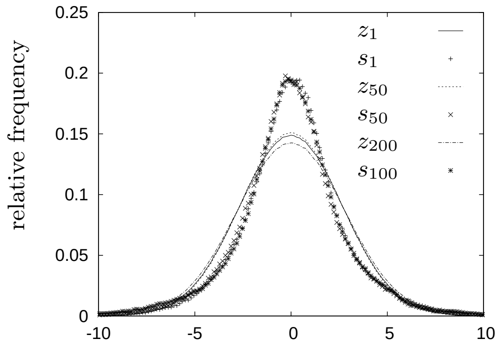
c) $F_{5.1}$, $N = 100$, $\sigma_\epsilon = 1$, $\mu = 30$



d) $F_{5.1}$, $N = 200$, $\sigma_\epsilon = 1$, $\mu = 30$



e) $F_{5.1}$, $N = 100$, $\sigma_\epsilon = 1$, $\mu = 40$

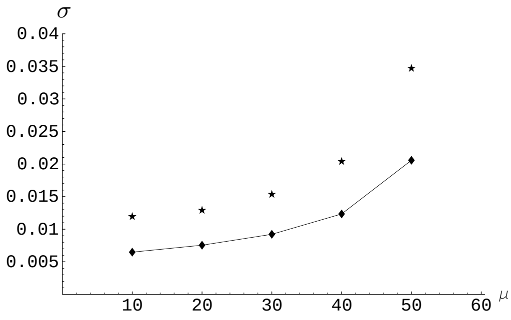


f) $F_{5.1}$, $N = 200$, $\sigma_\epsilon = 1$, $\mu = 40$

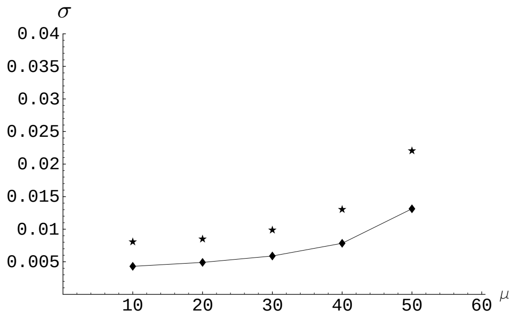
Concerning the validity of the lower bound (2.40), Figure 3.17 shows the lower bound for σ which was obtained by inserting the experimentally found and averaged σ_z-values (depicted in Fig. 3.18) into (2.40). There is a great difference between ES using σSA and CSA. In the former case, the increase of the dimension leads to higher σ_z values, although the relative deviation of the predicted lower bound does not change and generally stays at values around 0.4 to 0.5. Higher deviations are only found for small or high μ-values. Generally, σ_z decreases if more parents are used. In the case of ES using CSA, there is a minimal σ_z value which seems to be roughly at $\mu = \lambda/2$.

The assumption $\sigma_z \geq 1$ seems to be justified at least in the higher-dimensional search spaces. For $N = 30$, there exist some μ-values for which it is violated if σSA is used.

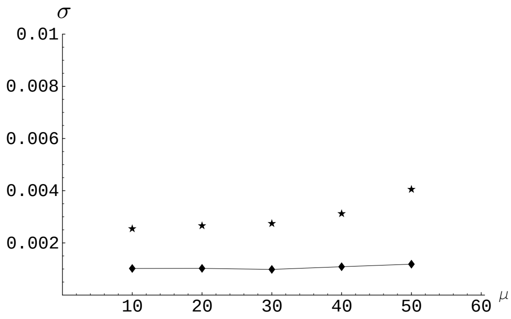
Fig. 3.17. Lower bound for the mutation strength σ obtained by Eq. (2.40) using the average σ_z in the steady state. The points indicate the average measured σ-values.



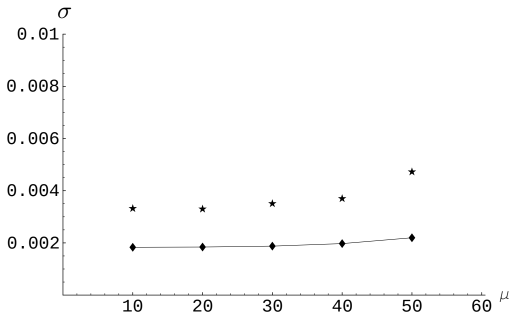
a) $F_{5.1}$, $N = 100$, $\sigma_\epsilon = 1$



b) $F_{5.1}$, $N = 200$, $\sigma_\epsilon = 1$

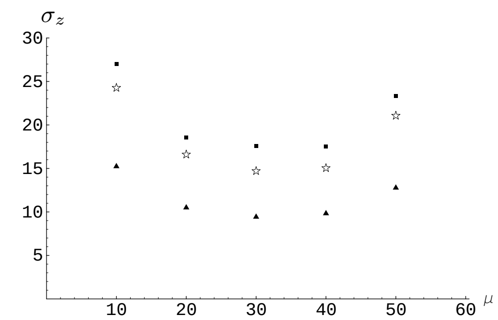


c) $F_{5.1}$, $N = 100$, CSA, $\sigma_\epsilon = 1$

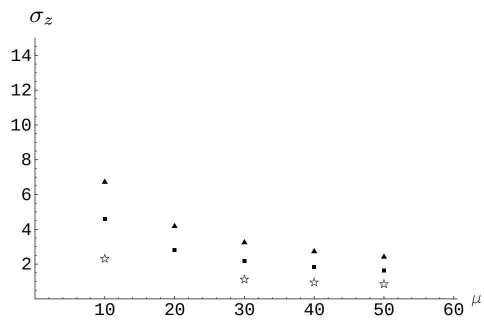


d) $F_{5.1}$, $N = 200$, CSA, $\sigma_\epsilon = 1$

Fig. 3.18. Average measured values for σ_z for several $(\mu/\mu_I, 60)$-ES runs.



a) $F_{5.1}$, CSA, $\sigma_\epsilon = 1$
$N = 30$ (stars), $N = 100$ (boxes),
$N = 200$ (triangles)



b) $F_{5.1}$, σSA, $\sigma_\epsilon = 1$
$N = 30$ (stars), $N = 100$ (boxes),
$N = 200$ (triangles)

Since the experiments show that the plots of the actual relative frequencies of the z_i tend to be steeper and narrower than those obtained by using the normal distribution, we introduce an alternative distribution. Considering Fig. 3.16, the double exponential distribution with the density function $p(z_i) = 1/(2\alpha)\mathrm{e}^{-|z_i|/\alpha}$ could be an appropriate choice. To obtain an estimate of the final fitness error, we start again with (2.37), i.e.

$$N\sigma - \frac{2\sigma}{\pi}\sum_{i=1}^{N}\mathrm{e}^{-z_i^2} \geq \left(\sum_{i=1}^{N} z_i(2\Phi(z_i)-1) + \frac{2}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z_i^2}{2}} - |z_i|\right)\frac{\sigma_\epsilon}{c_{1,\lambda}}.$$

In order to continue, we need to calculate the expected values of the terms in (2.37) which are given after some straightforward calculations as

$$\begin{aligned}
\mathrm{E}[|z_i|] &= \alpha, \\
\mathrm{E}[z_i(2\Phi(z_i)-1)] &= 2(\alpha - \frac{1}{\alpha})\Phi(-\frac{1}{\alpha})\mathrm{e}^{\frac{1}{2\alpha^2}} + \sqrt{\frac{2}{\pi}}, \\
\frac{2}{\sqrt{2\pi}}\mathrm{E}[\mathrm{e}^{-\frac{z_i^2}{2}}] &= \frac{2}{\alpha}\Phi(-\frac{1}{\alpha})\mathrm{e}^{\frac{1}{2\alpha^2}}, \text{ and} \\
\frac{2\sigma}{\pi}\mathrm{E}[\mathrm{e}^{-z_i^2}] &= \frac{2}{\sqrt{\pi}\alpha}\sigma\Phi(-\frac{1}{\sqrt{2}\alpha})\mathrm{e}^{\frac{1}{4\alpha^2}}.
\end{aligned} \tag{3.31}$$

