On the Steady State Analysis of Covariance Matrix Self-Adaptation Evolution Strategies on the Noisy Ellipsoid Model

Michael Hellwig and Hans-Georg Beyer

©2018 Elsevier. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

The final article version is going to be published in Elsevier’s Theoretical Computer Science.

For citations, please always refer to the published version available at https://doi.org/10.1016/j.tcs.2018.05.016
On the Steady State Analysis of Covariance Matrix Self-Adaptation Evolution Strategies on the Noisy Ellipsoid Model

Michael Hellwig, Hans-Georg Beyer

Abstract

This paper addresses the analysis of covariance matrix self-adaptive Evolution Strategies (CMSA-ES) on a subclass of quadratic functions subject to additive Gaussian noise: the noisy ellipsoid model. To this end, it is demonstrated that the dynamical systems approach from the context of isotropic mutations can be transferred to ES that also control the covariance matrix. Theoretical findings such as the component-wise quadratic progress rate or the self-adaptation response function can thus be reused for the CMSA-ES analysis. By deriving the steady state quantities approached on the noisy ellipsoid model for constant population size, a detailed description of the asymptotic CMSA-ES behavior is obtained. By providing self-adaptive ES with a population control mechanism, despite noise disturbances, the algorithm is able to realize continuing progress towards the optimum. Regarding the population control CMSA-ES (pcCMSA-ES), the analytical findings allow to specify its asymptotic long-term behavior and to identify influencing parameters. The finally obtained convergence rate matches the theoretical lower bound of all comparison-based direct search algorithms.

Keywords: evolution strategies, noisy optimization, self-adaptation, population control, dynamical systems approach, covariance matrix adaptation

1. Introduction

In many real-world applications the influences of various uncertainties increase the complexity of a corresponding optimization problem. Such uncertainties are summarized by the term noise. Noise may originate from sensory disturbances, randomized simulations or modeling insufficiencies. It impairs the
objective function evaluations or even the search space parameters of the optimization problem. In both cases, successive evaluations of the same parameter vector result in different objective function values. The task of finding an optimal solution for this kind of problems is a great challenge for optimization strategies.

In contrast to classical deterministic methods, direct optimization strategies such as Evolutionary Algorithms (EA) benefit from their intrinsically reduced exposure to noise disturbances. Their performance only relies on the observed objective function values of a candidate solution and, for example, the use of corrupted gradient information is evaded. In recent years, EA proved their ability to successfully deal with noisy optimization problems. This is endorsed by both empirical and theoretical investigations [1, 2].

The present paper focuses on the EA subclass of Evolution Strategies (ES) with mutative $\sigma$ self-adaptation ($\sigma$SA). Considering ES, a number of theoretical investigations on different test functions exists [3, 4, 5]. The theoretical analysis on such test functions is essential to understand the working principles of ES in the presence of noise. It allows to identify beneficial strategy parameter settings, supports the design of new successful algorithm variants, or discovers fundamental performance laws.

The work of [6] addressed the complete analysis of the $\sigma$SA-ES on the noise-free ellipsoid model. While the respective analysis only takes into account isotropic mutations within the procreation process of new candidate solutions, that analysis of the ellipsoid model provides generality with respect to arbitrary rotations of the search space. The respective analysis was transferred to the noisy ellipsoid model in [7]. Recent theoretical analyses on the ellipsoid model also confined themselves on ES with isotropic mutations [8, 9]. That is, those investigations omit the covariance matrix adaptation which is an integral part of state-of-the-art ES, like the Covariance Matrix Adaptation ES (CMA-ES) [10] or the Covariance Matrix Self-Adaptation ES (CMSA-ES) [11].

Despite the fact that ES exhibit a certain robustness in the presence of moderate noise disturbances, strong noise degrades the ES performance. At worst, noise may cause the ES to prematurely converge or even to diverge from the optimum. Two common ways to address this potential performance degradation are either resampling of objective function values or increasing the population size of the ES. The first approach aims at a reduction of the noise influence by averaging over a number of $\kappa$ objective function values. The second one uses growing population sizes in order to mitigate the noise impact via the internal intermediate recombination operator of the ES. Considering the $(\mu/\mu_I, \lambda)$-ES on quadratic functions increasing the population size turned out to be advantageous to performing resampling [12, 13]. On the downside, both methods result in escalating effort in terms of function evaluations.

In order to avoid excess of function evaluations adequate situations in the search process at which to apply resampling or to increase the population size have to be discovered. For detection of the right point to apply the above mentioned countermeasures, a number of strategies have been introduced over the years. These strategies range from simply reevaluating the objective function.
value of a single candidate solution several times to more sophisticated strategies like the uncertainty handling covariance matrix adaptation Evolution Strategy (UH-CMA-ES) [14] that considers rank changes within the offspring individuals after resampling their fitness ($\kappa = 2$). A small number of rank changes indicate low noise intensity and vice versa. While the latter approach realizes a tangible notion of the noise intensity, it still turns out to be too pessimistic [5]. Even a lot of rank changes can be tolerated by the ES still being able to realize progress towards the optimum due to the genetic repair effect induced by the intermediate recombination [15]. That is, the population size matters and must be controlled in order to approximate the optimum as efficiently as possible. Already previously, a residual error-based population size control rule was introduced in [13]. The respective strategy increased the population size if the fitness dynamics on average do not exhibit further progress. The respective work revealed that standard self-adaptive ES exhibit an advantageous behavior in the presence of additive fitness noise. Unlike Cumulative Step-size Adaptation (CSA) which shows a continuous mutation strength decrease, $\sigma$-Self-Adaptive ES ($\sigma$SA-ES) approach a steady state mutation strength. In fact, this bias towards larger step-sizes [16] is beneficial in noisy fitness environments as it helps to prevent premature convergence. For instance, the work of [17] revealed that a population control CMSA-ES variant, the pcCMSA-ES, does not behave like a “simple ES” [5] on the noisy ellipsoid model subject to additive Gaussian noise. That is, in the presence of strong noise the mutation strength of the pcCMSA-ES is not scaling with the distance to the optimum. Instead, the $\sigma$ dynamics approach a steady state $\sigma_{ss}$ as the strategy approaches the optimum. A first theoretical analysis of the pcCMSA-ES was sketched in [18].

As it turns out, on the general ellipsoid model the dynamical systems approach appears to be transferable to self-adaptive ES that control their covariance matrix. To this end, the covariance matrix adaptation can be interpreted as a transformation of the fitness environment rather than a change of the mutation distribution. Essential parts of the analyses [6, 7] can thus be applied in the context of covariance-adaptive ES, and particularly to the CMSA-ES. Section 3 of the present paper is devoted to the clarification of this proposition. Finding generalizations and providing additional insight into the working principles of the CMSA-ES in the presence of additive fitness noise on the ellipsoid model, the present paper is an extension of [18]. Apart from stating the motivations and the theoretical basis of our derivations more precisely, the present article elaborates on

(i) the transfer of earlier theoretical predictions from the context of $\sigma$SA-ES to CMSA-ES in Sec. 3.1.

(ii) the derivations of the CMSA-ES stationary states in Sec. 4 and Sec. 5. In particular, the paper calculates the general steady-state mutation strength including noisy ellipsoid models that were disregarded in [18].

(iii) the calculation of the pcCMSA-ES long-term dynamics and those of the corresponding convergence rate $CR$ in Sec. 6. To this end, the estimates
of the isolation time $G$ are substantiated and the convergence rate $CR$ in
the marginal case $\kappa = 1$ is refined.

(iv) the experimental verification of the theoretical predictions by taking into
account ellipsoid test functions with different conditioning numbers.

After an introduction of the noisy optimization problem in Sec. 2, Section 3.1
considers the transfer of the analysis approach from ES using isotropic mutations
towards covariance-adaptive ES. The concept of the dynamical systems approach is motivated and the fundamental equations from the context of isotropic ES [7] are presented. Parts of the theoretical analysis are addressing CMSA-ES with appropriate population control mechanism. The pcCMSA-ES [17] can be regarded as a standard example for these cases. Therefore, the working principles of the pcCMSA-ES are recapped in Sec. 3.3.

The first analysis step presented in Sec. 4 is the derivation of the steady
state mutation strength of CMSA-ES on the noisy ellipsoid model. While [18]
excluded some ellipsoid models with respect to the tractability of the analysis,
the steady state derivation is extended to the general ellipsoid model. Providing
a more convenient steady state representation for the ensuing analysis, the result
that was already sketched in [18] is presented in more detail.

Having determined the mutation strength steady state, Section 5 is con-
cerned with the description of the parameter vector dynamics of the CMSA-ES
in the presence of noise. Assuming a fixed population size, a steady state expression of the parameter vector components is derived. This allows to rediscover previously obtained steady state terms of the noisy fitness values [12] as well as
the residual distance from the optimum that is approached by the CMSA-ES algorithm without population adaptation. Up to this point, the analysis can (with respect to certain assumptions) be considered as an asymptotical examination of the CMSA-ES dynamics under additive Gaussian fitness noise.

Section 6 then takes up the concept of population control again. Given a
CMSA-ES that appropriately elevates the population size when the evolutionary
process is subject to strong noise disturbances, it is able to continuously reduce
its residual distance from the optimum. A prototype of such strategies is the
pcCMSA-ES. In this regard, Sec. 6 investigates the long-term behavior of the
pcCMSA-ES and provides a theoretical derivation of its convergence rate.

The paper closes with a discussion of the results obtained and points out potential research directions.

2. The noisy optimization problem

The paper considers the fitness environment defined by the noisy ellipsoid model. The ellipsoid model test function class is a subset of quadratic functions, particularly the family of positive definite quadratic forms (PDQFs). Consider the general PDQF formulation with positive definite matrix $B$

$$F(y) = y^\top B y + b^\top y + c.$$  \hspace{1cm} (1)
The linear term $b^\top y$ in (1) can be resolved by use of the translation

$$y \mapsto y' - \hat{c} \quad \text{with} \quad \hat{c} = \frac{1}{2} B^{-1} b.$$  

(2)

Consequently, the quadratic form becomes

$$F(y') = (y' - \hat{c})^\top B (y' - \hat{c}) + b^\top (y' - \hat{c}) + c$$

$$= y'^\top B y' - 2\hat{c}^\top B y' + \hat{c}^\top B \hat{c} + b^\top y' - b^\top \hat{c} + c.$$  

(3)

The linear term vanishes because of the choice of the translation vector $\hat{c}$ in (2). Applying some straightforward rearrangements one obtains

$$F(y') = y'^\top B y' + (\hat{c}^\top B \hat{c} - b^\top \hat{c} + c).$$  

(4)

The bracketed terms are constant and do not contribute to the problem complexity. That is, w.l.o.g. they can be neglected. Further, each positive definite symmetric matrix $B$ can be transformed into diagonal form $\bar{B} = \text{diag}(b_1, \ldots, b_N)$ by application of an appropriate basis transformation. Notice, that the diagonal elements $b_i$ of $B$ directly represent the eigenvalues of the matrix $B$.

Consequently, the notation

$$F(y) = \sum_{i=1}^{N} b_i y_i^2$$

already represents the general form of a quadratic function with a positive definite Hessian matrix. Equation (5) is referred to as the ellipsoid model. The parameter $N$ represents the corresponding search space dimensionality and $b_i$ are referred to as the coefficients of the ellipsoid model. Following the well-known sphere model, the name ellipsoid model originates from the shape of the contour lines of the objective function $F$ which defines ellipsoids in the $N$-dimensional search space. The corresponding unconstrained optimization problem of minimizing Eq. (5) over all $y \in \mathbb{R}^N$ reads

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

(6)

The optimum of (6) resides at the origin of the coordinate system, $y^* = 0$. Special cases of the ellipsoid model include the sphere model ($b_i = 1, \forall i$), the cigar function ($b_1 = 1, b_i = \xi, \forall i \neq 1$) and the discus function ($b_1 = \xi, b_i = 1, \forall i \neq 1$) with condition number $\xi > 1$. In order to take into account test problems with diverse magnitudes of the problem conditioning, the ellipsoid models with coefficients $b_i = i$ and $b_i = i^2$ will be considered to substantiate the theoretical predictions.

