Where Elitists Start Limping Evolution Strategies at Ridge Functions

Ahmet Irfan Oyman, Hans-Georg Beyer, Hans-Paul Schwefel

University of Dortmund, Germany,
{oyman, beyer, schwefel}@LS11.cs.uni-dortmund.de

Abstract. How well an optimization algorithm satisfies short-term and long-term goals, can be verified using appropriate test functions, respective convergence measures, theoretical analysis, and simulations. This paper analyses the convergence behavior of the evolution strategy (ES) at the parabolic ridge function using the standard $(1 \overset{+}{,} \lambda)$-ES. Some further results are given for the case of more general ridge functions. The results obtained are counter-intuitive and different from if not contrary to those obtained from the sphere model theory. Furthermore, using static analysis, we show that the progress rate and the quality gain possess entirely different characteristics.
Keywords. ridge functions, Evolution Strategy, elitist ES, progress rate, quality gain, convergence behavior.

1 Introduction

Short-term gains are normally *not* expected to be coherent with the long-term gain or success. In order to be able to measure these two kinds of gains, one needs local and global performance measures. These measures are called "quality gain" and "progress rate", respectively, in the terminology of the evolution strategies. Both of these measures are actually measured locally; however, they are of different nature. The consequences are noticeable e.g. at the performance of the elitist strategies.

The static as well as the dynamic behavior of the elitist and non-elitist Evolution Strategies with one parent are investigated at the parabolic ridge in this paper. Furthermore, some important results on the general case of ridge functions are mentioned. Simulation results are given and compared with those of the sphere model, where appropriate.

The following section (Section 2) serves as historical basis. The background for the optimization algorithms used is given in Section 3. Section 4 is devoted to the test function used in the analysis (parabolic ridge) and to other related functions. Section 5 introduces the convergence measures used, for some of which analytical formulae are derived in Section 6 (theory). These are compared with simulation results in Section 7. A summary is given in Section 8.

2 The historical basis

Evolution Strategies were first devised for solving experimental optimization problems with *discrete* variables, the numbers of which may be variable as well [Rec65,KS70]. Theoretical results, however, were first obtained for a fixed number of real variables [Rec71,Sch75]. For a detailed history see [BFM97].

Rechenberg introduced two types of N-dimensional functions in order to investigate the progress rate φ mainly depending on the problem size N and isotropic mutation strength σ, and to find rules for adjusting σ so that φ is maximized. These two functions, namely sphere model and corridor model, form the basis of the functions analysed in this paper: Ridge functions are closely related with the former one in their definition, and with the latter one by having an open success region. From another point of view, the latter one was formulated to model situations far from the optimum, and the former in its vicinity, respectively. In case of the spherical model, the mutation strength must be decreased permanently in order to come ever closer to the optimum, on the ridge it is sufficient to adjust σ once and for all.

The first $(1+1)$-ES was conceived with one parent and one descendant per generation, in an elitist manner. Later, algorithms with multiple parents, recombination (both intermediary and dominant) and even multiple parallel populations were introduced [Rec78,Her92]. All variables are mutated using normal distributions for real-valued ones, or using geometric distributions for integer variables, respectively. Furthermore, from the beginning, *self-adaptation* of the mutation strength σ was included in the algorithm.

Schwefel argued that a finite life span is advantageous for collectively self-adapting σ, and he devised a more general scheme which comprises plus and comma selection strategies as extreme cases [SR95]. *Aging* and self-adaptation are not considered in this paper.

3 Algorithms

Only the $(1,\lambda)$-ES and the $(1+\lambda)$-ES are considered in this paper. The self-adaptation and recombination mechanisms are not used. The algorithm starts at a random or user-defined location in the search space, $\mathbb{R}^N$. This initial parent generates λ descendants in its neighborhood. Mutations occur according to a Gaussian distribution with zero mean and standard deviation σ for all N variables. The fitness function values obtained for the isotropically scattered offspring are used in the selection. In the $(1,\lambda)$-ES, the descendant with the best fitness substitutes the parent, no matter whether it is better than the parent or not. In the elitist $(1+\lambda)$-ES, however, the parent survives if none of the descendants leads to an improvement. These steps (mutation, evaluation, and selection) are repeated until a termination criterion is satisfied.

