Towards a Theory of 'Evolution Strategies'. Some Asymptotical Results from the $(1,^+\lambda)$-Theory

Hans-Georg Beyer
beyer@ls11.informatik.uni-dortmund.de
University of Dortmund
Department of Computer Science, Chair of Systems Analysis
P. O. Box 500500
D-44227 Dortmund
Germany

December 1, 1993

Abstract

A method for the determination of the progress rate and the probability of success for the 'Evolution Strategy' (short ES) is presented . The new method is based on the asymptotical behavior of the χ-distribution and yields exact results in case of infinite-dimensional parameter spaces. The technique is demonstrated for the $(1,^+\lambda)$ ES using a spherical model including noisy quality functions. The results are used to discuss the convergence behavior of the ES.

Keywords

Evolution Strategy (ES)
- $(1,^+\lambda)$
- sperical model
- noisy fitness
- theory
- rate of progress
optimization
mutation-selection-principle

1 Introduction

For more than 25 years the *Evolution Strategy (ES)* has been living as a method of growing interest for technical and numerical optimization problems [1, 2, 3, 4].

Simple as its algorithm is (cf. section 1. 1) the mathematical analysis of the ES, however, seems to be a very hard task and up to now there are only a few exact results. E. g., the analysis of the optimization progress (also known as the *progress rate*, cf. section 1. 2) is of special interest, since it contains the key for the optimum working of the ES. The problem is, that this *progress rate* is (from mathematical point of view) the probability theory expectation of the successful mutations to be performed in high-dimensional spaces. It is the aim of this paper to present a new approach of tackling such high-dimensional problems, to add some new results to the ES-theory, and to initiate further investigations.

This article mainly deals with the so-called $(1, \lambda)$ ES and the $(1+\lambda)$ ES to be defined in the next section. Offspring are produced by replication and mutation. Recombination and other evolutionary mechanisms are not included in this analysis, although the approach to be presented here seems to be able of coping with such questions.

1.1 The Evolution Strategy Algorithm

Let $\hat{\mathbf{y}}$ be the vector to the location of a minimum to be searched for of a function

$$Q = Q(\mathbf{y}) \quad \text{with} \quad \mathbf{y} := (y_1, y_2, \ldots, y_i, \ldots, y_N)^T. \tag{1}$$

One starts with an initial vector $\mathbf{y}_{(0)}$ yielding a Q(uality)-value $Q_{(0)} = Q(\mathbf{y}_{(0)})$. This initial $\mathbf{y}_{(0)}$ serves as a 'parent' for the next generation of λ offspring $\mathbf{y}^1_{(1)}, \mathbf{y}^2_{(1)}, \ldots, \mathbf{y}^\lambda_{(1)}$. The reproduction process is 'disturbed' by random mutations $\mathbf{Z}$ preventing the offspring to be exact copies of their parent(s)

$$\mathbf{y}^1_{(1)} = \mathbf{y}_{(0)} + \mathbf{Z}^1, \quad \mathbf{y}^2_{(1)} = \mathbf{y}_{(0)} + \mathbf{Z}^2, \ldots \quad \mathbf{y}^\lambda_{(1)} = \mathbf{y}_{(0)} + \mathbf{Z}^\lambda. \tag{2}$$

The mutations $\mathbf{Z}^l$ are random vectors

$$\mathbf{Z}^l := (z^l_1, z^l_2, \ldots, z^l_i, \ldots, z^l_N)^T \tag{3}$$

with the components z^l_i all of them fluctuating independently and $(0, \sigma^l_i)$-normally distributed, i. e. the probability density of each component is given by

$$p(z^l_i) = \frac{1}{\sqrt{2\pi}\,\sigma^l_i}\ \exp -\frac{1}{2}\left(\frac{z^l_i}{\sigma^l_i}\right)^2. \tag{4}$$

The produced offspring $\mathbf{y}^l_{(1)}$ are now subject to selection. Therefore one has to determine their fitness values $Q^l_{(1)} = Q(\mathbf{y}^l_{(1)})$. At this 'phenotypic' level selection takes place. Since a minimum of $Q(\mathbf{y})$ is searched for selection has to prefer the smaller Q -values.

In the simplest case we only have one parent $\mathbf{y}_{(g)}$ in generation g generating only one offspring $\mathbf{y}_{(g+1)} = \mathbf{y}_{(g)} + \mathbf{Z}$. Selection can only take place between parent and offspring. The selection criterion of this strategy denoted by $(1 + 1)$ ES is therefore

$$Q(\mathbf{y}_{(g)} + \mathbf{Z}) \ \leq \ Q(\mathbf{y}_{(g)}) \ \Rightarrow \ \mathbf{y}_{(g+1)} := \mathbf{y}_{(g)} + \mathbf{Z} \tag{5}$$

$$Q(\mathbf{y}_{(g)} + \mathbf{Z}) \ > \ Q(\mathbf{y}_{(g)}) \ \Rightarrow \ \mathbf{y}_{(g+1)} := \mathbf{y}_{(g)}. \tag{6}$$

Beside this simplest ES there are a variety of ES different in the number of parents (denoted by μ) and offspring (λ), in selection: $(\mu+\lambda)$ or (μ,λ) ES, and other aspects [3].

In (+)-strategies parents are allowed to survive, if they are better than the offspring, whereas in (,)-strategies parents have to die *per definitionem*, even if all the offspring are much worse than all the parents. There are certain advantages of limited lifetime of the individuals, the reader is referred to the basic work by SCHWEFEL [3, 4].

1.2 The Rate of Progress and the Strategy Parameters

Strategy parameters are the number of parents and offspring and the standard deviations σ_i^l of the random vectors $\mathbf{Z}$ (cp. (3,4)). Since the strategy should have a maximum performance it is necessary to tune these parameters - a difficult and perhaps not completely solvable problem.

A strategy is called to be effective, if its strategy parameters are tuned in such a way that the rate of progress φ is a maximum. The rate of progress φ is defined as the expectation of the change in *Euclidian* distance to the point of optimum $\hat{\mathbf{y}}$ from generation g to $(g+1)$

$$\varphi(\lambda,\mu,\sigma,\ g) := E\{\|\hat{\mathbf{y}}-\mathbf{y}_{(g)}\| - \|\hat{\mathbf{y}}-\mathbf{y}_{(g+1)}\|\}. \tag{7}$$

Besides the dependencies λ, μ, σ, and the generation counter g there are also dependencies like the functional structure of $Q(\mathbf{y})$ and the dimensionality of $\mathbf{y}$ not explicitly displayed in (7). Since our main interest concerns the high dimensional parameter spaces ($N > 30$) the investigations will be done for the asymptotical case $N \to \infty$.

To determine the rate of progress it is necessary to know something about the functional structure of the function $Q(\mathbf{y})$ to be optimized. In order to get results of a certain generality it is desirable to have a model function which matches a wide range of Q-functions and moreover it should be simple to get analytical results. In his basic work [2] RECHENBERG has proposed two models - the *sphere* and the *corridor*. There is a third model - the *discus* - introduced by the author [7] for the investigation of his *general evolution method for dissipative systems* [8]. These three models are 'simple' in the sense that there is only one parameter describing the geometrical extent of the *domain of success* (that is the part of the parameter space defining the successful mutations).

The most important model is the sphere. The Q-function has spherical symmetry with the point of optimum $\hat{\mathbf{y}}$ as symmetry center, i. e. Q is a function of the distance R to the point of optimum. The set of constant Q-values defines a $(N-1)$-dimensional hypersphere in the N-dimensional parameter space of the $\mathbf{y}$-vectors. Thus, the only model parameter describing the domain of success is the radius R of the (hyper)sphere

$$Q(\mathbf{y}) = Q(\|\mathbf{y}-\hat{\mathbf{y}}\|) = Q(R)$$

.

1.3 The Goal

This paper is devoted to the spherical model for the $(1 \overset{+}{,} \lambda)$ ES[1]. The results concerning the ES without noise are partially known [11, 12]. The approach used by the author is new allowing the treatment of *'noisy'* Q-functions (noisy means that the virtual Q-value is disturbed by an erroneous measurement, numerical errors, etc.). For the *noisy ES* the author proposes the

$$(\tilde{\mu} \overset{+}{,} \tilde{\lambda})$$

nomenclature with the tilde indicating that the Q-values are disturbed. For example, $(\tilde{2} + \lambda)$ means that the parental fitness values are noisy whereas the offspring's fitness is exact.

An essential point of the approach will be given by the detailed analysis of the asymptotics of the mutation vectors $\mathbf{Z}$ in the next section. With these results the rate of progress for the $(1 + 1)$ ES can be determined. Then the $(1 \overset{+}{,} \lambda)$ ES can easily be analyzed.

Section 3 deals with the noisy ES. It will be necessary to make an agreement for the mathematical description of the noise. Furthermore one has to specify the local structure of the $Q(R)$-function, because both the mutations and the noise of the 'phenotype' have to be expressed in the 'Q-picture'.

2 The Basic Ideas for the Determination of the Rate of Progress

2.1 The Properties of the Mutation Vectors Z

It is provided that all the components of $\mathbf{Z}$ fluctuate independently normally distributed with equal standard deviation σ. For the probability density of $\mathbf{Z}$ one gets from (4)

$$p(\mathbf{Z}) = \frac{1}{(\sqrt{2\pi}\,\sigma)^N} \exp\left(-\frac{\mathbf{Z}^2}{2\,\sigma^2}\right) \tag{8}$$

with

$$\mathbf{Z}^2 = \sum_{i=1}^{N} z_i^2$$

. It is evident that this *Gauss*ian distribution depends only on the length of $\mathbf{Z}$, but not on the direction. The distribution has therefore a spherical symmetry invariant against the rotation group $\mathbf{O}(N)$. The transformation to new coordinates z'_k

$$z'_k = \sum_{i=1}^{N} a_{ki} z_i \tag{9}$$

with the orthogonal matrix a_{ki}

$$\sum_{i=1}^{N} a_{ki} a_{kj} = \delta_{ij} \;\Rightarrow\; (\mathbf{Z}')^2 = \mathbf{Z}^2 \tag{10}$$

[1] The $(\mu \overset{+}{,} \lambda)$ ES nomenclature is only a shortened form to refere to both the (μ, λ) ES and the $(\mu + \lambda)$ ES. It does *not* define a third kind of strategy.

leaves (8) unchanged. The length χ of the random vector $\mathbf{Z}$, $\chi := \sqrt{\mathbf{Z}^2}$, obeys the so-called χ-distribution [13, 14]

$$p(\chi) = \frac{2\,\chi^{N-1} \exp\left(-\frac{\chi^2}{2\,\sigma^2}\right)}{\sigma^N 2^{N/2}\Gamma(N/2)},$$

with the expectation ℓ

$$\ell := E\{\chi\} = \int_0^\infty \chi p(\chi)\, d\chi = \sqrt{2}\,\sigma\, \frac{\Gamma\left(\frac{N+1}{2}\right)}{\Gamma\left(\frac{N}{2}\right)} \tag{11}$$

(cp. [13]). The analysis of the asymptotical behavior $N \to \infty$ of (11) yields [7]

$$E\{\chi\} \sim \sigma \sqrt{N - \frac{1}{2}}. \tag{12}$$

For the variance $D^2\{\chi\}$ we find

$$\lim_{N\to\infty} D^2\{\chi\} = \frac{1}{2}\sigma^2. \tag{13}$$

The fluctuations of the length of the random vectors $\mathbf{Z}$ are in the range of $\sigma/\sqrt{2}$ (standard deviation). The average length ℓ of $\mathbf{Z}$ given by (12) behaves asymptotically like

$$\ell = \sigma\sqrt{N}. \tag{14}$$

A comparison of $\sigma/\sqrt{2}$ and $\sigma\sqrt{N}$ shows that the deviation from ℓ is small compared to the length. For almost all computations we shall therefore assume that the $\mathbf{Z}$-vectors are lying at the surface of the hypersphere with radius ℓ.