Regarding the problem of fitness noise in particular, the noise term directly affects the objective function values. The investigated noisy test problem is referred to as the noisy ellipsoid model. It is composed of the noise-free ellipsoid model and an additive noise term

$$\tilde{F}(y) = F(y) + \delta = \sum_{i=1}^{N} b_i y_i^2 + \sigma c \mathcal{N}(0, 1)$$

(7)
where \( \sigma \) denotes the noise strength, i.e. the standard deviation of the normally distributed noise term \( \delta \sim \sigma \epsilon N(0,1) \). One refers to \( F(y) \) as noise-free or ideal fitness value, and \( \tilde{F}(y) \) is denoted noisy or observed fitness value of a candidate solution \( y \), respectively. While other noise models exist, e.g. with respect to different noise distributions or even multiplicative fitness noise models, the present analysis confines to additive Gaussian fitness noise. For more information on ES applied to noisy optimization see [3, 1, 13].

Taking into account the transformation that leads from Eq. (1) to (4), we were able to show that an analysis of the ES behavior on the fitness environment defined by Eq. (5) or by Eq. (7) in the noisy case, respectively, already covers the general case of quadratic functions with positive definite Hessian matrix.

3. Dynamical systems approach and algorithms under investigation

3.1. Generalization towards covariance-adaptive ES

This section aims at the generalization of the dynamical systems approach [6] towards a CMSA-ES analysis. The analysis approach was introduced in the context of the \( \sigma \)SA-ES algorithm using isotropic mutations. Instead, the CMSA-ES adapts the covariance matrix of the offspring distribution and implicitly takes into account the structure of the underlying fitness landscape. This is due to the attempt of covariance matrix adaptation to learn a way of sampling mutation vectors that point in beneficial directions of the search space. Hence, the mutation within the CMSA-ES can no longer be assumed to be isotropically distributed. However, this section motivates that the dynamical systems approach can also be applied in the context of covariance-adaptive ES on the (noisy) ellipsoid model (5) or (7), respectively. To this end, the considerations that were only sketched in [18] are presented in detail.

Beginning with a brief recap of the CMSA-ES, its pseudo code is illustrated in Alg. 1. According to line 4, in each generation an individual standard deviation \( \sigma_l \in R \) is generated for each of the \( \lambda \) offspring. The standard deviation of the mutation step is also referred to as mutation strength. It is determined by the product of the parental mutation strength \( \sigma^{(g)} \) and a log-normally distributed random variate governed by the learning parameter \( \tau_{\sigma} \). The standard recommendation is \( \tau_{\sigma} = 1/\sqrt{2N} \). The mutation vector \( z_l \in \mathbb{R}^N \) is then drawn from a multivariate normal distribution with mean zero, standard deviation \( \sigma_l \) and covariance matrix \( C^{(g)} \). The covariance matrix \( C^{(g)} \) is determined by the distribution of previously generated successful candidate solutions. It contributes to the mutation operator by providing information of a potentially beneficial offspring distribution w.r.t. the respective fitness environment. Each offspring is generated by addition of the parental centroid \( y^{(g)} \) of the previous generation and the mutation vector \( z_l \). The fitness, i.e. the objective function value, \( \tilde{F}_l \) of the offspring is evaluated in line 8. The best offspring provide their individual mutation strengths as well as their mutation vectors to the next generation. To this end, the respective values of the \( \mu \) best offspring with the best fitness values \( \tilde{F}_{m\lambda}, m = 1, \ldots, \lambda \) are recombined in lines 11, 12, and 13. Notice,
Algorithm 1 The \((\mu/\mu_I, \lambda)\)-CMSA Evolution Strategy.

1: Initialization: \(\sigma^{(0)} \leftarrow \sigma^{(\text{init})}; \ y^{(0)} \leftarrow y^{(\text{init})}; \ C^{(0)} \leftarrow I; \ g \leftarrow 0\)

2: repeat
3: for \(l \leftarrow 1 \) to \(\lambda\) do
4: \(\sigma_l \leftarrow \sigma^{(g)} e^\tau_c N(0, 1)\)
5: \(s_l \leftarrow N(0, C^{(g)})\)
6: \(z_l \leftarrow \sigma_l s_l\)
7: \(y_l \leftarrow y^{(g)} + z_l\)
8: \(\tilde{F}_l \leftarrow \tilde{F}(y_l)\)
9: end for
10: \(g \leftarrow g + 1\)
11: \(\langle z \rangle \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} z_m; \lambda\)
12: \(\langle ss^\top \rangle \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} s_m; \lambda s_m; \lambda^\top\)
13: \(\sigma^{(g)} \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} \sigma_m; \lambda\)
14: \(y^{(g)} \leftarrow y^{(g-1)} + \langle z \rangle\)
15: \(C^{(g)} \leftarrow \left(1 - \frac{1}{\tau_C}\right) C^{(g-1)} + \frac{1}{\tau_c} \langle ss^\top \rangle\)
16: until termination condition
17: return \(y\)

\(m; \lambda\) denotes the \(m\)th best out of \(\lambda\) individuals. The parental centroid is then updated in line 14. In line 15, the CMSA-ES updates the covariance matrix \(C^{(g)}\) according to the best search direction samples. Here, \(\tau_c = 1 + N(N + 1)/2\mu\) is used as the standard learning parameter.

Let us compare the offspring procreation step of the CMSA-ES, line 4 to 7 in Alg. 1, with that of the \(\sigma\)SA-ES [6] using isotropic mutations. The generation of \(z_l\) within CMSA-ES addresses correlations between its components and thus involves the covariance matrix \(C^{(g)}\) of the distribution which is successively updated. In contrast, the respective isotropic mutation step, i.e. \(C = I = const.,\) within the \(\sigma\)SA-ES omits these interdependencies. It reads\(^1\)

\[
y_l = y + \sigma_l N(0, I).
\] (8)

In order to apply the dynamical systems approach [6], the offspring generation step in Alg. 1

\[
y_l = y + \sigma_l N(0, C)
\] (9)

must be reformulated. It can be rewritten as

\[
y_l = y + \sigma_l \sqrt{C} N(0, I).
\] (10)

Introducing the notation \(M := \sqrt{C},\) the creation of \(z_l\) can be formulated as a two-stage process. It can be represented by drawing a random variate from an

\(^1\)Note that the following derivations are assuming a single generation of Alg. 1. Hence, the superscripts \((g)\) are omitted for brevity until the end of this section.
isotropic normal distribution with mean zero and standard deviation $\sigma_l$

$$x_l = \sigma_l N(0, I),$$

which is then followed by a suitable transformation of the mutation vector $x_l$
(with respect to $M := \sqrt{C}$)

$$z_l = Mx_l.$$  

The offspring procreation step of the CMSA-ES thus only differs from those of the $\sigma$SA-ES by application of the transformation matrix $M$. This transformation only affects the way the fitness $F_l$ is evaluated within the ES.

The next step is concerned with resolving this issue. To this end, the fitness evaluation of the offspring candidate solutions with respect to Eq. (10) is investigated. Consider a quadratic fitness environment with positive definite $B\tilde{F}(y_l) = y_l^\top By_l + \delta,$ (13)

where $\delta = \sigma, N(0, 1)$ is an additive fitness noise term and $B = B^\top$.

Starting with the objective function (13), the fitness evaluation of an offspring can be reformulated. The fitness evaluation of a single offspring candidate solution (10) of Alg. 1 becomes

$$\tilde{F}(y_l + Mx_l) = (y_l + Mx_l)^\top B(y_l + Mx_l) + \delta$$

$$= y_l^\top By_l + 2y_l^\top B M x_l + x_l^\top M^\top B M x_l + \delta$$

(14)

By writing $d_l^\top = 2y_l^\top B M$ and $A = M^\top B M$, one obtains

$$\tilde{F}(y_l + Mx_l) = y_l^\top By_l + d_l^\top x_l + x_l^\top A x_l + \delta$$

(15)

Being the product of a positive definite matrix $B$ and two full-rank matrices $M^\top$ and $M$, the matrix $A$ itself is a positive definite matrix. Taking into account a single generation of the CMSA-ES, the covariance matrix $C$ as well as the parental centroid $y$ can be regarded as constant during the offspring procreation process. Analog to Eq. (2), the linear term $d_l^\top x_l$ can be removed by a linear translation operation. All $\lambda$ offspring of one generation are evaluated and selected based on the term $x_l^\top A x_l + \delta$ with $A = M^\top B M$ in Eq. (15). Hence, the offspring fitness evaluation according to (13) only differs from that in the $\sigma$SA-ES by a coordinate transformation with respect to $M := \sqrt{C}$. In this regard, covariance matrix adaptation can be interpreted as a change in the fitness environment perceived by an ES using isotropic mutations. Hence, transferring the analysis approach presented in [6] to “CMA-like” Evolution Strategies such as the CMSA-ES is reasonable. This result is also admissible to describe the CMSA-ES dynamics in the noise-free scenario.

Without loss of generality, the remainder of this paper assumes that the positive definite matrix $A$ is available in diagonal form $A = \text{diag}(a_1, a_2, \ldots, a_N)$. Hence, the trace of $A$ is denoted by $\Sigma a := \text{Tr}[A] = \sum_{i=1}^N a_i$ and the coefficients $a_i$ are associated with the eigenvalues of the transformed fitness environment (15).
Figure 1: Illustration of the (60/60, 200)-CMSA-ES dynamics on the noisy ellipsoid model with $b_i = i$, $N = 30$ and $\sigma_i = 1$. Standard parameters of the CMSA-ES and initial values $y^{(init)} = 0.6 \cdot 1$ and $\sigma^{(init)} = 0.3$ are used. For all dynamics, the geometrical mean of 50 independent CMSA-ES runs plotted against the number of generations. In (a) the ideal (noise-free) fitness dynamics and the mutation strength dynamics are plotted. Indicated by the triangles, the (red) line anticipates the theoretical prediction derived in Sec. 5.2, Eq. (68). On the other hand, (b) displays the normalized mutation strength $\sigma^*$ dynamics and the dynamical behavior of the corresponding noise-to-signal ratio $\vartheta := \sigma^*/\sigma^*$. These are obtained according to the normalizations in Sec. 4.2, Eqs. (33), (34), and (36). All dashed horizontal lines anticipate the respective steady state values to be derived in the remainder of this paper.

The usual dynamical $(\mu/\mu_1, \lambda)$-CMSA-ES behavior on the noisy ellipsoid model subject to additive Gaussian noise is illustrated in Fig. 1. Figure 1(a) provides the experimental noise-free fitness value dynamics of the parental centroid $F(y^{(g)})$, as well as the mutation strength dynamics $\sigma^{(g)}$. The red line (* markers) anticipates the prediction derived in Sec. 5.2, Eq. (68). Starting from a fitness measurement in generation $g = 500$, it provides a qualitative description of the CMSA-ES fitness dynamics when operating close to its steady state behavior and with constant population size $\mu$. Figure 1(b) illustrates the corresponding dynamics of the normalized mutation strength $\sigma^*$ and the signal-to-noise ratio $\vartheta = \sigma^*/\sigma^*$. The normalizations are obtained in Sec. 4.2 by use of Eqs. (33), (34), and (36). It can be observed that all four dynamics approach a stationary behavior with growing number of generations. The horizontal lines anticipate the respective steady state values to be derived in Sec. 4 and 5. In Fig. 1(b) a rapid growth of the signal-to-noise ratio $\vartheta$ is visible during the transient phase. This observation is in line with the assumption of sufficiently large $\vartheta$ values in Sec. 4. The rather large initial population size $\mu = 60$ is used to yield a sufficiently low steady state fitness value that in turn enables the algorithm to ensure a good agreement with the theoretically derived steady state values. It has to be mentioned that in all 50 runs the covariance matrix update was stalled after 500 generations in order to prevent a noise induced eigen-value blowup within the covariance matrix $C$. Otherwise, the noisy information gathered in $C$ can cause major deviations from the theoretical predictions. The accumulation of noise within the covariance matrix can be circumvented by turning off the $C$ update in time. For a more detailed discussion refer to Sec. 7.
3.2. Preliminary theoretical results

This section recaps the theoretical basis of the following CMSA-ES analysis steps. As a consequence of the considerations in Sec. 3.1, the analysis approach applied to isotropic mutation operators can be applied to the CMSA-ES on the noisy ellipsoid model. It allows for an analytical description of the dynamical CMSA-ES behavior subject to additive Gaussian noise. Further, the results can be used to specify the transition between consecutive states of the pcCMSA-ES algorithm [17] by use of mean value dynamics along the line of the dynamical systems approach [12].

