Toward a Theory of Evolution Strategies: On the Benefits of Sex - the $(\mu/\mu, \lambda)$ Theory

Hans-Georg Beyer,
University of Dortmund,
Department of Computer Science,
Systems Analysis Research Group,
D-44221 Dortmund,
Germany

beyer@LS11.informatik.uni-dortmund.de

Abstract
The multi-recombinant $(\mu/\mu, \lambda)$-Evolution Strategy (ES) is investigated for real-valued, N-dimensional parameter spaces. The analysis includes the intermediate recombination and the dominant recombination as well. These investigations are done for the spherical model first. The problem of the optimal population size depending on the parameter space dimension N is solved. A method extending the results obtained for the spherical model to non-spherical success domains is presented.

The power of sexuality is discussed and it is shown that this power does not mainly stem from the 'combination' of 'good properties' of the mates (building block hypothesis) but rather from *genetic repair* diminishing the influence of harmful mutations. The dominant recombination is analyzed by introduction of surrogate mutations leading to the concept of *species*. Conclusions for EAs including GAs are drawn.

Keywords
multi-membered Evolution Strategies (ES), progress rate theory, multi-recombination: intermediate and dominant/discrete, benefits of sexuality, genetic variety vs. genetic repair, building block hypothesis, non-spherical success domains, multi-recombinant GA

1 Introduction

One logical continuation to the analysis of Evolution Strategies (ES) presented so far (Beyer, 1993; Beyer, 1995) concerns the recombination. Most of the users of EA (Evolutionary Algorithm) would agree that one basic operator of an EA working with a population should be the recombination. And there are people who even belief that recombination is the most important (search) operator (this holds especially for the field of GAs, see e. g. (Goldberg, 1989)).

In order to shed light on the effect of recombination the so-called multi-recombinant $(\mu/\mu, \lambda)$-strategies are investigated. These are special variants of the $(\mu/\rho, \lambda)$-strategies. There are two variants of recombination - the intermediate and the dominant to be introduced in section 2.

The intermediate $(\mu/\mu, \lambda)$-strategy is analyzed in section 3 for the spherical fitness landscape. It can be easily tackled because intermediate recombination concentrates the parents in the 'center of mass'. $(\mu/\mu, \lambda)$-strategies exhibit remarkably high progress rates. This is due to the multi-recombination. The advantages of 'sexuality' will be discussed and it will be shown that its power stems mainly from *genetic repair*, whereas the building block hypothesis describes at best a side effect.

A rather difficult problem concerns the dominant $(\mu/\mu,\lambda)$-strategy to be discussed in section 4. This ES is very similar to GAs with uniform crossover. In the case of dominant recombination a simple mutation-recombination-model will be developed that provides an estimate for a 'mutation substitute' to be used in the progress formula of the intermediate strategy. A rough approximation for the progress rate will be derived.

The analysis of the dominant recombination reveals a very interesting *synergistic effect* concerning mutation and recombination: Iterative application of these two operators produces populations reminiscent of *species* observed in biology.

Section 5 contains information on how to generalize the results which were obtained for the sphere. A differential geometry approach will be outlined that provides a local approximation of the fitness landscape by a sphere. Thus, it will be possible to extend the validity of the model to non-spherical success domains.

The closing section 6 resumes the discussion on the benefits of recombination. Especially the dominant recombination will be considered. Since the concept is very general it will be applied to the GA theory, too.

This paper relies strongly on notations used and results presented in the last EC-Journal article of the author (Beyer, 1995). Especially in section 3.1.1, 3.2.1, and 3.2.2 we will take advantage from methods developed for the analysis of the (μ,λ)-ES. Therefore it will be good to have that EC-issue to hand.

2 The Intermediate and the Dominant $(\mu/\rho,\lambda)$ and $(\mu/\mu,\lambda)$-Algorithms

Sexuality, in the spirit of recombining hereditary factors, is a universal mechanism in biology. Therefore, it is not a surprise that recombination has been regarded as an important evolution mechanism for the ES, too. From SCHWEFEL's early work (Schwefel, 1974) onwards we can observe a continuos spreading of recombination techniques in ES algorithms.

Extending SCHWEFEL's (μ,λ) notation RECHENBERG has introduced the $(\mu/\rho,\lambda)$ notation for the (multi-)recombinant ES (Rechenberg, 1978). Here the new strategy parameter ρ determines the number of parents that multi-recombine to form a (one) new offspring. In contrast to the biological standard, $\rho=2$ (at least for the higher creatures), it is worth using multi-mixing, i. e. $\rho>2$, for ES applications.

As in biology, there are two different recombination patterns - *intermediate* and *dominant* recombination. The author proposes the

$$(\mu/\rho_I,\lambda) \qquad \text{and} \qquad (\mu/\rho_D,\lambda)$$

notation for their distinction.

Intermediate recombination is some kind of averaging. In $(\mu/\rho_I,\lambda)$ strategies ρ individuals $\mathbf{y}_{m_\nu}$, $\nu=1,\dots\rho$ from the parental pool $\{\mathbf{y}_1,\dots\mathbf{y}_\mu\}$ (consisting of μ parents) are chosen at random. The averaging is simply the determination of the *'center of mass'*

$$\texttt{intermediate } \mu/\rho \texttt{ recombination:} \qquad \langle\mathbf{y}\rangle=\frac{1}{\rho}\sum_{\nu=1}^{\rho}\mathbf{y}_{m_\nu}.$$

After recombination the mutation $\mathbf{Z}$ is additively applied. Thus, the offspring $\tilde{\mathbf{y}}_l$ can be expressed by[1]

$$\underline{(\mu/\rho_I,\lambda)\text{-}\mathtt{ES:}}\quad \tilde{\mathbf{y}}_l = \frac{1}{\rho}\sum_{\nu=1}^{\rho}\mathbf{y}_{m_\nu} + \mathbf{Z}, \qquad m_\nu = \mathtt{Random}\{1\,\dots\,\mu\}\ \wedge\ m_{\nu_1} \ne m_{\nu_2} \quad\Leftrightarrow\quad \nu_1 \ne \nu_2.$$

The mutation vector $\mathbf{Z}$ is - as usual - generated by iid normal random numbers with zero mean and standard deviation σ for each component. Selection acts on the $\tilde{\mathbf{y}}_l$-offspring and is equal to the (μ,λ)-strategies.

Because there are μ different parents, the theoretical investigation of the $(\mu/\rho_I,\lambda)$-strategies is at least as difficult as the (μ,λ) strategy. However, there is one exception - the $(\mu/\mu_I,\lambda)$ strategy. Because $\rho=\mu$ holds, there is no random mating. All parents are involved in the recombination

$$\underline{(\mu/\mu_I,\lambda)\text{-}\mathtt{ES:}}\quad \tilde{\mathbf{y}}_l = \frac{1}{\mu}\sum_{m=1}^{\mu}\mathbf{y}_m + \mathbf{Z}\,. \tag{1}$$

That is, each offspring $\tilde{\mathbf{y}}_l$ is generated from the 'center of mass individual' $\langle\mathbf{y}\rangle = \sum\mathbf{y}_m/\mu$ by mutation. The theoretical treatment of this strategy will be performed in the next section.

In order to describe the *dominant recombination* the $\mathbf{y}_m$-vectors are decomposed by an orthogonal basis $\{\mathbf{e}_i\}$

$$\mathbf{y} = \sum_{i=1}^{N} y_i\,\mathbf{e}_i.$$

An offspring $\tilde{\mathbf{y}}_l$ is constituted from ρ different, randomly chosen parents $\mathbf{y}_{\pi^l}$ with index π^l_ν

$$\pi^l_\nu := \mathtt{Random}\{1\,\dots\,\mu\}\ \wedge\ \pi^l_{\nu_1} \ne \pi^l_{\nu_2} \quad\Leftrightarrow\quad \nu_1 \ne \nu_2.$$

by

$$\underline{(\mu/\rho_D,\lambda)\text{-}\mathtt{ES:}}\quad \tilde{\mathbf{y}}_l = \sum_{i=1}^{N}\left(\mathbf{e}_i^T\mathbf{y}_{m_i}\right)\mathbf{e}_i + \mathbf{Z}, \qquad m_i = \mathtt{Random}\{\pi^l_1,\pi^l_2,\dots\pi^l_\rho\}.$$

That is, the coordinate i of the $\tilde{\mathbf{y}}_l$-vector is given by random choice of one of the coordinate-i-values from the mates of the index set $\{\pi^l_1,\pi^l_2,\dots\pi^l_\rho\}$ and an additive mutation z_i. Here $m_i = \mathtt{Random}\{\pi^l_1,\dots\pi^l_\rho\}$ indicates the random choice of an element from the set $\{\pi^l_1,\dots\pi^l_\rho\}$ with selection probability $1/\rho$. The constitution of the μ new parents by selection from the $\{\tilde{\mathbf{y}}_l\}$ is equal to the (μ,λ) ES.

The theoretical analysis of dominant recombination is still at the beginning. Even for the $\rho=\mu$ case

$$\underline{(\mu/\mu_D,\lambda)\text{-}\mathtt{ES:}}\quad \tilde{\mathbf{y}}_l = \sum_{i=1}^{N}\left(\mathbf{e}_i^T\mathbf{y}_{m_i}\right)\mathbf{e}_i + \mathbf{Z}, \qquad m_i = \mathtt{Random}\{1,2,\dots\,\mu\}, \tag{2}$$

only a rough approximation, up until now, has been developed to be discussed in section 4.

[1]In practice it is rather elaborate to produce ρ random indices m_ν all different from each other. A slightly different algorithm might as well allow 'self-fertilization' by sampling the ρ m_ν-numbers from a simple random number generator.

3 The $(\mu/\mu_I, \lambda)$-Theory

3.1 Basic Aspects of the $(\mu/\mu_I, \lambda)$-Theory

3.1.1 Definition of the Progress Rate $\varphi_{\mu/\mu,\lambda}$

Let us assume the optimum point $\hat{\mathbf{y}}$ of a spherical fitness landscape at $\hat{\mathbf{y}} = \mathbf{0}$. The parents $\mathbf{y}_m^{(g)}$ at generation (g) may form the center of mass $\langle \mathbf{y} \rangle^{(g)}$

$$\langle \mathbf{y} \rangle^{(g)} = \frac{1}{\mu} \sum_{m=1}^{\mu} \mathbf{y}_m^{(g)} =: \mathbf{R}. \tag{3}$$

Since $\hat{\mathbf{y}} = \mathbf{0}$ is assumed, $\langle \mathbf{y} \rangle^{(g)}$ is equal to the radius vector $\mathbf{R}$ measuring the distance of the center of mass to the optimum. The distance is given by $R = \|\langle \mathbf{y} \rangle^{(g)}\|$. According to (1) λ offspring $\tilde{\mathbf{y}}_l$ are generated from $\langle \mathbf{y} \rangle^{(g)}$ by mutations $\mathbf{Z}$. Their distance to the optimum may be denoted by $\tilde{R}_l$. From $\hat{\mathbf{y}} = \mathbf{0}$ it follows $\tilde{\mathbf{R}}_l = \tilde{\mathbf{y}}_l$. By fitness selection the best μ offspring $\tilde{\mathbf{y}}_l$ are chosen obeying the ordering

$$\tilde{R}_{1:\lambda} \le \tilde{R}_{2:\lambda} \le \ldots \le \tilde{R}_{m:\lambda} \le \ldots \le \tilde{R}_{\mu:\lambda}\,. \tag{4}$$

Let us denote the mth-best offspring by $\tilde{\mathbf{y}}_{m;\lambda}$. Then, the parents of the $(g+1)$ generation are given by $\mathbf{y}_m^{(g+1)} = \tilde{\mathbf{y}}_{m;\lambda}$. Hence, the new center of mass reads

$$\langle \mathbf{y} \rangle^{(g+1)} = \frac{1}{\mu} \sum_{m=1}^{\mu} \mathbf{y}_m^{(g+1)} = \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\mathbf{y}}_{m;\lambda} \tag{5}$$

and has the distance $\|\langle \mathbf{y} \rangle^{(g+1)}\|$ to the optimum. Progress is measured as the distance change $\Delta R = R - \|\langle \mathbf{y} \rangle^{(g+1)}\|$ from (g) to $(g+1)$. Thus, the progress rate φ is the expectation of ΔR

$$\varphi_{\mu/\mu,\lambda} = E\{\Delta R\} = E\left\{R - \|\langle \mathbf{y} \rangle^{(g+1)}\|\right\}. \tag{6}$$

From (1) and (3) we have for the mth-best offspring

$$\tilde{\mathbf{y}}_{m;\lambda} = \mathbf{R} + \mathbf{Z}_{m;\lambda},$$

i. e. the mth-best offspring is generated from the 'mth-best mutation'. $\mathbf{Z}_{m;\lambda}$ can be decomposed

$$\tilde{\mathbf{y}}_{m;\lambda} = R\mathbf{e}_R - x_{m;\lambda}\mathbf{e}_R + \mathbf{h}_{m;\lambda} \tag{7}$$

with $x_{m;\lambda}$, the component of $\mathbf{Z}_{m;\lambda}$ in $-\mathbf{R}$ direction, and the perpendicular part $\mathbf{h}_{m;\lambda}$. The center of mass at generation $(g+1)$ is obtained from (5) and (7)

$$\langle \mathbf{y} \rangle^{(g+1)} = (R - \langle x \rangle)\,\mathbf{e}_R + \langle \mathbf{h} \rangle \tag{8}$$

with

$$\langle x \rangle := \frac{1}{\mu} \sum_{m=1}^{\mu} x_{m;\lambda} \qquad \text{and} \qquad \langle \mathbf{h} \rangle := \frac{1}{\mu} \sum_{m=1}^{\mu} \mathbf{h}_{m;\lambda}. \tag{9}$$

Substitution of (8) into (6) yields

$$\begin{aligned} \varphi_{\mu/\mu,\lambda} &= E\left\{R - \sqrt{(R - \langle x \rangle)^2 + \langle \mathbf{h} \rangle^2}\right\} \\ \varphi_{\mu/\mu,\lambda} &= E\left\{R - \sqrt{R^2 + \langle \mathbf{h} \rangle^2 - 2R\,\langle x \rangle + \langle x \rangle^2}\right\}. \end{aligned} \tag{10}$$

3.1.2 Stochastical Approximation for $\varphi_{\mu/\mu,\lambda}$

The progress rate expression (10) can be further simplified

$$
\begin{aligned}
& R-\sqrt{R^2+\langle\mathbf{h}\rangle^2-2R\,\langle x\rangle+\langle x\rangle^2} \\
= \quad & R-\sqrt{R^2+\langle\mathbf{h}\rangle^2}\sqrt{1-\frac{2R\,\langle x\rangle}{R^2+\langle\mathbf{h}\rangle^2}+\frac{\langle x\rangle^2}{R^2+\langle\mathbf{h}\rangle^2}} \\
\approx \quad & R-\sqrt{R^2+\langle\mathbf{h}\rangle^2}\left(1-\frac{R\,\langle x\rangle}{R^2+\langle\mathbf{h}\rangle^2}-\frac{1}{2}\left(\frac{R\,\langle x\rangle}{R^2+\langle\mathbf{h}\rangle^2}\right)^2+\ldots\right).
\end{aligned}
\tag{11}
$$

Here we have used the assumptions

$$
E\{\langle x\rangle^2\}\ll E\{\langle\mathbf{h}\rangle^2\} \qquad \text{and} \qquad E\{|\langle x\rangle|\}\ll R, \tag{12}
$$

that can be easily justified by geometrical considerations and the Taylor expansion $\sqrt{1-a}=1-a/2-a^2/8+\ldots$. Neglecting the third term in the parenthesis of (11) gives for (10)

$$
\varphi_{\mu/\mu,\lambda}=E\left\{R\left(1-\sqrt{1+\frac{\langle\mathbf{h}\rangle^2}{R^2}}\right)+\frac{\langle x\rangle}{\sqrt{1+\frac{\langle\mathbf{h}\rangle^2}{R^2}}}\right\}. \tag{13}
$$

As can be seen from (10) or (13), φ depends on the two variates $\langle\mathbf{h}\rangle^2$ and $\langle x\rangle$.

