Analysis of the pcCMSA-ES on the noisy ellipsoid model

Hans-Georg Beyer
Research Centre PPE
Vorarlberg University of Applied Sciences
Hochschulstr. 1, 6850 Dornbirn, Austria
hans-georg.beyer@fhv.at

Michael Hellwig
Research Centre PPE
Vorarlberg University of Applied Sciences
Hochschulstr. 1, 6850 Dornbirn, Austria
michael.hellwig@fhv.at

ABSTRACT

Regarding the noisy ellipsoid model with additive Gaussian noise, the population control covariance matrix self-adaptation Evolution Strategy (pcCMSA-ES) by Hellwig and Beyer was empirically observed to exhibit a convergence rate (CR) close to the theoretical lower bound of -1 for all comparison-based direct search algorithms. The present paper provides the corresponding theoretical analysis of the pcCMSA-ES long-term behavior. To this end, the analysis from the context of isotropic mutations is transferred to the pcCMSA-ES that uses covariance matrix adaptation until significant noise influence is detected. The results allow for the computation of an upper bound on the number of generations between two consecutive test decisions of the pcCMSA-ES that ensures the observed performance. Further, the empirically observed convergence rate of $CR \simeq -1$ is theoretically derived.

CCS CONCEPTS

•Computing methodologies → Computational control theory; •Theory of computation → *Design and analysis of algorithms; Stochastic control and optimization;*

KEYWORDS

Evolution strategies, self-adaptation, population size control, dynamical systems approach, noisy ellipsoid model

ACM Reference format:
Hans-Georg Beyer and Michael Hellwig. 2017. Analysis of the pcCMSA-ES on the noisy ellipsoid model. In *Proceedings of GECCO '17, Berlin, Germany,*

1 INTRODUCTION

The ability of Evolutionary Algorithms (EAs) to successfully deal with noisy optimization problems has already been reported by, among others, [1, 9]. In this context, the work of [8] introduced an innovative noise handling approach for Evolution Strategies (ES). As discussed in [7], self-adaptation itself provides a favorable bias on the mutation strength adaptation process which helps to prevent premature convergence of the ES. Hence, the approach was based on the well-known covariance matrix self-adaptation Evolution Strategy (CMSA-ES) [6]. The CMSA-ES uses mutation strength self-adaptation and covariance matrix adaptation in order to evolve the population of candidate solutions towards the optimizer. Knowing that increasing the population size of the ES improves its ability to approach the optimizer in noisy environments [4], the CMSA-ES has been equipped with suitable mechanisms for noise identification (e.g. linear regression analysis of the observed fitness dynamics) and population size control. The approach in [8] is called *population control* covariance matrix self-adaptation Evolution Strategy (pcCMSA-ES). The corresponding pseudo code is presented in Alg. 1.

Building an integral part of Alg. 1, the lines 5 to 15, and line 30, represent the standard $(\mu/\mu,\lambda)$-CMSA-ES. In each generation λ offspring with individual mutation strengths σ_l are generated. The mutation strength σ_l is self-adaptively evolved using the learning parameter $\tau = 1/\sqrt{2N}$ (N – search space dimension). The mutation vector $\boldsymbol{z}_l$ of each offspring depends on the covariance matrix $\boldsymbol{C}$ which corresponds to the distribution of previously generated successful candidate solutions. The update rule can be found in line 30 where $\tau_C = 1 + N(N+1)/2\mu$ is used. After offspring creation, the noisy objective function (fitness) is evaluated. Based on the fitness values, the algorithm selects those μ of the λ offspring with the best fitness values $\tilde{F}_{m;\lambda}, m = 1,\ldots,\lambda$. Notice, $m;\lambda$ denotes the mth best out of λ individuals. Intermediate recombination is used to compute the centroids $\langle \boldsymbol{z} \rangle$ and $\langle \sigma \rangle$ of the μ best individuals, e.g. $\langle \sigma \rangle = \sum_{m=1}^{\mu} \sigma_{m;\lambda}/\mu$ is the mutation strength centroid.

The pcCMSA-ES specific noise identification and population size control are performed within lines 16 to 29. Line 4 ensures a constant truncation ratio $\nu = \mu/\lambda$. The algorithm gathers a list $\mathcal{F}$ of the noisy fitness values of the parental centroid. After a lead time *wait*, the last L entries of $\mathcal{F}$ are examined for noise influences on the strategy's convergence behavior using a linear regression approach. To this end, the piecewise trend of the sequence of L observed fitness values is approximated by a linear least squares regression model. The slope of the regression line governs the population control. That is, a negative slope is supposed to indicate ongoing positive progress towards the optimizer. Whereas 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, the real slope has to be bracketed in a confidence interval to ensure a correct decision. 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 $\tilde{\alpha}$ 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 μ 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 *adjC* to zero in line 22. 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 μ by the factor β. 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 *wait* will also be referred to as 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. A more detailed explanation of the pcCMSA-ES is provided in [8].

This paper addresses the theoretical analysis of the pcCMSA-ES, Alg. 1, for minimizing the *noisy ellipsoid model*

$$\tilde{F}(\boldsymbol{y}) = F(\boldsymbol{y}) + \delta = \sum_{i=1}^{N} b_i y_i^2 + \delta. \tag{1}$$

In (1) the parameters $b_i > 0$ are denoted as ellipsoid coefficients, and δ represents an unbiased noise term. While the pcCMSA-ES is applicable to different kinds of fitness noise, the analysis at hand concentrates on the case of a normally distributed noise term of constant noise strength σ_ϵ, i.e. $\delta \sim \sigma_\epsilon \mathcal{N}(0,1)$. Considering that respective noisy fitness environment, the empirical investigations in [8] revealed that the pcCMSA-ES is able to perform close to the theoretical lower bound convergence rate of all comparison-based direct search algorithms [11]. Our analysis intends to describe the pcCMSA-ES steady state behavior and to obtain the theoretical confirmation of the observed convergence rate.

