Analysis of the $(\mu/\mu, \lambda)$-ES on the Parabolic Ridge

A. Irfan Oyman
University of Dortmund
Department of Computer Science
Systems Analysis Research Group
D-44221 Dortmund, Germany
oyman@LS11.cs.uni-dortmund.de

Hans-Georg Beyer
University of Dortmund
Department of Computer Science
Systems Analysis Research Group
D-44221 Dortmund, Germany
beyer@LS11.cs.uni-dortmund.de

Abstract

The progress behavior of evolution strategies (ES) using recombination is analyzed in this paper on the parabolic ridge. This test function represents landscapes far from the optimum. The ES algorithms with intermediate and dominant recombination are considered in the analysis. The derivations are presented for intermediate recombination. Thereafter, the formulae for dominant recombination are obtained using the so-called surrogate mutation model. In the analysis, the formulae are derived for the progress rate φ and for the stationary distance $R^{(\infty)}$ to the ridge axis. As a result, it will be shown that the progress rate φ can increase if recombination is applied. Simulations are used to show the appropriateness of the formulae derived.

Keywords

Evolution strategy, EA theory, progress rate, progress behavior, recombination, parabolic ridge.

1 Introduction

The progress behavior of evolution strategies (ES) is an interesting, theoretical research direction. This analysis cannot, generally, be accomplished for an arbitrary test function because of mathematical difficulties. Therefore, one prefers typical fitness landscapes for this analysis. These landscapes should represent certain characteristics of other fitness functions, and the analysis should be tractable on them.

In the early years of ES theory, the so-called corridor model and the sphere model were investigated for the $(1+1)$-ES (Rechenberg, 1973) and the $(1, \lambda)$-ES (Schwefel, 1977). Ten years later, the Ph.D. work of Scheel (1985) resumed these investigations and provided new results for the $(1, \lambda)$-ES on the sphere model. Again, roughly ten years later, this field attracted new interest resulting in thorough analysis of the $(1 \stackrel{+}{,} \lambda)$-ES, the (μ, λ)-ES, and the $(\mu/\mu, \lambda)$ multirecombinant ES (see Beyer (1993, 1995a, 1995b) and Rechenberg (1994)). While these later investigations focus on the sphere model, the class of ridge functions received interest in simulation experiments performed by Herdy (1992) showing a different behavior compared to the sphere model class. However, a theory was still missing. Among the ridge functions, the parabolic ridge is the best investigated example. The parabolic ridge is a test function that models landscapes far from the optimum. The results to be obtained can be generalized to other landscapes as long as the local curvatures are similar.

Meanwhile, there are a lot of theoretical results on the performance of the ES on the ridge function class. Investigations on the $(1 \overset{+}{,} \lambda)$-ES have been published by Oyman et al. (1998) using a local geometrical model approach that led to surprisingly simple formulae for the progress rate φ. However, a probability theory approach yielding these results could not be presented there. Furthermore, the radial dynamics cannot be derived from those considerations. Filling this gap with a thorough theory was the aim of the paper by Oyman et al. (2000). There, the focus was on the $(1, \lambda)$-ES on the parabolic ridge. As already mentioned, the progress behavior of ES algorithms on the parabolic ridge is determined by two measures: the progress rate φ and the distance r of the center of mass to the ridge axis. In this work, we will focus on the $(\mu/\mu_{\mathrm{I}}, \lambda)$-ES, the ES algorithm with intermediate recombination, using probability theory techniques. In contrast, the investigations for dominant recombination, i.e., for the $(\mu/\mu_{\mathrm{D}}, \lambda)$-ES algorithm, will be carried out by analogy, relying on the so-called surrogate mutation model. The accuracy of the progress formulae obtained will be investigated by simulations.

The paper is organized as follows. The algorithms used and the test function considered are introduced in Section 2. Section 3 is devoted to the progress rate theory. In this section, the static and stationary progress rate formulae will be derived for the $(\mu/\mu_{\mathrm{I}}, \lambda)$-ES. The respective formulae for dominant recombination will be obtained by analogy. The appropriateness of this approach will be shown by simulations. Furthermore, the performance increasing effect of recombination will be discussed.

In Section 4, the radial dynamics, i.e., the evolution of the distance r to the ridge axis, will be investigated. The state equation of r will be derived for intermediate recombination. Its stationary solution $R^{(\infty)}$ will be derived for intermediate and dominant recombination and compared with simulations. Section 5 will summarize the paper and provide possible directions for future research.

2 The Algorithms and the Test Function

2.1 Algorithms

The evolution strategies belong to a class of algorithms, often called evolutionary algorithms (EAs), which imitate the processes of natural evolution. The main operators used are *mutation*, *crossover* (or recombination), and *selection*. The selection is performed by the environment, in our case the objective function to be optimized (the case of maximization is considered in this work without loss of generality), i.e.,

$$F(\mathbf{x}) \to \text{Max.} \qquad \text{with} \qquad \hat{\mathbf{x}} := \arg\max_{\mathbf{x}}[F(\mathbf{x})], \qquad \hat{F} := F(\hat{\mathbf{x}}) \tag{1}$$

yields the criterion for this purpose. Intermediate and dominant recombination will be considered.

In the $(\mu/\mu_{\mathrm{I}}, \lambda)$-ES algorithm analyzed in this paper, *intermediate recombination* serves as recombination operator. The $(\mu/\mu_{\mathrm{I}}, \lambda)$-ES algorithm can be found in Figure 1. It starts with the initialization of the parent population comprising μ individuals. Thereafter, the generation loop is executed until the termination criterion is fulfilled, which may be the generation of a definite number of populations or others.

The generation loop starts with the determination of the centroid, the center of mass, of the μ parents in the search space. This calculation of the centroid is called intermediate recombination. Thereafter, λ descendants are created on top of this centroid $\langle\mathbf{x}\rangle$. The

begin
  $g := 0$
  initialize($\mathbf{P}^{(0)}$) $\left[\mathbf{P}^{(0)} := \left\{\forall m \in (1,\ldots,\mu) : \left(\mathbf{x}_m^{(0)}, F(\mathbf{x}_m^{(0)})\right)\right\}\right]$
  while **not** *terminate*() **do**
    $\langle\mathbf{x}\rangle :=$ *recombine*($\mathbf{P}^{(g)}, \mu$)
    for $l := 1$ **to** λ **do**
      $\tilde{\mathbf{x}}_l :=$ *mutate*($\langle\mathbf{x}\rangle, \sigma$)
      $\tilde{F}_l := F(\tilde{\mathbf{x}}_l)$
    od
    $\tilde{\mathbf{P}}^{(g)} := \left\{\forall l \in \{1,\ldots,\lambda\} : (\tilde{\mathbf{x}}_l, \tilde{F}_l)\right\}$
    $\mathbf{P}^{(g+1)} :=$ *select*($\tilde{\mathbf{P}}^{(g)}$)
    $g := g + 1$
  od
end

Figure 1: The $(\mu/\mu_\mathrm{I},\lambda)$-ES algorithm.

creation of a descendant is done by mutation. The mutations are realized by adding an N-dimensional vector of independent, identically distributed, random numbers, sampled from the normal distribution with mean 0 and variance σ^2, having density

$$p_z(z) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{z}{\sigma}\right)^2\right], \tag{2}$$

to the parental centroid state $\langle\mathbf{x}\rangle$. As a result, one obtains offspring $\mathbf{x}_l$, which are evaluated according to the objective function $F(\mathbf{x})$. After the procreation of λ offspring, selection takes place. The best μ of these λ offspring individuals are selected according to their F values as the parents of the next generation. The last step of the generation loop is incrementing the generation counter g. Thereafter, the loop starts with the new parent individuals.

The $(\mu/\mu_\mathrm{D},\lambda)$-ES, the ES algorithm with *dominant recombination*, is also analyzed in this work. Differing from the $(\mu/\mu_\mathrm{I},\lambda)$-ES, which applies the mutations at the same point (the centroid of parental individuals), the $(\mu/\mu_\mathrm{D},\lambda)$-ES applies the mutations at a temporary state $\mathbf{x}_l$ that is generated anew for each individual. Therefore, the step of recombination is executed in the for-loop as the first line. In dominant recombination (see Figure 2), each variable of the temporary state $\mathbf{x}_l$ is sampled anew from the parental set. The parental individuals are selected using a uniform distribution. Using the orthonormal basis vectors $\mathbf{e}_i$, one can express dominant recombination with the following two equations:

$$r_i \;:=\; \mathtt{Uniform}\{1,\ldots,\mu\}, \tag{3}$$

$$\mathbf{x}_l \;:=\; \sum_{i=1}^{N}\left(\mathbf{e}_i^\mathrm{T}\mathbf{x}_{r_i}^{(g)}\right)\mathbf{e}_i. \tag{4}$$

The algorithms displayed in Figures 1 and 2 do not perform *self-adaptation*.[1] That is, the mutation strength σ is kept constant throughout the ES runs. This paper will touch aspects of self-adaptation only marginally. Investigations on the self-adaptive behavior are a topic of current research.

[1] For an introduction to self-adaptative ESs, see Bäck and Schwefel (1993) and Bäck (1997). Theoretical aspects are discussed in Beyer (1996).

```
begin
    g := 0
    initialize(P^(0))      [P^(0) := {∀m ∈ (1, ..., μ) : (x_m^(0), F(x_m^(0)))}]
    while not terminate() do
        for l := 1 to λ do
            E_l := mate(P^(g), μ)
            x_l := recombine(E_l, μ)
            x̃_l := mutate(x_l, σ)
            F̃_l := F(x̃_l)
        od
        P̃^(g) := {∀l ∈ {1, ..., λ} : (x̃_l, F̃_l)}
        P^(g+1) := select(P̃^(g))
        g := g + 1
    od
end
```

Figure 2: The $(\mu/\mu_{\mathrm{D}}, \lambda)$-ES algorithm.

