Why Noise May Be Good: Additive Noise on the Sharp Ridge

Silja Meyer-Nieberg
Faculty of Computer Science
Universität der Bundeswehr München
D-85577 Neubiberg, Germany
silja.meyer-nieberg@unibw.de

Hans-Georg Beyer
Department of Computer Science
Research Center Process and Product Engineering
Vorarlberg University of Applied Sciences
Hochschulstr. 1, A-6850 Dornbirn, Austria
Hans-Georg.Beyer@fhv.at

ABSTRACT

This paper considers self-adaptive $(\mu/\mu_I, \lambda)$-evolution strategies on the noisy sharp ridge. The evolution strategy (ES) is treated as a dynamical system using the so-called evolution equations to model the ES's behavior. The approach requires the determination of the one-generational expected changes of the state variables – the progress measures. For the analysis, the stationary state behavior of the ES on the sharp ridge is considered. Contrary to the usual perception of noise, it is shown that noise has a positive influence on the performance. An explanation for this astonishing behavior is given and conditions for the usefulness of noise in other fitness landscapes are discussed.

Categories and Subject Descriptors

G.1.6 [**Optimization**]: Unconstrained Optimization; I.2.8 [**Problem Solving, Control Methods, and Search**]: Heuristic Methods

General Terms

Algorithms, Performance

Keywords

Evolution strategies, self-adaptation, dynamical systems, ridge functions, noise

1. INTRODUCTION

Usually, noise is seen to have a detrimental influence on the outcome of an evolutionary process. On the sphere, for example, noise may cause a residual location error stopping the evolutionary algorithm from reaching its goal – the actual optimizer (see e.g. [4]). Further problems may include a loss of step-size control [12]. Surveys on noisy optimization and evolutionary algorithms can be found for example in [15]. To our knowledge, there are only three papers [16, 20, 7] which come to a different interpretation of the influence of noise. Bäck and Hammel [7] performed an experimental investigation of the effects of noise on three test functions. They found that low levels of noise w.r.t. the fitness do not influence the performance. Furthermore, in the case of one test function, a generalized Ackley function, small noise levels were found to lead to a greater convergence reliability. Levitan and Kauffman investigated adaptive walks under noise – simulating single agent hill-climbing walks on NK-landscapes [16]. They too found that small noise levels might be helpful. In [20], Rana et al. considered local and genetic search algorithms under noisy fitness evaluations on five test functions. Conducting a series of experiments, they identified two effects: A positive "soft annealing effect" meaning that noise smoothes the perceived fitness landscape which enables an algorithm to escape local optima and a negative effect of creating false optima. They found that it depends on the fitness function whether noise leads to positive or negative effects and conjectured that on simple fitness landscapes noise has little positive effects or even deteriorates the performance.

This paper is devoted to an analysis of mutative self-adaptation under noisy fitness evaluations. The adaptation of the mutation strength is a necessity in real-valued search spaces. To this end, several methods, e.g. the 1/5th rule [21] or the cumulative step-size and covariance matrix adaptation [19], have been introduced. One of the methods applied is self-adaptation as introduced by Rechenberg and Schwefel [21, 22]. The main principle of self-adaptation consists in treating the strategy parameters – the mutation strength in this case – in the same manner as the object parameters. They become part of an individual's genome and are evolved alongside with the object parameters.

In this paper, we consider self-adaptation on the noisy sharp ridge. Ridge functions are relatively simple test functions and can be seen as an extension of the sphere model. However, they have practical importance since evolutionary algorithms are often required to follow locally ridge-like function topologies [23]. Although the sharp ridge is a relatively simple function, it poses a problem for self-adaptive evolution strategies and other evolutionary algorithms. As noted by Herdy, they are prone to stagnation on the sharp ridge [14]. Up to now, several analyses of adaptation mechanisms on ridge functions have been presented starting with [14]. Most papers, however, focus on cumulative step-size

GECCO'08, July 12–16, 2008, Atlanta, Georgia, USA.

adaptation. Arnold, for instance, presented an analysis of intermediate evolution strategies on non-noisy ridge functions [3]. Arnold and Beyer [5] focused on CSA-ES on the noisy parabolic ridge. Lunacek and Whitley [17] considered self-adaptive $(1, \lambda)$-ES on the non-noisy parabolic and sharp ridge – gathering experimental data to test several hypotheses on the working principles of evolution strategies. In [6], Arnold and MacLeod compared several adaptation mechanisms on ridge functions. Self-Adaptation is included, but in a different form than in this paper.

This paper focuses on self-adaptation of intermediate evolution strategies on the noisy sharp ridge. It will be shown that the noisy sharp ridge is an example for an optimization problem where noise leads to an improvement. The paper is organized as follows. First, the fitness function and the noise model are presented. Afterwards, the working principles of self-adaptive intermediate evolution strategies are described. In the following section, the analysis approach is introduced and applied to the noisy sharp ridge. The results obtained are then compared with the results from experiments.

1.1 The Noisy Sharp Ridge

Ridge functions $F : \mathbb{R}^N \to \mathbb{R}$ are defined as follows

$$F(\mathbf{y}) \quad = \quad y_1 - d\Big(\sum_{i=2}^{N} y_i^2\Big)^{\alpha/2} \tag{1}$$

with parameters d and α. The parameter d weighs the loss part of the ridge function whereas the parameter α determines the topology of the fitness landscape. A ridge function with $\alpha = 1$ is called a *sharp ridge*. Ridge functions do not have a finite optimum. An algorithm is therefore required to increase the fitness perpetually. As (1) shows, there are two ways to improve the fitness: On the one hand, the linear gain part can be increased which equals a movement parallel to the ridge axis. On the other hand, the loss components can be decreased which equals a movement towards the ridge axis.

In the following, we consider the effects of additive normally distributed noise with zero mean and constant standard deviation (*noise strength*) σ_ϵ. The fitness that is used in the evaluation of an offspring is thus given by

$$\tilde{F}(\mathbf{y}) \quad = \quad y_1 - d\Big(\sum_{i=2}^{N} y_i^2\Big)^{\alpha/2} + \sigma_\epsilon \mathcal{N}(0,1) \tag{2}$$

with $\mathcal{N}(0, 1)$ denoting a standard normally distributed random variable. In the following, two new variables are introduced to simplify the notations in (1) and (2). For the remainder of the paper let $x := y_1$ and $\mathbf{R} := (y_2, \ldots, y_N)^{\mathrm{T}}$. The equations – for instance (2) – change to

$$\begin{aligned} \tilde{F}(x, \mathbf{R}) \quad &= \quad x - d\Big(\mathbf{R}^{\mathrm{T}}\mathbf{R}\Big)^{\alpha/2} + \sigma_\epsilon \mathcal{N}(0,1) \\ &= \quad x - dR^\alpha + \sigma_\epsilon \mathcal{N}(0,1) \end{aligned} \tag{3}$$

with $R := \|\mathbf{R}\|$.