Aiming at a description of the dynamical behavior, the analysis uses the approach presented in [5] in the context of isotropic mutations. In that situation the offspring creation reads

$$\tilde{\boldsymbol{y}}_l = \boldsymbol{y} + \boldsymbol{x}_l \quad \text{with} \quad \boldsymbol{x}_l \sim \sigma_l \mathcal{N}(0, \boldsymbol{I}). \tag{2}$$

Because the pcCMSA-ES adapts the problem specific covariance matrix $\boldsymbol{C}$, the mutation steps can no longer be regarded as isotropic. To apply the analysis approach [5] to the pcCMSA-ES, the offspring generation step in line 7 to 9 of Alg. 1 has to be considered a two step procedure. That is, a mutation vector $\boldsymbol{x}_l$ is drawn from an isotropic normal distribution $\sigma_l \mathcal{N}(0, \boldsymbol{I})$, and is then transformed into $\boldsymbol{z}_l = \boldsymbol{M}\boldsymbol{x}_l$ by multiplication with $\boldsymbol{M} = \sqrt{\boldsymbol{C}}$. Consequently, the fitness evaluation step reads

$$F_l = F(\boldsymbol{y} + \boldsymbol{M}\boldsymbol{x}_l) \quad \text{with} \quad \boldsymbol{x}_l \sim \sigma_l \mathcal{N}(0, \boldsymbol{I}). \tag{3}$$

As the noise-free ellipsoid model in Eq. (1) represents a positive definite quadratic form

$$F(\boldsymbol{y}) = \boldsymbol{y}^\top \boldsymbol{B}\boldsymbol{y} \quad \text{with} \quad \boldsymbol{B} = \operatorname{diag}(b_1, \ldots, b_N), \tag{4}$$

by taking into account (4), Eq. (3) becomes

$$\begin{aligned} F_l = F(\boldsymbol{y} + \boldsymbol{M}\boldsymbol{x}_l) &= (\boldsymbol{y} + \boldsymbol{M}\boldsymbol{x}_l)^\top \boldsymbol{B}(\boldsymbol{y} + \boldsymbol{M}\boldsymbol{x}_l) \\ &= \boldsymbol{y}^\top \boldsymbol{B}\boldsymbol{y} + \boldsymbol{d}^\top \boldsymbol{x}_l + \boldsymbol{x}_l^\top \boldsymbol{A}\boldsymbol{x}_l \end{aligned} \tag{5}$$

Algorithm 1 The pcCMSA Evolution Strategy. The linear regression analysis method for noise identification is implemented by use of the subroutine $\texttt{detection}(\boldsymbol{F}_{int}, \alpha)$.

1: **initialization:** $g \leftarrow 0$; $wait \leftarrow 0$; $adjC \leftarrow 1$; $\langle\sigma\rangle \leftarrow \sigma^{(init)}$;
2: $\boldsymbol{y} \leftarrow \boldsymbol{y}^{(init)}$; $\mu \leftarrow \mu^{(init)}$; $\mu_{\min} \leftarrow \mu^{(init)}$; $\boldsymbol{C} \leftarrow \boldsymbol{I}$; $\boldsymbol{\mathcal{F}} = \{\}$
3: **repeat**
4: $\quad \lambda \leftarrow \lfloor \mu/\nu \rfloor$
5: $\quad$ **for** $l \leftarrow 1$ to λ **do**
6: $\quad\quad \sigma_l \leftarrow \langle\sigma\rangle \mathrm{e}^{\tau \mathcal{N}(0,1)}$
7: $\quad\quad \boldsymbol{s}_l \leftarrow \sqrt{\boldsymbol{C}}\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
8: $\quad\quad \boldsymbol{z}_l \leftarrow \sigma_l \boldsymbol{s}_l$
9: $\quad\quad \tilde{\boldsymbol{y}}_l \leftarrow \boldsymbol{y} + \boldsymbol{z}_l$
10: $\quad\quad \tilde{F}_l \leftarrow \tilde{F}(\tilde{\boldsymbol{y}}_l)$
11: $\quad$ **end for**
12: $\quad g \leftarrow g + 1$
13: $\quad \langle z \rangle \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} z_{m;\lambda}$
14: $\quad \langle\sigma\rangle \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} \sigma_{m;\lambda}$
15: $\quad \boldsymbol{y} \leftarrow \boldsymbol{y} + \langle z \rangle$
16: $\quad$ **add** $\tilde{F}(\boldsymbol{y})$ **to** $\boldsymbol{\mathcal{F}}$
17: $\quad$ **if** $g > L \ \wedge \ wait = 0$ **then**
18: $\quad\quad \boldsymbol{\mathcal{F}}_{int} \leftarrow \boldsymbol{\mathcal{F}}(g - L : g)$
19: $\quad\quad td \leftarrow \texttt{detection}(\boldsymbol{\mathcal{F}}_{int}, \tilde{\alpha})$
20: $\quad\quad$ **if** $td = 0$ **then**
21: $\quad\quad\quad \mu \leftarrow \lfloor \mu\alpha \rfloor$
22: $\quad\quad\quad adjC \leftarrow 0$
23: $\quad\quad$ **else**
24: $\quad\quad\quad \mu \leftarrow \max(\mu_{\min}, \lfloor \mu/\beta \rfloor)$
25: $\quad\quad$ **end if**
26: $\quad\quad wait \leftarrow L$
27: $\quad$ **else**
28: $\quad\quad wait \leftarrow wait - 1$
29: $\quad$ **end if**
30: $\quad \boldsymbol{C} \leftarrow \left(1 - \frac{1}{\tau_c}\right)^{adjC} \boldsymbol{C} + \frac{adjC}{\tau_c} \langle \boldsymbol{s}\boldsymbol{s}^\top \rangle$
31: **until** termination condition
32: **return** $\boldsymbol{y}$

with $\boldsymbol{d}^\top = 2\boldsymbol{y}^\top \boldsymbol{B}\boldsymbol{M}$ and $\boldsymbol{A} = \boldsymbol{M}^\top \boldsymbol{B}\boldsymbol{M}$. As a product of a positive definite matrix $\boldsymbol{B}$ and two full-rank matrices $\boldsymbol{M}^\top$ and $\boldsymbol{M}$, the matrix $\boldsymbol{A}$ itself is a positive definite matrix, i.e. the eigenvalues a_i of $\boldsymbol{A}$ are all positive. Notice, that the matrix $\boldsymbol{C}$ and the parental centroid $\boldsymbol{y}$ can be regarded constant during a single offspring procreation process of the CMSA-ES. Thus the offspring fitness evaluation in Alg. 1 only differs from that in (2) by a linear transformation with respect to $\boldsymbol{M}$. Hence, the analysis approach [5] is applicable to "CMSA-like" Evolution Strategies alike the pcCMSA-ES. Instead of interpreting it as a change of the mutation distribution, the $\boldsymbol{C}$ adaptation is considered to transform the fitness environment perceived by an ES that uses isotropic mutations. This allows for the application of the analysis approach [5] especially in the steady state case of strong noise where the pcCMSA-ES turns off the $\boldsymbol{C}$ adaptation after the first noise identification (i.e. $\boldsymbol{A}$ does not change anymore).