4 Problems, test functions

The case of maximization is considered in this paper w.l.o.g. In the ES literature, the *sphere model* test function has been used for most theoretical studies, either *statically* [Bey93], i.e. keeping the parent individual fixed in the search space, or *dynamically*, i.e. over many generations in time and space, if self-adaptation is applied (see [Bey96,Rec94] among others):

$$F(\boldsymbol{x}) := -W\left(\sqrt{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}\right) =: -W(R) \tag{1}$$

NB, $\hat{x}_i$ are the coordinates of the optimizer, R denotes the Euclidean distance to the optimum, and N is the number of variables: Normally, $N \to \infty$ is assumed in the theoretical analysis; however, the results may serve as approximation for $N \geq 30$. $W(.)$ stands for an arbitrary monotonically increasing function. W.l.o.g., the origin ($\hat{\boldsymbol{x}} = \boldsymbol{0}$) can be used as optimizer.

The general case of the *ridge functions*[1] studied here can be stated as

$$F(\boldsymbol{x}) := x_0 - d\left(\sum_{i=1}^{N-1} x_i^2\right)^{\frac{\alpha}{2}} =: x_0 - dr^{\alpha}, \qquad \alpha \in \mathbb{R},\, d \in \mathbb{R}^+, N \in \mathbb{N} \;, \tag{2}$$

having the optimizer at $\hat{x}_0 = +\infty$ (therefore, the x_0-axis can be called *ridge axis* along which the progress is measured) and $\hat{x}_i = 0$, $\forall i \neq 0$, where r is the actual distance from the ridge axis. It has an open success region. Movements in

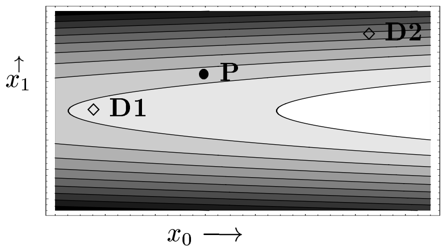

Fig. 1. Contour plot of the parabolic ridge test function, for $N = 2$. The brighter areas have higher fitness values, indicating that the optimum is at far right. The parent **P** and two descendants, **D1** and **D2**, are shown. No matter whether one uses a $(1+2)$-ES or $(1,2)$-ES, **D1** will be selected as the next parent because it has the best fitness value, although **D2** is the nearest one to the optimum.

x_0 direction are privileged linearly, whereas all deviations in other directions are punished by a multiplicative weight d and by a further weight α at the exponent. As α becomes larger, the optimum seeking process resembles an inclined razor riding process. Figure 1 shows a gentle case for $\alpha = 2$ (parabolic ridge). For $\alpha = 0$, one gets the *hyperplane* as the *limit* case, for $\alpha = 1$ the *sharp ridge*.

[1] NB, mathematicians use the same name for *another* function family [Pin97].

Due to the isotropy of the ES mutations chosen, epistatic effects do not degrade the performance even if the ridge or progress axis is *not* aligned with one of the coordinate axes [Sal96].

5 The convergence measures

Strictly speaking, the long-term behavior of an algorithm is of utmost interest in convergence analysis: One wants to know about the number of iterations that are necessary to reach (or come close to) the optimum (or the optimizer). Due to the simple Markovian character of ES, easier to achieve local measures may be used for estimating the long-term behavior. There are at least two of such measures:

$$\varphi^{(g)} := \mathrm{E}\{\|\hat{\boldsymbol{x}} - \boldsymbol{x}^{(g)}\| - \|\hat{\boldsymbol{x}} - \boldsymbol{x}^{(g+1)}\|\} \ , \tag{3}$$

the *progress rate* as defined by Rechenberg [Rec71], which measures the useful Euclidean distance traveled *toward* the optimum in the parameter space within one generation, where $\hat{\boldsymbol{x}}$ denotes the optimum vector. Due to the stochastic character of the process, one has to consider its expected value, of course. This progress rate varies not only over time, but also depends on the current position in the search space. Another measure,

$$\overline{Q}^{(g)} := \mathrm{E}\{F^{(g+1)} - F^{(g)}\} \ , \tag{4}$$

not less dependent from time and local position, is called the average *quality gain*. It has been analysed in [Bey94] for a wide class of fitness functions and mutation operators. The special case of uniform sphere mutations on the special case of $F(\boldsymbol{x}) = -\sum_{i=1}^{N} x_i^2$ has been analysed in [Rud97], too.

