Local Performance of the $(1+1)$-ES in a Noisy Environment

Dirk V. Arnold and Hans-Georg Beyer
Department of Computing Science XI
University of Dortmund
44221 Dortmund, Germany
{arnold,beyer}@ls11.cs.uni-dortmund.de

Abstract

While noise is a phenomenon present in many real-world optimization problems, the understanding of its potential effects on the performance of evolutionary algorithms is still incomplete. This paper investigates the effects of noise for the infinite-dimensional quadratic sphere and a $(1+1)$-ES with isotropic normal mutations. It is shown that overvaluation as a result of failure to reevaluate parental fitness leads to both reduced success probabilities and improved performance. Implications for mutation strength adaptation rules are discussed and optimal resampling rates are computed.

Keywords: Evolution strategy, noise, local performance, overvaluation

1 Introduction

Noise is present in many real-world optimization problems. It can stem from sources as different as measurement limitations, stochastic simulation procedures, and user input. It is important to understand and predict the effects of noise so as to be able to come to conclusions regarding good parameter settings of optimization algorithms and to compare the performance of different strategies.

A number of results regarding the influence of noise on the performance of evolutionary optimization algorithms have been published. In the realm of genetic algorithms, Fitzpatrick and Grefenstette [8] present empirical results and derive recommendations regarding resampling and population sizing. Miller and Goldberg [10] consider the fitness dynamics of the OneMax function and introduce a population sizing model which includes fitness noise. Rattray and Shapiro [13] investigate finite sampling size effects. Angeline [1] has studied the effects of noise on self-adaptive evolutionary programming. In the realm of evolution strategies, theoretical results regarding single-parent strategies have been published by Beyer [3, 6, 7] and by Rechenberg [15]. Some empirical results involving multi-parent strategies have been reported, among others, by Rechenberg [15] and by Nissen and Propach [11]. Arnold [2] and Beyer [7] both offer more extensive overviews of this research and pointers to related publications.

The focus of the present paper is on the local performance of $(1+1)$-evolution strategies (ES) operating in real-valued search spaces. The central question is what effects can occur if noisy fitness values can survive for more than a single generation. The references the most closely related are [3] and [15]. In the former, Beyer presents an analysis of the effects of noise on the local performance of $(1 \overset{+}{,} \lambda)$-ES in spherically symmetric fitness environments. In the latter, Rechenberg extends these results to a wider class of quadratic fitness functions. In both cases, the investigated algorithms differ from the one analyzed in what follows in that a noisy fitness value cannot persist for more than a single generation. For comma-strategies, this is a simple consequence of the fact that an individual cannot survive for more than a single generation. For plus-strategies, the analysis of Beyer presupposes that the fitness of the parent individual is reevaluated in every generation. However, as reevaluation of the parental fitness increases the computational costs it is likely not to be part of many ES implementations.

While this difference may at first appear minor, its effects on the performance of ES are considerable. The results presented in what follows show that failure to reevaluate the parental fitness leads to systematic overvaluation and in turn to reduced success probabilities and long periods of stagnation. However, comparing the local performance of the two strategies, it will turn out that reevaluating the parent not only increases the computational costs per generation, but also that the strategy which does not reevaluate the parent achieves a greater expected fitness gain per generation and prevents the deterioration of the quality of the solution for high noise strengths. On the downside, it has to be noted that failure to reevaluate the parental fitness can render a commonly employed success probability based mutation strength adaptation rule useless in the presence of noise.

This paper is organized as follows. Section 2 describes the (1+1)-ES with isotropic normal mutations and outlines the fitness environment for which the algorithm's performance is to be analyzed. Section 3 introduces the degree of overvaluation as a variable of critical influence on the performance of the ES. In Section 4, success probability and expected fitness gain are computed and the performance in case of optimally adjusted mutation strength is analyzed. The beneficial influence of resampling is addressed and the problem of mutation strength adaptation is discussed. Section 5 concludes with a summary and suggestions for future work.

2 Algorithm and Fitness Environment

Many of the problems considered in theoretical studies of ES are single-criterion optimization problems in $I\!R^n$, where typically n is large. In that case, without loss of generality it can be assumed that the task at hand is minimization of a fitness function $f : I\!R^n \to I\!R$. Note that high values of f correspond to low fitness and vice versa. Section 2.1 introduces the (1+1)-ES algorithm with isotropic normal mutations. Section 2.2 outlines the fitness environment for which the performance of the algorithm is analyzed in the succeeding sections.

2.1 The Algorithm

Using a $(1+1)$-ES, a parent with parameter space location $\mathbf{y} \in I\!R^n$ generates offspring at location $\mathbf{y}+\mathbf{z}$ that replaces the parent in the next generation if and only if it is superior in terms of fitness. Using isotropic normal mutations, mutation vector $\mathbf{z}$ is a random vector consisting of n independent normally distributed components with mean 0 and variance σ^2. That is, $\mathbf{z}$ has probability density

$$p_z(\mathbf{z}) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} \exp\left(-\frac{1}{2}\frac{\mathbf{z}^{\mathrm{T}}\mathbf{z}}{\sigma^2}\right).$$

The standard deviation σ of the components of the mutation vector is referred to as the *mutation strength*.

As the decision whether an offspring survives is made solely on the basis of how its fitness compares with that of its parent, the *fitness advantage*

$$q_{\mathbf{y}}(\mathbf{z}) = f(\mathbf{y}) - f(\mathbf{y}+\mathbf{z})$$

is an important quantity to consider. For fixed parent, the fitness advantage is a scalar random variate with a probability density $p_{q_{\mathbf{y}}}$ that can in some simple cases be given explicitly. In particular, in some fitness environments the distribution of $q_{\mathbf{y}}$ tends to normality as n tends to infinity. Beyer [4] has presented an approximation for $p_{q_{\mathbf{y}}}$ for arbitrary quadratic fitness functions which is based on Gram-Charlier expansions of probability distributions.

In Section 4, the expected fitness advantage after selection is used as measure for the local performance of the algorithm. Frequently, local performance is measured in parameter space rather than in fitness space. The corresponding performance measure is usually referred to as the *progress rate*. Technically, in many of the investigated environments that has the advantage that the tendency to normality of the offspring distribution is stronger, yielding better approximations for finite n. However, for the fitness environment introduced in Section 2.2 and using appropriate normalizations, in the limit $n \to \infty$ the corresponding performance measures agree exactly.

In what follows it is assumed that there is noise involved in the process of the evaluation of the fitness function. This form of noise has therefore been termed *fitness noise*. Loosely speaking, fitness noise deceives the selection mechanism. An individual at parameter space location $\mathbf{x}$ has an *ideal fitness* $f(\mathbf{x})$ and a *perceived fitness* which may differ from its ideal fitness. As a consequence, it is possible that an offspring that is superior in terms of ideal fitness is discarded instead of being selected to replace the parent if its perceived fitness compares unfavorably with that of the parent. Conversely, an offspring that is inferior in terms of ideal fitness may be selected to replace the parent due to a higher perceived fitness.

Fitness noise is commonly modeled by means of an additive, normally distributed random term with mean zero. That is, in a noisy environment, evaluation of the fitness function at parameter space location $\mathbf{x}$ yields perceived fitness $f(\mathbf{x}) + \sigma_\epsilon(\mathbf{x})N$, where N is a standard normally distributed random variable. Quite naturally, $\sigma_\epsilon(\mathbf{x})$ is referred to as the *noise strength*.

With selection based on perceived fitness rather than on ideal fitness, a parent with a perceived fitness that is much higher than its ideal fitness is likely to persist for many generations. Conversely, an offspring with a perceived fitness that is lower than its ideal fitness is likely not to be accepted at all even though its ideal fitness may compare favorably with that of its parent. As a consequence, after several iterations of the mutation-selection mechanism, the perceived fitness of the parent is likely to be higher than its ideal fitness. The discrepancy between ideal fitness and perceived fitness of the parent is in what follows referred to as the *degree of overvaluation* Ξ. Likewise, the discrepancy between ideal fitness and perceived fitness of an offspring is denoted by ξ. Due to the assumption of a normally distributed noise term with mean 0 and variance $\sigma_\epsilon^2(\mathbf{x})$, the probability distribution of the degree of overvaluation of an offspring at parameter space location $\mathbf{x}$ is normal with mean 0 and variance $\sigma_\epsilon^2(\mathbf{x})$. The degree of overvaluation of the parent cannot as easily be given as it is a result of the repeated application of the mutation-selection mechanism. An approximation of its probability distribution in the simple environment to be described in Section 2.2 will be derived in Section 3.1.

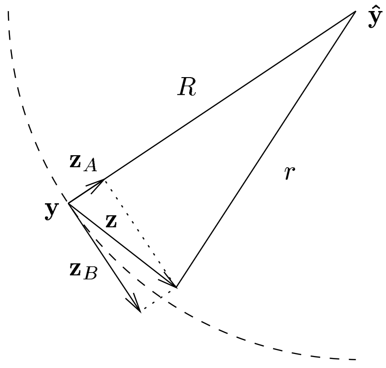

Figure 1: *Decomposition of a mutation vector* $\mathbf{z}$ *into two vectors* $\mathbf{z}_A$ *and* $\mathbf{z}_B$. *Vector* $\mathbf{z}_A$ *is parallel to* $\hat{\mathbf{y}} - \mathbf{y}$, *vector* $\mathbf{z}_B$ *is in the plane perpendicular to that.*

2.2 The Fitness Environment

Computing the expected fitness gain or the degree of overvaluation is a hopeless task for all but the most simple fitness functions. In ES theory, a repertoire of fitness functions simple enough to be amenable to mathematical analysis while still interesting enough to yield non-trivial results has been established. The most commonplace of these fitness functions is the quadratic sphere

$$f(\mathbf{y}) = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \tag{1}$$

which simply maps vectors $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$ in $I\!R^n$ to the square of their Euclidean distance to the optimum at location $\hat{\mathbf{y}} = (\hat{y}_1, \ldots, \hat{y}_n)^{\mathrm{T}}$. By formally letting the parameter space dimension n tend to infinity, assuming a particular dependence of the noise strength on parameter space location, and introducing appropriate normalizations as defined below, a number of simplifying conditions hold true. While admittedly being very simple, this fitness environment allows for a partially analytical treatment of the behavior of the $(1+1)$-ES and therefore contributes to its better understanding. In fact, systematic overvaluation is a phenomenon resulting from the failure to reevaluate the parental fitness, and it can be expected that qualitatively similar effects can be observed in other fitness environments as well as for other strategies if noisy fitness values can survive for more than a single generation.

Following an idea put forward in both [3] and [15], a mutation vector $\mathbf{z}$ can be written as the sum of two vectors $\mathbf{z}_A$ and $\mathbf{z}_B$, where $\mathbf{z}_A$ is parallel to $\hat{\mathbf{y}} - \mathbf{y}$ and $\mathbf{z}_B$ is in the plane perpendicular to that. Figure 1 illustrates this decomposition. Due to the isotropy of mutations, it can without loss of generality be assumed that $\mathbf{z}_A = (z_1, 0, \ldots, 0)^{\mathrm{T}}$ and $\mathbf{z}_B = (0, z_2, \ldots, z_n)^{\mathrm{T}}$, where the z_i are independent normally distributed random variables with mean 0 and variance σ^2.

Denoting the distance of parent and offspring to the location of the optimum by R and r, respectively, and using elementary geometry, from Figure 1 it can be seen that

$$\begin{aligned} r^2 &= (R - z_1)^2 + \|\mathbf{z}_B\|^2 \\ &= R^2 - 2Rz_1 + z_1^2 + \|\mathbf{z}_B\|^2. \end{aligned}$$

The distribution of z_1 is normal with mean 0 and variance σ^2. As $\|\mathbf{z}_B\|^2/\sigma^2$ is the sum of squares of $n-1$ independent standard normally distributed random variables, it is χ^2_{n-1}-distributed and thus has mean $n-1$ and variance $2(n-1)$. For $n \to \infty$, $\|\mathbf{z}_B\|^2$ therefore converges to $n\sigma^2$. Furthermore, z_1^2 can be neglected compared to $\|\mathbf{z}_B\|^2$. Consequently, the fitness advantage of the offspring in the limit $n \to \infty$ is

$$\begin{aligned} q_R &= R^2 - r^2 \\ &= -n\sigma^2 + 2Rz_1. \end{aligned}$$

Note that the index $\mathbf{y}$ indicating the parameter space location of the parent has been replaced by the parental distance to the optimum R as the exact location is irrelevant due to the spherical symmetry of the fitness function and the isotropy of mutations.