Taking the expectation of both sides of (2.37) therefore leads to

$$\begin{aligned}
\sigma\left(1 - \frac{2}{\sqrt{\pi}\alpha}\Phi(-\frac{1}{\sqrt{2}\alpha})\mathrm{e}^{\frac{1}{4\alpha^2}}\right) &\geq \frac{\sigma_\epsilon}{c_{1,\lambda}}\left(2(\alpha - \frac{1}{\alpha})\Phi(-\frac{1}{\alpha})\mathrm{e}^{\frac{1}{2\alpha^2}} + \sqrt{\frac{2}{\pi}} + \frac{2}{\alpha}\Phi(-\frac{1}{\alpha})\mathrm{e}^{\frac{1}{2\alpha^2}} - \alpha\right) \\
\Rightarrow \sigma &\geq \frac{\sigma_\epsilon}{c_{1,\lambda}}\left(\frac{2\alpha\Phi(-\frac{1}{\alpha})\mathrm{e}^{\frac{1}{2\alpha^2}} + \sqrt{\frac{2}{\pi}} - \alpha}{1 - \frac{2}{\sqrt{\pi}\alpha}\Phi(-\frac{1}{\sqrt{2}\alpha})\mathrm{e}^{\frac{1}{4\alpha^2}}}\right).
\end{aligned} \tag{3.32}$$

The final fitness error $\mathrm{E}[\Delta F] = \sum_{i=1}^{N} \mathrm{E}[|y_i|]$ is given as $\mathrm{E}[\Delta F] = N\sigma\alpha$. With (3.32), we obtain the inequality

$$E[\Delta F] \geq N\frac{\sigma_\epsilon}{c_{1,\lambda}}\left(\frac{\alpha^2(2\Phi(-\frac{1}{\alpha})\mathrm{e}^{\frac{1}{2\alpha^2}} - 1) + \sqrt{\frac{2}{\pi}}\alpha}{1 - \frac{2}{\sqrt{\pi}\alpha}\Phi(-\frac{1}{\sqrt{2}\alpha})\mathrm{e}^{\frac{1}{4\alpha^2}}}\right). \tag{3.33}$$

This lower bound for the final fitness error depends on the parameter α which is unknown. Therefore, we consider the function

$$h(x) = \frac{x^2(2\Phi(-\frac{1}{x})\mathrm{e}^{\frac{1}{2x^2}} - 1) + \sqrt{\frac{2}{\pi}}x}{1 - \frac{2}{\sqrt{\pi}x}\Phi(-\frac{1}{\sqrt{2}x})\mathrm{e}^{\frac{1}{4x^2}}}. \tag{3.34}$$

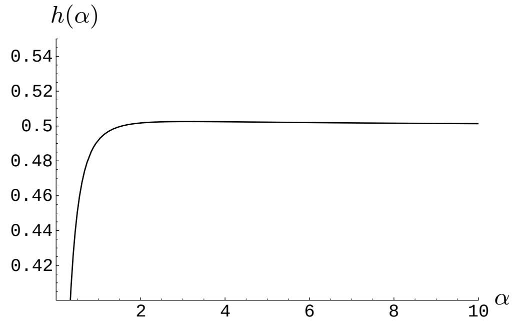
The denominator in (3.34) goes to 1 for $x \to \infty$, whereas $x^2(2\Phi(-\frac{1}{x})\mathrm{e}^{\frac{1}{2x^2}} - 1) + \sqrt{\frac{2}{\pi}}x$ approaches 1/2 which can be easily demonstrated by applying l'Hospital's Rule (see Appendix 5.1).

The plot of (3.34) is shown in Figure 3.19. When using the normal distribution, we assumed $\sigma \geq 1$. Since the variance of a double exponentially distributed variable z is given by $\mathrm{Var}[z] = 2\alpha^2$, we assume here $\alpha \geq 1/\sqrt{2}$. Since $h(x) \geq 0.47715$ for all $x \geq 1/\sqrt{2}$ (see Fig 3.19 or Appendix 5.1), we obtain the lower bound

$$\mathrm{E}[\Delta F] \geq 0.47715N\frac{\sigma_\epsilon}{c_{1,\lambda}}. \tag{3.35}$$

The derivation of (3.35) was analog to that of (2.43). We first introduced an equipartition assumption before deriving an inequality for the mutation strength. In both cases, this is given as a function of an unknown parameter for which a lower bound has to be derived. The only difference stems from using a different distribution. The predictive quality of (3.35) and (2.43) will be assessed in the next paragraph.

Fig. 3.19. The function $h(\alpha)$ obtained in the second derivation of the lower bound for function F_5 using the double exponential distribution.



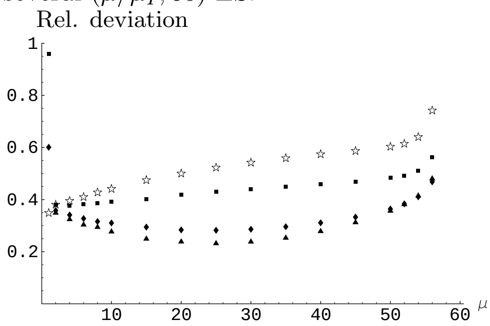
Predictive Quality of the Lower Bounds Eq. (2.43) and Eq. (3.35) The validity of the lower bound (2.43), $\mathrm{E}[\Delta F] \geq N\sigma_\epsilon/(\pi\mu c_{\mu/\mu_I,\lambda})$, was tested by conducting experiments with $(\mu/\mu_I, 60)$-ES for a noise strength of 1 using a 30-, 100-, 200, and a 300-dimensional search space. Function 5.1 (3.20) was again used as the test function. The experimental values were obtained by sampling over $900,000$ generations in the steady state. As σ control rules, σSA and CSA were used. As one can see in Fig. 3.21, the lower bound is not violated and seems to be a reasonable approximation of the real final fitness error. Figure 3.20 depicts the relative deviation of the lower bound from the actual measured values showing the decrease of the deviation when the dimension of the search space is increased. It is interesting that the increase from $N = 200$ to $N = 300$ leads towards relatively higher values of the final fitness error. Therefore, the relative deviation between prediction and experiments becomes greater although rather marginally. Generally, the CSA seems to lead towards smaller values especially for higher parental numbers where the deviation of the ES using σSA increases.

Figures 3.23 and 3.22 show the results obtained by (3.35), i.e. $\mathrm{E}[\Delta F] \geq 0.47715N \sigma_\epsilon/(\mu c_{\mu/\mu,\lambda})$, using the double exponential distribution instead of the normal distribution in the derivation. Generally, the approximation quality seems to be better compared to using (2.43), although the lower bound is violated for $N = 200$ and $N = 300$. This might be a hint that the double exponential distribution is not a appropriate choice for higher-dimensional search spaces although the deviations for $N = 300$ are smaller than for $N = 200$.

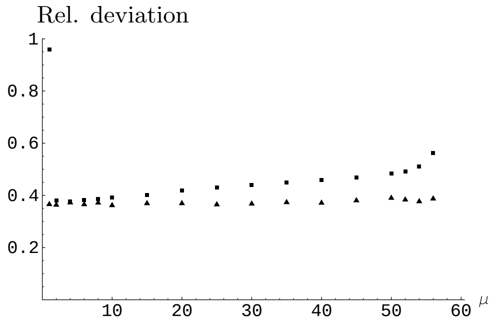
OneMax The predictive quality of (2.49), $\mathrm{E}[\Delta F] \approx \frac{N}{2}(1 - c_{1,\lambda}/(\sqrt{1+\sigma_\epsilon^2}))$, shall be assessed in this paragraph. Equation (2.49) describes the final fitness error of the function OneMax $F_4(\mathbf{y}) = \sum_{i=1}^{N} y_i$ under the conditions that N is sufficiently large and that the mutation rate $p_m = \frac{1}{N}$ is used.

The experimental data were obtained by sampling over $500,000$ generations in the steady state region. Figure 3.24 shows the final fitness error as a function of the offspring population size λ for several noise strengths. As already mentioned, (2.49) can only be applied if $c_{1,\lambda} \leq \sqrt{1+\sigma_\epsilon^2}$. This property can also be seen in the figure. The approximation quality is quite good as long as σ_ϵ is sufficiently high. For smaller noise strengths, i.e. $\sigma_\epsilon = 1$, the experimental data depart very soon from the curve obtained by (2.49).

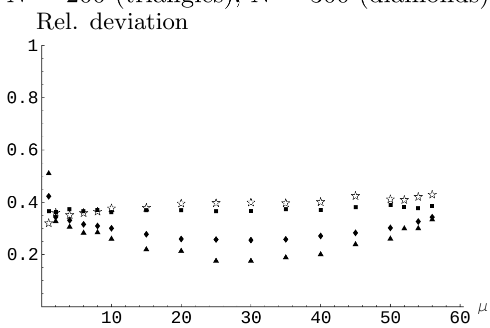
Fig. 3.20. The relative deviation of E[ΔF] as predicted by (2.43) from the experimental values for several $(\mu/\mu_I, 60)$-ES.