The deterministic evolution equations of the dynamical system are given by the component-wise quadratic progress rate [6]
\[
\varphi_{II}^I(y^{(g)}, \sigma^{(g)}) := E \left[ y_i^{(g)} - y_i^{(g+1)} \mid y^{(g)}, \sigma^{(g)} \right],
\]
and by the self-adaptation response (SAR) function [19]
\[
\psi(y^{(g)}, \sigma^{(g)}) := E \left[ \sigma^{(g+1)} - \sigma^{(g)} \mid y^{(g)}, \sigma^{(g)} \right].
\]

The respective expectations have already been derived in the context of the \( \sigma \)SA-ES using isotropic mutations on the noisy ellipsoid model in [7]. According to Sec. 3.1, by taking into account the transformation \( A = M^T BM \), the equations can also be applied in the case of covariance adaptation. In its asymptotically exact representation the component-wise quadratic progress rate reads
\[
\varphi_{II}^I(y^{(g)}, \sigma^{(g)}) \simeq \frac{2c_{\mu, \mu, \lambda} \sigma^{(g)} \sigma_i (y_i^{(g)})^2}{\sqrt{\frac{\sigma_i^2}{(\sigma^{(g)})^2} + \sum_{i=1}^{N} 2a_i^2 \left( 2(y_i^{(g)})^2 + (\sigma^{(g)})^2 \right)}}
\]
\[
- \frac{(\sigma^{(g)})^2}{\mu} \left[ 1 + \frac{((\mu - 1)c_{\mu, \mu, \lambda}^2 + c_{\mu, \mu, \lambda}^1) a_i^2 (y_i^{(g)})^2}{4(\sigma^{(g)})^2 + \sum_{i=1}^{N} a_i^2 \left( (y_i^{(g)})^2 + (\sigma^{(g)})^2 \right)} \right]
\]

and the asymptotical SAR function is given by
\[
\psi(y^{(g)}, \sigma^{(g)}) \simeq \tau^2 \left( \frac{1}{2} + \frac{c_{\mu, \mu, \lambda}^1}{\sigma_i^2 (\sigma^{(g)})^2} + \sum_{i=1}^{N} 2a_i^2 \left( 2(y_i^{(g)})^2 + (\sigma^{(g)})^2 \right) \right)
\]
\[
- \frac{c_{\mu, \mu, \lambda}^2 \sigma^{(g)}}{\sqrt{\frac{\sigma_i^2}{(\sigma^{(g)})^2} + \sum_{i=1}^{N} 2a_i^2 \left( 2(y_i^{(g)})^2 + (\sigma^{(g)})^2 \right)}}
\]

\[
\tau^2 \left( \frac{1}{2} + \frac{c_{\mu, \mu, \lambda}^1}{\sigma_i^2 (\sigma^{(g)})^2} + \sum_{i=1}^{N} 2a_i^2 \left( 2(y_i^{(g)})^2 + (\sigma^{(g)})^2 \right) \right)
\]
\[
- \frac{c_{\mu, \mu, \lambda}^2 \sigma^{(g)}}{\sqrt{\frac{\sigma_i^2}{(\sigma^{(g)})^2} + \sum_{i=1}^{N} 2a_i^2 \left( 2(y_i^{(g)})^2 + (\sigma^{(g)})^2 \right)}}
\]
Notice, that the term $c_{\mu/\mu,\lambda} := c_{\mu,\lambda}^{1,0}$ denotes the progress coefficient of the $(\mu/\mu I, \lambda)$-ES. It is a special case of the *generalized progress coefficients* introduced in [15]

$$c_{\mu,\lambda}^{a,b} = \frac{\lambda - \mu}{\sqrt{2\pi}} \left( \frac{\lambda}{\mu} \right) \int_{-\infty}^{\infty} (-t)^b e^{-\frac{a+b+1}{2}t^2} (1 - \Phi(t))^{\lambda - \mu - 1} \Phi(t)^{\mu - a} dt, \quad (20)$$

where $\Phi(t)$ is the cumulative distribution function of a standard normal variate.

It has to be kept in mind that the derivation of Eqs. (18) and (19) in [7] involved the approximation of random variables. In order to yield manageable integrals, these approximations rely on normal distributions. Due to Lindeberg’s condition this approximations improve for increasing $N$, and under the assumption of non-dominating eigenvalues (i.e. coefficients $a_i$ of the ellipsoid model). However, serving as a somewhat more rough approximation for small values of $N$, the Eqs. (18), (19) still provide good agreements with experimental results.

Being first interested in the description of the steady state long-term behavior of the ES algorithm with constant population size, it is necessary to find asymptotically correct solutions to

$$\psi(y^{(g)}, \sigma^{(g)}) = 0, \quad (21)$$

and

$$\varphi_{II}^{(g)}(y^{(g)}, \sigma^{(g)}) = 0. \quad (22)$$

A closed-form solution of this system of $N+1$ algebraic equations with condition $\sigma^{(g)} > 0$ is usually very hard to find. To this end, the analysis tackles the problem by using asymptotically correct simplifications. Further, the calculation of the asymptotical solutions to Eq. (21) and Eq. (22) is performed in a step-wise manner. First, the computation of the steady state mutation strength $\sigma_{ss}$ is addressed in Sec. 4. Afterwards, the derived $\sigma_{ss}$ is used in Sec. 5 to provide a steady state representations of the squared parameter vector components $(y_{ss})^2_i$.

Having found a solution that satisfies the system (21), (22) asymptotically, it is compared to experimental algorithm runs. This way the theoretically derived solution is verified experimentally.

### 3.3. Target of the analysis: The population control CMSA-ES

Aside extending the analysis approach towards self-adaptive ES equipped with covariance matrix adaptation, the present paper considers the analysis of the long-term behavior of an ES variant that uses population control for dealing with noisy fitness environments: the population control Covariance Matrix Self-Adaptation Evolution Strategy (pcCMSA-ES). Considering the problem of identification and mitigation of noise influences on the optimization process, the pcCMSA-ES was introduced in [17]. The pcCMSA-ES employs a mutation operator that relies on the adaptation of a problem specific covariance matrix. Depending on the pcCMSA-ES implementation, the covariance matrix is successively updated at least until a significant noise impact on the fitness evaluations
is detected. Thus the mutations can no longer be assumed to be isotropically distributed. Aiming at a description of the dynamical behavior, the analysis uses the approach presented in 3.2. The considerations in Sec. 3.1 are especially reasonable in the steady state case of strong noise where the pcCMSA-ES turns off the adaptation after the first noise identification (i.e. $A$ does not change anymore).

This section recap the operating principles of the pcCMSA-ES. The corresponding pseudo code can be found in Alg. 2. Building an integral part of Alg. 2, the lines 5 to 17, and line 34, represent the standard ($\mu/\mu, \lambda$)-CMSA-ES, Alg. 1. The pcCMSA-ES specific noise identification and population size control are performed within lines 18 to 33. Line 4 ensures a constant truncation ratio $\nu = \mu^{(t)}/\lambda^{(t)}$ within each isolation period $t$. The algorithm accumulates a sequence $F$ of observed fitness values by evaluating the parental centroid of each generation. After a lead time $\text{wait}$, the last $L$ entries of $F$ are examined for noise influences on the strategy’s convergence behavior. To this end, a linear regression approach is used. The piecewise trend of the sequence of $L$ observed fitness values is approximated by a linear least squares regression model. The slope of the estimated regression line governs the population control. That is, a negative slope is supposed to indicate on-going positive progress towards the optimum. Contrary, a non-negative slope implies progress stagnation or even divergence behavior of the ES. As the estimated slope of the regression line is itself a random variate, a confidence interval is calculated to ensure a correct decision with respect to a predefined significance level. Constructing a test statistic allows for the design of a hypothesis test that provides the decline requirement. The respective procedure is contained in the detection subroutine. There, the parameter $\bar{\alpha}$ refers to the significance level of the hypothesis test. As long as a negative trend is observed, the algorithm acts like the CMSA-ES. Having identified noise influences on the fitness sequence, detection returns the test decision $td = 0$. The performance degradation of the ES is addressed by increasing $\mu$ by the factor $\alpha > 1$. Further, the covariance matrix adaptation is turned off, once the algorithm has encountered significant noise impact by setting the parameter $\text{adjC}$ to zero in line 25. On the other hand, detection returns $td = 1$ if a significant negative trend is observed. In that case, the strategy tries to reduce the effort in terms of function evaluations and decreases the population size $\mu$ by the factor $\beta$. Such reductions are reasonable in fitness environments subject to distance proportional noise. After each test decision the lead time is reset to prevent the next hypothesis test from being biased by fitness values that rely on previous strategy parameter settings. Notice, since all strategy parameters are constant between two consecutive test decisions, the lead time $\text{wait}$ will also be referred to as an isolation period of the pcCMSA-ES in the following analysis. The pcCMSA-ES evolves the populations until a predefined termination condition (usually a maximal budget of function evaluations) is met. Notice, that two counters are inserted into the pcCMSA-ES pseudo code. One records the number of function evaluations $n$ (line 11 and 19), and the other one is used to log the number of isolation periods $t$ (line 30) completed by the algorithm. These lines build an interconnection with the cor-
Algorithm 2 The pcCMSA Evolution Strategy. The linear regression method for noise identification is implemented in the subroutine detection($F_{\text{int}}, \alpha$).

1: initialization: $g \leftarrow 0$; wait $\leftarrow L$; $\text{adj}C \leftarrow 1$; $\sigma^{(0)} \leftarrow \sigma^{(\text{init})}$; $y^{(0)} \leftarrow y^{(\text{init})}$;
2: $\mu^{(0)} \leftarrow \mu^{(\text{init})}$; $\mu_{\text{min}} \leftarrow \mu^{(\text{init})}$; $C^{(0)} \leftarrow I$; $n \leftarrow 0$; $t \leftarrow 0$; $\mathcal{F} = \emptyset$;
3: repeat
4: $\lambda^{(t)} \leftarrow \lfloor \mu^{(t)} / \nu \rfloor$
5: for $l \leftarrow 1$ to $\lambda^{(t)}$ do
6: $\sigma_l \leftarrow \sigma^{(g)} g^{\tau_C} N(0, 1)$
7: $s_l \leftarrow \sqrt{C^{(g)}} N(0, I)$
8: $z_l \leftarrow \sigma_l s_l$
9: $\tilde{y}_l \leftarrow y^{(g)} + z_l$
10: $\tilde{F}_l \leftarrow \tilde{F}(\tilde{y}_l)$
11: $n \leftarrow n + 1$
12: end for
13: $g \leftarrow g + 1$
14: $\langle z \rangle \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} z_{m, \lambda}$
15: $\langle ss^\top \rangle \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} s_{m, \lambda} s_{m, \lambda}^\top$
16: $\sigma^{(g)} \leftarrow \frac{1}{\nu} \sum_{m=1}^{\mu} \sigma_{m, \lambda}$
17: $y^{(g)} \leftarrow y^{(g-1)} + \langle z \rangle$
18: add $\tilde{F}(y^{(g)})$ to $\mathcal{F}$
19: $n \leftarrow n + 1$
20: if $g > L$ \& wait $= 0$ then
21: $\mathcal{F}_{\text{int}} \leftarrow \mathcal{F}(g - L : g)$
22: $td \leftarrow$ detection($\mathcal{F}_{\text{int}}, \bar{\alpha}$)
23: if $td = 0$ then
24: $\mu^{(t+1)} \leftarrow \lfloor \mu^{(t)} / \beta \rfloor$
25: $\text{adj}C \leftarrow 0$
26: else
27: $\mu^{(t+1)} \leftarrow \max(\mu_{\text{min}}, \lfloor \mu^{(t)} / \beta \rfloor)$
28: end if
29: wait $\leftarrow L$
30: $t \leftarrow t + 1$
31: else
32: wait $\leftarrow wait - 1$
33: end if
34: $C^{(g)} \leftarrow \left(1 - \frac{1}{\tau_C}ight)^{\text{adj}C} C^{(g-1)} + \frac{\text{adj}C}{\tau_C} (ss^\top)$
35: until termination condition
36: return $y^{(g)}$
Figure 2: Illustration of a single exemplary pcCMSA-ES run on the noisy ellipsoid models $b_i = i$ and $b_i = i^2$. Standard parameters of the pcCMSA-ES and initial values $y^{(\text{init})} = 1$ and $\sigma^{(\text{init})} = 1$ are chosen. The noise-free fitness dynamics of the parental centroid as well as the mutation strength dynamics are plotted against the number of function evaluations.

responding theoretical long-term analysis in Sec. 6 and have no impact on the pcCMSA-ES performance. For a more detailed presentation of the pcCMSA-ES refer to [17].