2.2 The $(1+1)$ ES for the Spherical Model

It is the aim of this section to demonstrate the new approach to the $(1+1)$-theory. One advantage of the derivation found by the author [7] is its clearness by taking full advantage of the asymptotical behavior of the mutation vectors $\mathbf{Z}$.

Let the parent $\mathbf{y} = \mathbf{P}$ be at a distance R from the point of optimum $\hat{\mathbf{y}} = \mathbf{O}$ of the spherical domain of success V_R. Due to the spherical symmetry it is not necessary to know the actual position of $\mathbf{P}$ in space, only the distance R is needed. From the point $\mathbf{P}$ offspring $\mathbf{P}+\mathbf{Z}$ are generated. Have a look at such a state $\mathbf{P}+\mathbf{Z}$ by producing a cut of the $N-1$-dimensional surface ∂V_R of the sphere with the two-dimensional plane spanned by the mutation vector $\mathbf{Z}$ and the line segment $\mathbf{PO}$. Fig. 1 shows the essential aspects, the surface ∂V_R projected onto the plane appears as the circle C_R with radius R and center point $\mathbf{O}$.

If $\mathbf{Z}$ is inside of C_R the mutation is accepted because the distance r of $\mathbf{P}+\mathbf{Z}$ to $\mathbf{O}$ is less than R. If $\mathbf{Z}$ is outside C_R we have $r > R$ and therefore the mutation is rejected. The difference $R - r$ is a random variable describing the progress of the ES per generation. Its expectation is the rate of progress defined in (7): $\varphi = E\{R - r\}$.

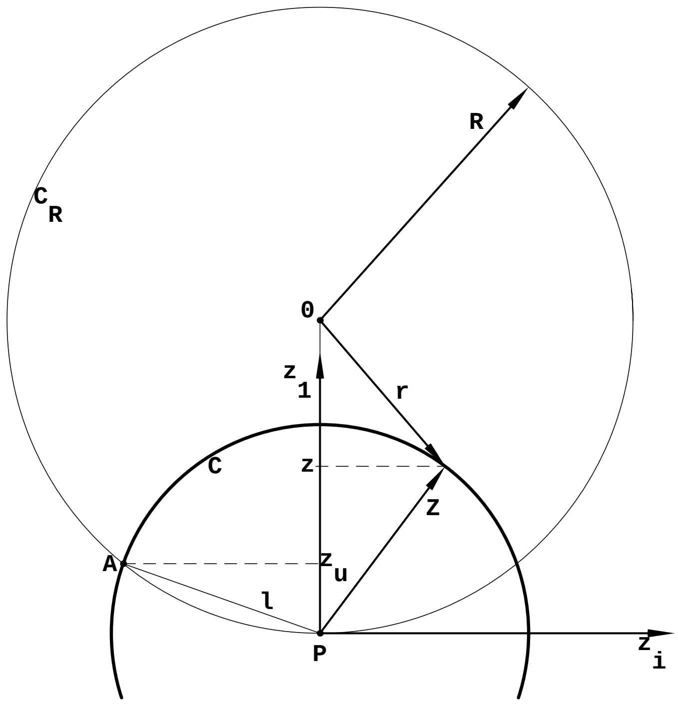

Figure 1: Projection of the surface ∂V_R of a N-dimensional sphere onto the two-dimensional plane spanned by the mutation $\mathbf{Z}$ and the line segment $\mathbf{PO}$.

In order to determine the rate of progress it is necessary to know the density of the random variable $R-r$. Since R is a constant the density of r is searched for. The random variable r depends upon the random vector $\mathbf{Z}$ which lies (as assumed) on the circle C of radius ℓ (cp. the asymptotical formulas (13,14) for the length of the mutation vector $\mathbf{Z}$). The projection of $\mathbf{Z}$ onto the line segment $\mathbf{PO}$ defines a random variable z. The random variable r depends upon the new random variable z. As can be seen from Fig. 1 the relation

$$r^2 = \ell^2 - z^2 + (R-z)^2 \tag{15}$$

holds. Due to the invariance properties (9, 10) of $p(\mathbf{Z})$ one finds for the density $p(z)$

$$p(z) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{z^2}{2\sigma^2}\right). \tag{16}$$

Thus, the rate of progress is given by

$$E\{R-r\} = \frac{R}{\sqrt{2\pi}\,\sigma} \int_{z_u}^{z_o} \left(1 - \sqrt{1 + \frac{\ell^2}{R^2} - \frac{2z}{R}}\right) \exp\left(-\frac{z^2}{2\sigma^2}\right) dz,$$

with the upper limit $z_o = \ell = \sigma\sqrt{N}$. The lower limit is given by point $\mathbf{A}$ in Fig. 1

$$z_u = \frac{\ell^2}{2R}. \tag{17}$$

By the usage of normalized quantities ([3, 7])

$$E^*\{R-r\} = \frac{N}{R}\,E\{R-r\} \qquad \text{and} \qquad \sigma^* = \frac{N}{R}\,\sigma \tag{18}$$

it is now possible to derive for $N \gg 1$ the asymptotical expression (for details see [10])

$$E^*\{R-r\} = \frac{\sigma^*}{\sqrt{2\pi}} \int_{\sigma^*/2}^{\sqrt{N}} \left(t - \frac{\sigma^*}{2}\right) \exp\left(-\frac{1}{2}t^2\right) dt. \tag{19}$$

Getting a deeper insight in (19) it is worth substituting back using (17,18) with the result

$$E\{R-r\} = \int_{z_u}^{\ell} (z - z_u) p(z)\, dz. \tag{20}$$

Now it is obvious that the rate of progress is given by the expectation of the admissible z-deviation $z - z_u \geq 0$ [2] (cf. Fig. 1)

$$E\{R-r\} = E\{z - z_u\}.$$

The integration in (19) is easily carried out and the normalized rate of progress is

$$E^* = E^*\{R-r\} = \frac{\sigma^*}{\sqrt{2\pi}} \exp\left(-\frac{(\sigma^*)^2}{8}\right) - \frac{(\sigma^*)^2}{2}\left[\frac{1}{2} - \Phi_0\left(\frac{\sigma^*}{2}\right)\right], \tag{21}$$

with the *Gauss*-integral[3] [13]

$$\Phi_0(x) := \frac{1}{\sqrt{2\pi}} \int_0^x \exp\left(-\frac{1}{2}t^2\right) dt. \tag{22}$$

Taking into account the definition (22) of Φ_0 and the associated footnote one sees that (21) is equal to the result in [2].

Besides the rate of progress the probability of success P_s is of interest for the (1+1) ES, that is the probability to have a successful (admissible) mutation. Since $p(z)$ is the density for the appearance of a z-value in direction to $\mathbf{O}$ (cf. Fig. 1) and the admissible mutations have to obey the relation $z_u \leq z \leq \ell$ (cf. (20)) one gets for P_s (for $N \to \infty$)

$$P_s = \int_{z_u}^{\ell} p(z)\, dz = \frac{1}{2} - \Phi_0\left(\frac{\sigma^*}{2}\right).$$

[2]In his doctoral thesis [7, 9] the author was misleaded by defining the rate of progress for the spherical model in terms of $E\{z\}$ instead of $E\{z - z_u\}$ discovered by SCHWEFEL [16]. This led to results differing from [2]. It is an intention of this paper to correct the former results [7, 9].

[3]In literature one often finds the so-called error function $erf(x)$ which is connected to the *Gauss*-integral by $\Phi_0(x) = \frac{1}{2} erf\left(\frac{x}{\sqrt{2}}\right)$.

2.3 The $(1 \overset{+}{,} \lambda)$ ES

The new method presented can be easily expanded to the multimembered ES. The characteristic of the $(1 \overset{+}{,} \lambda)$ strategies is the generation of λ offspring by one parent. It is assumed that the parent is located at the point $\mathbf{P}$ (cf. Fig. 1). Obviously, the λ mutation vectors $\mathbf{Z}$ are statistically independent from each other and for each mutation eq. (16) holds independently. The rate of progress is given by the expectation of the $\hat{z} - z_u$-deviation

$$E\{R - r\} = \int_{\hat{z}_u}^{\ell} (\hat{z} - z_u)\hat{p}_\lambda(\hat{z})\, d\hat{z}. \tag{23}$$

Here $\hat{p}_\lambda(\hat{z})$ is the probability density for the appearance of the offspring with the maximum z-value within a generation of λ offspring

$$\hat{p}_\lambda(\hat{z}) = \lambda p(\hat{z})[P(z < \hat{z})]^{\lambda-1} = \frac{d}{d\hat{z}}\,[P(z < \hat{z})]^\lambda \tag{24}$$

with

$$P(z < \hat{z}) = \int_{-\infty}^{\hat{z}} p(z)\, dz = \frac{1}{2} + \Phi_0\left(\frac{\hat{z}}{\sigma}\right), \tag{25}$$

the probability of finding a z-value less than $\hat{z}$ for one trial. To determine the rate of progress (23) it remains to specify the lower limit. For the $(1 + \lambda)$ ES only 'true' improvements are accepted, i. e. $\hat{z}$ must be greater than $z_u = \ell^2/2R$ (cp. (17)). For the $(1, \lambda)$ ES there are no restrictions by the spherical geometry: $\hat{z}_u = -\ell$.