2.2 The Parabolic Ridge

The parabolic ridge test function is defined as

$$F(\mathbf{x}) := x_0 - d \sum_{i=1}^{N-1} x_i^2. \tag{5}$$

The variable x_0 is named the linear part; the other x_i the quadratic part, respectively. The positive factor d is used to scale the effect of the quadratic part. An optimum (maximum) of this function does not exist in the usual mathematical terms, however, in the sense of a limit process, one can assume

$$\hat{x}_0 \to +\infty; \qquad \hat{x}_i = 0, \qquad 1 \le i \le N-1. \tag{6}$$

That is, the algorithm is expected to search for large values of x_0 while keeping other x_i as small as possible. The latter can be subsumed by

$$r = \sqrt{\sum_{i=1}^{N-1} x_i^2}, \tag{7}$$

i.e., r is the distance to the progress axis (or ridge axis).

Because there are two objectives, enlarging x_0 and decreasing r, there are two quantities to describe the ES process that are to be analyzed. First, one is interested in finding the expected change in the x_0 value in a single generation, the so-called progress rate φ. Second, a state equation of successive r values must be derived. Using this equation, the stationary r state can be determined.

In the general case of the parabolic ridge, the ridge axis can have any direction in the search space. This direction is specified by the vector $\mathbf{v}$, $\|\mathbf{v}\| = 1$. Furthermore, one can introduce a displacement $\mathbf{d}$ which shifts the coordinate origin. The general case of the parabolic ridge can be written as

$$F(\mathbf{x}) = \mathbf{v}^{\mathrm{T}}(\mathbf{x} - \mathbf{d}) - d\,[\|(\mathbf{v}^{\mathrm{T}}(\mathbf{x} - \mathbf{d}))\mathbf{v} - (\mathbf{x} - \mathbf{d})\|]^2. \tag{8}$$

3 The Progress Rate φ

3.1 Fundamentals

The progress rate φ will be derived in this section for the $(\mu/\mu, \lambda)$-ES on the parabolic ridge. It is defined as the expected progress along the ridge axis averaged over the μ best individuals selected for the next generation, i.e., between the centroids of two successive parental generations. If $z_{m;\lambda}$ is the random variate of the progress in x_0 direction for the m-th best individual, then the progress rate φ is given as the expected value of the average of those x_0 components which belong to the μ best mutations $\mathbf{z}$.

$$\varphi := \mathrm{E}\{\langle z \rangle\} = \mathrm{E}\left\{\frac{1}{\mu}\sum_{m=1}^{\mu} z_{m;\lambda}\right\} = \frac{1}{\mu}\sum_{m=1}^{\mu} \mathrm{E}\{z_{m;\lambda}\} \tag{9}$$

Here, z stands for the zeroth component of the mutation vector $\mathbf{z}$. Let us denote the probability density function of the random variate $z_{m;\lambda}$ as $p_{m;\lambda}(z)$. The expected value expression (9) becomes

$$\varphi = \frac{1}{\mu}\sum_{m=1}^{\mu} \int_{-\infty}^{\infty} z\, p_{m;\lambda}(z) \mathrm{d}z. \tag{10}$$

The structure of (10) is similar to expected value expressions already treated in the investigation of the $(\mu/\mu_\mathrm{I}, \lambda)$-ES on the sphere model investigated in Beyer (1995b, 81–111). Therefore, we will keep the explanations as short as possible. The density $p_{m;\lambda}(z)$ is formally given by

$$p_{m;\lambda}(z) = \frac{\lambda!}{(m-1)!(\lambda-m)!}\, p_z(z)\, P_{\mathrm{a}\, m;\lambda}(z). \tag{11}$$

It consists of: the $p_z(z)$ mutation density, which describes the $z \sim \mathcal{N}(0, \sigma^2)$ mutation process of the x_0-component of a single individual; the acceptance probability $P_{\mathrm{a}\, m;\lambda}(z)$, providing the probability that the z value considered belongs to the m-th best offspring; and a combinatorial factor counting the number of possible cases of being the m-th best out of μ individuals. The acceptance probability $P_{\mathrm{a}\, m;\lambda}(z)$ of the m-th best individual is obtained as an integral over all possible local quality values Q (for the definition of Q see Equation (17))

$$P_{\mathrm{a}\, m;\lambda}(z) = \int_{-\infty}^{\infty} p(Q|z)\, [P(Q)]^{\lambda-m}\, [1 - P(Q)]^{m-1}\, \mathrm{d}Q. \tag{12}$$

Here, $p(Q|z)$ is the conditional local fitness (quality) density produced by a $\mathbf{z} \sim \vec{\mathcal{N}}(0, \sigma^2)$ mutation vector, the 0-th coordinate of which is kept constant at $z_0 = z$. $P(Q)$ is the cumulative distribution function (cdf) of the nonconditional Q-distribution induced by an arbitrary mutation vector $\mathbf{z}$, i.e., $P(Q) = \Pr(Q(\mathbf{z}) \le Q)$, and $[1 - P(Q)]$ is its complement $\Pr(Q(\mathbf{z}) > Q)$. Since there are $(m-1)$ better Q values, we have the $(m-1)$ exponent at $\Pr(Q(\mathbf{z}) > Q)$, and there are $(\lambda - m)$ worse individuals leading to the $(\lambda - m)$ exponent at $\Pr(Q(\mathbf{z}) \le Q)$. Substituting (11) and (12) in (10), one obtains

$$\varphi = \frac{\lambda!}{\mu}\sum_{m=1}^{\mu} \int_{-\infty}^{\infty} z\, p_z(z) \int_{-\infty}^{\infty} p(Q|z)\, \frac{[P(Q)]^{\lambda-m}\, [1-P(Q)]^{m-1}}{(m-1)!(\lambda-m)!}\, \mathrm{d}Q\, \mathrm{d}z. \tag{13}$$

After exchanging the order of integration, one obtains

$$\varphi = \frac{\lambda!}{\mu} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} z\, p_z(z)\, p(Q|z)\, \mathrm{d}z \right) \sum_{m=1}^{\mu} \frac{[P(Q)]^{\lambda-m}\, [1-P(Q)]^{m-1}}{(m-1)!(\lambda-m)!}\, \mathrm{d}Q. \tag{14}$$

This is an *exact* progress rate integral measuring the advancement of the $(\mu/\mu_\mathrm{I}, \lambda)$-ES in an arbitrary search space direction on a local quality function Q. In order to investigate this formula further, one needs to specify $p(Q|z)$ and $P(Q)$. Since $p(Q|z)$ and $P(Q)$ depend on the local quality function measuring the local fitness change, Q must be specified first. This will be done for the parabolic ridge function in the next subsection. Thereafter, the conditional density $p(Q|z)$ will be approximated with a normal density, and then $P(Q)$ will be derived. Finally, the evaluation of the φ integral follows.

3.2 The Local Quality Function

Let the parental centroid be at the state $\mathbf{x}$ having the fitness value $F = F(\mathbf{x})$. According to the $(\mu/\mu_\mathrm{I}, \lambda)$-ES algorithm in Figure 1, offspring are generated by adding mutations $\mathbf{z}$ to this state $\mathbf{x}$ yielding fitness values $F(\mathbf{x}+\mathbf{z})$. Since the (μ, λ) truncation selection considers only the fitness change, the difference $F(\mathbf{x}+\mathbf{z}) - F(\mathbf{x})$ can be used to evaluate the *quality* of the mutation made. Thus, one gets the local quality function

$$Q_\mathbf{x}(\mathbf{z}) = F(\mathbf{x}+\mathbf{z}) - F(\mathbf{x}). \tag{15}$$

In the general case, $Q_\mathbf{x}(\mathbf{z})$ will be too complicated to be used in the further analysis. Therefore, one uses a quadratic approximation, which is obtained by Taylor expansion of $F(\mathbf{x}+\mathbf{z})$ at the parental state $\mathbf{x}$ assuming the mutations to be sufficiently small.

$$F(\mathbf{x}+\mathbf{z}) = F(\mathbf{x}) + \sum_{i=0}^{N-1} \frac{\partial F}{\partial x_i} z_i + \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{\partial^2 F}{\partial x_i \partial x_j} z_i z_j + \ldots \tag{16}$$

Dropping the higher order terms, the local quality function $Q_\mathbf{x}(\mathbf{z})$ (15) becomes

$$F(\mathbf{x}+\mathbf{z}) - F(\mathbf{x}) = Q_\mathbf{x}(\mathbf{z}) := \mathbf{a}^\mathrm{T}\mathbf{z} - \mathbf{z}^\mathrm{T}\mathbf{Q}\mathbf{z}. \tag{17}$$

By comparison of (17) and (16), one finds

$$(\mathbf{a})_i := \frac{\partial F}{\partial x_i}, \qquad (\mathbf{Q})_{ij} := -\frac{1}{2} \frac{\partial^2 F}{\partial x_i \partial x_j}. \tag{18}$$

For the parabolic ridge (5), one must note that approximation (16) is exact. The entries for the $\mathbf{a}$ vector and $\mathbf{Q}$ matrix are calculated using (18). One obtains

$$(\mathbf{a})_i = \begin{cases} 1 & i = 0 \\ -2dx_i & i \neq 0, \end{cases} \qquad (\mathbf{Q})_{ij} = \begin{cases} 0 & i \neq j \\ 0 & i = j = 0 \\ d & \text{otherwise} \end{cases} \tag{19}$$

and the quality function reads

$$Q_\mathbf{x}(\mathbf{z}) = z_0 - 2d \sum_{i=1}^{N-1} x_i z_i - d \sum_{i=1}^{N-1} z_i^2. \tag{20}$$

3.3 The Determination of the $p(Q|z)$ Density

In order to keep the integrals to be evaluated manageable, the conditional probability density function $p(Q|z)$ will be approximated by a normal density. The density describes the distribution of a Q variate produced according to Equation (20) with $\mathcal{N}(0, \sigma^2)$-distributed z_i $(i = 1, \ldots, N-1)$ under the condition that $z_0 = z$ is kept constant. Its normal approximation reads formally

$$p(Q|z) = \frac{1}{\sqrt{2\pi} S_{Q|z}} \exp\left[-\frac{1}{2} \frac{1}{S_{Q|z}^2} (Q - M_{Q|z})^2 \right]. \tag{21}$$