Considering the respective noisy fitness environment (7), the empirical investigations in [17] revealed that the pcCMSA-ES is able to perform close to the theoretical lower bound convergence rate of all comparison-based direct search algorithms [20]. The analysis to be presented intends to detail the description of the pcCMSA-ES steady state long-term behavior and to elaborate on the theoretical confirmation of the observed convergence rate obtained in [18].

All theoretical derivations presented in the remainder of this paper are substantiated by comparison to experimental pcCMSA-ES runs. The experiments are conducted on the noisy ellipsoid models $b_i = i$ and $b_i = i^2$, $\forall i = 1, \ldots, N$ in search space dimensionality $N = 30$ using a constant noise strength of $\sigma = 1$. The runs are initialized with $\sigma^{(\text{init})} = 1$, $y^{(\text{init})} = 1$, and population size $\mu^{(\text{init})} = 3$ using a truncation ratio of $\nu = 1/3$. The pcCMSA-ES uses the strategy parameters $\alpha = 2$, $\beta = \sqrt{\alpha}$, $\bar{\alpha} = 0.05$, and $L = \Sigma b$. The use of this isolation time matches the theoretical results obtained as long as the noise impact on the covariance matrix is relatively small. As the pcCMSA-ES is designed to stop the covariance matrix adaptation after the first noise detection, the choice of $L = \Sigma b$ also suffices to realize the predicted descent of the fitness dynamics in the strong noise scenario.

For an exemplary run of the pcCMSA-ES on both ellipsoidal test functions, the corresponding $F$ and $\sigma$ dynamics are displayed in Fig. 2. Using the mentioned setting, it becomes visible that the algorithm is in fact able to maintain a steady state mutation strength while the fitness dynamics are continuously

2For the purpose of testing and reproduction of results, the pcCMSA-ES code is provided in a GitHub repository (https://github.com/hellwigm/pcCMSA-ES).
decreasing with each population elevation. An analytical description of the observed dynamical behavior is provided in the following sections of this paper.

4. Steady state mutation strength derivation

Self-adaptive ES variants like the CMSA-ES show a different dynamical behavior than expected from other ES variants in the presence of additive fitness noise. After a transient phase, the $\sigma$ dynamics begin to fluctuate around a steady state $\sigma_{ss}$ as the strategy approaches a residual distance from optimum, cf. Fig. 1(a). Due to their intrinsic bias towards growing mutation strengths $\sigma$ [16], self-adaptive ES are less exposed to premature convergence in the presence of additive fitness noise. In contrast, the standard CMA-ES tends to erratically decrease the step-size in this noise scenario [21, 12].

Equipped with an appropriate population control mechanism, $\sigma$SA-ES variants are able to maintain progress towards the optimum while preserving the steady state mutation strength [13, 17]. For instance, [17] revealed that the pcCMSA-ES (Sec. 3.3) does not behave like a “simple ES” on the noisy ellipsoid model (7). The term “simple ES” was introduced in [5] referring to Evolution Strategies that scale their mutation strength proportional to the realized distance from the optimum in quadratic fitness environments. Instead, in the presence of strong additive Gaussian noise the $\sigma$ dynamics maintain a steady state $\sigma_{ss}$ while the strategy approaches the optimum (see Fig. 2).

Section 4.1 addresses the derivation of the CMSA-ES steady state mutation strength on the general noisy ellipsoid model (7). In this regard, it provides a generalization of the pcCMSA-ES analysis [18] that ruled out ellipsoid models with few dominating eigenvalues, e.g. the discus function (with condition number $\xi \gg 1$)

$$F(y) = \sum_{i=1}^{N} b_i y_i^2 \quad \text{with} \quad b_1 = \xi, \ b_i = 1 \ , \forall i \neq 1.$$ (23)

A detailed presentation of the more manageable steady-state result that was already provided in [18] for the pcCMSA-ES, follows in Sec. 4.2.

4.1. Steady state mutation strength on the general noisy ellipsoid model

A decisive requirement on the CMSA-ES for the derivation of the steady state mutation strength on the general noisy ellipsoid model (7) is the ability to realize continuous progress towards the optimum, i.e. $(y_i^{(g)})^2 \to 0$. Usually, ES disturbed by noise approach a residual distance depending on population size, noise strength, and search space dimension $N$. However, increasing the population size is able to reduce the residual distance from the optimum and thus realize an ongoing reduction of the fitness. In order to achieve a progressive reduction of the residual distance the ES needs to appropriately control its population sizes $\mu$ and $\lambda$. Note that a constant truncation ratio $\nu = \mu/\lambda$ of the parent and the offspring population is assumed.
Aiming at the derivation of the steady state mutation strength $\sigma_{ss}$, the following analysis only assumes that the considered ES is able to realize a sufficient population size control. On this understanding the condition $(y_i^{(g)})^2 \rightarrow 0$ is satisfied. In this regard, every CMSA-ES variant that successfully adapts its population size in a similar manner is also reflected in the analysis. In particular, the pC CMSA-ES can be regarded as a prototype of such strategies. Nevertheless, the corresponding steady state $\sigma$ derivation [18] was carried out on a subset of ellipsoid models and is now extended. Taking into account Sec. 2, the analysis makes use of the theoretical results derived for ES with isotropic mutations and applies those findings to the more general CMSA-ES. Hence, the covariance matrix adaptation is appropriately taken into consideration.

The steady state of the mutation strength can be defined as the expected mutation strength in the limit of a large number of generations, i.e. $g \rightarrow \infty$,

$$
\sigma_{ss} := \lim_{g \rightarrow \infty} E[\sigma^{(g)}]. \tag{24}
$$

Regarding self-adaptive ES, the mean value dynamics of the expected mutation strength are governed by the SAR function (17). The mutation strength transition from generation $g$ to $g + 1$ is thus expressed by

$$
E[\sigma^{(g+1)}] = \sigma^{(g)} (1 + \psi(y^{(g)}, \sigma^{(g)})), \tag{25}
$$

Considering the noisy ellipsoid model (7), the asymptotically derived SAR function is represented by (19). Taking into account (25), the steady state condition (24) can be equivalently formulated as

$$
\psi(y^{(g)}, \sigma_{ss}) \equiv 0. \tag{26}
$$

Assuming that the ES is able to continuously realize progress towards the optimum, i.e. $(y_i^{(g)})^2 \rightarrow 0$, the contribution of the respective terms to Eq. (19) can be neglected. Consequently, the steady state condition (26) becomes

$$
\frac{1}{2} + e_{1, \mu, \lambda}^{1, 1} \frac{2\sigma_{ss}^4 \sum_{i=1}^{N} a_i^2}{\sigma_i^2 + 2\sigma_{ss}^4 \sum_{i=1}^{N} a_i^2} \frac{2e_{\mu/ \mu, \lambda}^2 \sigma_{ss}^2 \Sigma a}{\sqrt{\sigma_i^2 + 2\sigma_{ss}^4 \sum_{i=1}^{N} a_i^2}} = 0. \tag{27}
$$

Referring to the steady state mutation strength $\sigma_{ss}$ the generation counters are rendered unnecessary. Now, rearranging the terms directly leads to the expression

$$
\frac{1}{4e_{\mu, \lambda}^{1, 1} \sum_{i=1}^{N} a_i^2} \frac{\sigma_{ss}^4}{\sigma_i^2 + 2\sigma_{ss}^4 \sum_{i=1}^{N} a_i^2} \frac{e_{\mu/ \mu, \lambda}^2 \Sigma a}{\sqrt{\sigma_i^2 + 2\sigma_{ss}^4 \sum_{i=1}^{N} a_i^2}} = 0. \tag{28}
$$

Making use of the substitution

$$
17$$
Figure 3: Comparison of the prediction quality of Eqs. (32) and (43). For the noisy ellipsoid model $a_i = i$, and the noisy discus function (23) (with $\xi = N$), the predicted $\sigma_{ss}$ values of the pcCMSA-ES are plotted against the search space dimension. The experimental mutation strength measurements are displayed by the green error bar plots. The dashed red lines illustrate the predicted steady state values of Eq. (32). Eq. (43) is illustrated by the solid blue lines. Note that the theoretical curves are obtained by use of sufficiently large population size $\lambda = 1000$ and a truncation ratio of $\nu = 1/3$.

\[ s := \frac{\sigma_{ss}^2}{\sqrt{\sigma^2 + 2\sigma_{ss}^4 \sum_{i=1}^{N} a_i^2}} \]  

Eq. (28) transforms into a quadratic equation in $s$

\[ s^2 - \frac{c_{\mu,\lambda} \Sigma a}{e_{\mu,\lambda} \sum_{i=1}^{N} a_i^2} s + \frac{1}{4e_{\mu,\lambda} \sum_{i=1}^{N} a_i^2} = 0. \]  

This quadratic equation can easily be solved for $s$ providing the solution $s$ with

\[ s = \frac{c_{\mu,\lambda} \Sigma a}{2e_{\mu,\lambda} \sum_{i=1}^{N} a_i^2} \left( 1 - \sqrt{1 - \frac{c_{\mu,\lambda} \sum_{i=1}^{N} a_i^2}{c_{\mu,\lambda} (\Sigma a)^2}} \right). \]  

Reconsidering (29) yields a description of the steady state mutation strength

\[ \sigma_{ss} = \frac{s \sigma}{\sqrt{1 - 2s^2 \sum_{i=1}^{N} a_i^2}}. \]  

\[ \text{Not yielding a real-valued } \sigma_{ss} \text{ in (32), the second root of Eq. (30) is discarded.} \]
Equation (32) provides a generalization of the steady state mutation strength derived in [18]. Hence, the result in Eq. (32) is also valid for ellipsoid models with few dominating eigenvalues (e.g. for the discus function (23)). The conformity of Eq. (32) with experimental measurements can be observed in Fig. 3.

To this end, two different noisy ellipsoid models (7) with noise strength $\sigma = 1$ are taken into consideration: the discus function (23) and the test function with coefficients $b_i = i, \forall i = 1, \ldots, N$. In Fig. 3 the predicted steady state mutation strength of the pcCMSA-ES is plotted against the search space dimension. The theoretical result of Eq. (32) is presented by the dashed red line. The mean value of the experimental $\sigma$ dynamics mutation strength, measured over the last 10% of the generations realized within $10^7$ function evaluations, as well as the corresponding standard deviations are displayed by the error bar plot. With growing dimensionality (N $\geq$ 20), the prediction shows a good agreement with the experimentally obtained data points. Notice, that the theoretical curves in Fig. 3 are obtained under the assumption that, due to a fixed truncation ratio $\nu = \mu/\lambda = \text{const.}$, the respective progress coefficients $e^{1,1}_{\mu,\lambda}$ and $e^{1,1}_{\mu,\lambda}$ are converging to a constant value for large population sizes $\lambda \rightarrow \infty$.