$$(1, \lambda)\ ES \Rightarrow \hat{z}_u = -\ell = -\sigma\sqrt{N}, \qquad (1 + \lambda)\ ES \Rightarrow \hat{z}_u = \frac{\sigma\sigma^*}{2}. \tag{26}$$

Taking the normalization (18) we obtain from (23) with (24,25)

$$E^*\{R - r\} = \sigma^* \int_{\hat{z}_u/\sigma}^{\sqrt{N}} \left(t - \frac{\sigma^*}{2}\right) \frac{d}{dt}\left[\frac{1}{2} + \Phi_0(t)\right]^\lambda dt. \tag{27}$$

2.3.1 The $(1, \lambda)$ Strategy

The result for the $(1, \lambda)$ ES can be achieved by integration of (27) by parts and taking the limit $N \to \infty$, one easily gets the well known result

$$E^*_{1,\lambda} = E^*_{1,\lambda}\{R - r\} = -\frac{(\sigma^*)^2}{2} + c_{1,\lambda}\sigma^* \tag{28}$$

where $c_{1,\lambda}$ is the so-called progress coefficient [11, 12]

$$c_{1,\lambda} = \int_{-\infty}^{\infty} \left\{\frac{1}{2} - \left[\frac{1}{2} + \Phi_0(t)\right]^\lambda\right\} dt. \tag{29}$$

An analytical treatment of (29) is difficult and successful, up to now, for the cases $\lambda = 1, 2, 3, 4, 5$ only (cf. [17], p. 48 or [10], p. 18). The progress coefficient $c_{1,\lambda}$ as a function of λ is shown in Fig. 2. The $c_{1,\lambda}$-coefficient is a slowly increasing function of λ.

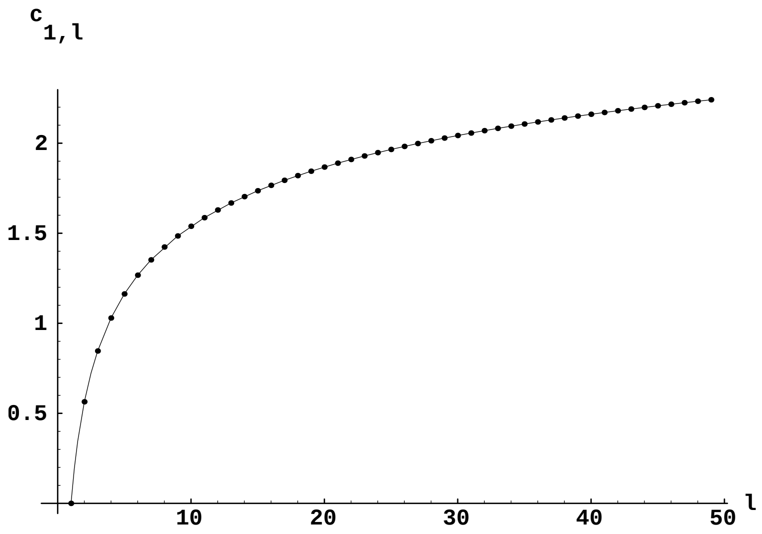

Figure 2: The progress coefficient $c_{1,l}$ defined by eq. (29).

Investigations done by the author have shown that $c_{1,\lambda}$ behaves like

$$c_{1,\lambda} \sim \sqrt{2 \ln \lambda}. \tag{30}$$

As one can derive from (28) the rate of progress run through a maximum at $\hat{\sigma^*} = c_{1,\lambda}$. The maximum rate of progress is therefore $\hat{E^*}_{1,\lambda} = (c_{1,\lambda})^2/2$. Taking the asymptotical behavior (30) into account the maximum rate of progress rises with the logarithm of the number of the descendants λ ($\lambda \to \infty$)

$$\hat{E}^*_{1,\lambda} \sim \ln \lambda.$$

2.3.2 The $(1 + \lambda)$ Strategy

The integral to be determined is given by (26,27). Integration by parts yields

$$E^*_{1+\lambda}\{R - r\} = \sigma^* c_{1+\lambda}(\sigma^*) - \frac{(\sigma^*)^2}{2}\left[1 - \left(\frac{1}{2} + \Phi_0\left(\frac{\sigma^*}{2}\right)\right)^\lambda\right]$$

with the σ^*-dependent progress 'coefficient' $c_{1+\lambda}(x)$

$$c_{1+\lambda}(x) = \frac{\lambda}{\sqrt{2\pi}} \int_{x/2}^{\infty} t \, \exp\left(-\frac{1}{2}t^2\right)\left[\frac{1}{2} + \Phi_0(t)\right]^{\lambda-1} dt. \tag{31}$$

The evaluation of (31) is difficult. For $\lambda = 2$ we obtain the new result

$$c_{1+2}(x) = \sqrt{\frac{2}{\pi}}\left[\exp\left(-\frac{x^2}{8}\right)\left(\frac{1}{2} + \Phi_0\left(\frac{x}{2}\right)\right) + \frac{1}{\sqrt{2}}\left(\frac{1}{2} - \Phi_0\left(\frac{x}{\sqrt{2}}\right)\right)\right].$$

Some cases $\lambda > 2$, treated by numerical integration, are displayed in Fig. 3.

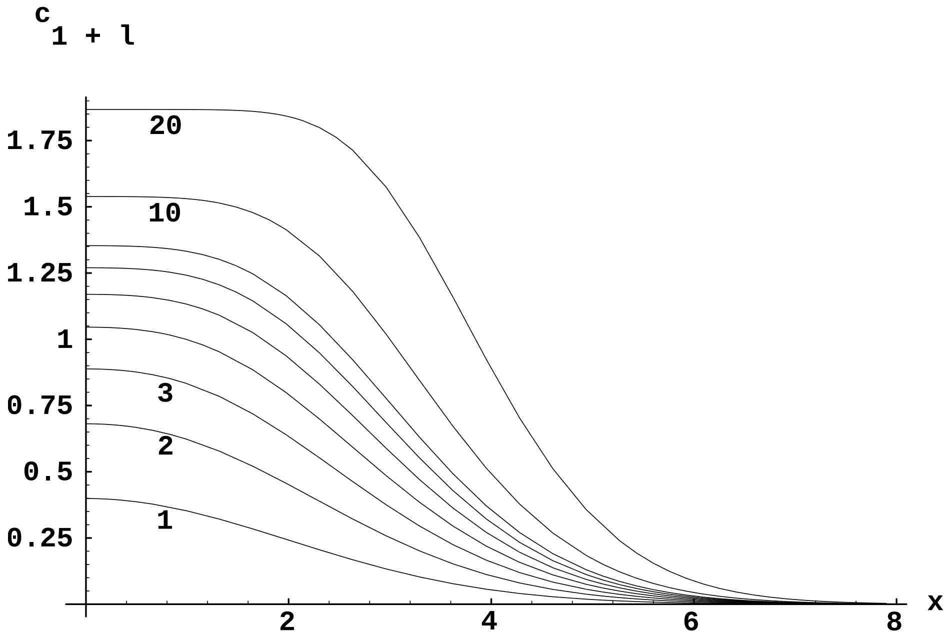

Figure 3: The progress function $c_{1+\lambda}(x)$. For higher λ-values a 'saturation' at $x \approx 0$ is observed. The saturation value is asymptotically equal ($\lambda > 10$) to the progress coefficient of the corresponding $(1, \lambda)$ strategy. The cases $l = \lambda = 1, 2, 3, 4, 5, 6, 7, 10, 20$ are shown.

2.4 The Determination of the Q-Gain

Besides the rate of progress $E\{R - r\}$ the gain in Q-values (fitness gain) or short - the Q-gain - is of certain interest. It is defined as the expectation of the change ΔQ of the Q-function for one generation

$$\overline{\Delta Q} := E\{Q(r) - Q(R)\} = \int_{\hat{z}_u}^{\ell} (Q(r(z)) - Q(R))\, p(z)\, dz. \tag{32}$$

The analysis of (32) using TAYLOR expansions ([10]) reveals for large N the new result

$$\overline{\Delta Q} = -Q' \frac{R}{N} E^*\{R - r\} \qquad \text{with} \qquad Q' = \left.\frac{dQ}{dr}\right|_R. \tag{33}$$

Equation (33) connects the rate of progress $E\{R-r\}$ with the Q-gain linearly. The factor of proportionality is the derivative of Q at R. This result is not a surprise if one takes into account that for $N \to \infty$ the fluctuations become small compared to R. The deviation of Q will be small and the linear approximation holds. For the case $Q_2(r) = c\,r^2$ the result (33) can be derived without TAYLOR expansion [7]

$$\overline{\Delta Q_2} = -\frac{2cR^2}{N} E^*\{R - r\}. \tag{34}$$

3 The Evolution Strategy with Noisy Data

3.1 The Description of the Noise

One of the claimed advantages of the ES is its robustness against disturbances of the determination of the Q-values. Such disturbances especially appear in engineering optimization tasks where the measurement of the Q-value to be optimized is affected by noise, etc. In numerical simulations one observes errors too, e. g., discretization errors in body shape optimizations, or an optimum is searched for which is stable against external disturbances [5, 6]. Up to now a theoretical treatment of noise in the ES was successful for the corridor model $(1 + 1)$ ES only [2].
In accordance to [2] it is assumed that a measurement $\tilde{Q}$ of Q is given by

$$\tilde{Q} = Q + \delta \tag{35}$$

where the (random) noise δ is *unbiased* normally distributed

$$p(\delta) = \frac{1}{\sqrt{2\pi}\,\sigma_\delta} \exp -\frac{1}{2}\left(\frac{\delta}{\sigma_\delta}\right)^2 . \tag{36}$$

The expectation of δ is zero (unbiased) since systematic errors should be excluded:

$$E\{\tilde{Q}\} = E\{Q\} = Q. \tag{37}$$

Substituting (35) into (36) we get

$$p(\tilde{Q}|Q) = \frac{1}{\sqrt{2\pi}\,\sigma_\delta} \exp -\frac{1}{2}\left(\frac{\tilde{Q} - Q}{\sigma_\delta}\right)^2 , \tag{38}$$

i. e. (35,36) define a conditional probability density. The $\tilde{Q}$-values depend on the density $p(Q)$ of the Q-mutations. For the calculations to come it is suitable to change from the *z-picture* used in section 2 to the *Q-picture*.