Obviously, q_R is normally distributed with mean $-n\sigma^2$ and variance $4\sigma^2R^2$. Introducing normalized mutation strength

$$\sigma^* = \sigma\frac{n}{R}, \tag{2}$$

the normalized fitness advantage of an offspring

$$q^* = q_R\frac{n}{2R^2}$$

is a normally distributed random variable with mean $E_{q^*} = -\sigma^{*2}/2$ and variance $D^2_{q^*} = \sigma^{*2}$ and therefore with probability density

$$p_{q^*}(x) = \frac{1}{\sqrt{2\pi}D_{q^*}}\exp\left(-\frac{1}{2}\left(\frac{x-E_{q^*}}{D_{q^*}}\right)^2\right). \tag{3}$$

Note that as the distribution of q^* is independent of the parental distance to the optimum, it has been possible to drop the index R. Assuming the existence of a mechanism that ensures that the normalized mutation strength σ^* is constant, the distribution of q^* is therefore constant throughout the optimization process. The problem of mutation strength adaptation is discussed in Section 4.4.3.

To preserve the homogeneity in the presence of noise, it is assumed that the noise strength scales with fitness such that the normalized noise strength

$$\sigma^*_\epsilon = \sigma_\epsilon\frac{n}{2R^2}$$

is constant throughout the parameter space. For finite normalized mutation strength and $n \to \infty$, $(R^2-r^2)/2R^2$ equals q^*/n and therefore tends to zero. Therefore, it can be assumed that the noise strength σ_ϵ at the location of an offspring is the same as that at the location of its parent. This fact significantly simplifies the theoretical analysis of the progress behavior.

Letting S^2 and s^2 denote the perceived fitness of parent and offspring, respectively, the offspring is accepted if and only if $s^2 < S^2$. For the present environment, the degrees of overvaluation of parent and offspring are $\Xi = R^2 - S^2$ and $\xi = r^2 - s^2$, respectively. Introducing normalizations

$$\Xi^* = \Xi\frac{n}{2R^2} \quad \text{and} \quad \xi^* = \xi\frac{n}{2R^2},$$

it follows that the degree of normalized overvaluation of the offspring is normally distributed with mean 0 and variance ${\sigma_\epsilon^*}^2$ and thus with probability density

$$p_{\xi^*}(x) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon^*} \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma_\epsilon^*}\right)^2\right). \tag{4}$$

Note that like the probability density of q^*, that of ξ^* is independent of the location in parameter space.

3 Overvaluation

Due to the homogeneity of the environment, the probability distributions of the normalized fitness advantage and of the degree of normalized overvaluation of an offspring are constant throughout the optimization process. The distribution of the degree of normalized overvaluation of the parent converges to a stable limit distribution. It is the goal of Section 3.1 to find an approximation for this distribution. In Section 3.2, several issues related to that distribution are discussed.

3.1 Obtaining the Distribution

An approximation to the limit distribution of the degree of normalized overvaluation Ξ^* can be obtained by an approach that has previously been employed by Beyer [5] to approximate the population distribution for a (μ, λ)-ES. The approach consists in expressing the unknown distribution in terms of its Gram-Charlier expansion with initially unknown cumulants. Then, the resulting cumulants of the distribution after application of the mutation and selection operators are computed. Finally, self-consistency conditions are imposed by demanding that the cumulants that define the shape of the distribution do not change from one generation to the next. In principle, any degree of accuracy can be achieved by considering sufficiently many terms in the expansion. In practice, for the problem of determining the distribution of the degree of normalized overvaluation, neglecting all but the first two cumulants will turn out to yield good results already. Effectively, this amounts to using a normal approximation for the limit distribution of Ξ^*.

Under the normal approximation, at time t, where t is sufficiently large for the distribution to have converged, the degree of overvaluation Ξ^* has probability density

$$p_{\Xi^*}(x) = \frac{1}{\sqrt{2\pi}D_{\Xi^*}} \exp\left(-\frac{1}{2}\left(\frac{x - E_{\Xi^*}}{D_{\Xi^*}}\right)^2\right), \tag{5}$$

where mean E_{Ξ^*} and standard deviation D_{Ξ^*} of the distribution remain to be determined.

Let $P^{(acc)}_{\xi^*|\Xi^*}(y|x)$ denote the chance of an offspring with degree of normalized overvaluation $\xi^* = y$ being accepted given a parent with degree of normalized overvaluation $\Xi^* = x$. Using the fact that q^* is normally distributed with mean E_{q^*} and variance $D^2_{q^*}$, this probability can be computed as

$$\begin{aligned}
P^{(acc)}_{\xi^*|\Xi^*}(y|x) &= \mathrm{Prob}[s^2 < S^2] \\
&= \mathrm{Prob}[r^2 - \xi < R^2 - \Xi] \\
&= \mathrm{Prob}[x - y < q^*] \\
&= \Phi\left(\frac{y - x + E_{q^*}}{D_{q^*}}\right),
\end{aligned} \tag{6}$$

where Φ stands for the probability distribution function of the standard normal distribution.

At time $t+1$, the degree of normalized overvaluation Ξ^*_+ equals Ξ^* if the offspring has been rejected, and it equals ξ^* if the offspring has been accepted. Thus, the degree of normalized overvaluation at time $t+1$ is in the interval $[x, x+\mathrm{d}x]$ if either the degree of normalized overvaluation of the offspring is in that interval and the offspring is accepted, or if the degree of normalized overvaluation of the parent is in that interval and the offspring is rejected. Therefore, the probability density of Ξ^*_+ can be given by the Chapman-Kolmogorov equation of the system as

$$p_{\Xi^*_+}(x) = \int_{-\infty}^{\infty} p_{\Xi^*}(y) p_{\xi^*}(x) P^{(acc)}_{\xi^*|\Xi^*}(x|y)\mathrm{d}y + \int_{-\infty}^{\infty} p_{\Xi^*}(x) p_{\xi^*}(y) \left[1 - P^{(acc)}_{\xi^*|\Xi^*}(y|x)\right] \mathrm{d}y. \tag{7}$$

Probability densities p_{Ξ^*} and p_{ξ^*} are given by Equations (5) and (4), respectively. Finally, self-consistency demands that

$$\mathcal{E}\left[\Xi^*_+\right] = \int_{-\infty}^{\infty} x p_{\Xi^*_+}(x)\mathrm{d}x = E_{\Xi^*} \tag{8}$$

and

$$\mathcal{D}^2\left[\Xi^*_+\right] = \int_{-\infty}^{\infty} x^2 p_{\Xi^*_+}(x)\mathrm{d}x - \left(\mathcal{E}\left[\Xi^*_+\right]\right)^2 = D^2_{\Xi^*}. \tag{9}$$

Equations (8) and (9) form a system of two non-linear equations that need to be solved for the two unknowns E_{Ξ^*} and $D^2_{\Xi^*}$.

Computation of the probability density $p_{\Xi^*_+}$ and of the mean and variance of Ξ^*_+ is cumbersome but straightforward. Detailed derivations of the results can be found in Appendices C.2, C.3, and C.4. Letting $\chi^{(k)}$ denote the kth derivative with respect to E_{Ξ^*} of

$$\chi\left(E_{\Xi^*}, D^2_{\Xi^*}\right) = \Phi\left(\frac{E_{\Xi^*} - E_{q^*}}{\sqrt{D^2_{q^*} + \sigma^{*2}_\epsilon + D^2_{\Xi^*}}}\right), \tag{10}$$

then, introducing symbols g_1 and g_2 for future reference, mean and variance of Ξ^*_+ are according to Equations (40) and (41)

$$\begin{aligned}
\mathcal{E}\left[\Xi^*_+\right] &= E_{\Xi^*}\chi^{(0)}\left(E_{\Xi^*}, D^2_{\Xi^*}\right) + \left(\sigma^{*2}_\epsilon + D^2_{\Xi^*}\right)\chi^{(1)}\left(E_{\Xi^*}, D^2_{\Xi^*}\right) \\
&= g_1\left(E_{\Xi^*}, D^2_{\Xi^*}\right)
\end{aligned} \tag{11}$$

and

$$\begin{aligned}
\mathcal{D}^2\left[\Xi^*_+\right] &= \left(\sigma^{*2}_\epsilon - D^2_{\Xi^*}\right)\left(1 - \frac{\partial g_1}{\partial E_{\Xi^*}}\right) + g_1\left(E_{\Xi^*}, D^2_{\Xi^*}\right)\left(E_{\Xi^*} - g_1\left(E_{\Xi^*}, D^2_{\Xi^*}\right)\right) + D^2_{\Xi^*}, \\
&= g_2\left(E_{\Xi^*}, D^2_{\Xi^*}\right),
\end{aligned} \tag{12}$$

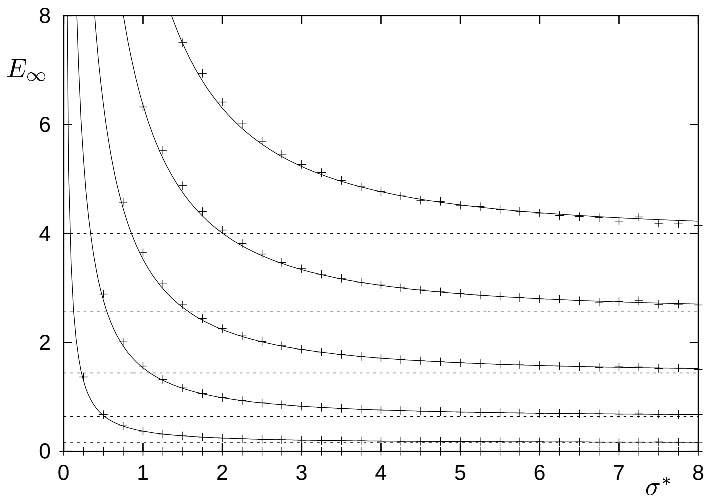

Figure 2: *Mean E_∞ of the degree of normalized overvaluation as a function of normalized mutation strength σ^* for the infinite-dimensional quadratic sphere. The solid curves display, from top to bottom, the results of numerical root finding of Equation* (15) *for normalized noise strengths $\sigma^*_\epsilon = 2.0$, 1.6, 1.2, 0.8, 0.4, and 0.0. The crosses represent data measured in experiments described in Appendix A. The dashed lines indicate the limit values for high normalized mutation strength derived in Section 3.2.1.*

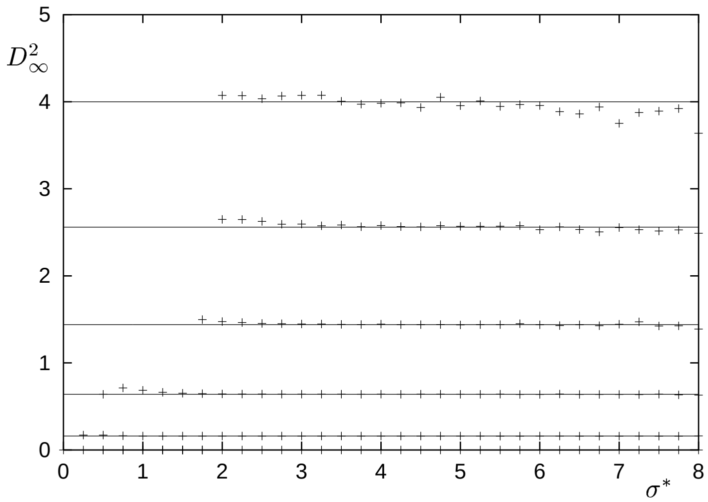

Figure 3: *Variance D^2_∞ of the degree of normalized overvaluation as a function of normalized mutation strength σ^* for the infinite-dimensional quadratic sphere. The curves display, from top to bottom, the results from Equation* (14) *for normalized noise strengths $\sigma^*_\epsilon = 2.0$, 1.6, 1.2, 0.8, 0.4, and 0.0. The crosses represent data measured in experiments described in Appendix A.*

respectively. Partial derivatives of χ are computed in Appendix C.1.

Equations (11) and (12) form a two-dimensional iterated map describing the change from one generation to the next of the macroscopic parameters E_{Ξ^*} and $D^2_{\Xi^*}$ determining the distribution of the degree of normalized overvaluation. After sufficiently many generations, convergence to a fixed point (E_∞, D^2_∞) can be observed. To find the fixed point, inserting Equation (12) into Equation (9) while making use of Equation (8) yields condition

$$\left(\sigma_\epsilon^{*2} - D^2_\infty\right)\left(1 - \left.\frac{\partial g_1}{\partial E_{\Xi^*}}\right|_{(E_\infty, D^2_\infty)}\right) = 0 \tag{13}$$

Clearly, Equation (13) is satisfied for

$$D^2_\infty = \sigma_\epsilon^{*2}. \tag{14}$$

Figure 3 shows a good agreement between Equation (14) and computer experiments described in Appendix A. Inserting Equation (11) into Equation (8) and using Equation (14), the other condition reads

$$\chi^{(1)}\left(E_\infty, \sigma_\epsilon^{*2}\right) = \frac{E_\infty}{2\sigma_\epsilon^{*2}}\left[1 - \chi^{(0)}\left(E_\infty, \sigma_\epsilon^{*2}\right)\right]. \tag{15}$$

The result of numerical root finding for E_∞ is displayed in Figure 2. Again, a good agreement between theory and experiments can be observed.