All theoretical derivations presented in this paper are substantiated by comparison to real pcCMSA-ES runs. The experiments are conducted on the noisy ellipsoid model $b_i = i, \forall i = 1, \ldots, N$ in search space dimension $N = 30$ using a constant noise strength of

$\sigma_\epsilon = 1$. The runs are initialized with $\sigma^{(init)} = 1$, $\boldsymbol{y}^{(init)} = \mathbf{1}$, and population size $\mu^{(init)} = 3$ 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 $L = \Sigma a$[1] (which matches the theoretical results).

The paper proceeds as follows: In Sec.2 the steady state mutation strength σ_{ss} of the pcCMSA-ES is derived. Assuming that the pcCMSA-ES operates with constant $\sigma \approx \sigma_{ss}$, its dynamical behavior during a single isolation period is addressed in Sec. 3. The results are used to obtain a description of the strategy's long-term fitness dynamics over multiple isolation periods in Sec. 4. Considering the corresponding expected growth of the number of function evaluations in Sec. 5, allows for the derivation of the pcCMSA-ES convergence rate in Sec. 6. The paper concludes with a discussion of the results obtained.

2 STEADY STATE MUTATION STRENGTH

The work of [8] revealed that the pcCMSA-ES on the noisy ellipsoid model (1) does not behave like a "simple ES", [2]. That is, in the presence of strong noise the mutation strength is not scaling with the distance to the optimizer. Instead, it approaches a steady state σ_{ss} as the strategy approaches the optimizer. The derivation of σ_{ss} is the aim of this section.

The steady state of the mutation strength is defined as the expected mutation strength in the limit of a large number of generations, i.e. $g \to \infty$, $\sigma_{ss} := \lim_{g\to\infty} \mathrm{E}[\sigma^{(g)}]$. Equivalently to this formulation, one can demand

$$\mathrm{E}[\sigma^{(g+1)}] = \mathrm{E}[\sigma^{(g)}] = \sigma_{ss} \tag{6}$$

for sufficiently large g values. The mean value dynamics of the mutation strength in σSA-ES are governed by the so-called Self-Adaptation Response (SAR) function [10]

$$\psi(\sigma^{(g)}) := \mathrm{E}\left[\frac{\sigma^{(g+1)} - \sigma^{(g)}}{\sigma^{(g)}} \,\middle|\, \boldsymbol{y}^{(g)}, \sigma^{(g)}\right]. \tag{7}$$

Hence, the transition from generation g to $g+1$ is given by

$$\mathrm{E}[\sigma^{(g+1)}] = \sigma^{(g)}(1 + \psi(\sigma^{(g)})). \tag{8}$$

Considering the noisy ellipsoid model (1), the respective SAR function has been derived for the asymptotic limit of large search space dimensions ($N \to \infty$) in [10]. Assuming that the considered ellipsoid model satisfies the requirement

$$\frac{\sum_{i=1}^N a_i^2}{(\Sigma a)^2} \overset{N\to\infty}{\to} 0 \quad \text{and} \quad 1 + \vartheta^2 \gg \frac{\sigma^{*2}}{2}\frac{\sum_{i=1}^N a_i^2}{(\Sigma a)^2} \tag{9}$$

does hold, the asymptotic SAR function reads [10]

$$\psi(\sigma^*) \simeq \tau^2 \left[\frac{1}{2} - \frac{c_{\mu/\mu,\lambda}\sigma^*}{\sqrt{1+\vartheta^2}}\left(1 - \frac{e^{1,1}_{\mu,\lambda}}{c_{\mu/\mu,\lambda}}\frac{1}{\sigma^*\sqrt{1+\vartheta^2}}\right)\right]. \tag{10}$$

Equation (10) contains the normalized mutation strength σ^*

$$\sigma^* = \sigma \Sigma a \Big/ \sqrt{\textstyle\sum_{j=1}^N a_j^2 y_j^2}\,, \tag{11}$$

and the noise-to-signal ratio ϑ

$$\vartheta = \frac{\sigma^*_\epsilon}{\sigma^*} = \frac{\sigma_\epsilon}{2\sigma\sqrt{\sum_{j=1}^N a_j^2 y_j^2}}. \tag{12}$$

[1]The term Σa is used to denote the trace of the positive definite matrix $\boldsymbol{A}$.

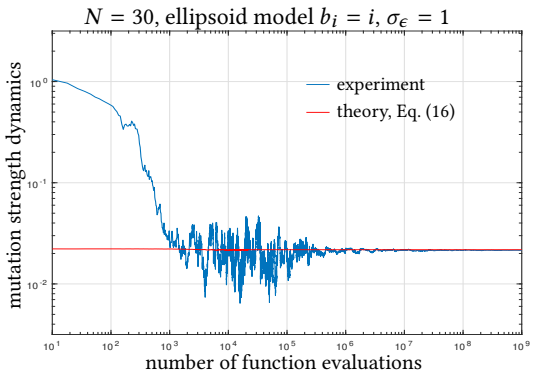

Figure 1: Comparison of the prediction (16) and experimental σ dynamics of the pcCMSA-ES plotted against the number of function evaluations. The experimental mutation strength measurements are displayed by the solid blue line. The red line illustrates the steady state values of (16).

Here, σ^*_ϵ denotes the normalized noise strength

$$\sigma^*_\epsilon = \sigma_\epsilon \frac{\Sigma a}{2\sum_{j=1}^N a_j^2 y_j^2}, \tag{13}$$

and y_i refers to the components of the parental centroid $\boldsymbol{y}$ at generation g. Notice, that Eqs. (10) to (13) omit the superscripts (g) for the sake of brevity. The term $c_{\mu/\mu,\lambda} := e^{1,0}_{\mu,\lambda}$ denotes the progress coefficient of the $(\mu/\mu_I,\lambda)$-ES. It is a special case of the *generalized progress coefficients* $e^{a,b}_{\mu,\lambda}$ introduced in [3]. Note that for a fixed truncation ratio $\nu = \mu/\lambda$, $c_{\mu/\mu,\lambda}$ converges to a constant value for $\lambda \to \infty$ (see [3], p. 249, Eq. 6.113). This also holds for $e^{1,1}_{\mu,\lambda}$.