3.2 The Q-$\tilde{Q}$-Picture

If the function $Q(\mathbf{y})$ to be optimized is a monotonic function of z, then one can transform $p(z)$ to $p(Q)$. The transformation law for density functions yields (see e. g. [14])

$$p(Q) = p(z(Q)) \left|\frac{dz}{dQ}\right| .$$

If one uses for sake of simplicity a quadratic r-dependence of Q taking (15) into account

$$Q \;=\; c(r(z))^2 \;=\; c\,(\ell^2 + R^2 - 2Rz), \tag{39}$$

one obtains with (16) the density for the Q-picture

$$p(Q) = \frac{1}{\sqrt{2\pi}\,2cR\sigma} \exp -\frac{1}{2}\left(\frac{Q - c(\ell^2 + R^2)}{2cR\sigma}\right)^2 . \tag{40}$$

The density for the $\tilde{Q}$-variate can be determined from the conditional density $p(\tilde{Q}|Q(r(z)))$ and the density of the z-fluctuations responsible for the $Q(z)$-density

$$p(\tilde{Q}) = \int_{-\infty}^{\infty} p(\tilde{Q}|Q(z))p(z)\,dz. \tag{41}$$

For the special function (39) the integration of (41) with (15, 16, 38) yields

$$p(\tilde{Q}) = \frac{1}{\sqrt{2\pi}\,\tilde{\sigma}} \exp -\frac{1}{2}\left[\frac{\tilde{Q} - c(\ell^2 + R^2)}{\tilde{\sigma}}\right]^2, \qquad \tilde{\sigma} = \sqrt{(\sigma_\delta)^2 + (2cR\sigma)^2}. \tag{42}$$

As expected $\tilde{Q}$ fluctuates more than Q (cf. 40). However, a direct comparison of σ to σ_δ is less informative. Therefore, normalized values will be introduced for σ and σ_δ. If the normalization (18) is used for σ a conceivable definition of σ_δ^* may be

$$\sigma_\delta^* = \frac{N}{2cR^2}\,\sigma_\delta \tag{43}$$

and $\tilde{\sigma}$ becomes

$$\tilde{\sigma} = \frac{2cR^2}{N}\sqrt{(\sigma^*)^2 + (\sigma_\delta^*)^2}\,. \tag{44}$$

The density (42) holds for the special quadratic Q-function $Q = cR^2$. It is possible to derive a general formula for $p(\tilde{Q})$ [10]. The result is

$$p(\tilde{Q}) = \frac{1}{\sqrt{2\pi}\,\tilde{\sigma}} \exp -\frac{1}{2}\left(\frac{\tilde{Q} - Q'\ell^2/(2R) - Q(R)}{\tilde{\sigma}}\right)^2 \tag{45}$$

with

$$\tilde{\sigma} = \frac{Q'R}{N}\sqrt{(\sigma_\delta^*)^2 + (\sigma^*)^2} \qquad \text{and} \qquad \sigma_\delta^* = \frac{N}{Q'R}\sigma_\delta. \tag{46}$$

The comparison of (45,46) with (42–44) reveals a correspondence between an arbitrary Q-function and the special case cr^2 displayed in Table 1.

cr^2	$2cR$	cR^2	$c\ell^2$	$\frac{2cR^2}{N}$
$Q(r)$	$Q' = \frac{dQ}{dr}\big\vert_R$	$Q(R)$	$Q'\frac{\ell^2}{2R}$	$Q'\frac{R}{N}$

Table 1: Correspondences between $Q(r) = cr^2$ and an arbitrary $Q(r)$-function for $\lim_{N\to\infty} R/N = 0$.

For sake of simplicity the following considerations will be performed for the case $Q = cr^2$ only. The probability density (42) is the starting point to calculate the density $p_\lambda(\tilde{Q})$ for the appearance of the best $\tilde{Q}$-value within a number of λ trials (in eq. (47) denoted by $\tilde{Q}_\lambda$). The formulas are similar to (24, 25). In the case of minimization the probability $P(\tilde{Q} > \tilde{Q}_\lambda)$ has to be used instead of (25). Taking (42) into account one obtains

$$P(\tilde{Q} > \tilde{Q}_\lambda) \;=\; \int_{\tilde{Q}_\lambda}^{\infty} p(\tilde{Q})\,dQ \;=\; \frac{1}{2} - \Phi_0\left(\frac{\tilde{Q}_\lambda - c(\ell^2 + R^2)}{\tilde{\sigma}}\right). \tag{47}$$

For the best $\tilde{Q}$-value we get therefore (writing $\tilde{Q}$ instead of $\tilde{Q}_\lambda$)

$$p_\lambda(\tilde{Q}) = -\frac{d}{d\tilde{Q}}\left[\frac{1}{2} - \Phi_0\left(\frac{\tilde{Q} - c(\ell^2 + R^2)}{\tilde{\sigma}}\right)\right]^\lambda. \tag{48}$$

This density will be used to calculate the 'virtual' Q-gain $E\{\Delta\tilde{Q}\}$ of the $(\tilde{1},^+\tilde{\lambda})$ ES.

3.3 The Determination of the 'virtual' Q-Gain

The 'virtual' Q-gain $E\{\Delta\tilde{Q}\}$ is defined in accordance to the undisturbed case (32)

$$E\{\Delta\tilde{Q}\} := E\{\tilde{Q}(r) - \tilde{Q}(R)\}. \tag{49}$$

If we take into account that for a fixed R-value eq. (37) holds we can write

$$E\{\Delta\tilde{Q}\} = E\{\tilde{Q}(r) - Q(R)\}. \tag{50}$$

A trial is treated as a successful mutation (in case of Q-minimization) if the virtual Q-value $\tilde{Q}$ is less than the expected (and undisturbed) Q-value at R, i. e. $\tilde{Q} < Q(R)$. The observed $\Delta\tilde{Q}$-gain (in the case of minimization: negative) does not necessarily imply that the mutation has brought an advance to the optimum point taking the z-picture as a basis, i. e. $z > R^2/2l$ does not necessarily hold since the real Q-value is disturbed by noise. That is why the author has denoted eq. (49) as the *virtual* Q-gain.
In the case of a quadratic Q-function $Q = cr^2$ it follows from eq. (50)

$$E_\lambda\{\Delta\tilde{Q}\} = \int_{c(R-\ell)^2}^{Q_0} (\tilde{Q} - cR^2) p_\lambda(\tilde{Q})\, d\tilde{Q} \tag{51}$$

with

$$(1,\lambda)\ ES:\ Q_0 = c(R+\ell)^2 \qquad \text{and} \qquad (1+\lambda)\ ES:\ Q_0 = cR^2,$$

respectively. If we insert (48) into (51) and change to normalized values we obtain finally for the $(1,\lambda)$ ES

$$E_{1,\lambda}\{\Delta\tilde{Q}\} = -\frac{2cR^2}{N}\left\{-\frac{(\sigma^*)^2}{2} + c_{1,\lambda}\sqrt{(\sigma_\delta^*)^2 + (\sigma^*)^2}\right\}.$$

For the $(1+\lambda)$ ES a similar procedure delivers

$$\begin{aligned} E_{1+\lambda}\{\Delta\tilde{Q}\} \quad = \quad -\frac{2cR^2}{N} & \left\{\sqrt{(\sigma_\delta^*)^2 + (\sigma^*)^2}\, c_{1+\lambda}\left(\frac{(\sigma^*)^2}{\sqrt{(\sigma_\delta^*)^2 + (\sigma^*)^2}}\right)\right. \\ & \left. -\frac{(\sigma^*)^2}{2}\left[1 - \left(\frac{1}{2} + \Phi_0\left(\frac{1}{2}\frac{(\sigma^*)^2}{\sqrt{(\sigma_\delta^*)^2 + (\sigma^*)^2}}\right)\right)^\lambda\right]\right\} \end{aligned}$$

there $c_{1+\lambda}(x)$ is the progress function (31) defined in section 2.3.2.

3.4 The $(\tilde{1}, \tilde{\lambda})$-Theory

It is the goal of the theory presented here to find a description for the *real* rate of progress. This can be done both in the z-picture and in the Q-picture. Due to eq. (34) it is possible to convert the Q-gain into the rate of progress and vice versa. We take the approach using the real Q-gain

$$\overline{\Delta Q} = E\{Q - cR^2\} = \int (Q - cR^2) p_{1,\lambda}(Q)\, dQ$$

$$\overline{\Delta Q} = \lambda \int_{c(R-\ell)^2}^{c(R+\ell)^2} (Q - cR^2)\, p(Q|cR^2)\, P\left(\tilde{Q}|Q < \left\{\tilde{Q}_2, \tilde{Q}_3, \ldots, \tilde{Q}_\lambda\right\}\Big|\, cR^2\right) dQ. \tag{52}$$

The probability density $p_{1,\lambda}(Q)$ in its form (52) can be derived from the following considerations: Let the first trial within a generation of λ offspring starting from the initial parental state $Q(R) = cR^2$ produce a z-value uniquely corresponding to a certain Q-value (cf. Fig. 1). The probability density $p(Q|cR^2)$ for such a Q is given by formula (40). In order to survive within the $(\tilde{1}, \tilde{\lambda})$ ES this trial must have a measured (and therefore disturbed) $\tilde{Q}$-value which has to be in the case of minimization less than all the other $\lambda - 1$ measured $\tilde{Q}$-values produced by mutation from the state cR^2. The probability for that was denoted by

$$P\left(\tilde{Q}|Q < \left\{\tilde{Q}_2, \tilde{Q}_3, \ldots, \tilde{Q}_\lambda\right\}\Big|\, cR^2\right) =: P_{1,\lambda}(Q) \tag{53}$$

in (52). Since the consideration presented above holds not only for the first trial but also for the other $\lambda - 1$ trials, we have λ independent cases with equal probability and therefore the factor λ before the integral in (52). The next goal is to determine the probability $P_{1,\lambda}(Q)$. It will be constructed from the statistically independent densities $p(\tilde{Q}|Q)$ and $p_{\lambda-1}(\hat{Q}|cR^2)$. Here, $p(\tilde{Q}|Q)$ is the conditional density describing the appearance of a $\tilde{Q}$-value under the condition Q, it is given by eq. (38). The density $p_{\lambda-1}(\hat{Q}|cR^2)$ represents the distribution of the best $\tilde{Q}$-value among $\lambda - 1$ trials all of them starting at the parental state $Q(R) = cR^2$. This probability was already obtained, we get from (48)

$$p_{\lambda-1}(\hat{Q}|cR^2) = -\frac{d}{d\hat{Q}} \left[\frac{1}{2} - \Phi_0\left(\frac{\hat{Q} - c(\ell^2 + R^2)}{\tilde{\sigma}}\right)\right]^{\lambda-1}. \tag{54}$$

Due to the statistical independence of $\hat{Q}$ and $\tilde{Q}$ we have for the joint density $p(\tilde{Q}, \hat{Q})$

$$p(\tilde{Q}, \hat{Q}) = p(\tilde{Q}|Q)\, p_{\lambda-1}(\hat{Q}|cR^2). \tag{55}$$

To get the probability (53) one has to determine the probability of $\tilde{Q}$ under the condition $-\infty \leq \hat{Q} \leq \infty$ and $-\infty \leq \tilde{Q} \leq \hat{Q}$, i. e. one has to determine the integral

$$P_{1,\lambda}(Q) = \int_{-\infty}^{\infty} \int_{-\infty}^{\hat{Q}} p(\tilde{Q}, \hat{Q})\, d\tilde{Q}\, d\hat{Q}. \tag{56}$$

The integration of (56) with (38, 54) and the treatment of (52) taking (40) into account yields after a longer calculation (presented in [10]) the real Q-gain of the $(\tilde{1}, \tilde{\lambda})$ ES

$$\overline{\Delta Q}_{1,\lambda} = -\frac{2cR^2}{N} \left\{-\frac{(\sigma^*)^2}{2} + c_{1,\lambda} \frac{(\sigma^*)^2}{\sqrt{(\sigma_\delta^*)^2 + (\sigma^*)^2}}\right\}. \tag{57}$$

Taking the relation (34) into account one gets for the rate of progress from (57)

$$E^*_{1,\lambda}\{R-r\} = -\frac{(\sigma^*)^2}{2} + c_{1,\lambda}\,\frac{(\sigma^*)^2}{\sqrt{(\sigma^*_\delta)^2+(\sigma^*)^2}}\,. \tag{58}$$

A comparison with (28) shows that the case of vanishing noise is included in (58).