1.2 The Self-Adaptive Intermediate Evolution Strategy

In this paper, self-adaptive intermediate $(\mu/\mu_I, \lambda)$ evolution strategies are considered. In the following, the basic working principle is described. Let $\mathbf{y}_m^{(g)}$ denote the object vector and $\varsigma_m^{(g)}$ the mutation strength of the mth parent in generation g. Based on a μ elemental parent population $(\mathbf{y}_m, \varsigma_m)_{m=1,\ldots,\mu}^{(g)}$, λ offspring are created by variation:

1. First, the mutation strength is varied. In this paper, intermediate ES are considered which also apply arithmetic recombination in the case of the mutation strengths and compute the mean $\langle \varsigma^{(g)} \rangle$ of the μ parental mutation strengths $\varsigma_m^{(g)}$. The mutation strength is then mutated which is realized by a multiplication with a random variable ϵ

$$\varsigma_l \quad = \quad \langle \varsigma^{(g)} \rangle \epsilon. \tag{4}$$

   A common choice for ϵ is a log-normally distributed random variable (see e.g. [13]). In this case, the random variable obeys $\epsilon \sim \exp(\tau \mathcal{N}(0, 1))$. The parameter τ is called *learning parameter*.

2. The object parameters are changed afterwards. First, recombination takes place and the centroid $\langle \mathbf{y}^{(g)} \rangle$ of the μ parental object vectors $\mathbf{y}_m^{(g)}$ is computed. The object parameters are then mutated by adding a normally distributed random vector with zero mean and the newly created mutation strength ς_l

$$\mathbf{y}_l \quad = \quad \langle \mathbf{y}^{(g)} \rangle + \varsigma_l \mathbf{N}(0,1). \tag{5}$$

Finally, based on the apparent fitness $\tilde{F}$, (3), the μ best offspring are chosen for the succeeding parental population.

1.3 Modeling the Evolution Strategy: The ES as a Dynamical System

Due to the form of the undisturbed fitness function $f(x, \mathbf{R}) = x - dR$, three variables are of interest: the x-component denoting the change parallel to the ridge axis, the lateral component $\mathbf{R}$ or its length R measuring the $(N-1)$-dimensional distance to the ridge, and the mutation strength ς. Therefore, the ES's time-dynamical behavior can be characterized by three state variables.

In the following, we follow the approach first introduced by Beyer in [8] and consider the ES as a dynamical system. This approach aims at modeling the change of a state variable $u^{(g)}$ from one generation to the next. This change can be divided into an expected or deterministic change and a random change ϵ_u

$$u^{(g+1)} \quad = \quad u^{(g)} + \mathrm{E}[u^{(g+1)} - u^{(g)}] + \epsilon_u(u^{(g)}). \tag{6}$$

Note that (6) neglects dependencies of u on the present state of other variables. To keep the notations short, these dependencies are only explicitely denoted when needed. Let us first address the expected changes. In the case of the state variables in the object parameters space, these are termed *progress rates* and denoted as $\varphi_x := \mathrm{E}[x^{(g+1)} - x^{(g)}]$ and $\varphi_R := \mathrm{E}[R^{(g)} - R^{(g+1)}]$. In the case of the mutation strength, the so-called *self-adaptation response function* (SAR) ψ must be obtained. Note that due to the multiplicative change, the change of the mutation strength in (4) is modeled by

$$\begin{aligned} \varsigma^{(g+1)} \quad &= \quad \varsigma^{(g)} \mathrm{E}\Big[\frac{\varsigma^{(g+1)} - \varsigma^{(g)}}{\varsigma^{(g)}}\Big] + \varsigma^{(g)} \epsilon_\sigma \\ &=: \quad \varsigma^{(g)} \psi(\varsigma^{(g)}) + \varsigma^{(g)} \epsilon_\sigma \end{aligned} \tag{7}$$

and ψ denotes thus the expected relative change of the random variable $\psi := \mathrm{E}[(\varsigma^{(g+1)} - \varsigma^{(g)})/\varsigma^{(g)}]$. In a first approach,

the random changes in (6) are neglected leading to the so-called *deterministic evolution equations*. The quality of the results obtained in this manner must be evaluated by experiments. While the approach is strongly simplifying the stochastic system's behavior, it will turn out that it serves quite well to extract the main characteristics of the ES's behavior in the stationary state.

As mentioned before, the deterministic evolution equations are used to investigate the change of the state variables x, R, and ς over time leading to

$$x^{(g+1)} = x^{(g)} + \varphi_x(\langle\varsigma^{(g)}\rangle, \sigma_\epsilon) \quad (8)$$

$$R^{(g+1)} = R^{(g)} - \varphi_R(R^{(g)}, \langle\varsigma^{(g)}\rangle, \sigma_\epsilon) \quad (9)$$

$$\langle\varsigma^{(g+1)}\rangle = \langle\varsigma^{(g)}\rangle\Big(1 + \psi(R^{(g)}, \langle\varsigma^{(g)}\rangle, \sigma_\epsilon)\Big). \quad (10)$$

In the following, the notations in (8) - (10) are simplified. First, we shorten the notations setting $R := R^{(g)}$, $r := R^{(g+1)}$, and $\sigma := \langle\varsigma^{(g)}\rangle$. Afterwards, similar normalization as in the sphere model case are introduced: $\varphi_x^* := N\varphi_x/R$, $\varphi_R^* := N\varphi_R/R$, $\sigma^* := N\sigma/R$, and $\sigma_\epsilon^* := N\sigma_\epsilon/R$. In the following, the main characteristics of the normalized progress measures are shortly discussed.

The progress rate $\varphi_x^* = N\mathrm{E}[x^{(g+1)} - x^{(g)}]/R$ was obtained for $\tau = 0$ and $N \to \infty$ (see e.g. [3, 6] for a derivation) as

$$\varphi_x^*(\sigma^*) = \frac{\sigma^{*2}}{\sqrt{(1+d^2)\sigma^{*2} + \sigma_\epsilon^{*2}}} c_{\mu/\mu,\lambda}. \quad (11)$$

Note, there is no influence of the value of x on its own expected change.

The progress rate $\varphi_R^* = N\mathrm{E}[R^{(g)} - R^{(g+1)}]/R^{(g)}$, i.e.,

$$\varphi_R^*(\sigma^*) = \frac{d\sigma^{*2}}{\sqrt{(1+d^2)\sigma^{*2} + \sigma_\epsilon^{*2}}} c_{\mu/\mu,\lambda} - \frac{\sigma^{*2}}{2\mu} \quad (12)$$

consists of a gain and a quadratic loss part and can be interpreted as a function of the mutation strength. Again, (12) is determined for $N \to \infty$ and $\tau = 0$. The ridge parameter d and noise strength σ_ϵ^* influence the gain part whereas the loss part is only influenced by the parent number μ.

