The Dynamics of Cumulative Step-Size Adaptation on the Ellipsoid Model

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

Michael Hellwig michael.hellwig@fhv.at
Research Center Process and Product Engineering, Vorarlberg University of Applied Sciences, Dornbirn, 6850, Austria

Abstract

The behavior of the $(\mu/\mu_I, \lambda)$-Evolution Strategy (ES) with cumulative step-size adaptation (CSA) on the ellipsoid model is investigated using dynamical systems analysis. At first a non-linear system of difference equations is derived that describes the mean-value evolution of the ES. This system is successively simplified to finally allow for deriving closed-form solutions of the steady state behavior in the asymptotic limit case of large search space dimensionalities. It is shown that the system exhibits linear convergence order. The steady state mutation strength is calculated and it is shown that compared to standard settings in σ self-adaptive ESs, the CSA control rule allows for an approximately μ-fold larger mutation strength. This explains the superior performance of the CSA in non-noisy environments. The results are used to derive a formula for the expected running time. Conclusions regarding the choice of the cumulation parameter c and the damping constant D are drawn.

Keywords

Evolution strategies, ellipsoid model, cumulative step size adaptation, mutation strength

1 Introduction

The performance of Evolution Strategies (ESs) depends crucially on the optimal control of the mutation strength σ that determines the step-length of the search steps used to generate offspring from parents. There are basically four established methods to learn/control the mutation strength: Rechenberg's 1/5-rule (Rechenberg, 1973), σ-Self-Adaptation (σSA) (Rechenberg, 1973; Schwefel, 1977), Meta-ES (Herdy, 1992; Rechenberg, 1994), and cumulative step size adaptation (CSA) (Ostermeier et al., 1994). Understanding and analyzing 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. The analysis approach, which turned out to be the most fruitful one up until now, considers the "ES + objective function" as a *dynamical system* (Meyer-Nieberg and Beyer, 2012). That is, it is the goal of the analysis to determine the time evolution of the system. However, since ES are probabilistic algorithms, this analysis concerns the dynamics of stochastic, most often non-linear, systems. Due to the difficulties of such an analysis, progress in this field is rather slow. The first fully analyzed algorithm concerned the $(1, \lambda)$-σSA-ES on the sphere (Beyer, 1996). In the following years progress was 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

(PDQFs), see e.g. Arnold and Beyer (2004); Jägersküpper (2006); Beyer and Meyer-Nieberg (2006); Arnold (2007); Beyer and Finck (2010).

The most advanced dynamical systems analysis approach on ESs has been presented by Beyer and Melkozerov (2014) where the $(\mu/\mu_I, \lambda)$-σSA-ES on the ellipsoid model has been investigated. In that paper, a new progress measure was introduced to model the dynamics of the quadratic distances of the parental state to the optimizer: the *quadratic progress rate*. While that work completed the analysis of isotropic self-adaptive standard ES, a similar analysis of CSA σ control is not advanced that much. The fitness models considered so far concerned the cigar function (Arnold, 2007; Arnold and Beyer, 2010) and another special case of PDQFs consisting of a mixture of two sphere models (Beyer and Finck, 2010). These analyses were performed along the line developed in Arnold (2002). Due to the inherent symmetries of those fitness models (cigar and mixture of two spheres, respectively), the dynamics in the $\mathbb{R}^N$ search space can be lumped together, thus reducing the dynamics to a few state variables describing the approach to the optimizer. This aggregation does not only concern the parental state in the $\mathbb{R}^N$ search space, but it also concerns the *evolution path cumulation* that is used to measure the average length of the actually realized change steps between two consecutive parental states, which are used to control the mutation strength σ. As a result, the analysis presented by Beyer and Melkozerov (2014) for the $(\mu/\mu_I, \lambda)$-σSA-ES cannot be directly transferred to the corresponding CSA-ES. It is the aim of this work to extend the analysis method developed in (Beyer and Melkozerov, 2014) to handle the path cumulation in CSA-ES. To this end, we will deviate from the standard analysis developed in (Arnold, 2002) and fall back to the analysis method that has been successful in the analysis of σSA. That is, we will derive a *self-adaptation response function* (SAR) for the cumulative step-size adaptation. This approach allows for an analysis of the CSA-ES similar to that of the $(\mu/\mu_I, \lambda)$-σSA-ES in (Beyer and Melkozerov, 2014).

The derivation of a SAR function for CSA is challenging, since a SAR function for CSA in the proper meaning of the word does not exist. The original SAR function (Beyer, 2001) is an *one-generation* progress measure that describes the expected relative σ change from generation g to $g + 1$. Path cumulation is, however, a process that is *non-local* in time, i.e., it is the result of the cumulation process that incorporates a fading record of parental steps taken. As a result, a SAR function for CSA must necessarily be a quantity that depends not only on the damping constant D (which corresponds to the learning parameter τ in the σSA, for details see below), but also on the cumulation time constant $1/c$.

The analysis to be presented can be regarded as an important preparatory step for an analysis of the CMA-ES (Hansen and Ostermeier, 2001) since it provides a system of difference equations that describe the σ-evolution for ellipsoid models. Since CMA-ES transforms quadratic models via the square root of the covariance matrix in another quadratic model, this analysis does also hold for these transformed models. That is, the analysis presented is a building block for a complete analysis of the CMA-ES.

The paper is organized as follows. After a short recap of the ideas of CSA and the ellipsoid model, results of the analysis of the σ self-adaptive ES are reviewed as the basis for the derivations to be presented in the subsequent sections. Next, the CSA relevant path cumulation is analyzed in Section 5 yielding a system of recurrence equations that are combined in Section 6 to form a system of evolution equations describing the mean-value dynamics of the $(\mu/\mu_I, \lambda)$-CSA-ES. In a next step, simplifications are introduced in order to obtain a system of tractable evolution equations that allows for the calculation of a function that describes the expected generational σ change in the steady state, similar to the self-adaptation respond function used in the analysis of the σSA-ES. The resulting system of evolution equations will then be treated by an Ansatz in Section 7.2 similar to the one used in (Beyer and Melkozerov, 2014). The steady state of the normalized mutation strength dynamics is considered in Section 7.3 and the influence of

the strategy parameters c and D are discussed. In Section 8 the steady state mutation strength and the convergence rates are calculated in terms of closed-form expressions. The results are used to estimate the expected running time in Section 8.3. Conclusions are drawn in the final Section 9.

2 The $(\mu/\mu_{\mathbf{I}}, \lambda)$-CSA-ES Algorithm

The focus of this paper is on the dynamical behavior of the $(\mu/\mu_I, \lambda)$-CSA-ES. As indicated by the notation, non-elitist selection (","-selection) rather than elitist selection ("+"-selection) is considered. That is, in each generation only the offspring population is involved in the selection process. The "comma"-selection is advantageous in real-valued parameter optimization since it allows for the use of greater mutation strengths which is for example beneficial when optimizing multi-modal objective functions. The $(\mu/\mu_I, \lambda)$-CSA-ES controls the mutation strength, also referred to as step-size, within the algorithm by *cumulative step-size adaptation* (CSA), see (Ostermeier et al., 1994). CSA gathers previously successful search steps in a fading search path history. The mutation strength is then adjusted depending on the length of this search path. Another option to adapt the mutation strength of the ES is σ-self-adaptation (σSA). In contrast to CSA, it provides each offspring with an individual mutation strength which is computed from the recombined mutation strengths of the best μ offspring of the previous generation. The σSA has been investigated in (Beyer and Melkozerov, 2014).

The pseudo-code of the $(\mu/\mu_I, \lambda)$-CSA-ES is presented in Tab. 1. First the initial parental parameter vector, also referred to as the *parental centroid* $\mathbf{y}^{(0)}$, the initial mutation strength $\sigma^{(0)}$, and the initial search path $\mathbf{s}^{(0)} = \mathbf{0} \in \mathbb{R}^N$ are specified. Then from line 3 to 5 λ offspring $\mathbf{y}_l$ are generated by adding the product of the mutation strength σ and an N-dimensional random mutation vector $\mathbf{z}_l$ to the parental centroid. The components of each $\mathbf{z}_l$ are independent and identically distributed standard normal variates. The corresponding fitness function value F_l of each offspring is calculated in line 6. In line 8 the mutation vectors of the μ best offspring w.r.t. fitness are recombined to create their centroid $\langle \mathbf{z} \rangle^{(g)}$. As indicated by the index I in $(\mu/\mu_I, \lambda)$ this centroid is generated using intermediate recombination. In this context the subscript $m;\lambda$

Initialize$(\mathbf{y}^{(0)}, \mathbf{s}^{(0)}, \sigma^{(0)})$	Line
$g \leftarrow 0$;	1
Repeat	2
For $l = 1$ To λ	3
$\mathbf{z}_l \leftarrow \mathcal{N}(0, \mathbf{1})$;	4
$\mathbf{y}_l \leftarrow \mathbf{y}^{(g)} + \sigma^{(g)} \mathbf{z}_l$;	5
$F_l \leftarrow F(\mathbf{y}_l)$;	6
End For	7
$\langle \mathbf{z} \rangle^{(g)} \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} \mathbf{z}_{m;\lambda}$;	8
$\mathbf{y}^{(g+1)} \leftarrow \mathbf{y}^{(g)} + \sigma^{(g)} \langle \mathbf{z} \rangle^{(g)}$;	9
$\mathbf{s}^{(g+1)} \leftarrow (1-c)\mathbf{s}^{(g)} + \sqrt{\mu c(2-c)} \langle \mathbf{z} \rangle^{(g)}$;	10
$\sigma^{(g+1)} \leftarrow \sigma^{(g)} \exp\left(\frac{\|\mathbf{s}^{(g+1)}\|^2 - N}{2DN} \right)$;	11
$g \leftarrow g + 1$;	12
Until(termination condition)	13

Table 1: Pseudo code of the $(\mu/\mu_I, \lambda)$-CSA-ES.

denotes the mth best of the λ offspring (i.e., in case of minimization the offspring with the mth smallest fitness value). On the one hand, the centroid of the μ best mutation vectors is used in line 9 to compose a new parental centroid $\mathbf{y}^{(g+1)}$, and on the other hand, to update the search path $\mathbf{s}^{(g+1)}$ in line 10. This search path contains a fading record of the strategy's previous steps. The length of its memory is determined by the choice of the constant parameter $c \in (0, 1)$, referred to as the *cumulation parameter*. The mutation strength $\sigma^{(g+1)}$ is then updated in line 11 by multiplication with an exponential value depending on the length of the search path $\mathbf{s}$ as well as the *damping parameter* D. $D > 0$ is a constant parameter that determines the magnitude of the mutation strength updates. The sign of $\|\mathbf{s}^{(g+1)}\|^2 - N$ determines whether the mutation strength is increased or decreased. Long search paths indicate that the steps made by the ES collectively point in one direction and could be replaced with fewer but longer steps. Short search paths suggest that the strategy steps back and forth and thus that smaller step sizes should be beneficial. After termination the strategy returns the current parental centroid which is considered as an approximation of the optimizer of the objective function $F(\mathbf{y})$.

3 The Ellipsoid Model

This work is concerned with the analysis of the $(\mu/\mu_I, \lambda)$-CSA-ES on the ellipsoid model

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

where N represents the search space dimensionality and a_i are the coefficients of the ellipsoid model. Throughout the investigations the coefficients a_i are chosen exemplarily as 1, i, and i^2. The optimization problem is formulated as

$$\min_{\mathbf{y} \in \mathbb{R}^N} F(\mathbf{y}). \tag{2}$$

Its optimizer resides at the origin of the coordinate system, $\mathbf{y} = 0$. The model (1) already represents the general case of positive definite quadratic forms for the $(\mu/\mu_I, \lambda)$-CSA-ES. This is due to the isotropy of the mutation vectors in line 5 which secure the algorithm's invariance w.r.t. arbitrary rotations of the coordinate system.

The dynamic behavior of the $(\mu/\mu_I, \lambda)$-CSA-ES on the ellipsoid model (1) is illustrated in Fig. 1. It presents the results of typical runs of the ES focusing on the squared components of the parental centroid $\mathbf{y}^{(g)}$, as well as the mutation strength σ dynamics. Approaching the optimizer the strategy continuously decreases the mutation strength with the passing number of generations.

4 Extending Previous Results to CSA-ES

This section deals with several results obtained within the analysis of the σSA-ES on the ellipsoid model in (Melkozerov and Beyer, 2010) and (Beyer and Melkozerov, 2014). These results keep their validity in the context of the CSA-ES and provide useful equipment for the further analysis. For reasons of clarity and comprehensibility we start with the abbreviation $\Sigma a := \sum_{i=1}^{N} a_i$, and introduce the mutation strength normalization

$$\sigma^{*(g)} := \frac{\sigma^{(g)} \Sigma a}{\sqrt{\sum_{j=1}^{N} a_j^2 y_j^{(g)^2}}}. \tag{3}$$

Next, the first-order progress rate is introduced:

Definition 1 *The progress rate of the $(\mu/\mu_I, \lambda)$-ES along the ith axis of the ellipsoid model (1) is the expected change of the parental parameter vector component y_i from generation g to generation $g+1$*

$$\varphi_i := E\left[y_i^{(g)} - y_i^{(g+1)} \,\middle|\, \mathbf{y}^{(g)} \right]. \tag{4}$$

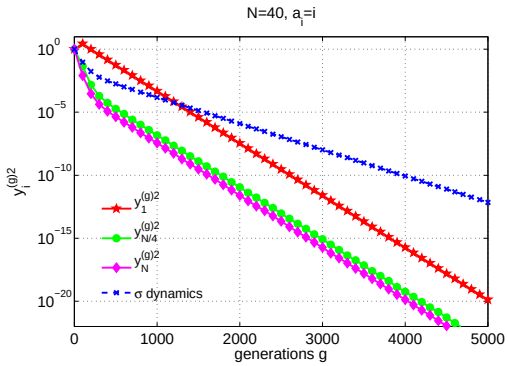

Figure 1: Dynamics of the $(3/3_I, 10)$-CSA-ES on the ellipsoid model with $a_i = i$ and $N = 40$. The quadratic deviation y_i^2 from the optimizer is displayed for $i = 1, 10, 40$. The blue dashed line represents the mutation strength σ. The CSA specific parameters are $c = 1/\sqrt{N}$ and $D = 1/c$. All curves are averaged over 10000 independent runs.

Using the progress rate normalization (Melkozerov and Beyer, 2010)

$$\varphi_i^* := \varphi_i \Sigma a, \tag{5}$$

the normalized progress rate reads

$$\varphi_i^*(\sigma^*) = \sigma^* c_{\mu/\mu,\lambda} a_i y_i. \tag{6}$$

The term $c_{\mu/\mu,\lambda} := e_{\mu,\lambda}^{1,0}$ denotes the progress coefficient of the $(\mu/\mu_I, \lambda)$-ES. It is a special case of the *generalized progress coefficients* (Beyer, 2001)

$$e_{\mu,\lambda}^{a,b} = \frac{\lambda - \mu}{\sqrt{2\pi}^{a+1}} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} (-t)^b \mathrm{e}^{-\frac{a+1}{2}t^2} \left(1 - \Phi(t)\right)^{\lambda-\mu-1} \Phi(t)^{\mu-a} \mathrm{d}t, \tag{7}$$

where $\Phi(t)$ is the cumulative distribution function of the standard normal variate.
Using the first-order progress rate one obtains good component-wise predictions of the expected approach of the parental centroid towards the optimizer as long as the distance to the optimizer is sufficiently large compared to the respective progress rate values. Otherwise the perturbations of the evolutionary process superimpose the mean value dynamics (4). That is, the predictive quality of the first-order progress rate decreases when approaching the optimizer. As a consequence, the more stable quadratic progress measure was conceived.