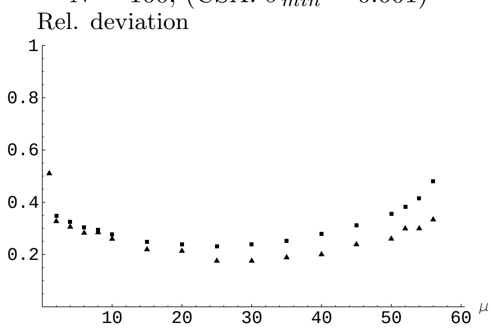
Influence of the dimension
σSA, $N = 30$ (stars), $N = 100$ (boxes),
$N = 200$ (triangles), $N = 300$ (diamonds)



Comparison between
σSA (boxes) and CSA (triangles)
$N = 100$, (CSA: $\sigma_{min} = 0.001$)

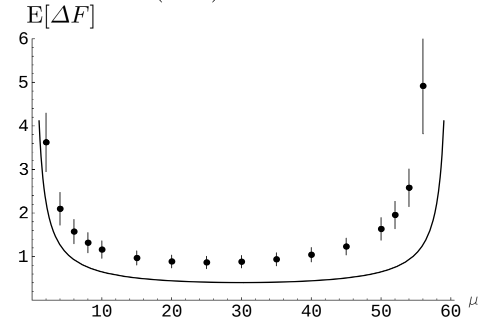


Influence of the dimension
CSA, $N = 30$ (stars), $N = 100$ (boxes),
$N = 200$ (triangles), $N = 300$ (diamonds)

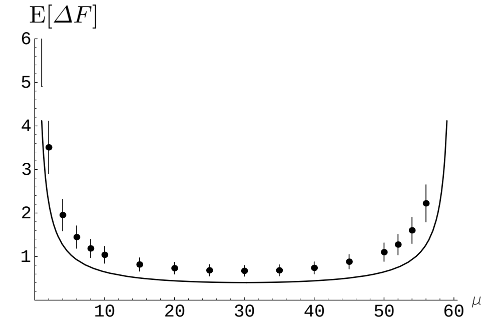


Comparison between
σSA (boxes) and CSA (triangles)
$N = 200$, (CSA: $\sigma_{min} = 0.0025$)

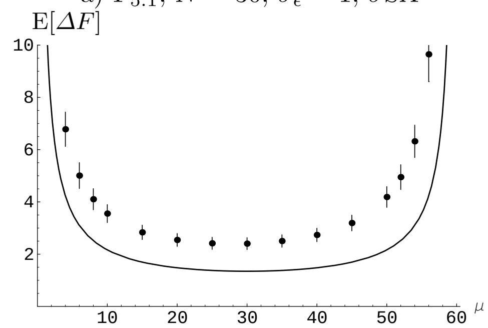
Fig. 3.21. $\mathrm{E}[\Delta F]$ for some $(\mu/\mu_I, 60)$-ES optimizing function F_5. The results were obtained using σSA and CSA as the control rule for the mutation strength adaptation. Also shown are the predictions of (2.43).The vertical bars indicate the $\pm$ standard deviations.



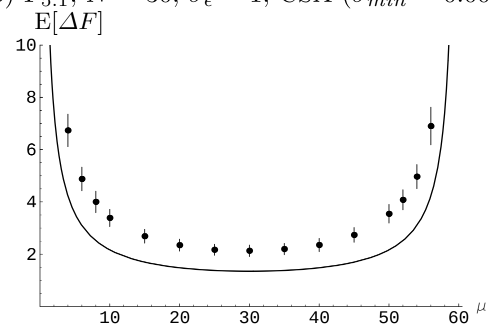
a) $F_{5.1}$, $N = 30$, $\sigma_\epsilon = 1$, σSA



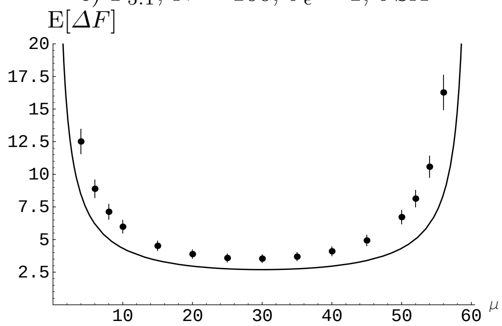
b) $F_{5.1}$, $N = 30$, $\sigma_\epsilon = 1$, CSA ($\sigma_{min} = 0.001$)



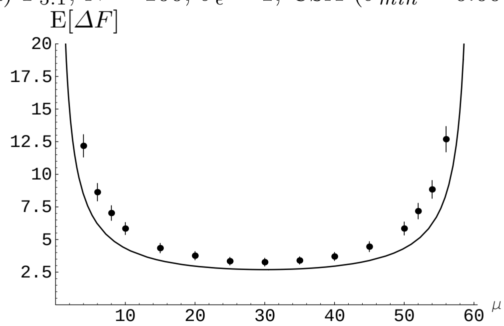
c) $F_{5.1}$, $N = 100$, $\sigma_\epsilon = 1$, σSA



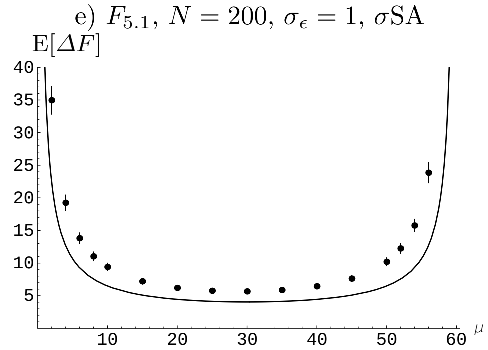
d) $F_{5.1}$, $N = 100$, $\sigma_\epsilon = 1$, CSA ($\sigma_{min} = 0.001$)



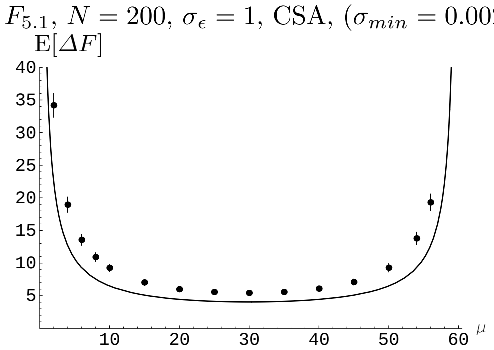
e) $F_{5.1}$, $N = 200$, $\sigma_\epsilon = 1$, σSA



f) $F_{5.1}$, $N = 200$, $\sigma_\epsilon = 1$, CSA, ($\sigma_{min} = 0.0025$)

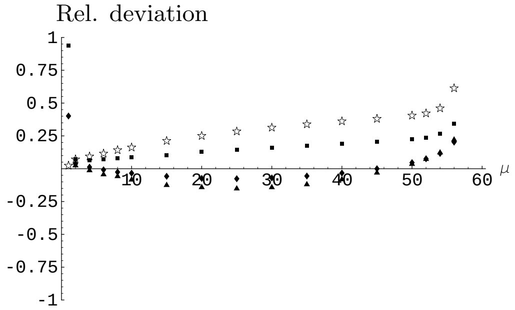


g) $F_{5.1}$, $N = 300$, $\sigma_\epsilon = 1$, σSA

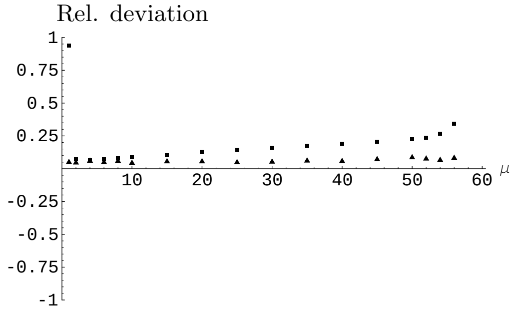


h) $F_{5.1}$, $N = 300$, $\sigma_\epsilon = 1$, CSA, ($\sigma_{min} = 0.0025$)

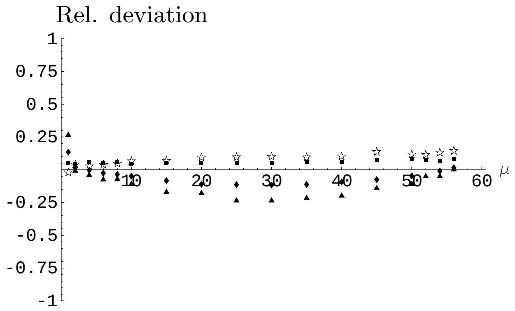
Fig. 3.22. The relative deviation of $\mathrm{E}[\Delta F]$ as predicted by (3.35) from the experimental values for several $(\mu/\mu_I, 60)$-ES.