4.2. Simplified steady state mutation strength derivation

While the Eqs. (31),(32) provide an analytical formulation of the steady state mutation strength on the general noisy ellipsoid model, the expressions are still cumbersome and complicate the establishment of further analysis steps. A more manageable formulation can be obtained under consideration of additional requirements. On the downside, the requirement excludes ellipsoid models with few dominating eigenvalues from the analysis.

The analysis makes use of the commonly used normalized quantities from the context of the noisy ellipsoid model (7), cf. [6],

$$\sigma^* = \sigma^{(g)} \frac{\Sigma a}{\sqrt{\sum_{j=1}^{N} a_j^2 (y_j^{(g)})^2}}$$  \hspace{1cm} (33)

and

$$\sigma^*_{\epsilon} = \sigma_{\epsilon} \frac{\Sigma a}{2 \sum_{j=1}^{N} a_j^2 (y_j^{(g)})^2}.$$  \hspace{1cm} (34)

By insertion of (33) and (34) into Eq. (19) the normalized SAR function is obtained as

$$\psi(\sigma^*) \simeq \tau^2 \left( \frac{1}{2} + c_{\mu,\lambda}^{1,1} \frac{1 + \frac{\sigma^* \sum_{i=1}^{N} a_i^2}{2 (\Sigma a)^2}}{1 + \vartheta^2 + \frac{\sigma^* \sum_{i=1}^{N} a_i^2}{2 (\Sigma a)^2}} - \frac{c_{\mu,\lambda,\lambda} \sigma^*}{\sqrt{1 + \vartheta^2 + \frac{\sigma^* \sum_{i=1}^{N} a_i^2}{2 (\Sigma a)^2}}} \right).$$  \hspace{1cm} (35)

Here, the normalized noise-to-signal ratio is denoted by

$$\vartheta = \frac{\sigma^*_{\epsilon}}{\sigma^*} = \frac{\sigma_{\epsilon}}{2\sigma^{(g)} \sqrt{\sum_{j=1}^{N} a_j^2 (y_j^{(g)})^2}}.$$  \hspace{1cm} (36)

19
Note that the normalized quantities as well as the normalized SAR function (35) omit the superscripts \((g)\) for the sake of brevity.

Taking into account the two conditions on the ellipsoid model,

\[
\frac{\sum_{i=1}^{N} a_i^2}{(\Sigma a)^2} \xrightarrow{N \rightarrow \infty} 0 \quad \text{and} \quad 1 + \vartheta^2 \gg \frac{\sigma^*}{2} \frac{\sum_{i=1}^{N} a_i^2}{(\Sigma a)^2},
\]

the asymptotic representation of the SAR function in normalized quantities (35) simplifies to

\[
\psi(\sigma^*) \simeq \tau^2 \left[ \frac{1}{2} - \frac{c_{\mu/\mu,\lambda} \sigma^*}{\sqrt{1 + \vartheta^2}} \left( 1 - \frac{e_{1,1}^{1,1}}{c_{\mu/\mu,\lambda} \sigma^* \sqrt{1 + \vartheta^2}} \right) \right].
\]

(38)

Note that Eq. (38) was already obtained in [7] in a slightly different formulation. Requiring a sufficiently large noise-to-signal ratio \(\vartheta \gg 1\), the second condition in (37) can be regarded as a relaxation of the requirement used in that respective paper. As a consequence, the second addend (containing the \(e_{1,1}^{1,1}\) term) in Eq. (35) can be neglected yielding

\[
\psi(\sigma^*) \simeq \tau^2 \left( \frac{1}{2} - \frac{c_{\mu/\mu,\lambda} \sigma^*}{\sqrt{1 + \vartheta^2}} \right).
\]

(39)

Notice, provided that a stable steady state mutation strength exists, the second condition in Eq. (37) is validated by reconsidering the definition (36). Due to the constant noise strength \(\sigma\), the noise-to-signal ratio continuously grows as the strategy approaches the optimum, i.e. as the term \(\sqrt{\sum_{j=1}^{N} a_j^2 (y_j(g))^2}\) decreases. If the initial population size is not large enough to approach a sufficiently small residual distance, at least the population control mechanism will ensure the desired \(\vartheta \gg 1\) by appropriately elevating the population size during the evolution. Hence, due to the first condition in (37), the applicability of the second condition will be validated by the final steady state solutions in Eqs. (57), (58) to be derived below. In fact, the conditions (37) hold for a wide range of noisy ellipsoid models (7), but exclude those with few dominate eigenvalues like the discus function (23). Being able to use (39) for the theoretical analysis of the steady state mutation strength significantly simplifies the derivations.

Together with the simplified representation of the normalized SAR function (39), the starting point for the calculation of the steady state mutation strength is the steady state condition (26) in normalized form

\[
\psi(\sigma_{ss}^*) \equiv 0.
\]

(40)

By inserting (39), one obtains an equation in \(\sigma_{ss}^*\)

\[
0 = \tau^2 \left( \frac{1}{2} - \frac{c_{\mu/\mu,\lambda} \sigma_{ss}^*}{\sqrt{1 + \vartheta^2}} \right).
\]

(41)
Neglecting the 1 compared to \( \vartheta \) \((\vartheta \gg 1)\) allows for a straightforward computation of the (asymptotic) normalized steady state mutation strength

\[
\sigma_{ss}^* \simeq \sqrt{\frac{\sigma^*_e}{2c_{\mu/\mu,\lambda}}}.
\]  
(42)

Renormalization with (33) and (34) finally yields the steady state mutation strength \(^4\)

\[
\sigma_{ss} \simeq \sqrt{\frac{\sigma_e}{4c_{\mu/\mu,\lambda}^2\sigma_a}}.
\]
(43)

For the CMSA-ES with constant population size, the prediction of Eq. (43) is displayed by the horizontal dashed (green) line in Fig. 1(a). One observes that the experimentally obtained \(\sigma\) dynamics fluctuate around the derived steady state mutation strength with growing number of generations.

Remembering that the calculation of (43) is based on condition (37), the resulting steady state mutation strength prediction is not exact for all ellipsoid models. For example, it is expected that the predictive quality will deteriorate in the case of the so-called discus function. This can be observed in Fig. 3, where the prediction of Eq. (43) is represented by the solid blue line. Consequently, the prediction quality of the steady state mutation strength obtained in Eq. (43) is less accurate when compared to the more general result in Eqs. (31),(32). However, Eq. (43) provides good predictions of the steady state mutation strength on a variety of ellipsoid models and allows to carry the theoretical analysis forward.

5. The expected progress for a fixed population size

In Sec. 4.2 an analytical expression for the steady state mutation strength of the CMSA-ES was derived under the conditions (37). The next step is concerned with the description of the CMSA-ES behavior using a fixed (but sufficiently large) population size \(\mu\). Regarding CMSA-ES variants with adequate population control such as the pcCMSA-ES in Alg. 2, the consideration of a fixed population size can be identified with the dynamical behavior during a single isolation period in between two consecutive test decisions.

Making use of \(\sigma_{ss}\) in Eq. (43), the investigations will assume that the strategy has already reached a steady state mutation strength \(\sigma^{(g)} \approx \sigma_{ss}\) in the limit of a sufficiently large noise-to-signal ratio \(\vartheta \gg 1\). That is, each \(\sigma_t\) produced in line 6 of Alg. 2 does only slightly fluctuate around \(\sigma_{ss}\) without distinct influence on the progress rate. Based on these assumptions, the expected \((y^{(g)}_i)^2\) dynamics

\(^4\)As a side note, it should be mentioned that (43) can also be derived from Eqs. (31) (32) by taking into account the first condition in (37).
can be determined. Finding the steady state expression of the component-wise search space parameter dynamics \((y_{ss})^2\) allows for the calculation of the expected steady state fitness dynamics within a single isolation period.

5.1. Deriving the steady state quantities

Remember the asymptotical representation of the component-wise quadratic progress rate (18). Considering the conditions (37) introduced in the context of the SAR function, a simplification of the asymptotical progress rate formulation in Eq. (18) can be obtained in a very similar way. Taking into account the normalizations (33), (34) and the progress rate normalization

\[
\varphi_i^{I*} = \varphi_i^{I} \Sigma a,
\]

the normalized progress rate becomes

\[
\varphi_i^{I*}(\sigma^*) \simeq \frac{2c_{\mu/\lambda} \sigma^* a_i y_i^2}{\sqrt{1 + \vartheta^2 + \frac{(\sigma^*)^2}{2} \sum_{i=1}^{N} a_i^2 / (\Sigma a)^2}} - \frac{(\sigma^*)^2}{\mu \Sigma a} \left[ \frac{\sum_{j=1}^{N} a_j^2 y_j^2}{\sum_{j=1}^{N} a_j^2 y_j^2} \right].
\]

Due to the second condition in (37) one obtains

\[
\varphi_i^{I*}(\sigma^*) \simeq \frac{2c_{\mu/\lambda} \sigma^* a_i y_i^2}{\sqrt{1 + \vartheta^2}} - \frac{(\sigma^*)^2}{\mu \Sigma a} \left[ \frac{\sum_{j=1}^{N} a_j^2 y_j^2}{\sum_{j=1}^{N} a_j^2 y_j^2} \right].
\]

The second addend in the square brackets of Eq. (46) consists of two fractions. While the order of the first fraction becomes \(O(1)\), omitting the squared brackets requires to ensure

\[
\sum_{j=1}^{N} a_j^2 y_j^2 \to 0.
\]

This can be regarded as an additional assumption on the ellipsoid model, i.e. implying \(\forall i = 1, \ldots, N : \sum_{j=1}^{N} a_j^2 y_j^2 \gg a_i^2 y_i^2\). However, regarding the steady

\[\text{Given a fixed truncation ratio } \mu/\lambda, e_{\mu,\lambda}^{2,0} \text{ and } e_{\mu,\lambda}^{1,1} \text{ converge to constants as } \mu \to \infty. \text{ Furthermore, } \vartheta \propto \sqrt{\mu}, \text{ as will be shown in Eq. (58) below.}\]
state solutions (53) and (55) one observes that condition (47) is asymptotically equal to
\[
\frac{a_i y_i^2}{\sum_{j=1}^{N} a_j y_j^2} \simeq \frac{a_i}{\sum a_i} \to 0. \tag{48}
\]
This condition is fulfilled provided that the first requirement in (37) is satisfied. After all, the second addend in the squared brackets of Eq. (46) can be neglected and the asymptotic approximation of the normalized component-wise quadratic progress rate is given by
\[
\phi_{II}^* (\sigma^*) \simeq \frac{2 c_{p/\mu,\lambda} \sigma_i^* a_i y_i^2}{\sqrt{1 + \vartheta^2}} - \frac{(\sigma^*)^2}{\mu \sum a_j y_j^2}. \tag{49}
\]
Switching to the renormalized quantities finally results in the representation
\[
\phi_{II}^* (y^{(g)}, \sigma^{(g)}) \simeq \frac{2 c_{p/\mu,\lambda} \sigma_i a_i y_i^{(g)}}{\sqrt{(1 + \vartheta^2) \sum_{j=1}^{N} a_j^2 (y_j^{(g)})^2}} - \frac{\sigma^2}{\mu}. \tag{50}
\]
On the noisy ellipsoid model (7), this asymptotic progress rate formulation was already derived in [7] in a slightly different way. In conjunction with (16), Eq. (50) can be used to calculate the \((y_i^{(g)})^2\) dynamics as carried out in Sec. 5.2.