For the sphere model, both φ and $\overline{Q}$ only depend on σ, N, and R, for the ridge models introduced in (2) they depend on σ, N, d, α, and r.

The *success rate* ω is the probability of getting a descendant with a better fitness function value than the parent in a *single* trial. For the calculation of ω we need the first two moments of the *quality difference* $\Delta F(z) := F(\boldsymbol{x}_{\mathrm{P}}+\boldsymbol{z}) - F(\boldsymbol{x}_{\mathrm{P}})$, of the parent at position $\boldsymbol{x}_{\mathrm{P}}$ and all possible descendants at $\boldsymbol{x}_{\mathrm{P}} + \boldsymbol{z}$. Let $M_{\Delta F} := \mathrm{E}\{\Delta F(\boldsymbol{z})\}$ and $S_{\Delta F} := \sqrt{\mathrm{E}\{\Delta F^2(\boldsymbol{z})\} - [\mathrm{E}\{\Delta F(\boldsymbol{z})\}]^2}$, respectively. ω can be approximated by (see [Bey94])

$$\omega \approx 1 - \Phi\left(\frac{0 - M_{\Delta F}}{S_{\Delta F}}\right) = \Phi\left(\frac{M_{\Delta F}}{S_{\Delta F}}\right) \ , \tag{5}$$

where "$\Phi(.)$" denotes the cumulative distribution function of the standard normal distribution, $\mathcal{N}(0, 1)$.

6 Theory

We first look closer to the formulation of ridge functions to explain the derivation semantically. The required stationary value for r is derived next. The approximate progress and success rate formulae will be derived in this section using simple local models.

The long-term and short-term aims can be formulated more precisely using (2): Minimization of r causes a short term gain, maximization of x_0 is a long term gain. The reason is simple: r can only be decreased to 0, whereas the values for x_0 are unlimited, but r has normally a greater effect on $\Delta F(\boldsymbol{z})$ than x_0.

Rechenberg already investigated the parabolic ridge ($\alpha = 2$). Under the condition $r = const$, he presents an approximate quality gain ($\overline{Q}$) formula for the $(1, \lambda)$-ES without derivation [Rec94, p. 66], and reasons about local and global progress [Rec94, p. 77]. In [OBS97], we provide derivations for φ, $\overline{Q}$, and ω, as well as the average distance to the ridge axis ($R^{(\infty)}$) for the $(1, \lambda)$-ES, and simulation results for the $(1+\lambda)$-ES, together with confirming simulation results for the $(1, \lambda)$-ES. Extensions to values other than $\alpha = 2$ and asymptotic results for $N \to \infty$ may be found in [OBS98]. The theoretical part of his paper is partially established on these results previously obtained.

A few words can be said here about r at ridge functions. If we start at an arbitrary $r^{(0)}$ value and apply the $(1, \lambda)$-ES algorithm for many generations, keeping σ constant, the $r^{(g)}$ value finally fluctuates around a typical mean value, which we will denote as $R^{(\infty)}$. A similar phenomenon is also observed at the sphere model for dynamic evaluation *without* self-adaptation. The $r \approx R^{(\infty)}$ case is called *stationary* in this paper. The number of generations required for this process within an $(1, \lambda)$-ES can be estimated for the case $\alpha = 2$ [OBS97, Subsection 13.10].