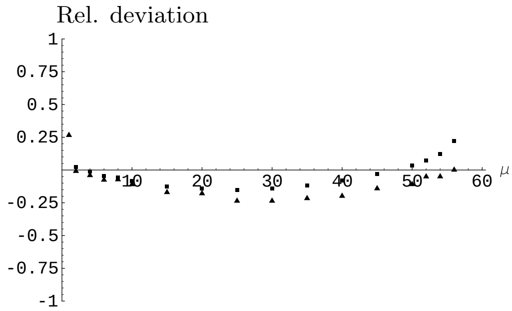
Influence of the dimension
σSA, $N = 30$ (stars), $N = 100$ (boxes),
$N = 200$ (triangles), $N = 300$ (diamonds)



Comparison between
σSA (boxes) and CSA (triangles)
$N = 100$,(CSA: $\sigma_{min} = 0.001$)

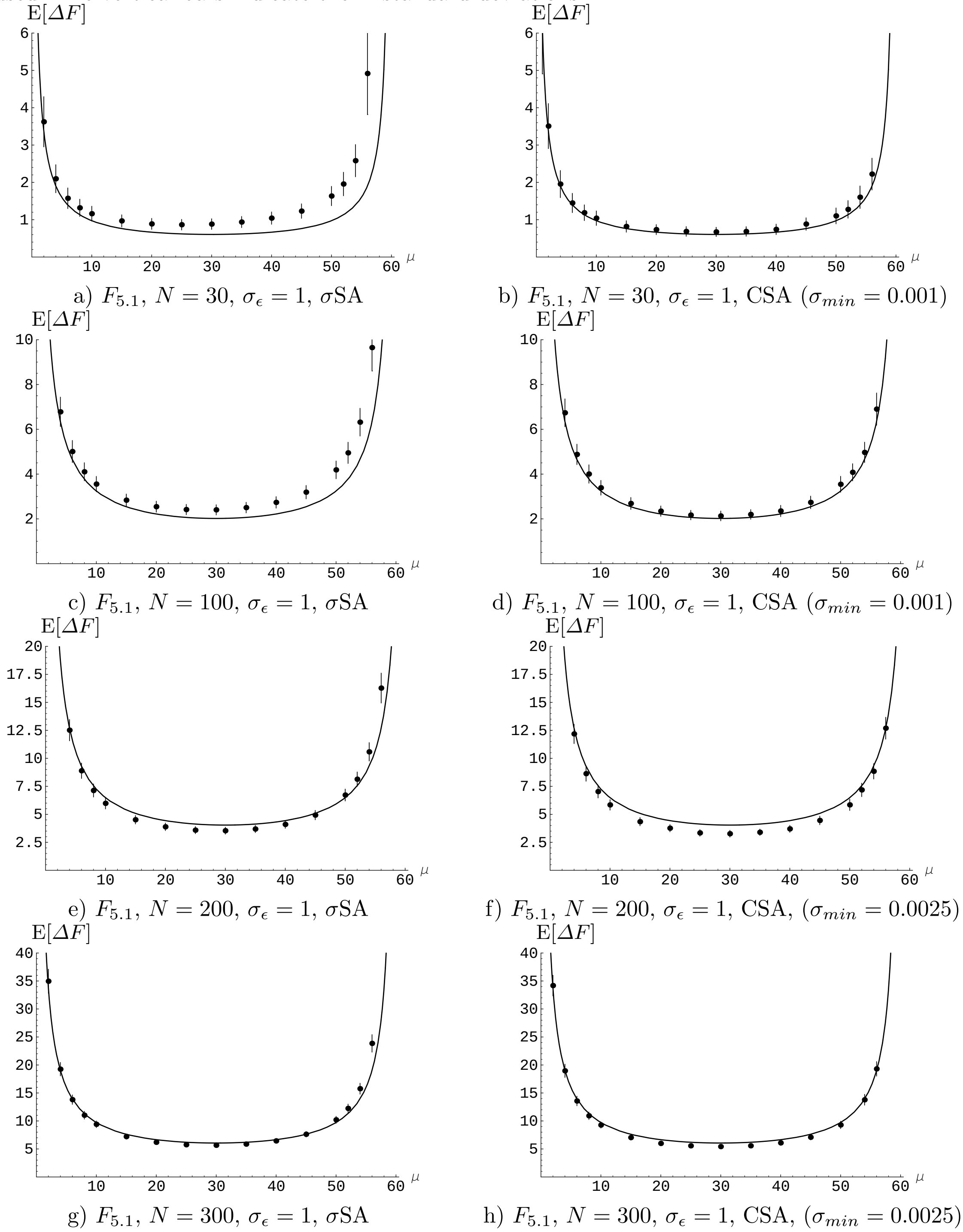


Influence of the dimension
CSA, $N = 30$ (stars), $N = 100$ (boxes),
$N = 200$ (triangles), $N = 300$ (diamonds)



Comparison between
σSA (boxes) and CSA (triangles)
$N = 200$, (CSA: $\sigma_{min} = 0.0025$)

Fig. 3.23. $\mathrm{E}[\Delta F]$ for some $(\mu/\mu_I, 60)$-ES optimizing function F_5. The results obtained by (3.35) are shown as curves. As control rule for the mutation strength adaptation, σSA and CSA were used. The vertical bars indicate the $\pm$ standard deviations.



a) $F_{5.1}$, $N = 30$, $\sigma_\epsilon = 1$, σSA

b) $F_{5.1}$, $N = 30$, $\sigma_\epsilon = 1$, CSA ($\sigma_{min} = 0.001$)

c) $F_{5.1}$, $N = 100$, $\sigma_\epsilon = 1$, σSA

d) $F_{5.1}$, $N = 100$, $\sigma_\epsilon = 1$, CSA ($\sigma_{min} = 0.001$)

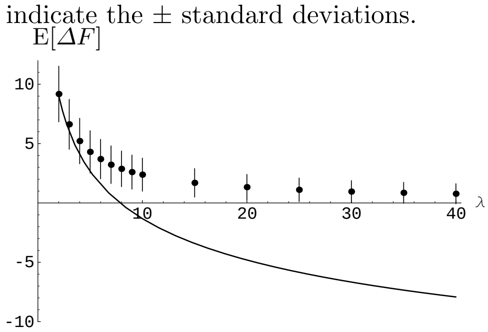
e) $F_{5.1}$, $N = 200$, $\sigma_\epsilon = 1$, σSA

f) $F_{5.1}$, $N = 200$, $\sigma_\epsilon = 1$, CSA, ($\sigma_{min} = 0.0025$)

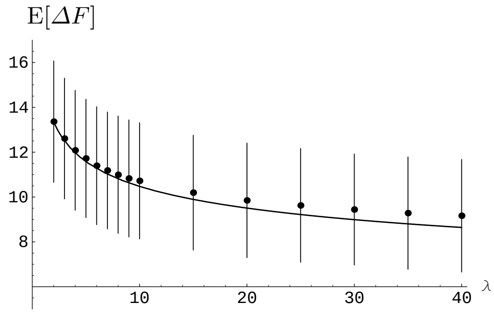
g) $F_{5.1}$, $N = 300$, $\sigma_\epsilon = 1$, σSA

h) $F_{5.1}$, $N = 300$, $\sigma_\epsilon = 1$, CSA, ($\sigma_{min} = 0.0025$)

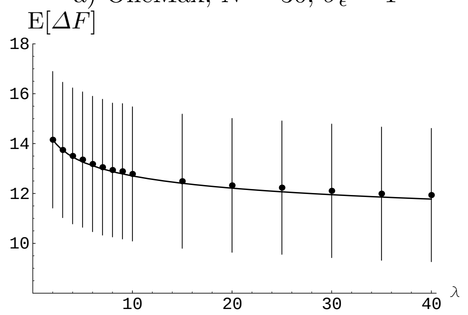
Fig. 3.24. Final fitness error for OneMax, depicting $\mathrm{E}[\Delta F]$ as a function of the offspring population size λ. Shown are the values obtained by (2.49) and the experimental data (points). The vertical bars indicate the ± standard deviations.