Definition 2 *The quadratic progress rate of the $(\mu/\mu_I, \lambda)$-ES along the ith axis of the ellipsoid model* (1) *is the expected change of squared component y_i^2 of the parental centroid between two consecutive generations g and $g+1$*

$$\varphi_i^{II} := E\left[y_i^{(g)2} - y_i^{(g+1)2} \,\middle|\, \mathbf{y}^{(g)} \right]. \tag{8}$$

It has first been introduced by Beyer and Melkozerov (2014) and reads

$$\varphi_i^{II}(\sigma^{(g)}) = \frac{2\sigma^{(g)} c_{\mu/\mu,\lambda} a_i y_i^{(g)2}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)2}} \sqrt{1 + \frac{\sigma^{(g)2} \sum_{j=1}^N a_j^2}{2\sum_{j=1}^N a_j^2 y_j^{(g)2}}}} - \frac{\sigma^{(g)2}}{\mu} \left(1 + \frac{\left((\mu-1)e_{\mu,\lambda}^{2,0} + e_{\mu,\lambda}^{1,1}\right) a_i^2 y_i^{(g)2}}{\sum_{j=1}^N a_j^2 y_j^{(g)2} \left(1 + \frac{\sigma^{(g)2} \sum_{j=1}^N a_j^2}{2\sum_{j=1}^N a_j^2 y_j^{(g)2}}\right)} \right). \tag{9}$$

Normalization with (5) and neglecting the $\sigma^2/2$ terms in the denominators yields

$$\varphi_i^{II*}(\sigma^*) = 2y_i\varphi_i^*(\sigma^*) - \frac{\sigma^{*2}}{\mu\Sigma a}\left[\sum_{j=1}^{N} a_j^2 y_j^2 + \left((\mu-1)e_{\mu,\lambda}^{2,0} + e_{\mu,\lambda}^{1,1}\right) a_i^2 y_i^2\right]. \tag{10}$$

The neglection of the $\sigma^2/2$ terms is admissible if $\sigma^{*2}\sum_{j=1}^{N} a_j^2 \Big/ \left(2(\Sigma a)^2\right) \ll 1$ holds, as demonstrated in (Beyer and Melkozerov, 2014). The quadratic progress rate formula depends on the first-order progress rate φ_i^*, cf. Eq. (6), as well as on a negative term which corresponds to the progress rate loss.

Because expression (10) is still too complex for the theoretical analysis, a simplified φ_i^{II*} formula was derived in (Beyer and Melkozerov, 2014). Under the assumption that the search space dimensionality is considerably greater than the parental population size and if there exists no dominating coefficient a_i, i.e. if the conditions

$$N \gg \mu \quad \text{and} \quad \forall i \,:\, \sum_{j\neq i} a_j^2 \gg a_i^2 \tag{11}$$

are fulfilled, the normalized progress rate is asymptotically

$$\varphi_i^{II*}(\sigma^*) = 2\sigma^* c_{\mu/\mu,\lambda} a_i y_i^2 - \frac{\sigma^{*2}}{\mu\Sigma a}\sum_{j=1}^{N} a_j^2 y_j^2. \tag{12}$$

Renormalization can be obtained by applying (3) and (5) and yields

$$\varphi_i^{II}(\sigma^{(g)}) = \frac{2\sigma^{(g)} c_{\mu/\mu,\lambda} a_i y_i^{(g)^2}}{\sqrt{\sum_{j=1}^{N} a_j^2 y_j^{(g)^2}}} - \frac{\sigma^{(g)^2}}{\mu}. \tag{13}$$

As shown in (Beyer and Melkozerov, 2014) the quadratic progress rate can be used to describe the expected approach to the optimizer for each squared component of the parental centroid. It exhibits the typical features of a well-defined progress measure, namely a gain and a loss term which depend on the mutation strength. This ensures the existence of an optimal mutation strength and of an evolution criterions (also referred to as convergence criterion), see below. Additionally, the loss term is inversely proportional to the parental population size μ. That is, the *genetic repair effect* of recombination can be observed on the ellipsoid model (Beyer and Melkozerov, 2014). The renormalized quadratic progress rate (13) will be useful within the derivation of the evolution equations of the CSA-ES in the next section.

Convergence to the optimizer in expectation requires according to (8) $\forall i = 1,\ldots,N$: $\varphi_i^{II*}(\sigma^*) > 0$. Therefore, $\sum_{i=1}^{N} a_i\varphi_i^{II*}(\sigma^*) > 0$ must hold for convergence in expectation. Using the asymptotic exact ($N \to \infty$) progress rate formula (12), this inequality yields directly the convergence criterion (Beyer and Melkozerov, 2014)

$$\sigma^* < 2\mu c_{\mu/\mu,\lambda}. \tag{14}$$

Hence, continuous positive progress to the optimizer requires that the normalized mutation strength σ^* resides in the interval $(0, 2\mu c_{\mu/\mu,\lambda})$.

5 The Mutation Strength Dynamics

In order to be able to build the evolution equations of the CSA-ES on the ellipsoid model (1) the description of the mutation strength adaptation from generation g to generation g+1 is necessary.

As it turns out, unlike the σ-SA-ES the CSA-ES lacks a simple formula for the expected change of the mutation strength σ. Due to the influence of the search path update on the mutation strength evolution, the description of the mutation strength dynamics is more complex. In the first place, we deal with this problem by describing the expected change of σ by use of a system of recurrence equations. The first difference equation concerns the mutation strength control. According to Tab. 1, line 11, the expected mutation strength in generation $g+1$ is

$$\mathrm{E}\left[\sigma^{(g+1)}\right] = \mathrm{E}\left[\sigma^{(g)} \exp\left(\frac{\|\mathbf{s}^{(g+1)}\|^2 - N}{2DN}\right)\right]. \tag{15}$$

In the asymptotic limit this can be expressed as

$$\mathrm{E}\left[\sigma^{(g+1)}\right] \simeq \sigma^{(g)} \exp\left(\frac{\mathrm{E}\left[\|\mathbf{s}^{(g+1)}\|^2\right] - N}{2DN}\right). \tag{16}$$

Obtaining a recurrence equation requires the knowledge of the squared length of search path in generation $g+1$. From line 10 of the algorithm in Tab. 1 it follows

$$\|\mathbf{s}^{(g+1)}\|^2 = (1-c)^2\|\mathbf{s}^{(g)}\|^2 + 2\sqrt{\mu}(1-c)\sqrt{c(2-c)}\mathbf{s}^{(g)\top}\langle\mathbf{z}\rangle^{(g)} + \mu c(2-c)\|\langle\mathbf{z}\rangle^{(g)}\|^2. \tag{17}$$

For the calculation of the expected change of the search path's squared norm $\mathrm{E}\left[\|\mathbf{s}^{(g+1)}\|^2 - \|\mathbf{s}^{(g)}\|^2\right]$ one needs to determine the expected values of the scalar product $\mathbf{s}^{(g)\top}\langle\mathbf{z}\rangle^{(g)}$ and the squared norm of the parental centroid's mutation vector $\|\langle\mathbf{z}\rangle^{(g)}\|^2$ in generation g. For the latter it holds

$$\mathrm{E}\left[\|\langle\mathbf{z}\rangle^{(g)}\|^2\right] = \mathrm{E}\left[\sum_{i=1}^{N}\left(\langle z\rangle_i^{(g)}\right)^2\right] = \sum_{i=1}^{N}\mathrm{E}\left[\left(\langle z\rangle_i^{(g)}\right)^2\right]. \tag{18}$$

Considering the expected value of the squared single components of the mutation vector results in

$$\mathrm{E}\left[\left(\langle z\rangle_i^{(g)}\right)^2\right] = \mathrm{E}\left[\left(\frac{1}{\mu}\sum_{m=1}^{\mu}(\mathbf{z}_{m;\lambda})_i\right)^2\right] = \frac{2}{\mu^2}\mathrm{E}\left[\sum_{l=2}^{\mu}\sum_{m=1}^{l-1}(\mathbf{z}_{l;\lambda})_i(\mathbf{z}_{m;\lambda})_i\right] + \frac{1}{\mu^2}\mathrm{E}\left[\sum_{m=1}^{\mu}(\mathbf{z}_{m;\lambda})_i^2\right] \tag{19}$$

These sums of product moments have been already derived in (Beyer and Melkozerov, 2014). Inserting the corresponding expressions into (19) and combining the fractions yields

$$\mathrm{E}\left[\left(\langle z\rangle_i^{(g)}\right)^2\right] = \frac{1}{\mu}\left(1 + \frac{a_i^2 y_i^{(g)2}}{\sum_{j=1}^{N} a_j^2 y_j^{(g)2}} \frac{(\mu-1)e_{\mu,\lambda}^{2,0} + e_{\mu,\lambda}^{1,1}}{1 + \frac{\sum_{j=1}^{N} a_j^2}{\sum_{j=1}^{N} a_j^2 y_j^{(g)2}} \frac{\sigma^{(g)2}}{2}}\right). \tag{20}$$

Finally, the aggregation of all N components of the mutation vector in generation $g+1$ yields

$$\mathrm{E}\left[\|\langle\mathbf{z}\rangle^{(g)}\|^2\right] = \frac{1}{\mu}\left(N + \frac{(\mu-1)e_{\mu,\lambda}^{2,0} + e_{\mu,\lambda}^{1,1}}{1 + \frac{\sum_{j=1}^{N} a_j^2}{\sum_{j=1}^{N} a_j^2 y_j^{(g)2}} \frac{\sigma^{(g)2}}{2}}\right). \tag{21}$$

The expressions (20) and (21) can be simplified considerably if the conditions in (11) hold. That is, provided that there exist no dominating ellipsoid coefficients a_i and that the search space

dimension is large compared to the parental population size μ, the term $a_i^2 y_i^{(g)^2}\left((\mu-1)e_{\mu,\lambda}^{2,0}+e_{\mu,\lambda}^{1,1}\right)$ in (20) can be neglected. The validity of this approach also requires that the y_i^2 dynamics behave "nicely". Depending on the initialization of the strategy this condition might be violated in the beginning, but it is always fulfilled in the asymptotic limit $g\to\infty$. The consistency of the approach was checked in (Beyer and Melkozerov, 2014) by reinserting the final y_i^2 results into the quadratic progress rate. Consequently, Eq. (20) becomes

$$\mathrm{E}\left[\left(\langle z\rangle_i^{(g)}\right)^2\right]\simeq\frac{1}{\mu}. \tag{22}$$

Similarly, Eq. (21) becomes asymptotically

$$\mathrm{E}\left[\|\langle\mathbf{z}\rangle^{(g)}\|^2\right]\simeq\frac{N}{\mu}. \tag{23}$$

The suitability of these asymptotically correct formulas will additionally be justified by comparisons of the two different iteratively generated dynamics (using Eq. (21) or Eq. (23), respectively) with the dynamics of real $(\mu/\mu_I,\lambda)$-CSA-ES runs in Sec. 6.

The next step is concerned with the calculation of the expected value of the scalar product $\mathrm{E}\left[\mathbf{s}^{(g)\top}\langle\mathbf{z}\rangle^{(g)}\right]$. Coming up with a closed-form solution turns out to be very difficult. Nonetheless, the change of the expected scalar product value between two consecutive generations can be formulated as a recurrence equation. In order to establish the difference equation the expected values of $\mathrm{E}\left[\langle z\rangle_i^{(g)}\right]$ and $\mathrm{E}\left[\langle z\rangle_i^{(g+1)}\right]$ need to be calculated. By use of the renormalized version of the first-order progress rate (4) we obtain

$$\frac{\sigma^{(g)}c_{\mu/\mu,\lambda}a_i y_i^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}=\mathrm{E}\left[y_i^{(g)}-y_i^{(g+1)}\right]=\mathrm{E}\left[y_i^{(g)}-\left(y_i^{(g)}+\sigma^{(g)}\langle z\rangle_i^{(g)}\right)\right]=-\sigma^{(g)}\mathrm{E}\left[\langle z\rangle_i^{(g)}\right], \tag{24}$$

or, respectively,

$$\mathrm{E}\left[\langle z\rangle_i^{(g)}\right]=\frac{-c_{\mu/\mu,\lambda}a_i y_i^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}. \tag{25}$$

Adding a term $\epsilon_{\langle z\rangle_i^{(g)}}$ which describes the stochastic fluctuations to the expected value $\mathrm{E}\left[\langle z\rangle_i^{(g)}\right]$ the ith component of the mutation vector of the parental centroid $\langle z\rangle_i^{(g)}$ in generation g can be modeled as

$$\langle z\rangle_i^{(g)}=\mathrm{E}\left[\langle z\rangle_i^{(g)}\right]+\epsilon_{\langle z\rangle_i^{(g)}}. \tag{26}$$

According to (25), the expected value of the ith component in generation $g+1$ is

$$\mathrm{E}\left[\langle z\rangle_i^{(g+1)}\right]=\frac{-c_{\mu/\mu,\lambda}a_i y_i^{(g+1)}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g+1)^2}}}. \tag{27}$$

For the further analysis it is desirable to represent the rhs of Eq. (27) by means of the ES state in the previous generation g. This is achieved by taking into account line 9 of the algorithm in Tab. 1 and the definition of the quadratic progress rate in Eq. (8)

$$\mathrm{E}\left[\langle z\rangle_i^{(g+1)}\right]\simeq\frac{-c_{\mu/\mu,\lambda}a_i\left(y_i^{(g)}+\sigma^{(g)}\langle z\rangle_i^{(g)}\right)}{\sqrt{\sum_{j=1}^N a_j^2\left(y_j^{(g)2}-\varphi_j^{II}(\sigma^{(g)})\right)}}=\frac{-c_{\mu/\mu,\lambda}a_i\left(y_i^{(g)}+\sigma^{(g)}\langle z\rangle_i^{(g)}\right)}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}\sqrt{1-\frac{\sum_{j=1}^N a_j^2\varphi_j^{II}(\sigma^{(g)})}{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}}. \tag{28}$$

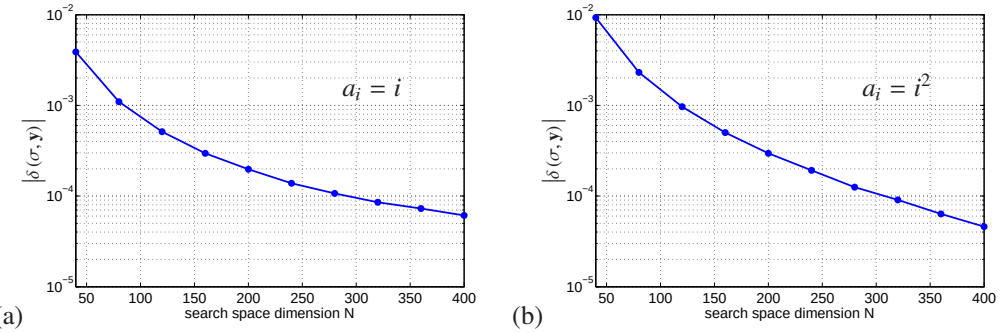

Figure 2: Steady State mean values of $\delta\left(\sigma^{(g)}, \mathbf{y}^{(g)}\right)$ plotted against the search space dimension N. Each data point is obtained by averaging over 100000 generations and 100 independent runs of a $(3/3_I, 10)$-CSA-ES.