Requiring $\vartheta \gg 1$, the second addend in (10) can be neglected and one obtains

$$\psi(\sigma^*) \simeq \tau^2\left(\frac{1}{2} - \frac{c_{\mu/\mu,\lambda}\sigma^*}{\sqrt{1+\vartheta^2}}\right). \tag{14}$$

Provided that a stable steady state mutation strength σ exists, the assumption $\vartheta \gg 1$ is validated by reconsidering the ϑ definition in (12). Due to the constant σ_ϵ, the noise-to-signal ratio ϑ grows as the strategy approaches the optimizer, i.e. as the $\sqrt{\sum_{j=1}^N a_j^2 y_j^2}$ term decreases. Actually, $\vartheta \to \infty$ holds as Alg. 1 continuously increases its population size μ in the case of additive Gaussian noise, Fig. 2.

Reconsidering (8) and the SAR function (14), the steady state condition (6) is equivalent to requiring $\psi(\sigma^*_{ss}) = 0$. That is, regarding (14) and neglecting the 1 compared to ϑ^2 allows for a straight forward computation of the (asymptotic) normalized steady state mutation strength

$$\sigma^*_{ss} \simeq \sqrt{\frac{\sigma^*_\epsilon}{2c_{\mu/\mu,\lambda}}}. \tag{15}$$

Renormalization with (11) and (13) yields the steady state σ

$$\boxed{\sigma_{ss} \simeq \sqrt{\frac{\sigma_\epsilon}{4c_{\mu/\mu,\lambda}\Sigma a}}.} \tag{16}$$

An illustration of the prediction quality of Eq. (16) on the noisy ellipsoid model $b_i = i$ in search space dimension $N = 30$ is presented in Fig. 1. After a transient phase, the experimental results begin to oscillate around the predicted steady state. The fluctuations decrease with growing population size.

Remembering that (16) relies on the conditions in (9), the resulting σ_{ss} prediction is not exact for all ellipsoid models. Since the first condition in (9) excludes ellipsoid models with only few dominant coefficients, the obtained results can deviate in those cases.

3 THE EXPECTED INNER PROGRESS

The next step is concerned with the description of the pcCMSA-ES behavior during a single isolation period in the later stage of the optimization process. Having obtained σ_{ss} in Sec. 2, the investigations will assume that the strategy has already reached a constant mutation strength $\sigma \approx \sigma_{ss}$ in the limit of a sufficiently large noise-to-signal ratio ϑ. Accordingly, the dynamical behavior during an initial transient phase is disregarded. Note that, the population size μ is constant during the isolation. Based on these assumptions, the expected y_i^2 dynamics within a single isolation period of the pcCMSA-ES can be determined.

The expected component-wise squared search space parameter transition over a single generation can be formulated as

$$\mathrm{E}[{y_i^{(g+1)}}^2] = \mathrm{E}[{y_i^{(g)}}^2] - \varphi_i^{II}, \tag{17}$$

where φ_i^{II} is the component-wise quadratic progress rate, see [5]. On the noisy ellipsoid model (1), its asymptotical representation was derived in [10]. Assuming the validity of (9) and demanding $\forall i : \sum_{j=1}^N a_j^2 y_j^2 \gg a_i^2 y_i^2$ it holds[2]

$$\varphi_i^{II}(\boldsymbol{y}^{(g)}) \simeq \frac{2c_{\mu/\mu,\lambda}\sigma a_i y_i^2}{\sqrt{(1+\vartheta^2)\sum_{j=1}^N a_j^2 y_j^2}} - \frac{\sigma^2}{\mu}. \tag{18}$$

Notice, that the expected value notations E[.] are omitted within the subsequent calculations for reasons of simplification.

Considering a constant mutation strength $\sigma \to \sigma_{ss}$, one can determine the component-wise y_i^2 dynamics as well as its expected steady state values $(y_{ss})_i^2$ within a single isolation period. The latter will be determined first by applying the steady state condition $\mathrm{E}[{y_i^{(g+1)}}^2] = \mathrm{E}[{y_i^{(g)}}^2] = (y_{ss})_i^2$, which holds for sufficiently large g, yielding

$$\varphi_i^{II}(\boldsymbol{y}_{ss}) \stackrel{!}{=} 0. \tag{19}$$

Consequently, inserting (18) into (19), and considering the renormalized representation of the noise-to-signal ratio (12), Eq. (19) can be solved for the component-wise steady state $(y_{ss})_i^2$. Remembering the prerequisite $\vartheta \gg 1$, the 1 in the denominator in (18) can be neglected. Thus, resolving (19) for $a_i y_i^2$ yield the steady state expression

$$a_i (y_{ss})_i^2 \simeq \frac{\sigma_\epsilon}{4c_{\mu/\mu,\lambda}\mu}, \qquad \forall i = 1,\ldots,N. \tag{20}$$

This steady state value of the pcCMSA-ES changes with each μ adaptation, i.e. it is different in each isolation period. Hence, the required number of generations to actually approach the steady state needs to be derived separately. Based on (20) the expected steady state fitness F_{ss} that is approached within a single isolation period can be determined. Taking into account the noisy ellipsoid model (1) with additive noise term $\delta \sim \mathcal{N}(0,\sigma_\epsilon^2)$, it follows that

$$F_{ss} := \lim_{g\to\infty} \mathrm{E}[\tilde{F}(\boldsymbol{y}^{(g)})] = \sum_{i=1}^N a_i (y_{ss})_i^2 \simeq \frac{N\sigma_\epsilon}{4c_{\mu/\mu,\lambda}\mu}. \tag{21}$$

The result of Eq. (21) was already obtained 12 years ago by use of the equipartition assumption and differential geometry [4]. As a byproduct of our considerations, the present derivation provides a proof based on first principles.