Due to findings presented in (Beyer, 1995, eqn. (8/9)), the standard deviation of $(\mathbf{h})^2$ is small compared to its expectation

$$
D\{(\mathbf{h})^2\}/E\{(\mathbf{h})^2\}\sim\sqrt{2/N}.
$$

That is, $\mathbf{h}^2$ exhibits only small fluctuations around its mean value. Therefore, this holds for the $\langle\mathbf{h}\rangle^2$ variate, too. If one further assumes that $\langle x\rangle$ and $\langle\mathbf{h}\rangle^2$ are nearly independent from each other, then the simplest form of the stochastical approximation can be applied: The curly brace from (13) as a function of $\langle\mathbf{h}\rangle^2$ and $\langle x\rangle$ is expanded in a Taylor series about $E\{\langle\mathbf{h}\rangle^2\}$ and $E\{\langle x\rangle\}$. Breaking off after the linear terms and taking the expectation yields

$$
\varphi_{\mu/\mu,\lambda}=R\left(1-\sqrt{1+\frac{\overline{\langle\mathbf{h}\rangle^2}}{R^2}}\right)+\frac{\overline{\langle x\rangle}}{\sqrt{1+\frac{\overline{\langle\mathbf{h}\rangle^2}}{R^2}}}\,, \tag{14}
$$

with the abbreviations

$$
\overline{\langle\mathbf{h}\rangle^2}:=E\left\{\langle\mathbf{h}\rangle^2\right\} \qquad \text{and} \qquad \overline{\langle x\rangle}:=E\left\{\overline{\langle x\rangle}\right\}. \tag{15}
$$

$\overline{\langle\mathbf{h}\rangle^2}$ can be further simplified. From (9) and (15) we have

$$
\begin{aligned}
\overline{\langle\mathbf{h}\rangle^2} &= E\left\{\left(\frac{1}{\mu}\sum_{m=1}^{\mu}\mathbf{h}_{m;\lambda}\right)^2\right\}=\frac{1}{\mu^2}E\left\{\sum_{m=1}^{\mu}\sum_{k=1}^{\mu}\left(\mathbf{h}_{k;\lambda}^T\,\mathbf{h}_{m;\lambda}\right)\right\} \\
&= \frac{1}{\mu^2}E\left\{\sum_{m=1}^{\mu}\left(\mathbf{h}_{m;\lambda}^T\,\mathbf{h}_{m;\lambda}\right)\right\}+\frac{1}{\mu^2}E\left\{\sum_{m=1}^{\mu}\sum_{k\neq m}\left(\mathbf{h}_{k;\lambda}^T\,\mathbf{h}_{m;\lambda}\right)\right\} \\
&= \frac{1}{\mu^2}\sum_{m=1}^{\mu}E\left\{\mathbf{h}_{m;\lambda}^2\right\}+\frac{1}{\mu^2}\sum_{m=1}^{\mu}\sum_{k\neq m}E\left\{\mathbf{h}_{k;\lambda}^T\,\mathbf{h}_{m;\lambda}\right\} \\
&= \frac{1}{\mu^2}\sum_{m=1}^{\mu}\overline{\mathbf{h}_{m;\lambda}^2}+\frac{1}{\mu^2}\sum_{m=1}^{\mu}\sum_{k\neq m}\overline{\mathbf{h}_{k;\lambda}^T\,\mathbf{h}_{m;\lambda}}\,.
\end{aligned}
\tag{16}
$$

It is interesting to note, that the expectation of the scalar product $\mathbf{h}_{k;\lambda}^T \mathbf{h}_{m;\lambda}$ vanishes

$$k \neq m \qquad \Longrightarrow \qquad E\left\{\mathbf{h}_{k;\lambda}^T \mathbf{h}_{m;\lambda}\right\} = 0. \tag{17}$$

The $\mathbf{h}_{k;\lambda}$-vectors are independent from each other. There is neutral selection normal to the $\mathbf{R}$ direction because of the symmetry of the spherical model. Each direction normal to $\mathbf{R}$ has equal likelihood. Thus, for fixed k and m

$$E\left\{\mathbf{h}_{k;\lambda}^T \mathbf{h}_{m;\lambda}\right\} = E\left\{\mathbf{h}_{k;\lambda}^T\right\} E\left\{\mathbf{h}_{m;\lambda}\right\} = 0$$

holds and from (16) only the first term

$$\overline{\langle\mathbf{h}\rangle^2} = \frac{1}{\mu^2}\sum_{m=1}^{\mu} \overline{\mathbf{h}_{m;\lambda}^2} =: \frac{1}{\mu}\langle\overline{\mathbf{h}^2}\rangle \tag{18}$$

remains. Therefore one finally obtains

$$\varphi_{\mu/\mu,\lambda} = R\left(1 - \sqrt{1 + \frac{\langle\overline{\mathbf{h}^2}\rangle}{\mu R^2}}\right) + \frac{\langle\overline{x}\rangle}{\sqrt{1 + \frac{\langle\overline{\mathbf{h}^2}\rangle}{\mu R^2}}}\,. \tag{19}$$

With (19) we have expressed the progress rate by the two expectancies $\langle\overline{\mathbf{h}^2}\rangle$ and $\langle\overline{x}\rangle$ to be determined in the next section.

3.2 Implementation of the Theory

3.2.1 Determination of $\langle\overline{x}\rangle$

As can be seen from (9) $\langle\overline{x}\rangle$ is given by the sum of the expected values of $x_{m;\lambda}$

$$\langle\overline{x}\rangle = \frac{1}{\mu}\sum_{m=1}^{\mu}\overline{x_{m;\lambda}} = \frac{1}{\mu}\sum_{m=1}^{\mu}\int_{-\infty}^{\infty} x p_{m;\lambda}(x)\,dx. \tag{20}$$

$p_{m;\lambda}(x)$ is the pdf of x belonging to the mth-best offspring. The ordering of the offspring obeys the ranking (4). The pdf $p(x)$ for a single trial x is given by

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}. \tag{21}$$

(x is a (rotated) component of the mutation vector $\mathbf{Z}$). Given a fixed x, the $\tilde{R}$ of the trial has the conditional pdf $p(r|x)$. Let this trial be the mth-best, then there are $(m-1)$ offspring with $(\tilde{R} \leq r)$ and $(\lambda - m)$ offspring with $(\tilde{R} > r)$, and there are $\lambda!/((m-1)!\,(\lambda-m)!)$ equivalent orderings. If integrated over the possible r-values one obtains the pdf $p_{m;\lambda}(x)$

$$p_{m;\lambda}(x) = p(x)\frac{\lambda!}{(m-1)!\,(\lambda-m)!}\int_{r=0}^{\infty} p(r|x)[P(\tilde{R} \leq r)]^{m-1}[P(\tilde{R} > r)]^{\lambda-m} dr. \tag{22}$$

$P(\tilde{R} \leq r) = P(r)$ has been already determined (see equation (11) in (Beyer, 1995))

$$P(r) = \frac{1}{2} + \Phi_0\left(\frac{r^2 - R^2 - \sigma^2 N}{2\sigma R\sqrt{1 + \frac{\sigma^2 N}{2R^2}}}\right), \tag{23}$$

with the Gauß-integral $\Phi_0(x) = \frac{1}{2}erf(x/\sqrt{2})$. It remains the determination of $p(r|x)$, which can be derived from the conditional cdf $P(\tilde{R} \le r|x) = P(\tilde{R}^2 \le r^2|x)$. From geometrical considerations provided in (Beyer, 1995, eq. (7)) one has

$$\tilde{R}^2 = \mathbf{h}^2 + R^2 - 2Rx \le r^2 \tag{24}$$

and therefore $\mathbf{h}^2 \le r^2 - R^2 + 2Rx$. The $\mathbf{h}^2$ variate obeys the pdf $p_{h^2}(v_2)$, (cp. eq. (10) in (Beyer, 1995))

$$p_{h^2}(v_2) = \frac{1}{\sqrt{2\pi}\,\sqrt{2N}\,\sigma^2} \exp -\frac{1}{2}\left(\frac{v_2 - N\sigma^2}{\sqrt{2N}\,\sigma^2}\right)^2, \tag{25}$$

thus we find

$$P(\tilde{R} \le r|x) = \int_{v_2=0}^{v_2=r^2-R^2+2Rx} p_{h^2}(v_2)\,dv_2 = \frac{1}{2} + \Phi_0\left(\frac{r^2 - R^2 - N\sigma^2 + 2Rx}{\sqrt{2N}\sigma^2}\right)$$

and

$$p(r|x) = \frac{d}{dr}P(\tilde{R} \le r|x) = \frac{2r}{\sqrt{2\pi}\,\sqrt{2N}\sigma^2} \exp -\frac{1}{2}\left(\frac{r^2 - R^2 - N\sigma^2 + 2Rx}{\sqrt{2N}\sigma^2}\right)^2. \tag{26}$$

Continuing with (20) one obtains after the exchange of the integration order

$$\langle \overline{x} \rangle = \frac{\lambda!}{\mu} \int_{r=0}^{\infty} \left(\int_{-\infty}^{\infty} x p(x) p(x|r)\,dx \right) \sum_{m=1}^{\mu} \frac{[P(r)]^{m-1}[1 - P(r)]^{\lambda-m}}{(m-1)!\,(\lambda-m)!}\,dr. \tag{27}$$

The expression in the parenthesis of (27) can be easily calculated. After the substitution $t = x/\sigma$ the $(\cdots)$-expression reads with (21) and (26)

$$(\cdots) = \frac{2r}{\sqrt{2\pi}\,\sqrt{2N}\sigma} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{1}{2}t^2} \exp -\frac{1}{2}\left(\frac{2R}{\sqrt{2N}\sigma}t + \frac{r^2 - R^2 - N\sigma^2}{\sqrt{2N}\sigma^2}\right)^2 dt.$$

This integral has been solved in (Beyer, 1994b)

$$(\cdots) = -\frac{1}{\sqrt{2\pi}} \frac{\sigma}{\sqrt{1 + \frac{N\sigma^2}{2R^2}}} \frac{2r}{2R\sigma\sqrt{1 + \frac{N\sigma^2}{2R^2}}} \frac{r^2 - R^2 - N\sigma^2}{2R\sigma\sqrt{1 + \frac{N\sigma^2}{2R^2}}} \exp -\frac{1}{2}\left(\frac{r^2 - R^2 - N\sigma^2}{2R\sigma\sqrt{1 + \frac{N\sigma^2}{2R^2}}}\right)^2. \tag{28}$$

By performing the substitution

$$y = \left(\frac{r^2 - R^2 - N\sigma^2}{2R\sigma\sqrt{1 + \frac{N\sigma^2}{2R^2}}}\right) \tag{29}$$

in (27) taking (28) and (23) into account one obtains[2]

$$\langle \overline{x} \rangle = -\frac{\sigma}{\sqrt{1 + \frac{N\sigma^2}{2R^2}}} \frac{1}{\sqrt{2\pi}} \frac{\lambda!}{\mu} \int_{y=-\infty}^{y=\infty} y e^{-\frac{1}{2}y^2} \sum_{m=1}^{\mu} \frac{\left(\frac{1}{2} + \Phi_0(y)\right)^{m-1} \left(\frac{1}{2} - \Phi_0(y)\right)^{\lambda-m}}{(m-1)!\,(\lambda-m)!}\,dy. \tag{30}$$

[2]The lower limit $y =: y_{lo}$ in (30) has been shifted to $y = -\infty$, because the upper bound $\hat{y}_{lo}$ of y_{lo} is $\hat{y}_{lo} = \mathtt{Max}\{y_l\} = -\sqrt{N/2}$ and the integrand is concentrated about $y = 0$.

If the sum in (30) is expressed by an integral (this technique has been demonstrated in (Beyer, 1995, section 3.1.2)) we get

$$\langle \overline{x} \rangle = -\frac{\sigma}{\sqrt{1+\frac{N\sigma^2}{2R^2}}} \frac{\lambda-\mu}{\sqrt{2\pi}} \binom{\lambda}{\mu} \int_{y=-\infty}^{y=\infty} \int_{x=0}^{\frac{1}{2}-\Phi_0(y)} y e^{-\frac{1}{2}y^2} x^{\lambda-\mu-1}(1-x)^{\mu-1}\, dx\, dy.$$

The change of the integration order and the substitution $x = \frac{1}{2} - \Phi_0(t)$ yields

$$\langle \overline{x} \rangle = \frac{-\sigma}{\sqrt{1+\frac{N\sigma^2}{2R^2}}} \frac{\lambda-\mu}{\sqrt{2\pi}} \binom{\lambda}{\mu} \int_{t=-\infty}^{t=\infty} \frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}} \left(\frac{1}{2}-\Phi_0(t)\right)^{\lambda-\mu-1} \left(\frac{1}{2}+\Phi_0(t)\right)^{\mu-1} \int_{y=-\infty}^{y=t} y e^{-\frac{1}{2}y^2}\, dy\, dt.$$

The y-integral gives $-\exp(-t^2/2)$ and after $t \to -t$ one obtains

$$\langle \overline{x} \rangle = \frac{\sigma}{\sqrt{1+\frac{N\sigma^2}{2R^2}}} e_{\mu,\lambda}^{1,0}\,, \tag{31}$$

with $e_{\mu,\lambda}^{1,0}$, a special case of the *generalized progress coefficients* introduced in (Beyer, 1995, eq. (82)). The integral representation of $e_{\mu,\lambda}^{1,0}$ can be found below, equation (40).