The symbols $M_{Q|z}$ and $S_{Q|z}$ stand for the mean value and standard deviation of this density, respectively. They can be obtained by determining the respective moments of the random function (20). If one recalls that for $i = 1, \ldots, N-1$, $\overline{z_i} = \overline{\mathcal{N}(0, \sigma^2)} = 0$ and $\overline{z_i^2} = \overline{\mathcal{N}^2(0, \sigma^2)} = \sigma^2$, for the first moment of Q one immediately obtains

$$M_{Q|z} = \overline{Q_{\mathbf{x}}(\mathbf{z})} = z - d(N-1)\sigma^2. \tag{22}$$

The calculation of the standard deviation $S_{Q|z}$ is a bit more difficult. However, its calculation has already been accomplished for the N-dimensional case in Beyer (1994, 61, Equation (15)). Adapting the respective equation one has

$$S_{Q|z} = \sigma \sqrt{\|\mathbf{a}'\|^2 + 2\sigma^2 \mathrm{Tr}[\mathbf{Q}'^2]}, \tag{23}$$

where $\mathbf{a}'$ is the $\mathbf{a}$ vector in (19) reduced by the first entry (index $i = 0$), and $\mathbf{Q}'$ is the Q matrix reduced by the first column and first row, because the first $\mathbf{z}$ coordinate z_0 is kept constant (condition $z_0 = z$). Thus, one obtains $\|\mathbf{a}'\|^2 = (2d)^2 \sum_{i=1}^{N-1} x_i^2$ and for the trace of $\mathbf{Q}'^2$: $\mathrm{Tr}[\mathbf{Q}'^2] = \sum_{i=1}^{N-1} d^2$. Therefore, $S_{Q|z}$ reads, taking (7) into account,

$$S_{Q|z} = \sigma \sqrt{(2dr)^2 + 2d^2(N-1)\sigma^2}. \tag{24}$$

The same method can be applied to derive the normal approximation for $P(Q)$. In that case, the complete $\mathbf{a}$ vector and $\mathbf{Q}$ matrix have to be considered, and one obtains

$$M_Q = -(N-1)d\sigma^2, \qquad S_Q = \sigma\sqrt{1 + (2dr)^2 + 2d^2(N-1)\sigma^2}. \tag{25}$$

Alternatively, $P(Q)$ can also be derived using the conditional $p(Q|z)$ density.

3.4 The Derivation of $P(Q)$

This derivation is presented here in order to show the consistency of the approximation approach taken. From the definition of $P(Q)$ and $p(Q|z)$ it follows

$$P(Q) = \int_{-\infty}^{Q'=Q} \int_{-\infty}^{\infty} p(Q'|z) p_z(z) \mathrm{d}z \, \mathrm{d}Q'. \tag{26}$$

The mutations are $z \sim \mathcal{N}(0, \sigma^2)$, obeying the density function $p_z(z)$ given by Equation (2). The inner integration is taken over the possible z values. The outer integration is due to the definition of the cumulative distribution, i.e., the probability for a sample of the Q' values to be smaller than a given Q.

After exchanging the order of integration, one first has to calculate the integral over Q' taking (21) and the left equation in (25) into account. This immediately produces the $\Phi[.]$ term in (27) below.

$$P(Q) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{1}{2}\left(\frac{z}{\sigma}\right)^2} \Phi\left[\underbrace{-\frac{\sigma}{S_{Q|z}}}_{a} \frac{z}{\sigma} + \underbrace{\frac{Q - M_Q}{S_{Q|z}}}_{b} \right] \mathrm{d}z, \tag{27}$$

where $\Phi[.]$ denotes the cumulative distribution function (cdf) of the standard normal distribution, i.e.,

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t=x} \mathrm{e}^{-\frac{1}{2}t^2} \mathrm{d}t. \tag{28}$$

After defining a and b as above, substituting $t := \frac{z}{\sigma}$, and using the integral formula (Beyer, 2000, Equation (A.10)),

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}t^2} \Phi[at + b] \mathrm{d}t = \Phi\left(\frac{b}{\sqrt{1 + a^2}} \right), \tag{29}$$

the resulting cdf $P(Q)$ reads, after using the right equation in (25),

$$P(Q) = \Phi\left(\frac{b}{\sqrt{1 + a^2}} \right) = \Phi\left(\frac{Q - M_Q}{S_Q} \right). \tag{30}$$

3.5 On the Derivation of the Progress Rate

The inner integral of (14) will be calculated next. After the substitutions

$$t := \frac{z}{\sigma}, \qquad \mathrm{d}t := \frac{\mathrm{d}z}{\sigma}, \qquad \text{and} \qquad s := \frac{Q - M_Q}{S_Q}, \qquad \mathrm{d}s := \frac{\mathrm{d}Q}{S_Q} \tag{31}$$

one first obtains

$$\frac{Q - M_Q - z}{S_{Q|z}} = \frac{S_Q \cdot s - z}{S_{Q|z}} = \underbrace{-\frac{\sigma}{S_{Q|z}}}_{a} t + \underbrace{\frac{S_Q}{S_{Q|z}}}_{b} s = at + b(s). \tag{32}$$

Considering this in $p(Q|z)$, the progress rate formula reads

$$\begin{aligned} \varphi \quad = \quad & \frac{\lambda!}{\sqrt{2\pi}\mu} \frac{\sigma S_Q}{S_{Q|z}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t\, \mathrm{e}^{-\frac{1}{2}t^2} \mathrm{e}^{-\frac{1}{2}(at + b(s))^2} \mathrm{d}t \\ & \times \sum_{m=1}^{\mu} \frac{[\Phi(s)]^{\lambda - m} [1 - \Phi(s)]^{m-1}}{(m-1)!(\lambda - m)!} \mathrm{d}s. \end{aligned} \tag{33}$$

Using the integral formula (Beyer, 2000, Equation (A.8))

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t\mathrm{e}^{-\frac{1}{2}t^2} \mathrm{e}^{-\frac{1}{2}(at+b)^2} \mathrm{d}t = \frac{-ab \exp\left(-\frac{1}{2} \frac{b^2}{1+a^2} \right)}{(1 + a^2)^{\frac{3}{2}}}, \tag{34}$$

one obtains

$$\varphi = \frac{\lambda!}{\sqrt{2\pi}\mu} \frac{\sigma^2}{S_Q} \int_{-\infty}^{\infty} s\, \mathrm{e}^{-\frac{1}{2}s^2} \sum_{m=1}^{\mu} \frac{[\Phi(s)]^{\lambda-m} [1-\Phi(s)]^{m-1}}{(m-1)!(\lambda-m)!} \,\mathrm{d}s. \tag{35}$$

The sum in (35) can be expressed by an integral. Using the identity (Abramowitz and Stegun, 1984, 83)

$$\sum_{m=1}^{\mu} \frac{P^{m-1} [1-P]^{\lambda-m}}{(m-1)!(\lambda-m)!} = \frac{1}{(\lambda-\mu-1)!(\mu-1)!} \int_0^{1-P(s)} x^{\lambda-\mu-1}(1-x)^{\mu-1} \,\mathrm{d}x, \tag{36}$$

for $P = 1 - \Phi(s)$, (35) becomes

$$\varphi = \frac{\sigma^2}{S_Q} \frac{\lambda-\mu}{\sqrt{2\pi}} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} s\, \mathrm{e}^{-\frac{1}{2}s^2} \int_0^{\Phi(s)} x^{\lambda-\mu-1}(1-x)^{\mu-1} \,\mathrm{d}x \,\mathrm{d}s. \tag{37}$$

After exchanging the integration order, one gets

$$\varphi = \frac{\sigma^2}{S_Q} \frac{\lambda-\mu}{\sqrt{2\pi}} \binom{\lambda}{\mu} \int_0^1 x^{\lambda-\mu-1}(1-x)^{\mu-1} \int_{\Phi^{-1}(x)}^{\infty} s\, \mathrm{e}^{-\frac{1}{2}s^2} \,\mathrm{d}s \,\mathrm{d}x. \tag{38}$$

The inner integration can be easily carried out, yielding $\exp\left[-\frac{1}{2}(\Phi^{-1}(x))^2\right]$. Using the substitution $x = \Phi(t)$, i.e., $\mathrm{d}x = \exp(-\frac{1}{2}t^2)\mathrm{d}t/\sqrt{2\pi}$, one obtains

$$\varphi = \frac{\sigma^2}{S_Q} \frac{\lambda-\mu}{2\pi} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} \mathrm{e}^{-t^2} [\Phi(t)]^{\lambda-\mu-1} [1-\Phi(t)]^{\mu-1} \,\mathrm{d}t. \tag{39}$$

Taking the S_Q definition in (25) into account and using the $c_{\mu/\mu,\lambda}$ definition

$$c_{\mu/\mu,\lambda} := \frac{\lambda-\mu}{2\pi} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} \mathrm{e}^{-t^2} [\Phi(t)]^{\lambda-\mu-1} [1-\Phi(t)]^{\mu-1} \,\mathrm{d}t \tag{40}$$

(see Beyer (1995b), for its definition and a table of $c_{\mu/\mu,\lambda}$ values), the static progress rate φ of the $(\mu/\mu_\mathrm{I}, \lambda)$-ES on the parabolic ridge is obtained

$$\boxed{\varphi = \frac{c_{\mu/\mu,\lambda}\sigma^2}{S_Q} = \frac{c_{\mu/\mu,\lambda}\sigma}{\sqrt{1+(2dr)^2+2d^2(N-1)\sigma^2}}.} \tag{41}$$

This formula shows the dependence of the progress rate on the distance r to the progress axis. The progress rate φ is always positive; it decreases down to zero as $r \to \infty$; and it attains its maximal value for $r = 0$.

