Mutation Strength Control via Meta Evolution Strategies on the Ellipsoid Model

Michael Hellwig[a,*], Hans-Georg Beyer[a]

[a]*Research Center Process and Product Engineering, Vorarlberg University of Applied Sciences, Hochschulstr. 1, Dornbirn, A-6850, Austria.*

Abstract

The ability of a hierarchically organized evolution strategy (meta evolution strategy) with isolation periods of length one to optimally control its mutation strength is investigated on convex-quadratic functions (referred to as ellipsoid model). Applying the dynamical systems analysis approach a first step towards the analysis of the meta evolution strategy behavior is conducted. A non-linear system of difference equations is derived to describe the mean-value evolution of the respective hierarchically organized strategy. In the asymptotic limit case of large search space dimensions this system is suitable to derive closed-form solutions which describe the longterm behavior of the meta evolution strategy. The steady state mutation strength is bracketed within an interval depending on the mutation strength control parameter. Compared to standard settings in cumulative step-length adaptation evolution strategies the meta evolution strategy realizes almost similar normalized mutation strengths. The performance of the meta evolution strategy turns out to be quite robust to the choice of its control parameters. The results allow for the derivation of the expected running time of the algorithm.

Keywords: evolution strategies, meta evolution strategies, ellipsoid model, mutation strength adaptation

1. Introduction

Evolution strategies (ES) build a subclass of evolutionary algorithms (EA) which is most typically concerned with the optimization of real-valued optimization problems. Albeit applications to combinatorial problems are of course possible, there are few publications regarding discrete optimization. For a more detailed overview of evolution strategies refer to [1]. The present work particularly focuses on the theoretical analysis and the assessment of ~~state-of-the-art~~ evolution strategies which has developed a long-standing tradition in years past. In the first place the scope of the theoretical analysis

*Corresponding author

Email addresses: Michael.Hellwig@fhv.at (Michael Hellwig), Hans-Georg.Beyer@fhv.at (Hans-Georg Beyer)

is not so much the application to complex optimization problems but the theoretical description of the evolution dynamics as well as the identification of advantageous strategy parameter settings. These examinations are typically performed on relatively simple test functions.

Aiming at (near) optimal performance, evolution strategies usually employ a mutation strength adaptation mechanism throughout the optimization process. The mutation strength σ determines the average step-length of the search step within the ES algorithm. The most commonly used step-size control methods are the $1/5$-th success rule [2], mutative self-adaptation (σSA) [3; 2], cumulative step length adaptation (CSA) [4], as well as the use of hierarchically organized evolution strategies [5; 6; 7]. Understanding the working principles of these adaptation techniques by considering the ES in conjunction with the objective functions to be optimized allows for a well-grounded choice of strategy specific parameters, such as the learning parameter, damping constant, etc.

Theoretical analyses of ES started with the investigation of the $(1 + 1)$-ES on the sphere model [2] and have been extended considering other ES variants and fitness environments, respectively. The first completely analyzed algorithm concerned the $(1, \lambda)$-σSA-ES on the sphere model. Further progress was then made in different directions concerning more complex, i.e., recombinative ES and also more complex objective functions such as ridge functions and a subset of positive definite quadratic forms. The most advanced analysis has been presented by [8] where the dynamics of the $(\mu/\mu_I, \lambda)$-σSA-ES on the ellipsoid model have been investigated. The fitness environment denoted as ellipsoid model represents the general case of positive definite quadratic forms (PDQF) for ES which employ isotropic mutations. This is due the algorithm being invariant w.r.t. arbitrary rotations of the coordinate system. Thus the results do also hold for the general fitness model $F(y) = y^\top Qy$ with Q as positive definite matrix (minimization considered). The results of the respective paper implicate the question how alternative σ control techniques do behave.

The use of hierarchically organized ES, similarly referred to as Meta-ES, can be motivated by interpreting the step-size control as an optimization problem. Consequently, an evolution strategy can be applied to learn the optimal step-size according to the underlying optimization problem itself. Acting this way, a two-level strategy is constructed. On the lower level multiple independent evolution strategies (inner ESs) with different step-sizes are operating in the search space of the original optimization problem. The upper level strategy (outer ES) operates in the strategy parameter space of the lower level strategies. That is, after some time of isolation the outer ES compares the performances of the inner ESs and adapts σ accordingly. Variation and selection are used on both levels. The formal Meta-ES notation can be found in [7] and reads

$$[\mu'/\rho', \lambda'(\mu/\rho, \lambda)^\gamma]\,. \tag{1}$$

According to (1), λ' populations conducting $(\mu/\rho, \lambda)$-ESs run in parallel over a number of γ generations. That is, the parameter γ determines the length of the so called isolation period of the inner ES. Each of these λ' inner ESs is realized with different strategy parameters that remain constant during the isolation period. The outer ES then selects those μ' populations which turn out to have the best strategy parameters w.r.t.

a previously defined selection criterion. The way of recombination is specified by the parameter ρ'.

Notice, that the adaptation by means of Meta-ES is not limited to step-size but can be applied to other strategy parameters. For example, Herdy [6] empirically investigated the problem of finding the optimal offspring population size and obtained near optimal values on hyperplane and sphere models. Herdy also expanded the approach of hierarchically organized strategies to a third level aiming at optimal control of the isolation time of the inner strategies. In principle, adding hierarchy levels for the optimization of more strategy parameters is possible. But it should be mentioned that adding higher levels to the Meta-ES becomes very expensive in terms of function evaluations.

The theoretical treatment devoted to Meta-ES is still scarce and so far addresses rather simple fitness models. A primary theoretical contribution to the topic of hierarchically organized strategies was presented in [9] investigating multiple step-size adaptation mechanisms including the $[1, 2(\mu/\mu_I, \lambda)^\gamma]$-Meta-ES on the class of ridge functions. Hence, hierarchically organized strategies appeared to demonstrate the best potential on ridge functions of all adaptation techniques considered. With increasing isolation time γ, nearly optimal performance was achieved. Also in the presence of noise, Meta-ES proved to be more robust than the other approaches.

Considering the sphere model, Beyer *et al.* [10] examined the ability of Meta-ES to optimally control the mutation strength, as well as the parental population size μ of the inner ESs. The theoretical analysis allowed for an improved understanding of the influence of the mutation strength control parameter on the performance of the Meta-ES. While the work of Arnold and MacLeod [9] excluded the sharp ridge the analysis of the $[1, 2(\mu/\mu_I, \lambda)^\gamma]$-Meta-ES on that respective fitness environment was accomplished in [11]. There, an estimate for the choice of the isolation time γ was derived that guarantees to prevent premature convergence on the ridge axis. As it turned out, the qualitative behavior of the Meta-ES on the sharp ridge function is not only determined by the isolation time γ but by the number of function evaluations devoted to the inner ES. In [12] the analysis on the sphere model was extended by adding a noise-term to the fitness landscape. In order to tackle the noisy problem a $[1, 4(\mu/\mu_I, \lambda)^\gamma]$-Meta-ES algorithm simultaneously dealing with the mutation strength control and population size adaptation was considered. Being able to predict the asymptotical growth of the normalized mutation strength the analysis allowed for deeper insight into the adaptation behavior of the Meta-ES in the presence of fitness noise.

This paper addresses the problem of analyzing the σ adaptation on more general quadratic objective functions than the sphere model. In particular the problem of analyzing the step-length control on the ellipsoid model is investigated. In this respect, it contributes to the current analysis of mutation strength adaptation mechanisms on the ellipsoid model which commenced with the work of Beyer and Melkozerov [8]. That paper introduced a new progress measure, the *quadratic progress rate*, allowing for the description of the dynamics of the quadratic distances of the parental centroid to the optimizer. Yielding an estimate for the optimal learning rate that differs from the known sphere model result, the work completed the analysis of isotropic self-adaptive standard ES.

Concerning the $(\mu/\mu_I, \lambda)$-CSA-ES which uses cumulative step-length adaptation

to control the mutation strength the respective analysis has been carried out in [13]. Deriving an approximation of the cumulation path allowed for the extension of the dynamical systems analysis approach developed in [8] to the analysis of the CSA-ES on the ellipsoid model. This way, it was possible to describe the longterm dynamics of the CSA-ES and identify reasonable recommendations for the choice of the strategy parameters.

The aim of this analysis is the derivation of formulas that predict the quantitative behavior of the Meta-ES. The analysis is based on the application of the dynamical systems analysis approach [1] to the $[1, 2(\mu/\mu_I, \lambda)^\gamma]$-Meta-ES. As a first step to more general investigations, the analysis presented here considers only the isolation time $\gamma = 1$, i.e. the $[1, 2(\mu/\mu_I, \lambda)^1]$-Meta-ES.

The transfer of the dynamical systems approach to hierarchical ES can be regarded as an important step towards a complete Meta-ES analysis on the ellipsoid model. Our paper augments the existing theoretical knowledge of mutation strength adaptation mechanisms for evolution strategies. On this basis the analysis approach might advance to allow handling more complex problem formulations or rather to deduce beneficial strategy parameter settings for the application of meta-ES to more difficult optimization problems.

After the introduction of the investigated Meta-ES variant our paper points out the considered optimization problem in Sec. 2. The basis of the analysis consists of the knowledge about the dynamics of the inner strategies. Section 3 recaps the required theory. The evolution equation of the mutation strength dynamics is modeled in Sec. 4. Together with the component-wise evolution equations of the search space parameters, it enables the description of the normalized mutation strength dynamics. In the long run the realized normalized mutation strength oscillates in limit cycles around an unstable fixed point. Being able to compute the limits of this oscillations and modeling the normalized mutation strength by mean value dynamics allows for the analysis of the algorithm's longterm dynamics. Finally, an estimate of the expected running time in terms of isolation periods of the Meta-ES is provided. Finally, Sec. 5 provides the conclusion of our analysis and discusses its implications on future work.

2. Meta-ES Algorithm and Problem Formulation

Throughout the paper a simple $[1, 2(\mu/\mu_I, \lambda)^\gamma]$-Meta-ES is used to control the mutation strength. The Meta-ES employs two inner $(\mu/\mu_I, \lambda)$-ESs which evolve from the same initial search space parameter vector $\mathbf{y}_p$ but with different mutation strengths. The mutation strengths are held constant within the inner strategies over the isolation period of γ generations. The outer ES which controls σ on a higher level is described in Fig. 1. In line 4 and 5 the two different σ values σ_1, σ_2 are generated by increasing and decreasing the parental mutation strength σ_p by the factor $\alpha > 1$, respectively. Consequently, one inner $(\mu/\mu_I, \lambda)$-ES runs with mutation strength $\sigma_1 = \alpha\sigma_p$ and one with $\sigma_2 = \sigma_p/\alpha$. Selection is performed in lines 8 and 9. The standard notation "$m; \lambda'$" indicates that the m-th best population out of all λ' populations is chosen. The populations evolve independently over γ generations. Afterwards they are ordered by the function values returned by the respective inner standard ES, displayed in Fig. 2. The mutation strength of the inner strategy which provides the better fitness function value

```
INITIALIZE [y_p, σ_p, α, μ, λ, γ];                          1
t ← 0;                                                      2
REPEAT                                                      3
  σ_1 ← σ_p α;                                              4
  σ_2 ← σ_p / α;                                            5
  [y_1, F(y_1)] ← ES(μ, λ, γ, σ_1, y_p);                    6
  [y_2, F(y_2)] ← ES(μ, λ, γ, σ_2, y_p);                    7
  σ_p ← σ_{1;2};                                            8
  y_p ← y_{1;2};                                            9
  t ← t + 1;                                                10
UNTIL [termination condition]                               11
```

Figure 1: Pseudo code of the $[1, 2(\mu/\mu_I, \lambda)^{\gamma}]$-Meta-ES. The Code of the inner ES is displayed in Fig. 2.

serves as the new σ_p. This procedure is repeated until the termination criterion is met (maximal number of function evaluations, fixed number of isolation periods, etc.).