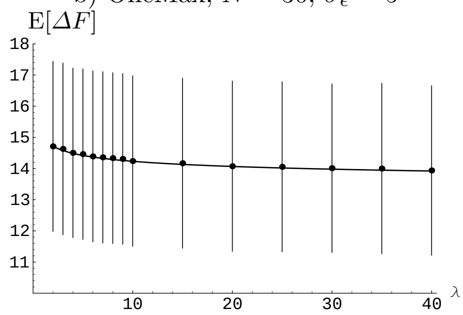
a) OneMax, $N = 30$, $\sigma_\epsilon = 1$



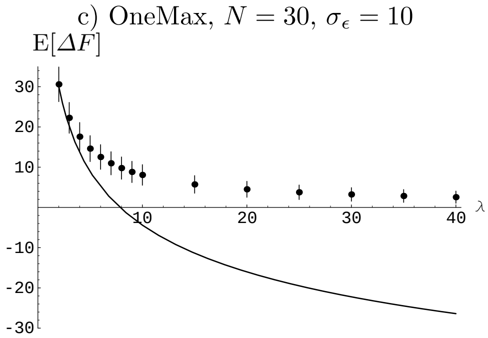
b) OneMax, $N = 30$, $\sigma_\epsilon = 5$



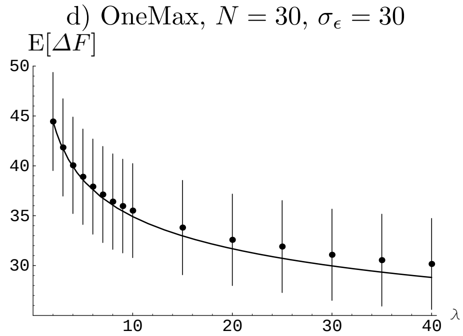
c) OneMax, $N = 30$, $\sigma_\epsilon = 10$



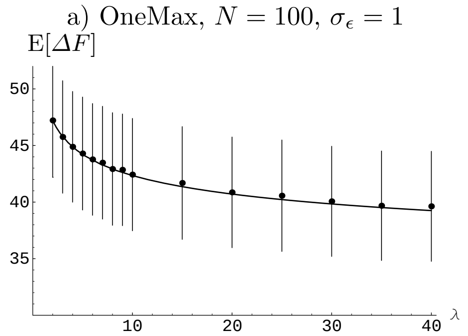
d) OneMax, $N = 30$, $\sigma_\epsilon = 30$



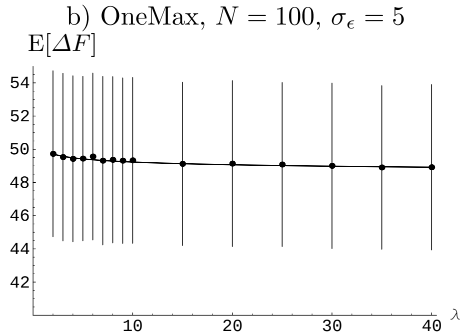
a) OneMax, $N = 100$, $\sigma_\epsilon = 1$



b) OneMax, $N = 100$, $\sigma_\epsilon = 5$



c) OneMax, $N = 100$, $\sigma_\epsilon = 10$



d) OneMax, $N = 100$, $\sigma_\epsilon = 100$

Chapter 4

Conclusions and Outlook

We have derived an approximate equation for the quality gain for arbitrary fitness functions and tested it on several fitness functions. Good results could be obtained for sufficiently small σ-values.

For many test functions, the quality gain consists of a gain and a loss part which in some cases can be interpreted as polynomials of σ. The gain part is generally of a lower degree than the loss term. Therefore, the quality gain increases first with the mutation strength – if the noise strength is sufficiently small – until the loss part gains more and more influence. The cubic functions are naturally an exception to this since they approach infinity and an increase of σ finally results in an increase of the quality gain.

While the quality gain is a local performance measure, it can be used to derive certain steady state conditions since the ES shows on average no progress if the steady state has been reached. Therefore, the quality gain becomes zero which was used to derive the final fitness error for several test function classes, i.e. quadratic, certain biquadratic functions, and OneMax. For the L_1-norm only a lower bound could be given. Other test functions still remain to be considered.

Except for the function OneMax the *equipartition assumption* was applied to arrive at the final formulae, which are formally very similar and simple expressions, generally given as $\mathrm{E}[\Delta F] = \sigma_\epsilon c N / c_{1,\lambda}$ where c is a constant that depends on the function class.

Only one greater deviation remained: The $(1,\lambda)$-σSA-ES tends to premature stagnation. Therefore, the final fitness error is considerably higher than it is predicted by the various equations. This behavior is caused by a fast reduction of the mutation strength. However, the reason why this occurs is not fully understood and will have to be investigated in the future.

One can use the equations describing the final fitness error for population sizing as it has been done in [4]. The equations predict the smallest final fitness error for a ratio of μ/λ around 0.5. This value, however, is not optimal as to the convergence velocity, for which $\mu/\lambda \approx 0.27$ is usually considered the optimal choice [7]. But since the curve of the final fitness error is generally quite flat in the interval $\mu/\lambda = 1/3$ to $\mu/\lambda = 2/3$ for the ES used, these values are also possible choices without leading a serious increase of the final fitness error.

The equations for the final fitness error were derived for $(1,\lambda)$ strategies. We considered the $(\mu/\mu_I,\lambda)$ variants by a simple modification of the respective equations. Although the prediction quality is quite good as we have seen, a formal derivation of the final fitness for $(\mu/\mu_I,\lambda)$ ES still needs to be developed.

In the case of the function OneMax, only $(1,\lambda)$-ES were considered. Equation (2.49) predicts the experimental data generally well but can only be used if the progress coefficient

$c_{1,\lambda}$ is sufficiently small in relation to the noise strength. For larger offspring population sizes, it tends to an underestimation of the final fitness error.

We applied the standard noise model assuming that the noise can be modeled simply by adding a normally distributed random variable to the fitness function. While this model is used in most publications there are several cases where it might not be applicable and other forms of modeling will have to be used. As an example, we refer to the optimization of aerodynamic structures (see [16]), where the noise has to be modeled inside the fitness function. The investigations will have to be extended to develop equations for these cases, which will require the application of novel techniques like the ones which were developed in [8].

Acknowledgements

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

Chapter 5

Appendix

5.1 Final Fitness Error of Function F_5

Inequality 1: $S_Q^2 \leq N\sigma^2 - 4\sigma^4 \sum_{i=1}^N \phi_{0,\sigma^2}^2(y_i)$
In order to prove the first inequality, we look at a single addend of S_Q^2, i.e. $\sigma^2 + 4y^2\Phi_{0,\sigma^2}(y)[1-\Phi_{0,\sigma^2}(y)] - 4\sigma^2\phi_{0,\sigma^2}(y)y[2\Phi_{0,\sigma^2}(y)-1] - 4\sigma^4\phi_{0,\sigma^2}^2(y)$, and show that

$$y^2\Phi_{0,\sigma^2}(y)[1-\Phi_{0,\sigma^2}(y)] \leq \sigma^2\phi_{0,\sigma^2}(y)y[2\Phi_{0,\sigma^2}(y)-1]. \tag{5.1}$$

Since the terms on both sides of the inequality are symmetric, it suffices to show the validity of (5.1) for $y \geq 0$. Using the transformation $z = \sigma y$, (5.1) simplifies to

$$z\Phi(z)(1-\Phi(z)) = z\Phi(z)\Phi(-z) \leq \phi(z)(2\Phi(z)-1), \tag{5.2}$$

where $\Phi(z) = \Phi_{0,1}(z)$ and $\phi(z) = \phi_{0,1}(z)$. Considering $f(z) := z\Phi(z)\Phi(-z) - \phi(z)(2\Phi(z)-1)$, we see that $f(0) = 0$ and that $f(z) \to 0$ for $z \to \infty$.