Alongside the steady state fitness description, Eq. (20) provides a steady state characterization of the normalized mutation strength dynamics σ_{ss}^*. Multiplying Eq. (20) with a_i and summing up all components $i = 1\ldots,N$, one obtains the term

$$\sum_{i=1}^N a_i^2 (y_{ss})_i^2 \simeq \frac{\sigma_\epsilon \Sigma a}{4c_{\mu/\mu,\lambda}\mu}. \tag{22}$$

Assuming that the ES operates in its steady state, inserting Eq. (22) into the denominator of the σ normalization in (11) yields

$$\sigma_{ss}^* \simeq \sigma_{ss}\Sigma a \left(\frac{\sigma_\epsilon \Sigma a}{4\mu c_{\mu/\mu,\lambda}}\right)^{-\frac{1}{2}}. \tag{23}$$

Further, by considering Eq. (16) one obtains

$$\sigma_{ss}^* \simeq \sqrt{\frac{\sigma_\epsilon}{4c_{\mu/\mu,\lambda}\Sigma a}} \cdot \sqrt{\frac{4\mu c_{\mu/\mu,\lambda}}{\sigma_\epsilon \Sigma a}} \cdot \Sigma a = \sqrt{\mu}. \tag{24}$$

Thus, during a single isolation period with constant μ, the σ_{ss}^* approaches the square root of the parental population size μ.

Regarding the noise-to-signal ratio ϑ in (12), its steady state dynamics can be characterized by the same token. Applying the normalizations (11), and (13), Eq. (22) as well as (23), yields the steady state of the noise-to-signal ratio

$$\vartheta_{ss} = \sigma_\epsilon^*/\sigma_{ss}^* \simeq 2c_{\mu/\mu,\lambda}\sqrt{\mu}. \tag{25}$$

Equation (25) reveals that ϑ_{ss} is also proportional to the square root of the population sizes μ.

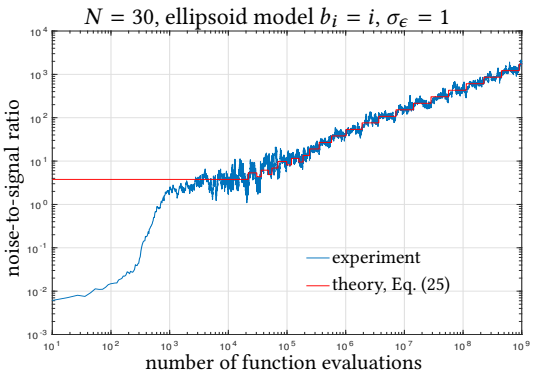

Figure 2: Illustration of the pcCMSA-ES ϑ dynamics plotted against the number of function evaluations. The solid blue lines depict the experimental measurements. The theoretical prediction of (25) is displayed by the red step function.

[2] Note that, Eq. (18) also holds for $\mu \gg N$ in the limit of large ϑ values.

Considering the pcCMSA-ES dynamics on the noisy ellipsoid model (subject to additive fitness noise of constant noise strength σ_ϵ), it can be observed that the parental population size μ is continuously increased as the strategy repeatedly identifies significant noise influences on the fitness dynamics. Acting this way, the pcCMSA-ES is able to approach closer residual distances from the optimizer and as a consequence the noise-to-signal ratio increases. That is, the requirement $\vartheta \to \infty$ for the steady state becomes better and better fulfilled as the parental population size increases. The predicted behavior is reflected by the illustrations in Fig. 2. It can be observed that the experimental dynamics follow the theoretical prediction. As inferred from (25), the noise-to-signal ratio ϑ increases with growing population size μ.

We proceed with a description of the generational progress during the isolation period. The analysis is based on the asymptotical component-wise quadratic progress rate in Eq. (18). Considering the renormalized version of ϑ in Eq. (12) as well as $\vartheta^2 \gg 1$, (18) becomes

$$\varphi_i^{II} \simeq \frac{2c_{\mu/\mu,\lambda}\sigma a_i y_i^2}{\sqrt{\vartheta^2 \sum_{j=1}^N a_j^2 y_j^2}} - \frac{\sigma^2}{\mu} = \frac{4c_{\mu/\mu,\lambda}\sigma^2 a_i y_i^2}{\sigma_\epsilon} - \frac{\sigma^2}{\mu}. \tag{26}$$

Inserting (26) into Eq. (17), one obtains a linear difference equation of the general type $z^{(g+1)} = z^{(g)}(1-u) + v$ with $z = \mathrm{E}[y_i^2]$,

$$u := \frac{4c_{\mu/\mu,\lambda}\sigma^2 a_i}{\sigma_\epsilon} \quad \text{and} \quad v := \frac{\sigma^2}{\mu}. \tag{27}$$

A closed-form solution of which can be derived provided that σ is constant during the isolation period. The solution reads[3]

$$z^{(g)} = z^{(0)}\,(1-u)^g + \frac{v}{u}\left[1-(1-u)^g\right]. \tag{28}$$

Plugging (27) into (28) one obtains the expected value dynamics of the y_i^2

$$\begin{aligned} y_i^{(g)^2} &= y_i^{(0)^2}\left(1 - \frac{4c_{\mu/\mu,\lambda}\sigma^2 a_i}{\sigma_\epsilon}\right)^g \\ &\quad + \frac{\sigma_\epsilon}{4\mu c_{\mu/\mu,\lambda} a_i}\left[1 - \left(1 - \frac{4c_{\mu/\mu,\lambda}\sigma^2 a_i}{\sigma_\epsilon}\right)^g\right]. \end{aligned} \tag{29}$$

At this point the abbreviation $\Theta := \sigma_\epsilon/(4\mu c_{\mu/\mu,\lambda})$ is introduced to reduce the length of the following equations.

Considering a sufficiently large parental population size μ the mutation strength σ in Eq. (29) can be approximately replaced with its steady state value (16), cf. Sec. 2. Consequently, the u-term in (27) asymptotically yields

$$u = \frac{4c_{\mu/\mu,\lambda}\sigma_{ss}^2 a_i}{\sigma_\epsilon} \simeq \frac{\sigma_\epsilon}{4c_{\mu/\mu,\lambda}\Sigma a}\frac{4c_{\mu/\mu,\lambda}a_i}{\sigma_\epsilon} = \frac{a_i}{\Sigma a}, \tag{30}$$

and thus the solution (29) of the linear difference equation becomes

$$y_i^{(g)^2} = \left(y_i^{(0)^2} - \frac{\Theta}{a_i}\right)\left(1 - \frac{a_i}{\Sigma a}\right)^g + \frac{\Theta}{a_i}. \tag{31}$$

[3]For sake of simplicity we use g as generation counter *within* the isolation phases. It should not be confused with the global g-counter in Alg. 1.