The inner ES is a standard $(\mu/\mu_I, \lambda)$ evolution strategy which operates with constant strategy parameters. A population of λ offspring is created by multiplying a vector of independent, standard normally distributed components with the mutation strength σ and then adding the product to the parental centroid $\langle \mathbf{y} \rangle$ of the previous generation. The new parental centroid $\langle \mathbf{y} \rangle$ is composed of the μ best candidates (w.r.t. their function values F_l). Having evolved over γ generations, the inner ES returns the new centroid $\langle \mathbf{y} \rangle$ and the associated fitness value $F(\langle \mathbf{y} \rangle)$].

This paper is concerned with the analysis of the Meta-ES' ability to adapt the optimal mutation strength on fitness environment referred to as the ellipsoid model. The respective objective function reads

$$F(\mathbf{y}) = \sum_{i=1}^{N} a_i y_i^2, \quad a_i > 0, \tag{2}$$

where N is the search space dimension and the a_i are the ellipsoid coefficients. Among others, the cigar function ($a_1 = 1,\ a_i = \xi\ \forall i \neq 1$ with condition number ξ) as well as

```
INITIALIZE [σ, ⟨y⟩, μ, λ, γ]                                 1
g ← 1;                                                      2
WHILE g ≤ γ                                                 3
  FOR l ← 1 TO λ                                            4
    y_l ← ⟨y⟩ + σ N_l(0, I);                                5
    F_l ← F(y_l);                                           6
  END FOR                                                   7
  ⟨y⟩ ← (1/μ) Σ_{m=1}^{μ} y_{m;λ};                          8
  g ← g + 1;                                                9
END WHILE                                                   10
RETURN [⟨y⟩, F(⟨y⟩)];                                       11
```

Figure 2: The standard $(\mu/\mu_I, \lambda)$-ES with constant mutation strength.

the sphere model ($a_i = 1$) represent special cases of the ellipsoid model. Throughout the analysis the ellipsoid test model cases $a_i = i$ and $a_i = i^2$ are considered in order to take into account different magnitudes of the problem conditioning. The optimization problem can be formulated as

$$\min_{\mathbf{y}\in\mathbb{R}^N} F(\mathbf{y}).$$

Its optimizer is located at the origin of the coordinate system, $\boldsymbol{y} = \mathbf{0}$. Notice, that Eq. (2) is a representation of the general case of positive definite quadratic forms for the $(\mu/\mu_I, \lambda)$-ES. This is a result of the isotropy of the mutation vectors of the inner algorithm (line 5) which guarantees invariance w.r.t. arbitrary rotations of the coordinate system.

3. The Inner ES Dynamics

The description of the dynamics of the inner ESs serves as the starting point of the Meta-ES analysis on the ellipsoid model (2). According to [8], the quadratic progress rate of an ES is component-wise defined as

$$\varphi_i^{II} \coloneqq E\left[y_i^{(g)^2} - y_i^{(g+1)^2} \,|\, \mathbf{y}^{(g)} \right]. \tag{3}$$

It specifies the expected one-generation change of the parameter vectors' squared distance to the optimizer for each component and is derived as

$$\begin{aligned} \varphi_i^{II}(\sigma^{(g)}) =& \frac{2\sigma^{(g)} c_{\mu/\mu,\lambda} a_i y_i^{(g)^2}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}} \\ & - \frac{\sigma^{(g)^2}}{\mu}\left(1 + \frac{\left((\mu-1)e_{\mu,\lambda}^{2,0} + e_{\mu,\lambda}^{1,1}\right) a_i^2 y_i^{(g)^2}}{\sum_{j=1}^N a_j^2 y_j^{(g)^2}} \right). \end{aligned} \tag{4}$$

In this context, the term $e_{\mu,\lambda}^{a,b}$ refers to the *generalized progress coefficients* of the $(\mu/\mu_I, \lambda)$-ES [14]

$$e_{\mu,\lambda}^{a,b} = \frac{\lambda-\mu}{\sqrt{2\pi}^{a+1}} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} (-t)^b \mathrm{e}^{-\frac{a+1}{2}t^2} \left(1-\Phi(t)\right)^{\lambda-\mu-1} \Phi(t)^{\mu-a} \mathrm{d}t, \tag{5}$$

where $\Phi(t)$ denotes the cumulative distribution function of a standard normal random variable and $c_{\mu/\mu,\lambda} \coloneqq e_{\mu,\lambda}^{1,0}$ as a special case represents the so called progress coefficient.

Assuming that the parental population size is considerably smaller than the search space dimensionality, and provided that there exists no dominating ellipsoid coefficient, i.e., $\forall i : \sum_{j\neq i} a_j \gg a_i$, the progress rate is asymptotically[1]

$$\varphi_i^{II}(\sigma^{(g)}) = \frac{2\sigma^{(g)} c_{\mu/\mu,\lambda} a_i y_i^{(g)^2}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}} - \frac{\sigma^{(g)^2}}{\mu}. \tag{6}$$

[1] In order to obtain (6) from (4) the y_i^2 dynamics are required to behave "nicely". That is, they have to exhibit a similar longterm behavior. As has been checked in [8] by reinserting the final y_i^2 results into the quadratic progress rate on the ellipsoid model this always holds in the asymptotic limit.

The readability of this paper is improved by using the abbreviation $\sum_{i=1}^{N} a_i =: \Sigma a$ for the sum of the ellipsoid coefficients. Further, the length of the equations can be reduced by denoting the square root in the denominator of the progress rate (6) as

$$R_a(\mathbf{y}^{(g)}) := \sqrt{\sum_{i=1}^{N} a_i^2 y_i^{(g)2}}. \tag{7}$$

Introducing the normalizations

$$\sigma^{*(g)} := \sigma^{(g)} \, \Sigma a / R_a(\mathbf{y}^{(g)}), \tag{8}$$

$$\varphi_i^{II*} := \varphi_i^{II} \Sigma a, \tag{9}$$

we obtain the asymptotic normalized progress rate from Eq. (6)

$$\varphi_i^{II*}(\sigma^*) = 2\sigma^* c_{\mu/\mu,\lambda} a_i {y_i}^2 - \frac{\sigma^{*2}}{\mu} \frac{\sum_{j=1}^{N} a_j^2 y_j^2}{\Sigma a}. \tag{10}$$

In what follows the asymptotic component-wise quadratic progress rate (6) and its normalized representation (10) will be considered. As can be inferred from [8] they provide an accurate description of the $(\mu/\mu_I, \lambda)$-ES dynamics.

4. Evolution Equations

In the next step, the evolution equations of the Meta-ES[2] on the general ellipsoid model will be analyzed considering an isolation time of $\gamma = 1$. The fitness values of the parental centroids returned from both inner strategies are computed. Comparing them after every single generation the mutation strength is adapted accordingly. After having derived the corresponding system of evolution equations for the $[1, 2(\mu/\mu_I, \lambda)^1]$-Meta-ES, the dynamics of the normalized mutation strength are analyzed. Since the normalized mutation strength dynamics exhibit a strong oscillatory behavior, their occurring oscillations are bracketed in lower and upper bounds. Finally, the longterm behavior of the Meta-ES can be described allowing for the prediction of the expected running time of the algorithm.

4.1. Derivation of the Evolution Equations

Following the dynamical systems approach in [15], the stochastic process of the ES from generation g to generation $g + 1$ can be subdivided into mean value dynamics and fluctuation terms. The mean value parts are directly given by Eq. (6). Letting ϵ_i denote the component-wise fluctuations of the stochastic process, one obtains the quadratic difference equation

$$y_i^{(g+1)2} = y_i^{(g)2} - \varphi_i^{II}(\sigma^{(g)}) + \epsilon_i. \tag{11}$$

[2]Unless indicated otherwise, thoughout the rest of the paper the term "the Meta-ES" will refer to the $[1, 2(\mu/\mu_I, \lambda)^1]$-Meta-ES variant with isolation time one.

It describes the one-generation change of the component-wise squared distance to the optimizer and serves as the starting point of the theoretical investigations.

The next task is calculating the evolution equation of the σ dynamics. Therefore, the inner evolution strategies which operate with different mutation strengths are investigated. The first strategy is equipped with a mutation strength which is increased by the multiplicative factor $\alpha > 1$, while the mutation strength within the second inner ES is decreased by division with the same factor α. We write

$$\begin{aligned} \sigma_1^{(g)} &:= \sigma^{(g)}\alpha, \\ \sigma_2^{(g)} &:= \sigma^{(g)}/\alpha. \end{aligned} \tag{12}$$

Considering the objective function (2) and making use of Eq. (6) the objective function values resulting from the two inner ESs can be computed. The fitness $F_1^{(g+1)} := F(\mathbf{y}^{(g+1)}, \sigma_1^{(g)})$ of the parental centroid returned by the first inner strategy operating with $\sigma_1^{(g)}$ for a single iteration is

$$F_1^{(g+1)} = \sum_{i=1}^{N} a_i \left({y_i^{(g)}}^2 - \frac{2\alpha\sigma^{(g)} c_{\mu/\mu,\lambda} a_i \left(y_i^{(g)}\right)^2}{R_a(\mathbf{y}^{(g)})} + \frac{\alpha^2 {\sigma^{(g)}}^2}{\mu} \right) + \epsilon_{F_1}. \tag{13}$$

By rearranging (13) it becomes

$$F_1^{(g+1)} = F(\mathbf{y}^{(g)}) - 2\alpha\sigma^{(g)} c_{\mu/\mu,\lambda} R_a(\mathbf{y}^{(g)}) + \frac{\alpha^2 {\sigma^{(g)}}^2}{\mu} \Sigma a + \epsilon_{F_1}. \tag{14}$$

In this context, the sum of the corresponding component-wise fluctuations is abbreviated by the term $\epsilon_{F_1} := \sum_{i=1}^{N} a_i \epsilon_i$. Analogously, the fitness value $F_2^{(g+1)} := F(\mathbf{y}^{(g+1)}, \sigma_2^{(g)})$ realized by the second inner ES using $\sigma_2^{(g)}$ can be calculated as

$$F_2^{(g+1)} = F(\mathbf{y}^{(g)}) - \frac{2\sigma^{(g)} c_{\mu/\mu,\lambda}}{\alpha} R_a(\mathbf{y}^{(g)}) + \frac{{\sigma^{(g)}}^2}{\alpha^2 \mu} \Sigma a + \epsilon_{F_2}. \tag{15}$$

According to line 8 of the Meta-ES algorithm, see Fig. 1, the difference of the fitness values governs the mutation strength adaptation in generation $g + 1$ in the following way

$$\begin{aligned} F_1^{(g+1)} - F_2^{(g+1)} > 0 \Rightarrow \quad \sigma^{(g+1)} &= \sigma_2^{(g)}, \\ F_1^{(g+1)} - F_2^{(g+1)} < 0 \Rightarrow \quad \sigma^{(g+1)} &= \sigma_1^{(g)}. \end{aligned} \tag{16}$$

Let $\epsilon_{\Delta F} := \epsilon_{F_1} - \epsilon_{F_2}$ denote the difference of the fluctuation sums in (14) and (15), then $F_1^{(g+1)} - F_2^{(g+1)}$ is calculated as

$$F_1^{(g+1)} - F_2^{(g+1)} = \frac{{\sigma^{(g)}}^2}{\mu} \Sigma a \left(\alpha^2 - \frac{1}{\alpha^2} \right) - 2\sigma^{(g)} c_{\mu/\mu,\lambda} R_a(\mathbf{y}^{(g)}) \left(\alpha - \frac{1}{\alpha} \right) + \epsilon_{\Delta F}. \tag{17}$$

By rearrangements and after normalization with (8), Eq. (17) becomes

$$F_1^{(g+1)} - F_2^{(g+1)} = 2c_{\mu/\mu,\lambda}\sigma^{(g)} R_a(\mathbf{y}^{(g)}) \left(\alpha - \frac{1}{\alpha} \right) \left[\left(\alpha + \frac{1}{\alpha} \right) \frac{\sigma^{*(g)}}{2\mu c_{\mu/\mu,\lambda}} - 1 \right] + \epsilon_{\Delta F}. \tag{18}$$

The term inside the squared brackets defines a discriminator function that controls the σ evolution

$$\Delta(\sigma^{*(g)}) := \left(\alpha + \frac{1}{\alpha}\right) \frac{\sigma^{*(g)}}{2\mu c_{\mu/\mu,\lambda}} - 1 + \epsilon_\Delta. \tag{19}$$

Here, the $\epsilon_{\Delta F}$ fluctuations from (18) are incorporated into the fluctuation term ϵ_Δ.