Fig. 4 depicts some curves from the $(\tilde{1},\tilde{5})$ ES. One observes a decrease of the progress rate with the growing noise level. If the relative noise level σ^*_δ crosses a certain threshold the $(\tilde{1},\tilde{5})$ ES cannot converge at all (this dynamical aspect of the ES will be discussed in section 3.7).

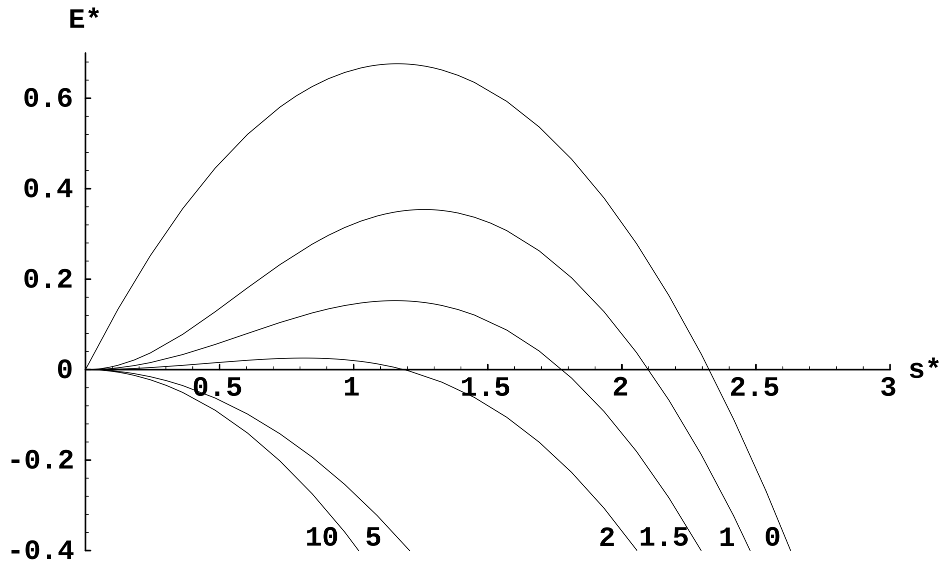

Figure 4: The normalized rate of progress for the noisy $(\tilde{1},\tilde{5})$ ES (NB $\sigma^* =: s^*$). Parameter is the normalized noise strength σ^*_δ: $\sigma^*_\delta = 0, 1, 1.5, 2, 5, 10$.

3.5 The $(\tilde{1}+\tilde{\lambda})$-Theory

In order to calculate the real progress rate for the $(\tilde{1}+\tilde{\lambda})$ ES we will proceed in the same manner as for the $(\tilde{1},\tilde{\lambda})$ theory. The rate of progress in the Q-picture (the Q-gain) is given by

$$\overline{\Delta Q} = \lambda \int_{c(R-\ell)^2}^{c(R+\ell)^2} (Q-cR^2)\, p(Q|cR^2)\, P_{1+\lambda}(Q)\, dQ. \tag{59}$$

The λ-factor in (59) has the same foundation as for formula (52) of the $(\tilde{1},\tilde{\lambda})$ ES. It is striking that even the upper limit of the integral is the same as for the (,)-strategy in contrast to the undisturbed $(1+\lambda)$ ES (cp. (26,27)). The upper limit takes into account that Q-values with $Q > cR^2$ may survive since not only the measured offspring's fitness

Q_o is disturbed but also the measured parental Q-value Q_p is disturbed too. The selection relys on the criterion $\tilde{Q}_o \le \tilde{Q}_p$.

How can $P_{1+\lambda}$ be derived?

For the $(\tilde{1}, \tilde{\lambda})$-strategy the probability $P_{1,\lambda}(Q)$ was calculated from the joint density $p(\tilde{Q}, \hat{Q})$ eq. (55). There were two independent variats: the $\tilde{Q}$-variat for the first trial and the variat $\hat{Q}$ standing for the second best trial. Now we have three players. The third independent variat comes from the measured parental fitness $\tilde{Q}_p$ with the density $p(\tilde{Q}_p|cR^2)$. The joint density is

$$p(\tilde{Q}, \hat{Q}, \tilde{Q}_p) = p(\tilde{Q}|Q)\, p_{\lambda-1}(\hat{Q}|cR^2)\, p(\tilde{Q}_p|cR^2). \tag{60}$$

If we assume that the parental $Q = cR^2$ is affected by the same 'kind of noise' as the measured values of the offspring, then formula (38) holds for $p(\tilde{Q}_p|cR^2)$ too

$$p(\tilde{Q}_p|cR^2) = \frac{1}{\sqrt{2\pi}\,\sigma_\delta} \exp -\frac{1}{2}\left(\frac{\tilde{Q}_p - cR^2}{\sigma_\delta}\right)^2.$$

The following considerations are performed in the space of the $\tilde{Q}, \hat{Q}, \tilde{Q}_p$-variats. Under what conditions survives a measured $\tilde{Q}$-mutation? Surely if (provided a minimization task is in consideration) $\tilde{Q} \le \hat{Q} \le \tilde{Q}_p$ holds. But also the disjoint case $\tilde{Q} \le \tilde{Q}_p \le \hat{Q}$ describes an admissible mutation. These two cases define the domains (A) and (B) in the space of the $\tilde{Q}, \hat{Q}, \tilde{Q}_p$-variats to be integrated over to get $P_{1+\lambda}(Q)$:

$$P_{1+\lambda}(Q) = P_A(Q) + P_B(Q)$$

with

$$P_A(Q) = \int_{-\infty}^{\infty}\int_{\hat{Q}}^{\infty}\int_{-\infty}^{\hat{Q}} p(\tilde{Q}|Q)\, p_{\lambda-1}(\hat{Q}|cR^2)\, p(\tilde{Q}_p|cR^2)\, d\tilde{Q}\, d\tilde{Q}_p\, d\hat{Q}$$

and

$$P_B(Q) = \int_{-\infty}^{\infty}\int_{-\infty}^{\hat{Q}}\int_{-\infty}^{\tilde{Q}_p} p(\tilde{Q}|Q)\, p_{\lambda-1}(\hat{Q}|cR^2)\, p(\tilde{Q}_p|cR^2)\, d\tilde{Q}\, d\tilde{Q}_p\, d\hat{Q}.$$

Thus one can find after a cumbersome calculation an expression for the real Q-gain (59) and finally the rate of progress

$$\begin{aligned} E^*_{1+\lambda} \;&=\; \frac{\lambda\,(\sigma^*)^2}{\sqrt{2\pi}\sqrt{(\sigma^*_\delta)^2 + (\sigma^*)^2}} \int_{-\infty}^{\infty} t \exp\left(-\frac{1}{2}t^2\right)\left[\frac{1}{2} + \Phi_0(t)\right]^{\lambda-1} \times \\ &\times \left[\frac{1}{2} + \Phi_0\left(\frac{\sqrt{(\sigma^*_\delta)^2 + (\sigma^*)^2}}{\sigma^*_\delta}\, t - \frac{(\sigma^*)^2}{2\sigma^*_\delta}\right)\right] dt \;-\; \frac{(\sigma^*)^2}{2}\, P_{s\,\lambda}(\sigma^*, \sigma^*_\delta), \end{aligned} \tag{61}$$

with the parental death probability $P_{s\,\lambda}(\sigma^*, \sigma^*_\delta)$ (for details cf. [10])

$$P_{s\,\lambda} = 1 \;-\; \frac{\sqrt{(\sigma^*_\delta)^2 + (\sigma^*)^2}}{\sqrt{2\pi}\,\sigma^*_\delta} \int_{-\infty}^{\infty} \exp -\frac{1}{2}\left(\frac{\sqrt{(\sigma^*_\delta)^2 + (\sigma^*)^2}}{\sigma^*_\delta}\, t - \frac{(\sigma^*)^2}{2\sigma^*_\delta}\right)^2 \left[\frac{1}{2} + \Phi_0(t)\right]^{\lambda} dt. \tag{62}$$

The analytical calculation of the integrals (61) and (62) is difficult and up to now the author succeeded only with the $(\tilde{1} + \tilde{1})$ ES presented in the next section. Due to the analytical difficulties we have to rely on results gathered from numerical integrations. As an example the case $\lambda = 5$ is displayed in Fig. 5.