3.6 The Stationary Progress Rate and Comparison with Experiments

Equation (41) gives the static progress rate for a given distance r to the ridge axis. Actually, the r value is an unknown random variate. Its determination will be carried out in Section 4. There, it will be shown that after a sufficiently large evolution time, the algorithm attains a steady state leading to a stationary distribution of the r values fluctuating around an expected stationary r value denoted by $R^{(\infty)}$. It is determined by Equation (93) below. After inserting the $R^{(\infty)}$ value in the static φ formula, one obtains an estimate for the stationary progress rate. The stationary formulae are presented in this section on the progress rate rather than in Section 4 in order to keep the φ formulae together. The result reads

$$\varphi = \frac{\mu c^2_{\mu/\mu,\lambda}}{\sqrt{\left(\frac{\mu c_{\mu/\mu,\lambda}}{\sigma}\right)^2 + \frac{[d(N-1)]^2}{2}\left(1+\sqrt{1+\left(\frac{2\mu c_{\mu/\mu,\lambda}}{d(N-1)\sigma}\right)^2}\right)}}. \tag{42}$$

This result can be normalized using

$$\sigma^* = d(N-1)\sigma, \tag{43}$$
$$\varphi^* = d(N-1)\varphi. \tag{44}$$

Thus, one obtains the normalized stationary progress rate formula

$$\varphi^* = \frac{\mu c^2_{\mu/\mu,\lambda}}{\sqrt{\left(\frac{\mu c_{\mu/\mu,\lambda}}{\sigma^*}\right)^2 + \frac{1}{2} + \frac{1}{2}\sqrt{1+\left(\frac{2\mu c_{\mu/\mu,\lambda}}{\sigma^*}\right)^2}}}. \tag{45}$$

Both Equations (42) and (45) also contain the $(1,\lambda)$ progress rate as a special case $\mu = 1$, because $c_{1,\lambda} \equiv c_{1/1,\lambda}$ holds per definition. As can be shown, Equation (45) holds also for the general parabolic ridge case (8).

The maximal value for the stationary progress rate is obtained for $\sigma \to \infty$ and $\sigma^* \to \infty$, respectively,

$$\hat{\varphi} = \lim_{\sigma\to\infty} \varphi = \frac{\mu c^2_{\mu/\mu,\lambda}}{d(N-1)}, \qquad \hat{\varphi}^* = \lim_{\sigma^*\to\infty} \varphi^* = \mu c^2_{\mu/\mu,\lambda}. \tag{46}$$

That is, as in the case of the $(1,\lambda)$-ES on the parabolic ridge (see Oyman et al. (1998)), the progress rate exhibits a saturation behavior. This is in contrast to the results on the sphere model (see, Beyer (1995b)).

The stationary progress rate formula (45) is compared with real ES runs in Figure 3. The simulations are carried out for the $(\mu/\mu_I, 10)$-ES with $\mu \in \{3,5,9\}$, $N = 1000$, and $d = 0.01$.[2] As one can see, the theoretical formula describes the simulation results correctly. The horizontal asymptote is attained in all three cases, although in the figure, the accordance is shown only for the $\mu = 9$ case.

The maximal progress rate values of the $(\mu/\mu_I,\lambda)$-ES and $(1,\lambda)$-ES can be compared for the same λ in order to assess the benefit obtained through recombination. Differing from the $(\mu/\mu_I,\lambda)$-ES, the $(1,\lambda)$-ES algorithm selects only the best descendant as the parent of the next generation. The maximal progress rate for the $(1,\lambda)$-ES was already obtained as (Oyman et al., 1997, 23)

$$\lim_{\sigma\to\infty} \varphi_{1,\lambda} = \frac{c^2_{1,\lambda}}{d(N-1)}, \qquad \lim_{\sigma^*\to\infty} \varphi^*_{1,\lambda} = c^2_{1,\lambda} \tag{47}$$

and is also contained as the special case $\mu = 1$ in Equation (46). In Figure 3, the comparison of progress rate values is done empirically. One can see that even the maximal progress

[2] Other values of d have been tested as well; one observes a similar predictive quality.

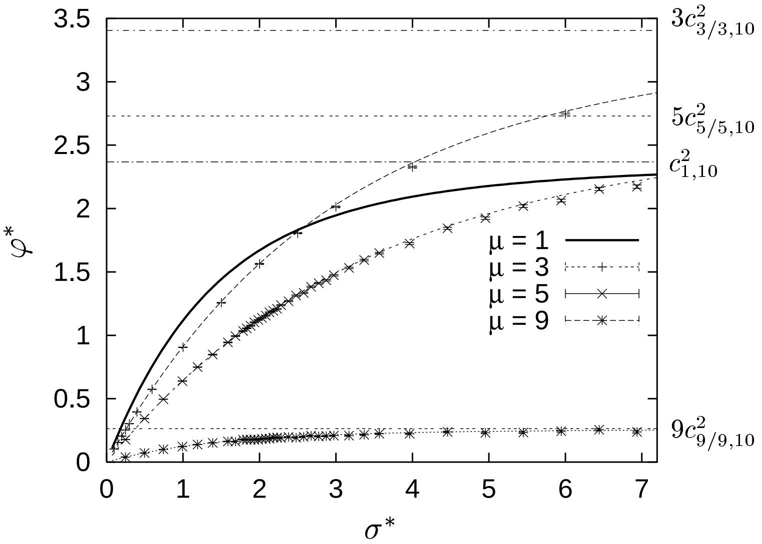

Figure 3: Normalized progress rate φ^* versus normalized mutation strength σ^* for the $(\mu/\mu_\mathrm{I}, 10)$-ES on the parabolic ridge, $\mu \in \{3, 5, 9\}$. The three asymptotic limits and the limit for the $(1, 10)$-ES are drawn as horizontal lines. The φ^* values are predicted correctly. In addition, the progress rate curve of the $(1, \lambda)$-ES is displayed as a thick continuous curve (without experimental data points).

of the $(5/5_\mathrm{I}, 10)$-ES is better than the $(1, 10)$-ES ($c_{1,10} \approx 1.53875$). The $(3/3_\mathrm{I}, 10)$-ES algorithm attains the largest maximal progress rate among all $(\mu/\mu_\mathrm{I}, 10)$-ES algorithms. From the investigation, one can conclude that intermediate recombination increases the progress rate beyond that of the $(1, \lambda)$-ES, provided that

$$\mu c^2_{\mu/\mu,\lambda} > c^2_{1,\lambda} \tag{48}$$

and σ is chosen sufficiently large. The latter condition (σ sufficiently large) can be assessed in Figure 3. It turns out that for small σ, the $(1, \lambda)$-ES performs better than the intermediate recombinant ES versions. This might have consequences when using self-adaptation techniques to learn the optimal σ. As for the parabolic ridge function, such techniques should be able to permanently increase the mutation strength in order to increase the progress rate. It is an open question whether there are simple self-adaptation rules, i.e., rules that rely on time locality, that are able to fulfill this desired behavior (for a thorough discussion, the reader is referred to Beyer and Deb (1999)).

3.7 On the Progress Rate of the $(\mu/\mu_\mathrm{D}, \lambda)$-ES

The progress rate of the ES with dominant recombination is obtained using an analogy approach based on the surrogate mutation model developed for the sphere model. Therefore, the application to the ridge function to be presented here is a more or less speculative one, because we do not have a proof for the correctness of our approach taken. The idea of surrogate mutations consists of substituting the effect of mutation and dominant recombination by surrogate mutations that are applied to the virtual center of mass point of the parental populations (for details, see Beyer (1995b)). For the asymptotic limit case $N \to \infty$ it has been shown for the sphere model that the influence of selection on the surrogate mutation strength can be neglected. Under this condition, the standard deviation σ_s of the surrogates becomes $\sigma_s = \sqrt{\mu}\sigma$. That is, neglecting the effect of selection the process of dominant recombination and (physical) mutation of strength, in the steady state σ yields an offspring population of variance $\mu\sigma^2$. Applying this surrogate mutation to the virtual population center of mass should have the same effect as the original dominant recombination and (physical) mutation process. Because the $(\mu/\mu_{\mathrm{I}}, \lambda)$-ES operates on the actual center of mass resulting in the φ formula (41), the progress rate of the $(\mu/\mu_{\mathrm{D}}, \lambda)$-ES is obtained from the $(\mu/\mu_{\mathrm{I}}, \lambda)$-ES formula by replacing σ by $\sqrt{\mu}\sigma$. One obtains from (41)

$$\boxed{\varphi = \frac{\sqrt{\mu}c_{\mu/\mu,\lambda}\sigma}{\sqrt{1 + (2dr)^2 + 2\mu d^2(N-1)\sigma^2}}.} \tag{49}$$

Similarly, for the stationary progress rate of the $(\mu/\mu_{\mathrm{D}}, \lambda)$-ES, one gets from (45)

$$\varphi^* = \frac{\mu c^2_{\mu/\mu,\lambda}}{\sqrt{\left(\frac{\sqrt{\mu}c_{\mu/\mu,\lambda}}{\sigma^*}\right)^2 + \frac{1}{2} + \frac{1}{2}\sqrt{1 + \left(\frac{2\sqrt{\mu}c_{\mu/\mu,\lambda}}{\sigma^*}\right)^2}}}. \tag{50}$$

It is important to note that these two formulae obtained for the $(\mu/\mu_{\mathrm{D}}, \lambda)$-ES are only valid if the direction $\mathbf{v}$ of the ridge axis in (8) is diagonal in the search space, e.g., $\mathbf{v} = \frac{1}{\sqrt{N}}(1, 1, \ldots, 1)^{\mathrm{T}}$. For other $\mathbf{v}$ directions the formula is unknown.

The progress rates obtained for intermediate and dominant recombination are compared in Figure 4. The results are obtained for a series of σ values and plotted for the $(5/5_{\mathrm{I}}, 10)$-ES and the $(5/5_{\mathrm{D}}, 10)$-ES. The axes are normalized using (43) and (44). The parabolic ridge parameters are $N = 1,000$ and $d = 0.01$. The simulations for the $(5/5_{\mathrm{I}}, 10)$-ES are compared to the theoretical formula (45) and for the $(5/5_{\mathrm{D}}, 10)$-ES to (50), respectively. Both simulation series are relatively well predicted by the analytically obtained formulae, although the accordance is better for intermediate recombination. As expected, there is a stronger discrepancy for the $(\mu/\mu_{\mathrm{D}}, \lambda)$-ES due to the rough surrogate mutation approach that strongly relies on $N \to \infty$ and the coordinate mixing behavior of the dominant recombination (diagonal $\mathbf{v}$ condition). As one can further see, the dominant version has a greater φ^* slope for small σ^* values. The slope can even be larger than that of the $(1, \lambda)$-ES (see also the discussion below Equation (48)). This may be regarded as an argument for using dominant recombination in self-adaptive ES versions.