If the x_0 component of ridge functions could be neglected, (2) becomes a special case of the sphere model with $W(r) = dr^\alpha$, $\hat{\boldsymbol{x}} = \mathbf{0}$, in $(N-1)$ dimensions. The $R^{(\infty)}$ value for large σ and d will be derived here for considering x_0 as *noise*. For the sphere model in case $\varphi^*_{sphere} \stackrel{!}{=} 0$, we obtain from the asymptotically $(N \to \infty)$ exact $\varphi^*_{sphere} = c_{1,\lambda}\sigma^* - {\sigma^*}^2/2$ for the $(1, \lambda)$-ES [Rec78,Bey93]

$$0 = c_{1,\lambda} - \frac{\sigma^*}{2}, \qquad \sigma\frac{N}{R} =: \sigma^* = 2c_{1,\lambda} \ , \tag{6}$$

($\varphi^* := \varphi N/R$, $\sigma^* := \sigma N/R$ are normalized, dimensionless items) and therefore $R^{(\infty)}_{sphere} = N\sigma/2c_{1,\lambda}$. Using $(N-1)$ instead of N, the corresponding $R^{(\infty)}$ limit for ridge functions reads

$$R^{(\infty)} \approx \frac{(N-1)\sigma}{2c_{1,\lambda}} \ , \tag{7}$$

The progress coefficient $c_{1,\lambda}$ for the $(1, \lambda)$-ES is the λ-th order statistics of λ samples generated using the standard normal distribution,

$$c_{1,\lambda} = \frac{\lambda}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t\mathrm{e}^{-\frac{1}{2}t^2} [\Phi(t)]^{\lambda-1} \mathrm{d}t \ . \tag{8}$$

The progress rate will be calculated next, where the $R^{(\infty)}$ value will be used to compute the asymptotic ($\sigma \to \infty$) value of φ. For the derivation of approximate φ, the local model is considered first.

The isometric surfaces of ridge functions can locally be approximated by *hyperplanes* for $r \approx R^{(\infty)}$ and corresponding σ (see Fig. 1). Consequently, we have a stationary progress in the direction of the gradient vector $\boldsymbol{a}$ as large as φ at the hyperplane, i.e. $\boldsymbol{\varphi_a} \approx c_{1,\lambda}\sigma\boldsymbol{a}/\|\boldsymbol{a}\|$. If we call the unit vector in x_0 direction as $\boldsymbol{e}_0$, one obtains $\varphi_{ridge} \approx \boldsymbol{e}_0^T \cdot \boldsymbol{\varphi_a}$. By partially differentiating (2) with respect to x_0 and r, the gradient vector $\boldsymbol{a}$ at the location (x_0, r) is obtained as

$$\boldsymbol{a} = \begin{pmatrix} \frac{\partial F}{\partial x_0} \\ \\ \frac{\partial F}{\partial r} \end{pmatrix} = \begin{pmatrix} 1 \\ -d\alpha r^{\alpha-1} \end{pmatrix} \quad \text{, therefore} \tag{9}$$

$$\varphi_{ridge} \approx \boldsymbol{e}_0^T \cdot \boldsymbol{\varphi_a} = \frac{c_{1,\lambda}\sigma}{\|\boldsymbol{a}\|}\boldsymbol{e}_0^T \cdot \boldsymbol{a} = \frac{c_{1,\lambda}\sigma}{\|\boldsymbol{a}\|} = \frac{c_{1,\lambda}\sigma}{\sqrt{1+(\alpha d r^{\alpha-1})^2}} \quad ; \tag{10}$$

$$\varphi_{ridge}\,|_{\alpha=2} = \frac{c_{1,\lambda}\sigma}{\sqrt{1+(2dr)^2}} \quad . \tag{11}$$

It is very remarkable that the negative quadratic term in the asymptotically $(N \to \infty)$ exact $\varphi_{1,\lambda}$ formula for the sphere model does not appear for the ridge function case.

The investigation of the φ limit for very large mutation strengths yields other remarkable results. Replacing r by $R^{(\infty)}$, i.e. inserting (7) into (10), one gets the asymptotic $(\sigma \to \infty)$ value of (10) as $\varphi_{ridge} = \infty$ for $\alpha < 2$, and $\varphi_{ridge} = 0$ for $\alpha > 2$, respectively. Using the normalizations

$$\varphi^* := d(N-1)\varphi \qquad \text{and} \qquad \sigma^* := d(N-1)\sigma \quad , \tag{12}$$

one gets from (11)