3.2.2 Determination of $\langle \overline{\mathbf{h}^2} \rangle$

$\langle \overline{\mathbf{h}^2} \rangle$ is defined by (18)

$$\langle \overline{\mathbf{h}^2} \rangle = \frac{1}{\mu} \sum_{m=1}^{\mu} \overline{\mathbf{h}^2_{m;\lambda}} = \frac{1}{\mu} \sum_{m=1}^{\mu} \int_{v_2=0}^{\infty} v_2 p_{m;\lambda}(v_2)\, dv_2. \tag{32}$$

with $v_2 := \mathbf{h}^2$ and $p_{m;\lambda}(v_2)$, the pdf of $\mathbf{h}^2$ for the mth-best trial. The pdf of $\mathbf{h}^2$ from a single trial is given by (25) $p(v_2) = p_{h^2}(v_2)$. By the same arguments used to derive (22) we obtain

$$p_{m;\lambda}(v_2) = p_{h^2}(v_2) \frac{\lambda!}{(m-1)!\,(\lambda-m)!} \int_{r=0}^{\infty} p(r|v_2)[P(r)]^{m-1}[1-P(r)]^{\lambda-m} dr. \tag{33}$$

$p(r|v_2) = p(r|\mathbf{h}^2) = \frac{d}{dr} P(\tilde{R} \le r|\mathbf{h}^2)$ can be derived from the conditional cdf $P(\tilde{R}^2 \le r^2|\mathbf{h}^2)$. From (24) we find $x \ge (\mathbf{h}^2 + R^2 - r^2)/2R$. The pdf of x is given by (21). Thus it is

$$P(\tilde{R} \le r|\mathbf{h}^2) = P(\tilde{R}^2 \le r^2|\mathbf{h}^2) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{x=\frac{\mathbf{h}^2+R^2-r^2}{2R}}^{\infty} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}} dx = \frac{1}{2} - \Phi_0\left(\frac{\mathbf{h}^2+R^2-r^2}{2R\sigma}\right)$$

and

$$p(r|v_2) = \frac{r}{\sqrt{2\pi}\,R\sigma} \exp -\frac{1}{2}\left(\frac{v_2+R^2-r^2}{2R\sigma}\right). \tag{34}$$

With (33), (34), and (25) one obtains for (32) analogously to (27)

$$\langle \overline{\mathbf{h}^2} \rangle = \frac{\lambda!}{\mu} \int_{r=0}^{\infty} \left(\int_{v_2=0}^{\infty} v_2 p_{h^2}(v_2) p(r|v_2)\, dv_2 \right) \sum_{m=1}^{\mu} \frac{[P(r)]^{m-1}[1-P(r)]^{\lambda-m}}{(m-1)!\,(\lambda-m)!}\, dr. \tag{35}$$

and the expression in the parenthesis reads

$$(\cdots) = \frac{r}{\sqrt{2\pi}\,R\sigma} \frac{1}{\sqrt{2\pi}\sqrt{2N\sigma^2}} \int_{v_2=0}^{\infty} v_2 \exp -\frac{1}{2}\left(\frac{v_2-\sigma^2 N}{\sqrt{2N\sigma^2}}\right)^2 \exp -\frac{1}{2}\left(\frac{v_2+R^2-r^2}{2R\sigma}\right)^2 dv_2.$$

From a straightforward calculation (Beyer, 1994b) it results

$$(\cdots) = -\frac{\sigma^2 N}{\sqrt{2\pi}} \frac{r}{\sigma R\sqrt{1+\frac{N\sigma^2}{2R^2}}} \left(1 - \frac{\sigma}{R\sqrt{1+\frac{N\sigma^2}{2R^2}}} \frac{R^2+\sigma^2 N - r^2}{2R\sigma\sqrt{1+\frac{N\sigma^2}{2R^2}}}\right) \exp -\frac{1}{2}\left(\frac{R^2+\sigma^2 N - r^2}{2R\sigma\sqrt{1+\frac{N\sigma^2}{2R^2}}}\right)^2 .$$

If inserted into (35) taking (23) into account one obtains by the substitution (29)

$$\langle\overline{\mathbf{h}^2}\rangle = \frac{\sigma^2 N}{\sqrt{2\pi}} \frac{\lambda!}{\mu} \int_{y=-\infty}^{\infty} \left(1 + \frac{\sigma}{R\sqrt{1+\frac{N\sigma^2}{2R^2}}}\, y\right) e^{-\frac{1}{2}y^2} \sum_{m=1}^{\mu} \frac{\left(\frac{1}{2}+\Phi_0(y)\right)^{m-1}\left(\frac{1}{2}-\Phi_0(y)\right)^{\lambda-m}}{(m-1)!\,(\lambda-m)!}\, dy.$$

The integration of the first part (i. e. the integrand belonging to the '1' in the parenthesis) gives $\sigma^2 N$. The second part is similar to (30). Hence, we find

$$\langle\overline{\mathbf{h}^2}\rangle = \sigma^2 N \left[1 - \frac{\sigma}{R\sqrt{1+\frac{N\sigma^2}{2R^2}}}\, e^{1,0}_{\mu,\lambda}\right]. \tag{36}$$

3.2.3 Assembly of the Progress Rate Formula

As usual, we introduce normalized quantities $\varphi^\star := \varphi N/R$ and $\sigma^\star := \sigma N/R$. The square root expression in (19) can be written with (36) as

$$\begin{aligned}
\sqrt{1+\frac{\langle\overline{\mathbf{h}^2}\rangle}{\mu R^2}} &= \sqrt{1+\frac{(\sigma^\star)^2}{\mu N} - \frac{(\sigma^\star)^2}{\mu N}\frac{\sigma^\star}{N\sqrt{1+\frac{(\sigma^\star)^2}{2N}}}\, e^{1,0}_{\mu,\lambda}} \\
&= \sqrt{1+\frac{(\sigma^\star)^2}{\mu N}}\sqrt{1-\frac{(\sigma^\star)^2}{\mu N\left(1+\frac{(\sigma^\star)^2}{\mu N}\right)}\frac{\sigma^\star}{N\sqrt{1+\frac{(\sigma^\star)^2}{2N}}}\, e^{1,0}_{\mu,\lambda}} \\
&\approx \sqrt{1+\frac{(\sigma^\star)^2}{\mu N}} - \frac{1}{2}\frac{(\sigma^\star)^3\, e^{1,0}_{\mu,\lambda}}{\mu N^2\sqrt{1+\frac{(\sigma^\star)^2}{\mu N}}\sqrt{1+\frac{(\sigma^\star)^2}{2N}}} .
\end{aligned}$$

The last line has been obtained by Taylor expansion $\sqrt{1-a} \approx 1 - a/2$. With the same order of approximation $1/\sqrt{1-a} \approx 1 + a/2$ one gets

$$\left(\sqrt{1+\frac{\langle\overline{\mathbf{h}^2}\rangle}{\mu R^2}}\right)^{-1} \approx \left(\sqrt{1+\frac{(\sigma^\star)^2}{\mu N}}\right)^{-1} + \frac{1}{2}\frac{(\sigma^\star)^3\, e^{1,0}_{\mu,\lambda}}{\mu N^2\sqrt{1+\frac{(\sigma^\star)^2}{\mu N}}\sqrt{1+\frac{(\sigma^\star)^2}{2N}}}$$

and with (19/31) the normalized progress rate $\varphi^\star$ reads

$$\begin{aligned}
\varphi^\star_{\mu/\mu,\lambda} = \quad & N\left(1-\sqrt{1+\frac{(\sigma^\star)^2}{\mu N}}\right) + \frac{\sigma^\star\, e^{1,0}_{\mu,\lambda}}{\sqrt{1+\frac{(\sigma^\star)^2}{\mu N}}\sqrt{1+\frac{(\sigma^\star)^2}{2N}}} \\
& + \frac{\sigma^\star\, e^{1,0}_{\mu,\lambda}}{\sqrt{1+\frac{(\sigma^\star)^2}{\mu N}}\sqrt{1+\frac{(\sigma^\star)^2}{2N}}}\frac{(\sigma^\star)^2}{2\mu N}\left(1+\frac{1}{N}\frac{\sigma^\star\, e^{1,0}_{\mu,\lambda}}{\sqrt{1+\frac{(\sigma^\star)^2}{2N}}}\right).
\end{aligned} \tag{37}$$

If one neglects the last term in (37), then we finally arrive at the *progress rate formula for the $(\mu/\mu_I, \lambda)$-ES*

$$\boxed{\varphi^\star_{\mu/\mu,\lambda}(\sigma^\star) \;=\; N\left(1-\sqrt{1+\frac{(\sigma^\star)^2}{\mu N}}\right) + \sigma^\star\, c_{\mu/\mu,\lambda}\,\frac{1+\frac{(\sigma^\star)^2}{2\mu N}}{\sqrt{1+\frac{(\sigma^\star)^2}{2N}}\sqrt{1+\frac{(\sigma^\star)^2}{\mu N}}}. \qquad (38)}$$

We have introduced the progress coefficient $c_{\mu/\mu,\lambda}$

$$c_{\mu/\mu,\lambda} := e^{1,0}_{\mu,\lambda} \qquad (39)$$

in order to have a 'notational continuity'.[3] $c_{\mu/\mu,\lambda}$ can be easily computed by numerical integration of (40), below. Table 1 gives a collection of some $c_{\mu/\mu,\lambda}$-values. In the next section we will

$\mu \setminus \lambda$	10	20	30	40	50	100	150	200	300
1	1.539	1.867	2.043	2.161	2.249	2.508	2.649	2.746	2.878
2	1.270	1.638	1.829	1.957	2.052	2.328	2.478	2.580	2.718
3	1.065	1.469	1.674	1.810	1.911	2.201	2.357	2.463	2.607
4	0.893	1.332	1.550	1.694	1.799	2.101	2.263	2.372	2.520
5	0.739	1.214	1.446	1.596	1.705	2.018	2.185	2.297	2.449
10	0.000	0.768	1.061	1.242	1.372	1.730	1.916	2.040	2.206
20	-	0.000	0.530	0.782	0.950	1.386	1.601	1.742	1.928
30	-	-	0.000	0.414	0.634	1.149	1.390	1.545	1.746
40	-	-	-	0.000	0.343	0.958	1.225	1.393	1.608
50	-	-	-	-	0.000	0.792	1.085	1.265	1.494
100	-	-	-	-	-	0.000	0.542	0.795	1.088

Table 1: Selected $c_{\mu/\mu,\lambda}$-coefficients.

derive asymptotic formulas for the $c_{\mu/\mu,\lambda}$-coefficients.

3.2.4 Asymptotics of the $c_{\mu/\mu,\lambda}$-Coefficients

For large λ values and moderate $\kappa := \mu/\lambda$ truncation ratios it is possible to derive asymptotic formulas for $c_{\mu/\mu,\lambda}$. The progress coefficient integral for $c_{\mu/\mu,\lambda}$ is given by

$$c_{\mu/\mu,\lambda} = e^{1,0}_{\mu,\lambda} = \frac{\lambda-\mu}{2\pi}\binom{\lambda}{\mu}\int_{-\infty}^{\infty} e^{-t^2}\left(\frac{1}{2}+\Phi_0(t)\right)^{\lambda-\mu-1}\left(\frac{1}{2}-\Phi_0(t)\right)^{\mu-1} dt. \qquad (40)$$

(NB, the generalized progress coefficients $e^{\alpha,\beta}_{\mu,\lambda}$ have been defined in (Beyer, 1995).)
If the substitution $z = \frac{1}{2} - \Phi_0(t)$ is applied, then one obtains

$$c_{\mu/\mu,\lambda} = \frac{\lambda}{\mu}\,\frac{1}{\sqrt{2\pi}}\,\frac{1}{B(\lambda-\mu,\mu)}\int_0^1 \exp -\frac{1}{2}\left[\Phi_0^{-1}\left(\frac{1}{2}-z\right)\right]^2 z^{\mu-1}(1-z)^{\lambda-\mu-1}\,dz. \qquad (41)$$

The quotient $(\lambda-1)!/((\lambda-\mu-1)!\,(\mu-1)!)$ has been expressed by the (complete) *beta function* $B(\lambda-\mu,\mu)$, which has the integral representation (Abramowitz & Stegun, 1984, p. 79)

$$B(\lambda-\mu,\mu) = \frac{(\lambda-\mu-1)!\,(\mu-1)!}{(\lambda-1)!} = \frac{\Gamma(\lambda-\mu)\,\Gamma(\mu)}{\Gamma(\lambda)} = \int_0^1 z^{\mu-1}(1-z)^{\lambda-\mu-1}\,dz. \qquad (42)$$

[3] The $c_{\mu/\mu,\lambda}$ notation was originally coined by (Rechenberg, 1993).