The SAR $\psi = \mathrm{E}[(\langle\varsigma^{*(g+1)}\rangle - \langle\varsigma^{*(g)}\rangle)/\langle\varsigma^{*(g)}\rangle]$,

$$\psi(\sigma^*) = \tau^2\Bigg(1/2 + e_{\mu,\lambda}^{1,1}\frac{(1+d^2)\sigma^{*2}}{(1+d^2)\sigma^{*2} + \sigma_\epsilon^{*2}} - c_{\mu/\mu,\lambda}\frac{d\sigma^{*2}}{\sqrt{(1+d^2)\sigma^{*2} + \sigma_\epsilon^{*2}}}\Bigg) \quad (13)$$

was determined under the assumption $\tau \ll 1$ and for $N \to \infty$. In the appendix, a short sketch of the derivation of the SAR is provided. It should be noted that the prediction quality of (13) deteriorates relatively fast with increasing σ^* for smaller values of N. The progress coefficients $e_{\mu,\lambda}^{1,1}$ and $c_{\mu/\mu,\lambda}$ in (11) to (13) are given by

$$e_{\mu,\lambda}^{\alpha,\beta} = \frac{\lambda-\mu}{\sqrt{2\pi}^{\alpha+1}}\binom{\lambda}{\mu} \times \int_{-\infty}^{\infty} t^\beta \mathrm{e}^{-\frac{\alpha+1}{2}t^2}\Phi(t)^{\lambda-\mu-1}\Big(1-\Phi(t)\Big)^{\mu-\alpha}\mathrm{d}t \quad (14)$$

with $c_{\mu/\mu,\lambda} := e_{\mu,\lambda}^{1,0}$ (see [10, p. 172]).

Since the required functions are given, the analysis can be started. First note that the evolution of the x-component does not influence the evolution of R and σ^*. Instead, the evolution parallel to the axis is governed by the evolution of the remaining two state variables. Therefore, it suffices to consider the system in R and σ^*. Note, $\langle\varsigma^{*(g+1)}\rangle := N\langle\varsigma^{*(g+1)}\rangle/r$ with $r = R(1 - \varphi_R^*/N)$. Using (9) and (10), the evolution equations read therefore

$$\begin{pmatrix} r \\ \langle\varsigma^{*(g+1)}\rangle \end{pmatrix} = \begin{pmatrix} R\Big(1 - \varphi_R^*(\sigma^*, \sigma_\epsilon^*)/N\Big) \\ \sigma^*\Big(\frac{1+\psi(\sigma^*,\sigma_\epsilon^*)}{1-\varphi_R^*(\sigma^*,\sigma_\epsilon^*)/N}\Big) \end{pmatrix}. \quad (15)$$

In (15), a third g-dependent variable appears: The normalized noise strength $\sigma_\epsilon^{*(g)}$ which changes with $R^{(g)}$. However, the (direct) influence of $R^{(g)}$ can be eliminated leading to the new evolution equation

$$\sigma_\epsilon^{*(g+1)} = \frac{\sigma_\epsilon^*}{1 - \varphi_R^*(\sigma^*, \sigma_\epsilon^*)/N}. \quad (16)$$

Due to the normalization, the evolution of R neither influences the evolution of the mutation strength nor the evolution of the noise strength. Its evolution can be decoupled and it suffices to analyze the two-dimensional evolution equations

$$\begin{pmatrix} \sigma_\epsilon^{*(g+1)} \\ \langle\varsigma^{*(g+1)}\rangle \end{pmatrix} = \begin{pmatrix} \frac{\sigma_\epsilon^*}{1-\varphi_R^*(\sigma^*,\sigma_\epsilon^*)/N} \\ \sigma^*\Big(\frac{1+\psi(\sigma^*,\sigma_\epsilon^*)}{1-\varphi_R^*(\sigma^*,\sigma_\epsilon^*)/N}\Big) \end{pmatrix}. \quad (17)$$

For the remainder of this paper, the evolution equations (17) are used.

2. A STATIONARY STATE

Evolution strategies show two kind of behaviors on the undisturbed sharp ridge. Depending on the size of the d-parameter, they either converge towards the axis or they diverge from the axis. The former situation is usually connected with a stagnation of the evolutionary process since the ES reduces the mutation strength in the approach of the axis. Noise changes the situation: If the fitness evaluations are overlaid with additive noise with constant noise strength, the ES cannot converge completely to the axis but fluctuates eventually at a distance to the axis. The evolution equations (17) can be used to determine the stationary state.

2.1 Determining Stationary States

Stationary points of (17) givem as $(\sigma_\epsilon^{*(g+1)}, \langle\varsigma^{*(g+1)}\rangle)^{\mathrm{T}} = (\sigma_\epsilon^*, \sigma^*)^{\mathrm{T}}$ can be determined in a straightforward way. The stationary solution of the evolution equation for σ_ϵ^* in (17) requires the progress rate (12) to be zero which results in two stationarity conditions

$$\varsigma_{stat_1}^* = 0 \bigvee \sqrt{(1+d^2){\varsigma_{stat_2}^*}^2 + \sigma_\epsilon^{*2}} = 2\mu c_{\mu/\mu,\lambda} d. \quad (18)$$

Therefore, the task is to find solutions of (18) which are also stationary points for the evolution of the mutation strength in (17). It can be shown that a stationary state of the system (17) is given by

$$\begin{pmatrix} \sigma_{\epsilon\,stat_1}^* \\ \varsigma_{stat_1}^* \end{pmatrix} = \begin{pmatrix} c \\ 0 \end{pmatrix} \quad (19)$$

with $c \in \mathbb{R}, c \geq 0$ or by

$$\begin{pmatrix} \sigma^*_{\epsilon\, stat} \\ \varsigma^*_{stat} \end{pmatrix} = \begin{pmatrix} 2d\mu c_{\mu/\mu,\lambda} \sqrt{\frac{d^2(4\mu c^2_{\mu/\mu,\lambda} - 2e^{1,1}_{\mu,\lambda} - 1) - 2e^{1,1}_{\mu,\lambda} - 1}{d^2(4\mu c^2_{\mu/\mu,\lambda} - 2e^{1,1}_{\mu,\lambda}) - 2e^{1,1}_{\mu,\lambda}}} \\ \frac{2d\mu c_{\mu/\mu,\lambda}}{\sqrt{d^2(4\mu c^2_{\mu/\mu,\lambda} - 2e^{1,1}_{\mu,\lambda}) - 2e^{1,1}_{\mu,\lambda}}} \end{pmatrix}. \quad (20)$$

Note, the stationary mutation strength ς^*_{stat} goes to zero for $d \to 0$ and to $2\mu c_{\mu/\mu,\lambda}/\sqrt{4\mu c^2_{\mu/\mu,\lambda} - 2e^{1,1}_{\mu,\lambda}}$ for $d \to \infty$. The normalized noise strength behaves in accordance to d: For $d \to 0$, $\sigma^*_{\epsilon\, stat}(d) \to 0$ and $\sigma^*_{\epsilon\, stat}(d) \to \infty$ for $d \to \infty$. Both variables are completely determined by the population parameter μ and λ and of course by the ridge parameter d.