Provided that the quotient

$$\delta\left(\sigma^{(g)}, \mathbf{y}^{(g)}\right) := \frac{\sum_{j=1}^{N} a_j^2 \varphi_j^{II}(\sigma^{(g)})}{\sum_{j=1}^{N} a_j^2 y_j^{(g)^2}} \tag{29}$$

is sufficiently small, Taylor expansion of the square root with respect to $\delta\left(\sigma^{(g)}, \mathbf{y}^{(g)}\right)$ transforms Eq. (28) into

$$\mathrm{E}\left[\langle z\rangle_i^{(g+1)}\right] \simeq \frac{-c_{\mu/\mu,\lambda} a_i \left(y_i^{(g)} + \sigma^{(g)} \langle z\rangle_i^{(g)}\right)}{\sqrt{\sum_{j=1}^{N} a_j^2 y_j^{(g)^2}}} \left(1 + \frac{1}{2}\delta\left(\sigma^{(g)}, \mathbf{y}^{(g)}\right)\right). \tag{30}$$

Figure 2 provides a verification of the assumption $\left|\delta\left(\sigma^{(g)}, \mathbf{y}^{(g)}\right)\right| \ll 1$ by plotting the experimentally generated mean values of the quotient δ against the search space dimensionality N. Consequently, the small contribution of the $\delta(\sigma^{(g)}, \mathbf{y}^{(g)})/2$ term to the expected value in Eq. (30) will be ignored in the following in order to keep the theoretical analysis tractable. Thus one obtains

$$\mathrm{E}\left[\langle z\rangle_i^{(g+1)}\right] \simeq \frac{-c_{\mu/\mu,\lambda} a_i \left(y_i^{(g)} + \sigma^{(g)} \langle z\rangle_i^{(g)}\right)}{\sqrt{\sum_{j=1}^{N} a_j^2 y_j^{(g)^2}}}. \tag{31}$$

Analogously to (26), the ith component of the mutation vector of the parental centroid $\langle z\rangle_i^{(g+1)}$ in generation $g+1$ can be described by the sum of its expected value and a term $\epsilon_{\langle z\rangle_i^{(g+1)}}$ denoting the stochastic fluctuations

$$\langle z\rangle_i^{(g+1)} = \mathrm{E}\left[\langle z\rangle_i^{(g+1)}\right] + \epsilon_{\langle z\rangle_i^{(g+1)}}. \tag{32}$$

Now, Eqs. (25) and (32) will be used to derive the difference equation for the scalar product. According to the algorithm, line 10, the search path components are updated by

$$s_i^{(g+1)} = (1-c) s_i^{(g)} + \sqrt{\mu c(2-c)} \langle z\rangle_i^{(g)}. \tag{33}$$

After multiplication with $\langle z\rangle_i^{(g+1)}$, Eq. (33) changes to

$$s_i^{(g+1)} \langle z\rangle_i^{(g+1)} = (1-c) s_i^{(g)} \langle z\rangle_i^{(g+1)} + \sqrt{\mu c(2-c)} \langle z\rangle_i^{(g)} \langle z\rangle_i^{(g+1)}, \tag{34}$$

and by insertion of (31) and (32), the ith addend of the scalar product $\mathbf{s}^\top\langle\mathbf{z}\rangle^{(g+1)}$ in generation $g+1$ reads

$$\begin{aligned} s_i^{(g+1)}\langle z\rangle_i^{(g+1)} \simeq (1-c)s_i^{(g)}&\left(\frac{-c_{\mu/\mu,\lambda}a_i\left(y_i^{(g)}+\sigma^{(g)}\langle z\rangle_i^{(g)}\right)}{\sqrt{\sum_{j=1}^N a_j^2 {y_j^{(g)}}^2}}+\epsilon_{\langle z\rangle_i^{(g+1)}}\right)\\ &+\sqrt{\mu c(2-c)}\langle z\rangle_i^{(g)}\left(\frac{-c_{\mu/\mu,\lambda}a_i\left(y_i^{(g)}+\sigma^{(g)}\langle z\rangle_i^{(g)}\right)}{\sqrt{\sum_{j=1}^N a_j^2 {y_j^{(g)}}^2}}+\epsilon_{\langle z\rangle_i^{(g+1)}}\right). \end{aligned} \tag{35}$$

Expansion of the products yields

$$\begin{aligned} s_i^{(g+1)}\langle z\rangle_i^{(g+1)} \simeq{}& (1-c)s_i^{(g)}\frac{-c_{\mu/\mu,\lambda}a_i y_i^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 {y_j^{(g)}}^2}}+(1-c)s_i^{(g)}\frac{-c_{\mu/\mu,\lambda}a_i\sigma^{(g)}\langle z\rangle_i^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 {y_j^{(g)}}^2}}\\ &+\sqrt{\mu c(2-c)}\langle z\rangle_i^{(g)}\frac{-c_{\mu/\mu,\lambda}a_i y_i^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 {y_j^{(g)}}^2}}+\sqrt{\mu c(2-c)}\langle z\rangle_i^{(g)}\frac{-c_{\mu/\mu,\lambda}a_i\sigma^{(g)}\langle z\rangle_i^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 {y_j^{(g)}}^2}}\\ &+\left((1-c)s_i^{(g)}+\sqrt{\mu c(2-c)}\langle z\rangle_i^{(g)}\right)\epsilon_{\langle z\rangle_i^{(g+1)}}. \end{aligned} \tag{36}$$

By rearranging the terms and making use of Eqs. (25) and (26) we obtain

$$\begin{aligned} s_i^{(g+1)}\langle z\rangle_i^{(g+1)} \simeq{}& (1-c)\left(1-\frac{c_{\mu/\mu,\lambda}a_i\sigma^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 {y_j^{(g)}}^2}}\right)s_i^{(g)}\langle z\rangle_i^{(g)}\\ &+\sqrt{\mu c(2-c)}\left(\langle z\rangle_i^{(g)}\frac{-c_{\mu/\mu,\lambda}a_i y_i^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 {y_j^{(g)}}^2}}+\frac{-c_{\mu/\mu,\lambda}a_i\sigma^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 {y_j^{(g)}}^2}}\left(\langle z\rangle_i^{(g)}\right)^2\right)\\ &+\left((1-c)s_i^{(g)}+\sqrt{\mu c(2-c)}\langle z\rangle_i^{(g)}\right)\epsilon_{\langle z\rangle_i^{(g+1)}}-(1-c)s_i^{(g)}\epsilon_{\langle z\rangle_i^{(g)}}. \end{aligned} \tag{37}$$

Considering expected values in Eq. (37), the perturbation terms $\epsilon_{\langle z\rangle_i^{(g+1)}}$ and $\epsilon_{\langle z\rangle_i^{(g)}}$ vanish by definition leading to

$$\begin{aligned} \mathrm{E}\left[s_i^{(g+1)}\langle z\rangle_i^{(g+1)}\right] \simeq{}& (1-c)\left(1-\frac{c_{\mu/\mu,\lambda}a_i\sigma^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 {y_j^{(g)}}^2}}\right)\mathrm{E}\left[s_i^{(g)}\langle z\rangle_i^{(g)}\right]\\ &+\sqrt{\mu c(2-c)}\left(\mathrm{E}\left[\langle z\rangle_i^{(g)}\right]\frac{-c_{\mu/\mu,\lambda}a_i y_i^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 {y_j^{(g)}}^2}}+\frac{-c_{\mu/\mu,\lambda}a_i\sigma^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 {y_j^{(g)}}^2}}\mathrm{E}\left[\left(\langle z\rangle_i^{(g)}\right)^2\right]\right). \end{aligned} \tag{38}$$

In order to justify the approximation of (37) by expected value expressions, the iterative dynamics resulting from Eq. (38) will be compared to experimental runs of the ES in Section 6. Now

considering Eq. (25), one obtains

$$\begin{aligned}\mathrm{E}\left[s_i^{(g+1)}\langle z\rangle_i^{(g+1)}\right] \simeq (1-c)\left(1-\frac{c_{\mu/\mu,\lambda}a_i\sigma^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}\right)\mathrm{E}\left[s_i^{(g)}\langle z\rangle_i^{(g)}\right] \\ +\ \sqrt{\mu c(2-c)}\left(\frac{c_{\mu/\mu,\lambda}^2 a_i^2 y_i^{(g)2}}{\sum_{j=1}^N a_j^2 y_j^{(g)^2}} - \frac{c_{\mu/\mu,\lambda}a_i\sigma^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}\mathrm{E}\left[\left(\langle z\rangle_i^{(g)}\right)^2\right]\right).\end{aligned} \tag{39}$$

Using (22), Eq. (39) transforms into

$$\begin{aligned}\mathrm{E}\left[s_i^{(g+1)}\langle z\rangle_i^{(g+1)}\right] \simeq (1-c)\left(1-\frac{c_{\mu/\mu,\lambda}a_i\sigma^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}\right)\mathrm{E}\left[s_i^{(g)}\langle z\rangle_i^{(g)}\right] \\ +\ \sqrt{\mu c(2-c)}\left(\frac{c_{\mu/\mu,\lambda}^2 a_i^2 y_i^{(g)2}}{\sum_{j=1}^N a_j^2 y_j^{(g)^2}} - \frac{c_{\mu/\mu,\lambda}a_i\sigma^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}\frac{1}{\mu}\right).\end{aligned} \tag{40}$$

Notice, the scalar product of the search path $\mathbf{s}^{(g)}$ and the centroid of the μ best mutation vectors $\langle\mathbf{z}\rangle^{(g)}$ in generation g is $\mathbf{s}^{(g)^\top}\langle\mathbf{z}\rangle^{(g)} = \sum_{i=1}^N s_i^{(g)}\langle z\rangle_i^{(g)}$. That is, Eq. (39) and (40), respectively, provide a component-wise representation of the difference equation that describes the one-generation change of the scalar product. The results in Eq. (20) and Eq. (39) allow for the compilation of a difference equation which describes the expected change of the squared length of the search path, see also Eq. (17),

$$\mathrm{E}\left[\|\mathbf{s}^{(g+1)}\|^2\right] = (1-c)^2\mathrm{E}\left[\|\mathbf{s}^{(g)}\|^2\right] + 2(1-c)\sqrt{\mu c(2-c)}\sum_{i=1}^N \mathrm{E}\left[s_i^{(g+1)}\langle z\rangle_i^{(g+1)}\right] + \mu c(2-c)\mathrm{E}\left[\|\langle\mathbf{z}\rangle^{(g)}\|^2\right]. \tag{41}$$

Together with Eq. (16) the mutation strength change from generation g to generation $g+1$ can be described by means of the difference equations (39) and (41) taking (21) into account.

6 Evolution Equations

In the case of the σSA-ES on the general quadratic fitness model the framework of the dynamical systems approach (Beyer, 2001) can be applied. It states that the stochastic process of the ES from generation g to generation $g+1$ can be divided into mean value parts and fluctuation terms (ϵ_i and ϵ_σ) as follows (Beyer and Melkozerov, 2014)

$$y_i^{(g+1)^2} = y_i^{(g)^2} - \varphi_i^{II}(\sigma^{(g)}) + \epsilon_i(\sigma^{(g)}), \tag{42}$$

$$\sigma^{(g+1)} = \sigma^{(g)}\left(1+\psi(\sigma^{(g)})\right) + \epsilon_\sigma(\sigma^{(g)}). \tag{43}$$

The mean value parts in Eq. (42) are directly given by the quadratic progress rate (9) and (13), respectively. The self-adaptation response (SAR) function $\psi(\sigma^{(g)})$ describes the mean value dynamics of Eq. (43). Unfortunately, in the context of CSA-ES the derivation of a closed SAR formula appears to be a hard task even in the steady state case. That is, in the first step it has to be substituted by the $N+2$ difference equations (16), (39), and (41) before in Sec. 7.1 an asymptotical approximation of the SAR function will be derived. In order to keep the analysis tractable throughout this paper the fluctuation terms will be disregarded and asymptotically correct simplifications will be derived and compared with experiments.

Iterative Scheme of Evolution Equations A

$$y_i^{(g+1)^2} \leftarrow y_i^{(g)^2} - \frac{2\sigma^{(g)} c_{\mu/\mu,\lambda} a_i y_i^{(g)^2}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}} \sqrt{1 + \frac{\sigma^{(g)^2} \sum_{j=1}^N a_j^2}{2\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}} + \frac{\sigma^{(g)^2}}{\mu}\left(1 + \frac{\left((\mu-1)e_{\mu,\lambda}^{2,0} + e_{\mu,\lambda}^{1,1}\right) a_i^2 y_i^{(g)^2}}{\sum_{j=1}^N a_j^2 y_j^{(g)^2}\left(1 + \frac{\sigma^{(g)^2}\sum_{j=1}^N a_j^2}{2\sum_{j=1}^N a_j^2 y_j^{(g)^2}}\right)}\right) \tag{A.1}$$

$$\mathrm{E}\left[s_i^{(g+1)}\langle z\rangle_i^{(g+1)}\right] \leftarrow \mathrm{E}\left[s_i^{(g)}\langle z\rangle_i^{(g)}\right]\left(1 - c - \frac{c_{\mu/\mu,\lambda}(1-c) a_i \sigma^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}\right) + \sqrt{\mu c(2-c)}\frac{c_{\mu/\mu,\lambda}^2 a_i^2 y_i^{(g)^2}}{\sum_{j=1}^N a_j^2 y_j^{(g)^2}} - \frac{\sqrt{\mu c(2-c)}}{\mu}\frac{c_{\mu/\mu,\lambda} a_i \sigma^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}\left(1 + \frac{\left((\mu-1)e_{\mu,\lambda}^{2,0} + e_{\mu,\lambda}^{1,1}\right) a_i^2 y_i^{(g)^2}}{\sum_{j=1}^N a_j^2 y_j^{(g)^2} + \frac{\sigma^{(g)^2}}{2}\sum_{j=1}^N a_j^2}\right) \tag{A.2}$$

$$\mathrm{E}\left[\|\mathbf{s}^{(g+1)}\|^2\right] \leftarrow \mathrm{E}\left[\|\mathbf{s}^{(g)}\|^2\right](1-c)^2 + 2(1-c)\sqrt{\mu c(2-c)}\sum_{i=1}^N \mathrm{E}\left[s_i^{(g)}\langle z\rangle_i^{(g)}\right] + c(2-c)\left(N + \left((\mu-1)e_{\mu,\lambda}^{2,0} + e_{\mu,\lambda}^{1,1}\right)\Big/\left(1 + \frac{\sum_{j=1}^N a_j^2}{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}\frac{\sigma^{(g)^2}}{2}\right)\right) \tag{A.3}$$

$$\sigma^{(g+1)} \leftarrow \sigma^{(g)} \exp\left(\frac{\mathrm{E}\left[\|\mathbf{s}^{(g+1)}\|^2\right] - N}{2DN}\right) \tag{A.4}$$

Table 2: Iterative scheme A summarizes the evolution equations of the $(\mu/\mu_I, \lambda)$-CSA-ES by use of the Eqs. (9), (39), (41), and (16).

A first representation of the strategy's evolution behavior is provided in iterative scheme A, Tab. 2. The one-generation behavior of the component-wise squared distance to the optimizer y_i^2 is modeled in (A.1) by use of Eq. (9). Using Eq. (20) to express the expected values $\mathrm{E}\left[(\langle z\rangle_i^{(g)})^2\right]$ in Eq. (39) yields the iterative relation (A.2) for the scalar product components $\mathrm{E}\left[s_i\langle z\rangle_i\right]$. The iterative description of the squared length of the search path $\mathbf{s}$ in (A.3) is obtained by inserting (21) into Eq. (41). Finally, the mutation strength adaptation is specified in (A.4) using Eq. (16).

Whether the modeling approach yields meaningful results can be checked by comparison of the iteratively generated dynamics of the system of evolution equations A, Tab. 2, with experimental results of real $(\mu/\mu, \lambda)$-CSA-ES runs in Fig. 3. The typical long-term behavior of the ES is observed for $a_i = i$ which is obtained by iteration of scheme A starting from $\sigma^{(0)} = 1$, $\mathbf{y}^{(0)} = \mathbf{1}$. Cumulation and damping parameter have been set to $c = 1/\sqrt{N}$ and $D = 1/c$. Considering a small dimensionality N, the experimental data of the ES slightly deviate from the theoretical predictions. These deviations diminish with increasing search space dimensionality. In fact, on both sides a good agreement of iterative and experimental dynamics can be observed.