For large values of μ and $\lambda-\mu$ the function $z^{\mu-1}(1-z)^{\lambda-\mu-1}$ forms a sharp peak. Its maximum may be at $z=\hat{z}$

$$\texttt{Max}\left\{z^{\mu-1}(1-z)^{\lambda-\mu-1}\right\} \quad \Leftrightarrow \quad z=\hat{z} \quad \Leftrightarrow \quad \hat{z}=\frac{\mu-1}{\lambda-2}. \tag{43}$$

Expanding

$$f(z) := \exp -\frac{1}{2}\left[\Phi_0^{-1}\left(\frac{1}{2}-z\right)\right]^2$$

at $z=\hat{z}$ in a Taylor series and substitution into (41) yields

$$c_{\mu/\mu,\lambda} = \frac{1}{\kappa}\frac{1}{\sqrt{2\pi}}\sum_{m=0}^{\infty}\frac{1}{m!}\left.\frac{d^m f(z)}{dz^m}\right|_{z=\hat{z}} \frac{1}{B(\lambda-\mu,\mu)}\int_0^1 (z-\hat{z})^m z^{\mu-1}(1-z)^{\lambda-\mu-1}\,dz\,. \tag{44}$$

The part after the derivative in (44) is an integral representation of the *hypergeometric function* $F(a,b;c;z)$ (Abramowitz & Stegun, 1984, p. 215). Therefore, (44) can be written as

$$c_{\mu/\mu,\lambda} = \frac{1}{\kappa}\frac{1}{\sqrt{2\pi}}\sum_{m=0}^{\infty}\frac{(-\hat{z})^m}{m!}\left.\frac{d^m f(z)}{dz^m}\right|_{z=\hat{z}} F\left(-m,\mu;\lambda;\frac{1}{\hat{z}}\right), \qquad \kappa=\frac{\mu}{\lambda}. \tag{45}$$

Below, we derive only the first two terms of the expansion (45).
For $m=0$ one finds with (42/44) $F\left(0,\mu;\lambda;\frac{1}{\hat{z}}\right)=1$. $m=1$ needs further considerations.
First, $\frac{df(z)}{dz}$ is determined. The derivative of $\Phi_0^{-1}(y)$ can be obtained by differentiation of the identity $\Phi_0^{-1}(y(t))=t,\ y(t)=\Phi_0(t)$. One finds $\frac{df}{dz}=\sqrt{2\pi}\,\Phi_0^{-1}\left(\frac{1}{2}-z\right)$ and therefore

$$\left.\frac{df}{dz}\right|_{z=\hat{z}} = \sqrt{2\pi}\Phi_0^{-1}\left(\frac{1}{2}-\frac{\mu-1}{\lambda-2}\right). \tag{46}$$

The integral part of (44) can be tackled by usage of (42)

$$-\hat{z}\,F\left(-1,\mu;\lambda;\frac{1}{\hat{z}}\right) = \frac{1}{B(\lambda-\mu,\mu)}\int_0^1 (z-\hat{z})\,z^{\mu-1}(1-z)^{\lambda-\mu-1}\,dz = \frac{\mu}{\lambda}-\hat{z}. \tag{47}$$

Collecting the parts (43/45/46/47) gives

$$\begin{aligned} c_{\mu/\mu,\lambda} \quad = \quad & \frac{1}{\kappa}\frac{1}{\sqrt{2\pi}}\left\{\exp -\frac{1}{2}\left[\Phi_0^{-1}\left(\frac{1}{2}-\kappa\left(\frac{1-1/\mu}{1-2/\lambda}\right)\right)\right]^2\right. \\ & \left. + \frac{\sqrt{2\pi}}{\lambda}\left(\frac{1-2\kappa}{1-2/\lambda}\right)\Phi_0^{-1}\left(\frac{1}{2}-\kappa\left(\frac{1-1/\mu}{1-2/\lambda}\right)\right) + \ldots\right\}. \end{aligned} \tag{48}$$

For large λ values ($\lambda \gtrapprox 200$), and moderate $\kappa=\mu/\lambda$ ($0.03 \lessapprox \kappa \lessapprox 0.97$) formula (48) may be approximately used instead of (40).[4] If the limit $\lambda\to\infty$ is taken in (48) one obtains the asymptotically exact formula

$$c_{\mu/\mu,\lambda} = \frac{\lambda}{\mu}\frac{1}{\sqrt{2\pi}}\exp -\frac{1}{2}\left[\Phi_0^{-1}\left(\frac{1}{2}-\frac{\mu}{\lambda}\right)\right]^2, \qquad 0<\frac{\mu}{\lambda}<1.$$

This approximation yields satisfactory results for $\lambda \gtrapprox 1000$.

[4] Instead of $\Phi_0^{-1}(y)$ the inverse error function erf^{-1} may be used: $\Phi_0^{-1}(y)=\sqrt{2}\,erf^{-1}(2y)$.

3.3 Results and Conclusions from the $(\mu/\mu_I, \lambda)$-Theory

3.3.1 Comparison with Experiments

The progress rate formula has been tested for different λ, μ, N - combinations. As an example results from the $(8/8_I, 30)$-ES for $N = 30$ and $N = 200$ are shown. The experimental technique used is similar to the $(1, \lambda)$-experiments described in (Beyer, 1995, section 2.3). G 'one-generation' experiments are performed from a fixed parent $\mathbf{y}$ (with distance R to the optimum) with fixed σ-mutation strength. 'One-generation' - according to section 2, eq. (1) - comprises the generation of λ offspring $\tilde{\mathbf{y}}_l$ by mutation from $\mathbf{y}$, selection of the best μ offspring $\tilde{\mathbf{y}}_{m;\lambda}$, and averaging these offspring to obtain $\|\langle\tilde{\mathbf{y}}\rangle\|$. The rest is analogous to the $(1, \lambda)$-experiments (op. cit.).

Fig. 1 depicts the results gathered from $G = 10,000$ 'one-generation' experiments per $\sigma^\star$-value and represented by dots. The curves are graphs from (38) with $c_{8/8,30} = 1.196$. The

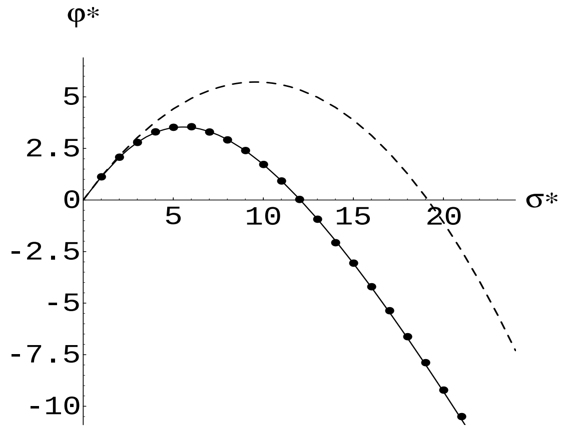


Figure 1: Comparison of theory and experiment for the $(8/8_I, 30)$ intermediate ES. The dashed curves are due to Rechenberg's formula. In the left picture the case $N = 30$ is displayed and in the right one $N = 200$.

dashed curves are due to Rechenberg's 'universal law' (Rechenberg, 1993)

$$\varphi^\star_{\mu/\mu,\lambda} = \sigma^\star c_{\mu/\mu,\lambda} - \frac{(\sigma^\star)^2}{2\mu} \, .$$

One can see that this formula holds mainly for small $\sigma^\star$ values: Like the case of the $(1, \lambda)$ theory the Rechenberg formula can be recovered by expanding the roots in (38) for small $(\sigma^\star)^2/N$-values.[5]

3.3.2 On the Benefits of Sexuality

The advantages of $(\mu/\mu, \lambda)$ strategies is their - at the first glance - surprisingly high progress rate in comparison with the simple (μ, λ)-ES. Suppose the strategies are tuned for maximum

[5]In his new book (Rechenberg, 1994, pp. 22, 23) Rechenberg emphasizes that his theory is an asymptotic one that holds for large N and small σ-values.

performance, e. g. by a σ-self-adaptation mechanism developed by SCHWEFEL (Schwefel, 1995). If we assume that such a mechanism provides (at least) locally the maximum progress rate $\hat{\varphi}^\star$ for fixed strategy parameters μ, λ, and N

$$\hat{\varphi}^\star(\mu, \lambda, N) = \mathtt{Max}_{[\sigma^\star]} \left\{ \varphi^\star(\sigma^\star, \mu, \lambda, N) \right\}, \tag{49}$$

then we can compare the (μ, λ) and the $(\mu/\mu, \lambda)$ ES. As an example the $\lambda = 50$ strategies for $N = 100$ are depicted in Fig. 2. The lower curve was obtained from (49) by maximization of the progress rate formula for $\varphi^\star_{\mu,\lambda}$ derived in (Beyer, 1995, eq. (98)) treating μ as a real parameter. The upper curve was computed by the same procedure applied to (38). As can be seen the

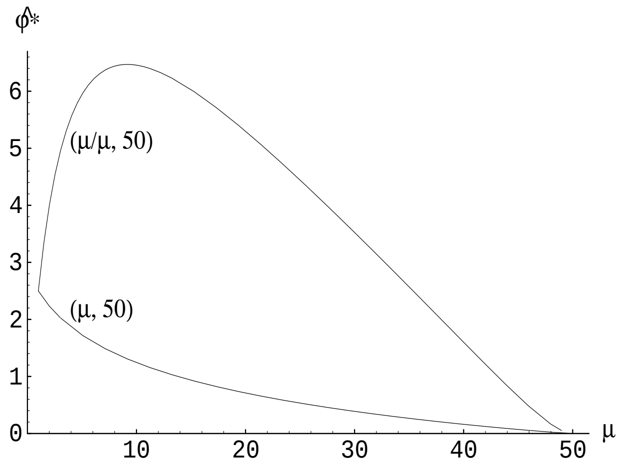


Figure 2: The maximum progress rate $\hat{\varphi}^\star$ versus the number of parents μ producing $\lambda = 50$ offspring. The parameter space dimension is $N = 100$. Note, the $(\mu/\rho_I, 50)$ strategies are expected to lie between theses two extremal curves.

$(\mu/\mu, \lambda)$ strategies are superior to the (μ, λ) strategies. What is the reason for this behavior?

In order to see the main difference between $\varphi^\star_{\mu,\lambda}$ and $\varphi^\star_{\mu/\mu,\lambda}$ we compare the simplified $\varphi^\star_{\mu,\lambda}$-formula (cp. Beyer, 1995, eq. (101))

$$\varphi^\star_{\mu,\lambda}(\sigma^\star) = N \left(1 - \sqrt{1 + \frac{(\sigma^\star)^2}{N}} \right) \; + \; c_{\mu,\lambda}\, \sigma^\star \sqrt{\frac{1 + \frac{(\sigma^\star)^2}{2N}}{1 + \frac{(\sigma^\star)^2}{N}}} \tag{50}$$

with (38). It is catching the eye that the $(\sigma^\star)^2/N$ of the first term in (50) is replaced by $(\sigma^\star)^2/\mu N$ in (38). If traced back to (18/19) one recognizes that the $1/\mu$ factor is due to the averaging of the $\mathbf{h}$ vectors of the best μ mutations. Note, we had decomposed the mutations $\mathbf{Z}$ (according to (7)) in a component x in optimum direction and in a perpendicular $\mathbf{h}$. By the intermediate recombination, i. e. the averaging of these $\mathbf{h}$ vectors, the length of the resulting $\langle \mathbf{h} \rangle$ is by an approximate factor of $1/\sqrt{\mu}$ smaller than the expected length of a single $\mathbf{h}$. This statement that holds exactly for the square of $\mathbf{h}$ was obtained from (16/17). Since the $\mathbf{h}$ vectors are not correlated (random mutations, and neutral selection normal to $\mathbf{R}$) the cross-correlation part in (16) vanishes. If one interprets the $\mathbf{h}$ components as the 'harmful part' of the mutation $\mathbf{Z}$ (because they increase the distance to the optimum), then a recombination producing $\langle \mathbf{h} \rangle$ decreases this part. The author will call this mechanism 'genetic repair'.

As far as intermediate recombination is concerned the statement often invoked by EA-users explaining the benefit of sexuality that recombination 'combines good properties' of the mates to form better offspring is wrong: In our model the 'good properties' are simply the $x_{m;\lambda}$. They are averaged according to (9). It is quite obvious that $\langle x \rangle$ cannot be larger than the x-value of the best offspring. The performance gain of the $(\mu/\mu, \lambda)$-ES has another reason: If one investigates the $\hat{\sigma}^\star$-values that maximizes $\varphi^\star$ (cp. e. g. Fig. 1), then it becomes clear that $(\mu/\mu, \lambda)$ strategies exhibit optimal performance for larger mutation strengths than (μ, λ) strategies. Larger mutations produce larger $x_{m;\lambda}$ components and $\mathbf{h}$ vectors as well. However, due to the *genetic repair* the 'harmful' influence of the $\mathbf{h}$ vectors is diminished. This reminds the author of a saying by SCHWEFEL (Schwefel, 1987):

> *Even if all of them were wrong they might as well be worthwhile to be recombined with each other.*

From (9/16) it becomes clear that the *genetic repair* exhibits the best performance, if all μ parents contribute to the recombination, i. e. $\rho = \mu$. The averaging of ρ parents with $\rho < \mu$ yields worse results. The case $\rho = 1$ is equivalent to the (μ, λ)-ES. Hence, the performance of all the other $(\mu/\rho_I, \lambda)$ strategies should lie between the two extremes (μ, λ) and $(\mu/\mu_I, \lambda)$.

Note, *genetic repair* relies on differing recombination mates, otherwise (17) does not hold. It is worth mentioning that in zoology a phenomenon is observed that may be called 'incest taboo'. In the higher animals, the begetting of offspring by close relatives is relatively seldom observed. And inbreeding often produces hereditary defects.

Up until now, the reasons for the sexual recombination are subject to controversial discussions in biology. Some scientists emphasize on the recombination as a source of genetic variation (for a discussion of the current hypotheses, cp. (Michod & Levin, 1988)), and we will see in section 4 that there is a certain counterpart in dominant $(\mu/\rho_D, \lambda)$ strategies. But this is only half the story. Genetic repair as pointed out by e. g. (Bernstein, Hopf & Michod, 1987) should also play an important role in the understanding of the universality of recombination. From the investigation of the $(\mu/\rho, \lambda)$ strategies the author makes a speculation somewhat different from the established ones (Michod & Levin, 1988):

In the author's opinion the benefit of sex mainly stems from the higher 'evolution speed' made possible by higher mutation rates in conjunction with genetic repair diminishing the influence of the 'harmful' part of the mutations. This does not necessarily imply that the physical mutations rate of a chromosome, genome, or species is always at a constant, high level (it may change depending on the environmental conditions). However, the results of the dominant recombination may be interpreted as large mutations generated from an average individual - the *wild-type*. Thus, it will be possible to transform a relatively small physical mutation rate into a larger mutation rate on the level of genetic information (to be discussed in section 4).

3.3.3 The Optimum μ-Choice

Unlike parents in real life asking on how many children they can afford, we have to determine the number of parents μ given a fixed number of offspring. This is because of the fact that producing offspring and calculating their fitness requires computer power which is restricted by the number of processors available. Usually the number of offspring is a multiple of the number of processors. Therefore, λ is more or less fixed and the question how to choose the number of parents μ is quite natural.