$$\lim_{\sigma\to\infty} \varphi_{ridge}\,|_{\alpha=2} = \frac{c_{1,\lambda}^2}{d(N-1)}, \qquad \lim_{\sigma^*\to\infty} \varphi^*_{ridge}\,|_{\alpha=2} = \hat{\varphi}^* = c_{1,\lambda}^2 \quad , \tag{13}$$

and $\hat{\sigma} = \hat{\sigma}^* = +\infty$, where the hat symbol denotes optimal values. The results in (13) are completely independent from σ or σ^*. This is in contrast to the sphere model, where $\lim_{\sigma^*\to\infty} \varphi^*_{sphere} = -\infty$, $\hat{\varphi}^*_{sphere} = c_{1,\lambda}^2/2$, and $\hat{\sigma}^*_{sphere} = c_{1,\lambda}$.

The computation of the success rate ω for the $(1, \lambda)$-ES and $\alpha = 2$ is straightforward. For the calculations, assuming $z_i \sim \mathcal{N}(0, \sigma^2)$, one has $\mathrm{E}\{z_i\} = \mathrm{E}\{z_i^3\} = 0$, $\mathrm{E}\{z_i^2\} = \sigma^2$, $\mathrm{E}\{z_i^4\} = 3\sigma^4$. Using the definition (5), one obtains step by step

$$\mathit{\Delta} F(\boldsymbol{z}) = F(\boldsymbol{x}_{\mathrm{P}} + \boldsymbol{z}) - F(\boldsymbol{x}_{\mathrm{P}}) = z_0 - d\sum_{i=1}^{N-1}(2x_i z_i + z_i^2) \tag{14}$$

$$M_{\mathit{\Delta} F} = -d(N-1)\sigma^2 \tag{15}$$

$$S_{\mathit{\Delta} F} = \sigma\sqrt{1+(2dr)^2 + 2d^2(N-1)\sigma^2} \tag{16}$$

$$\omega \approx \Phi\left(\frac{M_{\mathit{\Delta} F}}{S_{\mathit{\Delta} F}}\right) = \Phi\left(-\frac{d(N-1)\sigma}{\sqrt{1+(2dr)^2+2d^2(N-1)\sigma^2}}\right) \quad . \tag{17}$$

In (16), we used the statistical independence of the z_i. Asymptotically, after inserting (7) into (17), and using (12) one gets $\lim_{\sigma^* \to \infty} \omega = \Phi(-c_{1,\lambda})$. The theoretical results expressed in this section are valid for $N \geq 100$ and $r \approx R^{(\infty)}$.

7 Simulation results

All following simulation results were obtained for $\alpha = 2$, $d = 0.01$, $N = 100$, and $\lambda = 10$; $c_{1,10} \approx 1.54$. Mean values and error bars for 100 000 generations (after a transition phase of 2000 generations to reach stationarity) are plotted as dots for both versions of $(1 \overset{+}{,} 10)$-ES in Fig. 2 and Fig. 3. Continuous lines show the

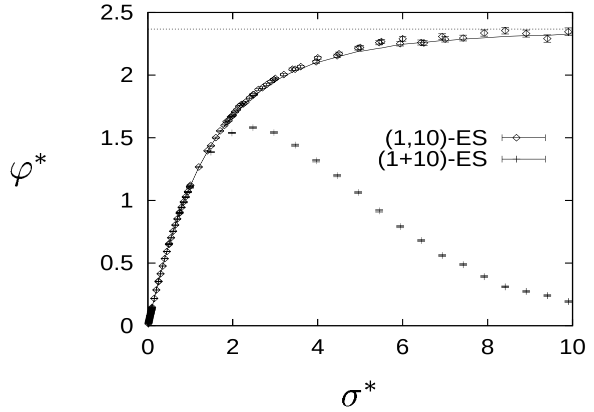

Fig. 2. Normalized progress rate φ^* versus normalized mutation strength σ^* curves for the $(1 \overset{+}{,} 10)$-ES. $d = 0.01$, $N = 100$. For increasing σ^*, the non-elitist strategy attains larger progress rates.

approximate theoretical results for the $(1, 10)$-ES, dotted horizontal lines limits. For the calculation of φ^* and ω over σ^* we did not use the $R^{(\infty)}$ formula (7), but the differing simulation results (right part of Fig. 3). Otherwise, theory and simulation results do not fit as well as they do in the range $2 < \sigma^* < 7$.