However, with regard to further theoretical investigations we are interested in a more convenient system of evolution equations. To this end, different approximative schemes are to be developed subsequently. In order to obtain the iterative scheme B, see Tab. 3, previously obtained simplifications are applied as follows. The progress rate representation (13) is utilized in (B.1) to express the y_i^2 dynamics. Instead of using (39), the dynamics of the scalar product com-

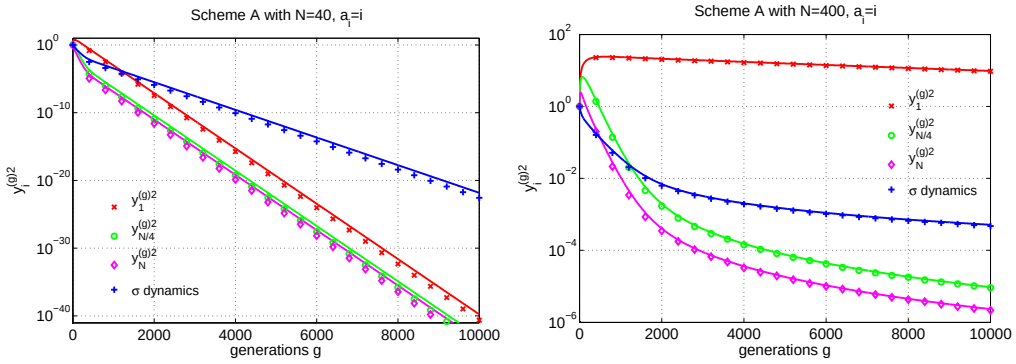

Figure 3: Dynamics generated by scheme A, Tab. 2, compared to real runs of the $(3/3_I, 10)$-CSA-ES using $a_i = i$, $N = 40$ and $N = 400$. The results of the evolution equations in A for the y_i^2 dynamics, as well as the σ dynamics, are depicted by the solid lines. The discrete data points display the corresponding experimental results which are averaged over 10000 independent runs of the ES.

ponents in (B.2) originate from the use of recurrence equation (40). Inserting (23) into Eq. (41) transforms (A.3) into (B.3). Moreover, using the first two terms of the Taylor expansion of the exponential function

$$\exp\left(\frac{\mathrm{E}\left[\|\mathbf{s}^{(g+1)}\|^2\right]-N}{2DN}\right) = 1 + \frac{\mathrm{E}\left[\|\mathbf{s}^{(g+1)}\|^2\right]-N}{2DN}\left(1 + \frac{1}{4D}\left(\frac{\mathrm{E}\left[\|\mathbf{s}^{(g+1)}\|^2\right]}{N} - 1\right) + \dots\right), \tag{44}$$

allows for approximating the mutation strength dynamics by

$$\sigma^{(g+1)} \simeq \sigma^{(g)}\left(1 + \frac{\mathrm{E}\left[\|\mathbf{s}^{(g+1)}\|^2\right]-N}{2DN}\right), \tag{45}$$

see also (B.4). Notice, higher order terms in (44) can be neglected assuming $D \to \infty$. This is true for standard CSA implementations in the asymptotic limit case, since D usually increases with the search space dimensionality N and $\|\mathbf{s}^{(g+1)}\|^2 \simeq N$. The simplifications within the modified iterative scheme B are justified by comparing its dynamics to the dynamics of the scheme A, Tab. 2. In Fig. 4 the iteratively generated dynamics are displayed for $a_i = i^2$, $c = 1/\sqrt{N}$, and $D = 1/c$. Both systems are iterated beginning with $\sigma^{(0)} = 1$, and $\mathbf{y}^{(0)} = \mathbf{1}$. The dynamics of the systems A and B show small deviations considering search space dimension $N = 40$. Using a higher dimensionality $N = 200$, the agreement of both iterative dynamics increases significantly. Both systems of evolution equations show the same long-term behavior and agree for sufficiently high dimensions N. Making use of the evolution equations B in Tab. 3 is reasonable because it allows for a tractable investigation of the CSA-ES. Thus it will be the basis for the derivations in the following sections. The compliance of the predictions from the iterative scheme B with experimental runs of the ES is illustrated in Fig. 5. The good approximation quality justifies the modeling approach by the use of the iterative scheme A or B, respectively. That is, for small dimensional problems the use of iterative scheme A will provide more precise predictions of the ES dynamics. Considering higher-dimensional spaces both systems provide a reliable modeling of the ES behavior.

In Fig. 5 two phases in the dynamics of the $(\mu/\mu, \lambda)$-CSA-ES can be observed. After the start of the optimization the ES dynamics enter a transient phase. This is followed by approach-

Iterative Scheme of Evolution Equations B

$$y_i^{(g+1)^2} \leftarrow y_i^{(g)^2}\left(1 - \frac{2\sigma^{(g)} c_{\mu/\mu,\lambda} a_i}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}\right) + \frac{\sigma^{(g)^2}}{\mu} \tag{B.1}$$

$$\begin{aligned}\mathrm{E}\left[s_i^{(g+1)}\langle z\rangle_i^{(g+1)}\right] \leftarrow\ & \mathrm{E}\left[s_i^{(g)}\langle z\rangle_i^{(g)}\right]\left(1 - c - \frac{c_{\mu/\mu,\lambda}(1-c) a_i \sigma^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}\right) \\ & + \frac{c_{\mu/\mu,\lambda}^2 \sqrt{\mu c(2-c)}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}\left(\frac{a_i^2 y_i^{(g)^2}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}} - \frac{a_i \sigma^{(g)}}{\mu c_{\mu/\mu,\lambda}}\right)\end{aligned} \tag{B.2}$$

$$\mathrm{E}\left[\|\mathbf{s}^{(g+1)}\|^2\right] \leftarrow \mathrm{E}\left[\|\mathbf{s}^{(g)}\|^2\right](1-c)^2 + 2(1-c)\sqrt{\mu c(2-c)}\sum_{i=1}^N \mathrm{E}\left[s_i^{(g)}\langle z\rangle_i^{(g)}\right] + c(2-c)N \tag{B.3}$$

$$\sigma^{(g+1)} \leftarrow \sigma^{(g)}\left(1 + \frac{\mathrm{E}\left[\|\mathbf{s}^{(g+1)}\|^2\right] - N}{2DN}\right) \tag{B.4}$$

Table 3: Iterative scheme B is obtained by simplification of the iterative scheme A making use of (13), (40), (41) with (23), as well as Eq. (45).

ing a steady state behavior. The transient period is characterized by a decrease of the y_i^2 curves and the σ values. The rate of this decline increases with i. That is, y_1 decreases significantly slower than y_N. The steady state behavior is featured by a slower decrease with the same rate for all single components y_i^2, $i = 1, \ldots, N$. In particular, the y_i^2 dynamics fall with a log-linear law. The steady state of the σ values exhibits a log-linear behavior as well, but with a different rate of decline.

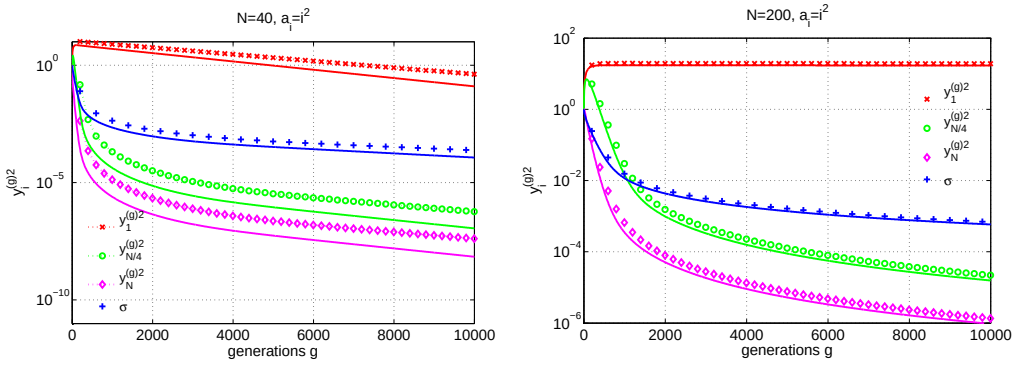

Figure 4: Comparison of the dynamics of the iterative schemes A and B simulating a $(3/3_I, 10)$-CSA-ES with $a_i = i^2$, $N = 40$, and $N = 200$ respectively. The iteratively generated y_i^2 dynamics (for $i = 1, N/4, N$) as well as the iterative σ dynamics are displayed. While the data points illustrate the results of the evolution equations from A, the results of B are represented by the solid lines. The CSA specific parameters have been set to $c = 1/\sqrt{N}$ and $D = 1/c$.

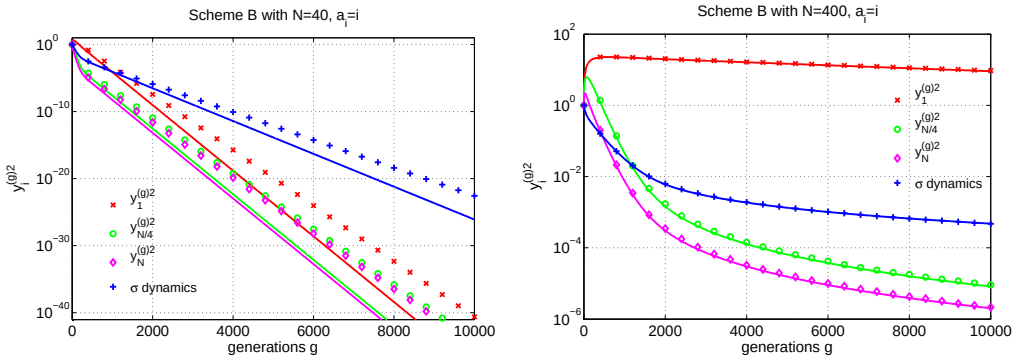

Figure 5: Iteratively generated dynamics of scheme B compared to real runs of the $(3/3_I, 10)$-CSA-ES using $a_i = i$, $N = 40$ and $N = 400$. The results of the evolution equations in B for the y_i^2 dynamics as well as the σ dynamics are illustrated by the solid lines. The discrete data points display the corresponding experimental results which are averaged over 10000 independent runs of the ES.

7 Steady State Dynamics

7.1 Derivation of Simplified Evolution Equations via Self-Adaptation Response of CSA

After having validated the modeling of the system of evolution equations B (Table 3), finding a closed form solution to the system of $2N + 2$ nonlinear difference equations still appears to be a difficult task. The goal of the subsequent section is to substitute the $N + 2$ equations which model the σ dynamics, i.e. (B.2) to (B.4), by one single evolution equation. As it turns out, considering the steady state dynamics allows for the derivation of a suitable approximation. That is, the search path $\mathbf{s}$ dynamics within the strategy's steady state can be used to provide an asymptotically exact approximation of a function analogous to the self-adaptation response (SAR) function of σ self-adaptation ES (Beyer and Melkozerov, 2014). To this end, the steady state dynamics of the difference equations (B.2) and (B.3) are investigated. In the first step the steady state dynamics of the scalar product components $\mathrm{E}\left[s_i^{(g)}\langle z\rangle_i^{(g)}\right]$ are examined. Afterwards the focus is on the resulting steady state dynamics of the complete search path. Using the steady state condition for the scalar product components

$$\mathrm{E}\left[s_i^{(g+1)}\langle z\rangle_i^{(g+1)}\right] = \mathrm{E}\left[s_i^{(g)}\langle z\rangle_i^{(g)}\right] =: (s_i\langle z\rangle_i)_{ss}, \tag{46}$$

and solving (B.2) for $(s_i\langle z\rangle_i)_{ss}$ yields

$$(s_i\langle z\rangle_i)_{ss} = c^2_{\mu/\mu,\lambda}\sqrt{\frac{\mu(2-c)}{c}}\,\frac{\dfrac{a_i^2 y_i^2}{\sum_{j=1}^N a_j^2 y_j^2} - \dfrac{a_i\sigma_{ss}}{\mu c_{\mu/\mu,\lambda}\sqrt{\sum_{j=1}^N a_j^2 y_j^2}}}{1 + \dfrac{(1-c)}{c}\dfrac{c_{\mu/\mu,\lambda}a_i\sigma_{ss}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^2}}}. \tag{47}$$

Here, $\sigma_{ss} = \sigma_{ss}^{(g)}$ denotes the steady state mutation strength which is attained by the ES after the transient phase (i.e., after a sufficiently long number of generations g). Applying the mutation strength normalization (3) one obtains

$$(s_i\langle z\rangle_i)_{ss} = c^2_{\mu/\mu,\lambda}\sqrt{\frac{\mu(2-c)}{c}}\,\frac{\dfrac{a_i^2 y_i^2}{\sum_{j=1}^N a_j^2 y_j^2} - \dfrac{a_i\sigma^*_{ss}}{\mu c_{\mu/\mu,\lambda}\Sigma a}}{1 + \dfrac{(1-c)}{c}\dfrac{c_{\mu/\mu,\lambda}a_i\sigma^*_{ss}}{\Sigma a}} \tag{48}$$

and the steady state scalar product $(\mathbf{s}^\top\langle\mathbf{z}\rangle)_{ss}$ becomes

$$\left(\mathbf{s}^\top\langle\mathbf{z}\rangle\right)_{ss} = c^2_{\mu/\mu,\lambda}\sqrt{\frac{\mu(2-c)}{c}}\sum_{i=1}^{N}\frac{\frac{a_i^2y_i^2}{\sum_{j=1}^N a_j^2y_j^2}-\frac{a_i\sigma^*_{ss}}{\mu c_{\mu/\mu,\lambda}\Sigma a}}{1+\frac{(1-c)}{c}\frac{c_{\mu/\mu,\lambda}a_i\sigma^*_{ss}}{\Sigma a}}. \tag{49}$$

Now that we have found a description of the steady state of the scalar product, the next step is concerned with the derivation of the steady state of the squared length of the search path vector $\|\mathbf{s}\|^2$. Again the use of the steady state condition

$$\mathrm{E}\left[\|\mathbf{s}^{(g+1)}\|^2\right] = \mathrm{E}\left[\|\mathbf{s}^{(g)}\|^2\right] =: \|\mathbf{s}\|^2_{ss} \tag{50}$$

combined with Eq. (41) and (23) (see also B.3 in Tab. 3) yields an asymptotically exact expression for the $\|\mathbf{s}\|^2$ dynamics in the proximity of the steady state

$$\|\mathbf{s}\|^2_{ss} = (1-c(2-c))\,\|\mathbf{s}\|^2_{ss} + 2(1-c)\sqrt{\mu c(2-c)}\left(\mathbf{s}^\top\langle\mathbf{z}\rangle\right)_{ss} + Nc(2-c). \tag{51}$$

Solving this for $\|\mathbf{s}\|^2_{ss}$ and plugging in (49), Eq. (51) yields

$$\|\mathbf{s}\|^2_{ss} = N + 2\frac{1-c}{c}\mu c^2_{\mu/\mu,\lambda}\sum_{i=1}^{N}\frac{\frac{a_i^2y_i^2}{\sum_{j=1}^N a_j^2y_j^2}-\frac{a_i\sigma^*_{ss}}{\mu c_{\mu/\mu,\lambda}\Sigma a}}{1+\frac{(1-c)}{c}\frac{c_{\mu/\mu,\lambda}a_i\sigma^*_{ss}}{\Sigma a}}. \tag{52}$$

The renormalized version of Eq. (52) using (3)

$$\|\mathbf{s}\|^2_{ss} = N + 2\frac{1-c}{c}\mu c^2_{\mu/\mu,\lambda}\sum_{i=1}^{N}\frac{\frac{a_i^2y_i^2}{\sum_{j=1}^N a_j^2y_j^2}-\frac{a_i\sigma_{ss}}{\mu c_{\mu/\mu,\lambda}\sqrt{\sum_{j=1}^N a_j^2y_j^2}}}{1+\frac{(1-c)}{c}\frac{c_{\mu/\mu,\lambda}a_i\sigma_{ss}}{\sqrt{\sum_{j=1}^N a_j^2y_j^2}}} \tag{53}$$

serves as an approximation for the short-term behavior of the search path. It describes the progress of the search path between two consecutive generations sufficiently well. Applying Eq. (53) to (45), recalling (50), the mutation strength dynamics can be modeled by the single evolution equation

$$\boxed{\sigma^{(g+1)} = \sigma^{(g)}\left(1+\frac{\mu c^2_{\mu/\mu,\lambda}}{DN}\frac{(1-c)}{c}\sum_{i=1}^{N}\frac{\frac{a_i^2y_i^{(g)2}}{\sum_{j=1}^N a_j^2y_j^{(g)2}}-\frac{a_i\sigma^{(g)}}{\mu c_{\mu/\mu,\lambda}\sqrt{\sum_{j=1}^N a_j^2y_j^{(g)2}}}}{1+\frac{(1-c)}{c}\frac{c_{\mu/\mu,\lambda}a_i\sigma^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2y_j^{(g)2}}}}\right).} \tag{54}$$

The second addend within the braces can be *interpreted* as the *self-adaptation response (SAR) function* $\tilde{\psi}$ of the $(\mu/\mu_I,\lambda)$-CSA-ES in the steady state. Using (3) the normalized version of $\tilde{\psi}$ reads

$$\boxed{\tilde{\psi}(\sigma^{*(g)}) = \frac{\mu c^2_{\mu/\mu,\lambda}}{DN}\frac{(1-c)}{c}\sum_{i=1}^{N}\frac{\frac{a_i^2y_i^{(g)2}}{\sum_{j=1}^N a_j^2y_j^{(g)2}}-\frac{a_i\sigma^{*(g)}}{\mu c_{\mu/\mu,\lambda}\Sigma a}}{1+\frac{(1-c)}{c}\frac{c_{\mu/\mu,\lambda}a_i\sigma^{*(g)}}{\Sigma a}}\quad.} \tag{55}$$

Considering (54) and (13), the strategy's mean value dynamics in the steady state can be described now by the iterative scheme C in Table 4. A comparison of the iteratively generated

Iterative Scheme of Evolution Equations C

$$y_i^{(g+1)2} \leftarrow y_i^{(g)2}\left(1 - \frac{2\sigma^{(g)} c_{\mu/\mu,\lambda} a_i}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)2}}}\right) + \frac{\sigma^{(g)2}}{\mu}\,, \tag{C.1}$$

$$\sigma^{(g+1)} \leftarrow \sigma^{(g)}\left(1 + \frac{\mu c_{\mu/\mu,\lambda}^2}{DN}\frac{(1-c)}{c}\sum_{i=1}^{N}\frac{\frac{a_i^2 y_i^{(g)2}}{\sum_{j=1}^N a_j^2 y_j^{(g)2}} - \frac{a_i \sigma^{(g)}}{\mu c_{\mu/\mu,\lambda}\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)2}}}}{1 + \frac{(1-c)}{c}\frac{c_{\mu/\mu,\lambda} a_i \sigma^{(g)}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)2}}}}\right). \tag{C.2}$$

Table 4: Iterative scheme C is composed of the difference equations describing the y_i^2 dynamics, cf. (B.1), as well as the σ dynamic which is now modeled by use of Eq. (54).

dynamics from scheme C to those of scheme B is displayed in Fig. 6. The good agreement of both dynamics for small dimension $N = 40$, as well as for high dimensionality $N = 400$, validates the use of the SAR function $\tilde{\psi}$ (55) to model the ES dynamics.