4 On the Radial Dynamics of the Parental Centroid

The distance r of the parental centroid to the ridge axis of the $(\mu/\mu, \lambda)$-ES will be investigated in this section. To this end, the state equation of r^2 for the intermediate recombination will be derived. Using this equation, the stationary $R^{(\infty)}$ formula will be derived. Thereafter, the stationary $R^{(\infty)}$ formula will be obtained for the $(\mu/\mu_{\mathrm{D}}, \lambda)$-ES using the surrogate mutation model.

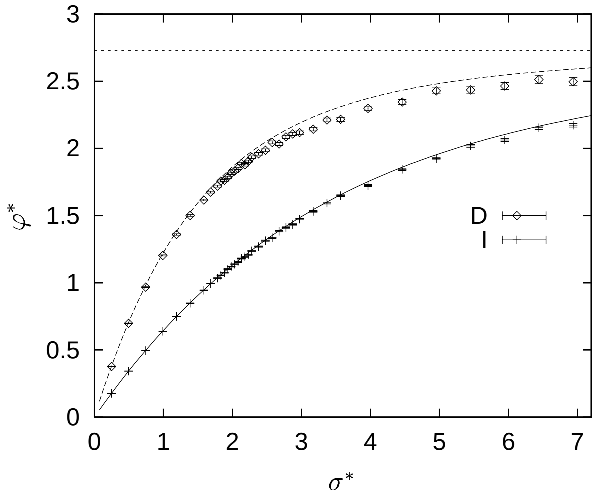

Figure 4: Normalized progress rate φ^* versus normalized mutation strength σ^* for the $(5/5_{\mathrm{I}}, 10)$-ES and the $(5/5_{\mathrm{D}}, 10)$-ES on the parabolic ridge with diagonal $\mathbf{v}$ vector. The common asymptotic limit $5c^2_{5/5,10}$ for $\sigma^* \to \infty$ is shown as a horizontal line.

4.1 Fundamentals

Let us recall the basics of the $(\mu/\mu_{\mathrm{I}}, \lambda)$-ES: Starting from the parental centroid $\langle \mathbf{x}^{(g)} \rangle$, offspring are generated according to

$$\tilde{\mathbf{x}}_l^{(g)} = \langle \mathbf{x}^{(g)} \rangle + \mathbf{z}_l^{(g)}. \tag{51}$$

The new parental centroid of generation $(g+1)$ is obtained by averaging over the μ best offspring, i.e.,

$$\langle \mathbf{x}^{(g+1)} \rangle = \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\mathbf{x}}_{m;\lambda}^{(g)}. \tag{52}$$

The first to the $(N-1)$-th component of $\langle \mathbf{x}^{(g+1)} \rangle$ determine the distance of the centroid to the ridge axis. The expected squared distance of the centroid at generation $g+1$ to the ridge axis is denoted by $\mathrm{E}\{{r^{(g+1)}}^2\}$

$$\mathrm{E}\left\{{r^{(g+1)}}^2\right\} := \mathrm{E}\left\{\sum_{i=1}^{N-1} \langle x_i^{(g+1)} \rangle^2\right\} \tag{53}$$

Here, the subscript index i denotes the respective components in the search space. The $\mathrm{E}\{{r^{(g+1)}}^2\}$ quantity is of special interest because it is used in the progress rate formula (41) as an estimate for r^2. Equation (53) can further be specified considering (52) and (51)

$$\mathrm{E}\left\{{r^{(g+1)}}^2\right\} = \mathrm{E}\left\{\sum_{i=1}^{N-1}\left(\frac{1}{\mu}\sum_{m=1}^{\mu} \tilde{x}_{i_{m;\lambda}}^{(g)}\right)^2\right\} = \frac{1}{\mu^2}\mathrm{E}\left\{\sum_{i=1}^{N-1}\left[\sum_{m=1}^{\mu}\left(\langle x_i^{(g)}\rangle + z_{i_{m;\lambda}}^{(g)}\right)\right]^2\right\}. \tag{54}$$

Unfortunately, this expression is difficult to evaluate analytically. Therefore, we start with another, analytically tractable expression and establish its relation to $\mathrm{E}\{r^{(g+1)^2}\}$

$$\mathrm{E}\left\{\langle r^{(g+1)^2}\rangle\right\} := \mathrm{E}\left\{\frac{1}{\mu}\sum_{m=1}^{\mu} r_m^{(g+1)^2}\right\}. \tag{55}$$

Note, the decisive difference to (53) is that the mean value calculation is applied to the length square of the μ parents of generation $(g+1)$. Since these parents are the selected μ best offspring of generation g one can write alternatively

$$\mathrm{E}\left\{\langle r^{(g+1)^2}\rangle\right\} = \frac{1}{\mu}\mathrm{E}\left\{\sum_{m=1}^{\mu}\left(\tilde{r}_{m;\lambda}^{(g)}\right)^2\right\} \tag{56}$$

$$= \frac{1}{\mu}\mathrm{E}\left\{\sum_{m=1}^{\mu}\sum_{i=1}^{N-1}\left(\tilde{x}_{i_{m;\lambda}}^{(g)}\right)^2\right\} = \frac{1}{\mu}\mathrm{E}\left\{\sum_{m=1}^{\mu}\sum_{i=1}^{N-1}\left(\langle x_i^{(g)}\rangle + z_{i_{m;\lambda}}^{(g)}\right)^2\right\}. \tag{57}$$

If one compares (54) with (57) one sees that both expressions contain the same vector components in the inner parentheses. The difference, however, comes from the order of summation and the square operation. To work it out, (54) and (57) will be evaluated further. For (57) one obtains by straightforward calculations, omitting the generation counter g,

$$\begin{aligned}
\mathrm{E}\left\{\langle r^{(g+1)^2}\rangle\right\} &= \frac{1}{\mu}\mathrm{E}\left\{\sum_{m=1}^{\mu}\sum_{i=1}^{N-1}\left[\langle x_i\rangle^2 + 2\langle x_i\rangle z_{i_{m;\lambda}} + z_{i_{m;\lambda}}^2\right]\right\} \\
&= \frac{1}{\mu}\mathrm{E}\left\{\sum_{m=1}^{\mu}\left[r^2 + \sum_{i=1}^{N-1}\left(2\langle x_i\rangle z_{i_{m;\lambda}} + z_{i_{m;\lambda}}^2\right)\right]\right\} \\
&= r^2 + \frac{2}{\mu}\mathrm{E}\left\{\sum_{m=1}^{\mu}\sum_{i=1}^{N-1}\langle x_i\rangle z_{i_{m;\lambda}}\right\} + \frac{1}{\mu}\mathrm{E}\left\{\sum_{i=1}^{N-1}\sum_{m=1}^{\mu} z_{i_{m;\lambda}}^2\right\} \\
&= r^2 + 2\mathrm{E}\left\{\sum_{i=1}^{N-1}\langle x_i\rangle\frac{1}{\mu}\sum_{m=1}^{\mu} z_{i_{m;\lambda}}\right\} + \mathrm{E}\left\{\sum_{i=1}^{N-1}\langle z_i^2\rangle\right\}.
\end{aligned} \tag{58}$$

Similarly, one gets step by step for (54)

$$\mathrm{E}\left\{r^{(g+1)^2}\right\} = \frac{1}{\mu^2}\mathrm{E}\left\{\sum_{i=1}^{N-1}\left(\mu\langle x_i\rangle + z_{i_{1;\lambda}} + \cdots + z_{i_{2;\lambda}} + z_{i_{\mu;\lambda}}\right)^2\right\}$$

$$= \frac{1}{\mu^2}\mathrm{E}\left\{\sum_{i=1}^{N-1}\left(\mu^2\langle x_i\rangle^2 + 2\mu\langle x_i\rangle\sum_{m=1}^{\mu} z_{i_{m;\lambda}} + \sum_{m=1}^{\mu} z_{i_{m;\lambda}}^2 + \sum_{m=1}^{\mu}\sum_{k\neq m} z_{i_{m;\lambda}} z_{i_{k;\lambda}}\right)\right\}$$

$$= r^2 + 2\mathrm{E}\left\{\sum_{i=1}^{N-1}\langle x_i\rangle\frac{1}{\mu}\sum_{m=1}^{\mu} z_{i_{m;\lambda}}\right\} + \frac{1}{\mu}\mathrm{E}\left\{\sum_{i=1}^{N-1}\langle z_i^2\rangle\right\} + T_4, \tag{59}$$

with the residual term T_4

$$T_4 = \frac{1}{\mu^2}\mathrm{E}\left\{\sum_{i=1}^{N-1}\sum_{m=1}^{\mu}\sum_{k\neq m} z_{i_{m;\lambda}} z_{i_{k;\lambda}}\right\} = \frac{1}{\mu^2}\mathrm{E}\left\{\sum_{m=1}^{\mu}\sum_{k\neq m} \mathbf{z'}_{m;\lambda}^{\mathrm{T}}\mathbf{z}'_{k;\lambda}\right\} \tag{60}$$

where $\mathbf{z}'$ denotes the $(N-1)$-dimensional subvector of $\mathbf{z}$ dropping the 0-th component, i.e., $\mathbf{z}' := (z_1, z_2, \ldots, z_{N-1})^{\mathrm{T}}$. In order to proceed, the $(N-1)$-dimensional $\mathbf{z}'$ vector is decomposed in two components. One component will be in $\mathbf{e}_R$ direction (the radial progress direction) and will have the magnitude y. The other component (denoted by $\mathbf{h}$) will cover the remaining $(N-2)$ directions that are selection-invariant, i.e., $\mathbf{z}' = y\mathbf{e}_R + \mathbf{h}$. One obtains for the T_4 term (60)

$$\begin{aligned} T_4 &= \frac{1}{\mu^2}\sum_{m=1}^{\mu}\sum_{k\neq m} \mathrm{E}\left\{(y_{m;\lambda}\mathbf{e}_R + \mathbf{h}_{m;\lambda})^{\mathrm{T}}(y_{k;\lambda}\mathbf{e}_R + \mathbf{h}_{k;\lambda})\right\} \\ &= \frac{1}{\mu^2}\sum_{m=1}^{\mu}\sum_{k\neq m} \mathrm{E}\left\{y_{m;\lambda}y_{k;\lambda}\right\} + \frac{1}{\mu^2}\sum_{m=1}^{\mu}\sum_{k\neq m} \mathrm{E}\left\{\mathbf{h}_{m;\lambda}^{\mathrm{T}}\mathbf{h}_{k;\lambda}\right\}. \end{aligned} \tag{61}$$

Here, the orthogonality of $\mathbf{e}_R$ and $\mathbf{h}$, i.e., $\mathbf{e}_R^{\mathrm{T}}\mathbf{h} = 0$, has been taken into account. The second term in (61) is zero since the $\mathbf{h}$ components of two mutations are uncorrelated and have the mean value zero. The absolute value of the first term can be bounded from above by $\kappa\sigma^2$, where κ is asymptotically ($N \to \infty$) independent of N. Thus, (61) can be bounded by $\kappa\sigma^2$, i.e., $|T_4| \leq \kappa\sigma^2$.