The most surprising overall result is the obvious superiority of the non-elitist ES over the elitist one. Whereas the progress rate of the latter goes down to zero as usual after a climax at the optimal mutation strength $\hat{\varphi}^* \approx 2.5$, it goes further up and never down for the comma version. This phenomenon is coupled with linearly increasing $R^{(\infty)}$ and a leveling out of the success rate at a constant limit. A closer investigation reveals the mechanism: The forgetting principle of non-elitism enhances larger, more frequent side steps from the ridge axis with short-term recessions, but better long-term chances to ride along the ridge that is more and more razor-like the larger N becomes.

To further clarify the situation, we use a static analysis in the following, i.e., we fix the distance r to the ridge axis and measure success rates and quality gains for a fixed mutation strength $\sigma = 2$ during 100 000 generations, all starting from the same position, and we do so for both $(1 \overset{+}{,} 10)$-ES under otherwise same conditions as before. Qualitative results of this experiment series are shown in Fig. 4. The theoretically upper limit $\varphi_{hyperplane}|_{\sigma=2} = 2c_{1,10}$ is also shown, as

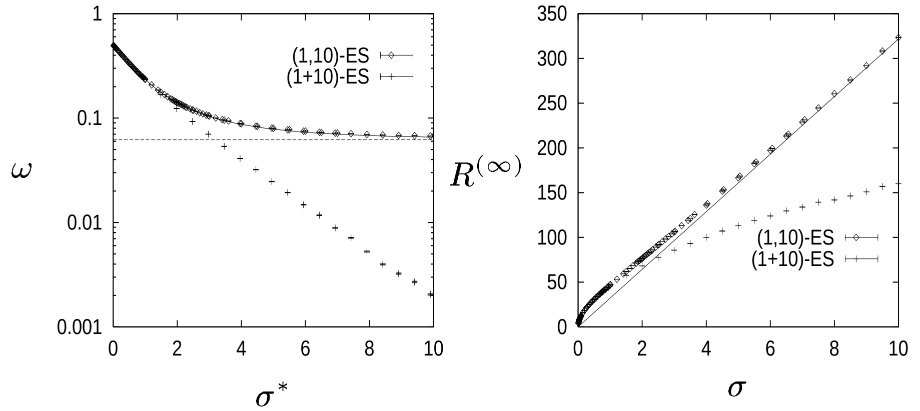

Fig. 3. Success rate ω versus normalized mutation strength σ^* (left picture) and stationary distance to ridge axis $R^{(\infty)}$ versus mutation strength σ curves (right picture) for the $(1 \overset{+}{,} 10)$-ES, obtained from the same simulation as for Fig. 2.

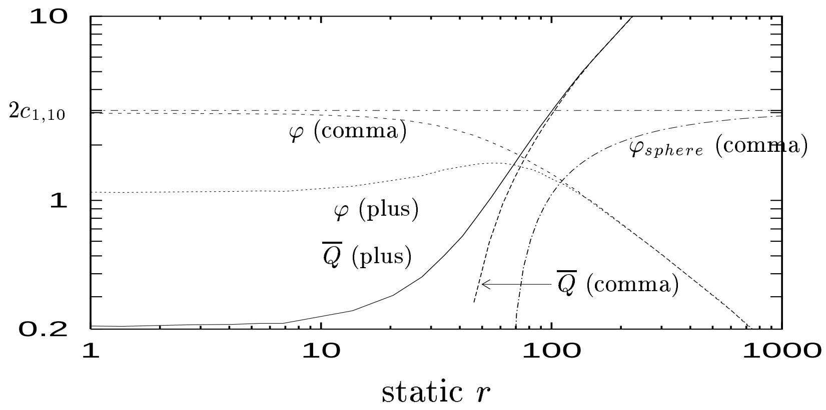

Fig. 4. Progress rate φ and quality gain $\overline{Q}$ over $r = const.$ *Static* analysis for given $\sigma = 2$, otherwise same conditions as before. Both axes are logarithmic. For the parabolic ridge, $\varphi \neq \overline{Q}$ if $r \neq R^{(\infty)}$, furthermore, φ and $\overline{Q}$ behave inversely over r. φ_{sphere} for the $(1, 10)$-ES is given for comparison.