The second equation in iterative scheme C, Tab. 4, is still too complicated to allow for the calculation of closed-form expressions of the steady state variables. The critical part in C.2 is the sum in the SAR function (55) the denominator of which depends on a_i. A simplification of Eq. (55) is obtained by requiring that the second addend in the denominator is sufficiently small leading to the condition

$$\frac{(1-c)}{c}\frac{c_{\mu/\mu,\lambda} a_i \sigma^*_{ss}}{\Sigma a} \ll 1. \tag{56}$$

Equation (56) has to be satisfied for all $i = 1 \ldots, N$, i.e., particularly for the largest coefficient $\max_{i=1,\ldots,N} a_i =: \hat{a}$, yielding

$$\frac{(1-c)}{c}\frac{c_{\mu/\mu,\lambda} \hat{a} \sigma^*_{ss}}{\Sigma a} \ll 1. \tag{57}$$

Requiring positive progress in direction of the optimizer, the normalized steady state mutation strength must be bounded according to (14) by $\sigma^*_{ss} < 2\mu c_{\mu/\mu,\lambda}$. Therefore, after inserting $\sigma^*_{ss} =$

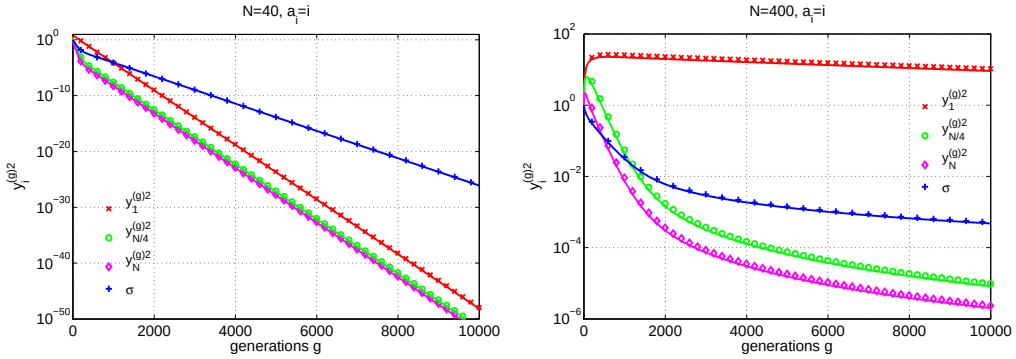

Figure 6: Comparison of the iterative systems B and C simulating a $(3/3, 10)$-CSA-ES. The dynamics of the component-wise squared distances to the optimizer y_i resulting from system B are illustrated by the solid lines. The results of scheme C are displayed by the discrete data points.

$2\mu c_{\mu/\mu,\lambda}$ in (57) and resolving for c yields the condition

$$c \gg \frac{1}{1 + \frac{\Sigma a}{2\mu c^2_{\mu/\mu,\lambda}\hat{a}}}. \tag{58}$$

That is, c must not be too small. Actually, for the ellipsoid models with coefficients $a_i = 1, i, i^2$ choices $c \propto \frac{1}{N}$ are too small. However, choices $c \propto \frac{1}{\sqrt{N}}$ will yield asymptotically exact results. Thus, using condition (58), the squared steady state search path length (52) becomes after a short calculation neglecting the lhs of (56) in the denominator of (52)

$$\|\mathbf{s}\|^2_{ss} = N + 2\frac{1-c}{c}\mu c^2_{\mu/\mu,\lambda}\left(1 - \frac{\sigma^*_{ss}}{\mu c_{\mu/\mu,\lambda}}\right). \tag{59}$$

Equation (59) has an interesting interpretation considering expected values: if $\sigma^*_{ss} < \mu c_{\mu/\mu,\lambda}$, then the steady state length square of the evolution path $\|\mathbf{s}\|^2_{ss}$ is greater than the length square of a random path (being N). Thus, by virtue of line 11, Tab. 1, σ is increased. As a result, σ^* can increase towards $\mu c_{\mu/\mu,\lambda}$, the optimal value for the sphere model. In the opposite case $\sigma^*_{ss} > \mu c_{\mu/\mu,\lambda}$, a decrease of σ^* towards $\mu c_{\mu/\mu,\lambda}$ happens. That is, in a static case (also referred to as scale-invariant case) the control rule in line 11, Tab. 1, drives σ to its optimal sphere model value. Note, however, in the real ES algorithm and its approximation schemes, e.g. Tab. 5, this does not happen since the y_i^2 dynamics influence the σ^* evolution and a steady state $\sigma^*_{ss} \neq \mu c_{\mu/\mu,\lambda}$ will result. Determining the real steady state will be done in the remainder of this paper.

Provided that (58) holds, also the SAR function (55) can be simplified yielding after renormalization with (3)

$$\tilde{\psi}(\sigma^{(g)}) = \frac{\mu c^2_{\mu/\mu,\lambda}}{DN}\frac{(1-c)}{c}\left(1 - \frac{\sigma^{(g)}\Sigma a}{\mu c_{\mu/\mu,\lambda}\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)2}}}\right). \tag{60}$$

In turn the evolution equation (54) simplifies as well. Thus, (C.2) in Tab. 4 changes to (D.2) in the new scheme D in Tab. 5.

The approximation quality of system D is validated in Fig. 7 for different choices of the cumulation parameter c and the search space dimensionality N. On the lhs of Fig. 7 the cumulation parameter c is set to $c = 1/N$ in such a way that condition (58) is not satisfied (provided that the ellipsoid coefficients are $a_i = i$). As a consequence one observes larger deviations between

Iterative Scheme of Evolution Equations D

$$y_i^{(g+1)2} \leftarrow y_i^{(g)2}\left(1 - \frac{2\sigma^{(g)}c_{\mu/\mu,\lambda}a_i}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)2}}}\right) + \frac{\sigma^{(g)2}}{\mu}\ , \tag{D.1}$$

$$\sigma^{(g+1)} \leftarrow \sigma^{(g)}\left(1 + \frac{\mu c^2_{\mu/\mu,\lambda}}{DN}\frac{(1-c)}{c}\left(1 - \frac{\sigma^{(g)}\Sigma a}{\mu c_{\mu/\mu,\lambda}\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)2}}}\right)\right). \tag{D.2}$$

Table 5: Iterative scheme D is composed of the y_i^2 difference equations, cf. (C.1), as well as the σ dynamics which is modeled using the simplified SAR function in (60) under the assumption that condition (58) holds.

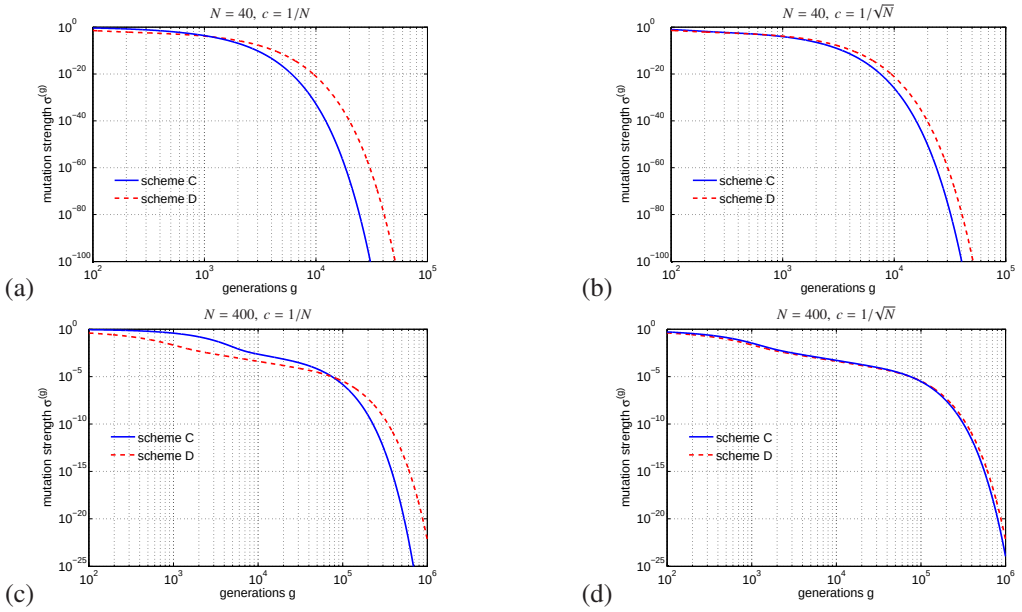

Figure 7: Comparison of the σ dynamics simulating a $(3/3, 10)$-CSA-ES by use of the iterative schemes C and D for the ellipsoid model $a_i = i$. While the cumulation parameter is set to $c = 1/N$ on the lhs, it is $c = 1/\sqrt{N}$ on the rhs of the illustration. Both upper figures display the results for search space dimension $N = 40$. In the figures below $N = 400$ is chosen.

the two iterative schemes C and D. However, these deviations are generally more pronounced in the transient phase of the evolutionary process that is emphasized in the plots due to the use of a logarithmic scale for the horizontal axes. The rhs displays a scenario where condition (58) is fulfilled ($c = 1/\sqrt{N}$). Especially with growing dimensionality a visually good agreement of both systems of evolution equation can be noticed.

Now that we have qualitatively validated the modeling approach, the deduction of closed-form solutions to the system C can be performed according to the approach introduced in (Beyer and Melkozerov, 2014). Taking a look at Fig. 6 one observes that after transition periods of different length all y_i^2 curves approach a log-linear behavior with the same slopes. The same log-linear trend holds true for the σ dynamics, but it exhibits a different slope. This observation suggests that the system C asymptotically reaches a linear systems behavior. Therefore, modeling closed-form solutions to the system by exponential functions is reasonable. Thus, in the proximity of the steady state the following *Ansatz* is used to solve the system of evolution equations C

$$y_i^{(g)^2} = b_i \mathrm{e}^{-\nu g}, \quad b_i > 0,\ \nu > 0 \tag{61}$$

$$\sigma^{(g)} = \sigma_0 \mathrm{e}^{-\frac{\nu}{2} g}, \quad \sigma_0 > 0. \tag{62}$$

This *Ansatz* has been introduced in (Beyer and Melkozerov, 2014) in order to solve the evolution equations of the σ-SA-ES in the asymptotic limit ($g \to \infty$). We will therefore only sketch the derivations in the remainder of this and in the following Sec. 7.2. The *Ansatz* (61), (62) already takes the observed different slopes of y_i^2 and σ correctly into account. Making use of the correct magnitudes ν and $\nu/2$, respectively, removes the time dependence within the asymptotic solutions of system C.

As a consequence, by inserting (61) and (62) into the normalization formula (3) one obtains the constant normalized steady state mutation strength

$$\sigma^* = \sigma_0 \frac{\Sigma a}{\sqrt{\sum_{j=1}^{N} a_j^2 b_j}} =: \sigma^*_{\rm ss}. \tag{63}$$

In the next step, the system C will be solved with the help of the *Ansatz* (61), (62). To that point, an eigenvalue problem will be established in the next section.

7.2 The Eigenvalue Problem

In order to keep this work self-contained, the eigenvalue problem arising from (61), (62) is considered in detail. This is done analogously to (Beyer and Melkozerov, 2014). Making use of the *Ansatz* (61), (62), the squared distance to the optimizer y_i^2 and the mutation strength σ in generation $g+1$ can be expressed by means of their states in generation g

$$\begin{aligned} y_i^{(g+1)^2} &= b_i \mathrm{e}^{-\nu(g+1)} = b_i \mathrm{e}^{-\nu g}\mathrm{e}^{-\nu} = y_i^{(g)^2}\mathrm{e}^{-\nu} \\ \sigma^{(g+1)} &= \sigma_0 \mathrm{e}^{-\frac{\nu}{2}(g+1)} = \sigma_0 \mathrm{e}^{-\frac{\nu}{2}g}\mathrm{e}^{-\frac{\nu}{2}} = \sigma^{(g)}\mathrm{e}^{-\frac{\nu}{2}} \end{aligned} \tag{64}$$

From Fig. 5 it can be deduced that ν must be rather small. That is, $\mathrm{e}^{-\nu}$ can be simplified using Taylor expansion $\mathrm{e}^{-\nu} = 1 - \nu + \mathrm{O}(\nu^2)$. Neglecting higher orders terms, (61), (62) transforms to

$$y_i^{(g+1)^2} = (1-\nu)\, b_i \mathrm{e}^{-\nu g}, \tag{65}$$

$$\sigma^{(g+1)} = \left(1 - \frac{\nu}{2}\right)\sigma_0 \mathrm{e}^{-\frac{\nu}{2}g}. \tag{66}$$

Inserting Eq. (65) and Eq. (66) into Eqs. (42) and (54), corresponding to (C.1) and (C.2) in Tab. 4, yields after modification

$$\nu b_i = \frac{2\sigma_0 c_{\mu/\mu,\lambda} a_i b_i}{\sqrt{\sum_{j=1}^{N} a_j^2 b_j}} - \frac{\sigma_0^2}{\mu}, \tag{67}$$

$$\nu = -2\frac{(1-c)}{DNc}\mu c^2_{\mu/\mu,\lambda} \sum_{i=1}^{N} \frac{\frac{a_i^2 b_i}{\sum_{j=1}^{N} a_j^2 b_j} - \frac{a_i \sigma_0}{\mu c_{\mu/\mu,\lambda}\sqrt{\sum_{j=1}^{N} a_j^2 b_j}}}{1 + \frac{(1-c)}{c}\frac{c_{\mu/\mu,\lambda} a_i \sigma_0}{\sqrt{\sum_{j=1}^{N} a_j^2 b_j}}}. \tag{68}$$

Taking Eq. (63) into account, one obtains a nonlinear system of $N+1$ equations

$$\nu b_i = \frac{2\sigma^*_{\rm ss} c_{\mu/\mu,\lambda} a_i b_i}{\Sigma a} - \frac{{\sigma^*_{\rm ss}}^2 \sum_{j=1}^{N} a_j^2 b_j}{\mu\,(\Sigma a)^2}, \tag{69}$$

$$\nu = -2\frac{(1-c)}{DNc}\mu c^2_{\mu/\mu,\lambda} \sum_{i=1}^{N} \frac{\frac{a_i^2 b_i}{\sum_{j=1}^{N} a_j^2 b_j} - \frac{a_i \sigma^*_{\rm ss}}{\mu c_{\mu/\mu,\lambda}\Sigma a}}{1 + \frac{(1-c)}{c}\frac{c_{\mu/\mu,\lambda} a_i \sigma^*_{\rm ss}}{\Sigma a}}, \tag{70}$$

where ν, b_i, and σ^*_{ss} are unknown time-independent steady state quantities. Notice, Eq. (70) contains the SAR function (55) revealing the relation

$$\boxed{\frac{\nu}{2} = -\tilde{\psi}\,(\sigma^*_{ss}).} \tag{71}$$

Comparing with Eq. (62), one sees that the rate by which σ evolves in the steady state is given by the (negative) value of the SAR function $\tilde{\psi}$.