3.2 Limit Behavior, Stability, and Convergence Rates

This section discusses several issues related to the limit distribution of the degree of normalized overvaluation found in the previous section. In particular, Section 3.2.1 analyzes the behavior of the mean degree of normalized overvaluation in the limit of very large normalized mutation strength. In Section 3.2.2, the stability of the limit distribution is investigated. Finally, in Section 3.2.3 the rate of convergence with which the limit distribution is approached is studied.

3.2.1 Limit behavior

For the infinite-dimensional quadratic sphere and finite normalized noise strength σ_ϵ^*, the expected degree of normalized overvaluation E_∞ approaches a finite constant value as the normalized mutation strength σ^* approaches infinity. The limit value can be computed by applying de l'Hôspital's rule and using the identities listed in Appendix C.1 in combination with $\lim_{\sigma^*\to\infty} \partial E_\infty/\partial\sigma^* = 0$ to obtain

$$\begin{aligned}
\lim_{\sigma^*\to\infty} & \frac{\chi^{(1)}\left(E_\infty, D^2_\infty\right)}{1 - \chi^{(0)}\left(E_\infty, D^2_\infty\right)} \\
&= \lim_{\sigma^*\to\infty} -\frac{\chi^{(2)}\left(E_\infty, D^2_\infty\right) + \chi^{(3)}\left(E_\infty, D^2_\infty\right)}{\chi^{(1)}\left(E_\infty, D^2_\infty\right) + \chi^{(2)}\left(E_\infty, D^2_\infty\right)} \\
&= \lim_{\sigma^*\to\infty} \frac{\sigma^{*2}/2 - \sigma^{*2}/4}{\sigma^{*2} - \sigma^{*2}/2} \\
&= \frac{1}{2}.
\end{aligned}$$

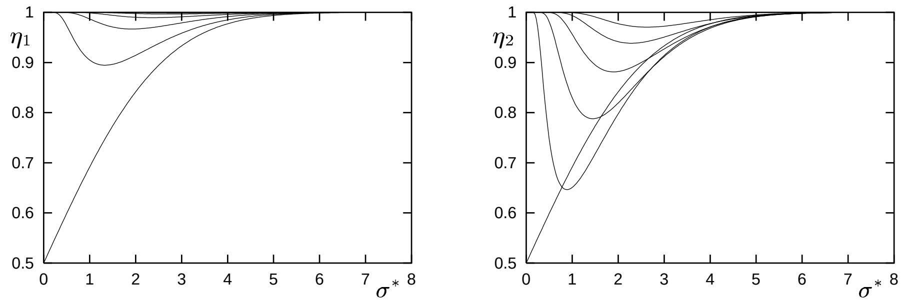

Figure 4: *Eigenvalues η_1 and η_2 of the Jacobian in Equation* (17) *as functions of the normalized mutation strength σ^* for the infinite-dimensional quadratic sphere. The curves display, from top to bottom, the results for noise strengths $\sigma_\epsilon^* = 2.0$, 1.6, 1.2, 0.8, 0.4, and* 0.0.

Thus, from Equation (15) it follows

$$\lim_{\sigma^* \to \infty} E_\infty = \sigma_\epsilon^{*2}. \tag{16}$$

These limit values are shown as dashed lines in Figure 2.

3.2.2 Stability

Section 3.1 has shown the existence of a fixed point (E_∞, D_∞^2) of the iterated map (g_1, g_2) given by Equations (11) and (12). The stability of this fixed point can be shown by linearizing the system at the location of the fixed point and showing that the resulting map is volume-contracting. For that purpose, the eigenvalues η_1 and η_2 of the Jacobian matrix

$$M = \left. \begin{pmatrix} \dfrac{\partial g_1}{\partial E_{\Xi^*}} & \dfrac{\partial g_1}{\partial D_{\Xi^*}^2} \\ \dfrac{\partial g_2}{\partial E_{\Xi^*}} & \dfrac{\partial g_2}{\partial D_{\Xi^*}^2} \end{pmatrix} \right|_{(E_\infty, D_\infty^2)} \tag{17}$$

have to be shown to be less than 1.0 in absolute value. For an introduction to stability theory of iterated maps see [9]. Computation of the eigenvalues is carried out in Appendix C.5. The results are displayed in Figure 4.

It can be observed that the eigenvalues of the system are indeed less than 1.0 in absolute value for any normalized mutation strength σ^*, $0 < \sigma^* < \infty$. Furthermore, they are always real, showing that the approach of the fixed point does not involve oscillatory behavior. Both eigenvalues tend to 1.0 for both $\sigma^* \to 0$ and $\sigma^* \to \infty$. Thus, for very small and for very large normalized mutation strengths the fixed point is almost neutral. A reason for that behavior will become obvious in the discussion of success probabilities in Section 4.1.

3.2.3 Convergence rates

While the behavior of the degree of normalized overvaluation after long time spans has been investigated in the previous sections, nothing has been said regarding the amount of time required to arrive at behavior which is statistically indistinguishable from that limit behavior. Numerical experiments indicate that that time can be very long especially for small normalized mutation strengths. For $\sigma^* = 0$ and $\sigma_\epsilon^* \neq 0$, at time t, E_{Ξ^*} is the maximum of t independent, normally distributed random variables with mean 0 and variance ${\sigma_\epsilon^*}^2$. Thus, the time behavior of E_{Ξ^*} is governed by order statistics. Using a bound derived by Beyer [6], p.76, the asymptotic behavior of the mean degree of normalized overvaluation is

$$E_{\Xi^*}(t) = \mathcal{O}\left(\sqrt{\log t}\right).$$

Thus, for $\sigma^* = 0$, the growth of the mean degree of normalized overvaluation is sublogarithmic. A very slow approach of the limit distribution can also be observed for small, non-zero normalized mutation strength.

However, it is also true that the initial approach of the fixed point is rather fast in the range of mutation strengths which will in Section 4 be seen to be the range of interest. At time step 0, the degree of normalized overvaluation is normally distributed with mean 0 and variance ${\sigma_\epsilon^*}^2$. Unless the mutation strength is chosen very high, a considerable portion of the expected final overvaluation builds up within the first few generations already. Thus, it can be expected that the effects of overvaluation become observable rather early during the evolution. This is especially true if the initial normalized mutation strength is chosen to be rather small and if it increases over time.

4 Performance

This section discusses the performance of the (1+1)-ES. In Section 4.1 the success probability is introduced, offering an explanation for the observations regarding the stability of the fixed point made in the previous section. In Section 4.2 the normalized fitness gain is computed, and in Section 4.3 optimal parameter settings are obtained. Section 4.4 discusses the value of overvaluation, the benefits of resampling, and problem of adapting the mutation strength.

4.1 Success Probabilities

The *success probability* P_{succ} is the probability with which a parent is replaced by an offspring. It can be computed from the conditional probability $P^{(acc)}_{\xi^*|\Xi^*}$ defined in Section 3.1 as

$$P_{succ} = \int_{-\infty}^{\infty} p_{\Xi^*}(x) \int_{-\infty}^{\infty} p_{\xi^*}(y) P^{(acc)}_{\xi^*|\Xi^*}(y|x)\mathrm{d}y\mathrm{d}x. \tag{18}$$

Evaluation of the integrals is once again straightforward and is carried out in Appendix C.6. According to Equation (43), the success probability reads

$$P_{succ} = 1 - \chi^{(0)}\left(E_{\Xi^*}, D^2_{\Xi^*}\right). \tag{19}$$

It is depicted in Figure 5 for the infinite-dimensional quadratic sphere in the limit case (E_∞, D^2_∞). Again, a good agreement between theory and experiments can be observed.

The reciprocal quantity $1/P_{succ}$ of the success probability is the average number of time steps that a parent survives. As for non-zero noise strength the success probability approaches zero both for small and for large normalized mutation strengths, long periods of stagnation can be observed for such parameter settings. This explains the approximate neutrality of the fixed point observed in Section 3.2.2.

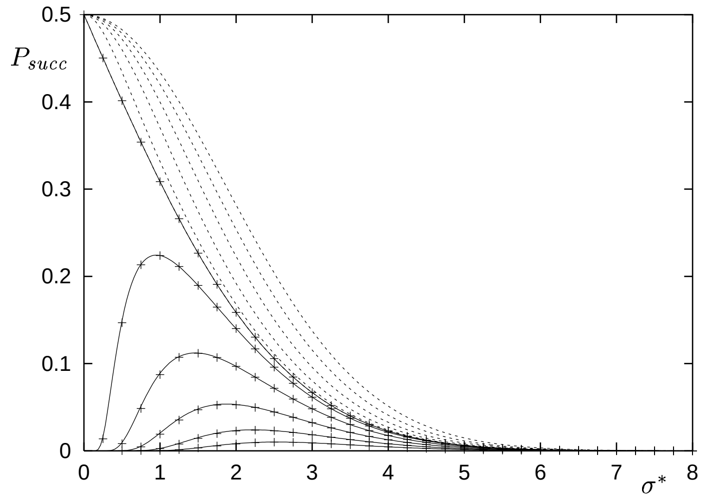

Figure 5: *Success probability P_{succ} as a function of normalized mutation strength σ^* for the infinite-dimensional quadratic sphere. The solid curves display, from top to bottom, the results from Equation* (19) *for noise strengths $\sigma_\epsilon^* = 0.0$, 0.4, 0.8, 1.2, 1.6, and 2.0. The crosses represent data measured in experiments described in Appendix A. The dashed curves display the corresponding results from [3] for a strategy which reevaluates the fitness of the parent at every time step. Note that for the latter strategy the noise strength increases from bottom to top.*

Comparing the dependence of the success probability on the normalized mutation strength, qualitative differences between the $(1+1)$-ES with and without reevaluation of the parental fitness can be observed. Results arrived at by Beyer [3] for a $(1+1)$-ES with reevaluation are included in Figure 5 for reference. Alternatively, the same results can be arrived at from Equation (19) with $E_{\Xi^*} = 0$ as reevaluating the parental fitness yields a degree of normalized overvaluation which is normal with mean 0 and variance σ_ϵ^{*2} due to the nature of fitness noise.

In general, with reevaluation, increasing the noise strength increases the success probability; without reevaluation, it decreases it. The difference between the two strategies is particularly pronounced for small normalized mutation strength. While with reevaluation the chance of accepting an offspring is close to 0.5, that chance is close to 0.0 — except in the absence of noise — if the parental fitness is not reevaluated. With reevaluation of the parental fitness, for $\sigma_\epsilon^* \to \infty$ and finite σ^*, the strategy performs a random walk in parameter space. Without reevaluation, the evolution simply stagnates.