well as the unnormalized, asymptotically $(N \to \infty)$ exact $\varphi_{sphere} = c_{1,10}\sigma - N\sigma^2/2R$ for the $(1, 10)$-ES. Obviously, $\varphi \neq \overline{Q}$ except for $r = R^{(\infty)}$. Remember that for $\sigma = 2$ or $\sigma^* = \sigma d(N-1) \approx 2$ (see Fig. 2) both ES versions yield nearly identical progress rates φ^* (or φ). However, the static analysis gives different results for the two progress measures:

$$\varphi_{1,\lambda} \geq \varphi_{1+\lambda}, \qquad \lim_{r\to\infty} \varphi_{1\overset{+}{,}\lambda} = 0, \qquad \lim_{r\to\infty} \overline{Q}_{1\overset{+}{,}\lambda} = +\infty \tag{18}$$

$$\overline{Q}_{1,\lambda} \leq \overline{Q}_{1+\lambda}, \quad \overline{Q}_{1,\lambda} \neq \varphi_{1,\lambda}, \quad \overline{Q}_{1+\lambda} \neq \varphi_{1+\lambda} \tag{19}$$

The limits for $r \to 0$ also differ from each other.

Figure 4 reveals the discrepancies between the long-term goal of creeping along the ridge axis, as measured by the progress rate φ, and the short-term goal of achieving improvement by coming closer to the ridge axis, as measured by the quality gain $\overline{Q}$: Whereas the quality gain values go to infinity far away from the ridge axis, go down for smaller r when using the elitist ES, but sharply down in case of the non-elitist ES, thus leading to believe in a superiority of elitism, progress rates go down to zero for both versions if r becomes too large, but yield a *sound advantage* of the forgetting strategy for all smaller values of r.

8 Conclusions & Outlook

For the parabolic ridge, one obtains $\hat{\varphi}^*_{1,\lambda} > \hat{\varphi}^*_{1+\lambda}$, $\lim_{\sigma^* \to \infty} \varphi^*_{1,\lambda} = \hat{\varphi}^*_{1,\lambda} = c^2_{1,\lambda}$, and $\hat{\omega} = \Phi(-c_{1,\lambda})$ using dynamic analysis. If the static case is considered, $\overline{Q}$ is not convertible to φ (except for $r = R^{(\infty)}$). At ridge functions, $\varphi_{1,\lambda} \geq 0$. All of these results are new with respect to the sphere model theory.

The $(1+\lambda)$-ES runs after short-term gains, tries implicitly to minimize r, and does this better than the $(1,\lambda)$-ES. This is a consequence of rejecting the replacement of the parent if the best descendant has a lower fitness value than the parent. As a result, the $(1+\lambda)$-ES stays at the same local conditions, attains larger $\overline{Q}$, but smaller φ. I.e. φ is just a byproduct of the selection process.

The analytical results obtained in this paper can be generalized to (μ, λ)-ES, $(\mu/\mu_{\mathrm{I}}, \lambda)$-ES, and under specific conditions to $(\mu/\mu_{\mathrm{D}}, \lambda)$-ES. Respective $R^{(\infty)}$ and φ values have already been derived analytically for these more general strategies. However, they are beyond the scope of this paper.

The results obtained so far may lead to more robust rules for controlling the mutation strength *endogenously*. It will be a challenging task to devise an appropriate mechanism in order to achieve self-adapting mutation strengths that yield maximal long-term performance.

9 Acknowledgments

This research is funded by a PhD scholarship of DAAD (German Academic Exchange Service), grant number A/95/11445. The second author is a Heisenberg fellow of the DFG (German Science Foundation) under grant Be 1578/4-1, and the third one is spokesman of the Collaborative Research Center on Computational Intelligence (SFB 531), sponsored by the DFG, as well. We are grateful to Peter Dittrich for helpful comments.

References

[Bey93] H.-G. Beyer. Toward a Theory of Evolution Strategies: Some Asymptotical Results from the $(1 \overset{+}{,} \lambda)$–Theory. *Evolutionary Computation*, 1(2):165–188, 1993.