Equation (69) can be rewritten in matrix form resulting in an eigenvalue problem

$$\mathbf{A} \cdot \mathbf{b} = \nu \mathbf{b}, \tag{72}$$

with eigenvector $\mathbf{b} = (b_1, b_2, \ldots, b_N)^\top$, steady state mutation strength $\sigma^*_{ss} = \text{const.}$, and an $N \times N$ matrix A component-wise given as

$$(\mathbf{A})_{ii} = 2\sigma^*_{ss} c_{\mu/\mu,\lambda} \frac{a_i}{\Sigma a} - \frac{{\sigma^*_{ss}}^2 a_i^2}{\mu(\Sigma a)^2}, \tag{73}$$

$$(\mathbf{A})_{ij} = \frac{-{\sigma^*_{ss}}^2 a_j^2}{\mu(\Sigma a)^2}, \quad i \neq j. \tag{74}$$

The matrix $\mathbf{A}$ has N eigenvalues ν and N eigenvectors $\mathbf{b}$. Due to the conditions of the *Ansatz* only those solutions of the eigenvalue problem with $\forall i : \; b_i > 0$ and $\nu > 0$ are acceptable. The *Ansatz* indicates that larger eigenvalues ν result in a faster decay of the y_i^2 and σ values. That is, in comparison to the smallest eigenvalue the impact of the larger ν values will decrease with $g \to \infty$. This corresponds to the faster decay of y_i^2-values in the initial and transient phase of the evolution process, see e.g. Fig. 1. Therefore, as for the steady state behavior, one is interested in the smallest eigenvalue ν. Additionally, the corresponding eigenvector $\mathbf{b}$ has to satisfy the condition $\forall i : \; b_i > 0$. Considering the models $a_i = 1, i, i^2$, and additionally $a_i = \sqrt{i}$, the corresponding smallest eigenvalues resulting from (72) depending on the normalized mutation strength are shown in Fig. 8. The ellipsoid model $a_i = \sqrt{i}$ is included to better assess the transition from the sphere model to the ellipsoid model. For $a_i = i$, as well as $a_i = i^2$, the numerically obtained points exhibit a linear growth over a wide range of σ^* values. While its coefficients are closer to those of the sphere model, even the $a_i = \sqrt{i}$ case exhibits this behavior to a certain extend. The general tendency of the eigenvalue dependence is characterized by a linear ascent and a sudden sharp drop in the vicinity of the zero progress rate value of the normalized mutation strength at $\sigma^* = 2\mu c_{\mu/\mu,\lambda}$, cf. Eq. 14. The numerical results presented in Fig. 8 are identical to the results presented in (Beyer and Melkozerov, 2014) in context of the self-adaptation ES.

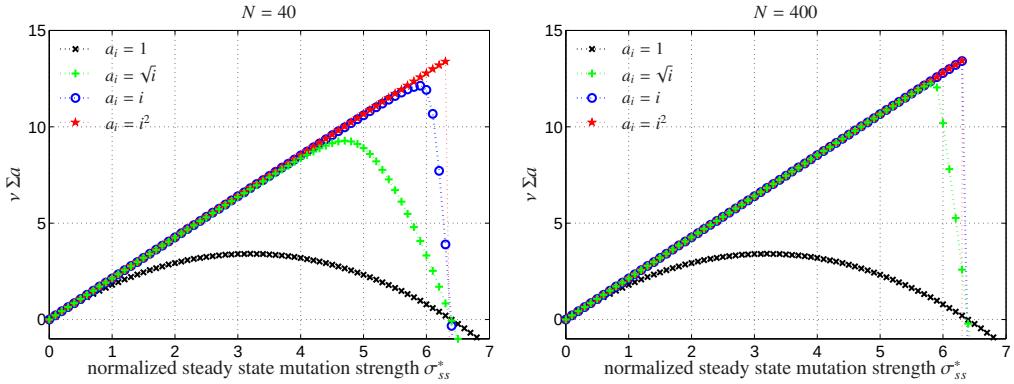

Figure 8: On the dependence of the numerically calculated smallest eigenvalues (multiplied by Σa) on the steady state mutation strength σ^*. The results of the (3/3, 10)-CSA-ES are presented for the ellipsoid model with coefficients $a_i = 1, \sqrt{i}, i, i^2$. The search space dimension is $N = 40$ (left figure) and $N = 400$.

Due to the linear behavior of the numerically obtained data points for sufficiently small values of σ^* it is possible to derive an analytical expression which describes the early growth behavior of the eigenvalues. Therefore, the quadratic σ^* terms within the eigenvalue problem (72) are neglected. This transforms Eqs. (72) and (73) into a diagonalized problem

$$\forall\, i \neq j:\ (\mathbf{A})_{ij} = 0 \qquad \text{and} \qquad (\mathbf{A})_{ii} = 2\sigma^*_{ss} c_{\mu/\mu,\lambda} \frac{a_i}{\Sigma a}. \tag{75}$$

Consequently, the N eigenvalues are identified as $\nu_i = 2\sigma^*_{ss} c_{\mu/\mu,\lambda} \frac{a_i}{\Sigma a}$. Since the steady state dynamics of the ES are governed by the smallest positive eigenvalue (see discussion above), the linear part for that $\nu(\sigma^*)$ corresponds to the smallest coefficient a_i. Writing

$$\check{a} := \min_{j=1,\ldots,N} a_j \tag{76}$$

for the smallest coefficient, the linear parts of the curves in Fig. 8 can be expressed by

$$\nu_{lin}(\sigma^*_{ss}) = 2\sigma^*_{ss} c_{\mu/\mu,\lambda} \frac{\check{a}}{\Sigma a}. \tag{77}$$

This linear approximation of the steady state mode eigenvalue was already derived in (Beyer and Melkozerov, 2014) and revealed good agreement with numerically obtained results for sufficiently small values of the normalized steady state mutation strength σ^*_{ss}.

7.3 The Normalized Mutation Strength in the Steady State

This section aims at the calculation of the normalized steady state mutation strength σ^*_{ss}. To this end, a difference equation which describes the change of the normalized mutation strength between two consecutive generations of the ES is derived. Starting point is the iterative scheme C of evolution equations. Considering (C.2) and the respective Eqs. (54) and (55), and proceeding to normalized quantities, Eq. (3), leads to

$$\sigma^{*(g+1)} \frac{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g+1)^2}}}{\Sigma a} = \sigma^{*(g)} \frac{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}{\Sigma a} \left(1 + \tilde{\psi}(\sigma^{*(g)})\right). \tag{78}$$

This equation can be rearranged to

$$\sigma^{*(g+1)} = \sigma^{*(g)} \frac{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g+1)^2}}} \left(1 + \tilde{\psi}(\sigma^{*(g)})\right). \tag{79}$$

Concentrating on the quotient of square roots taking (42) into account (neglecting fluctuations) yields

$$\frac{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g+1)^2}}} \simeq \sqrt{\frac{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}{\sum_{j=1}^N a_j^2 y_j^{(g)^2} - \sum_{j=1}^N a_j^2 \varphi_j^{II}(\sigma^{(g)})}} = \sqrt{\frac{1}{1 - \frac{\sum_{j=1}^N a_j^2 \varphi_j^{II}(\sigma^{(g)})}{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}}. \tag{80}$$

On the understanding that the term $\sum_{j=1}^N a_j^2 \varphi_j^{II}(\sigma^{(g)}) \big/ \sum_{j=1}^N a_j^2 y_j^{(g)^2}$ is sufficiently small, see also (29), Eq. (80) is approximated by

$$\frac{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}}{\sqrt{\sum_{j=1}^N a_j^2 y_j^{(g+1)^2}}} \simeq 1 + \frac{1}{2} \frac{\sum_{j=1}^N a_j^2 \varphi_j^{II}(\sigma^{(g)})}{\sum_{j=1}^N a_j^2 y_j^{(g)^2}}. \tag{81}$$

Making use of (12) and the normalization (5), the fraction on the rhs of (81) becomes

$$\frac{\sum_{i=1}^{N} a_i^2 \varphi_i^{II}(\sigma^{(g)})}{\sum_{j=1}^{N} a_j^2 y_j^{(g)2}} = \frac{2\sigma^{*(g)} c_{\mu/\mu,\lambda}}{\Sigma a} \frac{\sum_{i=1}^{N} a_i^3 y_i^{(g)2}}{\sum_{j=1}^{N} a_j^2 y_j^{(g)2}} - \frac{\sigma^{*(g)2}}{\mu} \frac{\sum_{i=1}^{N} a_i^2}{(\Sigma a)^2}, \tag{82}$$

and in conclusion

$$\frac{\sqrt{\sum_{j=1}^{N} a_j^2 y_j^{(g)2}}}{\sqrt{\sum_{j=1}^{N} a_j^2 y_j^{(g+1)2}}} \simeq 1 + \frac{\sigma^{*(g)} c_{\mu/\mu,\lambda}}{\Sigma a} \frac{\sum_{i=1}^{N} a_i^3 y_i^{(g)2}}{\sum_{j=1}^{N} a_j^2 y_j^{(g)2}} - \frac{\sigma^{*(g)2}}{2\mu} \frac{\sum_{i=1}^{N} a_i^2}{(\Sigma a)^2}. \tag{83}$$

Thus, the evolution equation (79) of the normalized mutation strength σ^* becomes asymptotically

$$\sigma^{*(g+1)} \simeq \sigma^{*(g)} \left(1 + \frac{\sigma^{*(g)} c_{\mu/\mu,\lambda}}{\Sigma a} \frac{\sum_{i=1}^{N} a_i^3 y_i^{(g)2}}{\sum_{j=1}^{N} a_j^2 y_j^{(g)2}} - \frac{\sigma^{*(g)2}}{2\mu} \frac{\sum_{i=1}^{N} a_i^2}{(\Sigma a)^2} \right) \left(1 + \tilde{\psi}(\sigma^{*(g)}) \right). \tag{84}$$

Applying the *Ansatz* (61), (62) to Eq. (84), the σ^* evolution equation after resolving the parentheses reads

$$\sigma^{*(g+1)} \simeq \sigma^{*(g)} \left(1 + \frac{\sigma^{*(g)} c_{\mu/\mu,\lambda}}{\Sigma a} \frac{\sum_{i=1}^{N} a_i^3 b_i}{\sum_{j=1}^{N} a_j^2 b_j} - \frac{\sigma^{*(g)2}}{2\mu} \frac{\sum_{i=1}^{N} a_i^2}{(\Sigma a)^2} + \tilde{\psi}(\sigma^{*(g)}) + \Delta(\sigma^{*(g)}) \right). \tag{85}$$

Notice, due to the complex form of $\tilde{\psi}(\sigma^{*(g)})$, Eq. (55), all mixed product terms of higher orders are aggregated in the error term $\Delta(\sigma^{*(g)})$. Provided that $cD = O(1)$, this error term vanishes at least by a factor of $1/N$ faster than the other terms in the parentheses of (85). This error term is neglected in the next steps in order to keep the further analysis manageable.

By demanding $\sigma^{*(g+1)} = \sigma^{*(g)} = \sigma^*_{ss}$, Eq. (85) can now be used to determine the steady state σ^*. This implies that the expression in the large parentheses becomes 1. Assuming $\Delta(\sigma^{*(g)}) \to 0$ and taking into account Eq. (71) yields the condition

$$\nu \stackrel{!}{=} 2 \frac{\sigma^*_{ss} c_{\mu/\mu,\lambda}}{\Sigma a} \frac{\sum_{i=1}^{N} a_i^3 b_i}{\sum_{j=1}^{N} a_j^2 b_j} - \frac{{\sigma^*_{ss}}^2}{\mu} \frac{\sum_{i=1}^{N} a_i^2}{(\Sigma a)^2}. \tag{86}$$

By factorization of the quotient $\sum_{i=1}^{N} a_i^2 \big/ (\Sigma a)^2$ this transforms into

$$\frac{\nu}{2} \stackrel{!}{=} \frac{\sum_{i=1}^{N} a_i^2}{(\Sigma a)^2} \left(\sigma^*_{ss} c_{\mu/\mu,\lambda} \frac{\Sigma a}{\sum_{i=1}^{N} a_i^2} \frac{\sum_{i=1}^{N} a_i^3 b_i}{\sum_{j=1}^{N} a_j^2 b_j} - \frac{{\sigma^*_{ss}}^2}{2\mu} \right). \tag{87}$$

The bracketed term on the rhs of (87)

$$\tilde{\varphi}^*(\sigma^*_{ss}) \coloneqq \sigma^*_{ss} c_{\mu/\mu,\lambda} \frac{\Sigma a}{\sum_{i=1}^{N} a_i^2} \frac{\sum_{i=1}^{N} a_i^3 b_i}{\sum_{j=1}^{N} a_j^2 b_j} - \frac{{\sigma^*_{ss}}^2}{2\mu}. \tag{88}$$

shares some similarities with the progress rate φ^*_{sp} of the sphere model (Beyer, 2001, p. 217)

$$\varphi^*_{sp}(\sigma^*) = c_{\mu/\mu,\lambda} \sigma^* - \frac{\sigma^{*2}}{2\mu}. \tag{89}$$

Actually, Eq. (89) is recovered in (88) for the sphere model, i.e. $\forall i : a_i = 1$.
Recalling the identity $\nu = -2\tilde{\psi}\,(\sigma^*_{ss})$, see (71), finally yields the steady state condition

$$\boxed{-\tilde{\psi}\,(\sigma^*_{ss}) = \frac{\sum_{i=1}^{N} a_i^2}{(\Sigma a)^2}\,\tilde{\varphi}^*\,(\sigma^*_{ss})\,.} \tag{90}$$

Introducing $a_i = 1$, $\forall i = 1, \ldots, N$ into (90) and considering the corresponding SAR function ψ^*_{sp} yields the well known sphere model steady state condition (Meyer-Nieberg and Beyer, 2005)

$$-\psi^*_{sp}(\sigma^*_{ss}) = \varphi^*_{sp}(\sigma^*_{ss})/N. \tag{91}$$

This is remarkable, the novel analysis approach presented describes the steady state by an equation (90) that is formally similar to the sphere model theory of the σSA-ES.

Condition (90) allows for the calculation of the normalized steady state mutation strength σ^*_{ss}. Regarding both sides of Eq. (90) as functions of σ^*_{ss} the curves intersect at the normalized steady state mutation strength of the respective ES. Considering the $(3/3_I, 10)$-ES on the sphere model ($a_i = 1$) as well as on the ellipsoid model $a_i = i^2$, the resulting graphs are shown in Fig. 9 using search space dimension $N = 400$. In each case, three different choices of the cumulation parameter are considered: $c = 3/\sqrt{N}$, $c = 1/\sqrt{N}$, and $c = 1/N$. The damping parameter is $D = 1/c$. The numerically computed solution of (90) is represented by the black dots. According to (88), the rhs of (90) is independent of the choice of the parameters c and D. On the other hand $\tilde{\psi}(\sigma^*_{ss})$ depends on c and D, see (55). Thus, variations in c lead to relocations of the intersection point. From this behavior the existence of an optimal c value can be conjectured which tunes the ES to operate at maximal progress rate $\tilde{\varphi}^*$.