In the first place, the steady state values \((y_{ss})^2\) are determined. Taking into account the steady state condition \(E[(y_i^{(g+1)})^2] = E[(y_i^{(g)})^2] = (y_{ss})^2\), which holds for a sufficiently large number of generations \(g\), Eq. (16) yields the condition
\[
\phi_{II}^* ((y_{ss})^2) = 0. \tag{51}
\]
By inserting (50) into (51), and considering the renormalized representation of the noise-to-signal ratio (36), Eq. (51) can be solved for the component-wise steady state \((y_{ss})^2\). One obtains the asymptotically correct solution
\[
a_i (y_{ss})^2 \simeq \frac{\sigma \varepsilon}{4 c_{p/\mu,\lambda} \mu} \sqrt{\frac{1}{\vartheta^2} + 1}, \quad \forall i = 1, \ldots, N. \tag{52}
\]
Remembering the prerequisite \(\vartheta \gg 1\), Eq. (52) is further simplified to
\[
a_i (y_{ss})^2 \simeq \frac{\sigma \varepsilon}{4 c_{p/\mu,\lambda} \mu}, \quad \forall i = 1, \ldots, N. \tag{53}
\]
This steady state expression depends on the noise strength as well as on the population size used by the CMSA-ES. Considering the pcCMSA-ES, the steady state value changes with each \(\mu\) adaptation, i.e. it is different in each isolation period. The required number of generations to actually approach the steady state needs to be examined separately in Sec. 5.2. Note that for a fixed truncation ratio \(\nu = \mu / \lambda\), the progress coefficients \(c_{p/\mu,\lambda}\) can roughly be regarded as
constant. That is, the influence of the population growth on the $c_{\mu/\mu,\lambda}$ values can be asymptotically neglected (see [15], p. 249, Eq. 6.113).

Based on (53) the expected steady state fitness $F_{ss}$ that is approached for a fixed population size (within a single isolation period) can be determined. Taking into account the noisy ellipsoid model (7) with additive noise term $\delta \sim \mathcal{N}(0, \sigma^2_{\epsilon})$, it follows that

$$F_{ss} := \lim_{g \to \infty} E[\hat{F}(y^{(g)})] = \sum_{i=1}^{N} a_i(y_{ss})^2 \simeq \frac{N \sigma_{\epsilon}}{4c_{\mu/\mu,\lambda} \mu} \quad (54)$$

The result of Eq. (54) was already obtained more than 12 years ago by use of the equipartition assumption and differential geometry [12]. As a byproduct of our considerations, the present derivation provides a proof based on first principles.

An illustration of the experimental fitness dynamics approaching the predicted steady state fitness is provided in Fig. 4. The fitness dynamics of the pcCMSA-ES are plotted against the number of function evaluations on the ellipsoid models $b_i = i$ and $b_i = i^2$ ($N = 30$). While the CMSA-ES in Alg. 1 would only approach steady state fitness indicated by the first horizontal line of the (red) step function, the pcCMSA steadily adjusts its population size. Thus the pcCMSA-ES is able to decrease its residual distance from the optimum and realize a continuing fitness reduction. It can be observed that the experimental results (solid blue curve) approach the predicted steady state fitness after each $\mu$ elevation. In this respect, the experiment validates the theoretically derived predictions.

Alongside the steady state fitness description, Eq. (53) provides a steady state characterization of the normalized mutation strength dynamics $\sigma^*_{ss}$. Multiplying Eq. (53) with $a_i$ and summing up all components $i = 1 \ldots, N$, one
Figure 5: The normalized steady state mutation strength dynamics of the pcCMSA-ES on the noisy ellipsoid model (7). The respective dynamics are plotted against the number of function evaluations. Experimental measurements are displayed by the solid blue line, while the red step function illustrates the theoretical steady state values with respect to Eq. (57).

obtains the term

\[ \sum_{i=1}^{N} a_i^2 (y_{ss})_i^2 \simeq \frac{\sigma_\epsilon \Sigma a}{4 \mu c_{\mu/\mu,\lambda}}. \]  \hspace{1cm} (55)

Assuming that the ES operates in its steady state, inserting Eq. (55) into the denominator of the \( \sigma \) normalization in (33) yields

\[ \sigma^*_{ss} \simeq \sigma_{ss} \Sigma a \left( \frac{\sigma_\epsilon \Sigma a}{4 \mu c_{\mu/\mu,\lambda}} \right)^{-\frac{1}{2}}. \] \hspace{1cm} (56)

Further, by considering Eq. (43) one obtains

\[ \sigma^*_{ss} \simeq \sqrt{\frac{\sigma_\epsilon}{4 c_{\mu/\mu,\lambda} \Sigma a} \cdot \sqrt{\frac{4 \mu c_{\mu/\mu,\lambda}}{\sigma_\epsilon \Sigma a}} \cdot \Sigma a = \sqrt{\mu}}. \] \hspace{1cm} (57)

Thus, during a single isolation period with constant \( \mu \), the \( \sigma^*_{ss} \) approaches the square root of the parental population size \( \mu \). This dynamical behavior is displayed in Fig. 5. It can be observed that the experimental results follow the theoretical steady state prediction. With growing population size \( \mu \) the normalized mutation strength dynamics exhibit a continuously increasing behavior.

Regarding the noise-to-signal ratio \( \vartheta \) in (36), its steady state dynamics can be characterized by the same token. Applying the normalizations (33), and (34), Eq. (55) as well as (57), yields the steady state of the noise-to-signal ratio

\[ \vartheta_{ss} = \sigma^*_\epsilon / \sigma^*_{ss} \simeq 2 c_{\mu/\mu,\lambda} \sqrt{\mu}. \] \hspace{1cm} (58)

Equation (58) reveals that \( \vartheta_{ss} \) is also proportional to the square root of the population sizes \( \mu \). Having a look at Fig. 1(b), the steady states (57) and (58) are
5.2. Description of the generational progress

The next analysis step is concerned with the description of the generational progress of the CMSA-ES during the isolation period. Still considering a fixed population size, the analysis is based on the asymptotical component-wise quadratic progress rate in Eq. (50). Making use of the renormalized version of the noise-to-signal ratio \( \vartheta \) in Eq. (36) and taking into account a sufficiently large noise-to-signal ratio \( \vartheta^2 \gg 1 \), (50) becomes

\[
\varphi^{I} (y^{(g)}, \sigma^{(g)}) \simeq \frac{2c_{\mu/\mu, \lambda} \sigma^{(g)} a_i y_i^2}{\sqrt{\vartheta^2 \sum_{j=1}^{N} \sigma_j^2 (y_j^{(g)})^2}} = \frac{\left(\sigma^{(g)}\right)^2}{\mu} - \frac{4c_{\mu/\mu, \lambda} \left(\sigma^{(g)}\right)^2 a_i y_i^{(g)}^2}{\sigma_{\epsilon}} - \frac{\left(\sigma^{(g)}\right)^2}{\mu}.
\]  

(59)
This representation of the progress rate allows to derive a linear difference equation that determines the component-wise expected value dynamics of the parameter vector \((y_i^{(g)})^2\) by inserting (59) into Eq. (16). Straightforward rearrangements yield

\[
y_i^{(g+1)} \approx y_i^{(g)} \left(1 - \frac{4c_{\mu/\mu,\lambda} (\sigma^{(g)})^2 a_i}{\sigma_e} \right) + \frac{(\sigma^{(g)})^2}{\mu},
\]

(60)

Note, despite the stochastic dynamics are approximated by the expected values, the expectation operator \(\mathbb{E}[\cdot]\) has been dropped here for brevity.

Keeping in mind the assumption that the ES is operating with a constant mutation strength close to its steady state value, i.e. \(\sigma = \sigma^{(g)} \approx \sigma_{ss}\), Eq. (60) represents a linear difference equation of the general type

\[
z^{(g+1)} = z^{(g)} (1 - u) + v
\]

with \(z = \mathbb{E}[y_i^{(g)}]^2\),

\[
u := \frac{4c_{\mu/\mu,\lambda} \sigma^2 a_i}{\sigma_e} \quad \text{and} \quad v := \frac{\sigma^2}{\mu}.
\]

(61)

A closed-form solution of which is provided by

\[
z^{(g_0+g)} = z^{(g_0)} (1 - u)^g + \frac{v}{u} [1 - (1 - u)^g],
\]

(62)

where \(g_0\) is the specific generation number at which a (new) isolation period starts. Considering (61), the closed-form solution of Eq. (60) is

\[
y_i^{(g_0+g)} = y_i^{(g_0)} \left(1 - \frac{4c_{\mu/\mu,\lambda} \sigma^2 a_i}{\sigma_e} \right)^g
\]

\[
+ \frac{\sigma_e}{4\mu c_{\mu/\mu,\lambda} a_i} \left[1 - \left(1 - \frac{4c_{\mu/\mu,\lambda} \sigma^2 a_i}{\sigma_e} \right)^g\right].
\]

(63)

At this point the abbreviation \(\Theta := \sigma_e/(4\mu c_{\mu/\mu,\lambda})\) is introduced to reduce the length of the following equations. For a sufficiently large parental population size \(\mu\) the mutation strength \(\sigma\) in Eq. (63) can be approximately replaced with its steady state value (43), cf. Sec. 4.2. Consequently, the \(u\)-term in (61) asymptotically yields

\[
u = \frac{4c_{\mu/\mu,\lambda} \sigma^2 a_i}{\sigma_e} \approx \frac{\sigma_e}{4\mu c_{\mu/\mu,\lambda} \Sigma a} \frac{4c_{\mu/\mu,\lambda} a_i}{\sigma_e} = \frac{a_i}{\Sigma a},
\]

(64)

and thus the solution (63) of the linear difference equation becomes

\[
y_i^{(g_0+g)} = \left(y_i^{(g_0)} - \Theta \frac{\Sigma a}{a_i}\right) \left(1 - \frac{a_i}{\Sigma a}\right)^g + \Theta \frac{a_i}{a_i}.
\]

(65)

Equation (65) characterizes the evolution of the squared components of the parental parameter vector \(y^{(g_0+g)}\) within a CMSA-ES that is considered close
to its steady state on the noisy ellipsoid model. Note that Eq. (65) disregards an eventual transient phase of the respective dynamics. Eq. (65) must not be regarded as a complete description of the dynamical CMSA-ES behavior. Nevertheless, the prediction quality of (65) does provide sufficiently good agreement if $\sigma$ has reached the steady state. The difference equation is particularly useful when it comes to describing the qualitative behavior of CMSA-ES variants that use population control to achieve continuous progress towards the optimum. Regarding the pcCMSA-ES, Eq. (65) suffices to predict the generational progress within an isolation period of the pcCMSA-ES starting from $y_i^{(g_0)}$.

Considering the noisy ellipsoid model (7) and taking into account (54), Eq. (65) allows for the derivation of the corresponding expected fitness value evolution

$$F^{(g_0+g)} = \sum_{i=1}^{N} \left( a_i y_i^{(g_0)} + \Theta \right) \left( 1 - \frac{a_i}{\Sigma a} \right)^g + F_{ss}. \quad (66)$$

Following this representation, the rate at which the fitness dynamics decline depends on the eigenvalues $a_i$ of $A$. Taking into account the component $i$ with the slowest change, allows for the estimation of an upper bound for the fitness value evolution. The slowest rate of change is provided by the smallest quotient $a_i/\Sigma a$ in (66). Hence, it is governed by

$$\tilde{a} := \min_{i=1,\ldots,N} a_i. \quad (67)$$

Using (67), the fitness dynamics (66) can be further estimated by

$$F^{(g_0+g)} \leq \left( F^{(g_0)} - F_{ss} \right) \left( 1 - \frac{\tilde{a}}{\Sigma a} \right)^g + F_{ss}. \quad (68)$$

The fitness dynamics of the pcCMSA-ES over multiple isolation periods of constant length $L = \Sigma b$ are illustrated in Fig. 7. Making use of the observed population sizes $\mu$ adapted in the real pcCMSA-ES run, the $F_{ss}$ term in (68) can be updated accordingly. This way, the fitness estimate (68) (solid green line) can be compared to the experimental fitness dynamics (solid blue line) over consecutive isolation periods. While the experimental measurements oscillate closely around the predicted steady state fitness values $F_{ss}$ (54), Eq. (68) gives a good prediction of the long-term dynamics. After a sufficient number of generations, both curves exhibit a similar rate of long-term decline. That is, the theoretical estimate (68) is a suitable upper bound for the long-term behavior of the expected fitness dynamics.