Since all other terms in (18) are positive ($\alpha > 1$), the sign of the difference $F_1^{(g+1)} - F_2^{(g+1)}$ in generation $g+1$ is equivalent to the sign of $\Delta(\sigma^{*(g)})$

$$\begin{aligned} F_1^{(g+1)} - F_2^{(g+1)} > 0 \Leftrightarrow \quad \Delta(\sigma^{*(g)}) > 0, \\ F_1^{(g+1)} - F_2^{(g+1)} < 0 \Leftrightarrow \quad \Delta(\sigma^{*(g)}) < 0. \end{aligned} \tag{20}$$

Because of the relation (16), this allows to formulate the evolution equation of the mutation strength

$$\sigma^{(g+1)} = \sigma^{(g)} \alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))}. \tag{21}$$

Having modeled the mutation strength evolution (21), the $[1, 2(\mu/\mu_I, \lambda)^1]$-Meta-ES is described by the system of $N+1$ evolution equation

$$\begin{aligned} {y_i^{(g+1)}}^2 &= {y_i^{(g)}}^2 - \varphi_i^{II}(\sigma^{(g)} \alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))}) + \epsilon_i, \\ \sigma^{(g+1)} &= \sigma^{(g)} \alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))}. \end{aligned} \tag{22}$$

Notice, the σ evolution equation (21) is exposed to fluctuations since the discriminator function $\Delta(\sigma^{*(g)})$ is subject to the ϵ_Δ fluctuations. In order to keep the analysis tractable the fluctuation terms ϵ_i, and consequently ϵ_Δ in (21), will be ignored in the following and system (22) becomes

$$\boxed{\begin{aligned} {y_i^{(g+1)}}^2 &= {y_i^{(g)}}^2 - \varphi_i^{II}(\sigma^{(g)} \alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))}), \\ \sigma^{(g+1)} &= \sigma^{(g)} \alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))}. \end{aligned}} \tag{23}$$

In order to assess the quality of the modeling approach the iteratively generated $\mathbf{y}$ and σ dynamics resulting from system (23) are compared to experimental runs of the Meta-ES. System (23) is based on the asymptotic progress rate (6) which was derived in the limit of large search space dimensions N. In order to check whether (23) allows for good conclusions about the ES behavior in smaller dimensions, its verification is carried out in both rather small ($N = 40$) as well as in larger dimensions ($N = 200$). Due to space restrictions not all experiments could be included in this publication.

In Fig. 3 the σ and y_i^2 dynamics (for the components $i = 1, N/4, N$) are displayed on the ellipsoid model $a_i = i^2$. Considering the search space dimension $N = 200$ the iterative dynamics are generated applying the asymptotic quadratic progress rate (6) to (23). The $[1, 2(3/3_I, 10)^1]$-Meta-ES using the adaptation parameter $\alpha = 1.2$ is initialized at $\sigma^{(0)} = 1$, $\mathbf{y}^{(0)} = \mathbf{1}$ and evolves over 10^4 isolation periods of length $\gamma = 1$. All experimental Meta-ES dynamics are averaged over 10^3 independent runs. The illustration in Fig. 3 reveals a good agreement between experimental dynamics and analytical predictions. It shows the typical behavior of the $[1, 2(\mu/\mu_I, \lambda)^1]$-Meta-ES on the ellipsoid model. At the beginning of the optimization the dynamics exhibit a transient phase that

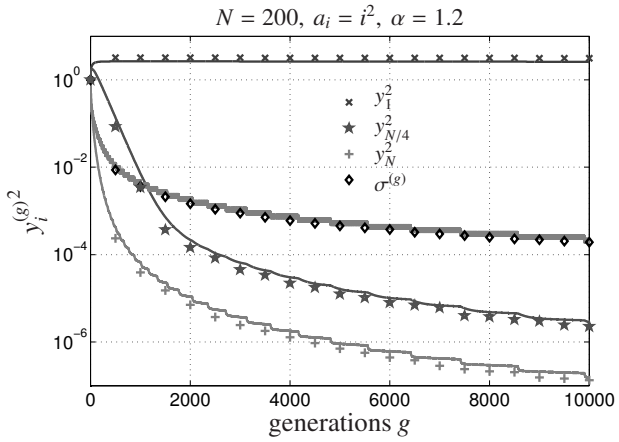

Figure 3: The dynamics of the $[1, 2(3/3_I, 10)^1]$-Meta-ES plotted against the number of generations. The solid lines represent the predictions of the iterative system (23). Additionally, the corresponding σ dynamics are displayed. All experimental results are averaged over 10^3 independent runs and illustrated by the data points.

is characterized by a rapid decline of the $y^2_{N/4}$, y^2_N, and σ curves. The y^2_1 curve decreases as well, but at a much smaller rate. Considering the longterm behavior of the Meta-ES all y^2_i dynamics decrease slower and at the same rate. In the long run the Meta-ES continuously decreases the component-wise squared distance to the optimizer as well as the mutation strength. The decline is characterized by step-wise oscillation phases. These oscillations are resulting from changes of the sign of the discriminator function $\Delta(\sigma^*)$ governed by the dynamics of the normalized mutation strength σ^*.

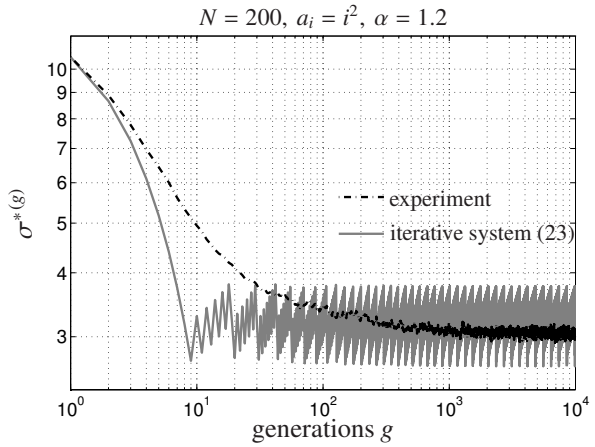

Figure 4: The normalized mutation strength dynamics of the $[1, 2(3/3_I, 10)^1]$-Meta-ES plotted against the number of generations. The predictions resulting from iterative system (23) are compared to real Meta-ES runs. The experimental results are averaged over 10^3 independent runs.

Regarding the normalized mutation strength in Fig. 4 both phases can be rediscovered. With increasing time the σ^* dynamics approach an unstable fixed point. That is, the normalized mutation strength exhibits a "steady state" like behavior oscillating in a limit cycle of large periodicity. Since the mutations within the inner ES influence the σ^* evolution the respective limit cycle is subject to mutative noise. In the following, we will refer to this noisy limit cycle as the steady state of the Meta-ES. Although being subject to the mentioned oscillations, after approaching its σ^* limit cycle the σ, and y_i^2 dynamics exhibit a log-linear longterm descend. This stable longterm behavior motivates our use of the term "steady state".

For $\alpha = 1.2$, the iterative and the experimental dynamics show a good agreement. But the agreement of the experimental results and theoretical predictions in Fig. 3 and Fig. 4 is not guaranteed when considering smaller values of the control parameter $1 < \alpha < 1.1$. The use of small α values may cause substantial deviations between predictions and real Meta-ES runs. Especially in low search space dimensions the deviations can be immense. Refer to Fig. 5 to visualize the impact of small α values on the prediction quality. These deviations are a result of the fluctuations that have been ignored in the iterative system (23). Their origin is investigated in more detail within Sec. 4.2. Even averaging over multiple Meta-ES runs is not sufficient to mitigate their influence on the experimental runs.

4.2. The ϵ_Δ Fluctuations

Regarding Fig. 5 the deviations of experimental and theoretical results for small values of α have to be examined more closely. Hence, one-generation Meta-ES experiments are considered in order to check the validity of the iterative dynamics. The one-generation experiments are performed in the following way:

a) A single iteration step of the $[1, 2(\mu/\mu_I, \lambda)^1]$-Meta-ES is executed for a given σ^* value and initial parameter vector $\mathbf{y}^{(0)}$. Therefore, the initial σ^* is renormalized to σ using Eq. (8). Then both inner strategies are iterated for one generation with mutation strength $\sigma\alpha$ and σ/α, respectively.

b) According to line 6 and 7 of the algorithm the function values returned by the inner strategies are used to compute the quantity

$$\tilde{\Delta}(\sigma^*) := \frac{(F(\mathbf{y}_1) - F(\mathbf{y}_2))}{\sigma \sum_{j=1}^{N} a_j^2 {y_j^{(0)}}^2}. \tag{24}$$

c) The steps a) and b) are repeated G times. Finally, the resulting samples are averaged.

The one-generation experiments for several values of $\sigma^* \in (0, 2\mu c_{\mu/\mu,\lambda}]$ are compared to theoretical predictions of $\tilde{\Delta}(\sigma^*)$ derived from Eq. (18) by straight forward rearrangements as

$$\tilde{\Delta}(\sigma^*) = 2c_{\mu/\mu,\lambda}\left(\alpha - \frac{1}{\alpha}\right)\left[\left(\alpha + \frac{1}{\alpha}\right)\frac{\sigma^{*(g)}}{2\mu c_{\mu/\mu,\lambda}} - 1\right]. \tag{25}$$

Notice, that (25) differs from (19) (after discarding ϵ_Δ) only by multiplication with $2c_{\mu/\mu,\lambda}\left(\alpha - \frac{1}{\alpha}\right)$. Provided $\alpha > 1$, this factor is always positive and does not change

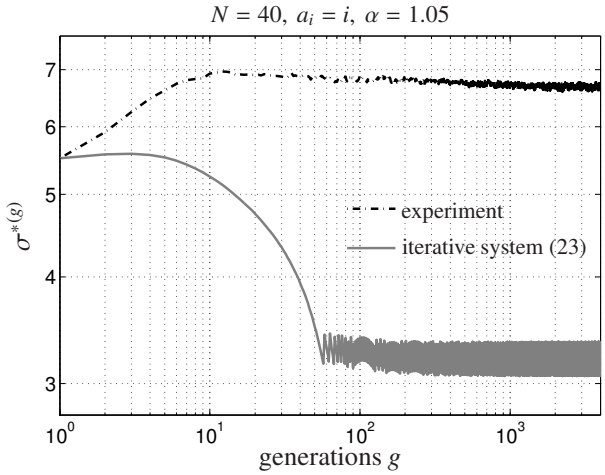

Figure 5: Illustration of the fluctuation impact on the dynamics of the $[1, 2(3/3_I, 10)^1]$-Meta-ES. Again, the solid line represent the predictions of the iterative system (23). The experimental dynamics are averaged over 10^3 independent runs and illustrated by the dashed line.

the sign of the discriminant function Δ. Thus, the predictions of (25) can also be used to characterize the mutation strength adaptation of the Meta-ES. Correspondingly, Eq. (24) is suitable for the experimental validation of this predictions.