The noise effectively stops the ES from approaching the ridge axis arbitrarily close. Using (20),

$$R_{stat} = \frac{N\sigma_\epsilon}{2d\mu c_{\mu/\mu,\lambda}} \times \sqrt{\frac{d^2(4\mu c^2_{\mu/\mu,\lambda} - 2e^{1,1}_{\mu,\lambda}) - 2e^{1,1}_{\mu,\lambda}}{d^2(4\mu c^2_{\mu/\mu,\lambda} - 2e^{1,1}_{\mu,\lambda} - 1) - 2e^{1,1}_{\mu,\lambda} - 1}} \quad (21)$$

is derived as the stationary distance to the axis.

As (19) and (20) show, system (17) comes to a halt either by a loss of step-size control in an arbitrary distance to the axis (19) or by attaining stationary values for the mutation strength and the distance.

The question remains under which conditions the latter stationary state exists. A closer look at (20) reveals that the weighting constant d is decisive w.r.t. μ and λ. Let $\mu \leq \lambda/2$. The stationary state is only defined for constants d which fulfill $d \geq d_{crit} = [(2e^{1,1}_{\mu,\lambda} + 1)/(4\mu c^2_{\mu/\mu,\lambda} - 2e^{1,1}_{\mu,\lambda} - 1)]^{(1/2)}$. This critical d-value has been encountered in the case of the undisturbed ridge [18]. There it decided about convergence or divergence of the ES to the axis. The reason for the reappearance is that only for $d > d_{crit}$, the ES moves towards the axis at all. In the case of $d < d_{crit}$, the distance to the axis and the mutation strength enlarge. Since the noise strength remains constant, it gradually looses its influence until the ES behaves as if it were optimizing the noise-free ridge. The constraint $\mu \leq \lambda/2$ is sufficient but not necessary. The equations generally hold unless $\mu \approx \lambda$ but a sharp boundary cannot be given.

Let us sum up our findings. Let us assume that the ES starts far away from the ridge axis. For $d > d_{crit}$, the ES moves towards the ridge axis as in the undisturbed case. Contrary to its behavior in the noise-free case, however, it converges to a stationary state that has a well-defined distance to the axis. The evolution of R comes to a halt on average and the ES travels parallel to the axis direction.

2.2 On the Beneficial Influence of Noise

Now, we are at the position to discuss the influence of noise on the performance. To this end, the stationary progress rate parallel to the axis is determined using (20) and the second condition in (18)

$$\varphi^*_{x\, stat} = \frac{{\varsigma^*_{stat}}^2}{2\mu d}. \quad (22)$$

As the normalized stationary mutation strength, the normalized stationary progress rate depends only on the stationary mutation strength in (20), the constant d, and on the population parameters.

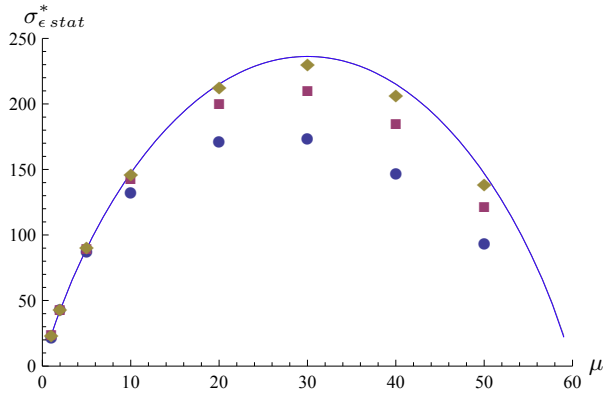

Figure 1: The stationary noise strength (20) and experimental results as functions of μ.

The non-normalized progress parallel to the ridge axis is obtained by plugging (20) and (21) into (22) as

$$\varphi^{st}_x(\sigma_\epsilon) = \sigma_\epsilon c_{\mu/\mu,\lambda} \sqrt{\frac{1}{d^2(4\mu c^2_{\mu/\mu,\lambda} - 2e^{1,1}_{\mu,\lambda} - 1) - 2e^{1,1}_{\mu,\lambda} - 1}} \times \sqrt{\frac{1}{d^2(4\mu c^2_{\mu/\mu,\lambda} - 2e^{1,1}_{\mu,\lambda}) - 2e^{1,1}_{\mu,\lambda}}} \quad (23)$$

The non-normalized progress rate increases linearly with the noise strength – a behavior that also holds for the non-normalized mutation strength

$$\varsigma_{stat} = \frac{\sigma_\epsilon}{\sqrt{d^2(4\mu c^2_{\mu/\mu,\lambda} - 2e^{1,1}_{\mu,\lambda} - 1) - (2e^{1,1}_{\mu,\lambda} + 1)}}. \quad (24)$$

As (23) and (24) reveal, noise improves the performance of the ES. The positive influence of the noise can be traced back to the general behavior of the ES on the sharp ridge. If the ES is far away from the ridge axis the influence of the noise is relatively small in comparison to that of R. Provided $d > d_{crit}$, the ES starts approaching the axis but is hindered in the convergence by the noise. Higher noise strengths result in larger distances to the axis and in larger stationary mutation strengths. Larger mutation strengths are connected with higher expected gains on the ridge axis. On the sharp ridge, therefore, noise with a constant standard deviation effectively stops the ES from optimizing the contained sphere model, i.e., the loss part of the ridge function, and enforces a more significant gain parallel to the axis. This only holds for sufficiently large ridge parameters d.

If d is too small, the ridge is not being tracked and a divergence of the distance occurs. The distance R increases which lessens the (relative) influence of the noise. In this case, the ES will gradually start to behave as if it were optimizing the undisturbed sharp ridge – striding away from the axis with a negative progress rate φ_R – but with an overall positive quality change, i.e., the gain parallel to the axis surpasses the loss due to the distances' increase. In short, noise either soon looses its influence or has an actual positive influence as it keeps the ES from optimizing only the sphere part.