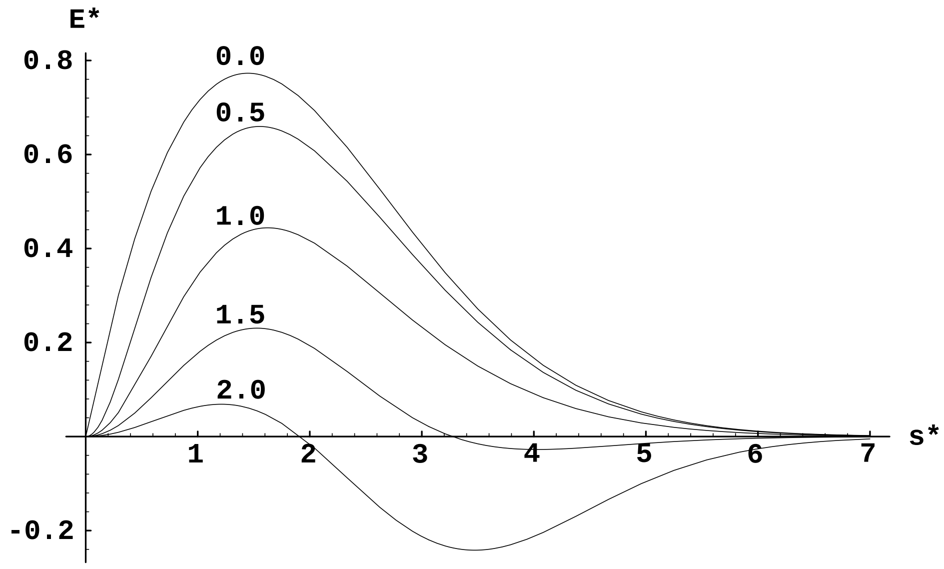

Figure 5: The normalized rate of progress vs. the normalized mutation strength $\sigma^* =: s^*$ for the noisy $(\tilde{1}+\tilde{5})$ ES. Parameter is the normalized noise strength σ^*_δ: $\sigma^*_\delta = 0, 0.5, 1, 1.5, 2$.

3.6 Analytical results to the $(\tilde{1}+\tilde{1})$ ES

3.6.1 The Rate of Progress and the Success Probability

For the $(\tilde{1}+\tilde{1})$ ES it is possible to calculate the integrals (61,62) analytically [10]. We find for the probability of success

$$P_s = \frac{1}{2} - \Phi_0\left(\frac{(\sigma^*)^2}{2\sqrt{2(\sigma^*_\delta)^2+(\sigma^*)^2}}\right). \tag{63}$$

The result for $E^*_{\tilde{1}+\tilde{1}}$ is

$$E^*_{\tilde{1}+\tilde{1}} = \frac{\sigma^*}{\sqrt{2\pi}}\,\frac{\sigma^*}{\sqrt{2(\sigma^*_\delta)^2+(\sigma^*)^2}}\,\exp\left[-\frac{(\sigma^*)^2}{8}\,\frac{(\sigma^*)^2}{2(\sigma^*_\delta)^2+(\sigma^*)^2}\right] - \frac{(\sigma^*)^2}{2}\,P_s\,. \tag{64}$$

Fig. 6 displays the results. As one can see in contrast to the undisturbed $(1+1)$ ES (cp. Fig. 6, the curve with $\sigma^*_\delta = 0$) the rate of progress may be negative. In general the rate decreases if one raises the noise level σ^*_δ. There is a certain upper limit $\check{\sigma}^*_\delta$ above which there is in average no evolutionary progress at all. Obviously, the limit is given at the point of vanishing (positive) maximum of $E^*(\sigma^*_\delta)$. A numerical analysis has shown that the limit is at

$$\check{\sigma}^*_\delta \approx 1.368. \tag{65}$$

To have a positive rate of progress σ^* must have a sufficient high value. E. g., one can infer from Fig. 6 for the case $\sigma^*_\delta = 1.25$ the mutation strength should be greater than 2.2.

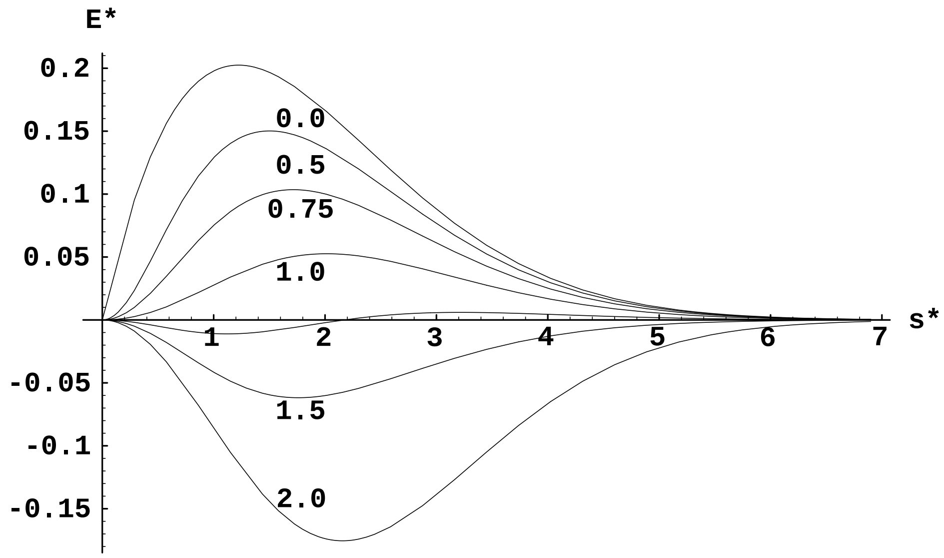

Figure 6: The normalized rate of progress for the noisy $(\tilde{1}+\tilde{1})$ ES (NB $\sigma^* =: s^*$). Parameter is the normalized noise strength σ_δ^*: $\sigma_\delta^* = 0, 0.5, 0.75, 1, 1.25, 1.5, 2$.

In other words, the region of reliable convergence is somewhat below the absolute limit $\check{\sigma}_\delta^*$. We denote this second limit by $\hat{\sigma}_\delta^*$. It is determined by the vanishing of the negative minimum of $E^*(\sigma^*)$ as one can see from Fig. 6. The limit is

$$\hat{\sigma}_\delta^* = \frac{2}{\sqrt{\pi}} \approx 1.12838. \tag{66}$$

For the region $\hat{\sigma}_\delta^* \leq \sigma_\delta^* \leq \check{\sigma}_\delta^*$ convergence or divergence of the ES depends on σ^*. What kind of σ-control should be used to get an (almost) optimum rate of progress?

3.6.2 The σ-Control and the 1/5-Rule

For the (1+1) ES Rechenberg [2] has given a *rule of thumb* how to control the mutation strength σ - the so-called 1/5-*rule* - the control of the mutation strength by tuning the probability of success P_s. Fig. 7 displays this probability (63) for the noisy $(\tilde{1}+\tilde{1})$ ES depending upon the normalized mutation strength σ^* for different levels of noise.

As one can see from Fig. 7 P_s is a monotonically decreasing function of the mutation strength σ^* (provided σ_δ^* to be a constant). Therefore eq. (63) can be inverted to σ^*. If we insert the result into the formula for the progress rate (64) we get the normalized rate of progress depending upon the probability of success displayed in Fig. 8.

For all $\sigma_\delta^* < \check{\sigma}_\delta^*$ there is an optimum probability of success guaranteeing a maximum E^*. The special case $\sigma_\delta^* = 0$ was already treated in [2]. In accordance to this early result we can find in Fig. 8 an optimum probability $P_s \approx 0.27$.

Fig. 9 shows the optimum success probability $\hat{P}_s$ depending upon the normalized noise strength σ_δ^*. One can see that the probability of success for an optimum working of the

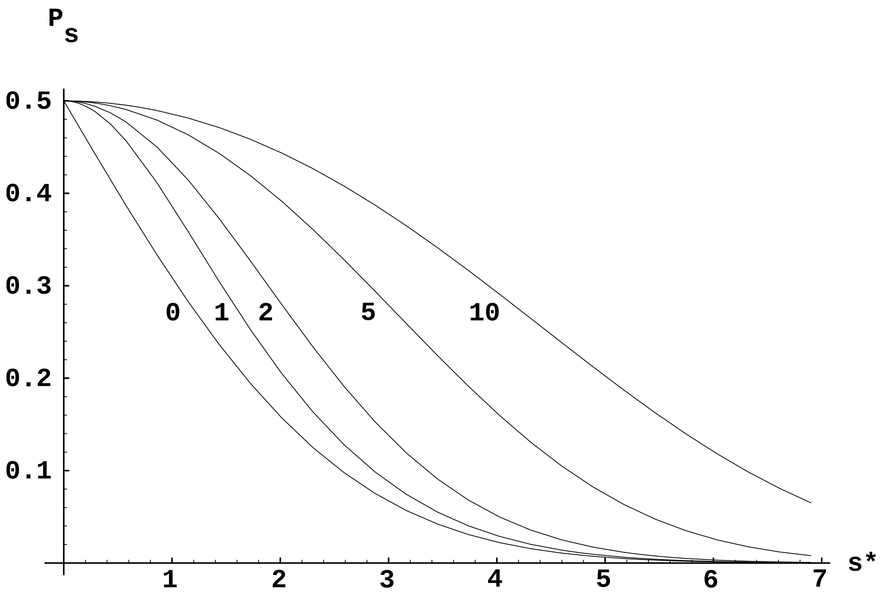

Figure 7: The probability of success P_s for the noisy spherical $(\tilde{1} + \tilde{1})$ ES versus the normalized mutation strength $\sigma^* =: s^*$ with the normalized noise level σ^*_δ as a parameter: $\sigma^*_\delta = 0, 1, 2, 5, 10$.

$(\tilde{1} + \tilde{1})$ ES should obey the condition

$$P_s < 0.27.$$

The case $\hat{P}_s = 1/5$ is in the 'moderate' range, i. e. the 1/5-rule applied to the noisy ES yields acceptable results. (To be exact $P_s = 1/5$ is the optimum probability for the normalized noise strength $\sigma^*_\delta \approx 1.02$, see Fig. 9.) RECHENBERG's proposal $\hat{P}_s = 1/10$ for the noisy ES (cf. [2], p. 123), however, works unsatisfactory: If one compares the resulting progress rates for an admissible noise level (in order to have convergence) one has an increase in the admissible noise level of about 20% only. If we are not yet in the critical range $\sigma^*_\delta > \hat{\sigma}^*_\delta$ which is bounded by $\sigma^*_\delta < \check{\sigma}^*_\delta$ (cf. eq. (65)), anyway, the choice $P_s = 1/10$ first of all yields a decrease in the progress rate. It will be better to decrease the relative (normalized) noise strength by suitable methods (cf. next section).

How does the σ-control with the 1/5-rule work?