From the viewpoint of genetic repair the number of parents should be large, e. g. $\mu = \rho = \lambda$, yielding the minimum length $\langle \mathbf{h} \rangle$ vector. However, this would produce small or vanishing $\langle x \rangle$-

values ($\langle x \rangle = 0 \;\Leftrightarrow\; \mu = \lambda$). Therefore, there is bound to be a trade-off between $\langle x \rangle$ (should be large) and $\langle \mathbf{h} \rangle$ (should be small). There will be an optimum μ-choice guaranteeing a maximum progress rate. This is what Fig. 2 showed for the $(\mu/\mu_I, 50)$-ES.

Given fixed parameter space dimension N and λ, eqn. (38/40) can be used to determine the optimum $\hat{\varphi}^\star$, $\hat{\sigma}^\star$, and $\hat{\mu}$ values. This has been done numerically by maximization of $\varphi^\star = \varphi^\star(\sigma^\star, \mu, \lambda, N)$ with respect to $\sigma^\star$ and μ

$$\hat{\varphi}^\star(\lambda, N) = \mathtt{Max}_{[\sigma^\star,\mu]} \left\{\varphi^\star(\sigma^\star, \mu, \lambda, N)\right\} \quad \Leftrightarrow \quad \sigma^\star = \hat{\sigma}^\star, \quad \mu = \hat{\mu}. \tag{51}$$

The results are depicted in Fig. 3 and in the next figure.

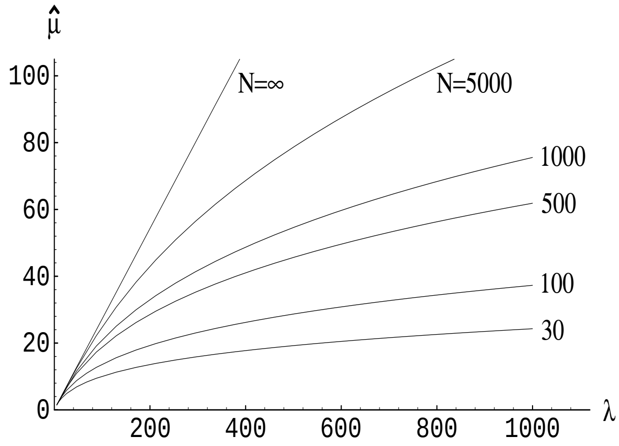


Figure 3: The optimum μ-choice versus the offspring number λ for different parameter space dimensions.

Fig. 3 shows the optimal number of parents $\hat{\mu}$ as a function of λ in the range $\lambda = 5 \ldots 1000$. It is important to note, that there is a relatively strong dependence on N. The curves for $N < \infty$ are sublinear, whereas the limit case $N \to \infty$ provides a linear one with inclination ≈ 0.27. I. e., for $N \to \infty$ the optimal number of parents should be $\hat{\mu} = 0.27\,\lambda$. For $N < \infty$ the inequality $\hat{\mu} < 0.27\,\lambda$ holds. For practical applications Tab. 2 gives a collection of optimal parent numbers.

Fig. 4 displays the $\varphi^\star$ and $\sigma^\star$ values belonging to the optimum μ-choice. It is striking that for $N \to \infty$ the optimum $(\mu/\mu_I, \lambda)$ strategies exhibit an almost *linear speed-up* of the progress rate as to the number of offspring.[6] However, for 'normally sized' problems the speed-up scales sooner or later sub-linear, but will be still on a remarkable level. Therefore, *as far as 'fast hill climbing' is desired, the $(\hat{\mu}/\hat{\mu}_I, \lambda)$-ES will be the fastest one.*

4 A Rough Estimate for the Dominant $(\mu/\mu_D, \lambda)$-ES

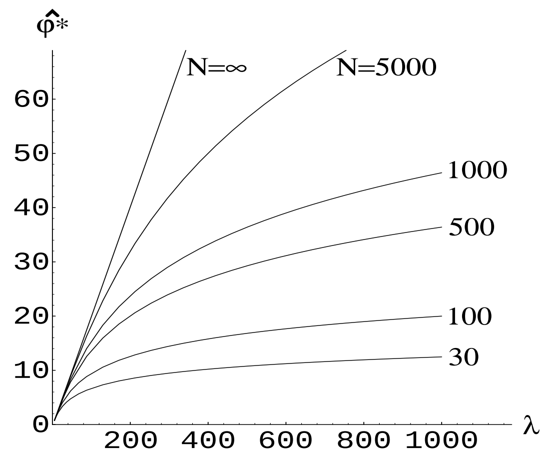
4.1 A Simple Model for the $(\mu/\mu_D, \lambda)$ Theory

The dominant recombinant $(\mu/\mu_D, \lambda)$ strategy cannot be analyzed as easily as the intermediate one. This is so because the mutations $\mathbf{Z}$ do not act on a single intermediate parent $\langle \mathbf{y} \rangle$ but on a parent distribution (cp. eq. (2)) just like the (μ, λ)-ES. Hence, one has to determine parental

[6] Actually, a slightly *super-linear speed-up* for $N \to \infty$ is forecasted (cp. Tab. 2).

	$\hat{\mu}$				$\hat{\sigma}^\star$				$\hat{\varphi}^\star$			
$\lambda \setminus N$	30	100	1000	∞	30	100	1000	∞	30	100	1000	∞
10	3	3	3	3	2.788	3.035	3.177	3.196	1.529	1.638	1.695	1.703
20	4	5	5	6	4.016	5.172	5.936	6.663	2.805	3.261	3.629	3.699
30	5	7	8	8	4.826	6.763	9.055	9.568	3.692	4.545	5.503	5.721
40	6	8	10	11	5.451	7.697	11.38	13.05	4.370	5.597	7.246	7.746
50	7	9	12	14	5.966	8.468	13.49	16.54	4.916	6.469	8.865	9.766
60	8	10	14	16	6.407	9.131	15.41	19.42	5.369	7.214	10.37	11.79
70	8	11	16	19	6.562	9.717	17.16	22.91	5.754	7.864	11.76	13.81
80	9	12	18	22	6.928	10.24	18.77	26.40	6.095	8.440	13.06	15.84
90	10	13	19	25	7.258	10.72	19.86	29.88	6.394	8.955	14.26	17.86
100	10	14	21	27	7.359	11.17	21.26	32.77	6.665	9.420	15.39	19.89
150	12	17	28	41	8.126	12.57	26.09	49.60	7.706	11.25	20.10	30.01
200	14	19	33	54	8.731	13.46	29.29	65.83	8.444	12.58	23.75	40.13
250	15	21	38	68	9.084	14.21	31.96	82.67	9.016	13.61	26.71	50.25
300	16	23	42	81	9.391	14.86	34.01	98.90	9.481	14.46	29.19	60.37
350	17	25	45	95	9.666	15.45	35.59	115.7	9.872	15.17	31.33	70.50
400	18	26	49	108	9.917	15.81	37.25	132.0	10.21	15.79	33.21	80.62
450	19	27	52	122	10.15	16.15	38.55	148.8	10.51	16.33	34.88	90.74
500	19	29	55	135	10.23	16.62	39.75	165.0	10.77	16.82	36.39	100.9
600	21	31	60	162	10.63	17.18	41.71	198.1	11.22	17.66	39.01	121.1
700	22	33	64	189	10.88	17.69	43.26	231.2	11.61	18.37	41.24	141.4
800	23	34	68	216	11.10	18.01	44.68	264.2	11.93	18.98	43.18	161.6
900	23	36	72	243	11.18	18.44	45.98	297.3	12.22	19.52	44.90	181.8
1000	24	37	76	270	11.37	18.71	47.19	330.3	12.48	19.99	46.43	202.1
1500	28	43	90	406	12.10	19.99	51.35	496.3	13.45	21.81	52.33	303.3
2000	30	47	101	541	12.50	20.81	54.26	661.6	14.13	23.09	56.50	404.5

Table 2: The optimal μ-choice in $(\mu/\mu_I, \lambda)$-recombinant Evolution Strategies. $\hat{\varphi}^\star$ is the theoretically achievable progress rate, if the strategy works at the optimum mutation strength $\hat{\sigma}^\star$. The optimum $\hat{\mu}$ depends on the parameter space dimension N. The cases $N = 30,\ 100,\ 1000$, and ∞ are displayed.



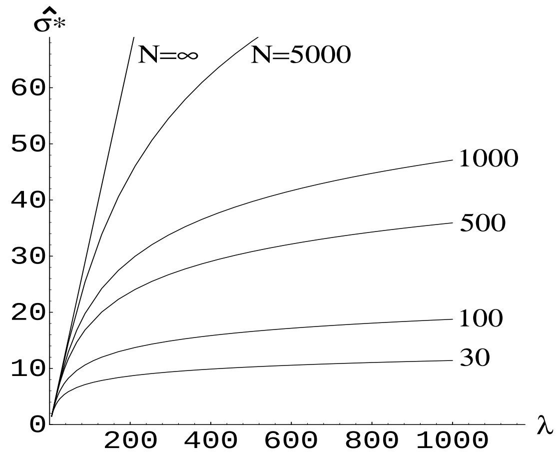


Figure 4: Left picture: The maximum progress rate $\hat{\varphi}^\star$ reachable by a strategy with optimally chosen $\mu = \hat{\mu}$ and optimally adapted mutation strength $\sigma = \hat{\sigma}$ (displayed in the right picture). Right picture: The optimum (normalized) standard deviation $\sigma^\star$ of the mutations versus the number of offspring for optimally chosen number of parents.

distributions by the self-consistent method. The full analysis remains as a task for the future. However, by a simple model presented below, it will be possible to give a first estimate for the progress rate φ. Similar results - without derivation - can be found in (Rechenberg, 1994).

The definition of the progress rate can be taken from the intermediate theory (3 – 6). φ measures the expected distance change of the parental center of mass from generation (g) to $(g+1)$

$$\varphi_{\mu/\mu_D,\lambda} := E\left\{R - \tilde{R}\right\} = E\left\{\langle \mathbf{y}\rangle^{(g)} - \langle \mathbf{y}\rangle^{(g+1)}\right\}, \qquad \langle \mathbf{y}\rangle = \frac{1}{\mu}\sum_{m=1}^{\mu} \tilde{\mathbf{y}}_{m;\lambda}.$$

Given the parental distribution at (g) the process recombination-mutation-selection generates the new parental distribution at $(g+1)$. From the mathematical point of view, there is no reason to separate the processes recombination/mutation. They are both random processes without bias or correlation. Indeed, one may interpret the resulting distribution as if it were generated by a special kind of mutations starting from the parental average state $\langle \mathbf{y}\rangle^{(g)}$.[7] We may call this newly defined random (mutation) vector starting from $\langle \mathbf{y}\rangle^{(g)}$ the surrogate mutation $\mathbf{S}$ of the recombination-mutation process. If $p^{(g)}(\mathbf{S})$, i. e. the N-dimensional pdf of $\mathbf{S}$, would be known, then one could proceed as in the case of the intermediate theory: $\mathbf{S}_{m;\lambda}$ can be decomposed according to (7) and the progress rate expression (10) exactly holds. Even the next steps from 3.1.2 are applicable and the formula (19) is valid.

All the difficulties in the $(\mu/\mu_D, \lambda)$ theory are concentrated in the self-consistent estimation of $p^{(g)}(\mathbf{S})$. A first - very rough - approximation to $p^{(g)}(\mathbf{S})$ may be obtained by neglecting the influence of selection on $p^{(g)}(\mathbf{S})$. A system evolving by a mutation-recombination process, i. e. driven by symmetrical mutations normally distributed and dominant random recombination,

[7]This is exactly what has been done for one dimension - the r-direction - in the case of the (μ, λ)-ES (cp. Beyer, 1995, section 3.1.3).

generates an *isotropic* normally distributed $p^{(g)}(\mathbf{S})$ with iid components

$$p(s_i) = \frac{1}{\sqrt{2\pi}\,\sigma_s} e^{-\frac{1}{2}\left(\frac{s_i}{\sigma_s}\right)^2} \tag{52}$$

measured from the center of mass $\langle \mathbf{y} \rangle^{(g)}$. The only unknown in this process is the mutation strength σ_s of the *isotropic surrogate mutations* $\mathbf{S}$. We will tackle this problem in the next section. If σ_s is known, one can apply the $(\mu/\mu_I, \lambda)$-standard theory to obtain the progress rate formula: One only has to exchange the mutation strength σ producing $\langle \overline{\mathbf{h}^2} \rangle$ and $\langle \overline{x} \rangle$ in (19) by $\sigma_s = \sigma_s(\mu, \sigma)$.