2.3 Comparison with Experiments

Figures 1 - 3 show a comparison between the normalized stationary values (20) and (22) and experimental data for

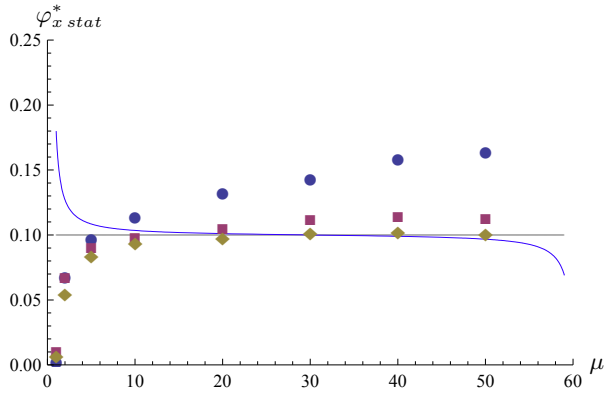

Figure 2: The stationary progress rate (22), the approximation (27) (thin grey line), and experimental results as functions of μ.

three search space dimensionalities $N = 30$, $N = 100$, and $N = 500$. The results for $N = 30$ are denoted by dots, those for $N = 100$ by squares, whereas the results for $N = 500$ are given by diamonds. The constant d was set to $d = 5$. The experimental results were averaged over several runs with different choices of σ_ϵ with $\sigma_\epsilon/N = 1, 2, 3$, and 5. In the case of $N = 30$ each data point was sampled over $100,000$ generations for each noise strength and then averaged over all noise strengths, i.e., over a total of $4 \times 100,000$ generations. For $N = 100$ and $N = 500$ $4 \times 200,000$ generations were used. The prediction quality improves with the search

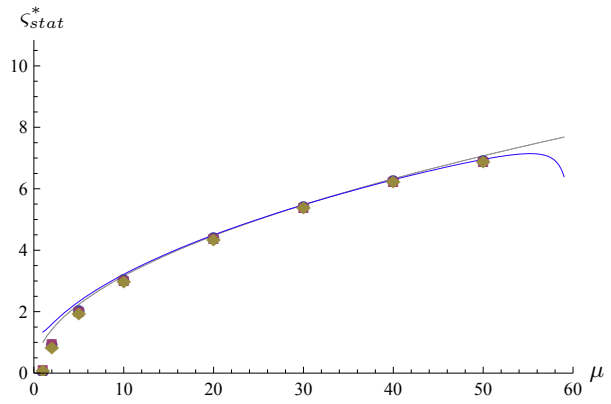

Figure 3: The stationary mutation strength (20), the approximation (25) (thin grey line), and experimental results as functions of μ.

space dimensionality with the exception of the stationary mutation strength in (20), Fig. 3. In this case, the agreement is good even for the lower dimensional search space $N = 30$ and does not improve visibly if N increases. It can be seen though that (20) tends to overestimate the stationary mutation strength if the parent number is relatively small. This probably causes in turn the greater deviations of (22) from the experimental progress rates for these μ values (see Fig. 2). While the agreement of (22) with the experiments is quite good for large N in general, the experimental results for $\mu = 1$ and $\mu = 2$ are far lower than predicted. Figures 4 compares the non-normalized values (21), (23), and (24) with the results of experiments. The results for $N = 30$ are denoted by dots, those for $N = 100$ by squares, whereas the results for $N = 500$ are given by diamonds. The search space dimensionalities were $N = 30$, $N = 100$,

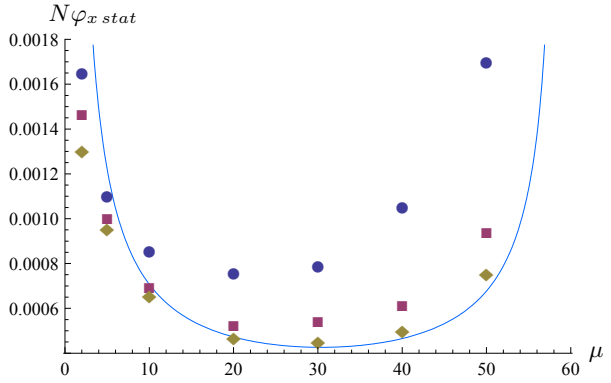

Figure 4: The stationary progress rate (23) and experimental results for constant noise $\sigma_\epsilon = 1/N$ as a function of μ.

and $N = 500$. In the case of $N = 30$ each data point was sampled over $100,000$ generations, whereas for $N = 100$ and $N = 500$ $200,000$ generations were used. Again, the prediction quality is relatively poor for $N = 30$ and improves with the search space dimensionality.

2.4 The Effects of Recombination

The effects of recombination remain to be addressed. Figure 4 shows the stationary progress rate as a function of the parent number μ. As it reveals, switching from $\mu = 1$ to $\mu > 1$ is not beneficial. To find out why, let us start with the normalized mutation and noise strength (20). Provided that the size of the offspring population is not small, (20) shows an interesting scaling behavior with respect to μ. If $2\mu c^2_{\mu/\mu,\lambda} \gg e^{1,1}_{\mu,\lambda}$ holds, the stationary point can be approximated with

$$\begin{pmatrix} \sigma^*_{\epsilon\ appr} \\ \varsigma^*_{appr} \end{pmatrix} = \begin{pmatrix} 2d\mu c_{\mu/\mu,\lambda} \\ \sqrt{\mu} \end{pmatrix}. \tag{25}$$

Equation (25) holds for $\mu \not\approx 1$ and $\mu \not\approx \lambda$ and large λ. Interestingly, it equals the scaling behavior on the noisy sphere with only one exception, the ridge parameter d, which appears in the case of the noise strength. Apparently, the ES behaves very similarly on the noise sharp ridge as on the noisy sphere. Recombination increases the mutation strength in (25) and (23) proportionally to $\sqrt{\mu}$. An increase also occurs in the case of the normalized noise strength which increases with $2\mu c_{\mu/\mu,\lambda}$. This results in smaller distances to the ridge axis. With similar arguments as before, the scaling behavior of the distance to the ridge w.r.t. μ reads

$$R_{appr} = \frac{N\sigma_\epsilon}{2d\mu c_{\mu/\mu,\lambda}}. \tag{26}$$

The approximate distance (26) is also the minimal possible distance that can be obtained. This can be verified by using the second stationary condition in (18), $(1+d^2){\varsigma^*_{stat_2}}^2 + {\sigma^*_\epsilon}^2 = 4d^2\mu^2 c^2_{\mu/\mu,\lambda}$ and letting $\varsigma^*_{stat_2} \to 0$.

While the decrease of the distance was beneficial on the sphere, it has the opposite effect on the ridge. The normalized stationary progress (22) reveals the problem: If $\mu \not\approx 1$

and $\mu \not\approx \lambda$, the influence of μ on the normalized progress rate is negligible. Because of $\varsigma^*_{stat} \propto \sqrt{\mu}$, the normalized progress rate does not differ much from

$$\varphi^*_{x\,appr} = \frac{1}{2d} \tag{27}$$

provided that λ is not small.

Since the normalized progress rate stays nearly constant, a problem is encountered in the case of the non-normalized variables. The non-normalized progress rate (23) scales approximately with $1/(2\mu c_{\mu/\mu,\lambda})$ and drops sharply if recombination is introduced. As we have seen, the increase of the normalized mutation strength with $\sqrt{\mu}$ is not sufficient to change the normalized progress rate. The decrease of the progress rate with the decreasing distance is not counteracted in any way.