According to Eq. (68), the fitness values in expectation drop at least with the factor $(1 - \tilde{a}/\Sigma a)^g$. Hence, the number of generations $G$ required to obtain a fitness reduction by a factor of $\delta < 1$, can be computed by solving the equation

$$\delta \leq \left( 1 - \frac{\tilde{a}}{\Sigma a} \right)^G \quad (69)$$
Figure 7: Illustration of the fitness dynamics plotted against the number of generations of the pcCMSA-ES. The solid blue lines depict the experimental measurements, while the steady state fitness (54) is represented by the solid red step function. Each step corresponds to a single isolation period of constant \( \mu \). The solid green curve displays the upper bound expected fitness dynamics (68). In order to appropriately take into account the dynamics over consecutive isolation periods, the steady state fitness value \( F_{ss} \) used in (68) is updated with respect to (54) after each \( \mu \) adjustment.

for \( G \). Applying the logarithm on both sides results in

\[
\ln \delta = G \cdot \ln \left( 1 - \frac{\bar{a}}{\Sigma a} \right). \tag{70}
\]

Considering the power series of the logarithm, the right-hand side (rhs) of Eq. (70) can roughly be estimated by

\[
\ln(1 - \frac{\bar{a}}{\Sigma a}) = \sum_{k=1}^{\infty} -\frac{1}{k} \left( \frac{\bar{a}}{\Sigma a} \right)^k \leq -\frac{\bar{a}}{\Sigma a}, \tag{71}
\]

with \(|\bar{a}/\Sigma a| < 1\). Using (71), \( G \) is estimated on the basis of (70) as

\[
G \leq \frac{\ln(\delta^{-1})}{\bar{a}} \Sigma a. \tag{72}
\]

Equation (72) provides an upper bound for the running time \( G \), i.e. the required length of the isolation period, to reach a desired fitness reduction by a factor of \( \delta \in (0, 1) \). Notice, that regarding the pcCMSA-ES in Alg. 2, the application of \( L \geq G \) suffices to realize the desired dynamical behavior.

6. The expected long-term dynamics of the pcCMSA-ES

The steady state quantities derived in Sec. 5 for the CMSA-ES on the noisy ellipsoid model (7) have been obtained under the assumption of a sufficiently large and constant parental population size and a constant mutation strength
close to its steady state value. This section is devoted to population size control. Adapting the population size properly enables the CMSA-ES to steadily reduce the residual distances to the optimum. That is, each increase of the population size comes with a reduction of the residual distance. In between two consecutive population adjustments, i.e. during an isolation period, such CMSA-ES variants exhibit the behavior derived in Sec. 5.1. Hence, the prior analysis steps are applicable to CMSA-ES with a population control mechanism that is able to appropriately elevate the population size if strong noise disturbs the evolutionary process. A prototype of a population control CMSA-ES is the pcCMSA-ES, cf. Alg. 2.

Accordingly, the analysis in Sec. 5 and Sec. 4 already includes the theoretical description of the pcCMSA-ES dynamics within a single isolation period. The description of the pcCMSA-ES long-term dynamics over multiple isolation periods and the theoretical derivation of its empirically observed convergence rate is tackled in this section.

6.1. Long-term fitness dynamics

The theoretical analysis is based on the following assumptions: The length of the pcCMSA-ES isolation periods is assumed to comply with the running time bound Eq. (72) which ensures a reduction of the fitness dynamics by a factor of at least $\delta$. During the isolation period the algorithm operates with fixed population sizes $\mu$, $\lambda$ and truncation ratio $\nu = \mu / \lambda$. Furthermore, it is assumed that the strategy continuously increases $\mu$ after each isolation period $t$. This assumption corresponds to the empirically observed behavior of the pcCMSA-ES on the noisy ellipsoid model with constant additive fitness noise. Notice, occasionally wrong test decisions are disregarded within the theoretical analysis as they exhibit minor influence on the overall algorithm dynamics. The population size increase follows the rule (cf. line 23 in Alg. 2)

$$\mu(t+1) = \alpha \mu(t) = \alpha^t \mu_0, \quad \text{with} \quad \mu(0) = \mu_0, \quad \alpha > 1.$$  

(73)

Taking into account sufficiently large parental population sizes $\mu$, the noise-to-signal ratio is supposed to increase, see Sec. 5. Accordingly, one can assume that the strategy is operating with a (near) constant mutation strength $\sigma$ close to the steady state derived in Eq. (43). Further, a constant (co-)variance of the mutation step is assumed: $\bar{y}_t = \bar{y} + \sigma \mathcal{N}(0, C)$. Concerning the long-term behavior of the pcCMSA-ES in the context of (7) subject to additive noise of constant variance $\sigma_{\epsilon}$, a constant matrix $C$ is reasonable since the strategy turns off the covariance matrix adaptation after the first noise detection (cf. line 24 in Alg. 2).

Being based on these assumptions, the analysis starts from the description of the inner fitness dynamics (68). Letting $F(t)$ denote the expected fitness value at the end of the $t$th isolation period, and taking into account (69), one obtains

$$F(t) \leq (F(t-1) - F_{ss}(t)) \delta + F_{ss}(t).$$  

(74)
The respective steady state fitness is provided by (54) and consideration of (73) yields
\[ F_{ss}(t) = \frac{\sigma_e N}{4c_{\mu/\lambda,t}\mu(t)} = \frac{\sigma_e N}{4c_{\mu/\lambda,0}\mu_0 t}. \] (75)

Simple rearrangements transform Eq. (74) into
\[ F(t) \leq F(t-1)\delta + F_{ss}(t)(1 - \delta). \] (76)

Introducing the abbreviation
\[ \omega := \sigma_e N(1 - \delta)/(4c_{\mu/\lambda,0}), \] (77)

Eq. (76) is reformulated as
\[ F(t) \leq F(t-1)\delta + \omega\alpha^{-t}. \] (78)

Regarding the iterative relation between two consecutive fitness values (78), the fitness at the end of isolation period \( t \) can be successively affiliated with the initial fitness. Following the recursion
\[ F(t) \leq F(t-1)\delta + \omega\alpha^{-t} \leq F(t-2)\delta^2 + \omega\alpha^{-(t-1)}\delta + \omega\alpha^{-t} \leq \ldots, \] (79)

yields an estimate of the fitness after the \( t \)th isolation period
\[
F(t) \leq F(0)\delta^t + \omega \left( \alpha^{-1}\delta^{t-1} + \alpha^{-2}\delta^{t-2} + \cdots + \alpha^{-t} \right) \\
\leq F(0)\delta^t + \frac{\omega\delta^{t-1}}{\alpha} \left( (\alpha\delta)^{-1} + \cdots + (\alpha\delta)^{-(t-1)} \right). \] (80)

The bracketed term on the rhs of Eq. (80) is the \( (t - 1) \)th partial sum of a geometric series. Hence, one obtains
\[
\left( (\alpha\delta)^{-1} + \cdots + (\alpha\delta)^{-(t-1)} \right) = \left( \frac{1}{\alpha\delta} \right)^{t-1} - \frac{1}{\alpha\delta - 1}. \] (81)

Replacing the respective sum in (80) with (81) yields
\[ F(t) \leq F(0)\delta^t + \frac{\omega\delta^{t-1}}{\alpha} \left( \frac{1}{\alpha\delta} \right)^{t-1} - \frac{1}{\alpha\delta - 1}. \] (82)

Finally, rearranging the terms leads to
\[
F(t) \leq F(0)\delta^t + \frac{\omega}{1 - \alpha\delta} \delta^t \left[ \left( \frac{1}{\alpha\delta} \right)^t - 1 \right]. \] (83)

While the parameters \( \alpha > 1 \) and \( \delta < 1 \) can be set independently, the magnitude of the product \( \alpha\delta = \kappa \) governs the further analysis. Therefore, the two cases \( \kappa < 1 \) and \( \kappa > 1 \), respectively, have to be distinguished.
The $\kappa < 1$ case. Considering $t \to \infty$ drives $1/(\alpha \delta)^t$ towards infinity. That is, for sufficiently large $t$ the term considerably exceeds 1, and Eq. (83) transforms into

$$ F(t) \leq F(0) \delta^t + \frac{1}{(\alpha \delta)^t}. \quad (84) $$

Taking a closer look at the rhs of Eq. (84), one observes that

$$ F(0) \delta^t + \frac{\omega}{1 - \alpha \delta} \delta^t \frac{1}{(\alpha \delta)^t} = \left( F(0) \kappa^t + \frac{\omega}{1 - \alpha \delta} \right) \frac{1}{\alpha^t}. \quad (85) $$

After factoring out $1/\alpha^t$ on the rhs of (85), considering that $\kappa^t \to 0$ for $t \to \infty$ yields

$$ F(0) \delta^t + \frac{\omega}{1 - \alpha \delta} \delta^t \frac{1}{(\alpha \delta)^t} \simeq \frac{\omega}{1 - \alpha \delta \alpha^t}. \quad (86) $$

Applying the logarithm on both sides of (84) after insertion of (86) results in

$$ \ln F(t) \lesssim \ln \left( \frac{\omega}{1 - \kappa} - t \ln \alpha \right). \quad (87) $$

A comparison of the theoretically obtained fitness estimates is displayed in Fig. 8. There, it is observed that the analysis is able to predict the experimentally obtained rate of decline. According to assumption (73), the predictions are only displayed for those isolation periods where a continuous increase of the population size is realized. The respective isolation period after which a steady population growth is observed is marked by the vertical dashed line. The transient time needed to establish the desired descent is much longer on ill-conditioned ellipsoid model $b_i = i^2$. However, after a sufficient number of isolation periods the downward slopes of the experimental measurements and those of the predictions show a good agreement. Both theoretical results in Fig. 8 almost overlap as the solid green curve from (83) approaches the solid red line representing Eq. (87) with growing $t$. Hence, both representations provide a suitable description of the actual pcCMSA-ES fitness decline.

The $\kappa > 1$ case. In this situation, the term $1/(\alpha \delta)^t$ approaches zero for $t \to \infty$. Consequently, Eq. (83) becomes

$$ F(t) \leq \left[ F(0) + \frac{\omega}{\alpha \delta - 1} \right] \delta^t. \quad (88) $$

Applying the logarithm and reconsidering $\kappa = \delta \alpha$ results in

$$ \ln F(t) \leq \ln \left( F(0) + \frac{\omega}{\kappa - 1} \right) - t (\ln \alpha - \ln \kappa). \quad (89) $$

32
6.2. Number of function evaluations

Based on the asymptotic fitness dynamics, the next step is concerned with the description of the number of fitness function evaluations. The number of function evaluations after the $t$th period is denoted by $n(t)$. Considering the truncation ratio $\nu = \mu/\lambda$ and referring to the number of generations within a single period as $G$, the number of function evaluations within an isolation period is $G \cdot \lambda(t) + 1 = G \cdot \mu(t)/\nu + 1$. This includes the additional fitness evaluation of the parental centroid, cf. line 18 of Alg. 2. Aggregating these values over all preceding periods, the cumulative number of function evaluations is

$$n(t) = \sum_{k=0}^{t-1} \mu(k) \frac{G}{\nu} + t + \frac{G}{\nu} \mu_0 \sum_{k=0}^{t-1} \alpha^k = t + \frac{G}{\nu} \mu_0 \frac{\alpha^t - 1}{\alpha - 1}. \quad (90)$$

Requiring a sufficiently large number of isolation periods, the linear $t$ can be neglected compared to $\alpha^t$. Hence, applying the logarithm to Eq. (90) asymptotically yields

$$\ln n(t) \simeq \ln \frac{G\mu_0}{\nu(\alpha - 1)} + \ln(\alpha^t - 1). \quad (91)$$

Further, for large $t$ the first term of the sum on the rhs of (91) can be neglected when compared to the second term. By taking into account $\alpha^t - 1 \simeq \alpha^t$ (for large $t$), in the asymptotic limit the logarithm of the number of function evaluations linearly increases with the number of isolation periods $t$

$$\ln n(t) \simeq t \cdot \ln \alpha. \quad (92)$$

The comparison of the obtained predictions with experimental measurements from the pcCMSA-ES runs on the noisy ellipsoid model is provided in Fig. 9. By considering different intercepts for the asymptotical obtained slope (92), it
is visible that the experimental measurements approach the predicted behavior with time. After a transient phase, the experimental dynamics qualitatively match the behavior predicted by Eq. (91) as well as by Eq. (92). That is, both curves exhibit the same asymptotical growth. The delay of the theoretical and experimental results is due to the negligence of the transient phase in Eq. (92).