According to Eq. (55) there is also an influence of D displayed in Fig. 10 for the $N = 40$ case considering the sphere and the $a_i = i$ ellipsoid. The D values are varied holding the cumulation parameter $c = 1/\sqrt{N}$ and $c = 1/N$, respectively, constant. The red solid line in both figures corresponds to the rhs in Eq. (90) together with (88). The $-\tilde{\psi}\,(\sigma^*)$ curves, which depend on the parameter D, are represented by the marked blue lines. As one can see, D influences the slope of $-\tilde{\psi}\,(\sigma^*_{ss})$. That is, increasing D, while keeping c constant, leads to a decrease of the $-\tilde{\psi}$ slope. As a consequence the intersection point of both curves moves to the right, i.e.,

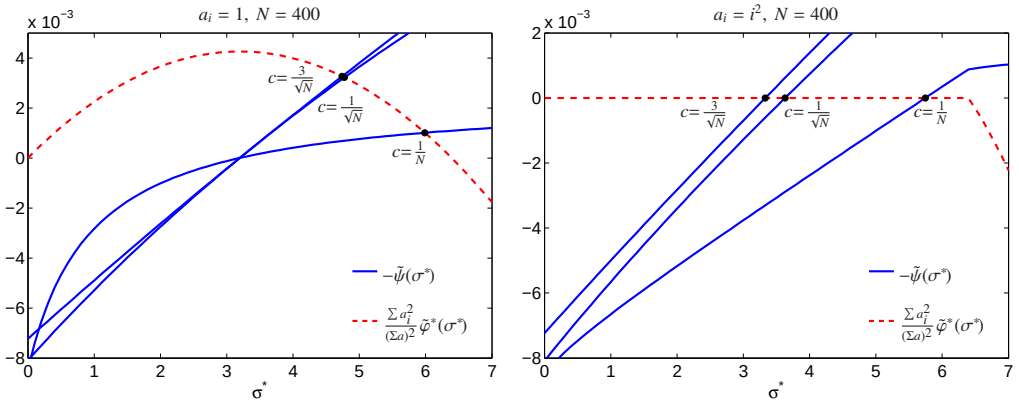

Figure 9: Illustration of steady state condition (90) for a $(3/3_I, 10)$-CSA-ES with $D = 1/c$. Both sides of (90) are plotted against the normalized mutation strength σ^*. The normalized steady state mutation strength σ^*_{ss} is localized at the intersection of the respective curves.

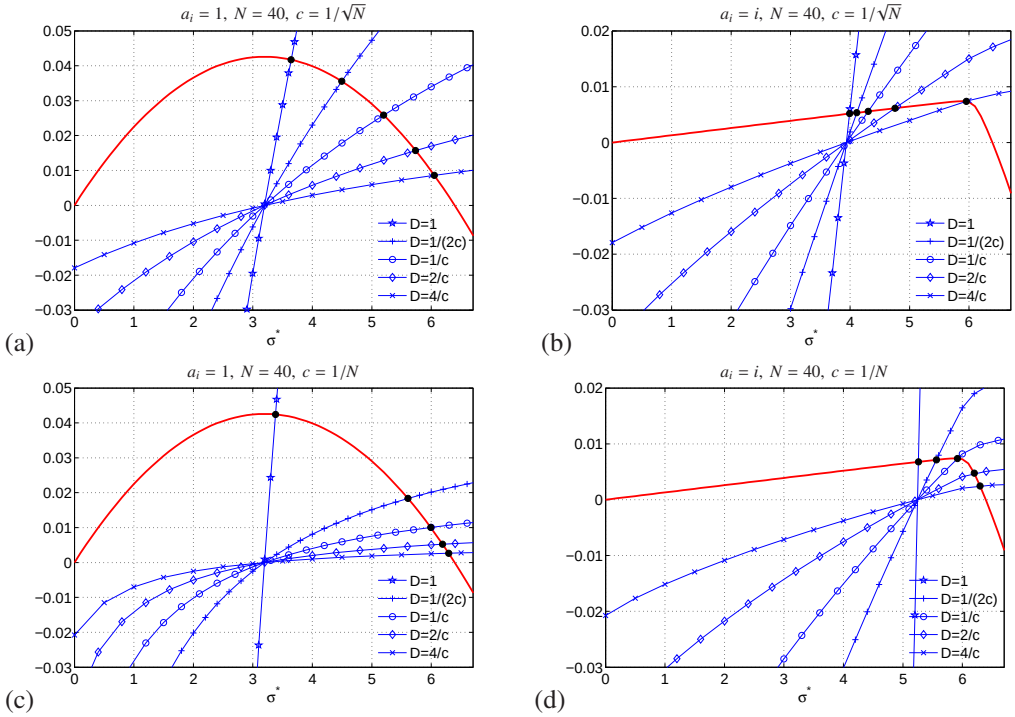

Figure 10: The influence of the damping parameter D on the numerical solution of Eq. (90). The solid (red) line depicts the rhs of (90). The $(3/3_I, 10)$-CSA-ES operates on the sphere, $a_i = 1$, and the ellipsoid model with $a_i = i$, using cumulation parameter $c = 1/\sqrt{N}$ (figures a and b) and $c = 1/N$ (figures c and d) for $N = 40$.

the normalized steady state mutation strength is increased. Independent of the choice of the damping parameter D, all $-\tilde{\psi}(\sigma^*)$ graphs intersect in the same point σ_0^* on the x-axis being the zero of $\tilde{\psi}$. Considering the sphere model this intersection point is independent of c. Since $\forall i : a_i = 1$, one obtains for the root of (55) $\sigma_0^* = \mu c_{\mu/\mu,\lambda}$. In the case of the ellipsoid model with $a_i = i$ this zero varies with the cumulation parameter c. It shifts to the right for smaller c values. However, the corresponding steady state σ_{ss}^* can only be obtained from Eq. (55) by numerical root finding.

The solid (red) line in Fig. 10 represents the rhs of (90), which is by virtue of (88) and (87) equal to half the steady state ν_{ss}. The latter determines via (61) the rate by which the ES approaches the optimizer in the steady state. Since $\nu_{ss}/2$ is determined by σ_{ss}^*, it depends in turn on the choice of D and c. Fig. 11 displays these dependencies. To this end, ν_{ss} is multiplied with the term $\frac{\Sigma a}{2\mu c_{\mu/\mu,\lambda}\check{a}}$ in order to reduce the impact of the considered ellipsoid model as well as the impact of the population sizes on the realized progress. The resulting values are then plotted versus $\tau = 1/c$ being the cumulation time constant that influences the fading of the search path memory within the CSA-ES. The sphere model and the ellipsoid $a_i = i^2$ case are considered. As one can see in Fig. 11 for the ellipsoid with $a_i = i^2$, there is almost no influence of the different damping constant D formulae on the progress rate towards the optimizer in the steady state. This is different to the case of the sphere model. The ellipsoid case $a_i = i$, not displayed in this paper, lies in between these two models.

As for the sphere model, Fig. 11a and b, $D = 1$ seems to be the better choice of the damping

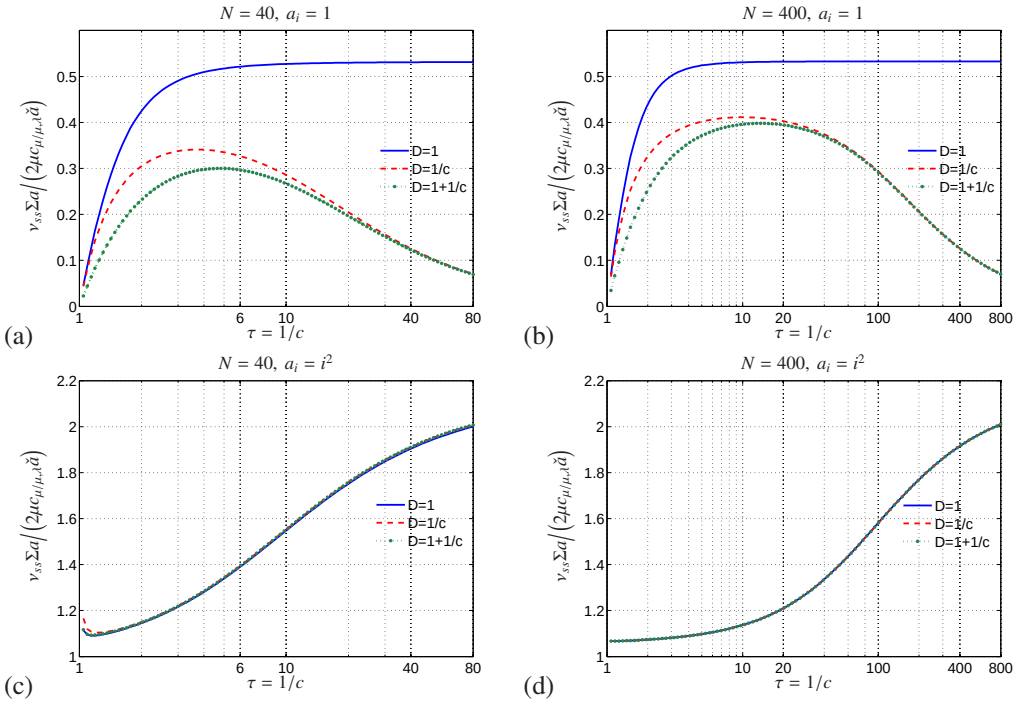

Figure 11: The numerically calculated steady state mode eigenvalue ν_{ss} multiplied by the term $\frac{\Sigma a}{2\mu c_{\mu/\mu,\lambda}\check{a}}$ is plotted against the parameter $\tau = 1/c$. The curves for three different choices of D in search space dimension $N = 40$ and $N = 400$ are illustrated using a $(3/3_I, 10)$-CSA-ES on the sphere (figures a and b) and the ellipsoid with $a_i = i^2$ (figures c and d). Note, the curves regarding the ellipsoid overlap almost completely.

parameter compared to the standard recommendation in (Hansen and Ostermeier, 2001). However, this ignores the effect of possible oscillations that have been neglected by considering the asymptotic solution of the iterative schemes using the *Ansatz* (61), (62). Using small D-values in line 11, Tab. 1, such as $D = 1$, results in large generational σ changes being the driving force of σ^* oscillations already observed by Hansen (1998). These oscillations can lead to considerable regression of the strategy's progress. That is why, larger D values such as $D = \sqrt{N}$ are recommended.

8 Derivation of Closed-Form Expressions for the Steady State

Within this section the normalized steady state mutation strength will be derived which in turn yields the convergence rate ν_{ss} and the expected runtime of the CSA-ES. In order to find a closed-form solution it is necessary to start with assumption (56)

$$\frac{(1-c)}{c}\frac{c_{\mu/\mu,\lambda}a_i\sigma^*_{ss}}{\Sigma a} \ll 1.$$

This allows for the use of the simplified SAR function (60) from iterative scheme D (Table 5) which after normalization (3) reads

$$\tilde{\psi}\left(\sigma^*_{ss}\right) = \frac{\mu c^2_{\mu/\mu,\lambda}}{DN}\frac{(1-c)}{c}\left(1 - \frac{\sigma^*_{ss}}{\mu c_{\mu/\mu,\lambda}}\right) \tag{92}$$

in the strategy's steady state. In the first step, considering only the sphere model ($\forall i = 1, \ldots, N : a_i = 1$) we are able to replicate previous findings from Arnold and Beyer (2004). Afterwards, using the linear approximation of the steady state mode eigenvalue the results are extended to more general ellipsoid models.

8.1 Derivation for the Sphere Model

On the basis of Eq. (90), using (92) and (88), and inserting the coefficients of the sphere model, i.e. $a_i = 1$, the equality

$$-\frac{\mu c_{\mu/\mu,\lambda}^2}{ND}\frac{(1-c)}{c}\left(1-\frac{\sigma_{ss}^*}{\mu c_{\mu/\mu,\lambda}}\right) = \frac{1}{N}\left(c_{\mu/\mu,\lambda}\sigma_{ss}^* - \frac{{\sigma_{ss}^*}^2}{2\mu}\right). \tag{93}$$

is obtained. In order to solve Eq. (93) for σ_{ss}^* it is rearranged to

$${\sigma_{ss}^*}^2 - 2\mu c_{\mu/\mu,\lambda}\left(1-\frac{(1-c)}{cD}\right)\sigma_{ss}^* - 2\mu^2 c_{\mu/\mu,\lambda}^2\frac{(1-c)}{cD} = 0. \tag{94}$$

Resolving for the normalized steady state mutation strength yields after a short calculation

$$\boxed{\sigma_{ss}^* = \mu c_{\mu/\mu,\lambda}\left[1-\tfrac{(1-c)}{cD}+\sqrt{1+\left(\tfrac{(1-c)}{cD}\right)^2}\right].} \tag{95}$$

There are different recommendations regarding the choice of D and c given in (Hansen, 1998)

$$D = \frac{1}{c} \quad \text{with} \quad c = \frac{1}{\sqrt{N}} \tag{96}$$

and (Hansen and Ostermeier, 2001)

$$D = 1 + \frac{1}{c} \quad \text{with} \quad c = \frac{4}{N+4} \tag{97}$$

which do not influence the outcome of (95) as $N \to \infty$. In all these case one gets $cD \to 1$ and $(1-c) \to 1$, thus, yielding

$$\sigma_{ss}^* = \sqrt{2}\mu c_{\mu/\mu,\lambda}. \tag{98}$$

This estimate for the normalized steady state mutation strength of the $(\mu/\mu_I, \lambda)$-CSA-ES on the sphere model was already obtained in (Arnold and Beyer, 2004) using another approach. This result indicates that the steady state σ^* is by a factor of $\sqrt{2}$ too large compared to the optimal sphere model value $\sigma_{opt}^* = \mu c_{\mu/\mu,\lambda}$ that guarantees maximal convergence rate ν towards the optimizer for the sphere model. As we have seen in Sec. 7.3, Fig. 11a and b, choosing larger D can improve the situation as far as σ_{ss}^* is concerned. Equation (95) can be used to tune D to a certain extend to a target mutation strength $\sigma_{target}^* = \mu c_{\mu/\mu,\lambda}(1+\epsilon)$ $(0 < \epsilon < 1)$. Resolving (95) with $\sigma_{target}^* = \sigma_{ss}^*$ for D yields

$$D = 2\frac{\epsilon}{1-\epsilon^2}\frac{(1-c)}{c}. \tag{99}$$

The applicability of the resulting D in real CSA-ES algorithm must, however, be taken with care. As have been already discussed in Sec. 7.3, D-values too small can result in oscillatory σ^* behavior.