Something similar can be observed in the case of the non-normalized mutation strength, itself. The normalized noise scales with $2\mu c_{\mu/\mu,\lambda}$ and the distance with $1/(2\mu c_{\mu/\mu,\lambda})$. This outperforms the increase of the normalized mutation strength with $\sqrt{\mu}$: The non-normalized mutation strength decreases with $1/(2\sqrt{\mu} c_{\mu/\mu,\lambda})$. This decreasing is necessary on the sphere. Since the ES is able to approach the optimizer more closely, the mutation strength must reflect this and decrease accordingly. On the ridge, however, this means that the mutation strength is decreased because the subgoal of optimizing the sphere is better realized. This does not result in a better achievement of the overall goal.

On first sight, recombination does not have any benefits. However, the $(1, \lambda)$-ES looses step-size control in the stationary state and reduces the mutation strength to extremely small values – a behavior not predictable by the deterministic evolution equations. Therefore, as a rule the $(1, \lambda)$-ES exhibits long stagnation phases. Recombination with a small number of parents is therefore necessary. This is a further similarity to the noisy sphere [12]. There, it could be shown that the $(1, \lambda)$-ES performed basically a random walk – biased, however, towards smaller mutation strengths. However, the exact causes for the behavior on the sharp ridge remain to be investigated.

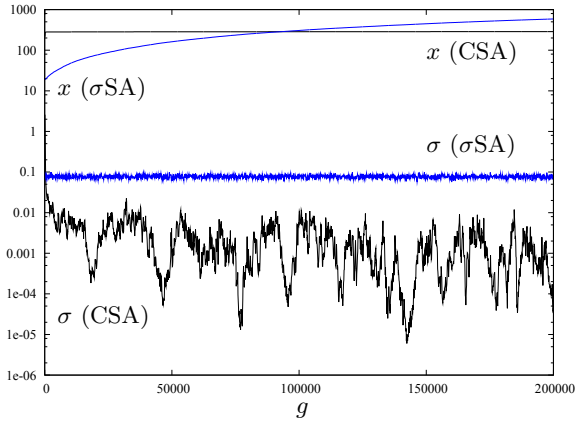

Figure 5: Comparison of a $(10, 10_I, 60)$-CSA-ES and self-adaptive $(10, 10_I, 60)$-ES on the noisy sharp ridge with $d = 5$, $\sigma_\epsilon = 100/30$, and $N = 30$. Shown are the the evolutions of the mutation strength and the x-component (every 100th value) for the first $200,000$ generations. The results were averaged over 10 runs.

3. CONCLUSIONS

In this paper, noisy ridge functions were investigated using the standard noise model of additive normally distributed noise. It has been shown using the evolution equations, that noise has a positive effect and improves the performance. Unlike the findings in [7, 16, 20], this does not only hold for small noise levels but even more for higher levels.

Optimizing ridge functions comprises two subgoals: a) minimizing the $(N-1)$-dimensional sphere, thus, reaching the ridge axis. b) travelling along (i.e., parallel to) the ridge axis. If d is sufficiently large, subgoal a) dominates the evolution process. As a result the mutation strength is permanently reduced and the fitness gain w.r.t. subgoal b) reduces, too. Additive noise prevents the ES from perfectly reaching subgoal a). The ES stays away from the ridge axis and a steady state distance as well as a steady state mutation strength is obtained on average. Therefore, no premature convergence occurs. Instead, due to the steady state mutation strength evolution strategies show on average a constant progress parallel to the axis direction. Of course, this only holds if d is sufficiently large so that the axis acts as an attractor. If d is too small and the ES diverges, the effects of the noise are soon diluted until it behaves as if it were optimizing the undisturbed ridge. If d is sufficiently large, a stationary state of the distance and the mutation strength exists. Here, we find that the larger the noise strength, the larger the stationary distance and the mutation strength. This results in larger progress parallel to the axis direction. Additive noise is beneficial on the sharp ridge: Because of the noise the finite subgoal of optimizing the sphere cannot be realized.

Recombination has a similar effect as in the case of the sphere model. It reduces the distance to the axis. This decrease with μ is stronger than the increase of the normalized stationary mutation strength (w.r.t. the distance and search space dimensionality). These responses eventually cause a performance degradation: The normalized progress stays constant and the non-normalized progress rate decreases. However, recombination on the ridge is necessary since the $(1, \lambda)$-ES looses step-size control.

Having analyzed the ES on the noisy sharp ridge, we have found a first example where it can be shown analytically that noise can help to improve the performance of the ES. Due to the deeper understanding of the underlying evolutionary process, we can now envision the conditions a fitness landscape should obey in order to be a candidate for noise-improved EA performance: The noise prevents the ES from being attracted by local optimal basins and the mutation strength is kept at a reasonable level. If there is in addition a global fitness landscape topology that has a global tendency to better attractors, then the ES can have the chance to evolve to more promising regions of the fitness landscape. The depth of the attractors that can be passed by can be controlled by the noise strength used. Actually, assuming a simple local landscape structure, the noise strength needed to avoid a local attractor of predefined depth can already be estimated [11]. Therefore, it should be possible to search for attractors of greater depth, thus, having a strategy for more global search. However, two caveats must be mentioned here. First, this idea can only work if there is an appropriate global fitness topology. Second, at the end of the evolutionary process, the user must reduce the noise level in order to get close to the local optimizer which is

then hopefully the global one. In other words, a guarantee for reaching the global optimizer cannot be given.

The analysis presented here can and should be extended in several points: First of all, the progress measures obtained for the evolution equations hold exactly only for $N \to \infty$. All results obtained using these progress measures hold only approximately in low-dimensional search spaces. Therefore, one aim should be to use progress measures obtained for finite N. In addition, an inclusion of the perturbation parts of the evolution equations would be interesting. Furthermore, a comparison with other adaptation schemes as the CSA or the 1/5th rule is of interest. Figure 5, for example, shows a comparison of a CSA-ES to a self-adaptive ES. In the case of the CSA, a problem occurs: The CSA-ES shows a loss of step-size control on the sharp ridge. This erratic behavior of the mutation strength is also observed for other choices of μ. This has consequences for the optimization of the x-component. While the CSA-ES surpasses the self-adaptive ES in the early stages of the evolutionary run, the progress stagnates in the steady state whereas the self-adaptive ES progresses still. The speed of the self-adaptive ES depends on the size of the steady state mutation strength and therefore on the noise strength. The higher the noise the sooner the self-adaptive ES surpasses the CSA-ES. Finding a predictive model for these observations remains a task for future research.

4. REFERENCES

[1] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja. *A First Course in Order Statistics*. Wiley, New York, 1992.

[2] D. V. Arnold. *Noisy Optimization with Evolution Strategies*. Kluwer Academic Publishers, Dordrecht, 2002.

[3] D. V. Arnold. Cumulative step length adaptation on ridge functions. In T. Runarsson et al., editors, *Parallel Problem Solving from Nature – PPSN IX*, pages 11–20. Springer Verlag Heidelberg, 2006.