The algorithm starts with an estimated σ-value. Taking this strategy parameter a sufficient number of evolution steps (generations) k is carried out as described in section 1.1. Let k_s be the number of the successful mutations an estimate for the probability of success is

$$P_s \approx \frac{k_s}{k}.$$

Using eq. (63) it is possible to estimate σ^* and with (18) it follows

$$\sigma^*(P_s) = \frac{N}{R}\,\sigma. \qquad (67)$$

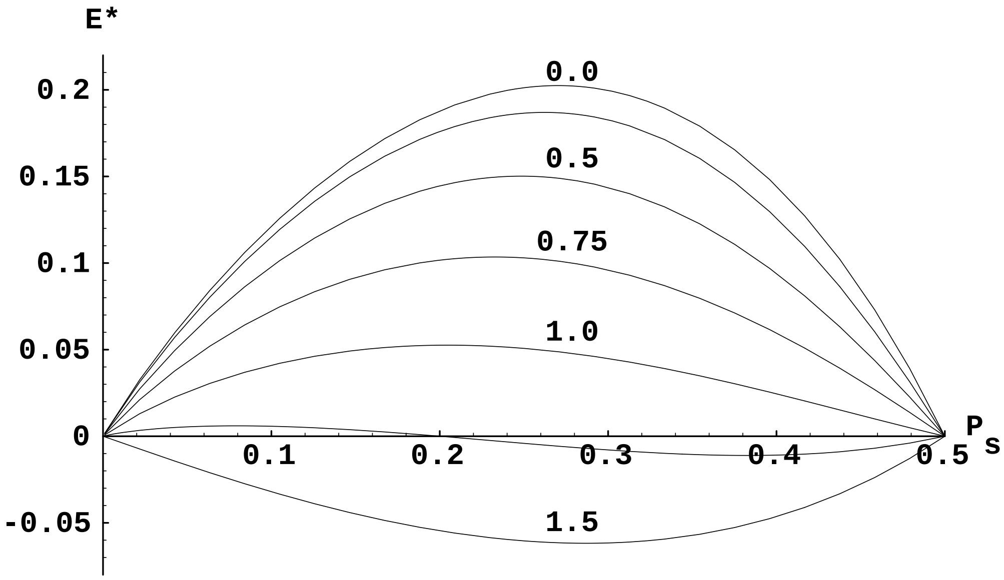

Figure 8: The normalized rate of progress $E^*_{\tilde{1}+\tilde{1}}$ versus the success probability P_s with the normalized noise level σ^*_δ as parameter, $\sigma^*_\delta = 0, 0.25, 0.5, 0.75, 1, 1.25, 1.5$.

Suppose the strategy parameter is at optimum, then the equation

$$\sigma^*(\hat{P}_s) = \frac{N}{R}\,\sigma_{opt} \tag{68}$$

would hold. The combination of (67, 68) yields for the optimum σ-value

$$\sigma_{opt} = \frac{\sigma^*(\hat{P}_s)}{\sigma^*(P_s)}\,\sigma,$$

provided that the noise strength is known. Due to the monotony of $P_s(\sigma^*)$ (cp. Fig. 7) we have

$$\begin{aligned} \frac{\sigma^*(\hat{P}_s)}{\sigma^*(P_s)} < 1 \quad &\Leftrightarrow \quad P_s < \hat{P}_s \\ \frac{\sigma^*(\hat{P}_s)}{\sigma^*(P_s)} > 1 \quad &\Leftrightarrow \quad P_s > \hat{P}_s. \end{aligned}$$

I. e., the parameter σ is to increase, if the measured success probability P_s is greater than $\hat{P}_s$ and to decrease, if P_s is less than the given $\hat{P}_s$, respectively. That is the message of the 1/5-rule:

$$\left.\begin{array}{l} P_s < 1/5 \;\Rightarrow\; \sigma := \sigma/a \\ P_s > 1/5 \;\Rightarrow\; \sigma := \sigma\,a \end{array}\right\}. \tag{69}$$

If there is not an extreme deviation of P_s from $\hat{P}_s$, then the values of a will be in a moderate range $a = 1 \ldots 2$. The influence of the noise strength σ^*_δ may be small and one

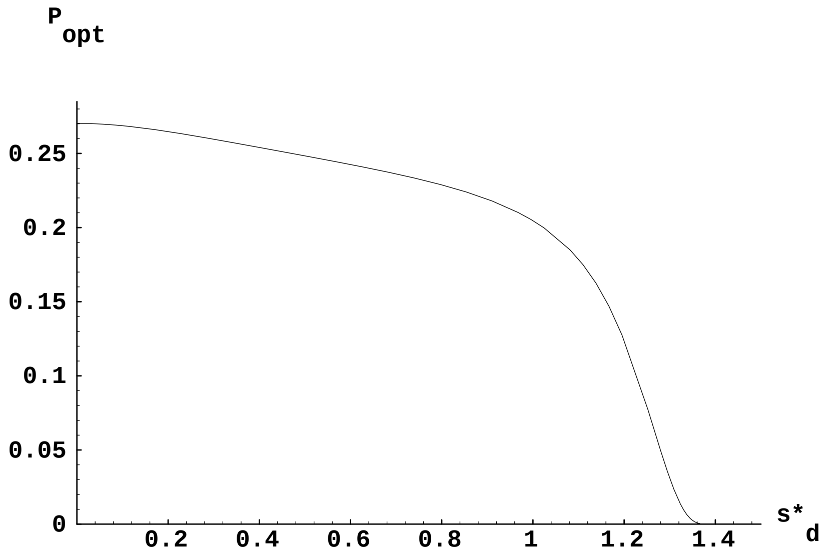

Figure 9: The optimum probability of success $\hat{P}_s =: P_{opt}$ for the $(\tilde{1} + \tilde{1})$ ES versus the normalized noise strength $\sigma_\delta^* =: s_d^*$.

can assume $a = const.$ This somehow 'fuzzy' argumentation holds especially for those cases where the real domain of success of the function Q to be optimized is unknown, i. e. one has rather to guess the optimum strategy parameter than to use the exact values from the spherical model.

3.7 Some Dynamical Aspects of the Evolution Strategy

In this section some applications of the progress rate to dynamical aspects of the ES will be discussed. The rate of progress describes the average progress towards the optimum within a generation, whereas the notion *dynamics* means the description of the distance to the optimum over an arbitrary range of generations. (The rate of progress describes the 'microscopic aspects' of the ES, whereas the dynamics stands for the 'macroscopic aspects' of the ES.) One goal of dynamics is to calculate $R(g)$, i. e. the distance to the optimum as a function of the generation number g with the initial (start) condition $R(g = 0) =: R_0$. The relation between an arbitrary state $R(g)$ at generation g and the next generation $R(g + 1) = \overline{r}$ is given by eq. (18)

$$R(g+1) - R(g) = \Delta R = \overline{r} - R(g) = -E^* \frac{R(g)}{N}.$$

This equation is asymptotic equivalent to the differential equation

$$\frac{dR}{dg} = -E^*(\sigma^*, \sigma_\delta^*)\ \frac{R}{N}.$$

If one takes the relations (18) and (43) into account it can be seen

$$\frac{dR}{dg} = -E^*\left(\sigma \frac{N}{R},\ \sigma_\delta \frac{N}{2cR^2}\right)\ \frac{R}{N} = -f(R;\, \sigma, \sigma_\delta) \tag{70}$$

that this (in general) *non-linear* differential equation depends not only on the R-value but also on the mutation strength σ. In the general case σ itself obeys a dynamics depending on R. Therefore, for a general treatment it would be necessary to formulate a differential equation for the mutation strength σ. This is a future task including a 'microscopic' analysis analogous to the treatment of the progress rate for the R-dynamics.

At this point it is only possible to discuss the $(\tilde{1} + \tilde{1})$ ES with P_s-control and in a second section to illuminate some basic aspects of the stationary $(\tilde{1}, \tilde{\lambda})$ strategy.

3.7.1 The Dynamics of the $(\tilde{1} + \tilde{1})$ ES with P_s-Control

It is assumed that there is a σ-P_s-control mechanism (cp. section 3.6) guaranteeing the relation

$$P_s = const.$$

If we insert (63) and (64) into the differential equation (70) we have

$$\frac{dR}{dg} = -\frac{R}{N}\left\{\frac{\kappa}{\sqrt{2\pi}}\exp\left(-\frac{\kappa^2}{8}\right) - \frac{P_s}{2}\left[\frac{\kappa^2}{2} + \sqrt{\frac{\kappa^4}{4} + 2\,\kappa^2\,(\sigma_\delta^*)^2}\right]\right\} = -\frac{R}{N}E^*(P_s, \sigma_\delta^*) \quad (71)$$

with

$$\kappa = 2\,\Phi_0^{-1}\left[\frac{1}{2} - P_s\right] = const. > 0$$

where $\Phi_0^{-1}(x)$ denotes the inverse function to the Gauss integral (cp. eq. (22)). Assuming a constant relative noise level

$$\sigma_\delta^* = const. \quad (72)$$

equation (71) can be solved

$$R(g) = R_0 \exp\left(-\frac{E^*}{N}g\right).$$

If $E^* > 0$ holds, then the ES converges exponentially with the number of generations g to the point of optimum. Fig. 8 and Fig. 9, respectively, can be used to estimate the P_s, σ_δ^*-convergence range. The limit of reliable convergence is given by relation (66). For that range there are no restrictions on the P_s.
To the case $E^* < 0$. If condition (72) is preserved

$$\sigma_\delta^* = \sigma_\delta \frac{N}{2cR^2} = const. \quad (73)$$

we have $R \to \infty$ for $g \to \infty$, i. e. the ES diverges (this statement holds only for a quadratic Q-function, otherwise one has to take the correspondence Table 1). Eq. (73) can be interpreted in the following way: Since $Q = cR^2$ to fulfil (73) we have to postulate $\sigma_\delta = Q(R)\,\delta = cR^2\,\delta$. The standard deviation σ_δ of the measuring error must be proportional to the expectation of the measured value. The value δ is therefore the *relative measuring error*

$$\delta := \frac{\sigma_\delta}{Q(R)}. \quad (74)$$

For the $(\tilde{1}+\tilde{1})$ ES there is a maximum relative measuring error determining the limit of convergence. With eq. (66) we find the limit for reliable convergence

$$\delta \leq \frac{4}{\sqrt{\pi}\, N}. \tag{75}$$

To have reliable convergence above limit (75) the only method is to repeat the measurement, say m times, keeping the state $\mathbf{y}_g^l$ of the offspring constant (cf. eq. (1,2)) and take the average

$$Q_m := \frac{1}{m} \sum_{i=1}^{m} \tilde{Q}(\mathbf{y}), \qquad \mathbf{y} = const. \tag{76}$$

The selection (5,6) must be performed, naturally, at the stage of the Q_m-values. Since σ_δ is the standard deviation of the $\tilde{Q}$-values the standard deviation of the average (76) is given by

$$\sigma\{Q_m\} = \frac{\sigma_\delta}{\sqrt{m}} \tag{77}$$

(see e. g. [14], p. 218f). Thus we get instead of (75)

$$\delta \leq \frac{4\sqrt{m}}{\sqrt{\pi}\, N}.$$

By enlarging the number of measurements m it is now possible to enforce the convergence of the $(\tilde{1}+\tilde{1})$ ES.