4.2 Isotropic Surrogate Mutations, Progress Rate, and is there Genetic Repair in Dominant Recombinant Strategies?

4.2.1 Isotropic Surrogate Mutations

Because dominant recombination acts on the level of independent coordinates the consideration for one component $s = s_i$ suffices. For sake of simplicity we will describe the recombination-mutation process starting from a fixed point

$$\langle s \rangle^{(0)} = s_1^{(0)} = s_2^{(0)} = \ldots = s_m^{(0)} = \ldots = s_\mu^{(0)} = 0.$$

The superscript index counts the generations whereas the subscript index labels the parents. Recombination and mutation are independent processes. Therefore the effect of recombination can be described by a single pdf. Since the (physical) mutations are $N(0, \sigma)$ normally distributed and recombination is simply a sampling from certain normal variates generated by theses mutations the resulting offspring are normally distributed, too. Let $\sigma_s^{(g+1)}$ be the standard deviation of a single component of the surrogate mutations $\mathbf{S}^{(g+1)}$ that produce the $(g+1)$ generation from the center of mass $\langle \mathbf{S} \rangle^{(g)}$ which reads for a single component

$$\langle s \rangle^{(g)} = \frac{1}{\mu} \sum_{m=1}^{\mu} s_m^{(g)}. \tag{53}$$

$\sigma_s^{(g+1)}$ is composed from the recombinative $\sigma_R^{(g)}$ and the (physical) mutational part σ (independent from each other)

$$\sigma_s^{(g+1)} = \sqrt{\left(\sigma_R^{(g)}\right)^2 + \sigma^2}. \tag{54}$$

The recombinative part $\sigma_R^{(g)}$ is an estimate of the sampling from the distribution of the $s_m^{(g)}$ characterized by the standard deviation of the surrogate mutations $\sigma_s^{(g)}$. Thus we have for $\sigma_R^{(g)}$

$$\sigma_R^{(g)} = \sqrt{E\left\{\left(s^{(g)} - \langle s \rangle^{(g)}\right)^2\right\}}. \tag{55}$$

Note, $\langle s \rangle^{(g)}$ as defined by (53) is a random variate itself. The $s_m^{(g)}$ variates generated by mutation and recombination have the expectation

$$E\left\{s_m^{(g)}\right\} = 0$$

and are not correlated

$$E\left\{s_k^{(g)}\,s_l^{(g)}\right\} = \left\{\begin{array}{rl} 0, & k \neq l \\ \left(\sigma_s^{(g)}\right)^2, & k = l \end{array}\right. .$$

For a fixed parent m one obtains from (55)

$$\begin{aligned}
E\left\{\left(s_m^{(g)} - \langle s\rangle^{(g)}\right)^2\right\} &= E\left\{\left(s_m^{(g)} - \frac{1}{\mu}\sum_{k=1}^{\mu} s_k^{(g)}\right)^2\right\} \\
&= E\left\{\left(\left(1-\frac{1}{\mu}\right)s_m^{(g)} - \frac{1}{\mu}\sum_{k\neq m} s_k^{(g)}\right)^2\right\} \\
&= E\left\{\left(1-\frac{1}{\mu}\right)^2\left(s_m^{(g)}\right)^2 - 2\left(1-\frac{1}{\mu}\right)\frac{1}{\mu}\sum_{k\neq m} s_m^{(g)}\,s_k^{(g)} + \frac{1}{\mu^2}\sum_{k\neq m}\sum_{l\neq m} s_k^{(g)}\,s_l^{(g)}\right\} \\
&= \left(1-\frac{1}{\mu}\right)^2 E\left\{s_m^{(g)}\,s_m^{(g)}\right\} + \frac{1}{\mu^2}\sum_{k\neq m}\sum_{l\neq m} E\left\{s_k^{(g)}\,s_l^{(g)}\right\} \\
&= \left(1-\frac{1}{\mu}\right)^2 E\left\{\left(s_m^{(g)}\right)^2\right\} + \frac{\mu-1}{\mu^2} E\left\{\left(s_m^{(g)}\right)^2\right\} \\
&= \frac{\mu-1}{\mu} E\left\{\left(s_m^{(g)}\right)^2\right\} = \frac{\mu-1}{\mu}\left(\sigma_s^{(g)}\right)^2 .
\end{aligned}$$

This gives for (54) the iterative scheme

$$\sigma_s^{(g+1)} = \sqrt{\frac{\mu-1}{\mu}\left(\sigma_s^{(g)}\right)^2 + \sigma^2}\,. \tag{56}$$

Since $\sigma_s^{(0)} = 0$ holds (cp. (52)) one finds by recurrence

$$\sigma_s^{(g+1)} = \sigma\sqrt{1 + \frac{\mu-1}{\mu} + \ldots + \left(\frac{\mu-1}{\mu}\right)^{g-1} + \left(\frac{\mu-1}{\mu}\right)^{g}}\,.$$

Within the square root there is a sum of geometric progression

$$\sigma_s^{(g+1)} = \sigma\sqrt{\left[1-\left(\frac{\mu-1}{\mu}\right)^{g+1}\right]\Big/\left[1-\left(\frac{\mu-1}{\mu}\right)\right]} \tag{57}$$

and the steady state case $g \to \infty$ yields[8]

$$\sigma_s = \lim_{g\to\infty}\sigma_s^{(g)} = \sigma\sqrt{\mu}. \tag{58}$$

This result is very remarkable. By the dominant recombination physical mutations with standard deviation σ are transformed into larger surrogate mutations. However, although there is no selective pressure in this model the resulting mutation strength (58) is restricted. The population is concentrated around the center of mass (the 'wild-type') forming a (what biologist would call) '*species*'. NB, the derivation given does not rely on a special kind of distribution (e. g. normal distribution), but on the statistical independence of mutation and recombination only. Therefore, the result is very general and should hold , e. g., for GAs too. The author will come back to this point in the conclusions of this article.

[8]This formula might as well be obtained by the self-consistent method solving (56) for $\sigma_s^{(g+1)} = \sigma_s^{(g)} = \sigma_s$.

4.2.2 The Progress Rate Formula

Now we can finish with the $\varphi_{\mu/\mu_D,\lambda}$ formula. As pointed out at the end of 4.1 eq. (19) holds for the dominant case, too. The unknown values are $\langle\overline{\mathbf{h}^2}\rangle$ and $\langle\overline{x}\rangle$, now produced by the surrogate mutations with standard deviation $\sigma_s = \sqrt{\mu}\,\sigma$. Because the actual influence of the selection process is not known the simplest approximation for $\langle\overline{\mathbf{h}^2}\rangle$ is it to assume no selection. I. e., each component of the $\mathbf{h}$ vectors is $N(0, \sigma_s)$ normally distributed. Therefore, the expectation for $\langle\overline{\mathbf{h}^2}\rangle$ becomes (cp. Beyer, 1995, eq. (8))

$$\langle\overline{\mathbf{h}^2}\rangle \approx N(\sigma_s)^2 = \mu N \sigma^2. \tag{59}$$

For $\langle\overline{x}\rangle$ we assume a simple selection process taking the best μ offspring of a sample from the $N(0, \sigma_s)$-distribution comprising λ offspring. Formula (20) holds with $p_{m;\lambda}(x) = p_{m:\lambda}(x)$, i. e. these are the simple order statistics of the normal variate x: $p(x) = \exp(-\frac{1}{2}x^2/\sigma_s^2)/\sqrt{2\pi}\sigma_s$. The result is similar to (31)

$$\langle\overline{x}\rangle = \sigma_s\, e^{1,0}_{\mu,\lambda} = \sigma_s\, c_{\mu/\mu,\lambda} = \sqrt{\mu}\,\sigma\, c_{\mu/\mu,\lambda}. \tag{60}$$

If eqn. (59/60) are inserted into (19), one finally obtains

$$\boxed{\varphi^\star_{\mu/\mu_D,\lambda} = N\left(1 - \sqrt{1 + \frac{(\sigma^\star)^2}{N}}\right) + \frac{\sqrt{\mu}\, c_{\mu/\mu,\lambda}\, \sigma^\star}{\sqrt{1 + \frac{(\sigma^\star)^2}{N}}}.} \tag{61}$$

It turns out that the progress rate reachable by dominant recombinant strategies are comparable with those from the intermediate one (cf. point 4.3 below).

4.2.3 Is there Genetic Repair in Dominant Recombinant Strategies?

The mechanism responsible for high progress rates in intermediate strategies is quite clear - the *genetic repair* (i. e. building the center of mass), *explicitly performed by the algorithm* (cp. eq. (1)). However, in dominant strategies there is not such an algorithmic averaging of the genes. What is the reason for their benefits then? It is indeed *genetic repair*, too, but *implicitly performed.* This - at the first glance - astonishing answer becomes clear if one interprets the dominant recombination as a sampling process estimating the distribution of the parents. Thus, the information from the best μ offspring forms an '*internal model*' of the fitness landscape as already conjectured by Schwefel (Schwefel, 1987). But, up until now, it was not quite clear what the notion 'internal model' really means and how an algorithm - simple as the dominant ES - can take advantage from it. As it was shown in the case of the (μ, λ)-ES (Beyer, 1995, section 3.1.3), estimating a distribution is equivalent to determining its moments. In other words, estimating the internal model by dominant recombination includes the implicit estimation of the first moment of the parental distribution. However, the first moment of the parental distribution is nothing else but the center of mass. Now, the wheel turns full circle - dominant recombination estimates the center of mass, thus it performs (estimated or approximative) *implicit genetic repair.*

Returning to the debate regarding the alternative '*genetic variety* OR *genetic repair*' we thus come to an remarkable résumé. Actually there is no alternative. The dominant recombinant ES produces *genetic variety* AND performs (implicit) *genetic repair.* Both mechanisms work together in a synergistic fashion.

4.3 Comparison with Experiments

In order to produce the dotted graphs in Fig. 5 the author has used a simulation technique that reconstructs the actual progress rate $\varphi^\star$ and $\sigma^\star$ from ES-runs by classification. That is, the mutation strength σ is held constant during the evolution run and the observed $\varphi^\star$ and $\sigma^\star$ values are collected in $\sigma^\star$-classes. The class averages are computed and displayed. The left

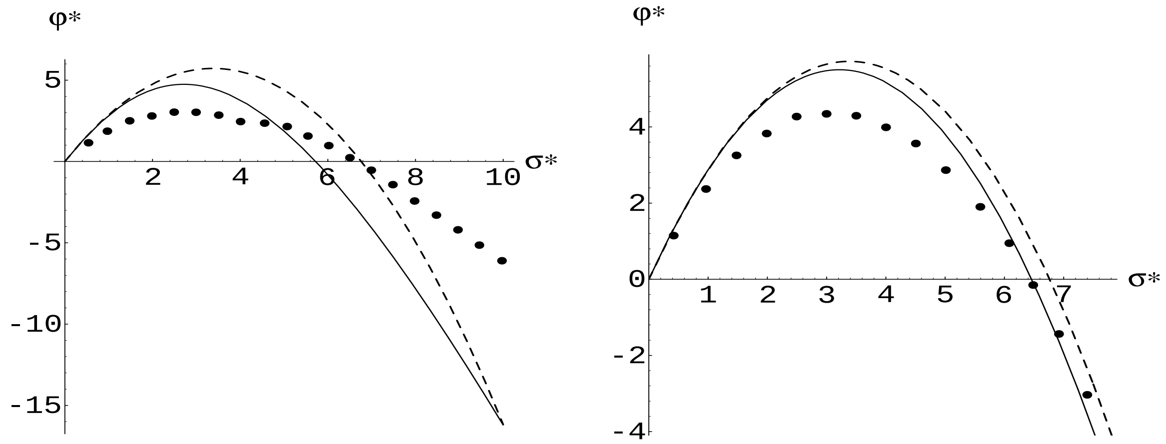


Figure 5: Left picture: $(8/8_D, 30)$-ES with $N = 30$. Right picture: $(8/8_D, 30)$-ES with $N = 200$.

picture is a reconstruction from 100 ES-runs and the right one from 200 ES-runs. The graphs of (61) (solid curves) deviate considerably from the experiments as expected, however, the order of magnitude is correct. The same holds for the Rechenberg formula (dashed curves)

$$\varphi^\star_{\mu/\mu_D,\lambda} = \sqrt{\mu}\, c_{\mu/\mu,\lambda}\sigma^\star - \frac{(\sigma^\star)^2}{2}$$

which can be obtained from (61) by Taylor expansion.

The large deviations just in the interesting $\sigma^\star$-range ($\varphi^\star > 0$) are due to the simple isotropic surrogate mutation model used. The real $p(\mathbf{Z})$ distribution is not isotropic, but compressed in $\mathbf{R}$-direction. A refined surrogate mutation model should take into account different distributions for x and $\mathbf{h}$ in (10/19). This is a task for the future.

5 Non-Spherical Fitness Landscapes

The effort necessary for the determination of progress rates on spherical models is large, however, the results would be of relatively small value if they were restricted to the fitness function $F(\mathbf{y}) = \mathbf{y}^2$ only. This arises the question about the extendibility of the spherical model to non-spherical success domains. And indeed, under certain conditions such an extension can be obtained straightforward by RIEMANNian geometry.

It is the basic idea to approximate the fitness landscape locally by a hypersphere surface. That way, the problem reduces to the task determining the (local) *mean radius* $R_{\overline{\kappa}}$ of the success domain at the parental point. This is a differential geometry problem to be solved in 5.2.

Note, in this section the author uses *Einstein's summation convention* for Greek dummy indices running from 1 to $N-1$, e. g.

$$a_\alpha b^\alpha := \sum_{\alpha=1}^{N-1} a_\alpha b^\alpha .$$

It is assumed that the reader is familiar with the tensor calculus and some basics from Riemannian geometry (cp. e. g. Kreyszig, 1968).

5.1 The Basic Aspects of the Differential Geometry Approach

5.1.1 The Local Fitness Function $Q(\mathbf{x})$

Given the fitness function $F(\mathbf{y})$ the local behavior of F at the parental state $\mathbf{y} = \mathbf{y}^{(g)}$ may be approximated by a Taylor expansion

$$F(\mathbf{y}+\mathbf{x}) = F(\mathbf{y}) + \sum_{i=1}^{N} \frac{\partial F}{\partial y_i} x_i + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\partial^2 F}{\partial y_i \, \partial y_j} x_i x_j + \ldots$$

breaking off after the quadratic term. Such an approximation is permissible if F is differentiable at $\mathbf{y}$ and $\mathbf{x}$ is not too large, usually fulfilled for the mutations $\mathbf{x} = \mathbf{Z}$. Therefore, the local behavior of F can be approximated by the bilinear form

$$\begin{aligned} Q_{\mathbf{y}}(\mathbf{x}) &= F(\mathbf{y}+\mathbf{x}) - F(\mathbf{y}) = \sum_{i=1}^{N} \frac{\partial F}{\partial y_i} x_i + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\partial^2 F}{\partial y_i \, \partial y_j} x_i x_j \\ &= \sum_{i=1}^{N} a_i x_i - \sum_{i=1}^{N} \sum_{j=1}^{N} q_{ij} x_i x_j \end{aligned} \tag{62}$$

with

$$a_i = a_i(\mathbf{y}) = \frac{\partial F}{\partial y_i} \quad \text{and} \quad q_{ij} = q_{ij}(\mathbf{y}) = -\frac{1}{2} \frac{\partial^2 F}{\partial y_i \, \partial y_j}, \quad q_{ij} = q_{ji}.$$

$Q_{\mathbf{y}}(\mathbf{x}) = 0$ defines the boundary ∂Q of the *local success domain*, i. e. that part of the $\mathbb{R}^N$ parameter space containing individuals (if produced) fitter than the parent $\mathbf{y} = \mathbf{y}^{(g)}$. The boundary ∂Q: $Q(\mathbf{x}) = 0$, geometrically interpreted, is a hypersurface in the $\mathbb{R}^N$. Introducing the Cartesian system $\{\mathbf{e}_i\}$, $i = 1 \ldots N$, the hypersurface ∂Q may be described by the vector function

$$\mathbf{x}(u^1, u^2, \ldots u^{N-1}) = \sum_{i=1}^{N} x_i(u^1, u^2, \ldots u^{N-1}) \, \mathbf{e}_i \tag{63}$$

depending on $(N-1)$ (contravariant) coordinates u^α $(\alpha = 1 \ldots N-1)$. Because the local geometry of ∂Q at the parental position $\mathbf{x} = \mathbf{0}$ is of interest, a natural coordinate system is given by

$$x_1 = u^1, \quad x_2 = u^2, \; \ldots \; x_{N-1} = u^{N-1}, \quad x_N = f(u^1, u^2, \ldots u^{N-1}). \tag{64}$$

Hence, x_N is obtained by solving $(Q = Q(\mathbf{x}(u^1, \ldots u^{N-1})) = 0$ for x_N. It is obvious that x_N depends on the $(N-1)$ coordinates u^α and $u^\alpha = 0$ $(\alpha = 1, \ldots N-1)$ $\Rightarrow$ $\mathbf{x} = \mathbf{0}$.