[4] D. V. Arnold and H.-G. Beyer. Local performance of the $(\mu/\mu_I, \lambda)$-ES in a noisy environment. In W. Martin and W. Spears, editors, *Foundations of Genetic Algorithms, 6*, pages 127–141, San Francisco, CA, 2001. Morgan Kaufmann.

[5] D. V. Arnold and H.-G. Beyer. Evolution strategies with cumulative step length adaptation on the noisy parabolic ridge. *Natural Computing*, 2007. To appear.

[6] D. V. Arnold and A. MacLeod. Step length adaptation on ridge functions. Technical Report CS-2006-08, Dalhousie University, Faculty of Computer Science, August 2006.

[7] T. Bäck and U. Hammel. Evolution strategies applied to perturbed objective functions. In Z. Michalewicz et al., editors, *Proc. First IEEE Conf. Evolutionary Computation*, pages 40–45. IEEE Press, Piscataway NJ, 1994.

[8] H.-G. Beyer. Toward a theory of evolution strategies: Self-adaptation. *Evolutionary Computation*, 3(3):311–347, 1996.

[9] H.-G. Beyer. On the performance of $(1, \lambda)$-evolution strategies for the ridge function class. *IEEE Transactions on Evolutionary Computation*, 5(3):218–235, 2001.

[10] H.-G. Beyer. *The Theory of Evolution Strategies*. Natural Computing Series. Springer, Heidelberg, 2001.

[11] H.-G. Beyer, D. V. Arnold, and S. Meyer-Nieberg. A new approach for predicting the final outcome of evolution strategy optimization under noise. *Genetic Programming and Evolvable Machines*, 6(1):7–24, 2005.

[12] H.-G. Beyer and S. Meyer-Nieberg. Self-adaptation of evolution strategies under noisy fitness evaluations. *Genetic Programming and Evolvable Machines*, 7(4):295–328, 2006.

[13] H.-G. Beyer and H.-P. Schwefel. Evolution strategies: A comprehensive introduction. *Natural Computing*, 1(1):3–52, 2002.

[14] M. Herdy. Reproductive isolation as strategy parameter in hierachically organized evolution strategies. In R. Männer and B. Manderick, editors, *Parallel Problem Solving From Nature, PPSN II,*, Berlin, 1992. Springer-Verlag.

[15] Y. Jin and J. Branke. Evolutionary optimization in uncertain environments – a survey. *IEEE Transactions on Evolutionary Computation*, 9(3):303–317, 2005.

[16] B. Levitan and S. Kauffman. Adaptive walks with noisy fitness measurements. *Molecular Diversity*, 1(1):53–68, 1995.

[17] M. Lunacek and D. Whitley. Searching for balance: Understanding self-adaptation on ridge functions. In T. Runarsson et al., editors, *Parallel Problem Solving from Nature – PPSN IX*, pages 82–91. Springer Verlag Heidelberg, 2006.

[18] S. Meyer-Nieberg and H.-G. Beyer Mutative self-adaptation on the sharp and parabolic ridge. In C. Stephens et al., editors, *Foundations of Genetic Algorithms IX (FOGA 2007)*, pages 70–96, Heidelberg, 2007. Springer Verlag.

[19] A. Ostermeier, A. Gawelczyk, and N. Hansen. A derandomized approach to self-adaptation of evolution strategies. *Evolutionary Computation*, 2(4):369–380, 1995.

[20] S. Rana, L. D. Whitley, and R. Cogswell. Searching in the presence of noise. In H.-M. Voigt et al., editors, *Parallel Problem Solving from Nature, 4*, pages 198–207, Heidelberg, 1996. Springer.

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

[22] H.-P. Schwefel. *Evolution and Optimum Seeking*. Wiley, New York, NY, 1995.

[23] L. Whitley, M. Lunacek, and J. Knight. Ruffled by ridges: How evolutionary algorithms can fail. In K. Deb et al., editors, *Genetic and Evolutionary Computation–GECCO 2004, part II*, pages 294–306. Springer, Heidelberg, 2004.

APPENDIX

A. ON THE DETERMINATION OF THE SELF-ADAPTATION RESPONSE

The section sketches the derivation of the self-adaptation response function (SAR) which denotes the expected relative

change of the mutation strength

$$\psi(\sigma) \;=\; \mathrm{E}\Big[\frac{\langle\varsigma\rangle-\sigma}{\sigma}\Big] = \frac{1}{\mu}\sum_{m=1}^{\lambda}\mathrm{E}\Big[\frac{\varsigma_{m;\lambda}-\sigma}{\sigma}\Big]. \quad (28)$$

The random variable $\varsigma_{m;\lambda}$ is the mutation strength associated with the mth best perceived fitness change in λ trials. Using the concept of induced order statistics [1, 2], the pdf of the random variable is given by

$$p_{m;\lambda}(\varsigma|\sigma) \;=\; \frac{\lambda!p_\sigma(\varsigma|\sigma)}{(m-1)!(\lambda-m)!}\int_{-\infty}^{\infty}p_Q(Q|\varsigma) \times \mathrm{P}_Q(Q|\sigma)^{\lambda-m}\Big(1-\mathrm{P}_Q(Q|\sigma)\Big)^{m-1}\mathrm{d}Q. (29)$$

The derivation of the SAR can be divided into two main parts: First, an expression for the fitness change induced by a mutation has to be determined. Afterwards, the resulting equations have to be transformed and simplified. Let us first address the fitness change. In the case of the undisturbed sharp ridge, it has been obtained by Beyer in [9]. A straightforward adaptation for noise leads to $P_Q(Q|\varsigma) = \Phi\Big(\frac{Q+d\sqrt{R^2+\varsigma^2N}-R}{\sqrt{\sigma_\epsilon^2+\varsigma^2(1+d^2\frac{R^2+\varsigma^2N/2}{R^2+\varsigma^2N})}}\Big)$. In the following, we use a more general notation setting $P_Q(Q|\varsigma) = \Phi([Q+h(\varsigma)]/g(\varsigma))$.