4.2 Fitness Gain

As a progress measure, let the normalized *fitness gain* q_+^* of a generation equal zero if the offspring is rejected, and let it equal the normalized fitness advantage of the offspring if it is accepted. The expected normalized fitness gain is then

$$\mathcal{E}[q_+^*] = \int_{-\infty}^{\infty} p_{\Xi^*}(x) \int_{-\infty}^{\infty} y p_{q^*}(y) P_{q^*|\Xi^*}^{(acc)}(y|x) \mathrm{d}y \mathrm{d}x, \tag{20}$$

where, using the fact that ξ^* is normally distributed with mean 0 and variance σ_ϵ^{*2}, the probability of accepting an offspring with a normalized fitness advantage $q^* = y$ given a

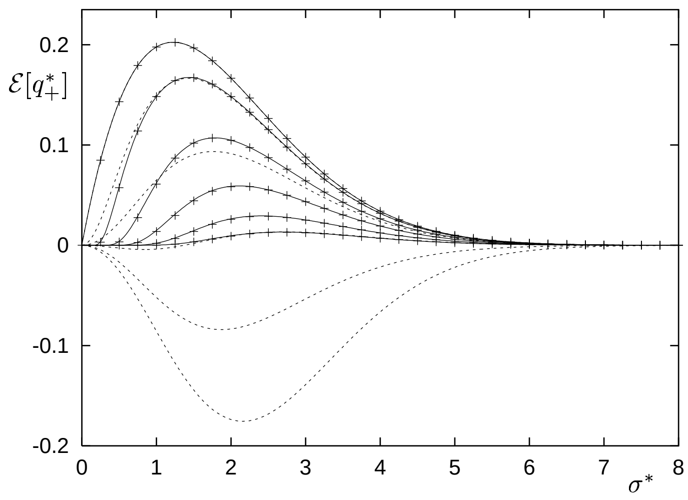

Figure 6: *Expected normalized fitness gain $\mathcal{E}[q^*_+]$ as a function of normalized mutation strength σ^* for the infinite-dimensional quadratic sphere. The solid curves display, from top to bottom, the results for normalized noise strengths $\sigma^*_\epsilon = 0.0$, 0.4, 0.8, 1.2, 1.6, and 2.0. The crosses represent data measured in experiments described in Appendix A. The dashed lines represent the corresponding results from [3] for a strategy reevaluating the parental fitness at every time step.*

parent with degree of normalized overvaluation $\Xi^* = x$ is

$$\begin{aligned} P^{(acc)}_{q^*|\Xi^*}(y|x) &= \text{Prob}(s^2 < S^2) \\ &= \text{Prob}(r^2 - \xi < R^2 - \Xi) \\ &= \text{Prob}(x - y < \xi^*) \\ &= \Phi\left(\frac{y-x}{\sigma^*_\epsilon}\right). \end{aligned} \tag{21}$$

Solving the integrals in Equation (20) is straightforward. The computations are carried out in Appendix C.7. According to Equation (44), the result reads

$$\mathcal{E}[q^*_+] = E_{q^*}\left[1 - \chi^{(0)}\left(E_{\Xi^*}, D^2_{\Xi^*}\right)\right] + D^2_{q^*}\chi^{(1)}\left(E_{\Xi^*}, D^2_{\Xi^*}\right). \tag{22}$$

It is depicted in Figure 6 for the infinite-dimensional quadratic sphere in the limit case (E_∞, D^2_∞). Again, a good agreement between theory and experiments can be observed.

Comparing the dependence of the expected normalized fitness gain on the normalized mutation strength, qualitative differences between the $(1+1)$-ES with and without reevaluation of the parental fitness can be observed. Results arrived at by Beyer [3] for a $(1+1)$-ES with reevaluation are included in Figure 6 for reference. The same results can be obtained from Equation (22) with $E_{\Xi^*} = 0$ as explained in Section 4.1.

With reevaluation of the parental fitness, the expected normalized fitness gain is negative for a wide range of mutation strengths if the normalized noise strength exceeds a value of about 1.0. For σ^*_ϵ greater than about 1.4, the expected normalized fitness gain is negative except for very high normalized mutation strengths. Without reevaluation on the other hand, the expected normalized fitness gain is non-negative for any normalized mutation strength.

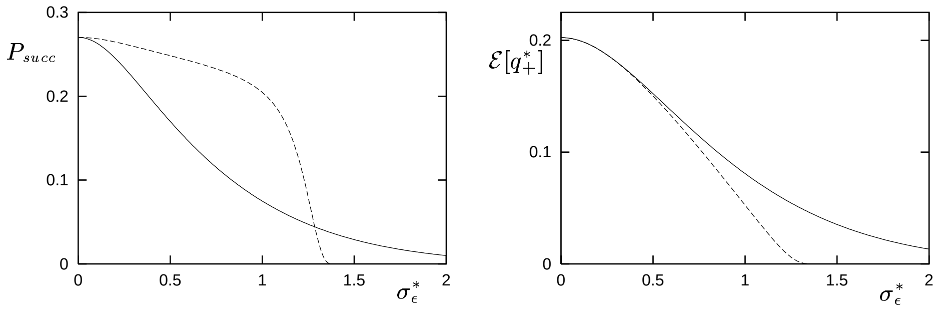

Figure 7: *Success probability P_{succ} and expected normalized fitness gain $\mathcal{E}[q_+^*]$ in the case of optimally chosen normalized mutation strength as functions of normalized noise strength σ_ϵ^* for the infinite-dimensional quadratic sphere. The solid and the dashed curves display the results for the strategies without and with reevaluation of the parental fitness, respectively.*

4.3 Optimal Parameter Settings

The performance of the $(1+1)$-ES for optimally chosen normalized mutation strength, i.e. for the normalized mutation strength that maximizes the expected normalized fitness gain, is of particular interest. It can be obtained by computing the derivative of the expected normalized fitness gain with respect to σ^* and finding a root thereof.

Figure 7 displays the success probability and the fitness gain for optimally chosen normalized mutation strength for both the strategies with and without reevaluation of the parental fitness. The right graph shows that the strategy without reevaluation is never inferior, but is clearly superior to that with reevaluation for high noise strengths. In that case, the strategy with reevaluation requires very high mutation strengths to effectively reduce the success probability to zero so as to avoid negative expected normalized fitness gain. The strategy without reevaluation on the other hand is still capable of producing measurable positive normalized fitness gain. Considering efficiency, i.e. expected fitness gain per evaluation of the fitness function, the strategy without reevaluation fares even better as it requires only one fitness evaluation per time step, as compared to the two required for the strategy with reevaluation.

4.4 Discussion

This section discusses the benefits of a positive degree of normalized overvaluation, the use of resampling to improve efficiency, and the problem of adapting the mutation strength.

4.4.1 Overvaluation revisited

The previous section has shown that a $(1+1)$-ES which does not reevaluate the parental fitness locally outperforms one which does on the noisy infinite-dimensional quadratic sphere. Failure to reevaluate the parental fitness has been seen to lead to systematic overvaluation. For zero mutation strength, it has been shown that the degree of normalized overvaluation grows sublogarithmically with time. For non-zero mutation strength, from time to time a gain in ideal fitness acts to reduce the degree of normalized overvaluation, which in turn approaches a stable limit distribution.

The reason for the improved performance of the strategy without reevaluation are the reduced success probabilities that are a consequence of systematic overvaluation of the parental fitness. A small gain in perceived fitness is often due to noise rather than to a gain in ideal fitness. Positive overvaluation raises the bar for an offspring to be accepted, preventing the strategy from making steps purely due to noise that are likely to lead it away from the location of the optimum.

However, positive overvaluation also leads to the rejection of offspring which are superior in terms of ideal fitness and which would be accepted if no overvaluation were present. Clearly, both zero overvaluation and infinite overvaluation are less than optimal. It is interesting to ask what degree of normalized overvaluation would be optimal if it could be selected deliberately instead of being a result of the interplay between mutation and selection. As, using Equation (22),

$$\frac{\partial \mathcal{E}\left[q_+^*\right]}{\partial E_{\Xi^*}} = \sigma^{*2}\left(\frac{1}{2}\chi^{(1)} + \chi^{(2)}\right),$$

optimal E_{Ξ^*} necessarily requires $\chi^{(1)}/2 + \chi^{(2)} = 0$ and therefore, using the identities listed in Appendix C.1, $E_{\Xi^*} = \sigma_\epsilon^{*2}$. This value agrees exactly with the degree of normalized overvaluation that is obtained for $\sigma^* \to \infty$. For finite normalized mutation strength however, as can be seen from Figure 2, the expected degree of normalized overvaluation that results from the interplay between mutation and selection is higher than optimal. Therefore, occasional reevaluation of the parental fitness may be preferable to both no reevaluation at all and to reevaluation in every generation. How frequently the parental fitness should be reevaluated depends on the accuracy with which the mutation strength can be adapted.

4.4.2 Resampling

Naturally, resampling can be employed as a way of reducing the noise strength. Averaging over k independent evaluations of the fitness function at any one parameter space location effectively reduces the noise strength to $\sigma_\epsilon/\sqrt{k}$, albeit at the cost of a k-fold increase in computational costs even if the overhead resulting from averaging is ignored. Let the expected fitness gain per fitness function evaluation in case of optimally adjusted mutation strength and k-times sampling be denoted as η_k. That is, define

$$\eta_k(\sigma_\epsilon^*) = \frac{1}{k}\max_{\sigma^* \in I\!R_+}\left(\mathcal{E}[q_+^*]\big|_{\sigma_\epsilon^*/\sqrt{k}}\right).$$

Then for any given normalized noise strength σ_ϵ^* there is a $k = k_{opt}$ which maximizes $\eta_k(\sigma_\epsilon^*)$. Table 1 lists the values of k_{opt} for a range of normalized noise strengths for a strategy which does not reevaluate parental fitness.

k_{opt}	1	2	3	4	5	6	7	8	9	10
σ_ϵ^*	1.44	1.90	2.26	2.57	2.85	3.10	3.33	3.55	3.75	3.94

Table 1: *Optimal number k_{opt} of fitness evaluations per offspring. The values in the lower row indicate the normalized noise strength up to which the number of fitness function evaluations in the upper row is preferable to the next higher number.*

4.4.3 Mutation strength adaptation

As pointed out in Section 2, the analysis of the preceding sections relies on the constancy of the normalized mutation strength σ^*. Due to the nature of the normalization in Equation (2) and the fact that the distance R to the location of the optimum changes, a mechanism for the adaptation of the mutation strength σ is required.

In $(1 + 1)$-ES, mutation strength adaptation mechanisms are commonly based on the measuring of success probabilities. The mutation strength adaptation mechanism proposed in the seminal book by Rechenberg [14] relies on the observation that for the fitness functions investigated, the success probabilities in case of optimally adjusted mutation strength are in a range of values centered at about one fifth, and that generally increasing the mutation strength reduces the success probability and vice versa. Thus, Rechenberg's recommendation is to monitor success probabilities by averaging over a number of generations, and to increase the mutation strength if the observed estimate of the success probability exceeds 0.2 and to decrease the mutation strength if the success probability is below 0.2. Schwefel [16], p.112, suggests the following implementation of that rule:

> After every n mutations, check how many successes have occurred over the preceding $10n$ mutations. If this number is less than $2n$, multiply the step lengths by the factor 0.85; divide them by 0.85 if more than $2n$ successes occurred.

It has to be noted that originally, this rule has been suggested only for the noise-free case. However, as shown by Beyer [3] and also as plausible from the left graph in Figure 7 in combination with Figure 5, the rule is useful for the noisy, infinite-dimensional quadratic sphere as well if the normalized noise strength does not exceed a value of about 1.2 and if the parental fitness is reevaluated in every generation. However, the figures also suggest that it is not suitable for strategies that do not reevaluate parental fitness except for relatively moderate normalized noise strengths. For normalized noise strengths exceeding about 0.47, a success probability of 0.2 is not achieved for any mutation strength. Even worse, it is not true that decreasing the mutation strength always increases the success probability and vice versa. As a result, after a number of mutation strength adaptations, the mutation strength will be so small that the normalized mutation strength is in the regime to the left of the maximum in Figure 5. Further reductions of the mutation strength act to further reduce the success probability and therefore lead in turn to a further reduction of the mutation strength. After a while, the decay of the normalized mutation strength becomes essentially exponential.

Figure 8 illustrates that this effect is not merely academic but that it is observable also in finite-dimensional search spaces and on relatively short time scales. It shows results from two typical runs of a $(1+1)$-ES using a mutation strength adaptation mechanism that differs from that suggested by Schwefel only in that an estimate of the success probability is obtained by averaging over n generations rather than over $10n$. The change has been made so as to be able to adapt the mutation strength earlier during the evolution and is rather insignificant for the performance of the algorithm. The objective function is a quadratic sphere with parameter space dimension $n = 40$. The strategy does not reevaluate the parental fitness. Experiments have been conducted with normalized noise strengths $\sigma^*_\epsilon = 0.4$ and $\sigma^*_\epsilon = 0.8$. For $\sigma^*_\epsilon = 0.4$, the mutation strength adaptation rule works reasonably well and maintains a normalized mutation strength which ensures continued progress. For $\sigma^*_\epsilon = 0.8$ however, the normalized mutation strength tends to zero after a number of mutation strength adaptations. After time step 500, in only two time steps a positive gain of ideal fitness is achieved. The right hand

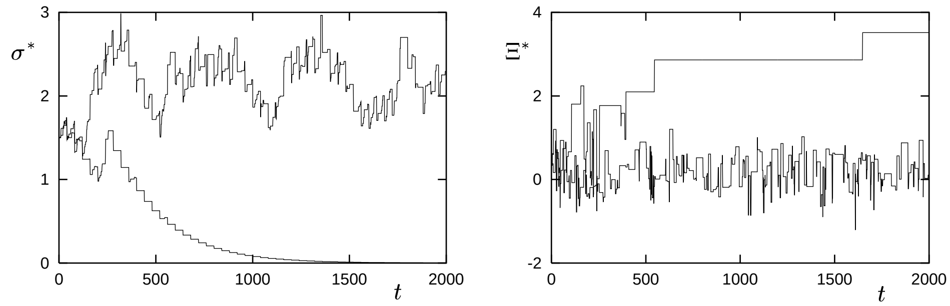

Figure 8: *Results of two typical runs of a* $(1+1)$*-ES with mutation strength adaptation based on success probabilities. Normalized noise strengths are* $\sigma_\epsilon^* = 0.4$ *and* $\sigma_\epsilon^* = 0.8$*, respectively. The left graph shows the development of the normalized mutation strength* σ^* *over time* t*, the right graph that of the degree of normalized overvaluation* Ξ^**. In the left graph, the upper curve corresponds to the lower normalized noise strength, in the right graph the lower one.*

graph in Figure 8 shows that overvaluation which has been identified as the underlying reason for the stagnation is indeed present as expected. While the limit state is not reached within 2000 generations, overvaluation already has a significant impact on the performance of the strategy.