The one-generation experiments of the $[1, 2(3/3_I, 10)^1]$-Meta-ES on the ellipsoid model $a_i = i$ are illustrated in Fig. 6 considering two different values of α. There, the mean values of the real Meta-ES runs are obtained by averaging over 2×10^5 one-generation experiments and displayed together with their standard deviations by the error bar plot. The experimental data for $\alpha = 1.02$ in (a) are averaged over 5×10^5 one-generation experiments. The dashed black line represents the theoretical prediction obtained from Eq. (25). The figures show a good agreement of the theoretical predictions with the experimentally obtained data. On account of space restrictions the one-generation experiments are illustrated for $N = 40$. Nevertheless, the conformity of theoretical and experimental results increases with growing search space dimensionality N.

One observes that the slope of $\tilde{\Delta}(\sigma^*)$-curve increases with growing α. This indicates larger differences of the $\tilde{\Delta}(\sigma^*)$-values realized by two inner strategies operating with different normalized mutation strength. Greater values of the mutation strength control parameter α increase the ability of the Meta-ES to distinguish the inner strategies, see Fig. 6 (b). Independently of the choice of α or the search space dimension the magnitudes of the standard deviations exhibit only minor changes. That is, the much lower slope of the $\tilde{\Delta}(\sigma^*)$ curve for small α indicates an increasing influence of the occurring fluctuations on the Meta-ES dynamics. As can be observed in Fig. 6 (a), the standard deviations are considerably larger than the mean values. This corresponds to a much smaller signal-to-noise ratio. It explains the huge deviations of the experimental and theoretical dynamics appearing in Fig. 5. Due to the high fluctuations for small values of α almost every second decision to increase or decrease the mutation strength

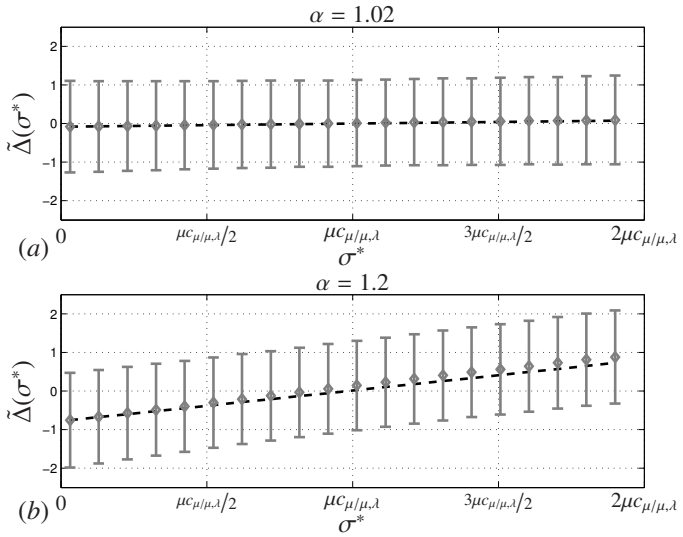

Figure 6: One-generation experiments of the $[1, 2(3/3_I, 10)^1]$-Meta-ES on the ellipsoid model $a_i = i$, $N = 40$. The initial parameter vector is $\mathbf{y} = \mathbf{1}$. The experimental measurements (24) are compared to the predictions (25) and illustrated for $\alpha = 1.02$ in (a), as well as $\alpha = 1.2$ in (b).

is wrong.

Another way to examine whether the deviations of the experimental and theoretical dynamics originate from neglecting the fluctuations within the iterative dynamics (21) is to incorporate a smoothing mechanism into the $[1, 2(3/3_I, 10)^1]$-Meta-ES algorithm, see Fig.1. The mutation strength control is governed by the $\tilde{\Delta}$ values corresponding to the difference $(F(\mathbf{y}_1) - F(\mathbf{y}_2))$. For the purpose of smoothing this decision value a cumulation approach is added to the algorithm. The $\tilde{\Delta}$ cumulation within the Meta-ES algorithm aims at smoothing the occurring ϵ_Δ fluctuations in the experimental dynamics. This approach should increase the experiments' agreement with the theoretical predictions. The decision whether the mutation strength is increased or decreased is no longer based directly on the current sign of the difference of the function values returned by the inner strategies. Instead, the measured $\tilde{\Delta}$ values are accumulated to

$$\hat{\Delta} \leftarrow (1 - \varrho)\hat{\Delta} + \varrho\tilde{\Delta}. \tag{26}$$

Containing the fading record of the past $\tilde{\Delta}$ values $\hat{\Delta}$ governs the mutation strength adaptation. In this process the parameter $0 < \varrho < 1$ controls the memory of the Δ-cumulation. Making use of $\varrho = 1/\sqrt{3N}$, an illustration of the cumulation approach is provided in Fig. 7. Apart from the incorporation of the $\hat{\Delta}$ cumulation into the Meta-ES algorithm the same configurations as in Fig. 5 have been used. The figure reveals that the mitigation of the ϵ_Δ fluctuations in fact moves the experimental Meta-ES dynamics towards their theoretical predicted behavior. However, the cumulation parameter ϱ has to be set appropriately because a very small ϱ on the one hand stretches the transient phase of the algorithm and on the other hand it needs considerably more time to compensate wrong decisions caused by very large fluctuations. Using too large values of ϱ

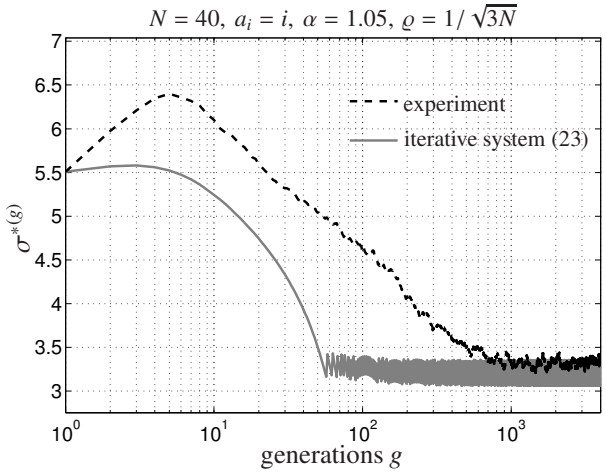

Figure 7: Illustration of the effect of $\hat{\Delta}$ cumulation (26) on the normalized mutation strength dynamics of the $[1, 2(3/3_I, 10)^1]$-Meta-ES. Again, the solid line represent the predictions of the iterative system (23). The experimental results are averaged over 10^3 independent Meta-ES runs and illustrated by the dashed line.

of course would decrease the dampening effect of the Δ-cumulation.

Conclusively, the deviations in fact originate from neglecting the ϵ_Δ fluctuations within the theoretical model of the evolution equations (19) and (23), respectively. It turns out that small values of the control parameter α decrease the ability of the Meta-ES to distinguish the fitness values returned by the two inner strategies. Using larger α reduces the influence of the ϵ_Δ fluctuations on the σ adaptation and improves the agreement of the theoretical model with the experimental results. That is, provided that α is chosen appropriately even in low search spaces the iterative system (23) represents a suitable model of the $[1, 2(\mu/\mu_I, \lambda)^1]$-Meta-ES. In the following analysis α is assumed to fit this requirement.

4.3. Normalized Mutation Strength

Having validated the modeling approach that resulted in system (23) the next step aims at taking a closer look at the normalized mutation strength dynamics depicted in Fig. 4. As already mentioned, the iteratively generated σ dynamics decrease over time revealing a step-wise oscillatory behavior. These oscillations are a result of changes in the sign of $\Delta(\sigma^{*(g)})$, cf. Eq. (19). Since the σ dynamics of the real Meta-ES runs are subject to large fluctuations, and also due to the averaging over 1000 independent runs, the theoretical oscillations are not exactly reflected by the experimental dynamics. But regarding the overall behavior, at least for sufficiently large values of α the dynamics show a good agreement. In order to gain further insight into the σ dynamics, the evolution of the normalized mutation strength σ^* has to be analyzed. For the description of the normalized mutation strength σ^* we recall the mutation strength normalization

in (8) and the progress rate (6). The σ^* dynamics are then obtained as

$$\sigma^{*(g+1)} = \sigma^{(g+1)} \frac{\Sigma a}{R_a(\mathbf{y}^{(g+1)})} = \frac{\sigma^{(g)} \alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))} \Sigma a}{\sqrt{\sum_{j=1}^{N} a_j^2 \left(y_j^{(g+1)}\right)^2}}. \tag{27}$$

Due to the progress rate definition the sum in the denominator can be rewritten as

$$\sum_{j=1}^{N} a_j^2 y_j^{(g+1)2} = \sum_{j=1}^{N} a_j^2 y_j^{(g)2} - \sum_{j=1}^{N} a_j^2 \varphi_j^{II}(\sigma^{(g)} \alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))}). \tag{28}$$

Accordingly, by factoring out $\sqrt{\sum_{i=1}^{N} a_i^2 y_i^2}$, Eq. (27) becomes

$$\sigma^{*(g+1)} = \frac{\sigma^{(g)} \alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))} \Sigma a}{\sqrt{\sum_{j=1}^{N} a_j^2 y_j^{(g)2}} \sqrt{1 - \frac{\sum_{j=1}^{N} a_j^2 \varphi_j^{II}(\sigma^{(g)} \alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))})}{\sum_{j=1}^{N} a_j^2 y_j^{(g)2}}}}, \tag{29}$$

and thus the evolution equation of the normalized mutation strength reads

$$\sigma^{*(g+1)} = \frac{\sigma^{*(g)} \alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))}}{\sqrt{1 - \frac{\sum_{j=1}^{N} a_j^2 \varphi_j^{II}(\sigma^{(g)} \alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))})}{\sum_{j=1}^{N} a_j^2 y_j^{(g)2}}}}. \tag{30}$$

Recurrence equations like Eq. (30) can exhibit qualitatively different dynamics: stable fixed points, limit cycles, or chaotic behaviors. Since the influence of the ellipsoid model, as well as the influence of parameters like α and N, on Eq. (30) is not evident one is interested in further simplifications. That is, considering (6) and (8), after some straight forward operations the fraction $\sum_{j=1}^{N} a_j^2 \varphi_j^{II}(\sigma^{(g)} \alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))}) / \sum_{j=1}^{N} a_j^2 y_j^{(g)2}$ within the square root of (30) becomes

$$\frac{\sum_{j=1}^{N} a_j^2}{(\Sigma a)^2} \left(2 c_{\mu/\mu,\lambda} \sigma^{*(g)} \alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))} \frac{\Sigma a \sum_{j=1}^{N} a_j^3 y_j^{(g)2}}{\sum_{j=1}^{N} a_j^2 \sum_{j=1}^{N} a_j^2 y_j^{(g)2}} - \frac{\left(\sigma^{*(g)} \alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))}\right)^2}{\mu} \right). \tag{31}$$

For reasons of clarity and comprehensibility the abbreviation

$$Q \coloneqq \frac{\Sigma a \sum_{j=1}^{N} a_j^3 y_j^2}{\sum_{j=1}^{N} a_j^2 \sum_{j=1}^{N} a_j^2 y_j^2}. \tag{32}$$

is introduced. It contains the fraction depending on the y_i^2 dynamics. That is, (31) can be reformulated as

$$\frac{\sum_{j=1}^{N} a_j^2}{(\Sigma a)^2} \left(2 c_{\mu/\mu,\lambda} \sigma^{*(g)} \alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))} Q - \frac{\left(\sigma^{*(g)} \alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))}\right)^2}{\mu} \right). \tag{33}$$