We need to show that $f(z) \leq 0$ for all $z \geq 0$. If f has only a minimum on $(0,\infty)$, the inequality is proven. The first derivative of f is given as $f'(z) = \Phi(z)\Phi(-z) - 2\phi^2(z) := h(z) - g(z)$ with $f'(0) < 0$ and $\lim_{z\to\infty} f'(z) = 0$. The second derivative is given by $f''(z) = h''(z) - g''(z) = -\phi(z)(2\Phi(z)-1) + 4z\phi^2(z)$. We will first consider the zero points of $f''(z)$, i.e. the solutions of

$$\begin{aligned} h''(z) = g''(z) &\iff \phi(z)(2\Phi(z)-1) = 4z\phi^2(z) \\ &\iff 2\Phi(z) - 1 = 4z\phi(z). \end{aligned} \tag{5.3}$$

Naturally, we have $f''(0) = 0$. Setting $k(z) = 2\Phi(z) - 1$ and $l(z) = 4z\phi(z)$, we note that k with $k'(z) = 2\phi(z)$ is monotonically increasing and approaches one. In contrast to this, l with $l'(z) = 4\phi(z)(1-z^2)$ grows until $z = 1$ and then decreases monotonically approaching zero. We also note that $l'(z) \geq k'(z)$ for $z \leq 1/\sqrt{2}$ whereas $l'(z) \leq k'(z)$ for $z \geq 1/\sqrt{2}$. Therefore, there is exactly one $z_0 > 1/\sqrt{2}$, with $l(z_0) = k(z_0)$ or $f''(z_0) = 0$. Since $f''(z) \geq 0$ for $z \leq z_0$ and $f''(z) \leq 0$ for $z_0 \leq z$, $f'(z)$ has its only maximum in z_0. We know that $f'(z_0) > 0$, because f' decreases monotonically towards zero for $z \geq z_0$. Since $f'(0) < 0$, there is exactly one $z_l \in (0, z_0)$ with $f'(z_l) = 0$. Therefore, f has only one extremum (minimum) in $(0,\infty)$ which is necessarily smaller than zero.

Equation 2: $|M_Q| = \sum_{i=1}^N y_i(2\Phi_{0,\sigma^2}(y_i) - 1) + 2\sigma^2\phi_{0,\sigma^2}(y_i) - |y_i|$
To show the validity of the second equation, it suffices to prove $y(2\Phi_{0,\sigma^2}(y) - 1) + 2\sigma^2\phi_{0,\sigma^2}(y) \geq |y|$. Since all terms are symmetric, we will only consider $y \geq 0$. Using the transformation

$z = \sigma y$ again, the inequality is given as $2z\Phi(z) + 2\phi(z) \geq 2z$. Considering the function $f(z) := z\Phi(z) + \phi(z) - z = z(\Phi(z) - 1) + \phi(z)$, we see that $f(0) > 0$ and $f(z) \to 0$ for $z \to \infty$. The derivative is given by $f'(z) = \Phi(z) - 1 \leq 0$. The function f is therefore monotonically decreasing approaching zero in the limit case which shows $f(z) \geq 0$ for all $z \geq 0$.

Inequality 3: $h(x) = \frac{\sqrt{x^4+x^2}-x^2}{1-\frac{2}{\pi\sqrt{2x^2+1}}} \geq 0.5$ for $x \geq 1$ which can be shown as follows.

$$\begin{aligned} h(x) &= \frac{\sqrt{x^4+x^2}-x^2}{1-\frac{2}{\pi\sqrt{2x^2+1}}} \geq 0.5 \\ &\Rightarrow \sqrt{x^4+x^2}-x^2 \geq \frac{1}{2}(1-\frac{2}{\pi\sqrt{2x^2+1}}) \\ &\Longleftrightarrow \sqrt{2x^2+1}(\sqrt{x^4+x^2}-x^2) \geq \frac{1}{2}(\sqrt{2x^2+1}-\frac{2}{\pi}) \\ &\Longleftrightarrow 2\sqrt{2x^6+3x^4+x^2} - 2x^2\sqrt{2x^2+1} - \sqrt{2x^2+1} \geq -\frac{2}{\pi} \\ &\Longleftrightarrow \sqrt{8x^6+12x^4+4x^2} - \sqrt{(2x^2+1)^3} \geq -\frac{2}{\pi} \\ &\Longleftrightarrow \sqrt{(2x^2+1)^3-(2x^2+1)} - \sqrt{(2x^2+1)^3} \geq -\frac{2}{\pi} \end{aligned} \tag{5.4}$$

Setting $t = 2x^2 + 1$, we consider $g(t) = \sqrt{t^3 - t} - \sqrt{t^3}$, $t \geq 3$. We will show that g is a monotonically increasing function. Since $g(3) > -2/\pi$ this will prove $h(x) \geq 0.5$ for all $x \geq 1$. The derivative is given as $g'(t) = \frac{3t^2-1}{2\sqrt{t^3-t}} - \frac{3t^2}{2\sqrt{t^3}}$. Setting $g'(t) = 0$ leads to

$$\begin{aligned} &\frac{3t^2-1}{\sqrt{t^3-t}} = \frac{3t^2}{\sqrt{t^3}} \\ \Rightarrow\ & (3t^2-1)\sqrt{t^3} = 3t^2\sqrt{t^3-t} \\ \Longleftrightarrow\ & (9t^4-6t^2+1)t^3 = 9t^4(t^3-t) \\ \Longleftrightarrow\ & 9t^4-6t^2+1 = 9t^4-9t^2 \ \text{ since } t \geq 3 \\ \Rightarrow\ & 3t^2+1 = 0 \end{aligned} \tag{5.5}$$

Therefore, $g'(t) \neq 0$ for all $t \geq 3$. Since $g'(3) \geq 0$, the function g is monotonically increasing which proves $h(x) \geq 0.5$ for all $x \geq 1$.

Alternative distribution for F_5: The behavior of function $h(x)$=$(x^2(2\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}} - 1)$ $+ \sqrt{\frac{2}{\pi}}x)/(1 - \frac{2}{\sqrt{\pi}x}\Phi(-\frac{1}{\sqrt{2}x})e^{\frac{1}{4x^2}})$ (3.34) will be considered in the following.

We will start with determining the limit value of $h(x)$ showing that $h(x) \to 1/2$ for $x \to \infty$. Since the denominator approaches 1 as it can easily be seen, it suffices to show that $x^2(2\Phi(-1/x)e^{\frac{1}{2x^2}} - 1) + x\sqrt{2/\pi} \to 1/2$.

As a first step, we calculate the limit value of $x(2\Phi(-1/x)e^{\frac{1}{2x^2}} - 1) + \sqrt{2/\pi}$.

We know that $2\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}} - 1) \to 0$ for $x \to \infty$. Writing $x(2\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}} - 1)$ as $(2\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}} - 1)/\frac{1}{x}$, we can apply l'Hospital's rule leading to

$$\frac{\sqrt{\frac{2}{\pi}}\frac{1}{x^2} - \frac{2}{x^3}\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}}}{-\frac{1}{x^2}} = -\sqrt{\frac{2}{\pi}} + \frac{2}{x}\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}} \tag{5.6}$$

which approaches $-\sqrt{2/\pi}$. Therefore, $x(2\Phi(-1/x)e^{\frac{1}{2x^2}}-1)+\sqrt{2/\pi}\to 0$ for $x\to\infty$.

We will now determine the limit value of the nominator. Setting $f(x)=x(2\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}}-1)+\sqrt{2/\pi}$ and $g(x)=1/x$, we have $f(x)\to 0$ and $g(x)\to 0$.

Applying l'Hospital's rule again, we get

$$\lim_{x\to\infty}\frac{f(x)}{g(x)}=\lim_{x\to\infty}\frac{f'(x)}{g'(x)}=\lim_{x\to\infty}\frac{2\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}}-1+\sqrt{\frac{2}{\pi}}\frac{1}{x}-\frac{2}{x^2}\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}}}{-\frac{1}{x^2}}. \tag{5.7}$$

Since we have again $f'(x)\to 0$ and $g'(x)\to 0$, we continue with the second derivative and arrive at

$$\lim_{x\to\infty}\frac{f'(x)}{g'(x)}=\lim_{x\to\infty}\frac{f''(x)}{g''(x)}=\lim_{x\to\infty}\frac{\frac{2}{x^3}\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}}-\sqrt{\frac{2}{\pi}}\frac{1}{x^4}+\frac{2}{x^5}\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}}}{\frac{2}{x^3}} \tag{5.8}$$

$$=\lim_{x\to\infty}\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}}-\frac{1}{x\sqrt{2\pi}}+\frac{1}{x^2}\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}}=\frac{1}{2}. \tag{5.9}$$

Therefore, $h(x)\to 1/2$ for $x\to\infty$.