5.1.2 The Metric Tensor $g_{\alpha\beta}$

The local metric on the $(N-1)$-dimensional Riemannian manifold is given by the first fundamental form

$$(ds)^2 = g_{\alpha\beta} du^\alpha \, du^\beta \tag{65}$$

established by the (covariant) metric tensor $g_{\alpha\beta}$

$$g_{\alpha\beta} = \mathbf{x}_{u^\alpha}^T \mathbf{x}_{u^\beta}; \qquad \mathbf{x}_{u^\alpha} := \frac{\partial}{\partial u^\alpha}\mathbf{x}$$

constituted by the scalar products of the tangent vectors $\mathbf{x}_{u^\alpha}$, $\mathbf{x}_{u^\beta}$. For the tangent vectors $\mathbf{x}_{u^\alpha}$ one finds from (63/64)

$$\mathbf{x}_{u^\alpha} = \mathbf{e}_\alpha + \mathbf{e}_N \frac{\partial f}{\partial u^\alpha} \tag{66}$$

Thus, the metric tensor reads

$$g_{\alpha\beta} = \delta_{\alpha\beta} + \frac{\partial f}{\partial u^\alpha}\frac{\partial f}{\partial u^\beta} \tag{67}$$

with Kronecker's delta symbol $\delta_{\alpha\beta}$. For further calculations the contravariant metric tensor will be needed. Because of

$$g_{\alpha\beta} g^{\beta\gamma} = g_\alpha{}^\gamma = \begin{cases} 1, & \alpha = \gamma \\ 0, & \alpha \neq \gamma \end{cases} , \tag{68}$$

$g^{\beta\gamma}$ is the inverse tensor to $g_{\alpha\beta}$. $(g_{\beta\gamma})^{-1}$ can be formally obtained by a matrix power series, adapted from the simple Taylor expansion

$$(1+c)^{-1} = 1 - c + c^2 - c^3 + c^4 - \ldots, \tag{69}$$

for the matrix $\mathbf{E} + \mathbf{C}$ with $\mathbf{E}$ as identity matrix

$$(\mathbf{E} + \mathbf{C})^{-1} = \mathbf{E} - \mathbf{C} + \mathbf{C}\,\mathbf{C} - \mathbf{C}\,\mathbf{C}\,\mathbf{C} + \mathbf{C}^4 - \ldots. \tag{70}$$

Let

$$\{\mathbf{C}\}_{\beta\gamma} := \frac{\partial f}{\partial u^\beta}\frac{\partial f}{\partial u^\gamma} \qquad \text{and} \qquad \{\mathbf{E}\}_{\beta\gamma} = \delta_{\beta\gamma} ,$$

one obtains from (70)

$$\begin{aligned}
(g_{\beta\gamma})^{-1} &= \left(\delta_{\beta\gamma} + \frac{\partial f}{\partial u^\beta}\frac{\partial f}{\partial u^\gamma}\right)^{-1} \\
&= \delta_{\beta\gamma} - \frac{\partial f}{\partial u^\beta}\frac{\partial f}{\partial u^\gamma} + \frac{\partial f}{\partial u^\beta}\left(\sum_\alpha \frac{\partial f}{\partial u^\alpha}\frac{\partial f}{\partial u^\alpha}\right)\frac{\partial f}{\partial u^\gamma} \\
&\quad - \frac{\partial f}{\partial u^\beta}\left(\sum_\alpha \frac{\partial f}{\partial u^\alpha}\frac{\partial f}{\partial u^\alpha}\right)\left(\sum_\alpha \frac{\partial f}{\partial u^\alpha}\frac{\partial f}{\partial u^\alpha}\right)\frac{\partial f}{\partial u^\gamma} + \ldots \\
(g_{\beta\gamma})^{-1} &= \delta_{\beta\gamma} - \frac{\partial f}{\partial u^\beta}\frac{\partial f}{\partial u^\gamma}\left[1 - \left(\sum_\alpha \frac{\partial f}{\partial u^\alpha}\frac{\partial f}{\partial u^\alpha}\right) + \left(\sum_\alpha \frac{\partial f}{\partial u^\alpha}\frac{\partial f}{\partial u^\alpha}\right)^2 - \ldots\right].
\end{aligned} \tag{71}$$

If (69) is compared with the bracket in (71) one reads

$$g^{\beta\gamma} = (g_{\beta\gamma})^{-1} = \delta_{\beta\gamma} - \frac{\frac{\partial f}{\partial u^\beta}\frac{\partial f}{\partial u^\gamma}}{1 + \sum_{\alpha=1}^{N-1}\frac{\partial f}{\partial u^\alpha}\frac{\partial f}{\partial u^\alpha}}. \tag{72}$$

The correctness of (72) can be easily proved by contraction with (67) yielding (68).

5.1.3 The Second Fundamental Form

Consider a curve $\mathcal{C}$ through the parental point $\mathbf{x} = \mathbf{0}$ on the surface ∂Q parametrized by its arc length s

$$\mathcal{C}: \quad \mathbf{x}_{\mathcal{C}} = \mathbf{x}\left(u^1(s), u^2(s), \ldots u^{N-1}(s)\right).$$

The principal normal $\mathbf{h}$ to $\mathcal{C}$ is given by the second derivative of $\mathbf{x}$

$$\mathbf{h} = \frac{\mathbf{x}''}{\kappa}, \quad \kappa = \|\mathbf{x}''\|, \quad (\cdots)' = \frac{d}{ds}(\cdots), \tag{73}$$

with the curvature κ of $\mathcal{C}$. For the second derivative one finds

$$\mathbf{x}'' \;=\; \mathbf{x}_{u^\alpha u^\beta} u^{\alpha\prime} u^{\beta\prime} + \mathbf{x}_{u^\alpha} u^{\alpha\prime\prime}. \tag{74}$$

If the curve $\mathcal{C}$ is of such a kind that in the parental point $\mathbf{x} = \mathbf{0}$ the principal normal is parallel to the surface normal $\mathbf{n}$ of ∂Q, then one speaks of a normal section curve $\mathcal{C}_n$ with the normal curvature κ_n. The scalar product of $\mathbf{h}$ and $\mathbf{n}$ for such a curve is $\mathbf{h}^T\mathbf{n} = 1$. Hence, we obtain from (73/74)

$$\mathcal{C}_n: \quad \mathbf{h}^T\mathbf{n} = 1 \quad \Longrightarrow \quad \kappa_n = \mathbf{n}^T\mathbf{x}'' = \mathbf{n}^T\mathbf{x}_{u^\alpha u^\beta} u^{\alpha\prime} u^{\beta\prime} + \mathbf{n}^T\mathbf{x}_{u^\alpha} u^{\alpha\prime\prime}. \tag{75}$$

The normal vector $\mathbf{n}$ on ∂Q is given by

$$\mathbf{n} = \frac{\nabla Q}{\|\nabla Q\|}, \qquad \nabla Q := \sum_{i=1}^{N} \mathbf{e}_i \frac{\partial}{\partial x_i}. \tag{76}$$

The differentiation of $Q\left(\mathbf{x}(u^1, u^2, \ldots u^{N-1})\right) = 0$ yields

$$\frac{\partial Q}{\partial u^\alpha} = 0 \quad \Longrightarrow \quad 0 = \sum_{i=1}^{N} \frac{\partial Q}{\partial x_i}\frac{\partial x_i}{\partial u^\alpha} = \nabla^T Q\, \mathbf{x}_{u^\alpha},$$

and therefore $\mathbf{n}^T\mathbf{x}_{u^\alpha} = 0$ holds. Thus, one obtains from (75) the second fundamental form for the normal curvature κ_n

$$\kappa_n = b_{\alpha\beta} u^{\alpha\prime} u^{\beta\prime}, \tag{77}$$

$$b_{\alpha\beta} := \mathbf{n}^T\mathbf{x}_{u^\alpha u^\beta}. \tag{78}$$

Note, the normal curvature κ_n depends on the tangent direction of the curve in the parental point. There are $N-1$ independent directions forming the basis for the calculation of the mean curvature.

5.1.4 The Mean Curvature $\overline{\kappa}$ of ∂Q

The normal curvature κ_n still depends on the tangent direction of the curve in the parental point. We are interested, however, in a curve-independent characterization of the behavior of ∂Q at the parental state $\mathbf{x} = \mathbf{0}$. This can be achieved by determination of the principal curvatures κ_p defined as the extreme values of κ_n. To proceed, we first re-write (77)

$$\kappa_n = b_{\alpha\beta} u^{\alpha\prime} u^{\beta\prime} = b_{\alpha\beta}\frac{du^\alpha}{ds}\frac{du^\beta}{ds} = \frac{b_{\alpha\beta} du^\alpha du^\beta}{(ds)^2}.$$

By application of the first fundamental form (65) one obtains

$$\kappa_n = \frac{b_{\alpha\beta} du^\alpha du^\beta}{g_{\alpha\beta} du^\alpha du^\beta}.$$

As can be seen κ_n depends only on the tangent direction given by du^α. Therefore, the principal curvatures κ_p are determined by the extremal condition

$$\kappa_p = \kappa_n \quad \Longleftrightarrow \quad \frac{\partial \kappa_n}{\partial l^\alpha} = 0$$

for the function

$$\kappa_n(l^1, \ldots l^{N-1}) = \frac{b_{\alpha\beta} l^\alpha l^\beta}{g_{\alpha\beta} l^\alpha l^\beta}.$$

Performing the derivation yields $b_{\alpha\beta} l^\beta = \kappa_p g_{\alpha\beta} l^\beta$. With $l^\beta = g^{\beta\gamma} l_\gamma$ and (68) we obtain

$$b_\alpha{}^\gamma l_\gamma = \kappa_p l_\alpha,$$

which is an eigenvalue problem. Because of (78) $b_\alpha{}^\gamma$ is a symmetric matrix. Therefore, the eigenvalues $\kappa_1, \kappa_2, \ldots \kappa_{N-1}$ will be real and there is an orthogonal system of eigenvectors defining the $N-1$ principal directions. Each of this tangent directions has its specific normal curvature $\kappa_n = \kappa_p$, called principal curvature. The sum of these principal curvatures defines the so-called *mean curvature* M^* (cp. Kreyszig, 1968). For our purpose the author will give a slightly different definition interpreting the 'mean curvature' as *averaging the* $N-1$ *principal curvatures*

$$\overline{\kappa} = \frac{1}{N-1} \sum_{p=1}^{N-1} \kappa_p.$$

The sum over the eigenvalues κ_p can be easily obtained. From the eigenvalue theory it is well-known that the sum of the eigenvalues is equal to the trace of the matrix

$$\sum_{p=1}^{N-1} \kappa_p = Tr\{b_\alpha{}^\gamma\} = b_\alpha{}^\alpha = b_{\alpha\beta} g^{\alpha\beta}.$$

This is a very satisfactory result because the trace of a second order tensor as a contraction by means of $g^{\alpha\beta}$ is a scalar invariant representing a property of the 'inner' geometry independent from the coordinate system used. Furthermore, the mean curvature $\overline{\kappa}$ does not depend on the curve $\mathcal{C}$ chosen on ∂Q. And for the hypersphere, $\overline{\kappa}$ is simply the reciprocal to its radius R. Therefore, for non-spherical success domains we may use the *mean radius* $R_{\overline{\kappa}}$

$$R_{\overline{\kappa}} = \frac{1}{\overline{\kappa}} = \frac{N-1}{b_{\alpha\beta} g^{\alpha\beta}} \tag{79}$$

as an approximation describing the local behavior of the fitness landscape by a hypersphere.

5.2 Determination of the Mean Radius $R_{\overline{\kappa}}$

In order to calculate $R_{\overline{\kappa}}$, $g^{\alpha\beta}$ and $b_{\alpha\beta}$ are to be determined for the manifold (64) at the parental state $\mathbf{x} = \mathbf{0}$, i. e. for $u^1 = u^2 = \ldots = u^{N-1} = 0$.