If the condition (73) is not fulfilled the dependence of the noise strength σ_δ^* upon the R-value has to be taken into account in eq. (71). The steady state R_∞ can be determined from (71) by the condition $\frac{dR}{dg} = 0$, i. e., in general a *non-zero distance* $R_\infty \neq 0$ *to the real optimum* remains at the end of the evolution process (even though σ converges $\sigma \to 0$, for details cf. [10]).

3.7.2 Some Basic Properties of the $(\tilde{1}, \tilde{\lambda})$ ES

Due to eq. (70) with (58) the R-dynamics of the $(\tilde{1}, \tilde{\lambda})$ ES is governed by the differential equation

$$\frac{dR}{dg} = \sigma^2 \left[\frac{N}{2R} - \frac{2cRc_{1,\lambda}}{\sqrt{\sigma_\delta^2 + (2cR\sigma)^2}} \right]. \tag{78}$$

Since up to now a theory describing the dynamics of the mutation strength σ does not exist it is only possible to investigate the steady state behavior and general convergence properties.

Convergence of the ES is given, if at first

$$\frac{dR}{dg} < 0 \tag{79}$$

and at second

$$\frac{dR}{dg} \to 0 \quad \text{for} \quad g \to \infty \tag{80}$$

holds. Provided that $\sigma < \infty$ we get from (78,79) the

$$\underline{\textit{sufficient evolution condition:}} \quad \sigma_\delta^2 + (2cR\sigma)^2 < \left(\frac{4cR^2 c_{1,\lambda}}{N}\right)^2. \tag{81}$$

If we assume the convergence of σ, i. e. $\lim_{g\to\infty} \sigma = 0$, then we can derive the

$$\underline{\textit{necessary evolution condition:}} \quad \sigma_\delta < \frac{4cR^2 c_{1,\lambda}}{N} \tag{82}$$

and on the supposition of a relative error δ (74)

$$\delta < \frac{4c_{1,\lambda}}{N}. \tag{83}$$

Since $c_{1,\lambda}$ is a monotonically increasing function of λ (cf. Fig. 2) the convergence of the ES can be guaranteed by increasing the number of offspring. The same effect may be achieved by averaging the measurements (i. e. by repeated measuring without mutational changes). With (74, 77, 83) we get the criterion

$$\delta < \frac{4\sqrt{m}\, c_{1,\lambda}}{N}. \tag{84}$$

In order to ensure convergence the question arises whether it would be better to raise the number of descendants λ or to increase the number of repetitions m per offspring. Since the progress coefficient $c_{1,\lambda}$ raises only with the logarithm of λ (cp. eq. (30)) and $\sigma\{Q_m\}$ obeys a square root law (77) it is expected that for larger $m\,\lambda$ products the averaging method should be preferred. However, for the range $m\,\lambda \leq 22$ it may be better to raise the number of offspring. This results can be achieved from the relation

$$1 \leq \frac{c_{1\,,m\,\lambda}}{\sqrt{m}\, c_{1,\lambda}}, \tag{85}$$

combining the evolution conditions (83) and (84). If (85) holds, then it will be better to use the $(1\,, m\,\lambda)$ ES instead of the $(1,\lambda)$ ES with m-times averaging. The results of the numerical analysis of (85) are displayed in Table 2.

λ	2	3
m	< 12	< 4

Table 2: Combination of parameters for which a $(1\,, m\,\lambda)$-strategy would be better than a $(1,\lambda)$ ES with averaging over m scores.

We can draw a remarkable conclusion for the $(\tilde{1}, \tilde{\lambda})$ - spherical ES: For an effective working of the ES one should prefer strategies with a low number of descendants but the selection should rely on an averaging process over a high number of measurements. That means, philosophically interpreted, it seems that the *democracy principle* (the averaging) yields better results than an *elite principle* (selection of the seemingly best offspring) in a community being based on deception or which is affected by disinformation, noise, etc. The question whether this statement exceeds the limit given by the spherical model remains for future investigations.

The non-validity of (81) does not necessarily led to a divergence $R \to \infty$. If (82) holds for a certain $R > R_l$ the steady state condition (80) can be used to determine the final distance R_∞. Resolving eq. (78) for $R = R_\infty$ yields

$$R_\infty = \frac{1}{\sqrt{2}} \frac{\sigma N}{2c_{1,\lambda}} \sqrt{1 + \sqrt{1 + \left(\frac{2c_{1,\lambda}\sigma_\delta}{cN\sigma^2}\right)^2}}.$$

Assuming the convergence of σ, i. e. $\sigma \to 0$ for $g \to \infty$ we get from eq. (78)

$$R_\infty = \frac{1}{2}\sqrt{\frac{\sigma_\delta N}{c_{1,\lambda}\, c}}.$$

4 Summary and Conclusions

A new approach to the progress rate, the success probability, and the Q-gain for the spherical model has been presented. Starting from the asymptotical behavior of the mutation vectors for nearly infinite-dimensional parameter spaces it was possible to derive the integrals for the $(1 \overset{+}{,} \lambda)$ ES by simple geometrical argumentation. The results obtained are consistent to former findings [2, 3, 11, 12, 17] in case of the $(1+1)$ ES and the $(1, \lambda)$ ES. For the $(1 + \lambda)$ ES and the Q-gain new results have been obtained. The main part of this paper was devoted to the noisy ES. Using the argumentation from section 2 it was possible to extend the method to the $(\tilde{1} \overset{+}{,} \tilde{\lambda})$ strategy by describing the selection process in the Q-$\tilde{Q}$-picture. In this picture the *virtual* Q-gain can be easily derived. But this observed $\tilde{Q}$-gain has not a direct connection to the *real* progress rate. Therefore a $(\tilde{1}, \tilde{\lambda})$ and a $(\tilde{1} + \tilde{\lambda})$ theory have been developed with surprisingly simple analytical results in case of the $(\tilde{1}, \tilde{\lambda})$ ES and a closed expression for the $(\tilde{1} + \tilde{1})$ ES. These results have been finally used to discuss the convergence behavior of the noisy ES.

As one can see, the determination of analytical expressions for the progress rate, the convergence behavior, etc. is a difficult task and it seems likely that results will be achieved for simple models only. The method presented, however, points in a direction which may be fruitful for further investigations. By taking full advantage of the asymptotical behavior of the mutation vectors it is possible to get analytical results not only for the spherical model but also for the discus model introduced by the author [7]. The application of the method to the $(\mu \overset{+}{,} \lambda)$ ES seems possible.

The description of the σ-dynamics (the σ-control problem) is another direction for future investigations. Up to now only the 1/5-rule can be treated mathematically. But this rule is of minor importance, useful especially for the $(\tilde{1} + \tilde{1})$ ES. For higher order evolutionary schemes an evolutionary adaptation of the strategy parameter σ seems rather suitable (Schwefel [3, 4]) since such a method can work model independently. The investigation of such an evolutionary parameter tuning should be a future task of high priority, because it contains the key for the understanding of the global dynamics of the Evolution Strategy.

Acknowledgements

The author would like to thank G. Rudolph for critical remarks and Hans-Paul Schwefel for his kind interest and helpful discussions.

The revised version was prepared at the Computing Center of the University of Mannheim. The author is grateful to Hans W. Meuer for his support.

References

[1] Rechenberg, I.: *Cybernetic Solution Path of an Experimental Problem.* Royal Aircraft Establishment, Library Translation 1122, Farnborough, 1965.

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

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

[4] Schwefel, H.-P.: *Collective Phenomena in Evolutionary Systems.* Paper presented at the *International Conference on 'Problems of Constancy and Change'*, 31st Annual Meet. International Soc. for General System Research, Budapest, June 1987, Vol. II, pp. 1025–1032.

[5] Beyer, H.-G. et al.: *Minimization of Multibunch-BBU in a LINAC by Evolutionary Strategies.* In Rossbach, J. (ed.): *15th International Conference on High Energy Accelerators HEACC'92.* Int. J. Mod. Phys. A (Proc. Suppl.) **2B**(1993), pp. 848–850.

[6] Beyer, H.-G.: *Some Aspects of the 'Evolution Strategy' for Solving TSP-Like Optimization Problems.* In Männer, R.; Manderick, B. (eds.): *Parallel Problem Solving from Nature, 2.* Elsevier, Amsterdam, 1992, pp. 361–370.

[7] Beyer, H.-G.: *Ein Evolutionsverfahren zur mathematischen Modellierung stationärer Zustände in dynamischen Systemen.* Doctoral thesis, HAB Weimar, 1989.

[8] Beyer, H.-G.: *Simulation of Steady States in Dissipative Systems by Darwin's Paradigm of Evolution.* J. Non-Equilib. Thermodyn., Vol. **15**(1990), pp. 45–58.

[9] Beyer, H.-G.: *On a General Evolution Strategy for Dissipative Systems.* In Voigt, H.-M.; Mühlenbein, H.; Schwefel, H.-P. (eds.): *Evolution and Optimization '89: Selected Papers on Evolution Theory, Combinatorial Optimization, and Related Topics.* (Series: *Mathematical Ecology and Evolution*), Akademie-Verlag, Berlin, 1990, pp. 69–78.

[10] Beyer, H.-G.: *Towards a Theory of 'Evolution Strategies'.* In series: *Report of the Systems Analysis Research Group*, University of Dortmund, No. SYS-5/92.

[11] Rechenberg, I.: *The Evolution Strategy. A Mathematical Model of Darwinian Evolution.* In Frehland, E. (ed.): *Synergetics - From Microscopic to Macroscopic Order.* Springer Series in Synergetics, Vol. **22**, New York, 1984, pp. 122–132.

[12] Muth, C.: *Introduction to Evolution Strategy - a Tutorial Paper.* Regelungstechnik, *30.* Jahrgang (1982), No. 9, pp. 297–303.

[13] Bronshtein, I. N.; Semendyayev, K. A.: *Handbook of Mathematics.* Verlag Harry Deutsch, Frankfurt, 1985.

[14] Wadsworth, G. P.; Bryan, J. G.: *Introduction to Probability and Random Variables.* McGraw-Hill, New York, Toronto, London, 1960.

[15] Magnus, W.; Oberhettinger. F.; Soni , R. P.: *Formulas and Theorems for the Special Functions of Mathematical Physics.* Springer-Verlag, Berlin, Heidelberg, New York, 1966.

[16] Schwefel, H.-P.: *private communication.* Nov. 1990.

[17] Scheel, A.: *Beitrag zur Theorie der Evolutionsstrategie.* Doctoral thesis, TU Berlin, 1985.