To summarize, great care has to be taken if a success probability based mutation strength adaptation rule is used in a noisy environment if noisy fitness values can survive for more than a single generation. The presence of systematic overvaluation can lead to the violation of the assumptions on which rules such as the one quoted above rest. Of course averaging over a number of samples can always be used so as to reduce the noise strength enough to make success probability based rules work, but neither is it easy to determine whether a particular resampling rate is sufficient, nor make the computational costs involved this choice very appealing. While overvaluation has been shown to have the potential to have beneficial effects for the local performance of the $(1+1)$-ES, reevaluation of the parental fitness may be necessary to make success probability based mutation strength adaptation viable.

5 Conclusion

In this paper, the performance of a $(1+1)$-ES with isotropic normal mutations and without reevaluation of the parental fitness has been investigated for the infinite-dimensional quadratic sphere and additive Gaussian noise. The algorithm has been shown to exhibit a behavior qualitatively different from that of previously analyzed strategies which rely on reevaluation of the parental fitness in every generation. Overvaluation has been identified as the source of the observed differences, and it has been shown that while overvaluation can lead to long periods of stagnation, it can also be beneficial for the performance of the $(1+1)$-ES in the environment that has been studied. In particular, it can prevent the ES from a deterioration of the quality of the solution due to noise. However, it has also been pointed out that overvaluation can render commonly used success probability based mutation strength adaptation schemes useless.

Possible future work includes exploring possibilities of devising mutation strength adaptation algorithms which work in the presence of overvaluation. An investigation of the degree of overvaluation and its influence on the performance of a $(1+1)$-ES for finite-dimensional quadratic spheres and for other fitness functions, such as the ridge which has in the absence of noise been studied by Oyman et al. [12], are of interest. The latter fitness function does not require the dynamic adaptation of the mutation strength. Finally, modifications to the algorithm which lead to an average degree of normalized overvaluation which is closer to its optimal value are worth being investigated.

Appendix

Part A of this appendix describes the computer experiments that have been used to verify the results obtained in Sections 3 and 4. Part B introduces some mathematical basics and results useful for the understanding of the method employed for approximating the distribution of the degree of normalized overvaluation in Section 3. Also introduced are some identities useful for the solution of the integrals that have been omitted in the text and are contained in Part C of this appendix.

A Running Computer Experiments

The computer experiments the results of which have been used in Sections 3 and 4 to verify the obtained results consist of 10^9 one-generation experiments for each combination of fixed σ^* and σ_ϵ^* values examined. For that purpose, the degree of normalized overvaluation Ξ^* is initialized by generating a random sample from a normal distribution with mean E_{Ξ^*} and variance $D^2_{\Xi^*}$, where E_{Ξ^*} and $D^2_{\Xi^*}$ are determined from Equations (15) and (14). Then, in each of the 10^9 generations, normalized fitness advantage q^* and degree of normalized overvaluation of an offspring ξ^* are generated by randomly sampling from normal distributions with mean E_{q^*} and variance $D^2_{q^*}$ and mean 0 and variance σ_ϵ^{*2}, respectively. If $\Xi^* < q^* + \xi^*$, the offspring is accepted and Ξ^* is replaced by ξ^*. Otherwise, Ξ^* remains unchanged. Mean and variance of the degree of normalized overvaluation, the success probability, and the expected normalized fitness gain are obtained by averaging over all but the first 40,000 generations.

As noted in Section 4.1, for $\sigma_\epsilon^* \neq 0$ success probabilities can be very small for large and in particular for small normalized mutation strengths. As a consequence, for such parameter settings stagnation times can be very long compared with the number of one-generation experiments performed, and averaging does not yield good estimates for the mean and variance of the degree of normalized overvaluation. Such unsatisfactory results have been omitted from Figures 2 and 3.

B Some Mathematical Basics

Sections B.1 and B.2 introduce some basics from probability theory and Hermite polynomials as a preparation for the treatment of Gram-Charlier expansions in Section B.3. A more extensive account of the matter can be found in [17]. Section B.4 lists and proves some integral identities that are useful in Part C of this appendix.

B.1 Cumulants of Probability Distributions

Like moments, cumulants are descriptive constants that can be useful for measuring the properties of and, in certain circumstances, for specifying probability distributions. The *cumulant generating function* ψ of a probability distribution with density $p(x)$ is defined as

$$\psi(t) = \log\left(\int_{-\infty}^{\infty} p(x)\mathrm{e}^{itx}\mathrm{d}x\right)$$

and equals thus the logarithm of the characteristic function. If $\psi(t)$ can be expanded into a power series in t as

$$\psi(t) = \sum_{k=1}^{\infty} \left.\frac{\mathrm{d}^k\psi}{\mathrm{d}t^k}\right|_{t=0} \frac{(it)^k}{k!},$$

then the kth cumulant κ_k of the distribution is equal to the coefficient of $t^k/k!$ in the expansion. That is,

$$\kappa_k = \frac{1}{i^k} \left.\frac{\mathrm{d}^k\psi}{\mathrm{d}t^k}\right|_{t=0}$$

It can be shown that the first cumulant of a probability distribution is equal to its mean, and that the second cumulant is equal to its variance. The third and fourth cumulants are measures for the skewness and kurtosis of the distribution, respectively. The cumulants of a distribution can also be computed from its moments as detailed in [17]. The cumulant generating function of the standard normal distribution is

$$\psi(t) = it - \frac{1}{2}t^2.$$

All cumulants κ_k with $k \geq 3$ equal zero.

Cumulants possess transformation properties that make them an attractive choice for many purposes. All cumulants but the first are invariant under a change of origin. If the variate values are multiplied by a constant c, the kth cumulant κ_k is multiplied by c^k. The kth cumulant of a sum of independent random variables is the sum of the kth cumulants of its components.

B.2 Hermite Polynomials

Hermite polynomials occur naturally in connection with successive derivatives of the probability density function of the standard normal distribution and are defined by the identity

$$\left(-\frac{\mathrm{d}}{\mathrm{d}x}\right)^k \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x^2} = \mathrm{He}_k(x)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x^2}.$$

Explicit calculation shows that the first five Hermite polynomials are

$$\begin{aligned}
\mathrm{He}_0(x) &= 1\\
\mathrm{He}_1(x) &= x\\
\mathrm{He}_2(x) &= x^2 - 1\\
\mathrm{He}_3(x) &= x^3 - 3x\\
\mathrm{He}_4(x) &= x^4 - 6x^2 + 3.
\end{aligned}$$

Generally, the kth Hermite polynomial $\mathrm{He}_k(x)$ is of degree k in x and the coefficient of x^k is unity. Furthermore, it can be shown that for $k \geq 1$ the identities

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{He}_k(x) = k\mathrm{He}_{k-1}(x) \tag{23}$$

and

$$\mathrm{He}_{k+1}(x) = x\mathrm{He}_k(x) - k\mathrm{He}_{k-1}(x) \tag{24}$$

hold. Moreover, the orthogonality property

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{He}_k(x)\mathrm{He}_l(x)\mathrm{e}^{-\frac{1}{2}x^2}\mathrm{d}x = \delta_{kl}k!, \tag{25}$$

where δ_{kl} denotes the Kronecker delta, is valid for any $k, l \geq 0$.

B.3 Gram-Charlier Expansions

Gram-Charlier expansions seek to represent a given density function as a series in the derivatives of the normal density function. The approach relies on the expectation that if two distributions have a certain number of cumulants in common, they will bear some resemblance to each other. If cumulants up to order k are identical it is often the case that as k tends to infinity the distributions approach each other. Thus, the hope is that by equating the lower moments of two distributions they can be brought to approximate equality.

For the sake of notational simplicity, it is assumed that the random variable the density of which is to be approximated is standardized. A random variable can be standardized by means of a linear transformation. The transformation properties of cumulants are outlined in Section B.1. Supposing that the density function $p(x)$ can be expanded formally in a series of derivatives of the standard normal density as

$$p(x) = \frac{1}{\sqrt{2\pi}}\sum_{k=0}^{\infty} c_k \mathrm{He}_k(x)\mathrm{e}^{-\frac{1}{2}x^2}, \tag{26}$$

the coefficients c_k can be identified by means of multiplying the series with $\mathrm{He}_k(x)$, integrating from $-\infty$ to ∞, and using Equation (25). Inserting the results in Equation (26), the beginning of the series reads

$$p(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x^2}\left[1 + \frac{\kappa_3}{3!}\mathrm{He}_3(x) + \frac{\kappa_4}{4!}\mathrm{He}_4(x) + \frac{\kappa_5}{5!}\mathrm{He}_5(x) + \frac{\kappa_6 + 10\kappa_3^2}{6!}\mathrm{He}_6(x) + \dots\right].$$

The cumulants of the series equal the cumulants κ_k of the original probability distribution. Uniform convergence of the series can be shown under certain general conditions. For example, if $p(x)$ has a continuous derivative such that

$$\int_{-\infty}^{\infty}\left(\frac{\mathrm{d}p}{\mathrm{d}x}\right)^2 \mathrm{e}^{\frac{1}{2}x^2}\mathrm{d}x$$

converges, convergence of the series is assured. For details see [17].

In practical applications, usually only a finite number of terms in Equation (26) can be considered. The important question is not whether an infinite series can represent a density function, but whether a finite number of terms can serve as a satisfactory approximation. While this is often the case, it has to be kept in mind that the sum of a finite number of terms may give negative values, and that the sum may behave irregularly in the sense that the sum of k terms may give a worse fit than the sum of $k-1$ terms. The finite series approach is useful only in cases of moderate skewness of the distribution, and it is of little use if it is the tails of the distribution that are of interest.

B.4 Some Integral Identities

For any non-negative integer k, the following three identities hold:

1.

$$\frac{1}{\sqrt{2\pi}} \int \mathrm{He}_k(x) \mathrm{e}^{-\frac{1}{2}x^2} \mathrm{d}x = \begin{cases} \Phi(x) & \text{if } k = 0 \\ -\frac{1}{\sqrt{2\pi}} \mathrm{He}_{k-1}(x) \mathrm{e}^{-\frac{1}{2}x^2} & \text{if } k > 0 \end{cases} \tag{27}$$

2.

$$\begin{aligned} &\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{He}_k(x) \mathrm{e}^{-\frac{1}{2}x^2} \exp\left(-\frac{1}{2}\left(\frac{x-E}{D}\right)^2\right) \mathrm{d}x \\ &\qquad = \frac{D}{(1+D^2)^{(k+1)/2}} \mathrm{He}_k\left(\frac{E}{(1+D^2)^{1/2}}\right) \exp\left(-\frac{1}{2}\frac{E^2}{1+D^2}\right) \end{aligned} \tag{28}$$

3.

$$\begin{aligned} &\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{He}_k(x) \mathrm{e}^{-\frac{1}{2}x^2} \Phi\left(\frac{x-E}{D}\right) \mathrm{d}x \\ &= \begin{cases} 1 - \Phi\left(\frac{E}{(1+D^2)^{1/2}}\right) & \text{if } k = 0 \\ \frac{1}{\sqrt{2\pi}} \frac{1}{(1+D^2)^{k/2}} \mathrm{He}_{k-1}\left(\frac{E}{(1+D^2)^{1/2}}\right) \exp\left(-\frac{1}{2}\frac{E^2}{1+D^2}\right) & \text{if } k > 0 \end{cases} \end{aligned} \tag{29}$$

B.4.1 Proof of the first identity

To prove Equation (27) it is sufficient to show that the derivative with respect to x of the right hand side equals $\mathrm{He}_k(x)\exp(-x^2/2)/\sqrt{2\pi}$. For $k = 0$ and $k = 1$, this is immediately clear. For $k > 1$,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}\left[-\frac{1}{\sqrt{2\pi}} \mathrm{He}_{k-1}(x) \mathrm{e}^{-\frac{1}{2}x^2}\right] &= -\frac{1}{\sqrt{2\pi}} \left[(k-1)\mathrm{He}_{k-2}(x) - x\mathrm{He}_{k-1}(x)\right] \mathrm{e}^{-\frac{1}{2}x^2} \\ &= \frac{1}{\sqrt{2\pi}} \mathrm{He}_k(x) \mathrm{e}^{-\frac{1}{2}x^2}, \end{aligned}$$

where Equations (23) and (24) have been used in the first and in the second step, respectively.