It remains to be shown that $h(x)\geq 0.47715$ for all $x\geq 1/\sqrt{2}$. Setting $f(x)=x^2(2\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}}-1)+\sqrt{\frac{2}{\pi}}x$ and $g(x)=1-\frac{2}{\sqrt{\pi}x}\Phi(-\frac{1}{\sqrt{2}x})e^{\frac{1}{4x^2}}$, we know that $h(x)\geq f(x)$.

In the following, we show that f is a monotonic function on $[1/\sqrt{2},\infty)$. The derivative is given by $f'(x)=2x(2\Phi(-1/x)e^{\frac{1}{2x^2}}-1)-2/x\Phi(-1/x)e^{\frac{1}{2x^2}}+2\sqrt{2/\pi}$. Setting $f'(x)=0$, we get

$$2x(2\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}}-1)-\frac{2}{x}\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}}=-2\sqrt{\frac{2}{\pi}}\iff$$
$$1-2\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}}+\frac{1}{x^2}\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}}=\frac{1}{x}\sqrt{\frac{2}{\pi}}. \tag{5.10}$$

Setting $s=1/x$, we now consider the function $k(s)=1-2\Phi(-s)e^{\frac{s^2}{2}}+s^2\Phi(-s)e^{\frac{s^2}{2}}-s\sqrt{2/\pi}$ on $(0,\sqrt{2}]$. Since $k(0)=0$ and $k(\sqrt{2})<0$, we show that $k(s)<0$ for all $s\in(0,\sqrt{2}]$.

Let us consider the derivative of k, i.e. $k'(s)=s^2(s\Phi(-s)e^{\frac{s^2}{2}}-1/\sqrt{2\pi})$. Setting $k'(s)=0$, leads towards $s=0$ or $s\Phi(-s)e^{\frac{s^2}{2}}=1/\sqrt{2\pi}$. Let us assume in the following that there exists a $c\in(0,\sqrt{2})$ with $c\Phi(-c)e^{\frac{c^2}{2}}=1/\sqrt{2\pi}$. We have to show that $k(c)\leq 0$. Substituting c into k, we get $m(c)=1-2/(c\sqrt{2\pi})-c/\sqrt{2\pi}$ for all $c\in(0,\sqrt{2})$. Since $\lim_{c\to 0}m(c)<0$ and $\lim_{c\to\sqrt{2}}m(c)<0$, the proposition is proven if m' has no zero points on $(0,\sqrt{2})$. Since $m'(c)=2/(\sqrt{2\pi}c^2)-1/\sqrt{2\pi}$, we have $m'(c)=0$ for $c=\pm\sqrt{2}$. Therefore, $k(s)<0$ for all $s\in(0,\sqrt{2}]$ and $f'(x)\neq 0$ for all $x\in[1/\sqrt{2},\infty)$. Since $f'(1/\sqrt{2})>0$, the derivative is positive.

Since $f(x)\geq 0.47715$ for $x\geq 12$ and $h(1/\sqrt{2})>0.47715$, it only remains to verify $h(x)>0.47715$ for $x\in(1/\sqrt{2},12)$. Setting $c=0.47715$, let us consider $l(x)=f(x)-cg(x)$ for $x\in(1/\sqrt{2},12)$.

$$l(x)=x^2(2\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}}-1)+\sqrt{\frac{2}{\pi}}x-c+c\frac{2}{\sqrt{\pi}x}\Phi(-\frac{1}{\sqrt{2}x})e^{\frac{1}{4x^2}}\geq 0$$
$$\iff(2\Phi(-\frac{1}{x})e^{\frac{1}{2x^2}}-1)+\frac{\sqrt{2}}{\sqrt{\pi}x}-\frac{c}{x^2}+c\frac{2}{\sqrt{\pi}x^3}\Phi(-\frac{1}{\sqrt{2}x})e^{\frac{1}{4x^2}}\geq 0 \tag{5.11}$$

Setting $s = 1/x$, we now consider

$$h(s) = 2\Phi(-s)e^{\frac{s^2}{2}} - 1 + \frac{s\sqrt{2}}{\sqrt{\pi}} - s^2c + c\frac{2s^3}{\sqrt{\pi}}\Phi(-\frac{s}{\sqrt{2}})e^{\frac{s^2}{4}} \geq 0 \tag{5.12}$$

on $I = (1/12, \sqrt{2})$. We decompose I into smaller intervals, i.e. $I = \bigcup_k I_k$ and show that $h(s) \geq 0$ for all $s \in I_k = [b_k, b_{k+1}]$. Since $h(s)$ is given by

$$h(s) = h(b_k) + \int_{b_k}^{s} h'(t)\,\mathrm{d}t, \tag{5.13}$$

we will derive a lower bound for $h'(t)$ which can be determined as

$$h'(t) = t(2\Phi(-t)e^{\frac{t^2}{2}} + \frac{c}{\sqrt{\pi}}\Phi(-\frac{t}{\sqrt{2}})e^{\frac{t^2}{4}}(6t + t^3) - 2c - \frac{c}{\pi}t^2) = tu(t). \tag{5.14}$$

Since $\Phi(-t)e^{\frac{t^2}{2}}$ is a decreasing function, we have $\Phi(-t)e^{\frac{t^2}{2}} \geq \Phi(-b_{k+1})e^{\frac{b_{k+1}^2}{2}} = w_{k+1}$ and $\Phi(-\frac{t}{\sqrt{2}})e^{\frac{t^2}{4}} \geq \Phi(-\frac{b_{k+1}}{\sqrt{2}})e^{\frac{b_{k+1}^2}{4}} = v_{k+1}$. Thus, we obtain

$$u(t) \geq 2w_{k+1} + \frac{c}{\sqrt{\pi}}v_{k+1}(6t + t^3) - 2c - \frac{c}{\pi}t^2 \tag{5.15}$$

as a lower bound for u on I_k. In order to find a minimal value for that lower bound $\tilde{u}$ in I_k, we consider the possible zero points of $\tilde{u}'(t)$.

$$\begin{aligned} \tilde{u}'(t) &= \frac{c}{\sqrt{\pi}}v_{k+1}(6 + 3t^2) - \frac{2c}{\pi}t = 0 \\ &\iff t^2 + 2 - \frac{2}{3\sqrt{\pi}v_{k+1}}t = 0 \\ &\Rightarrow\ t_{1,2} = \frac{1}{3\sqrt{\pi}v_{k+1}} \pm \sqrt{\frac{1}{9\pi v_{k+1}^2} - 2} \end{aligned} \tag{5.16}$$

The root is only a real number if $1/(9\pi v_{k+1}^2) \geq 2$ or $1/(18\pi) \geq v_{k+1}^2$. The smallest possible value for v_{k+1} is $\Phi(-1)\sqrt{e}$ which is greater than $1/\sqrt{18\pi}$. Therefore, $\tilde{u}'$ does not have any zero points on I_k and $\tilde{u}$ is a monotone function. Depending on the sign of $\tilde{u}'$, $\tilde{u}$ assumes its minimum either at b_k or at b_{k+1}. Setting $c_k = \min_{t\in I_k} \tilde{u}(t)$, a lower bound for $h(s)$ is thus given by

$$h(s) \geq h(b_k) + \int_{b_k}^{s} tc_k\,\mathrm{d}t = h(b_k) + \frac{c_k}{2}(s^2 - b_k^2). \tag{5.17}$$

Depending on the sign of c_k, we finally obtain the lower bound

$$h(s) \geq h_k = \begin{cases} h(b_k) & \text{if } c_k > 0 \\ h(b_k) + \frac{c_k}{2}(b_{k+1}^2 - b_k^2) & \text{if } c_k < 0 \end{cases}. \tag{5.18}$$

The intervals chosen and the values of h_k are given in Table 5.1. All values were obtained by using Mathematica. As one can see, $h_k \geq 0$ for all k which proves $h(s) \geq 0$ on I.

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

The $c_{\mu/\mu,\lambda}$–coefficients used in the paper were obtained by using the generalized progress coefficients $e_{\mu,\lambda}^{\alpha,\beta}$ (see [6], p. 172). The coefficient $e_{\mu,\lambda}^{1,0}$ can be identified with $c_{\mu/\mu,\lambda}$ and determined numerically. Table 5.2 shows some selected $c_{\mu/\mu,\lambda}$–values used in this paper.