6.3. The asymptotic convergence rate of the pcCMSA-ES

The calculation of the asymptotically exact convergence rate of the pcCMSA-ES is presented in this section. It is based on the assumptions in Sec. 6 as well as on the asymptotic $F(t)$ and $n(t)$ dynamics in Eqs. (87), (89), and (92), respectively. Defining the asymptotic convergence rate $CR$ by the quotient of the expected logarithmic fitness and the logarithmic number of function evaluations in the limit $t \to \infty$, one obtains

$$CR := \lim_{t \to \infty} \frac{\ln F(t)}{\ln n(t)}.$$  (93)

As the expected fitness dynamics depend on the magnitude of the parameter $\kappa$, the two cases described in Sec. 6.1 must be considered.

**The $\kappa < 1$ case.** Inserting the asymptotical dynamics (87) and Eq. (92) into the convergence rate definition (93), one obtains

$$CR = \lim_{t \to \infty} \frac{\ln F(t)}{\ln n(t)} = \lim_{t \to \infty} \frac{\ln \left( \frac{\omega}{1 - \alpha \delta} - t \ln \alpha \right)}{t \ln \alpha} = -1.$$  (94)
Hence, a convergence rate of \(-1\) can be obtained under the condition that \(\kappa = \alpha \delta \leq 1\). Reconsidering \(\delta\) from (69), the condition \(\alpha \delta \leq 1\) allows for the computation of a bound on the isolation time \(G\) which ensures a maximal asymptotic convergence rate of \(CR = -1\). Applying the logarithm, using (71), and solving the resulting equation for \(G\) yields

\[
G \geq -\frac{\ln \alpha}{\ln(1 - \hat{a}/\Sigma a)}. \tag{95}
\]

Consequently, choosing the isolation time \(G = \ln \alpha \Sigma a/\hat{a}\) satisfies the condition in Eq. (95), and ensures that the strategy approaches a convergence rate of \(CR \simeq -1\). The sufficiency of this particular choice of \(G\) can be substantiated by verifying whether the assumption \(\kappa = \alpha \delta \leq 1\) holds. To this end, the representation of \(\delta\) in Eq. (69) is considered. By application of the logarithm, one obtains

\[
\ln(\kappa) = \ln(\alpha \delta) = \ln(\alpha) + G \ln \left(1 - \frac{\hat{a}}{\Sigma a}\right). \tag{96}
\]

Making use of the estimate presented in Eq. (71) yields

\[
\ln(\kappa) \leq \ln(\alpha) - G \frac{\hat{a}}{\Sigma a}, \tag{97}
\]

and by insertion of the suggested isolation length of \(G = \ln \alpha \Sigma a/\hat{a}\), Eq. (97) becomes

\[
\ln(\kappa) \leq \ln(\alpha) - \ln \alpha \frac{\Sigma a}{\hat{a}} = 0. \tag{98}
\]

Hence, the initial assumption holds for \(G = \ln \alpha \Sigma a/\hat{a}\). Note that the marginal case of \(\kappa = 1\) is separately investigated at the end of this section.

Equation (95) indicates that the use of population control parameters \(1 < \alpha < e\) reduces the upper bound on the isolation time. However, choosing \(\alpha\) too close to 1 lowers the reliability of the noise hypothesis test within the pcCMSA-ES such that using \(\alpha \geq 2\) is recommended. Notice, that a convergence rate realization of \(CR = -1\) represents the theoretically derived lower bound for all comparison based search algorithms [20]. On the noisy ellipsoid model this descent was empirically discovered in runs of the pcCMSA-ES [17]. The corresponding illustration is provided in Fig. 10. The solid blue lines refer to the experimental fitness measurements \(F(t)\). The experimentally obtained cumulative number of function evaluations \(n(t)\) is displayed by the solid green line. Using initial values \(10^i, i \in \{-2, \ldots, 4\}\), the dashed magenta lines represent the reciprocal of the asymptotically derived \(n(t)\) law in Eq. (92). In the limit of large \(t\), both experimental dynamics approach the predicted decline. That is, the fitness dynamics decrease proportional to \(1/n(t)\) indicating a convergence rate of \(CR \simeq -1\), cf. Eq. (94).
The $\kappa > 1$ case. Considering the results from Eq. (89) and Eq. (92), the convergence rate (93) is obtained as

$$CR = \lim_{t \to \infty} \frac{\ln \left( F(0) + \frac{\omega}{\alpha \delta - 1} \right) + t(\ln \kappa - \ln \alpha)}{t \ln(\alpha)}$$  \hfill (99)

Accordingly, in the limit of a large number of isolation periods $t$ the convergence rate $CR$ becomes

$$CR = \frac{(\ln \kappa - \ln \alpha)}{\ln(\alpha)} = \frac{\ln \kappa}{\ln \alpha} - 1 > -1.$$  \hfill (100)

Looking at Eq. (100), it is possible to tune the convergence rate towards $-1$. Making use of $\kappa = \alpha \delta$, one obtains

$$CR = \frac{\ln(\alpha \delta)}{\ln(\alpha)} - 1 = \frac{\ln \alpha + \ln \delta}{\ln \alpha} - 1 = \frac{\ln \delta}{\ln \alpha}.$$  \hfill (101)

Reconsidering the $\delta$ definition (69), Eq. (101) transforms into

$$CR \approx G \cdot \frac{\ln(1 - \frac{\hat{a}}{\Sigma a})}{\ln(1 - \frac{\hat{a}}{\Sigma a})}. $$  \hfill (102)

Resolving Eq. (102) for $G$, results in

$$G \approx CR \frac{\ln \alpha}{\ln(1 - \frac{\hat{a}}{\Sigma a})} \approx -CR \frac{\ln \alpha}{\hat{a} \Sigma a}. $$  \hfill (103)

Choosing $G$ in accordance with (103), the convergence rate of the pcCMSA-ES can be driven towards a desired (but suboptimal) value $CR > -1$. 

Figure 10: Relation of the fitness dynamics and the reciprocal of the number of function evaluations. All dynamics are plotted against the number of isolation periods of $L = 465$ on the ellipsoid model $b_i = i$ generations each, and $L = 9455$ on the ellipsoid model $b_i = i^2$, respectively.
The $\kappa = 1$ case. Notice that, the marginal case $\kappa = 1$ also yields a convergence rate $CR = -1$. Beginning with Eq. (80), and considering $\kappa := \alpha \delta = 1$, the bracketed term simply sums up to

$$\left(1 + (\alpha \delta)^{-1} + (\alpha \delta)^{-2} + \cdots + (\alpha \delta)^{-(t-1)}\right) = t$$

Consequently, the estimate of the fitness dynamics in Eq. (82) becomes

$$F(t) \leq F(0)\delta^t + \frac{t \omega \delta^t}{\alpha \delta} = (F(0) + t \omega) \delta^t.$$  

(105)

Applying the logarithm and considering the relation $\delta^t = (\kappa/\alpha)^t$ yields

$$\ln F(t) \leq \ln (F(0) + t \omega) - t (\ln \alpha - \ln \kappa) = \ln (F(0) + t \omega) - t \ln \alpha.$$  

(106)

Regarding the convergence rate definition (93), and taking into account Eq. (92), one obtains the desired convergence rate

$$CR = \lim_{t \to \infty} \frac{\ln (F(0) + t \omega) - t \ln \alpha}{t \ln \alpha} = -1.$$  

(107)

7. Conclusions

This paper carried out an asymptotical analysis of the CMSA-ES on the noisy ellipsoid model subject to additive Gaussian noise. To this end, the concept of covariance matrix adaptation had to be covered by the analysis approach. It turned out that the progress rate theory from [7] can be transferred to covariance-adaptive ES on quadratic functions with positive definite Hessian. On this basis, an expression for the expected steady state mutation strength of the CMSA-ES was determined in two ways. The first result avoids limitations on the considered ellipsoid models. This can be regarded a generalization over the second $\sigma_{ss}$ derivation that was already sketched in [18]. On the other hand, the second steady state mutation strength representation appeared more convenient for the successive analysis steps.

Making use of the mutation strength result obtained in Sec. 4, a steady state description of the CMSA-ES parameter vector components was provided assuming a fixed population size. The asymptotically correct solutions to the system (21), (22) are verified by comparison to experimental (pc)CMSA-ES runs. Based on these solutions, the steady state expression of the fitness dynamics, that was priorly derived in [12], was obtained in a different manner. Furthermore, the residual steady state distance approached by the CMSA-ES was computed and the normalized mutation strength $\sigma^*$, as well as the noise-to-signal ratio $\vartheta$, were found to be proportional to the square root of the parental population size $\sqrt{\mu}$. In retrospect, this substantiates the assumption of a sufficiently large $\vartheta$ value needed for the steady state derivation when taking into

---

6Notice that the quotient $\ln(F(0) + t \omega)/(t \ln \alpha) \to 0$ with increasing $t$. 

37
account large population sizes. This asymptotic characterization of the CMSA-ES dynamics was obtained while keeping the population size of the CMSA-ES constant. Regarding the concept of population size control, the preliminary findings also apply to the dynamical behavior within a single isolation period of the pcCMSA-ES.

Continuing this way, the paper presented the detailed theoretical analysis of the asymptotical pcCMSA-ES long-term dynamics on the ellipsoid model subject to additive Gaussian noise. The empirically observed \( CR = -1 \) convergence rate has been derived. To reach this theoretically proven best convergence rate, the ES must not behave like a “simple ES” as shown in [5]. That is, the ES must not be scale-invariant w.r.t. the mutation strength when approaching the optimum subject to strong noise. According to Sec. 4, the mutation strength \( \sigma \) of self-adaptive ES reaches a steady state value that does not change with the parental population size \( \mu \) and the distance from the optimum. However, “violating” the scale-invariance is not sufficient to reach \( CR = -1 \). As shown in this paper, the population increasing factor \( \alpha \) and the isolation time \( L \geq G \) have influence on the \( F \)-dynamics. The isolation time must be chosen not too small to assure a sufficient approach to the steady state \( F (\mu = \text{const.}) \) that also depends on the choice of \( \alpha \). \( G \) itself depends on the eigenvalue spectrum of the matrix \( A \) – a combination of the Hessian of the ellipsoid model and the actual transformation matrix \( \sqrt{C} \). Since the covariance matrix \( C \) evolves during the evolution, the eigenvalue spectrum evolves as well. This does not present a problem in the noiseless case. Actually, transforming the general quadratic model successively into a sphere model (approximatively), it presents a desired behavior. However, in the strong noise case \( C \) accumulates random noise that results in a fast eigenvalue blowup and the condition number as well as the trace \( \Sigma a \) of the whole system increase over time. Consequently, under strong noise the \( C \)-update reveals counterproductive effects and \( G \) cannot be fixed any longer. As a consequence, \( CR = -1 \) cannot be reached. Therefore, it is algorithmically important (and implemented in the pcCMSA-ES, Alg. 2) to stall the \( C \)-update if strong noise is detected.

Acknowledgements

This work was supported by the Austrian Science Fund FWF under grant P29651-N32.

References