[Bey94] H.-G. Beyer. Towards a Theory of 'Evolution Strategies': Progress Rates and Quality Gain for $(1,^{+} \lambda)$-Strategies on (Nearly) Arbitrary Fitness Functions. In Y. Davidor, R. Männer, and H.-P. Schwefel, editors, *Parallel Problem Solving from Nature, 3*, pages 58–67, Heidelberg, 1994. Springer.

[Bey96] H.-G. Beyer. Toward a Theory of Evolution Strategies: Self-Adaptation. *Evolutionary Computation*, 3(3):311–347, 1996.

[BFM97] Th. Bäck, D. B. Fogel, and Z. Michalewicz. *Handbook of Evolutionary Computation*. Oxford University Press, New York, and Institute of Physics Publishing, Bristol, 1997.

[Her92] M. Herdy. Reproductive Isolation as Strategy Parameter in Hierarchically Organized Evolution Strategies. In R. Männer and B. Manderick, editors, *Parallel Problem Solving from Nature, 2*, pages 207–217, Amsterdam, 1992. Elsevier.

[KS70] J. Klockgether and H.-P. Schwefel. Two-phase nozzle and hollow core jet experiments. In D. G. Elliott, editor, *Proc. Eleventh Symp. Engineering Aspects of Magnetohydrodynamics*, pages 141–148, Pasadena CA, March 24-26 1970. California Institute of Technology.

[OBS97] A. I. Oyman, H.-G. Beyer, and H.-P. Schwefel. Analysis of a Simple ES on the "Parabolic Ridge". Technical Report SyS–2/97, University of Dortmund, Department of Computer Science, Systems Analysis Research Group, August 1997. `http://ls11-www.cs.uni-dortmund.de/people/oyman/TR.ps.gz`.

[OBS98] A. I. Oyman, H.-G. Beyer, and H.-P. Schwefel. Convergence Behavior of the $(1 \overset{+}{,} \lambda)$ Evolution Strategy on the Ridge Functions. Technical Report SyS–1/98, University of Dortmund, Department of Computer Science, Systems Analysis Research Group, February 1998. `http://ls11-www.cs.uni-dortmund.de/people/oyman/TR/TR2.ps.gz`.

[Pin97] A. Pinkus. Approximating by Ridge Functions. In A. Le Mehaute, C. Rabut, and L. L. Schumaker, editors, *Surface Fitting and Multiresolution Methods*, pages 279–292. Vanderbilt University Press, Nashville, 1997.

[Rec65] I. Rechenberg. Cybernetic solution path of an experimental problem. Library translation 1122, Royal Aircraft Establishment, Farnborough, 1965.

[Rec71] I. Rechenberg. *Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Dr.-Ing. Dissertation (PhD Thesis), Technical University of Berlin, Department of Process Engineering, 1971.

[Rec78] I. Rechenberg. Evolutionstrategien. In B. Schneider and U. Ranft, editors, *Simulationsmethoden in der Medizin und Biologie*, pages 83–114. Springer, Berlin, 1978.

[Rec94] I. Rechenberg. *Evolutionsstrategie '94*, volume 1 of *Werkstatt Bionik und Evolutionstechnik*. Frommann–Holzboog, Stuttgart, 1994.

[Rud97] G. Rudolph. *Convergence Properties of Evolutionary Algorithms*. Verlag Dr. Kovač, Hamburg, 1997.

[Sal96] R. Salomon. Re-evaluating genetic algorithm performance under coordinate rotation of benchmark functions. A survey of some theoretical and practical aspects of genetic algorithms. *BioSystems*, 39(3):263–278, 1996.

[Sch75] H.-P. Schwefel. *Evolutionsstrategie und numerische Optimierung*. Dr.-Ing. Dissertation (PhD Thesis), Technical University of Berlin, Department of Process Engineering, 1975.

[SR95] H.-P. Schwefel and G. Rudolph. Contemporary Evolution Strategies. In F. Morán, A. Moreno, J. J. Merelo, and P. Chacón, editors, *Advances in Artificial Life. Third International Conference on Artificial Life*, pages 893–907. Springer, Berlin, 1995.