Table 5.1. The lower bounds h_k for the interval decomposition considered. Also shown are the values for the derivative of $\tilde{u}$ and the c_k-values.

b_k	b_{k+1}	$\tilde{u}'(b_k + (b_{k+1} - b_k)/2)$	c_k	h_k
0.05	0.1	0.743357	0.0084933	0.0000573665
0.1	0.15	0.7114	0.00939039	0.00022919
0.15	0.2	0.681914	0.00976756	0.000512478
0.2	0.25	0.654726	0.00967363	0.00090116
0.25	0.3	0.629675	0.00915364	0.0013865
0.3	0.35	0.606612	0.00824899	0.00195748
0.35	0.4	0.585401	0.00699779	0.00260111
0.4	0.45	0.565914	0.00543505	0.00330276
0.45	0.5	0.548037	0.00359291	0.00404641
0.5	0.55	0.531658	0.00150086	0.00481491
0.55	0.6	0.51668	-0.000814083	0.00556674
0.6	0.65	0.503009	-0.00332715	0.00624933
0.65	0.7	0.490558	-0.00601568	0.00688195
0.7	0.75	0.479249	-0.00885894	0.00744421
0.75	0.8	0.469008	-0.011838	0.00791558
0.8	0.85	0.459764	-0.0149355	0.00827546
0.85	0.9	0.451456	-0.0181358	0.00850329
0.9	0.95	0.444023	-0.0214243	0.00857861
0.95	1	0.437411	-0.0247881	0.00848115
1	1.05	0.431568	-0.0282151	0.00819085
1.05	1.1	0.426447	-0.0316945	0.00768791
1.1	1.05	0.422003	-0.0352165	0.00695286
1.15	1.2	0.418195	-0.038772	0.00596657
1.2	1.25	0.414985	-0.0423531	0.00471029
1.25	1.3	0.412336	-0.0459523	0.00316564
1.3	1.35	0.410214	-0.049563	0.00131466
1.35	1.37	0.413664	-0.038805	0.0017404
1.37	1.39	0.413097	-0.0402718	0.00087403
1.39	1.4	0.41428	-0.0368903	0.000607786
1.4	1.41	0.414053	-0.0376248	0.000142087
1.41	1.414	0.414844	-0.03543814214607244	$5.27635490428548 * 10^{-6}$
1.414	$\sqrt{2}$	0.415395	-0.0338832	$5.246888030107535 * 10^{-6}$

Table 5.2. Some $c_{\mu/\mu,\lambda}$–coefficients

μ	1	2	4	6	10	15	20	25	30
$c_{\mu/\mu,60}$	2.319	2.127	1.882	1.713	1.472	1.252	1.076	0.924	0.788
μ	35	40	45	50	54	56	58	59	
$c_{\mu/\mu,60}$	0.660	0.538	0.417	0.294	0.190	0.134	0.073	0.039	

5.3 Description of the ES Used

For the simulation of the dynamic behavior of $(\mu/\mu_I, \lambda)$–ES under systematic noise, the ES has to control the endogenous strategy parameter σ. We used a standard approach in this paper: the σ self–adaptation (see [17], [18], or [9]). Due to the findings in [7] we we widely refrained from using the cumulative step size adaptation (see [13] and [14]) only considering it as an alternative in the case of the L_1-norm. To prevent premature convergence, we introduced a minimal mutation strength.

The σ self-adaptation makes use of a coupled inheritance of object and strategy parameters. For a parameter $\mathbf{a}$, we denote with $\langle \mathbf{a} \rangle^{(g)}$ the centroid of the μ best offspring, i.e.

$$\langle \mathbf{a} \rangle^{(g)} := \frac{1}{\mu} \sum_{m=1}^{\mu} \mathbf{a}_{m;\lambda}^{(g)} \tag{5.19}$$

The algorithm can be expressed as

$$\forall l = 1, \ldots, \lambda : \begin{cases} \sigma_l^{(g+1)} := \langle \sigma \rangle^{(g)} \mathrm{e}^{\tau \mathcal{N}(0,1)} \\ \mathbf{y}_l^{(g+1)} := \langle \mathbf{y} \rangle^{(g)} + \sigma_l^{(g+1)} \mathcal{N}(\mathbf{0}, \mathbf{1}), \end{cases} \tag{5.20}$$

where τ is the so–called learning parameter. For the simulations $\tau = \frac{1}{\sqrt{N}}$ was chosen. As can be seen in (5.19) the mutation strengths σ_l are calculated for each offspring l independently. This σ_l is then used as the mutation parameter to obtain the corresponding object parameter.

In contrast to the σSA-algorithm, the cumulative step-size adaptation [12, 14, 1] uses a single mutation strength parameter σ per generation to generate the offspring population. The parameter is changed by a deterministic rule which is controlled by certain statistics gathered over the course of generations. The update rule reads

$$\left. \begin{array}{l} \forall l = 1, \ldots, \lambda : \mathbf{y}_l^{(g+1)} := \langle \mathbf{y} \rangle^{(g)} + \sigma^{(g)} \mathbf{z}_l, \text{ where } \mathbf{z}_l \sim \mathcal{N}(0,1) \\ \mathbf{s}^{(g+1)} := (1-c)\mathbf{s}^{(g)} + \sqrt{(2-c)c}\sqrt{\mu} \left(\langle \mathbf{z} \rangle^{(g+1)} \right) \\ \sigma^{(g+1)} := \sigma^{(g)} \exp \left(\frac{||\mathbf{s}^{(g+1)}||^2 - N}{2DN} \right) \end{array} \right\} . \tag{5.21}$$

The initial value $\mathbf{s}^{(0)}$ was chosen as $\mathbf{s}^{(0)} = 0$. The recommended and used standard settings for the cumulation parameter c and the damping constant D are $c = \frac{1}{\sqrt{N}}$ resp. $D = \sqrt{N}$. The term N in the numerator equals the mean squared length of the progress vector if consecutive steps are independend. As in [1], the algorithm was slightly changed by taking the the squared length of the accumulation vector instead of the length as in [12].

Bibliography

[1] D. V. Arnold. *Noisy Optimization with Evolution Strategies.* Kluwer Academic Publishers, Dordrecht, 2002.

[2] D. V. Arnold and H.-G. Beyer. Investigation of the (μ, λ)-ES in the Presence of Noise. In *Proceedings of the CEC'01 Conference*, pages 332–339, Piscataway, NJ, 2001. IEEE.

[3] D. V. Arnold and H.-G. Beyer. Local Performance of the $(1 + 1)$-ES in a Noisy Environment. *IEEE Transactions on Evolutionary Computation*, 6(1):30–41, 2002.

[4] D. V. Arnold and H.-G. Beyer. On the Benefits of Populations for Noisy Optimization. *Evolutionary Computation*, 11(2):111–127, 2003.

[5] 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.

[6] H.-G. Beyer. *The Theory of Evolution Strategies.* Natural Computing Series. Springer, Heidelberg, 2001.

[7] 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. et al. Cantú-Paz, editor, *GECCO-2003: Proceedings of the Genetic and Evolutionary Computation Conference*, pages 525–536, Berlin, Germany, 2003. Springer.

[8] 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.

[9] H.-G. Beyer and H.-P. Schwefel. Evolution Strategies: A Comprehensive Introduction. *Natural Computing*, 1(1):3–52, 2002.

[10] J. Branke. *Evolutionary Optimization in Dynamic Environments.* Kluwer Academic Publishers, Dordrecht, 2001.

[11] J.M. Fitzpatrick and J.J. Grefenstette. Genetic Algorithms in Noisy Environments. In P. Langley, editor, *Machine Learning: Special Issue on Genetic Algorithms*, volume 3, pages 101–120. Kluwer Academic Publishers, Dordrecht, 1988.

[12] N. Hansen. *Verallgemeinerte individuelle Schrittweitenregelung in der Evolutionsstrategie.* Doctoral thesis, Technical University of Berlin, Berlin, 1998.

[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. Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation*, 9(2):159–195, 2001.

[15] B. L. Miller. *Noise, Sampling, and Efficient Genetic Algorithms*. PhD thesis, University of Illinois at Urbana-Champaign, Urbana, IL 61801, 1997. IlliGAL Report No. 97001.

[16] M. Olhofer, T. Arima, T. Sonoda, M. Fischer, and B. Sendhoff. Aerodynamic Shape Optimisation Using Evolution Strategies. In I. Parmee and P. Hajela, editors, *Optimisation in Industry*, pages 83–94. Springer Verlag, 2002.

[17] I. Rechenberg. *Evolutionsstrategie '94*. Frommann-Holzboog Verlag, Stuttgart, 1994.

[18] H.-P. Schwefel. *Numerical Optimization of Computer Models*. Wiley, Chichester, 1981.