5.2.1 The Metric Tensor

The metric tensor $g^{\alpha\beta}$ is given by (72). This needs the calculation of $\frac{\partial f}{\partial u^\alpha}$ at $u^\alpha = 0$. Since $f(u^1, \dots u^{N-1})$ is the root of x_N to the equation $Q\left(\mathbf{x}(u^1, \dots u^{N-1})\right) = 0$ (cp. eq. (64)), the direct determination of $\frac{\partial f}{\partial u^\alpha}$ is rather tedious. Implicit differentiation will be used instead. If $Q(u^1, \dots u^{N-1}) = 0$ is differentiated taking (64) into account one obtains

$$\frac{\partial Q}{\partial u^\alpha} = \frac{\partial Q}{\partial x_\alpha} + \frac{\partial Q}{\partial x_N}\frac{\partial x_N}{\partial u^\alpha} = \frac{\partial Q}{\partial x_\alpha} + \frac{\partial Q}{\partial x_N}\frac{\partial f}{\partial u^\alpha} = 0 \tag{80}$$

and therefore

$$\left.\frac{\partial f}{\partial u^\alpha}\right|_{u^{(\cdot)}=0} = -\left(\left.\frac{\partial Q}{\partial x_\alpha}\right|_{\mathbf{x}=\mathbf{0}}\right)\Big/\left(\left.\frac{\partial Q}{\partial x_N}\right|_{\mathbf{x}=\mathbf{0}}\right).$$

For the fitness bilinear form (62) we thus obtain

$$\left.\frac{\partial f}{\partial u^\alpha}\right|_{u^{(\cdot)}=0} = -\frac{a_\alpha}{a_N}, \tag{81}$$

and the components of the contravariant metric tensor (72) read

$$g^{\alpha\beta} = \delta_{\alpha\beta} - \frac{a_\alpha a_\beta / a_N^2}{1 + \frac{1}{a_N^2}\sum_{\gamma=1}^{N-1} a_\gamma a_\gamma} = \delta_{\alpha\beta} - \frac{a_\alpha a_\beta}{\sum_{i=1}^{N}(a_i)^2}. \tag{82}$$

5.2.2 The $b_{\alpha\beta}$-Tensor

$b_{\alpha\beta}$ is defined by (78). First, the surface normal vector $\mathbf{n}$ at $\mathbf{x} = \mathbf{0}$ is determined. From (76) with (62) one finds

$$\nabla Q|_{\mathbf{x}=\mathbf{0}} = \sum_{i=1}^{N} a_i \mathbf{e}_i \quad \Longrightarrow \quad \mathbf{n} = \frac{\sum_{i=1}^{N} a_i \mathbf{e}_i}{\sqrt{\sum_{i=1}^{N}(a_i)^2}}. \tag{83}$$

The determination of $\mathbf{x}_{u^\alpha u^\beta}$ needs further considerations. Applying $\partial/\partial u^\beta$ on (66) yields

$$\mathbf{x}_{u^\alpha u^\beta} = \frac{\partial}{\partial u^\beta}\mathbf{x}_{u^\alpha} = \mathbf{e}_N \frac{\partial^2 f}{\partial u^\alpha \partial u^\beta}. \tag{84}$$

The second derivatives of f are implicitly obtained. Differentiation of (80) taking (64) into account gives

$$\begin{aligned}\frac{\partial^2 Q}{\partial u^\beta \partial u^\alpha} &= \frac{\partial^2 Q}{\partial x_\beta \partial x_\alpha} + \frac{\partial^2 Q}{\partial x_N \partial x_\alpha}\frac{\partial f}{\partial u^\beta} + \frac{\partial^2 Q}{\partial x_\beta \partial x_N}\frac{\partial f}{\partial u^\alpha} \\ &\quad + \frac{\partial^2 Q}{\partial x_N \partial x_N}\frac{\partial f}{\partial u^\beta}\frac{\partial f}{\partial u^\alpha} + \frac{\partial Q}{\partial x_N}\frac{\partial^2 f}{\partial u^\beta \partial u^\alpha} = 0.\end{aligned}$$

At $\mathbf{x} = \mathbf{0}$ (i. e. $u^{(\cdot)} = 0$) we have for the bilinear case (62) with (81)

$$\begin{aligned}\left.\frac{\partial^2 Q}{\partial u^\beta \partial u^\alpha}\right|_{u^{(\cdot)}=0} &= -2q_{\alpha\beta} - 2q_{\alpha N}\left(-\frac{a_\beta}{a_N}\right) - 2q_{\beta N}\left(-\frac{a_\alpha}{a_N}\right) \\ &\quad - 2q_{NN}\left(-\frac{a_\alpha a_\beta}{a_N^2}\right) + a_N \left.\frac{\partial^2 f}{\partial u^\beta \partial u^\alpha}\right|_{u^{(\cdot)}=0} = 0,\end{aligned}$$

and (84) becomes

$$\mathbf{x}_{u^\alpha u^\beta} = \mathbf{e}_N \frac{2}{a_N} \left(q_{\alpha\beta} - \frac{q_{\alpha N} a_\beta + q_{\beta N} a_\alpha}{a_N} + \frac{q_{NN}}{a_N^2} a_\alpha a_\beta \right). \tag{85}$$

From (78), (83), and (85) we obtain the second fundamental tensor $b_{\alpha\beta}$

$$b_{\alpha\beta} = \mathbf{n}^T \mathbf{x}_{u^\alpha u^\beta} = \frac{2}{\sqrt{\sum_{i=1}^N a_i^2}} \left(q_{\alpha\beta} - \frac{q_{\alpha N} a_\beta + q_{\beta N} a_\alpha}{a_N} + \frac{q_{NN}}{a_N^2} a_\alpha a_\beta \right). \tag{86}$$

5.2.3 Calculation of the Mean Radius $R_{\overline{\kappa}}$

The contraction of $b_{\alpha\beta}$ with the metric tensor (82) gives

$$\begin{aligned} b_{\alpha\beta} g^{\alpha\beta} \quad = \quad & \frac{2}{\sqrt{\sum_{i=1}^N a_i^2}} \Bigg\{ \sum_{\alpha=1}^{N-1} \left(q_{\alpha\alpha} - 2\frac{q_{\alpha N} a_\alpha}{a_N} + \frac{q_{NN}}{a_N^2} a_\alpha^2 \right) \\ & - \frac{1}{\sum_{i=1}^N a_i^2} \left[\sum_{\alpha=1}^{N-1} \sum_{\beta=1}^{N-1} a_\alpha q_{\alpha\beta} a_\beta - \frac{2}{a_N} \left(\sum_{\alpha=1}^{N-1} q_{\alpha N} a_\alpha \right) \left(\sum_{\beta=1}^{N-1} a_\beta^2 \right) \right. \\ & \left. + \frac{q_{NN}}{a_N^2} \left(\sum_{\alpha=1}^{N-1} a_\alpha^2 \right) \left(\sum_{\beta=1}^{N-1} a_\beta^2 \right) \right] \Bigg\}. \end{aligned}$$

By extending the sums from $N-1$ to N one obtains

$$b_{\alpha\beta} g^{\alpha\beta} \quad = \quad \frac{2}{\sqrt{\sum_{i=1}^N a_i^2}} \left[\sum_{i=1}^N q_{ii} - \frac{\sum_{i=1}^N \sum_{k=1}^N a_i q_{ik} a_k}{\sum_{i=1}^N a_i^2} \right]. \tag{87}$$

This is an astonishing simple expression, especially, if expressed by matrix algebra $\{\mathbf{Q}\}_{ik} := q_{ik}$, $\{\mathbf{a}\}_i := a_i$, and $\|\mathbf{a}\| = \left(\sum_{i=1}^N a_i^2\right)^{1/2}$, then eq. (87) reads[9]

$$b_{\alpha\beta} g^{\alpha\beta} = \frac{2}{\|\mathbf{a}\|} \left(Tr\{\mathbf{Q}\} - \frac{\mathbf{a}^T \mathbf{Q}\, \mathbf{a}}{\|\mathbf{a}\|^2} \right). \tag{88}$$

Therefore, the mean radius (79) can be expressed by

$$R_{\overline{\kappa}} = \frac{N-1}{2} \frac{\|\mathbf{a}\|}{Tr\{\mathbf{Q}\} - \frac{\mathbf{a}^T \mathbf{Q}\, \mathbf{a}}{\|\mathbf{a}\|^2}}. \tag{89}$$

The normalized mutation strength $\sigma^\star = \sigma N / R_{\overline{\kappa}}$ and the progress rate $\varphi^\star = \varphi N / R_{\overline{\kappa}}$ becomes (assuming $N/(N-1) \to 1$ for large N)

$$\boxed{\sigma^\star = \sigma \frac{2}{\|\mathbf{a}\|} \left(Tr\{\mathbf{Q}\} - \frac{\mathbf{a}^T \mathbf{Q}\, \mathbf{a}}{\|\mathbf{a}\|^2} \right), \quad \varphi^\star = \varphi \frac{2}{\|\mathbf{a}\|} \left(Tr\{\mathbf{Q}\} - \frac{\mathbf{a}^T \mathbf{Q}\, \mathbf{a}}{\|\mathbf{a}\|^2} \right). \qquad (90)}$$

[9] $Tr\{\mathbf{Q}\}$ is the *trace* of $\mathbf{Q}$: $Tr\{\mathbf{Q}\} = \sum_i q_{ii}$.

Thus, the progress rate formulas obtained for the spherical model can be approximately used for non-spherical success domains.

The quality of the $R_{\overline{\kappa}}$-approximation depends on the local fitness function $Q(\mathbf{x})$ to be approximated. For the discussion of experimental results the reader is referred to (Beyer, 1994a). The quality mainly depends on the eigenvalue spectrum of Q. If the eigenvalues are concentrated around $Tr\{\mathbf{Q}\}/(N-1)$, then the results will be excellent. In such a case, provided $\mathbf{Q}$ is definite, one can even drop the second term in (88). This is due to the second term - also known as the Rayleigh quotient - that is bounded by the minimum/maximum of the $\mathbf{Q}$ spectrum.

6 Conclusions and Outlook

This paper has dealt with multi-recombinant Evolution Strategies. The theory developed applies to the spherical model, but can be extended to non-spherical fitness landscapes by introducing a mean curvature/radius depending on the local properties of the parental (or center of mass) state.

Unlike the (μ, λ)-theory developed in the last issue of this journal, the $(\mu/\mu_I, \lambda)$ intermediate recombinant strategy has been easily analyzed. This was due to the multi-recombinant averaging concentrating the parents in the center of mass of the best μ offspring. The progress formula is similar to the $(1, \lambda)$ formula.

For $(\mu/\mu_I, \lambda)$ one observes astonishing large progress rates. Genetic repair accounts for this phenomenon. This led the author to some speculations concerning the benefits of sexuality (in the sense of recombination). In biology, arguments against and also in favour of the genetic repair hypothesis are concentrated mainly on the advantage for the *single individual* produced by recombination/crossover. This is questionable in the authors opinion. At least for the EA-theory, *genetic repair is mainly a statistic phenomenon acting on the whole population/species.*

For the ES-practice it is important to know the optimum number of parents producing a fixed number of offspring. Because this is an N-dependent problem, a table has been presented meeting the needs for most of the applications.

A rather difficult problem concerns the $(\mu/\mu_D, \lambda)$ dominant recombinant strategies. It was the basic idea to interpret the results of the recombination as if it were produced by a mutation substitute, called the surrogate mutation, performed from the 'center of mass parent'. Note, from the probability theory there is not a big difference between recombination and mutation. Both can be described by densities. However, the question arises on how to determine this recombination/mutation density. A very first approach, used in this paper, neglects the selection. It was interesting to discover that the system mutation-recombination without selection, but iteratively performed, generates a steady state distribution of the parents concentrated around an 'average parent'. (One might as well expect a successive dispersal of the parental distribution!).

The isotropic surrogate mutation approach yields results lying in the order of magnitude of the experiments. Further research effort will be necessary to improve the progress rate formula. However, the principle and the reason for the benefits of dominant recombination should have become clear.

Unlike most of the EA researchers explaining the advantages of recombination (in parameter optimizing ESs and GAs as well) by some kind of building block hypothesis (see, for example, Goldberg, 1989), the author advocates for genetic variety and genetic repair implicitly performed by the dominant recombination process. This statement is of general validity. It should hold for GAs, too, and we may draw three conclusions for the GA-theory/practice (omitting a fourth one concerning the selection)

First, *mutation is the motor of evolution.* It is the source for the variation of the genome. Therefore, never underestimate the power of mutation: An EA without recombination can work, whereas *an EA without mutations will get stuck* (NB, in real GAs the population size is always limited).

Second, recombination is a process forming a 'species' crowded around the 'wild-type parent'. This is equivalent to estimating the parent/offspring distribution, including implicitly the estimation of the center of mass (the wild-type). Thus, the dominant recombination performs implicitly genetic repair.

Third, dominant recombination has a triple effect: *genetic repair*, transformation of physical mutations into larger recombinant mutations (*genetic variation*), and building a *species by variational restriction.* Note, as in biology, recombination of different species (different building blocks) does not work in general, because of the violation of the species condition.[10] However, another concept may work well in GAs, too - the multi-recombination. I. e. producing an offspring from more than two parents. Multi-mixing is only rarely observed in biology[11] due to the high costs of nowadays sexuality (at least for the higher creatures: the problem of gender). However, this is not a problem for GAs as far as the 'group sex members' fit into the processor's local cache. It will be worth investigating such multi-mixing GAs.

Acknowledgments

The author would like to thank Hans-Paul Schwefel for valuable hints. This work was funded by the Deutsche Forschungsgemeinschaft, DFG grant Be 1578/1-1.

References

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

Bernstein, H., Hopf, F. A., & Michod, R. E. (1987). The Molecular Basis of the Evolution of Sex. *Advances in Genetics*, 24, 323–370.

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

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

Beyer, H.-G. (1994b). Towards a Theory of 'Evolution Strategies': Results from the N-dependent (μ, λ) and the Multi-Recombinant $(\mu/\mu, \lambda)$ Theory. Technical Report SYS–5/94, Department of Computer Science, University of Dortmund.

Beyer, H.-G. (1995). Toward a Theory of Evolution Strategies: The (μ, λ)-Theory. *Evolutionary*

[10]This does not exclude - under certain, but very limiting conditions - that different building blocks may recombine to fitter ones, e. g., if the fitness is an additive function of the different gene loci. However, this linear case is not very interesting for real world optimizations.

[11]It seems to be a very archaic mechanism, today mainly observed in bacteria species (Margulis & Sagan, 1988).

Computation, 2(4), xxx–yyy.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning.* Reading, MA: Addison-Wesley.

Kreyszig, E. (1968). *Differentialgeometrie.* Leipzig, Germany: AVG Geest & Portig K.-G.

Margulis, L., & Sagan, D. (1988). The Cannibalistic Legacy of Primordial Androgynes. In R. Belling & G. Stevens (Eds.), *Nobel Conference XXIII (1987): The Evolution of Sex*, 23–40. San Francisco, CA: Harper and Row.

Michod, R. E., & Levin, B. R. (Eds.) (1988). *The Evolution of Sex: an Examination of Current Ideas.* Sunderland, MA: Sinauer Associates Inc.

Rechenberg, I. (1978). Evolutionsstrategien. In B. Schneider & U. Ranft (Eds.), *Simulationsmethoden in der Medizin und Biologie*, 83–114. Berlin, Germany: Springer-Verlag.

Rechenberg, I. (1993). *Evolutionsstrategie '93.* Manuscript version of the book: *Evolutionsstrategie '94.*

Rechenberg, I. (1994). *Evolutionsstrategie '94.* Stuttgart, Germany: Frommann–Holzboog Verlag.

Schwefel, H.-P. (1974). Adaptive Mechanismen in der biologischen Evolution und ihr Einfluß auf die Evolutionsgeschwindigkeit. Abschlußbericht zum DFG-Vorhaben Re 215/2. Technical Report, Technical University of Berlin.

Schwefel, H.-P. (1987). Collective Phenomena in Evolutionary Systems. In P. Checkland & I. Kiss (Eds.), *Preprints of the 31st Annual Meeting of the International Society for General Systems Research*, 1025–1032. Budapest.

Schwefel, H.-P. (1995). *Evolution and Optimum Seeking.* New York, NY: Wiley.