The first point in the determination of the SAR is to eliminate the integral expression $\mathrm{P}_Q(Q|\sigma) = \int_0^\infty P_Q(Q|\varsigma)p_\sigma(\varsigma|\sigma)\,\mathrm{d}\varsigma$ in (29). As shown in [10], substituting $\mathrm{P}_Q(Q|\sigma)$ with $P_Q(Q|\sigma)$ introduces an error of order τ^2. Since we consider the case of $\tau \ll 1$, the error may be neglected. First, the argument of the distribution function is standardized – introducing $z = -[Q+h(\sigma)]/g(\sigma)$. The SAR changes to

$$\psi(\sigma) \;=\; \int_0^\infty\Big(\frac{\varsigma-\sigma}{\sigma}\Big)\frac{p_\sigma(\varsigma|\sigma)}{\mu}\sum_{m=1}^{\mu}\frac{\lambda!}{(m-1)!(\lambda-m)!} \times\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{g(\sigma)}{g(\varsigma)}\mathrm{e}^{-\frac{1}{2}\left(\frac{g(\sigma)z-(h(\varsigma)-h(\sigma))}{g(\varsigma)}\right)^2} \times\Big(1-\Phi(z)\Big)^{\lambda-m}\Phi(z)^{m-1}\,\mathrm{d}z\,\mathrm{d}\varsigma. \quad (30)$$

In the following, several transformations and calculations have to be performed. In the next step, the order of the summation and the inner integration is swapped. The sum in (30) itself represents a regularized incomplete beta function [10, p. 147f] and can be substituted by an integral. The integral is then plugged into the SAR and the integration order of the inner integrals is changed which results in

$$\psi(\sigma) \;=\; \int_0^\infty\frac{g(\sigma)}{g(\varsigma)}\Big(\frac{\varsigma-\sigma}{\sigma}\Big)p_\sigma(\varsigma|\sigma)(\lambda-\mu)\binom{\lambda}{\mu} \times\int_{-\infty}^{\infty}\Big(1-\Phi(t)\Big)^{\lambda-\mu-1}\Phi(t)^{\mu-1}\frac{\mathrm{e}^{-\frac{t^2}{2}}}{2\pi} \times\int_{-\infty}^{t}\mathrm{e}^{-\frac{1}{2}\left(\frac{g(\sigma)z-(h(\varsigma)-h(\sigma))}{g(\varsigma)}\right)^2}\,\mathrm{d}z\,\mathrm{d}t\,\mathrm{d}\varsigma. \quad (31)$$

This is the point to introduce further simplifications in order to solve the integrals. The starting point is the innermost integral over z which leads the distribution function $\Phi(((g(\sigma)/g(\varsigma))t-(h(\varsigma)-h(\sigma))/g(\sigma))$. This function is then expanded into its Taylor series around σ. Plugging the resulting series into the SAR, three integrals are obtained: one containing the normal distribution function at σ, one comprising the first derivation and a quadratic $(\varsigma-\sigma)$-term, and one with higher derivations and polynomials in $(\varsigma-\sigma)$ with degree three or higher. First of all, the integration over ς is addressed. The expectation of $(\varsigma-\sigma)^k$ leads to a series in τ^{2l}. It can be shown that the expectation of $[(\varsigma-\sigma)/\sigma]^k$ does not include any τ^{2l}-terms with $2l+1<k$. In this paper, the series is expanded to the precision of τ^2, thus the expectations of terms $[(\varsigma-\sigma)/\sigma]^k$, $k\geq 3$, enter the error term. Since the last integral contains only polynomials in $(\varsigma-\sigma)$ with degree three or higher, only the first two terms need to be taken into account

$$\psi(\sigma) = \tau^2\Bigg(\frac{(\lambda-\mu)}{2}\binom{\lambda}{\mu}\int_{-\infty}^{\infty}\Big(1-\Phi(t)\Big)^{\lambda-\mu-1}\frac{\Phi(t)^\mu\mathrm{e}^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\,\mathrm{d}t +\sigma(\lambda-\mu)\binom{\lambda}{\mu}\int_{-\infty}^{\infty}\Big(-\frac{g'(\sigma)}{g(\sigma)}t-\frac{h'(\sigma)}{g(\sigma)}\Big) \times\Big(1-\Phi(t)\Big)^{\lambda-\mu-1}\Phi(t)^{\mu-1}\frac{\mathrm{e}^{-t^2}}{2\pi}\,\mathrm{d}t\Bigg)+\mathcal{O}(\tau^4). \quad (32)$$

The value of the first integral is one. The other integral cannot be solved analytically. Instead, the generalized progress coefficients $e_{\mu;\lambda}^{\alpha,\beta}$ (14) are used. The SAR is finally given by

$$\psi(\sigma) = \tau^2\Bigg(\frac{1}{2}+\sigma\Big(e_{\mu,\lambda}^{1,1}\frac{g'(\sigma)}{g(\sigma)}-c_{\mu/\mu,\lambda}\frac{h'(\sigma)}{g(\sigma)}\Big)\Bigg)+\mathcal{O}(\tau^4). \quad (33)$$

The self-adaptation response (33) has been derived under the assumption $\tau\ll 1$. In the case of the sharp ridge, the functions g and h read $h(\varsigma) = d\sqrt{R^2+\varsigma^2N}-R)$ and $g(\varsigma) = \sqrt{\sigma_\epsilon^2+\varsigma^2(1+d^2(R^2+\varsigma^2N/2)/(R^2+\varsigma^2N)}$. Obtaining the first derivatives, introducing normalizations, and plugging the results into the SAR leads to

$$\psi(\sigma^*) = \tau^2\Bigg(\frac{1}{2}-\frac{c_{\mu/\mu,\lambda}d\sigma^{*2}}{\sqrt{1+\frac{\sigma^{*2}}{N}}\sqrt{\sigma_\epsilon^{*2}+\sigma^{*2}\Big(1+d^2\Big(\frac{1+\frac{\sigma^{*2}}{2N}}{1+\frac{\sigma^{*2}}{N}}\Big)\Big)}} +\frac{e_{\mu,\lambda}^{1,1}\sigma^{*2}\Big(1+d^2\Big(\frac{1+\frac{\sigma^{*2}}{2N}}{1+\frac{\sigma^{*2}}{N}}\Big)\Big)}{\Big(\sigma_\epsilon^{*2}+\sigma^{*2}+d^2\sigma^{*2}\Big(\frac{1+\frac{\sigma^{*2}}{2N}}{1+\frac{\sigma^{*2}}{N}}\Big)\Big)} -\frac{e_{\mu,\lambda}^{1,1}d^2\frac{\sigma^{*4}}{N}\Big(\frac{1}{\big(1+\frac{\sigma^{*2}}{N}\big)^2}\Big)}{2\Big(\sigma_\epsilon^{*2}+\sigma^{*2}+d^2\sigma^{*2}\Big(\frac{1+\frac{\sigma^{*2}}{2N}}{1+\frac{\sigma^{*2}}{N}}\Big)\Big)}\Bigg)+\mathcal{O}(\tau^4). \quad (34)$$

Letting $N\to\infty$ gives finally (11)

$$\psi(\sigma^*) \;=\; \tau^2\Big(\frac{1}{2}+\frac{e_{\mu,\lambda}^{1,1}\sigma^{*2}(1+d^2)}{\sigma^{*2}(1+d^2)+\sigma_\epsilon^{*2}} -\frac{c_{\mu/\mu,\lambda}d\sigma^{*2}}{\sqrt{\sigma^{*2}(1+d^2)+\sigma_\epsilon^{*2}}}\Big)+\mathcal{O}(\tau^4).$$