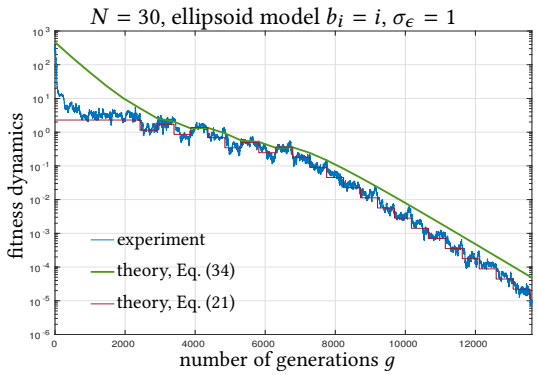

Figure 3: Illustration of the fitness dynamics plotted against the number of generations of the pcCMSA-ES. The solid blue lines depict the experimental measurements, while the steady state fitness (21) is represented by the solid red step function. Each step corresponds to a single isolation period of constant μ. The solid green curve displays the upper bound expected fitness dynamics (34).

This equation characterizes the component-wise evolution of the squared parental parameter vector $\boldsymbol{y}^{(g)^2}$ within an isolation period of the pcCMSA-ES starting from $y_i^{(0)}$.

Considering the noisy ellipsoid model (1) and taking into account (21), Eq. (31) allows for the derivation of the corresponding expected fitness value evolution

$$F^{(g)} = \sum_{i=1}^N \left(a_i y_i^{(0)^2} - \Theta\right)\left(1 - \frac{a_i}{\Sigma a}\right)^g + F_{ss}. \tag{32}$$

Following this representation, the rate at which the fitness dynamics decline depends on the eigenvalues a_i of $\boldsymbol{A}$. 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 (32). Hence, it is governed by

$$\check{a} := \min_{i=1,\ldots,N} a_i. \tag{33}$$

Using (32), the fitness dynamics can be further estimated by

$$F^{(g)} \leq \left(F^{(0)} - F_{ss}\right)\left(1 - \frac{\check{a}}{\Sigma a}\right)^g + F_{ss}. \tag{34}$$

In Fig. 3, the fitness estimate (34) (solid green line) is compared to the experimental fitness dynamics (solid blue line). After a sufficient number of generations, both curves exhibit a similar rate of long-term decline. That is, the theoretical estimate (34) is suitable to upper bound the qualitative long-term behavior of the fitness dynamics. Making use of the observed population sizes μ adapted in the real pcCMSA-ES run, the solid red line predicts the step-wise decrease of the steady state fitness over the corresponding isolation periods well. It is observed that the experimental measurements oscillate closely around the predicted steady state values.

According to Eq. (34), the fitness values in expectation drop at least with the factor $(1-\check{a}/\Sigma a)^g$. Hence, the number of generations G required to obtain a fitness reduction of at least $(1-\delta)$

percent, can be computed by solving

$$\delta \stackrel{!}{=} \left(1 - \frac{\breve{a}}{\Sigma a}\right)^G \tag{35}$$

for G. Applying the logarithm on both sides results in

$$\ln \delta = G \cdot \ln\left(1 - \frac{\breve{a}}{\Sigma a}\right). \tag{36}$$

Considering the power series of the logarithm, the right-hand side (rhs) of Eq. (36) can roughly be estimated by

$$\ln(1 - \frac{\breve{a}}{\Sigma a}) = \sum_{k=1}^{\infty} -\frac{1}{k}\left(\frac{\breve{a}}{\Sigma a}\right)^k \leq -\frac{\breve{a}}{\Sigma a}, \tag{37}$$

with $|\breve{a}/\Sigma a| < 1$. Using (37), G is estimated on the basis of (36) as

$$G \leq \frac{\ln(\delta^{-1})}{\breve{a}}\Sigma a. \tag{38}$$

Equation (38) 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 $\delta \in (0,1)$.[4] The result in Eq. (38) is obtained under the assumption of a sufficiently large and constant parental population size. Moreover, the previous analysis only considered the dynamics within a single isolation period. The description of the overall pcCMSA-ES dynamics is tackled in the next section.

4 LONG-TERM FITNESS DYNAMICS

Having provided the necessary aspects in the previous sections, now the focus is on analyzing the long-term fitness value dynamics of the pcCMSA-ES. 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. (38) which ensures a reduction of the fitness dynamics by a factor of at least δ. During the isolation period the algorithm operates with fixed population sizes μ, λ and truncation ratio $\nu = \mu/\lambda$. Furthermore, it is assumed that the strategy continuously increases μ 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 decision 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 21 in Alg. 1)

$$\mu(t+1) = \alpha\mu(t) = \alpha^t\mu_0, \qquad \text{with} \quad \mu(0) = \mu_0,\ \alpha > 1. \tag{39}$$

Taking into account sufficiently large parental population sizes μ, the noise-to-signal ratio is supposed to increase, see Sec. 3. Accordingly, one can assume that the strategy is operating with a (near) constant mutation strength σ close to the steady state derived in Eq. (16). Further, a constant (co-)variance of the mutation step is assumed: $\tilde{\boldsymbol{y}}_l = \boldsymbol{y} + \sigma\mathcal{N}(0,\boldsymbol{C})$. Concerning the long-term behavior of the pcCMSA-ES in the context of (1) subject to additive noise of constant variance σ_ϵ, a constant matrix $\boldsymbol{C}$ is reasonable since the strategy turns off the covariance matrix adaptation after the first noise detection (cf. line 22 in Alg. 1).