As for the third term in (59), the expected value $\mathrm{E}\{\sum_{i=1}^{N-1}\langle z_i^2\rangle\}$ can be calculated using induced order statistics similar to the derivation of φ or $\mathrm{E}\{\langle r^{(g+1)^2}\rangle\}$ (for details see Oyman (1999, Appendix A)). It will be approximated by $(N-1)\sigma^2$ (asymptotically exact) neglecting the influence of selection on the length of $\mathbf{h}$:

$$\mathrm{E}\left\{\sum_{i=1}^{N-1}\langle z_i^2\rangle\right\} \sim (N-1)\sigma^2. \tag{62}$$

That is, this value is asymptotically ($N \to \infty$) the length square of the $(N-1)$-dimensional mutation vector *before* selection.

Since $|T_4| \leq \kappa\sigma^2$, (61) can be neglected in (59) as compared to (62). After these considerations, one obtains by subtracting (58) from (59)

$$\mathrm{E}\left\{r^{(g+1)^2}\right\} \simeq \mathrm{E}\left\{\langle r^{(g+1)^2}\rangle\right\} - \left(1 - \frac{1}{\mu}\right)(N-1)\sigma^2. \tag{63}$$

Therefore, the desired expected value can be calculated by determining $\mathrm{E}\{\langle r^{(g+1)^2}\rangle\}$.

4.2 On the Derivation of $\mathrm{E}\left\{\langle r^{(g+1)^2}\rangle\right\}$

4.2.1 Getting the Integral Representation

Denoting the random variable for the r^2 value of the m-th best descendant by u, one has for (56)

$$\mathrm{E}\left\{\langle r^{(g+1)^2}\rangle\right\} = \frac{1}{\mu}\sum_{m=1}^{\mu}\mathrm{E}\left\{\left(\tilde{r}_{m;\lambda}^{(g)}\right)^2\right\} = \frac{1}{\mu}\sum_{m=1}^{\mu}\int_0^{\infty} u p_{m;\lambda}(u)\,\mathrm{d}u. \tag{64}$$

According to its structure, this expected value integral is similar to (10). The density $p_{m;\lambda}(u)$ can be specified further by analogy to (11) and (12)

$$p_{m;\lambda}(u) = \frac{\lambda!}{(m-1)!(\lambda-m)!}p(u)P_{\mathrm{a}\,m;\lambda}(u), \tag{65}$$

with the density $p(u)$ of the random variate $u = r^2$ being the square of the distance of an offspring to the ridge axis given the (parental) centroid distance $R^2 := {r^{(g)}}^2$, and $P_{\mathrm{a}\, m;\lambda}(u)$ representing the acceptance probability for having this offspring as the m-th best leading to

$$P_{\mathrm{a}\, m;\lambda}(u) = \int_{-\infty}^{\infty} p(Q|u)\, [P(Q)]^{\lambda-m}\, [1-P(Q)]^{m-1}\, \mathrm{d}Q. \tag{66}$$

After substituting (66) in (65) and inserting this in (64) one obtains by exchanging the order of integrations

$$\mathrm{E}\left\{\langle {r^{(g+1)}}^2 \rangle\right\} = \frac{\lambda!}{\mu} \sum_{m=1}^{\mu} \int_{-\infty}^{\infty} \underbrace{\left(\int_{0}^{\infty} u\, p(u)\, p(Q|u)\, \mathrm{d}u \right)}_{I(Q)} \frac{[P(Q)]^{\lambda-m}\, [1-P(Q)]^{m-1}}{(m-1)!(\lambda-m)!}\, \mathrm{d}Q. \tag{67}$$

4.2.2 The Determination of $p(Q|u)$ and $p(u)$

The conditional density $p(Q|u)$ is easily obtained by considering the local quality function (15) using (5) and (7) with $R^2 = \left(r^{(g)}\right)^2$ and $u = \tilde{r}^2$, i.e.,

$$F(\mathbf{x}) = x_0 - dR^2, \qquad F(\mathbf{x}+\mathbf{z}) = x_0 + z_0 - du \tag{68}$$

leading to

$$Q_{\mathbf{x}}(\mathbf{z}) := F(\mathbf{x}+\mathbf{z}) - F(\mathbf{x}) = z_0 - d(u - R^2). \tag{69}$$

Since R^2 and u are fixed, Q depends on the $\mathcal{N}(0,\sigma^2)$ distributed random variate z_0 only. Thus, Q is $\mathcal{N}(-d(u-R^2),\sigma^2)$ distributed leading to the density

$$p(Q|u) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2} \left(Q + d(u - R^2)\right)^2 \right]. \tag{70}$$

As for the determination of $p(u)$, it shall be recalled that u is the length square of distance of an offspring to the ridge axis given the parental distance R. The distribution of the u variate has already been derived as a normal approximation in Beyer (1995a, 383–384). It reads for the case of $N-1$ dimensions, which make up the quadratic part of the parabolic ridge,

$$P(u) = \Phi\left(\frac{u - R^2 - (N-1)\sigma^2}{\sigma\sqrt{4R^2 + 2(N-1)\sigma^2}} \right). \tag{71}$$

After introducing the substitution

$$\mathrm{S} := \sqrt{4R^2 + 2(N-1)\sigma^2}, \tag{72}$$

the $p(u)$ density is obtained by differentiation of $P(u)$ with respect to u

$$p(u) = \frac{1}{\sqrt{2\pi}\sigma \mathrm{S}} \exp\left[-\frac{1}{2\sigma^2 \mathrm{S}^2} \left(u - R^2 - (N-1)\sigma^2\right)^2 \right]. \tag{73}$$

4.2.3 The Determination of the Inner Integral $I(Q)$

In order to evaluate the $I(Q)$ in (67), the expressions (73) and (70) are inserted in the inner integral. Using the substitution

$$t := \frac{u - R^2 - (N-1)\sigma^2}{\sigma \mathrm{S}} \quad \Rightarrow \quad \mathrm{d}u = \sigma \mathrm{S} \mathrm{d}t, \qquad \frac{u - R^2}{\sigma} = \mathrm{S}t + (N-1)\sigma, \tag{74}$$

one gets

$$\begin{aligned} I(Q) &= \frac{\mathrm{S}}{2\pi} \int_{-\infty}^{\infty} t\,\mathrm{e}^{-\frac{1}{2}t^2} \exp\left[-\frac{1}{2}\left(d\mathrm{S}t + \frac{Q}{\sigma} + d(N-1)\sigma\right)^2\right] \mathrm{d}t \\ &+ \frac{R^2 + (N-1)\sigma^2}{2\pi\sigma} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}t^2} \exp\left[-\frac{1}{2}\left(d\mathrm{S}t + \frac{Q}{\sigma} + d(N-1)\sigma\right)^2\right] \mathrm{d}t. \end{aligned} \tag{75}$$

The lower integration limit $t_l = -[R^2 + (N-1)\sigma^2]/(\sigma\mathrm{S}) \xrightarrow{N\to\infty} -\infty$. It was extended to $-\infty$ as an approximation for the $N < \infty$ finite-dimensional case. The equal sign used in (75) and the following formulae derived from (75) are, therefore, asymptotically ($N \to \infty$) correct. After introducing

$$a := d\mathrm{S} \qquad \text{and} \qquad b := \frac{Q}{\sigma} + d(N-1)\sigma \tag{76}$$

and considering the two integral formulae (Beyer, 2000, Equations (A.6, A.8))

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}t^2} \mathrm{e}^{-\frac{1}{2}(at+b)^2}\,\mathrm{d}t &= \frac{1}{\sqrt{1+a^2}} \exp\left(-\frac{1}{2}\frac{b^2}{1+a^2}\right), &(77)\\ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t\mathrm{e}^{-\frac{1}{2}t^2} \mathrm{e}^{-\frac{1}{2}(at+b)^2}\,\mathrm{d}t &= -\frac{ab}{(1+a^2)^{\frac{3}{2}}} \exp\left(-\frac{1}{2}\frac{b^2}{1+a^2}\right), &(78) \end{aligned}$$

the result reads

$$I(Q) = \left(\frac{\mathrm{S}}{\sqrt{2\pi}}\frac{-ab}{1+a^2} + \frac{R^2 + (N-1)\sigma^2}{\sqrt{2\pi}\sigma}\right) \frac{1}{\sqrt{1+a^2}} \exp\left[-\frac{1}{2}\frac{b^2}{1+a^2}\right]. \tag{79}$$

Considering M_Q and S_Q from (25), one can simplify

$$\begin{aligned} 1 + a^2 &= 1 + d^2\mathrm{S}^2 = 1 + d^2(4R^2 + 2(N-1)\sigma^2) = \frac{S_Q^2}{\sigma^2}, &(80)\\ b &= \frac{Q}{\sigma} + \frac{d(N-1)\sigma^2}{\sigma} = \frac{1}{\sigma}(Q - M_Q). &(81) \end{aligned}$$

This gives for (79) the final result

$$I(Q) = \left(-\frac{d\mathrm{S}^2\sigma}{S_Q^2}(Q - M_Q) + \frac{R^2 + (N-1)\sigma^2}{\sigma}\right) \frac{\sigma}{\sqrt{2\pi}S_Q} \exp\left[-\frac{1}{2}\left(\frac{Q - M_Q}{S_Q}\right)^2\right]. \tag{82}$$