Denoting the bracketed expression in (33) as

$$\tilde{\varphi}^{II*}(\sigma^{*(g)}\alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))}) := 2c_{\mu/\mu,\lambda}\sigma^{*(g)}\alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))}Q - \frac{\left(\sigma^{*(g)}\alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))}\right)^2}{\mu}, \tag{34}$$

the evolution of the normalized mutation strength can be condensed to

$$\sigma^{*(g+1)} = \frac{\sigma^{*(g)}\alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))}}{\sqrt{1 - \frac{\sum_{j=1}^{N} a_j^2}{(\Sigma a)^2}\tilde{\varphi}^{II*}(\sigma^{*(g)}\alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))})}}. \tag{35}$$

Notice, that notation $\tilde{\varphi}$ in (34) is chosen because the expression on the rhs of (34) resembles the usual structure of the progress rate of an ES. Multiplying the normalized progress rate (10) with a_i^2, taking the sum of all N components and considering (33) yields the following relation to the normalized progress rate (10)

$$\frac{\sum_{j=1}^{N} a_j^2}{(\Sigma a)^2}\tilde{\varphi}^{II*}(\sigma^*) = \frac{1}{\Sigma a \sum_{j=1}^{N} a_j^2 y_j^2}\sum_{i=1}^{N} a_i^2 \varphi_i^{II*}(\sigma^*). \tag{36}$$

Returning to Eq. (35), it represents an iterative mapping of the form $\sigma^{*(g+1)} = f_\sigma\left(\sigma^{*(g)};\alpha\right)$. Like the recurrence equation (30) it is still difficult to analyze due to its complexity. But assuming that the Meta-ES system operates in its steady state and following the analysis in [13] the term Q can be approximated by the expression

$$Q(\sigma^{*(g)}) \simeq \frac{\sigma^{*(g)}}{2\mu c_{\mu/\mu,\lambda}} + \Sigma a \frac{\breve{a}}{\sum_{i=1}^{N} a_i^2}. \tag{37}$$

Notice that $\breve{a} := \min\left\{a_j \,|\, j = 1,\ldots,N\right\}$ denotes the smallest ellipsoid coefficient.

The use of this approximation is justified by comparing the Q value approximation (37) to measurements within experimental Meta-ES runs. The validation is illustrated in Fig. 8 considering the $[1, 2(3/3, 10)^1]$-Meta-ES ($\alpha = 1.2$) on the ellipsoid model $a_i = i$ in dimension $N = 40$. There, represented by the dashed black " $\times$ " line the Q values corresponding to (32) are calculated from the experimental y_i^2 dynamics. On the other hand measurements of the normalized mutation strengths σ^* are utilized to compute the approximated Q dynamics (37) depicted as dotted gray " + " curve. All measurements are obtained from experimental Meta-ES runs that have been averaged over 1000 independent runs. The good agreement of the results displayed in Fig. 8 validates the use of approximation (37) provided that the Meta-ES has reached its steady state limit cycle.

By making use of the approximation (34) and inserting $Q(\sigma^{*(g)}\alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))})$ into Eq. (35) one obtains a recurrence equation of the normalized mutation strength which is independent of $\mathbf{y}^{(g)}$

$$\boxed{\sigma^{*(g+1)} \simeq \frac{\sigma^{*(g)}\alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))}}{\sqrt{1 - \frac{2c_{\mu/\mu,\lambda}\sigma^{*(g)}\alpha^{-\mathrm{sign}(\Delta(\sigma^{*(g)}))}\breve{a}}{\Sigma a}}}.} \tag{38}$$

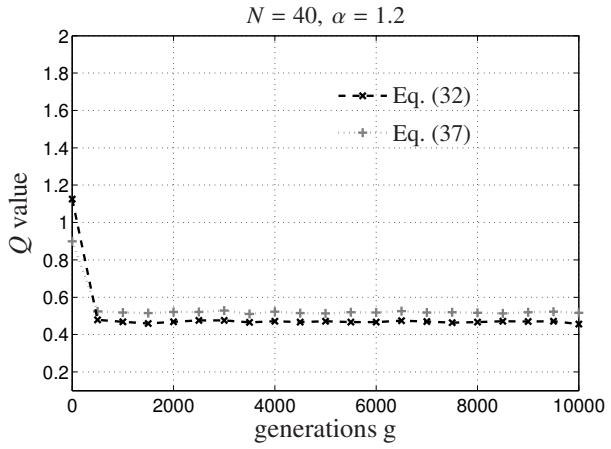

Figure 8: The Q value dynamics of the $[1, 2(3/3_I, 10)^1]$-Meta-ES on the ellipsoid model ($a_i = i$). The strategy is initialized at $y_i^{(0)} = 1\ \forall i$ with $\sigma_0 = 1$.

Being an iterative mapping of the form $\sigma^{*(g+1)} = \tilde{f}_\sigma(\sigma^{*(g)}; \alpha)$, Eq. (38) allows for the illustration of the normalized mutation strength. In Fig. 9, the σ^* values are plotted using the control parameter $\alpha = 1.2$. Part (a) displays the normalized mutation strength dynamics as a function of g. One observes that the normalized mutation strength dynamics approach a limit cycle, i.e., the σ^* values exhibit an oscillatory behavior. Illustrating the iterative mapping (38) as a function of σ^*, Fig. 9 (b) visualizes the unstable

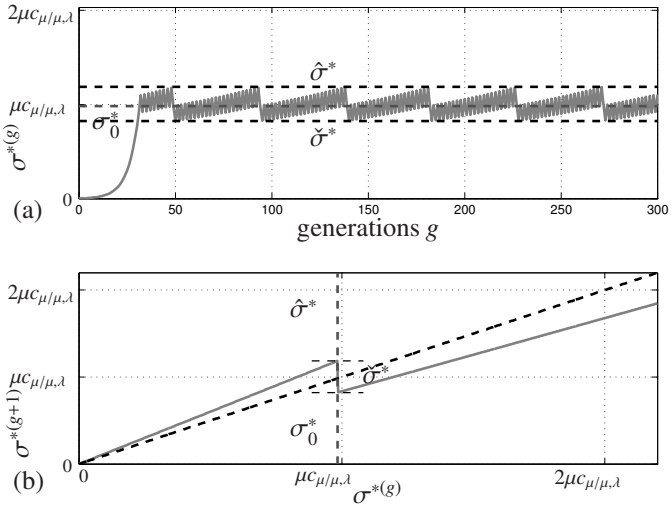

Figure 9: The σ^* dynamics resulting from the iterative mapping (38) considering the $[1, 2(3/3_I, 10)^1]$-Meta-ES with $N = 40$ on the ellipsoid model $a_i = i$ and $\alpha = 1.2$. The limit cycle in part (a) corresponds to the unstable fixed point observed in (b), i.e., the σ^* values oscillate between $\hat{\sigma}^*$ and $\check{\sigma}^*$.

fixed point $\sigma^* = \sigma^*_f$ that corresponds to the limit cycle displayed in (a). The σ^* dynamics do not intersect the dashed line $\sigma^{*(g+1)} = \sigma^{*(g)}$. Instead in (38) we observe a point of discontinuity at $\sigma^* = \sigma^*_0$. This corresponds to the change of increasing σ to decreasing σ from generation g to generation $g+1$. The transition point is the root of the discriminant function $\Delta(\sigma^*)$. Neglecting the ϵ_Δ term within the discriminator function $\Delta(\sigma^*)$, cf. (19), the point of discontinuity σ^*_0 is obtained by solving $\Delta(\sigma^*) \stackrel{!}{=} 0$

$$\sigma^*_0 = \frac{2\mu c_{\mu/\mu,\lambda}\,\alpha}{\alpha^2+1}. \tag{39}$$

Notice, for the sphere model σ^*_0 has already been derived in [10] and the result is equal to the one in Eq. (39).

Taking a look at Fig. 9, we refer to the upper limit of the oscillations as $\hat{\sigma}^*$ and to the lower limit as $\check{\sigma}^*$. That is, after reaching the limit cycle attractor, the normalized mutation strength dynamics are confined in an oscillation interval $[\check{\sigma}^*, \hat{\sigma}^*]$. The interval boundaries are derived by use of the iterative mapping (38).

Considering the point of discontinuity (39), the left-sided limit of the σ^* oscillation interval is obtained as

$$\begin{aligned}\hat{\sigma}^* &:= \lim_{\sigma^* \to \sigma^*_0 - 0} \tilde{f}_\sigma(\sigma^*) = \frac{\sigma^*_0 \alpha}{\sqrt{1 - \frac{2c_{\mu/\mu,\lambda}\alpha\sigma^*_0 \check{a}}{\Sigma a}}} \\ &\simeq \frac{2\mu c_{\mu/\mu,\lambda}}{1+\alpha^2} \frac{\alpha^2}{\sqrt{1 - 4\mu c^2_{\mu/\mu,\lambda} \frac{\alpha^2}{1+\alpha^2} \frac{\check{a}}{\Sigma a}}},\end{aligned} \tag{40}$$

and the right-sided limit is calculated as

$$\begin{aligned}\check{\sigma}^* &:= \lim_{\sigma^* \to \sigma^*_0 + 0} \tilde{f}_\sigma(\sigma^*) = \frac{\sigma^*_0}{\alpha\sqrt{1 - \frac{2c_{\mu/\mu,\lambda}\sigma^*_0 \check{a}}{\alpha\Sigma a}}} \\ &\simeq \frac{2\mu c_{\mu/\mu,\lambda}}{1+\alpha^2} \frac{1}{\sqrt{1 - 4\mu c^2_{\mu/\mu,\lambda} \frac{1}{1+\alpha^2} \frac{\check{a}}{\Sigma a}}}.\end{aligned} \tag{41}$$

The limits $\hat{\sigma}^*$ and $\check{\sigma}^*$ of the steady state oscillation interval are plotted against the control parameter α in Fig. 10. They are displayed by the solid gray lines. The point of discontinuity σ^*_0 resides within these limits and is indicated by the dashed line. In order to validate the predicted limits they are compared to measurements of the normalized mutation strength relying on both the iteratively generated dynamics and experimental runs of the $[1, 2(3/3_I, 10)^1]$-Meta-ES algorithm. These measurements are displayed by the error bars and connected by dotted lines. A single data point represents the mean value of σ^* realized by the Meta-ES after having approached its limit cycle attractor. It is obtained by running the Meta-ES algorithm for t_{max} isolation periods of $\gamma = 1$ generations using the respective α value to control the mutation strength. Choosing t_{max} sufficiently large to ensure that the strategy is operating in its steady state limit cycle the mean and standard deviation of the σ^* values are measured over the last 25% of the isolation periods. The gray "★" error bar plot illustrates the normalized mutations

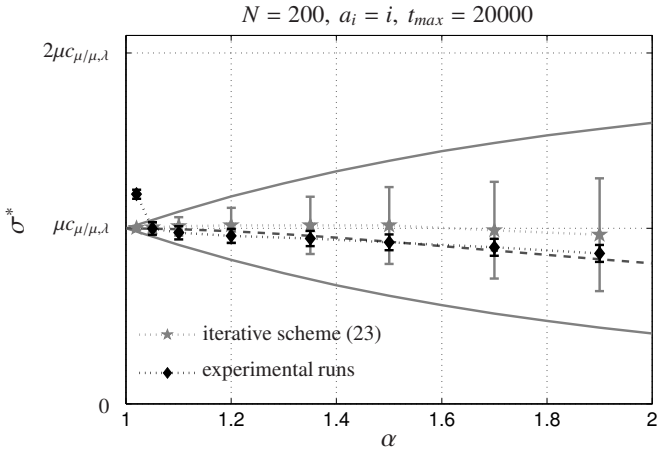

Figure 10: On the influence of α on the steady state normalized mutation strength σ^*. Represented by the solid gray lines, the limits (40) and (41) of the iterative mapping (38) are plotted against the control parameter α. The gray error bar plot illustrates the mean of the normalized mutations strength with corresponding standard deviation realized by iteration of system (23). The results of experimental Meta-ES runs averaged over 100 generations are displayed by the black error bars. The dashed line depicts the σ_0^* values.

strength realized by iteration of system (23). The standard deviation of the iterative results grows with the magnitude of α. The corresponding measurements from experimental runs of the $[1, 2(3/3_I, 10)^1]$-Meta-ES algorithm are displayed by the black "♦" error bar plot. By averaging over 100 independent runs the effect of growing standard deviations is reduced. The error bars corresponding to the experiments indicate the position of the mean values after averaging over multiple Meta-ES runs instead of displaying the actual standard deviation of the σ^* dynamics resulting from a single Meta-ES run.