Being based on these assumptions, the analysis starts from the description of the inner fitness dynamics (34). Letting $F(t)$ denote the expected fitness value at the end of the tth isolation period, and taking into account (35), one obtains

$$F(t) \leq (F(t-1) - F_{ss}(t))\,\delta + F_{ss}(t). \tag{40}$$

The respective steady state fitness is provided by (21) and consideration of (39) yields

$$F_{ss}(t) = \frac{\sigma_\epsilon N}{4c_{\mu/\mu,\lambda}\mu(t)} = \frac{\sigma_\epsilon N}{4c_{\mu/\mu,\lambda}\mu_0\alpha^t}. \tag{41}$$

Simple rearrangements transform Eq (40) into

$$F(t) \leq F(t-1)\delta + F_{ss}(t)(1-\delta). \tag{42}$$

Introducing the abbreviation $\omega := \sigma_\epsilon N(1-\delta)/(4c_{\mu/\mu,\lambda}\mu_0)$, Eq. (42) is reformulated as

$$F(t) \leq F(t-1)\delta + \omega\alpha^{-t}. \tag{43}$$

Regarding the iterative relation between two consecutive fitness values (43), the fitness at the end of isolation period t can be successively affiliated with the initial fitness. Following the recursion

$$\begin{aligned} F(t) &\leq F(t-1)\delta + \omega\alpha^{-t} \\ &\leq F(t-2)\delta^2 + \omega\alpha^{-(t-1)}\delta + \omega\alpha^{-t} \leq \ldots, \end{aligned} \tag{44}$$

yields an estimate of the fitness after the tth isolation period

$$\begin{aligned} F(t) &\leq F(0)\delta^t + \omega\left(\alpha^{-1}\delta^{t-1} + \alpha^{-2}\delta^{t-2} + \cdots + \alpha^{-t}\right) \\ &\leq F(0)\delta^t + \frac{\omega\delta^{t-1}}{\alpha}\left(1 + (\alpha\delta)^{-1} + \cdots + (\alpha\delta)^{-(t-1)}\right). \end{aligned} \tag{45}$$

The bracketed term on the rhs of Eq. (45) is the $(t-1)$th partial sum of a geometric series. Hence, one obtains

$$\left(1 + (\alpha\delta)^{-1} + \cdots + (\alpha\delta)^{-(t-1)}\right) = \frac{(\frac{1}{\alpha\delta})^t - 1}{\frac{1}{\alpha\delta} - 1}. \tag{46}$$

Replacing the respective sum in (45) with (46) yields

$$F(t) \leq F(0)\delta^t + \frac{\omega\delta^{t-1}}{\alpha}\,\frac{(\frac{1}{\alpha\delta})^t - 1}{\frac{1}{\alpha\delta} - 1}. \tag{47}$$

Finally, rearranging the terms leads to the representation

$$\boxed{F(t) \leq F(0)\delta^t + \frac{\omega}{1-\alpha\delta}\,\delta^t\left[(\tfrac{1}{\alpha\delta})^t - 1\right].} \tag{48}$$

While the parameters $\alpha > 1$ and $\delta < 1$ can be set independently, the magnitude of the product $\alpha\delta =: \varkappa$ governs the further analysis. Therefore, the two cases $\varkappa < 1$ and $\varkappa > 1$, respectively, have to be distinguished.

The $\varkappa < 1$ 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. (48) transforms into

$$F(t) \leq F(0)\delta^t + \frac{\omega}{1-\alpha\delta}\,\delta^t\,\frac{1}{(\alpha\delta)^t}. \tag{49}$$

Taking a closer look at the rhs of Eq. (49), one observes that

$$F(0)\delta^t + \frac{\omega}{1-\alpha\delta}\,\delta^t\,\frac{1}{(\alpha\delta)^t} = \left(F(0)\varkappa^t + \frac{\omega}{1-\alpha\delta}\right)\frac{1}{\alpha^t}. \tag{50}$$

[4] Notice, that regarding the pcCMSA-ES in Alg. 1, the application of $L \geq G$ suffices to realize the desired dynamical behavior.

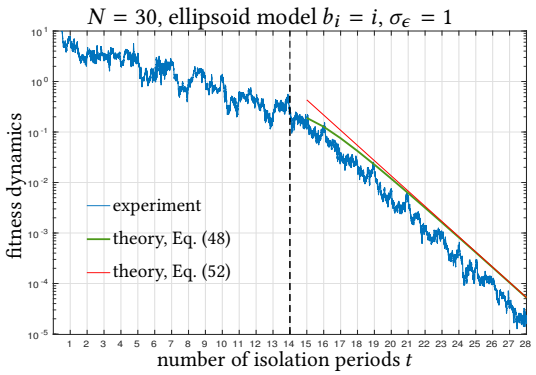

Figure 4: Illustration of the overall fitness dynamics plotted against the number of isolation periods. The theoretical prediction of (48) is represented by the solid green line. The asymptotically obtained prediction of Eq. (52) ($\varkappa < 1$) is illustrated by the red line. The experimental fitness measurements per generation are displayed by the solid blue line.

After factoring out $1/\alpha^t$ on the rhs of (50), considering that $\varkappa^t \to 0$ for $t \to \infty$ yields

$$F(0)\delta^t + \frac{\omega}{1-\alpha\delta}\delta^t \frac{1}{(\alpha\delta)^t} \simeq \frac{\omega}{1-\alpha\delta}\frac{1}{\alpha^t}. \tag{51}$$

Consequently, applying the logarithm on both sides of (49) after insertion of (51) results in

$$\boxed{\ln F(t) \lesssim \ln \frac{\omega}{1-\varkappa} - t \ln \alpha.} \tag{52}$$

The $\varkappa > 1$ case: In this situation, the term $1/(\alpha\delta)^t$ approaches zero for $t \to \infty$. Consequently, Eq. (48) becomes

$$F(t) \leq \left(F(0) + \frac{\omega}{\alpha\delta - 1}\right)\delta^t. \tag{53}$$

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

$$\boxed{\ln F(t) \leq \ln\left(F(0) + \frac{\omega}{\varkappa - 1}\right) - t(\ln\alpha - \ln\varkappa).} \tag{54}$$

A comparison of the theoretically obtained fitness estimates for the $\varkappa < 1$ case is displayed in Fig. 4. There, it is observed that the analysis is able to predict the experimentally obtained rate of decline. According to assumption (39), the predictions are only displayed for those isolation periods where a continuous increase of the population size is ensued. The theoretical results in Fig. 4 almost overlap as the solid green curve from (48) approaches the solid red line representing Eq. (52) with growing t.