4.2.4 Determining $\mathrm{E}\{\langle r^{(g+1)^2}\rangle\}$

Now, the outer integration in (67) can be tackled. After substituting (82) back in (67), we have

$$\begin{aligned}\mathrm{E}\left\{\langle r^{(g+1)^2}\rangle\right\} &= \frac{\lambda!}{\mu}\frac{\sigma}{\sqrt{2\pi}S_Q}\sum_{m=1}^{\mu}\int_{-\infty}^{\infty}\frac{[P(Q)]^{\lambda-m}\,[1-P(Q)]^{m-1}}{(m-1)!(\lambda-m)!} \\ &\times\left[\frac{R^2+(N-1)\sigma^2}{\sigma}-\frac{d\mathrm{S}^2\sigma}{S_Q^2}\,(Q-M_Q)\right]\exp\left[-\frac{(Q-M_Q)^2}{2S_Q^2}\right]\,\mathrm{d}Q. \qquad (83)\end{aligned}$$

Taking (30) into account and applying the substitution for s defined in (31), one obtains

$$\begin{aligned}\mathrm{E}\left\{\langle r^{(g+1)^2}\rangle\right\} &= \frac{\lambda!}{\sqrt{2\pi}\mu}\int_{-\infty}^{\infty}\left[R^2+(N-1)\sigma^2-\frac{d\mathrm{S}^2\sigma^2}{S_Q}\,s\right]\mathrm{e}^{-\frac{1}{2}s^2} \\ &\times\sum_{m=1}^{\mu}\frac{[\Phi(s)]^{\lambda-m}\,[1-\Phi(s)]^{m-1}}{(m-1)!(\lambda-m)!}\,\mathrm{d}s. \qquad (84)\end{aligned}$$

Using the identity in (36) for $P=1-\Phi(s)$, one gets

$$\begin{aligned}\mathrm{E}\left\{\langle r^{(g+1)^2}\rangle\right\} &= \frac{\lambda-\mu}{\sqrt{2\pi}}\binom{\lambda}{\mu}\int_{-\infty}^{\infty}\left[R^2+(N-1)\sigma^2-\frac{d\mathrm{S}^2\sigma^2}{S_Q}\,s\right]\mathrm{e}^{-\frac{1}{2}s^2} \\ &\times\int_0^{\Phi(s)}x^{\lambda-\mu-1}(1-x)^{\mu-1}\,\mathrm{d}x\,\mathrm{d}s. \qquad (85)\end{aligned}$$

In the next step, the order of integrations is exchanged yielding

$$\begin{aligned}\mathrm{E}\left\{\langle r^{(g+1)^2}\rangle\right\} &= \frac{\lambda-\mu}{\sqrt{2\pi}}\binom{\lambda}{\mu}\int_0^1 x^{\lambda-\mu-1}(1-x)^{\mu-1} \\ &\times\int_{\Phi^{-1}(x)}^{\infty}\left[R^2+(N-1)\sigma^2-\frac{d\mathrm{S}^2\sigma^2}{S_Q}\,s\right]\mathrm{e}^{-\frac{1}{2}s^2}\,\mathrm{d}s\,\mathrm{d}x. \qquad (86)\end{aligned}$$

After the substitution $x=\Phi(y)$, $\mathrm{d}x=\exp(-\frac{1}{2}y^2)\mathrm{d}y/\sqrt{2\pi}$, Equation (86) becomes

$$\begin{aligned}\mathrm{E}\left\{\langle r^{(g+1)^2}\rangle\right\} &= \frac{\lambda-\mu}{2\pi}\binom{\lambda}{\mu}\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{1}{2}y^2}[\Phi(y)]^{\lambda-\mu-1}[1-\Phi(y)]^{\mu-1} \\ &\times\int_y^{\infty}\left[R^2+(N-1)\sigma^2-\frac{d\mathrm{S}^2\sigma^2}{S_Q}\,s\right]\mathrm{e}^{-\frac{1}{2}s^2}\,\mathrm{d}s\,\mathrm{d}y. \qquad (87)\end{aligned}$$

The inner integral can be calculated easily. One obtains

$$\sqrt{2\pi}(1-\Phi(y))(R^2+(N-1)\sigma^2)-\frac{d\mathrm{S}^2\sigma^2}{S_Q}\mathrm{e}^{-\frac{1}{2}y^2}. \qquad (88)$$

Taking the definition (40) of the $c_{\mu/\mu,\lambda}$ progress coefficient into account and the integral identity

$$\frac{\lambda-\mu}{\sqrt{2\pi}}\binom{\lambda}{\mu}\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{1}{2}y^2}[\Phi(y)]^{\lambda-\mu-1}[1-\Phi(y)]^{\mu}\mathrm{d}y\;\equiv\;1, \qquad (89)$$

one finally obtains

$$\mathrm{E}\left\{\langle r^{(g+1)^2}\rangle\right\}=R^2+(N-1)\sigma^2-\frac{d\mathrm{S}^2\sigma^2}{S_Q}c_{\mu/\mu,\lambda}. \qquad (90)$$

4.3 The Radial Dynamics of the $(\mu/\mu_I, \lambda)$-ES

4.3.1 Determination of the $\mathrm{E}\{r^{(g+1)^2}\}$

After the lengthy calculations, we are now able to investigate the expected value dynamics of the squared distance of the parental centroid to the ridge axis. Taking the asymptotically $(N \to \infty)$ correct formula (63), one obtains with (90) and $R^2 = r^{(g)^2}$

$$\mathrm{E}\left\{r^{(g+1)^2}\right\} = r^{(g)^2} - d\frac{\mathrm{S}^2\sigma^2}{S_Q}c_{\mu/\mu,\lambda} + \frac{1}{\mu}(N-1)\sigma^2, \tag{91}$$

and finally by back substitution of S and S_Q using (72) and (25)

$$\boxed{\mathrm{E}\left\{r^{(g+1)^2}\right\} = r^{(g)^2} - \frac{d\left(4r^{(g)^2} + 2(N-1)\sigma^2\right)}{\sqrt{1+d^2\left(4r^{(g)^2}+2(N-1)\sigma^2\right)}}c_{\mu/\mu,\lambda}\sigma + \frac{1}{\mu}(N-1)\sigma^2.} \tag{92}$$

This asymptotically exact equation also holds for the generalized parabolic ridge (8). The state equation can be used for several purposes. The mean value dynamics of r^2 of the parental centroid over generations can be investigated. The expected value of $r^{(g+1)^2}$ can be determined for a given $r^{(g)^2}$, or alternatively, the expected progress in r direction. Furthermore, (92) can be used to investigate the steady state behavior of the $(\mu/\mu_I, \lambda)$-ES on the parabolic ridge to be discussed next.

4.3.2 Steady State Behavior and Simulations

If the $(\mu/\mu_I, \lambda)$-ES operates with constant mutation strength σ, one observes in real ES runs after a certain transient time the establishment of a stable expected value $R^{(\infty)}$ of the centroid's distance to the ridge axis. This expected value $R^{(\infty)} = \mathrm{E}\{r^{(\infty)}\}$ can be estimated using Equation (92). Provided that the fluctuations of r around $R^{(\infty)}$ are sufficiently small, one can assume $R^{(\infty)} \approx \sqrt{\mathrm{E}\{r^{(\infty)^2}\}}$. The expected value $\mathrm{E}\{r^{(\infty)^2}\}$ is obtained from (92) by inserting the steady state condition $\mathrm{E}\{r^{(g+1)^2}\} \overset{!}{=} \mathrm{E}\{r^{(g)^2}\} =: x$ in (92) resulting in a quadratic equation in x. This equation can be straightforwardly solved for x, and with $R^{(\infty)} = \sqrt{x}$, one finally obtains

$$\boxed{R^{(\infty)} = \frac{(N-1)\sigma}{2\mu c_{\mu/\mu,\lambda}}\sqrt{\frac{1}{2}\left[1+\sqrt{1+\left(\frac{2\mu c_{\mu/\mu,\lambda}}{d(N-1)\sigma}\right)^2}\right] - \frac{2\mu^2 c^2_{\mu/\mu,\lambda}}{N-1}}.} \tag{93}$$

Interestingly, (93) contains, as a limit case, the residual distance formula of the $(N-1)$ dimensional sphere model $(N \to \infty)$. In order to see this, let us recall that parabolic ridge, defined by (5), gets more and more spherical as d increases. For $d \to \infty$, the sphere model is asymptotically reached. If we perform this limit operation in (93), one obtains asymptotically the residual distance formula of the sphere model

$$R^{(\infty)}_{sphere} = \frac{(N-1)\sigma}{2\mu c_{\mu/\mu,\lambda}}. \tag{94}$$

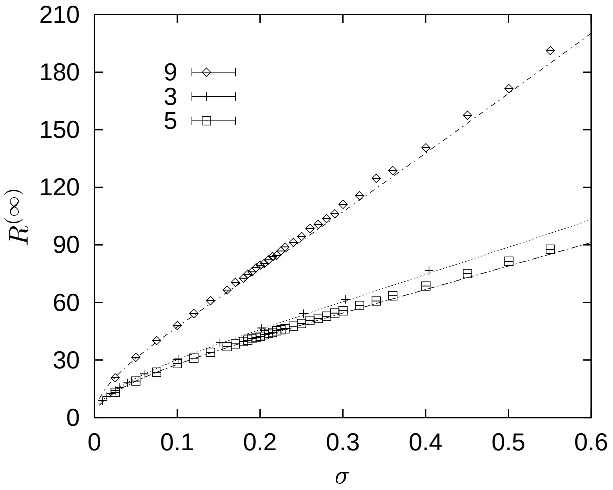

Figure 5: The stationary distance $R^{(\infty)}$ versus the mutation strength σ for the $(\mu/\mu_{\rm I}, 10)$-ES on the parabolic ridge, $\mu \in \{3, 5, 9\}$. The $R^{(\infty)}$ values are predicted correctly.

As one can see in (93) and even better in (94), the main effect of recombination is in the reduction of the residual distance to the ridge axis. This reduction can be up to a factor of $1/\mu$ compared to the (μ, λ)-ES. This reduction of the harmful components, also known as the genetic repair (GR) effect (see Beyer (1997)), is also the reason for the improved progress behavior in axis direction measured by φ. As one can see in Equation (41), a reduced $R^{(\infty)}$ results in a reduced denominator leading to a larger φ.