As a matter of course, the iteratively generated results always reside in the oscillation interval. Due to the huge influence of the ϵ_Δ fluctuations that are not incorporated into the iterative model (23) the measurements from the experimental runs deviate significantly from the predicted oscillation interval for $\alpha = 1.02$. Regarding smaller search space dimensions N the influence of the fluctuations can be observed up to values about $\alpha = 1.2$. The impact of the ϵ_Δ fluctuations has been discussed in Sec. 4.2.

Interestingly, with exception of the experimental results corresponding to very small α values the measured mean values are remarkably close to the point of discontinuity σ_0^*. Hence, we deduce that the magnitude of the control parameter α has only minor influence on the expected normalized steady state mutation strength and thus on the overall performance of the Meta-ES. That is, operating with larger α parameters improves the ability to discriminate the two inner strategies without degrading the algorithm's performance considerably.

The distribution of the experimentally obtained normalized steady state mutation strength is illustrated using histogram plots in Fig. 11. The figure displays four choices of the control parameter α on the ellipsoid model $a_i = i$ in dimension $N = 200$ using

the same configuration as those displayed in Fig. 10. The distributions of the σ^* values are measured considering the final 25% generations obtained from the experimental Meta-ES runs. Regarding the distribution corresponding to $\alpha = 1.02$ the influence of the ϵ_Δ fluctuations is observable again. The mean value deviates significantly from the point of discontinuity and the measurements are not evenly distributed around their mean value. In contrast to the $\alpha = 1.02$ case, for $\alpha \geq 1.1$ the deviations between the mean value of the measurements and σ_0^* get reduced. That is, the theoretical model fits more accurately to the experiments.

4.4. The Steady State Dynamics

The point of discontinuity σ_0^* always resides within the oscillation interval $[\check{\sigma}^*, \hat{\sigma}^*]$ of the normalized mutation strength. For instance, in Fig. 10 it is represented by the dashed line. According to Eq. (39), σ_0^* depends on the population size of the Meta-ES, as well as on the control parameter α. With growing α the point of discontinuity slowly decreases. It governs the σ adaptation of the Meta-ES in the sense that the strategy will increase the mutation strength as long as $\sigma^* < \sigma_0^*$. On the other hand the Meta-ES will instantly reduce σ when the condition $\sigma_0^* < \sigma^*$ is fulfilled. This behavior results in the limit cycles around σ_0^* that are observed in Fig. 9.

As displayed in Fig. 10, after having approached its steady state the normalized mutation strength of the Meta-ES fluctuates in limit cycles around the point of discontinuity. Operating in this limit cycle around σ_0^* the strategy's expected normalized mutation strength will not exceed values of $\sigma_0^*\alpha$, neither will it fall below values of σ_0^*/α. Notice, considering small α parameters the described behavior is disturbed by the huge impact of the ϵ_Δ fluctuations mentioned in Sec. 4.2. However, in the range

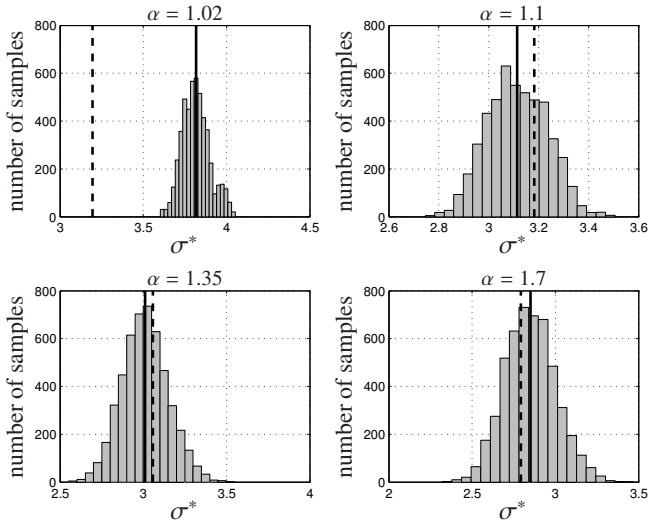

Figure 11: Illustration of the distribution of σ^* after the strategy has approached its steady state behavior. The solid vertical line corresponds to the measured mean value. The dashed vertical line displays the point of discontinuity σ_0^* according to Eq. (39).

of control parameters $\alpha \geq 1.2$ the Δ fluctuations are rather small. Further, the histogram plots in Fig. 11 suggest that in the steady state the σ^* values are sort of evenly distributed around their mean value. For $\alpha \geq 1.2$ this mean value is located near σ_0^*. Thus the Meta-ES on average operates with a normalized mutation strength in close proximity to the point of discontinuity. This gives the motivation to model the steady state dynamics of the normalized mutation strength in terms of a mean value $\overline{\sigma}_{ss}^*$ and corresponding fluctuation parts $\epsilon_{\sigma_{ss}^*}$

$$\sigma_{ss}^* \simeq \overline{\sigma}_{ss}^* + \epsilon_{\sigma_{ss}^*}. \tag{42}$$

Assuming that the mean value $\overline{\sigma}_{ss}^*$ dynamics in Eq. (42) are characterized sufficiently well by the point of discontinuity σ_0^* allows for the approximation of the steady state behavior. By omission of the noise term $\epsilon_{\sigma_{ss}^*}$ one obtains

$$\sigma_{ss}^* \simeq \sigma_0^* = 2\mu c_{\mu/\mu,\lambda} \frac{\alpha}{1+\alpha^2}. \tag{43}$$

Hence, the steady state y_i^2 dynamics of the $[1, 2(\mu/\mu_I, \lambda)^1]$-Meta-ES can be derived using $\varphi_i^{II*}(\sigma_{ss}^*)$, cf. Eq. (10). Accordingly, ignoring the fluctuations the one-generation change in the component-wise distance to the optimizer in the strategy's steady state is approximated by

$$y_i^{(g+1)^2} \simeq y_i^{(g)^2} - \varphi_i^{II*}(\sigma_{ss}^*). \tag{44}$$

Taking into account (10), we obtain the difference equation

$$y_i^{(g+1)^2} \simeq y_i^{(g)^2} - 2\sigma_{ss}^* c_{\mu/\mu,\lambda} \frac{a_i}{\Sigma a} y_i^{(g)^2} + \frac{{\sigma_{ss}^*}^2}{\mu} \frac{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}{(\Sigma a)^2}. \tag{45}$$

Making use of the assumption (43) transforms (45) into

$$\begin{aligned} y_i^{(g+1)^2} \simeq y_i^{(g)^2} &- 4\mu c_{\mu/\mu,\lambda}^2 \frac{\alpha}{(1+\alpha^2)} \frac{a_i}{\Sigma a} y_i^{(g)^2} \\ &+ 4\mu c_{\mu/\mu,\lambda}^2 \frac{\alpha^2}{(1+\alpha^2)^2} \frac{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}{(\Sigma a)^2}. \end{aligned} \tag{46}$$

Having obtained the steady state approximation of the y_i^2 dynamics, the next task addresses the corresponding σ evolution. Inserting assumption (43) into the mutation strength normalization (8) the σ dynamics are calculated as

$$\sigma^{(g)} = \sigma_{ss}^* \frac{R_a(\mathbf{y}^{(g)})}{\Sigma a} = 2\mu c_{\mu/\mu,\lambda} \frac{\alpha}{1+\alpha^2} \frac{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}{\Sigma a}. \tag{47}$$

Consequently, the equations (46) and (47) build an iterative scheme describing the system's steady state evolution behavior

$$\begin{aligned} y_i^{(g+1)^2} &\simeq y_i^{(g)^2} - \frac{4\mu c_{\mu/\mu,\lambda}^2 \alpha}{(1+\alpha^2)\Sigma a} \left(a_i y_i^{(g)^2} - \frac{\alpha \sum_{j=1}^N a_j^2 y_j^{(g)^2}}{(1+\alpha^2)\Sigma a} \right), \\ \sigma^{(g+1)} &\simeq \frac{2\mu c_{\mu/\mu,\lambda} \alpha}{1+\alpha^2} \frac{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g+1)^2}}}{\Sigma a}. \end{aligned} \tag{48}$$

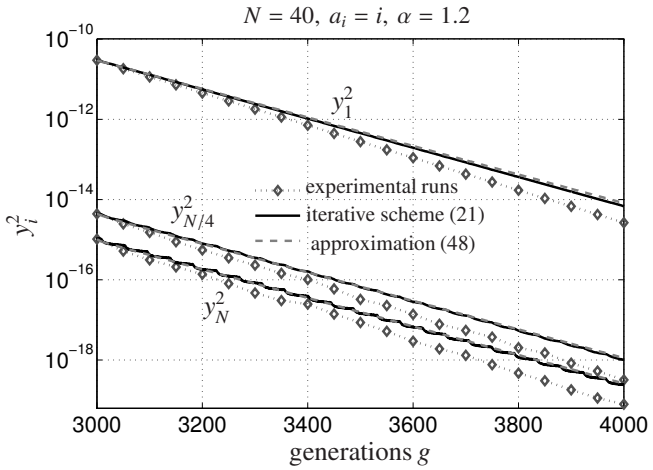

Figure 12: Validation of the approximated steady state dynamics according to (48). The dynamics of the $[1, 2(3/3_I, 10)^1]$-Meta-ES with control parameter $\alpha = 1.2$ on the ellipsoid model $a_i = i$ in dimension $N = 40$ are illustrated starting at $\sigma^0 = 1$ and $\mathbf{y}^{(0)} = \mathbf{1}$. The experimental curves are averaged over 10^3 independent Meta-ES runs.

The steady state dynamics resulting from system (48) are validated in Fig. 12. Therefore, they are compared to the iterative dynamics resulting system (23), as well as to experimental runs of the Meta-ES algorithm. In order to ensure that the Meta-ES has reached its steady state the iterative system (23) is iterated over a sufficiently large number of generations. Within the steady state the iterative dynamics are compared to the analytically obtained steady state predictions (48).

Figure 12 illustrates the y_i^2 dynamics after applying the $[1, 2(3/3_I, 10)^1]$-Meta-ES using control parameter $\alpha = 1.2$ on the ellipsoid model $a_i = i$ in dimension $N = 40$. All real Meta-ES runs are initialized in the iteratively generated state after 3000 generations in order to shorten the transient phase. Even in rather low search space dimension the predicted steady state dynamics (48) show a good agreement with the iterative dynamics resulting from system (21). That is, the modeling of the steady state dynamics only slightly deviates from the initial iterative model (23). The agreement of the predictions and the experimentally obtained dynamics improves with growing dimensionality. Having validated the approximation quality the system (48) can be used to analyze the steady state dynamics.