B.4.2 Proof of the second identity

Proof of Equation (28) is by induction. By means of quadratic completion,

$$\begin{aligned} x^2 + \left(\frac{x-E}{D}\right)^2 &= \frac{x^2(1+D^2) - 2Ex + E^2}{D^2} \\ &= \frac{x^2 - 2Ex/(1+D^2) + E^2/(1+D^2)^2}{D^2/(1+D^2)} + \frac{E^2/(1+D^2) - E^2/(1+D^2)^2}{D^2/(1+D^2)} \\ &= \left(\frac{x - E/(1+D^2)}{D/(1+D^2)^{1/2}}\right)^2 + \frac{E^2}{1+D^2}. \end{aligned} \tag{30}$$

Let A_k and B_k denote the left and right hand sides of Equation (28), respectively. Then, for $k = 0$,

$$\begin{aligned}
A_0 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}x^2} \exp\left(-\frac{1}{2}\left(\frac{x-E}{D}\right)^2\right) \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{E^2}{1+D^2}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{x-E/(1+D^2)}{D/(1+D^2)^{1/2}}\right)^2\right) \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \frac{D}{(1+D^2)^{1/2}} \exp\left(-\frac{1}{2}\frac{E^2}{1+D^2}\right) \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}y^2} \mathrm{d}y \\
&= \frac{D}{(1+D^2)^{1/2}} \exp\left(-\frac{1}{2}\frac{E^2}{1+D^2}\right) \\
&= B_0,
\end{aligned}$$

where in the first step Equation (30) and in the second step the substitution $y = (x - E/(1+D^2))/(D/(1+D^2)^{1/2})$ have been used.

Similarly, for $k = 1$,

$$\begin{aligned}
A_1 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x\mathrm{e}^{-\frac{1}{2}x^2} \exp\left(-\frac{1}{2}\left(\frac{x-E}{D}\right)^2\right) \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{E^2}{1+D^2}\right) \int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2}\left(\frac{x-E/(1+D^2)}{D/(1+D^2)^{1/2}}\right)^2\right) \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \frac{D}{(1+D^2)^{1/2}} \exp\left(-\frac{1}{2}\frac{E^2}{1+D^2}\right) \int_{-\infty}^{\infty} \left(\frac{E}{1+D^2} + \frac{Dy}{(1+D^2)^{1/2}}\right) \mathrm{e}^{-\frac{1}{2}y^2} \mathrm{d}y \\
&= \frac{D}{1+D^2} \frac{E}{(1+D^2)^{1/2}} \exp\left(-\frac{1}{2}\frac{E^2}{1+D^2}\right) \\
&= B_1,
\end{aligned}$$

where again in the first step Equation (30) and in the second step the substitution $y = (x - E/(1+D^2))/(D/(1+D^2)^{1/2})$ have been used.

Given its validity for $k-1$ and k, where $k > 1$, the validity of Equation (28) for $k+1$, can be shown by computing the derivatives of A_k and B_k with respect to E. Using straightforward calculations, the derivative of the left hand side can be computed as

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}E} A_k &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{He}_k(x) \mathrm{e}^{-\frac{1}{2}x^2} \frac{\mathrm{d}}{\mathrm{d}E} \exp\left(-\frac{1}{2}\left(\frac{x-E}{D}\right)^2\right) \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{x-E}{D^2} \mathrm{He}_k(x) \mathrm{e}^{-\frac{1}{2}x^2} \exp\left(-\frac{1}{2}\left(\frac{x-E}{D}\right)^2\right) \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}D^2} \int_{-\infty}^{\infty} \left[\mathrm{He}_{k+1}(x) + k\mathrm{He}_{k-1}(x) - E\mathrm{He}_k(x)\right] \mathrm{e}^{-\frac{1}{2}x^2} \exp\left(-\frac{1}{2}\left(\frac{x-E}{D}\right)^2\right) \mathrm{d}x \\
&= \frac{1}{D^2} A_{k+1} + \frac{k}{D^2} A_{k-1} - \frac{E}{D^2} A_k, \qquad (31)
\end{aligned}$$

where in the second step Equation (24) has been used.

Similarly, the derivative of the right hand side can be computed as

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}E}B_k &= \frac{D}{(1+D^2)^{(k+1)/2}}\frac{\mathrm{d}}{\mathrm{d}E}\left[\mathrm{He}_k\left(\frac{E}{(1+D^2)^{1/2}}\right)\exp\left(-\frac{1}{2}\frac{E^2}{1+D^2}\right)\right] \\
&= \frac{D}{(1+D^2)^{(k+1)/2}}\left[\frac{k}{(1+D^2)^{1/2}}\mathrm{He}_{k-1}\left(\frac{E}{(1+D^2)^{1/2}}\right)\right. \\
&\qquad\left. -\frac{E}{1+D^2}\mathrm{He}_k\left(\frac{E}{(1+D^2)^{1/2}}\right)\right]\exp\left(-\frac{1}{2}\frac{E^2}{1+D^2}\right) \\
&= \frac{k}{1+D^2}B_{k-1} - \frac{E}{1+D^2}B_k, \qquad (32)
\end{aligned}
$$

where in the first step Equation (23) has been used.

From the validity of Equation (28) for k it follows equality of the derivatives. That is, $\mathrm{d}A_k/\mathrm{d}E = \mathrm{d}B_k/\mathrm{d}E$, and from Equations (31) and (32) it follows

$$
\begin{aligned}
A_{k+1} &= EA_k - kA_{k-1} + \frac{D^2k}{1+D^2}B_{k-1} - \frac{D^2E}{1+D^2}B_k \\
&= \frac{E}{1+D^2}B_k - \frac{k}{1+D^2}B_{k-1} \\
&= B_{k+1},
\end{aligned}
$$

where in the first step the identities $A_{k-1} = B_{k-1}$ and $A_k = B_k$ have been used. The final step is an immediate consequence of Equation (24).

B.4.3 Proof of the third identity

Equation (29) can now easily be proven by integrating both sides of Equation (28) with respect to E. Let as above A_k and B_k denote the left and right hand sides of Equation (28), and let C_k and D_k denote the left and right hand sides of Equation (29), respectively. Then, for the integral of the left hand side of Equation (28), it follows

$$
\begin{aligned}
\int A_k \mathrm{d}E &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{He}_k(x)\mathrm{e}^{-\frac{1}{2}x^2}\int \exp\left(-\frac{1}{2}\left(\frac{x-E}{D}\right)^2\right)\mathrm{d}E\mathrm{d}x \\
&= D\int_{-\infty}^{\infty}\mathrm{He}_k(x)\mathrm{e}^{-\frac{1}{2}x^2}\left[1-\Phi\left(\frac{x-E}{D}\right)\right]\mathrm{d}x \\
&= \begin{cases} \sqrt{2\pi}D(1-C_k) & \text{if } k=0 \\ -\sqrt{2\pi}DC_k & \text{if } k>0 \end{cases}, \qquad (33)
\end{aligned}
$$

where in the final step Equation (27) has been used. For the integral of the right hand side it follows

$$\begin{aligned}\int B_k \mathrm{d}E &= \frac{D}{(1+D^2)^{(k+1)/2}} \int \mathrm{He}_k\left(\frac{E}{(1+D^2)^{1/2}}\right)\exp\left(-\frac{1}{2}\frac{E^2}{1+D^2}\right)\mathrm{d}E \\ &= \begin{cases} \sqrt{2\pi}D\Phi\left(\frac{E}{(1+D^2)^{1/2}}\right) & \text{if } k=0 \\ -\frac{D}{(1+D^2)^{k/2}}\mathrm{He}_{k-1}\left(\frac{E}{(1+D^2)^{1/2}}\right)\exp\left(-\frac{1}{2}\frac{E^2}{1+D^2}\right) & \text{if } k>0 \end{cases} \\ &= \begin{cases} \sqrt{2\pi}D(1-D_k) & \text{if } k=0 \\ -\sqrt{2\pi}DD_k & \text{if } k>0 \end{cases}, \end{aligned} \tag{34}$$

where in the first step Equation (27) has been used.

From the validity of Equation (28) it follows equality of the integrals, i.e. $\int A_k \mathrm{d}E = \int B_k \mathrm{d}E$. Therefore, from Equations (33) and (34) it follows Equation (29).

C Evaluation of the Integrals

This section is organized as follows. Section C.1 lists and proves two useful identities which make it possible to state the results to be derived much more concisely. In Section C.2 the probability density of Ξ_+^* is obtained. Then, in Sections C.3 and C.4, the mean and variance of Ξ_+^* are computed. Sections C.6, C.5, and C.7 contain derivations of the eigenvalues of the Jacobian, the success probability, and the expected fitness gain, respectively.

C.1 Partial Derivatives of χ

Defining

$$\chi = \Phi\left(\frac{E_{\Xi^*} - E_{q^*}}{\sqrt{D_{q^*}^2 + \sigma_\epsilon^{*2} + D_{\Xi^*}^2}}\right)$$

as in Equation (10) and writing E as an abbreviation for $E_{q^*} - E_{\Xi^*}$ and D as an abbreviation for $\sqrt{D_{q^*}^2 + \sigma_\epsilon^{*2} + D_{\Xi^*}^2}$, the kth derivative, $k > 0$, of χ with respect to E_{Ξ^*} is

$$\chi^{(k)} = \frac{\partial^k \chi}{\partial E_{\Xi^*}^k} = \frac{1}{\sqrt{2\pi}D^k}\mathrm{He}_{k-1}\left(\frac{E}{D}\right)\exp\left(-\frac{1}{2}\left(\frac{E}{D}\right)^2\right). \tag{35}$$

The identity can easily be shown by induction. Straightforward calculation shows validity of Equation (35) for $k = 1$. To infer its validity for $k + 1$ from its validity for k, it is easily verified that

$$
\begin{aligned}
\frac{\partial^{k+1}}{\partial E_{\Xi^*}^{k+1}}\chi &= \frac{\partial \chi^{(k)}}{\partial E_{\Xi^*}} \\
&= \frac{1}{\sqrt{2\pi}D^k}\frac{\partial}{\partial E_{\Xi^*}}\left[\mathrm{He}_{k-1}\left(\frac{E}{D}\right)\exp\left(-\frac{1}{2}\left(\frac{E}{D}\right)^2\right)\right] \\
&= \frac{1}{\sqrt{2\pi}D^k}\left[\frac{E}{D^2}\mathrm{He}_{k-1}\left(\frac{E}{D}\right) - \frac{k-1}{D}\mathrm{He}_{k-2}\left(\frac{E}{D}\right)\right]\exp\left(-\frac{1}{2}\left(\frac{E}{D}\right)^2\right) \\
&= \frac{1}{\sqrt{2\pi}D^{k+1}}\mathrm{He}_k\left(\frac{E}{D}\right)\exp\left(-\frac{1}{2}\left(\frac{E}{D}\right)^2\right),
\end{aligned}
$$

where Equations (23) and (24) have been used in the second and third steps, respectively.

Straightforward calculation also shows that

$$\frac{\partial \chi}{\partial D_{\Xi^*}^2} = \frac{1}{2}\chi^{(2)}. \tag{36}$$

As a simple corollary,

$$\frac{\partial \chi^{(k)}}{\partial D_{\Xi^*}^2} = \frac{1}{2}\chi^{(k+2)}. \tag{37}$$

Furthermore, for the steady state straightforward calculations show that

$$\frac{\partial \chi}{\partial \sigma^*} = \left(\frac{\partial E_\infty}{\partial \sigma^*} + \sigma^*\right)\chi^{(1)} + \sigma^*\chi^{(2)}, \tag{38}$$

where the relationships $E_{q^*} = -\sigma^{*2}/2$, $D_{q^*} = \sigma^{*2}$, and $D_\infty^2 = \sigma_\epsilon^{*2}$ which are valid for the infinite-dimensional quadratic sphere have been used.