In Figure 5, the predictions of the stationary $R^{(\infty)}$ formula (93) are compared with real simulations. The simulations were performed for the $(\mu/\mu_{\rm I}, 10)$-ES with $\mu \in \{3, 5, 9\}$, $N = 1,000$, and $d = 0.01$. The simulation results are well predicted by the theoretical formula. The three curves have lines as asymptotes, as can be inferred from the figure. Its slope can be easily calculated from (93) considering the σ asymptotics.

4.4 The Steady State Behavior of the $(\mu/\mu_{\rm D}, \lambda)$-ES

The stationary distance $R^{(\infty)}$ to the ridge axis of the $(\mu/\mu_{\rm D}, \lambda)$-ES is obtained using the same surrogate mutation approach demonstrated in Subsection 3.7 for the φ calculation. By replacing σ by $\sqrt{\mu}\sigma$ in (93) for the $(\mu/\mu_{\rm I}, \lambda)$-ES, one gets

$$\boxed{R^{(\infty)} = \frac{(N-1)\sigma}{2\sqrt{\mu}c_{\mu/\mu,\lambda}} \sqrt{\frac{1}{2}\left[1 + \sqrt{1 + \left(\frac{2\sqrt{\mu}c_{\mu/\mu,\lambda}}{d(N-1)\sigma}\right)^2}\,\right] - \frac{2\mu^2 c^2_{\mu/\mu,\lambda}}{N-1}}.} \tag{95}$$

Note that this formula is only valid if $\mathbf{v}$ in (8) is diagonal in the search space. The stationary $R^{(\infty)}$ values obtained for intermediate and dominant recombination are compared in Figure 6. The results are obtained for $N = 1,000$ and $d = 0.01$ on the parabolic ridge for a series of σ values. The simulations for the $(5/5_{\rm I}, 10)$-ES are compared with the theoretical formula (93) and for the $(5/5_{\rm D}, 10)$-ES with (95), respectively. The analytically

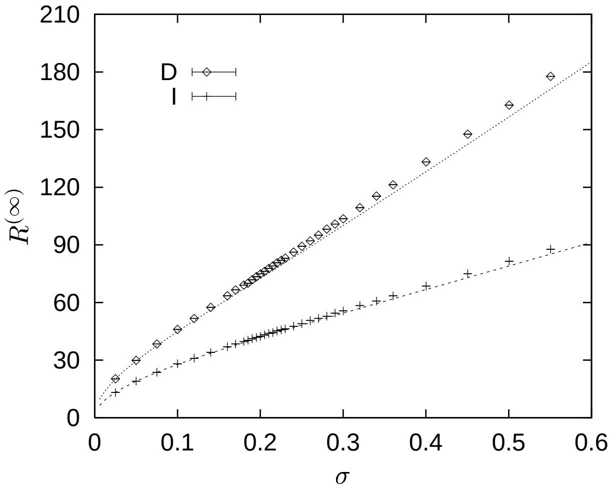

Figure 6: The stationary distance $R^{(\infty)}$ versus the mutation strength σ for the $(5/5_\mathrm{I}, 10)$-ES and the $(5/5_\mathrm{D}, 10)$-ES on the parabolic ridge. The $R^{(\infty)}$ values are predicted better for the $(5/5_\mathrm{I}, 10)$-ES case.

derived formulae describe the simulation results well; however, the one for intermediate recombination yields less estimation error. The asymptotic limit is a line in both cases, which can be inferred from the figure.

5 Summary and Outlook

The progress behavior of the $(\mu/\mu, \lambda)$-ES has been investigated in this work on the parabolic ridge test function. The theoretical analysis was carried out for the intermediate recombination (i.e., the $(\mu/\mu_\mathrm{I}, \lambda)$-ES). Thereafter, the formulae for dominant recombination, i.e., the $(\mu/\mu_\mathrm{D}, \lambda)$-ES, were obtained from the $(\mu/\mu_\mathrm{I}, \lambda)$-ES formulae using the surrogate mutation approach, which is valid for $N \to \infty$.

The progress behavior was investigated using the progress rate φ and the distance r to the ridge axis. Concerning the latter one, the state equation for the $(\mu/\mu_\mathrm{I}, \lambda)$-ES was derived. Furthermore, the stationary $R^{(\infty)}$ value was obtained for the $(\mu/\mu_\mathrm{I}, \lambda)$-ES as well as for the $(\mu/\mu_\mathrm{D}, \lambda)$-ES.

The static progress rate formula was derived for the $(\mu/\mu_\mathrm{I}, \lambda)$-ES. The stationary progress rate formula was obtained thereafter using the stationary $R^{(\infty)}$ value. These two formulae were obtained for the $(\mu/\mu_\mathrm{D}, \lambda)$-ES using the surrogate mutation model from the respective formulae for the $(\mu/\mu_\mathrm{I}, \lambda)$-ES.

The dependence of the progress rate performance on μ was investigated for the example of the $(\mu/\mu_\mathrm{I}, 10)$-ES, and the theoretical results were compared with the simulations. It has also been shown that recombination increases the steady state progress rate over the $(1, \lambda)$-ES, provided that $\mu c^2_{\mu/\mu,\lambda} > c_{1,\lambda}$ and that σ is sufficiently large. It turns out that the performance surplus of the recombinant strategies is due to the reduction of the residual distance of the population to the ridge axis. The progress rates of intermediate and dominant recombination were compared to each other for the example $(5/5, 10)$-ES. The stationary

$R^{(\infty)}$ value was compared with simulation results for different μ values and intermediate recombination, and thereafter for intermediate and dominant recombination.

There remains a lot of future work. The results obtained are exact for $N \to \infty$. Its predictive value for small N should be further improved. Besides these technical challenges, the progress behavior of (μ, λ)-ES algorithms, i.e., ES algorithms with $\mu > 1$ but without recombination, remains to be investigated as well as the inclusion of self-adaptation. This work considered only local performance measures. However, it can be extended to global measures addressing questions concerning the computational time complexity of the ES. This topic will be treated in a forthcoming paper. Furthermore, one can extend the analysis beyond the parabolic ridge to other ridge functions.

Acknowledgments

The authors thank the anonymous reviewers and Dirk V. Arnold for their helpful comments, remarks, and corrections.

The first author was partially supported by a Ph.D. scholarship from DAAD (German Academic Exchange Service) and by the Deutsche Forschungsgemeinschaft (DFG) grant Be1578/6-1. The second author acknowledges support as a Heisenberg Fellow of the DFG under grant Be1578/4-1.

References

Abramowitz, M. and Stegun, I. A. (1984). *Pocketbook of Mathematical Functions.* Verlag Harri Deutsch, Thun, Germany.

Bäck, T. (1997). Self-adaptation. In Bäck, T., Fogel, D. and Michalewicz, Z., editors, *Handbook of Evolutionary Computation*, pages C7.1:1–C7.1:15, Oxford University Press, New York, New York.

Bäck, T. and Schwefel, H.-P. (1993). An Overview of Evolutionary Algorithms for Parameter Optimization. *Evolutionary Computation*, 1(1):1–23.

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

Beyer, H.-G. (1994). Towards a Theory of 'Evolution Strategies'. Progress Rates and Quality Gain for $(1 \overset{+}{,} \lambda)$–Strategies on Nearly Arbitrary Fitness Functions. In Davidor, Y., Männer, R. and Schwefel, H.-P., editors, *Parallel Problem Solving from Nature (PPSN III) – International Conference on Evolutionary Computation*, pages 58–67, Springer, Heidelberg, Germany.

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

Beyer, H.-G. (1995b). Toward a Theory of Evolution Strategies: On the Benefits of Sex-the (μ, λ)–Theory. *Evolutionary Computation*, 3(1):81–111.

Beyer, H.-G. (1996). Toward a Theory of Evolution Strategies: Self-Adaptation. *Evolutionary Computation*, 3(3):311–347.

Beyer, H.-G. (1997). An Alternative Explanation for the Manner in which Genetic Algorithms Operate. *BioSystems*, 41(1):1–15.

Beyer, H.-G. (2000). *The Theory of Evolution Strategies*. Springer, Heidelberg, Germany.

Beyer, H.-G. and Deb, K. (1999). On the Analysis of Self-Adaptive Evolutionary Algorithms. Technical Report CI 69/99, University of Dortmund, Dortmund, Germany.

Herdy, M. (1992). Reproductive Isolation as Strategy Parameter in Hierarchically Organized Evolution Strategies. In Männer, R. and Manderick, B., editors, *Parallel Problem Solving from Nature 2*, pages 207–217, Elsevier, Amsterdam, The Netherlands.

Oyman, A. I. (1999). *Convergence Behavior of Evolution Strategies on Ridge Functions*. Ph.D. thesis, Department of Computer Science, University of Dortmund, Dortmund, Germany.

Oyman, A. I., Beyer, H.-G. and Schwefel, H.-P. (1997). Analysis of a Simple ES on the "Parabolic Ridge". Technical Report SyS-2/97, Department of Computer Science, Systems Analysis Research Group, University of Dortmund, Dortmund, Germany.

Oyman, A. I., Beyer, H.-G. and Schwefel, H.-P. (1998). Where Elitists Start Limping: Evolution Strategies at Ridge Functions. In Eiben, A. E., Bäck, T., Schoenauer, M. and Schwefel, H.-P., editors, *Parallel Problem Solving from Nature (PPSN V) - International Conference on Evolutionary Computation*, pages 34–43, Springer, Berlin, Germany.

Oyman, A. I., Beyer, H.-G. and Schwefel, H.-P. (2000). Analysis of the $(1, \lambda)$-ES on the Parabolic Ridge. *Evolutionary Computation*, 8(3):249–265.

Rechenberg, I. (1973). *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Verlag Frommann-Holzboog, Stuttgart-Bad Cannstatt, Germany. (In German).

Rechenberg, I. (1994). *Evolutionsstrategie'94*. Frommann-Holzboog, Stuttgart, Germany. (In German).

Scheel, A. (1985). *Beitrag zur Theorie der Evolutionsstrategie*. Doctoral thesis, Technical University of Berlin, Berlin, Germany. (In German).

Schwefel, H.-P. (1977). *Numerische Optimierung von Computer–Modellen mittels der Evolutionsstrategie*, volume 26 of *Interdisciplinary Systems Research*. Birkhäuser, Basel, Switzerland. (In German).