The typical longterm behavior of the Meta-ES dynamics is observed in Fig. 3 and Fig. 12, respectively. The illustrations suggest that the y_i^2 and σ dynamics in their steady state exhibit a log-linear decline. Thus the system (48) can be solved using the following exponential *Ansatz*:

$$y_i^{(g)^2} = b_i \mathrm{e}^{-\nu g}, \quad b_i > 0, \nu > 0. \tag{49}$$

This *Ansatz* was first introduced in [8] considering the analysis of the σSA-ES on the ellipsoid model. As the mutation strength evolution (48) depends on the y_i^2 dynamics,

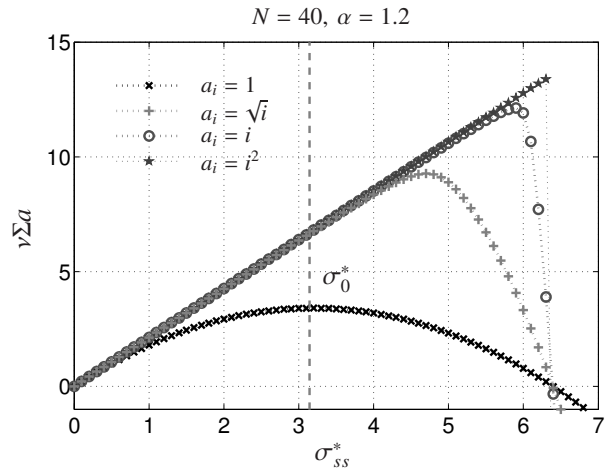

Figure 13: Numerical solutions of the eigenvalue problem (51) plotted against the normalized steady state mutation strength σ^*_{ss}. The results of four different ellipsoid models, $a_i = 1$, $a_i = \sqrt{i}$, $a_i = i$, and $a_i = i^2$, are displayed using the $[1, 2(3/3_I, 10)^1]$-Meta-ES with $\alpha = 1.2$ in search space dimension $N = 40$. The point of discontinuity (39) is illustrated by the vertical dashed line.

it can be expressed in terms of (49)

$$\begin{aligned}\sigma^{(g)} &= \frac{2\mu c_{\mu/\mu,\lambda}\alpha}{1+\alpha^2}\frac{\sqrt{\sum_{j=1}^{N} a_j^2 y_j^{(g)^2}}}{\Sigma a} \\ &= \frac{2\mu c_{\mu/\mu,\lambda}\alpha}{1+\alpha^2}\frac{\sqrt{\sum_{j=1}^{N} a_j^2 b_j}}{\Sigma a}\mathrm{e}^{-\frac{\nu}{2}g}.\end{aligned} \tag{50}$$

By use of (49) the system's state at generation $g+1$ is directly connected to the state at generation g, e.g., $y_i^{(g+1)^2} = b_i \mathrm{e}^{-\nu g}\mathrm{e}^{-\nu} = y_i^{(g+1)^2}\mathrm{e}^{-\nu}$. Since the slopes of the y_i^2 dynamics are rather small, the term $\mathrm{e}^{-\nu}$ can be simplified to $\mathrm{e}^{-\nu} = 1 - \nu + O(\nu^2)$ using Taylor expansion around zero. Applying this operations to the first equation of system (48) using the representation from (45) yields the eigenvalue problem

$$\nu b_i \simeq 2\sigma^*_{ss} c_{\mu/\mu,\lambda}\frac{a_i}{\Sigma a} b_i - \frac{{\sigma^*_{ss}}^2}{\mu}\frac{\sum_{j=1}^{N} a_j^2 b_j}{(\Sigma a)^2}, \tag{51}$$

where ν is the eigenvalue and the b_i are the components of the corresponding eigenvector.

Although the approximation of the Meta-ES steady state is only applicable assuming that the Meta-ES is operating with a normalized mutation strength in close proximity to the point of discontinuity, i.e., $\sigma^*_{ss} \simeq \sigma^*_0$, the eigenvalue problem can be solved numerically for general values of the normalized steady state mutation strength σ^*_{ss}. This gives an impression how the smallest eigenvalue depends on the normalized mutation strength σ^*. Taking into account the ellipsoid models $a_i = 1$, $a_i = \sqrt{i}$, $a_i = i$ and $a_i = i^2$, after multiplication with Σa the resulting eigenvalues are presented in Fig. 13.

Except for the sphere model ($a_i = 1$) the curves of the other ellipsoid models show a similar behavior. For the non-spherical ellipsoid models the illustration reveals that the eigenvalues increase linearly for a wide range of σ^* values including the vicinity of $\sigma^* = \sigma_0^* = \sigma_{ss}^*$.

That is, the eigenvalue can be approximated considering only the linear parts within the eigenvalue problem (51). Thus the quadratic term in (51) can be neglected and the eigenvalue problem simplifies considerably

$$\nu b_i \simeq 2\sigma_{ss}^* c_{\mu/\mu,\lambda} \frac{a_i}{\Sigma a} b_i. \tag{52}$$

Inserting $\sigma_{ss}^* \simeq \sigma_0^*$ into this representation the eigenvalues are directly obtained as

$$\nu \simeq 4\mu c_{\mu/\mu,\lambda}^2 \frac{\alpha}{(1+\alpha^2)} \frac{a_i}{\Sigma a}. \tag{53}$$

The *Ansatz* (49) indicates that larger values of ν result in a faster decline of the y_i^2 dynamics. Regarding the steady state dynamics, for $g \to \infty$, the impact of the larger ν values is neglectable compared to the smallest eigenvalue ν. The smallest positive eigenvalue corresponds to the slowest mode of the Meta-ES dynamics and consequently governs the steady state.

Denoting the smallest ellipsoid coefficient $\breve{a} := \min(a_i)$, the steady state mode eigenvalue is derived as

$$\boxed{\nu \simeq 4\mu c_{\mu/\mu,\lambda}^2 \frac{\alpha}{(1+\alpha^2)} \frac{\breve{a}}{\Sigma a}.} \tag{54}$$

The quality of formula (54) is verified by comparison to experimental $[1, 2(3/3, 10)^1]$-Meta-ES runs considering different choices of the parameter α. To this end, 100 independent Meta-ES runs are performed with fixed normalized steady state mutation strength $\sigma_{ss}^* = \sigma_0^* = 2\mu c_{\mu/\mu,\lambda}\alpha/(1+\alpha^2)$. The mutation strength $\sigma^{(g)}$ is obtained by renormalization in each generation. Then, a linear polynomial $\ln y_i^2 = -\nu g + \ln b_i$ is fitted to the averaged y_i^2 data points yielding N experimental eigenvalues ν, i.e., one for each y_i^2 curve. The deviations between these N eigenvalues are very small. For instance, on the ellipsoid model $a_i = i$ with $N = 40$ the smallest ν varies from the remaining $N-1$ eigenvalues by a maximum of 2%. Accordingly, the ν values corresponding to the y_1^2 dynamics are plotted in Fig. 14 (scaled up by multiplication with Σa) and compared to the analytical prediction in (54). Due to the normalization with Σa the prediction does neither depend on the search space dimensionality nor on the ellipsoid coefficients a_i.

Taking into account the magnitude of the corresponding normalization factor Σa, the predictions show a good agreement with the experimental data (the maximal relative error is less than 6%). On the ellipsoid model $a_i = i$ the rather large deviations for $N = 40$ decrease with growing search space dimensionality. Also considering the ellipsoid model $a_i = i^2$ improves the agreement of the data points with the theoretical points. The impact of the control parameter α on the steady state mode eigenvalue is rather small. On the considered interval of α parameters the resulting eigenvalues

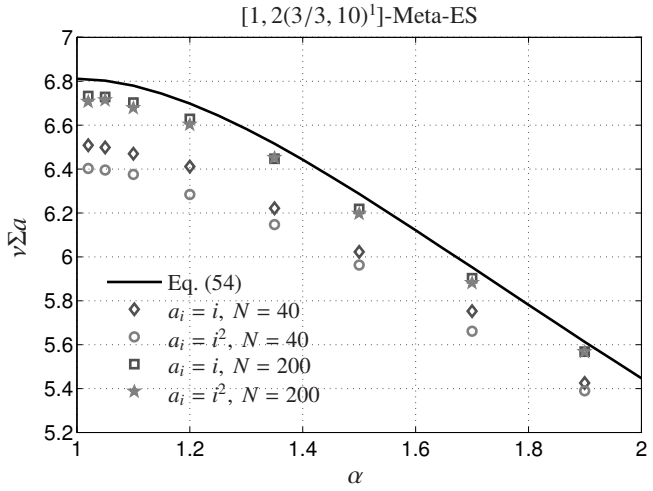

Figure 14: Product of the eigenvalues (54) and the sum of the ellipsoid coefficients Σa plotted against the interval of control parameters α. The experimentally obtained values on the ellipsoid models $a_i = i$, and $a_i = i^2$, are displayed as data points in search space dimensions $N = 40$ and $N = 200$.

differ in a maximal factor of about $1.4/\Sigma a$. In consequence of the huge fluctuations for small α that are observable in the experimental runs of the Meta-ES algorithm, selecting $\alpha = 1.2$ seems to be an appropriate choice to efficiently control σ.

From Fig. 13 it is evident that the linear approximation of the eigenvalue problem (52) is not applicable to the sphere model ($a_i = 1$). Thus Eq. (54) is unqualified for the prediction of the steady state progress on the sphere. Nevertheless, a prediction of the steady state mode eigenvalue on the sphere model can be obtained directly from Eq. (51). Requiring $\forall i : \; a_i = 1$ in (51) and taking the sum over all $i = 1, \ldots, N$ yields after minor rearrangements

$$\nu \simeq \frac{2c_{\mu/\mu,\lambda}\sigma^*_{ss}}{N} - \frac{{\sigma^*_{ss}}^2}{\mu N}. \tag{55}$$

Considering the normalized steady state mutation strength approximation $\sigma^*_{ss} \simeq \sigma^*_0$, and considering (39), one obtains the steady state mode eigenvalue

$$\boxed{\nu \simeq \frac{4\mu c^2_{\mu/\mu,\lambda}}{N}\frac{\alpha}{1+\alpha^2}\left(1 - \frac{\alpha}{1+\alpha^2}\right).} \tag{56}$$

This formula represents an estimate for the steady state progress of the $[1, 2(\mu/\mu_I, \lambda)^\gamma]$-Meta-ES on the sphere model ($a_i = 1$). Notice, that Eq. (56) is only accessible provided that the Meta-ES operates with a normalized steady state mutation strength in the proximity of σ^*_0. On the sphere model, this holds true if the mutation strength control parameter satisfies $\alpha > \alpha_0$. The critical value α_0 has already been derived in [10] as

$$\alpha_0 = 1 + \frac{\mu c^2_{\mu/\mu,\lambda}}{2N - \mu c^2_{\mu/\mu,\lambda}}. \tag{57}$$

Given the $\alpha < \alpha_0$ case, the normalized mutation strength of the $[1, 2(\mu/\mu_I, \lambda)^1]$-Meta-ES on the sphere model approaches a stable fixed point. This stable fixed point resides in the vicinity of $\sigma^* \lessapprox 2\mu c_{\mu/\mu,\lambda}$. According to [10], it is not possible to tune α adequately to shift this stable fixed point to the optimal normalized mutation strength $\sigma^*_{opt} = \mu c_{\mu/\mu,\lambda}$ on the sphere model. Thus (56) represents an description of the steady state progress provided that the Meta-ES operates close to its optimal performance.