C.2 Computation of the Probability Density $p_{\Xi_+^*}$

The probability density of Ξ_+^* is given by Equation (7). Using Equations (5), (4), and (6), for the first integral therein it follows

$$
\begin{aligned}
&\int_{-\infty}^{\infty} p_{\Xi^*}(y)p_{\xi^*}(x)P_{\xi^*|\Xi^*}^{(acc)}(x|y)\mathrm{d}y \\
&\quad= \frac{1}{2\pi\sigma_\epsilon^* D_{\Xi^*}}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}\left(\frac{y-E_{\Xi^*}}{D_{\Xi^*}}\right)^2\right)\exp\left(-\frac{1}{2}\left(\frac{x}{\sigma_\epsilon^*}\right)^2\right)\Phi\left(\frac{x+E_{q^*}-y}{D_{q^*}}\right)\mathrm{d}y \\
&\quad= \frac{1}{2\pi\sigma_\epsilon^*}\exp\left(-\frac{1}{2}\left(\frac{x}{\sigma_\epsilon^*}\right)^2\right)\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{1}{2}z^2}\Phi\left(\frac{x+E_{q^*}-E_{\Xi^*}-D_{\Xi^*}z}{D_{q^*}}\right)\mathrm{d}z \\
&\quad= \frac{1}{\sqrt{2\pi}\sigma_\epsilon^*}\exp\left(-\frac{1}{2}\left(\frac{x}{\sigma_\epsilon^*}\right)^2\right)\Phi\left(\frac{x+E_{q^*}-E_{\Xi^*}}{\sqrt{D_{q^*}^2+D_{\Xi^*}^2}}\right),
\end{aligned}
$$

where in the second step the substitution $z = (y - E_{\Xi^*})/D_{\Xi^*}$ and in the third step Equation (29) have been used.

For the second integral in Equation (7) it follows analogously

$$\begin{aligned}
&\int_{-\infty}^{\infty} p_{\Xi^*}(x) p_{\xi^*}(y) \left[1 - P^{(acc)}_{\xi^*|\Xi^*}(y|x)\right] \mathrm{d}y \\
&\quad = \frac{1}{2\pi\sigma_\epsilon^* D_{\Xi^*}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{x - E_{\Xi^*}}{D_{\Xi^*}}\right)^2\right) \exp\left(-\frac{1}{2}\left(\frac{y}{\sigma_\epsilon^*}\right)^2\right) \left[1 - \Phi\left(\frac{y - x + E_{q^*}}{D_{q^*}}\right)\right] \mathrm{d}y \\
&\quad = \frac{1}{2\pi D_{\Xi^*}} \exp\left(-\frac{1}{2}\left(\frac{x - E_{\Xi^*}}{D_{\Xi^*}}\right)^2\right) \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}z^2} \left[1 - \Phi\left(\frac{\sigma_\epsilon^* z - x + E_{q^*}}{D_{q^*}}\right)\right] \mathrm{d}z \\
&\quad = \frac{1}{\sqrt{2\pi} D_{\Xi^*}} \exp\left(-\frac{1}{2}\left(\frac{x - E_{\Xi^*}}{D_{\Xi^*}}\right)^2\right) \Phi\left(\frac{x - E_{q^*}}{\sqrt{D_{q^*}^2 + \sigma_\epsilon^{*2}}}\right),
\end{aligned}$$

where in the second step the substitution $z = y/\sigma_\epsilon^*$ and in the third step Equation (29) have been used.

Putting together the results it follows

$$\begin{aligned}
p_{\Xi_+^*}(x) = {} & \frac{1}{\sqrt{2\pi}\sigma_\epsilon^*} \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma_\epsilon^*}\right)^2\right) \Phi\left(\frac{x + E_{q^*} - E_{\Xi^*}}{\sqrt{D_{q^*}^2 + D_{\Xi^*}^2}}\right) \\
& + \frac{1}{\sqrt{2\pi} D_{\Xi^*}} \exp\left(-\frac{1}{2}\left(\frac{x - E_{\Xi^*}}{D_{\Xi^*}}\right)^2\right) \Phi\left(\frac{x - E_{q^*}}{\sqrt{D_{q^*}^2 + \sigma_\epsilon^{*2}}}\right)
\end{aligned} \tag{39}$$

for the probability density of Ξ_+^*.

C.3 Computation of the Mean of Ξ_+^*

With the probability density $p_{\Xi_+^*}$ expressed as the sum of two terms as in Equation (39), computation of the mean

$$\mathcal{E}\left[\Xi_+^*\right] = \int_{-\infty}^{\infty} x p_{\Xi_+^*}(x) \mathrm{d}x$$

involves the evaluation of two integrals. Evaluating the first of these yields

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi}\sigma_\epsilon^*} \int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma_\epsilon^*}\right)^2\right) \Phi\left(\frac{x + E_{q^*} - E_{\Xi^*}}{\sqrt{D_{q^*}^2 + D_{\Xi^*}^2}}\right) \mathrm{d}x \\
&\qquad = \frac{\sigma_\epsilon^*}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y \mathrm{e}^{-\frac{1}{2}y^2} \Phi\left(\frac{\sigma_\epsilon^* y + E_{q^*} - E_{\Xi^*}}{\sqrt{D_{q^*}^2 + D_{\Xi^*}^2}}\right) \mathrm{d}y \\
&\qquad = \frac{\sigma_\epsilon^{*2}}{\sqrt{2\pi}\sqrt{D_{q^*}^2 + \sigma_\epsilon^{*2} + D_{\Xi^*}^2}} \exp\left(-\frac{1}{2}\left(\frac{E_{q^*} - E_{\Xi^*}}{\sqrt{D_{q^*}^2 + \sigma_\epsilon^{*2} + D_{\Xi^*}^2}}\right)^2\right) \\
&\qquad = \sigma_\epsilon^{*2} \chi^{(1)},
\end{aligned}$$

where in the first step the substitution $y = x/\sigma_\epsilon^*$ and in the second step Equations (29) and (35) have been used.

Analogously, evaluation of the second integral yields

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi} D_{\Xi^*}} \int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2}\left(\frac{x - E_{\Xi^*}}{D_{\Xi^*}}\right)^2\right) \Phi\left(\frac{x - E_{q^*}}{\sqrt{D_{q^*}^2 + \sigma_\epsilon^{*2}}}\right) \mathrm{d}x \\
&\qquad = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (E_{\Xi^*} + D_{\Xi^*} y) \mathrm{e}^{-\frac{1}{2}y^2} \Phi\left(\frac{D_{\Xi^*} y - E_{q^*} + E_{\Xi^*}}{\sqrt{D_{q^*}^2 + \sigma_\epsilon^{*2}}}\right) \mathrm{d}y \\
&\qquad = E_{\Xi^*} \Phi\left(\frac{E_{\Xi^*} - E_{q^*}}{\sqrt{D_{q^*}^2 + \sigma_\epsilon^{*2} + D_{\Xi^*}^2}}\right) \\
&\qquad\qquad + \frac{D_{\Xi^*}^2}{\sqrt{2\pi}\sqrt{D_{q^*}^2 + \sigma_\epsilon^{*2} + D_{\Xi^*}^2}} \exp\left(-\frac{1}{2}\left(\frac{E_{q^*} - E_{\Xi^*}}{\sqrt{D_{q^*}^2 + \sigma_\epsilon^{*2} + D_{\Xi^*}^2}}\right)^2\right) \\
&\qquad = E_{\Xi^*} \chi^{(0)} + D_{\Xi^*}^2 \chi^{(1)},
\end{aligned}$$

where in the first step the substitution $y = (x - E_{\Xi^*})/D_{\Xi^*}$ and in the second step Equations (29), (10), and (35) have been used.

Altogether, it follows

$$\mathcal{E}\left[\Xi_+^*\right] = E_{\Xi^*} \chi^{(0)} + \left(\sigma_\epsilon^{*2} + D_{\Xi^*}^2\right) \chi^{(1)}. \tag{40}$$

for the mean of Ξ_+^*.

C.4 Computation of the Variance of Ξ_+^*

With the probability density $p_{\Xi_+^*}$ expressed as the sum of two terms as in Equation (39), computation of the variance

$$\mathcal{D}^2\left[\Xi_+^*\right] = \int_{-\infty}^{\infty} x^2 p_{\Xi_+^*}(x) \mathrm{d}x - \left(\mathcal{E}\left[\Xi_+^*\right]\right)^2$$

involves the evaluation of two integrals. Evaluating the first of these yields

$$
\begin{aligned}
&\frac{1}{\sqrt{2\pi}\sigma_\epsilon^*}\int_{-\infty}^{\infty} x^2 \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma_\epsilon^*}\right)^2\right)\Phi\left(\frac{x+E_{q^*}-E_{\Xi^*}}{\sqrt{D_{q^*}^2+D_{\Xi^*}^2}}\right)\mathrm{d}x \\
&\qquad = \frac{\sigma_\epsilon^{*2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left(1+\left(y^2-1\right)\right)\mathrm{e}^{-\frac{1}{2}y^2}\Phi\left(\frac{\sigma_\epsilon^* y+E_{q^*}-E_{\Xi^*}}{\sqrt{D_{q^*}^2+D_{\Xi^*}^2}}\right)\mathrm{d}y \\
&\qquad = \sigma_\epsilon^{*2}\left[1-\Phi\left(\frac{E_{\Xi^*}-E_{q^*}}{\sqrt{D_{q^*}^2+\sigma_\epsilon^{*2}+D_{\Xi^*}^2}}\right)\right] \\
&\qquad\qquad - \frac{\sigma_\epsilon^{*4}(E_{q^*}-E_{\Xi^*})}{\sqrt{2\pi}\sqrt{D_{q^*}^2+\sigma_\epsilon^2+D_{\Xi^*}^2}^3}\exp\left(-\frac{1}{2}\left(\frac{E_{q^*}-E_{\Xi^*}}{\sqrt{D_{q^*}^2+\sigma_\epsilon^2+D_{\Xi^*}^2}}\right)^2\right) \\
&\qquad = \sigma_\epsilon^{*2}\left(1-\chi^{(0)}\right)-\sigma_\epsilon^{*4}\chi^{(2)},
\end{aligned}
$$

where substitution $y = x/\sigma_\epsilon^*$ has been used in the first step and Equation (29) in the second.

Analogously, evaluation of the second integral yields

$$
\begin{aligned}
&\frac{1}{\sqrt{2\pi}D_{\Xi^*}}\int_{-\infty}^{\infty} x^2 \exp\left(-\frac{1}{2}\left(\frac{x-E_{\Xi^*}}{D_{\Xi^*}}\right)^2\right)\Phi\left(\frac{x-E_{q^*}}{\sqrt{D_{q^*}^2+\sigma_\epsilon^{*2}}}\right)\mathrm{d}x \\
&\qquad = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left(\left(E_{\Xi^*}^2+D_{\Xi^*}^2\right)+2E_{\Xi^*}D_{\Xi^*}y+D_{\Xi^*}^2\left(y^2-1\right)\right) \\
&\qquad\qquad\qquad \mathrm{e}^{-\frac{1}{2}y^2}\Phi\left(\frac{D_{\Xi^*}y-E_{q^*}+E_{\Xi^*}}{\sqrt{D_{q^*}^2+\sigma_\epsilon^{*2}}}\right)\mathrm{d}y \\
&\qquad = \left(E_{\Xi^*}^2+D_{\Xi^*}^2\right)\Phi\left(\frac{E_{\Xi^*}-E_{q^*}}{\sqrt{D_{q^*}^2+\sigma_\epsilon^{*2}+D_{\Xi^*}^2}}\right) \\
&\qquad\qquad + \frac{2E_{\Xi^*}D_{\Xi^*}^2}{\sqrt{2\pi}\sqrt{D_{q^*}^2+\sigma_\epsilon^2+D_{\Xi^*}^2}}\exp\left(-\frac{1}{2}\left(\frac{E_{q^*}-E_{\Xi^*}}{\sqrt{D_{q^*}^2+\sigma_\epsilon^2+D_{\Xi^*}^2}}\right)^2\right) \\
&\qquad\qquad + \frac{D_{\Xi^*}^4(E_{q^*}-E_{\Xi^*})}{\sqrt{2\pi}\sqrt{D_{q^*}^2+\sigma_\epsilon^{*2}+D_{\Xi^*}^2}^3}\exp\left(-\frac{1}{2}\left(\frac{E_{q^*}-E_{\Xi^*}}{\sqrt{D_{q^*}^2+\sigma_\epsilon^2+D_{\Xi^*}^2}}\right)^2\right) \\
&\qquad = \left(E_{\Xi^*}^2+D_{\Xi^*}^2\right)\chi^{(0)}+2E_{\Xi^*}D_{\Xi^*}^2\chi^{(1)}+D_{\Xi^*}^4\chi^{(2)},
\end{aligned}
$$

where the substitution $y = (x-E_{\Xi^*})/D_{\Xi^*}$ has been used in the first step and Equation (29) in the second.