8.2 Derivation for the Ellipsoid Model

Considering other ellipsoid models than the special case of the sphere model, the analytical derivation of the normalized steady state mutation strength from Eq. (90) is not possible due to the non-availability of a closed-form expression for (88). Therefore, Eq. (71) is used as starting point and the linear $\nu(\sigma^*)$ approximation for the smallest eigenvalue, Eq. (77), will be applied. That is, $-\tilde{\psi}(\sigma^*_{ss}) = \nu_{lin}(\sigma^*_{ss})/2$ must be solved for σ^*_{ss}. Using the respective expressions (92) and (77), one obtains

$$-\frac{\mu c^2_{\mu/\mu,\lambda}}{DN}\frac{(1-c)}{c}\left(1-\frac{\sigma^*_{ss}}{\mu c_{\mu/\mu,\lambda}}\right) = c_{\mu/\mu,\lambda}\sigma^*_{ss}\frac{\check{a}}{\Sigma a}. \tag{100}$$

Equation (100) can then be solved for the normalized steady state mutation strength σ^*_{ss}:

$$-\frac{\mu c_{\mu/\mu,\lambda}}{DN}\frac{(1-c)}{c} + \frac{\sigma^*_{ss}}{DN}\frac{(1-c)}{c} = \sigma^*_{ss}\frac{\check{a}}{\Sigma a}, \tag{101}$$

yielding finally the normalized steady state mutation strength

$$\boxed{\sigma^*_{ss} = \frac{\mu c_{\mu/\mu,\lambda}}{\left(1-\frac{cDN}{(1-c)}\frac{\check{a}}{\Sigma a}\right)}.} \tag{102}$$

This equation depends only on the ellipsoid model coefficients as well as the choice of strategy parameters of the ES considered. Inserting (102) into (77) yields the linear approximation of the steady state mode eigenvalue

$$\boxed{\nu_{ss} = \frac{\check{a}}{\Sigma a}\frac{2\mu c^2_{\mu/\mu,\lambda}}{\left(1-\frac{cDN}{(1-c)}\frac{\check{a}}{\Sigma a}\right)}.} \tag{103}$$

That is, provided that the linear eigenvalue approximation holds, the CSA-ES approaches by virtue of Eq. (61) the optimizer with the convergence rate (103). With the help of (102), a condition on the choice of D can be derived. Requiring the convergence criterion (14), i.e. $\sigma^*_{ss} < 2\mu c_{\mu/\mu,\lambda}$, one obtains

$$D < \left(\frac{1}{c}-1\right)\frac{\Sigma a}{2N\check{a}}. \tag{104}$$

According to Fig. 8 the linear ν approximation (77) is valid up to a certain $\check{\sigma}^* < 2\mu c_{\mu/\mu,\lambda}$ only. Therefore, the validity of the Eqs. (102, 103, 104) is restricted, too. Additionally, the derivation of the equations by use of (92) is only admissible assuming that condition (58) holds which constrains the range of the cumulation parameter

$$\frac{2\mu c^2_{\mu/\mu,\lambda}\hat{a}}{2\mu c^2_{\mu/\mu,\lambda}\hat{a} + \Sigma a} < c < 1. \tag{105}$$

This must be kept in mind when applying these formulae. Figure 12 compares the theoretical predictions of the steady state convergence rate ν_{ss} with measurements from real $(3/3_I, 10)$-CSA-ES runs as well as with iteratively generated results by making use of scheme A. The ellipsoid models $a_i = i$ and $a_i = i^2$ have been considered. The experimental convergence rates have been obtained by running the $(3/3_I, 10)$-CSA-ES over a sufficient long time until it reaches the steady state. Then the y^2_1 values of the last 25% of generations have been averaged over 100 independent runs. After that, a linear polynomial $\ln y^2_1 = -\nu g + \ln b_1$ has been fitted to the

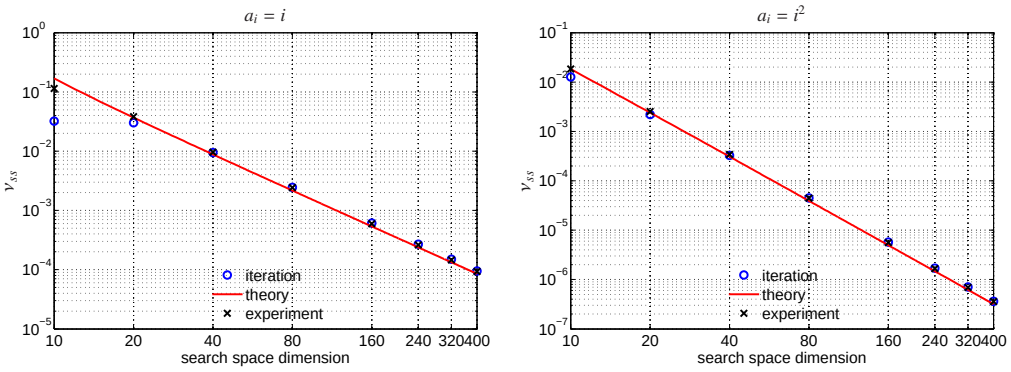

Figure 12: The convergence rates (steady state eigenvalues) realized by the $(3/3_I, 10)$-CSA-ES are plotted against the search space dimension N for the ellipsoid models $a_i = i$ (left figure) and $a_i = i^2$ (right figure). Parameters are $c = 1/\sqrt{N}$ and $D = 1/c$. The experimental data are averaged over 100 independent runs.

experimental y_1^2 data yielding ν in Fig. 12. This curve fitting techniques has also been applied to the iterative y_1^2 values resulting from scheme A of Tab. 2. As one can see, there is a good agreement of the linearized theory with the real ES runs. That is why, the equations obtained can be used to estimate the expected runtime of the CSA-ES.

8.3 Expected Runtime of the CSA-ES on Ellipsoid Models

The strategy's steady state dynamics are governed by an exponential decrease of the y_i components given by Eq. (61) where the inverse time constant ν is determined by (77). Inserting (61) into the ellipsoid fitness model (1) the steady state fitness dynamics can be determined. Starting from generation g_0 sufficiently large such that transient effects have vanished, the fitness value after g generations is

$$F\left(\mathbf{y}^{(g_0+g)}\right) = \sum_{i=1}^{N} a_i b_i \mathrm{e}^{-\nu(g_0+g)} = F\left(\mathbf{y}^{(g_0)}\right)\mathrm{e}^{-\nu g}. \tag{106}$$

Consequently, the objective function value decreases exponentially fast with time constant $1/\nu$. Equation (106) allows for the estimation of the expected running time G in which the fitness value is improved by a factor of $2^{-\beta}$. Considering $F\left(\mathbf{y}^{(g_0+G)}\right)/F\left(\mathbf{y}^{(g_0)}\right)$, from (106) one obtains $\mathrm{e}^{-\nu G} = 2^{-\beta}$ and resolving this for G results in $G = \beta\ln(2)/\nu$. Inserting (77) finally yields

$$G = \frac{\beta\ln(2)}{2\sigma^*_{ss} c_{\mu/\mu,\lambda}} \frac{\Sigma a}{\breve{a}}. \tag{107}$$

That is, the expected runtime G is asymptotically proportional to the quotient of the sum of the ellipsoid coefficients Σa and the smallest coefficient $\breve{a} = \min_i(a_i)$. For the two considered ellipsoid models this means that the expected running time increases with order N^2 for $a_i = i$, and with N^3 for $a_i = i^2$, respectively. The same result had been obtained for the $(\mu/\mu_I, \lambda)$-σSA-ES on the ellipsoid model in (Beyer and Melkozerov, 2014). However, the σSA-ES realizes a different normalized steady state mutation strength σ^*_{ss}. Taking into account the estimation of the normalized steady state mutation strength for non-spherical ellipsoid models in (102) the minimal expected running time becomes

$$\boxed{\breve{G} = \frac{\beta\ln(2)}{2\mu c^2_{\mu/\mu,\lambda}}\left(\frac{\Sigma a}{\breve{a}} - \frac{cDN}{(1-c)}\right).} \tag{108}$$

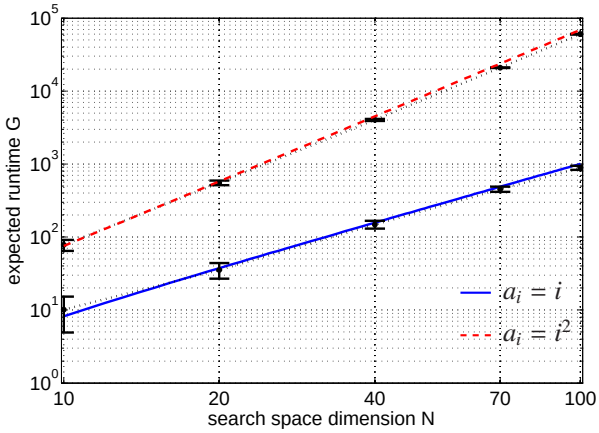

Figure 13: Expected runtime for reducing the objective function value by a factor of $1/4$ by the $(3/3_I, 10)$-CSA-ES on the ellipsoids $a_i = i$ and $a_i = i^2$ using $D = 1/c$ and $c = 1/\sqrt{N}$. The predictions of (108) are displayed by the solid and the dashed line, respectively. Mean and standard deviation of 10 experimental runs are represented by the error bars.

Notice, in order to guarantee validity, the cumulation parameter c and the damping parameter D in Eq. (108) have to be chosen according to (105) and (104), respectively. Smaller choices of c, e.g., would contradict the use of the iterative scheme D. Additionally, it has to be kept in mind that (108) is not applicable to the sphere model since the derivation relies on the use of the linear approximation of the steady state mode eigenvalue (77) which is not appropriate for the sphere (see also Fig. 8). In Fig. 13 the expected runtimes of the $(3/3_I, 10)$-CSA-ES on the ellipsoid models $a_i = i, i^2$ are plotted versus the search space dimension N. Using cumulation parameter $c = 1/\sqrt{N}$ and damping parameter $D = 1/c$, the theoretical predictions for $\beta = 2$ are validated by comparison with real ES runs. The experimental measurements are averaged over 100 independent ES runs. The standard deviation of the runs is depicted by error bars. A good agreement between theoretical and experimental results is observed. Furthermore, with growing search space dimension N the standard deviation decreases.

9 Conclusions

In this paper we have extended the dynamical systems analysis approach developed for the $(\mu/\mu_I, \lambda)$-σSA-ES in (Beyer and Melkozerov, 2014) to the analysis of the CSA-ES on the ellipsoid model. To this end, an analysis approach for the dynamics of the search path cumulation of the CSA-ES has been developed being the basis for the analysis of the σ-evolution in the CSA-ES.

The dynamics of the $(\mu/\mu_I, \lambda)$-CSA-ES on the ellipsoid model are characterized by a transient and a steady state phase. This behavior is very similar to the dynamics of the self-adaptive $(\mu/\mu_I, \lambda)$-σSA-ES. The main difference concerns the steady state normalized mutation strength $\sigma^*_{\rm ss}$ when using standard strategy parameter settings. While for typical strategy parameter choices and learning parameter $\tau \propto 1/\sqrt{N}$, the mutation strength $\sigma^*_{\rm ss}$ is in the vicinity of 1 for the σSA-ES (Beyer and Melkozerov, 2014), the CSA-ES yields $\sigma^*_{\rm ss} \propto \mu c_{\mu/\mu,\lambda}$. That is, in contrast to the σSA-ES, the normalized mutation strength in CSA-ES is proportional to the number of parents μ. As a result, using standard choices for the strategy parameters, the CSA performs up to a factor of μ better than the σSA-ES. As have been shown in (Beyer and Melkozerov, 2014), one can tune the σSA-ES for performance optimality, thus, reaching the same performance as

the CSA-ES. However, to this end, the learning parameter must be chosen proportionally to the square root of the quotient of the smallest eigenvalue of the Hessian and its trace - quantities that are usually not known in a black-box optimization scenario.

While there is this sensitive dependency of the performance of the σSA-ES on the learning parameter τ that depends on the Hessian, there are also strategy parameter dependencies in the CSA-ES. There are two parameters D and c that influence the performance of the CSA-ES. Given a fixed c, one can infer from Fig. 10 and Eq. (55) that D values greater than 1 result in flatter ψ-curves. These in turn shift the steady state σ^* to the right. Depending on the fitness model, this can increase the steady state ν, thus increasing the performance of the ES, but it can also result in smaller ν values. This is especially striking for the sphere model, where by virtue of Eq. (95) the standard recommendation $D = 1 + 1/c$ (Hansen and Ostermeier, 2001) results in a $\sigma^*_{\text{ss}} = \sqrt{2}\mu c_{\mu/\mu,\lambda}$ that is by a factor of $\sqrt{2}$ larger than the optimal σ^*. Therefore, from viewpoint of the sphere model, choosing $D = 1$ should be preferred to the choice $D = 2 + N/4$ that results from Eq. (98). By implication of slower progress, such a choice would also work for the ellipsoid model and would even simplify the σ update in the pseudo code of the CSA-ES in Tab. 1, line 11. However, using $D = 1$ is bought at the expense of oscillatory behavior of σ^* when approaching fitness landscape conditions similar to a sphere model.

The results obtained for the CSA-ES may also have some implications for the CMA-ES. Since this is designed to evolve a covariance matrix that, from the viewpoint of the CSA, transforms the initial optimization problem locally into a spherical one, such oscillations may likely occur at the end of the learning phase. As a consequence, the performance may degrade. That is why, using $D = 1$ cannot be generally recommended. Alternatively, one may use the D formula (99) for algorithm tuning. Yet, considering problems of very high search space dimensionality, using a constant D instead of an N dependent value might be an option to slightly increase the performance of the CMA-ES after having reached the steady state.

Besides the choice of D, there is still a certain freedom of choice concerning c. The reciprocal of which controls the time horizon of path length cumulation. Our investigations have shown that the recommended setting of c within $c \propto [1/\sqrt{N}, 1/N]$ has a rather limited influence on the performance of the ES: Smaller c-values result in a somewhat longer transient time, but in a slightly better steady state performance.

The analysis presented concerned the non-noisy fitness case. As a next step, one should investigate noisy fitness functions as well as dynamically changing fitness functions. Furthermore, the analysis approach should also be able to tackle the dynamics of the CMA-ES. The first step has already been completed: the modeling of the cumulative step size adaptation.

Acknowledgements

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

References

Arnold, D. (2002). *Noisy Optimization with Evolution Strategies*. Kluwer Academic Publishers, Dordrecht.

Arnold, D. (2007). On the Use of Evolution Strategies for Optimising Certain Positive Definite Quadratic Forms. In *GECCO-2007: Proceedings of the Genetic and Evolutionary Computation Conference*, pages 634–641, New York. GECCO '07, ACM.

Arnold, D. and Beyer, H.-G. (2004). Performance Analysis of Evolutionary Optimization With Cumulative Step Length Adaptation. *IEEE Transactions on Automatic Control*, 49(4):617–622.

Arnold, D. and Beyer, H.-G. (2010). On the Behaviour of Evolution Strategies Optimising Cigar Functions. *Evolutionary Computation*, 18(4):661–682.

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

Beyer, H.-G. (2001). *The Theory of Evolution Strategies*. Natural Computing Series. Springer, Heidelberg.

Beyer, H.-G. and Finck, S. (2010). Performance of the $(\mu/\mu_I, \lambda)$-σSA-ES on PDQFs. *IEEE Transactions on Evolutionary Computation*, 14(3):400–418.

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

Beyer, H.-G. and Meyer-Nieberg, S. (2006). Self-Adaptation on the Ridge Function Class: First Results for the Sharp Ridge. In Runarsson, T. et al., editors, *Parallel Problem Solving from Nature 9*, pages 71–80, Berlin. Springer.

Hansen, N. (1998). *Verallgemeinerte individuelle Schrittweitenregelung in der Evolutionsstrategie. Eine Untersuchung zur entstochastisierten,koordinatensystemunabhängigen Adaptation der Mutationsverteilung*. Mensch und Buch Verlag, Berlin. ISBN 3-933346-29-0.

Hansen, N. and Ostermeier, A. (2001). Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation*, 9(2):159–195.

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

Jägersküpper, J. (2006). How the (1+1)-ES Using Isotropic Mutations Minimizes Positive Definite Quadratic Forms. *Theoretical Computer Science*, 361(1):38–56.

Melkozerov, A. and Beyer, H.-G. (2010). On the Analysis of Self-adaptive Evolution Strategies on Elliptic Model: First Results. In *Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation*, GECCO '10, pages 369–376, New York, NY, USA. ACM.

Meyer-Nieberg, S. and Beyer, H.-G. (2005). On the analysis of self-adaptive recombination strategies: first results. In *Congress on Evolutionary Computation*, pages 2341–2348. IEEE.

Meyer-Nieberg, S. and Beyer, H.-G. (2012). The Dynamical Systems Approach – Progress Measures and Convergence Properties. In Rozenberg, G., Bäck, T., and Kok, J., editors, *Handbook of Natural Computing*, pages 741–814. Springer, Berlin.

Ostermeier, A., Gawelczyk, A., and Hansen, N. (1994). Step-Size Adaptation Based on Non-Local Use of Selection Information. In Davidor, Y., Männer, R., and Schwefel, H.-P., editors, *Parallel Problem Solving from Nature, 3*, pages 189–198, Heidelberg. Springer-Verlag.

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

Rechenberg, I. (1994). *Evolutionsstrategie '94*. Frommann-Holzboog Verlag, Stuttgart.

Schwefel, H.-P. (1977). *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie*. Interdisciplinary systems research; 26. Birkhäuser, Basel.