4.5. Expected Running Time

Having derived the steady state mode eigenvalue allows for the calculation of the expected running time of the Meta-ES algorithm. Therefore, the steady state fitness dynamics are determined by use of the Eqs. (2) and (49) starting from generation g_0 for an evolution interval g

$$F(\mathbf{y}^{(g_0+g)}) = \sum_{j=1}^{N} a_j^2 b_j \mathrm{e}^{-\nu(g_0+g)} = F(\mathbf{y}^{(g_0)})\mathrm{e}^{-\nu g}. \tag{58}$$

That is, in the strategy's steady state the objective function drops exponentially fast with time constant $\tau = 1/\nu$. The representation of the fitness dynamics provides an estimate for the expected running time G of the algorithm needed to improve the objective function value by a factor $2^{-\beta}$. Notice, that G in the context of Meta-ES refers to the iteration number of the outer strategy. From (58) one obtains

$$2^{-\beta} = \frac{F(\mathbf{y}^{(g_0+G)})}{F(\mathbf{y}^{(g_0)})} = \mathrm{e}^{-\nu G}, \tag{59}$$

and by taking the logarithm and applying (54), G becomes

$$\boxed{G \simeq \frac{\beta \ln(2)}{4\mu c^2_{\mu/\mu,\lambda}} \frac{(1+\alpha^2)}{\alpha} \frac{\Sigma a}{\check{a}}.} \tag{60}$$

Within each iteration the outer strategy performs 2λ objective function evaluations. Thus the expected number of function evaluations needed to improve the fitness value by a factor of $2^{-\beta}$ is proportional to G. Considering the $[1, 2(3/3_I, 10)^1]$-Meta-ES on the ellipsoid models $a_i = i, i^2$, it is displayed in Fig. 15 for different search space dimensions N. Using control parameter $\alpha = 1.2$ the theoretical predictions for $\beta = 2$ are verified using real ES runs. All experimental results are averaged over 100 independent ES runs and illustrated by error bars with their corresponding standard deviations. One observes a good agreement between experimental results and theoretical predictions. The standard deviation decreases with growing dimensionality N. Compared to the $(3/3, 10)$-CSA-ES investigated in [13] with standard parameters $c = \frac{1}{\sqrt{N}}$, $D = \frac{1}{c}$, the Meta-ES in expectation needs about twice the number of function evaluations.

The approximated expected running time G is asymptotically proportional to the quotient of the sum of the ellipsoid coefficients Σa and the smallest coefficient $\check{a}$. In particular, regarding the fitness model (2) the result can be extended to the general fitness model $F(\boldsymbol{y}) = \boldsymbol{y}^\top \boldsymbol{Q}\boldsymbol{y}$ with positive definite matrix $\boldsymbol{Q} \in \mathbb{R}^{N\times N}$. In this situation, $\check{a} = \min(a_i)$ is identified with the smallest eigenvalue $\check{\kappa}$ of the corresponding eigenvalue problem $\boldsymbol{Q}\boldsymbol{u} = \kappa\boldsymbol{u}$. As the trace of $\boldsymbol{Q}$ consists of the sum of its eigenvalues we obtain

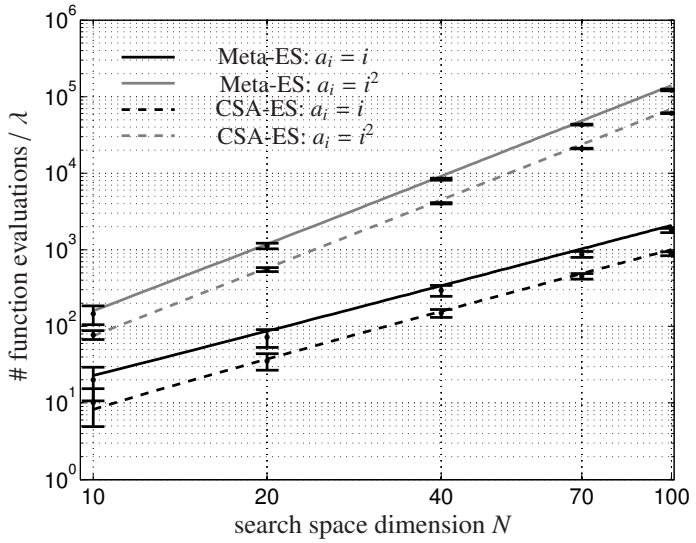

Figure 15: The expected number of function evaluations per λ plotted against the search space dimension from $N = 10$ to $N = 100$. The $[1, 2(3/3, 10)^1]$-Meta-ES uses the mutation strength control parameter $\alpha = 1.2$. The theoretical results are displayed by solid lines. For comparison, the corresponding predictions of the $(3/3, 10)$-CSA-ES are provided as dashed lines. All experimentally obtained data on the ellipsoid models $a_i = i$, and $a_i = i^2$, are displayed by the error bar plots.

$\Sigma a = \mathrm{Tr}[\boldsymbol{Q}]$. That is, considering the $[1, 2(\mu/\mu_I, \lambda)^\gamma]$-Meta-ES on the general fitness model yields: $G \propto \mathrm{Tr}[\boldsymbol{Q}]/\check{\kappa}$. This means that the expected running time of the two mentioned ellipsoid models increases with order N^2 for $a_i = i$, and with N^3 for $a_i = i^2$, respectively.

5. Conclusion

The paper applies the dynamical systems analysis approach which was developed for the $(\mu/\mu_I, \lambda)$-σSA-ES in [8] to the analysis of mutation strength adaptation via Meta-ES on the ellipsoid model. To this end, the evolution equations for the considered $[1, 2(\mu/\mu_I, \lambda)^1]$-Meta-ES have been derived. They allow for the derivation of the normalized mutation strength evolution. For the normalized mutation strength dynamics on the ellipsoid model we distinguish a transient and a "steady state" like phase. The latter is characterized by an oscillating behavior around the point of discontinuity σ_0. By identifying the point of discontinuity with the mean value dynamics of the normalized mutation strength one is able to describe the expected log-linear longterm behavior of the Meta-ES algorithm. Finally, an estimate concerning the expected running time of the Meta-ES has been presented.

Similar results have been obtained for the $(\mu/\mu_I, \lambda)$-σSA-ES in [8] and for the $(\mu/\mu_I, \lambda)$-CSA-ES in [13], respectively. However, depending on the choice of the specific strategy parameters these strategies realize different normalized steady state mutation strengths σ^*_{ss}. The optimal normalized mutation strength on the ellipsoid model

resides in the interval $[\mu c_{\mu/\mu,\lambda}, 2\mu c_{\mu/\mu,\lambda})$. Beyer and Melkozerov [8] showed that the σ-SA-ES performs sub-optimally on non-spherical ellipsoid models approaching σ^*_{ss} in the vicinity of 1 when using standard parameter settings. In contrast to σ-SA, mutation strength control by CSA, cf. [13], for typical parameter choices yields a normalized steady state mutation strength proportional to the number of parents, i.e., $\sigma^*_{ss} \propto \mu c_{\mu/\mu,\lambda}$. Although it is unlikely in black-box optimization scenarios, given knowledge of the fitness landscape the performance of the σ-SA-ES can be tuned in direction of the CSA-ES performance. In the same way choosing the strategy parameters of the CSA-ES appropriate to the considered fitness environment could possibly improve its performance.

On the ellipsoid model, the $[1, 2(\mu/\mu_I, \lambda)^1]$-Meta-ES turns out to be more robust to the choice of the control parameter α since it has only minor influence on the overall performance. Using the standard choice of the control parameter, $\alpha = 1.2$, the Meta-ES approaches a normalized steady state mutation strength close to $\sigma^*_{ss} = \mu c_{\mu/\mu,\lambda}$. That is, in expectation the Meta-ES realizes a σ^*_{ss} similar to the CSA-ES. But employing two inner ESs in each isolation period the Meta-ES requires twice the number of function evaluations to achieve about the same progress.

Notice, that (60) is unsuitable for the sphere model because the derivation is based on the use of the linear approximation of the steady state mode eigenvalue (49) which is not applicable to the sphere model (cf. Fig. 13). The analysis of the $[1, 2(\mu/\mu_I, \lambda)^\gamma]$-Meta-ES on the sphere model was carried out by Beyer *et al.* [10] using a different approach. In contrast to the non-spherical ellipsoid model, the optimal normalized mutation strength on the sphere model is $\sigma^* = \mu c_{\mu/\mu,\lambda}$. On the sphere the Meta-ES approaches a stable normalized mutation strength $\sigma^* > \mu c_{\mu/\mu,\lambda}$ considering sufficiently small parameters α. For larger α values σ^* fluctuates in limit cycles around the point of discontinuity σ^*_0 close to $\mu c_{\mu/\mu,\lambda}$. Thus the Meta-ES almost adapts the optimal σ on the sphere model provided that α is chosen appropriately. In this situation, we were able to provide a formula that allows for the prediction of the corresponding steady state progress.

Although the work considers only the $\gamma = 1$ case the presented analysis approach can be extended to longer isolation periods $\gamma > 1$. In that situation, one is able to provide evidence that larger γ values reduce the occurring ϵ_Δ fluctuations and thus improve the detectability of the best inner ES even for rather small α values. This generalization is regarded as the next step within the analysis of the Meta-ES for mutation strength adaptation.

Furthermore, the analysis approach should also be able to address the problem of controlling other strategy parameters of the inner $(\mu/\mu_I, \lambda)$-ES such as the parental population size μ. This way the analysis might be extended to investigate noisy fitness functions. A first step into this direction has been presented in [12] considering the noisy sphere model. Another reasonable approach could be the application of Meta-ES to dynamically changing fitness functions.

Acknowledgements

This work was supported by the Austrian Science Fund FWF under grant P22649-N23.

References

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

[2] I. Rechenberg, Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution, Problemata, 15, Frommann-Holzboog, 1973.

[3] H.-P. Schwefel, Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie, Interdisciplinary systems research; 26, Birkhäuser, Basel, 1977.

[4] A. Ostermeier, A. Gawelczyk, N. Hansen, Step-Size Adaption Based on Non-Local Use of Selection Information, in: Parallel Problem Solving from Nature - PPSN III, Jerusalem, Israel, October 9-14, 1994, pp. 189–198.

[5] I. Rechenberg, Evolutionsstrategien 8 (1978) 83–114.

[6] M. Herdy, Reproductive Isolation as Strategy Parameter in Hierarichally Organized Evolution Strategies, in: Parallel Problem Solving from Nature 2, PPSN-II, Brussels, Belgium, September 28-30, 1992, pp. 207–217.

[7] I. Rechenberg, Evolutionsstrategie 94, Vol. 1 of Werkstatt Bionik und Evolutionstechnik, Frommann-Holzboog, Stuttgart, 1994.

[8] H.-G. Beyer, A. Melkozerov, The Dynamics of Self-Adaptive Multi-Recombinant Evolution Strategies on the General Ellipsoid Model, IEEE Transactions on Evolutionary Computation 18 (5) (2014) 764–778.

[9] D. V. Arnold, A. MacLeod, Step length adaptation on ridge functions, Evolutionary Computation 16 (2) (2008) 151–184.

[10] H.-G. Beyer, M. Dobler, C. Hämmerle, P. Masser, On strategy parameter control by Meta-ES, in: Genetic and Evolutionary Computation Conference, GECCO 2009, Proceedings, Montreal, Québec, Canada, July 8-12, 2009, pp. 499–506.

[11] H.-G. Beyer, M. Hellwig, Mutation strength control by Meta-ES on the sharp ridge, in: Genetic and Evolutionary Computation Conference, GECCO '12, Philadelphia, PA, USA, July 7-11, 2012, pp. 305–312.

[12] H.-G. Beyer, M. Hellwig, Controlling population size and mutation strength by Meta-ES under fitness noise, in: Foundations of Genetic Algorithms XII, FOGA '13, Adelaide, SA, Australia, January 16-20, 2013, pp. 11–24.

[13] H.-G. Beyer, M. Hellwig, The Dynamics of Cumulative Step Size Adaptation on the Ellipsoid Model, Evolutionary Computation (2015) 1–33.

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

[15] S. Meyer-Nieberg, H.-G. Beyer, The Dynamical Systems Approach - Progress Measures and Convergence Properties., in: G. Rozenberg, T. Bäck, J. N. Kok (Eds.), Handbook of Natural Computing, Springer, 2012, pp. 741–814.

[16] M. Herdy, The Number of Offspring As Strategy Parameter in Hierarchically Organized Evolution Strategies, SIGBIO Newsl. 13 (2) (1993) 2–9.

[17] A. Melkozerov, H.-G. Beyer, On the analysis of self-adaptive evolution strategies on elliptic model: first results., in: M. Pelikan, J. Branke (Eds.), GECCO, ACM, 2010, pp. 369–376.