Altogether, making use of Equations (35) and (40), it follows

$$\begin{aligned}
\mathcal{D}^2\left[\Xi_+^*\right] &= \sigma_\epsilon^{*2} + \left(E_{\Xi^*}^2 + D_{\Xi^*}^2 - \sigma_\epsilon^{*2}\right)\chi^{(0)} + 2E_{\Xi^*}D_{\Xi^*}^2\chi^{(1)} - \left(\sigma_\epsilon^{*4} - D_{\Xi^*}^4\right)\chi^{(2)} - \left(\mathcal{E}\left[\Xi_+^*\right]\right)^2 \\
&= D_{\Xi^*}^2 + E_{\Xi^*}\left[E_{\Xi^*}\chi^{(0)} + \left(\sigma_\epsilon^{*2} + D_{\Xi^*}^2\right)\chi^{(1)}\right] - \left(\mathcal{E}\left[\Xi_+^*\right]\right)^2 \\
&\qquad + \left(\sigma_\epsilon^{*2} - D_{\Xi^*}^2\right)\left[1 - \chi^{(0)} - E_{\Xi^*}\chi^{(1)} - \left(\sigma_\epsilon^{*2} + D_{\Xi^*}^2\right)\chi^{(2)}\right] \\
&= D_{\Xi^*}^2 + E_{\Xi^*}\mathcal{E}\left[\Xi_+^*\right] - \left(\mathcal{E}\left[\Xi_+^*\right]\right)^2 + \left(\sigma_\epsilon^{*2} - D_{\Xi^*}^2\right)\left(1 - \frac{\partial\mathcal{E}\left[\Xi_+^*\right]}{\partial E_{\Xi^*}}\right)
\end{aligned} \tag{41}$$

for the variance of Ξ_+^*.

C.5 Determination of the Eigenvalues of the Jacobian

With g_1 and g_2 given by Equations (11) and (12), respectively, the entries of the Jacobian defined in Equation (17) are

$$\begin{aligned}
m_{11} &= \left.\frac{\partial g_1}{\partial E_{\Xi^*}}\right|_{(E_\infty, D_\infty^2)} \\
&= \chi^{(0)}\left(E_\infty, D_\infty^2\right) + E_\infty\chi^{(1)}\left(E_\infty, D_\infty^2\right) + 2D_\infty^2\chi^{(2)}\left(E_\infty, D_\infty^2\right)
\end{aligned}$$

$$\begin{aligned}
m_{12} &= \left.\frac{\partial g_1}{\partial D_{\Xi^*}^2}\right|_{(E_\infty, D_\infty^2)} \\
&= \chi^{(1)}\left(E_\infty, D_\infty^2\right) + \frac{1}{2}E_\infty\chi^{(2)}\left(E_\infty, D_\infty^2\right) + D_\infty^2\chi^{(3)}\left(E_\infty, D_\infty^2\right)
\end{aligned}$$

$$\begin{aligned}
m_{21} &= \left.\frac{\partial g_2}{\partial E_{\Xi^*}}\right|_{(E_\infty, D_\infty^2)} \\
&= g_1\left(E_\infty, D_\infty^2\right) + \left(E_\infty - 2g_1\left(E_\infty, D_\infty^2\right)\right)\left.\frac{\partial g_1}{\partial E_{\Xi^*}}\right|_{(E_\infty, D_\infty^2)} - \left(\sigma_\epsilon^{*2} - D_\infty^2\right)\left.\frac{\partial^2 g_1}{\partial E_{\Xi^*}^2}\right|_{(E_\infty, D_\infty^2)} \\
&= (1 - m_{11})E_\infty
\end{aligned}$$

$$\begin{aligned}
m_{22} &= \left.\frac{\partial g_2}{\partial D_{\Xi^*}^2}\right|_{(E_\infty, D_\infty^2)} \\
&= \left(E_\infty - 2g_1\left(E_\infty, D_\infty^2\right)\right)\left.\frac{\partial g_1}{\partial D_{\Xi^*}^2}\right|_{(E_\infty, D_\infty^2)} + \left.\frac{\partial g_1}{\partial E_{\Xi^*}}\right|_{(E_\infty, D_\infty^2)} \\
&\qquad - \left(\sigma_\epsilon^{*2} - D_\infty^2\right)\left.\frac{\partial^2 g_1}{\partial E_{\Xi^*}\partial D_{\Xi^*}^2}\right|_{(E_\infty, D_\infty^2)} \\
&= m_{11} - m_{12}E_\infty,
\end{aligned}$$

where Equations (35), (36), (37), (14), and (15) have been used. It follows

$$\eta_{1,2} = \frac{m_{11} + m_{22}}{2} \pm \sqrt{\left(\frac{m_{11} - m_{22}}{2}\right)^2 + m_{12}m_{21}} \tag{42}$$

for the eigenvalues of (m_{ij}).

C.6 Determination of the Success Probability

The success probability P_{succ} is given by Equation (18). Reversing the order of the integrations,

$$P_{succ} = \int_{-\infty}^{\infty} p_{\xi^*}(y) \int_{-\infty}^{\infty} p_{\Xi^*}(x) P^{(acc)}_{\xi^*|\Xi^*}(y|x) \mathrm{d}x \mathrm{d}y.$$

Using Equations (4), (5), and (6), it follows for the inner integral

$$\begin{aligned}
\int_{-\infty}^{\infty} p_{\Xi^*}(x) P^{(acc)}_{\xi^*|\Xi^*}(y|x) \mathrm{d}x &= \frac{1}{\sqrt{2\pi} D_{\Xi^*}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{x - E_{\Xi^*}}{D_{\Xi^*}}\right)^2\right) \Phi\left(\frac{y - x + E_{q^*}}{D_{q^*}}\right) \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}z^2} \Phi\left(\frac{y + E_{q^*} - E_{\Xi^*} - D_{\Xi^*} z}{D_{q^*}}\right) \mathrm{d}z \\
&= \Phi\left(\frac{y + E_{q^*} - E_{\Xi^*}}{\sqrt{D_{q^*}^2 + D_{\Xi^*}^2}}\right),
\end{aligned}$$

where the substitution $z = (x - E_{\Xi^*})/D_{\Xi^*}$ has been used in the second step and Equation (29) in the third. Therefore,

$$\begin{aligned}
P_{succ} &= \frac{1}{\sqrt{2\pi}\sigma_\epsilon^*} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{y}{\sigma_\epsilon^*}\right)^2\right) \Phi\left(\frac{y + E_{q^*} - E_{\Xi^*}}{\sqrt{D_{q^*}^2 + D_{\Xi^*}^2}}\right) \mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}z^2} \Phi\left(\frac{\sigma_\epsilon^* z + E_{q^*} - E_{\Xi^*}}{\sqrt{D_{q^*}^2 + D_{\Xi^*}^2}}\right) \mathrm{d}z \\
&= 1 - \Phi\left(\frac{E_{\Xi^*} - E_{q^*}}{\sqrt{D_{q^*}^2 + \sigma_\epsilon^{*2} + D_{\Xi^*}^2}},\right) \\
&= 1 - \chi^{(0)},
\end{aligned} \tag{43}$$

where the substitution $z = y/\sigma_\epsilon^*$ has been used in the first step and Equation (29) in the second.

C.7 Determination of the Expected Fitness Gain

The expected fitness gain is given by Equation (20). Reversing the order of the integrations,

$$\mathcal{E}\left[q_+^*\right] = \int_{-\infty}^{\infty} y p_{q^*}(y) \int_{-\infty}^{\infty} p_{\Xi^*}(x) P^{(acc)}_{q^*|\Xi^*}(y|x) \mathrm{d}x \mathrm{d}y.$$

Using Equations (5) and (21), it follows for the inner integral

$$\begin{aligned}\int_{-\infty}^{\infty} p_{\Xi^*}(x) P^{(acc)}_{q^*|\Xi^*}(y|x)\mathrm{d}x &= \frac{1}{\sqrt{2\pi}D_{\Xi^*}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{x-E_{\Xi^*}}{D_{\Xi^*}}\right)^2\right) \Phi\left(\frac{y-x}{\sigma^*_\epsilon}\right) \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}z^2} \Phi\left(\frac{y-E_{\Xi^*}-D_{\Xi^*}z}{\sigma^*_\epsilon}\right) \mathrm{d}z \\ &= \Phi\left(\frac{y-E_{\Xi^*}}{\sqrt{\sigma^{*2}_\epsilon + D^2_{\Xi^*}}}\right),\end{aligned}$$

where in the second step the substitution $z = (x - E_{\Xi^*})/D_{\Xi^*}$ and in the third step Equation (29) have been used. Therefore,

$$\begin{aligned}\mathcal{E}\left[q^*_+\right] &= \frac{1}{\sqrt{2\pi}D_{q^*}} \int_{-\infty}^{\infty} y \exp\left(-\frac{1}{2}\left(\frac{y-E_{q^*}}{D_{q^*}}\right)\right) \Phi\left(\frac{y-E_{\Xi^*}}{\sqrt{\sigma^{*2}_\epsilon + D^2_{\Xi^*}}}\right) \mathrm{d}y \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (E_{q^*} + D_{q^*}z)\, \mathrm{e}^{-\frac{1}{2}z^2} \Phi\left(\frac{E_{q^*}+D_{q^*}z-E_{\Xi^*}}{\sqrt{\sigma^{*2}_\epsilon + D^2_{\Xi^*}}}\right) \mathrm{d}z \\ &= E_{q^*}\left[1 - \Phi\left(\frac{E_{\Xi^*}-E_{q^*}}{\sqrt{D^2_{q^*} + \sigma^{*2}_\epsilon + D^2_{\Xi^*}}}\right)\right] \\ &\qquad + \frac{D^2_{q^*}}{\sqrt{2\pi}\sqrt{D^2_{q^*} + \sigma^{*2}_\epsilon + D^2_{\Xi^*}}} \exp\left(-\frac{1}{2}\left(\frac{E_{q^*}-E_{\Xi^*}}{\sqrt{D^2_{q^*} + \sigma^{*2}_\epsilon + D^2_{\Xi^*}}}\right)^2\right) \\ &= E_{q^*}\left(1-\chi^{(0)}\right) + D^2_{q^*}\chi^{(1)}, \end{aligned} \tag{44}$$

where in the first step substitution $z = (y - E_{\Xi^*})/D_{\Xi^*}$ and in the second step Equations (28) and (29) have been used.

Acknowledgements

Support by the Deutsche Forschungsgemeinschaft (DFG) under grants Be 1578/4-1 and Be 1578/6-1 is gratefully acknowledged. The second author is a Heisenberg fellow of the DFG.

References

[1] P. J. Angeline, "The Effects of Noise on Self-Adaptive Evolutionary Optimization", in L. J. Fogel, P. J. Angeline, and T. Bäck (eds.), *Proceedings of the Fifth Annual Conference on Evolutionary Programming*, 432-439, (MIT Press, Cambridge, 1996).

[2] D. V. Arnold, "Evolution Strategies in Noisy Environments – A Survey of Existing Work", in *Proceedings of the Second EvoNet Summer School on Theoretical Aspects of Evolutionary Computing*, (Springer, Heidelberg, 1999).

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

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

[5] H.-G. Beyer, "Toward a Theory of Evolution Strategies: The (μ, λ)-Theory", *Evolutionary Computation*, 2(4), 381-407, (1995).

[6] H.-G. Beyer, *Zur Analyse der Evolutionsstrategien*, Habilitationsschrift, Universität Dortmund, (1996).

[7] H.-G. Beyer, "Evolutionary Algorithms in Noisy Environments: Theoretical Issues and Guidelines for Practice", in *Computer Methods in Applied Mechanics and Engineering*, in print, (2000).

[8] J. M. Fitzpatrick and J. J. Grefenstette, "Genetic Algorithms in Noisy Environments", in P. Langley (ed.), *Machine Learning*, 3, 101-120, (Kluwer, Dordrecht, 1988).

[9] D. G. Luenberger, *Introduction to Dynamic Systems*, (Wiley, New York, 1979).

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

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

[12] A. I. Oyman, H.-G. Beyer, and H.-P. Schwefel, "Where Elitists Start Limping: Evolution Strategies at Ridge Functions", in A. E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel (eds.), *Parallel Problem Solving from Nature*, 5, 34-43, (Springer, Berlin, 1998).

[13] L. M. Rattray and J. Shapiro, "Noisy Fitness Evaluation in Genetic Algorithms and the Dynamics of Learning", in R. K. Belew and M. D. Vose (eds.), *Foundations of Genetic Algorithms*, 4, (Morgan Kaufmann, San Mateo, 1997).

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

[15] I. Rechenberg, *Evolutionsstrategie '94*, (Frommann-Holzboog, Stuttgart, 1994).

[16] H.-P. Schwefel, *Evolution and Optimum Seeking*, (Wiley, New York, 1995).

[17] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics*, Sixth edition, Volume 1: Distribution Theory, (Arnold, London, 1994).