5 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 tth 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 $G \cdot \lambda(t) = G \cdot \mu(t)/\nu$. Aggregating these values over all preceding periods, the cumulative number of function evaluations is

$$n(t) = \sum_{k=0}^{t-1} \mu(k)\frac{G}{\nu} = \frac{G}{\nu}\mu_0 \sum_{k=0}^{t-1} \alpha^k = \frac{G}{\nu}\mu_0 \frac{\alpha^t - 1}{\alpha - 1}. \tag{55}$$

Applying the logarithm to Eq. (55) yields

$$\ln n(t) = \ln \frac{G\mu_0}{\nu(\alpha - 1)} + \ln(\alpha^t - 1). \tag{56}$$

Requiring a sufficiently large number of isolation periods, one obtains $\alpha^t - 1 \simeq \alpha^t$. Further, for large t the first term of the sum on the rhs of (56) can be neglected when compared to the second term. Hence, in the asymptotic limit the logarithm of the number of function evaluations linearly increases with the number of isolation periods t

$$\ln n(t) \simeq t \cdot \ln \alpha. \tag{57}$$

6 ASYMPTOTIC CONVERGENCE RATE

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. 4 as well as on the asymptotic $F(t)$ and $n(t)$ dynamics in Eqs. (52), (54) and (57), 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

$$CR := \lim_{t\to\infty} \frac{\ln F(t)}{\ln n(t)}. \tag{58}$$

As the expected fitness dynamics depend on the magnitude of the parameter $\varkappa$, the two cases described in Sec. 4 must be considered.

The $\varkappa < 1$ case: Inserting the asymptotical dynamics (52) and Eq. (57) into the convergence rate definition (58), one obtains

$$CR = \lim_{t\to\infty} \frac{\ln F(t)}{\ln n(t)} = \lim_{t\to\infty} \frac{\ln \frac{\omega}{1-\alpha\delta} - t\ln\alpha}{t\ln\alpha} = -1. \tag{59}$$

Hence, a convergence rate of -1 can be obtained under the condition that $\varkappa = \alpha\delta \leq 1$. Reconsidering δ from (35), 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 (37), and solving the resulting equation for G yields

$$G \geq -\frac{\ln\alpha}{\ln(1 - \check{a}/\Sigma a)} \leq \frac{\ln\alpha}{\check{a}}\Sigma a. \tag{60}$$

Conclusively, choosing an isolation time of $G = \ln\alpha\Sigma a/\check{a}$ ensures that the strategy approaches a convergence rate of $CR \simeq -1$. Equation (60) indicates that the use of population control parameters $1 < \alpha < \mathrm{e}$ reduces the upper bound on the isolation time.[5] Notice, that a convergence rate realization of $CR = -1$ represents the theoretically derived lower bound for all comparison based search algorithms [11]. On the noisy ellipsoid model this descent was empirically discovered in runs of the pcCMSA-ES [8]. The corresponding illustration is provided in Fig. 5. The solid blue lines refer to the

[5]However, choosing α too close to 1 lowers the reliability of the noise hypothesis test within the pcCMSA-ES such that using $\alpha \geq 2$ is recommended.

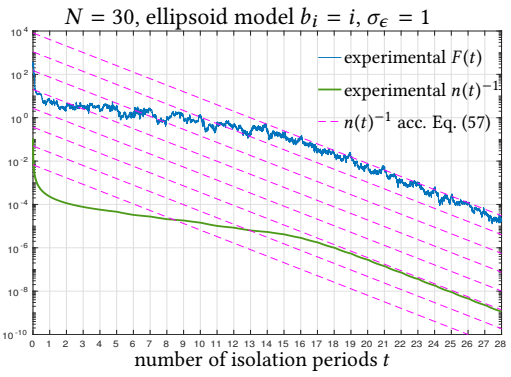

Figure 5: Illustration of the dynamics of the cumulated number of function evaluations. All dynamics are plotted against the number of isolation periods of $L = 465$ generations each.

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. (57). 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. (59).

The $\varkappa > 1$ case: Considering the results from Eq. (54) and Eq. (57), the convergence rate (58) is obtained as

$$CR = \lim_{t\to\infty} \frac{\ln\left(F(0) + \dfrac{\omega}{\alpha\delta - 1}\right) + t(\ln\varkappa - \ln\alpha)}{t\ln(\alpha)} \tag{61}$$

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

$$CR = \frac{(\ln\varkappa - \ln\alpha)}{\ln(\alpha)} = \frac{\ln\varkappa}{\ln\alpha} - 1 > -1. \tag{62}$$

Looking at Eq. (62), it is possible to tune the convergence rate towards -1. Making use of $\varkappa = \alpha\delta$, one obtains

$$CR = \frac{\ln(\alpha\delta)}{\ln\alpha} - 1 = \frac{\ln\alpha + \ln\delta}{\ln\alpha} - 1 = \frac{\ln\delta}{\ln\alpha}. \tag{63}$$

Reconsidering the δ definition (35), Eq. (63) transforms into

$$CR \simeq G \cdot \frac{\ln(1 - \dfrac{\breve{a}}{\Sigma a})}{\ln\alpha}. \tag{64}$$

Resolving Eq. (64) for G, results in

$$G \simeq CR\frac{\ln\alpha}{\ln(1 - \dfrac{\breve{a}}{\Sigma a})} \simeq -CR\frac{\ln\alpha}{\breve{a}}\Sigma a. \tag{65}$$

Choosing G in accordance with (65), the convergence rate of the pcCMSA-ES can be driven towards a desired (suboptimal) value $CR > -1$. Notice that, considering Eq. (54), the case $\varkappa = 1$ is also covered by (62) yielding $CR = -1$.

7 CONCLUSIONS

This paper presented the theoretical analysis of the asymptotical pcCMSA-ES long-term dynamics on the ellipsoid model subject to additive Gaussian noise. We were able to derive the empirically observed $CR = -1$ convergence rate using the progress rate theory from [10]. To reach this theoretically proven best convergence rate, the ES must not behave like a "simple ES" as shown in [2]. That is, the ES must not be scale-invariant w.r.t. the mutation strength when approaching the optimizer. According to Sec. 2, the mutation strength σ of the pcCMSA-ES reaches a steady state value that does not change with the parental population size μ and the distance from the optimizer. However, "violating" the scale-invariance is not sufficient to reach $CR = -1$. As shown in this paper, the population increasing factor α 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 = $ const.) that also depends on the choice of α. G itself depends on the eigenvalue spectrum of the matrix $\boldsymbol{A}$ – a combination of the Hessian of the ellipsoid model and the actual transformation matrix $\sqrt{\boldsymbol{C}}$. Since the covariance matrix $\boldsymbol{C}$ evolves during the evolution, the eigenvalue spectrum evolves as well. This does not present a problem in the noiseless case, but in the strong noise case $\boldsymbol{C}$ accumulates random noise that results in a fast eigenvalue blow-up. Hence, G cannot be fixed any longer and $CR = -1$ cannot be reached. Therefore, it is algorithmically important (and implemented in the pcCMSA-ES, Alg. 1) to stall the $\boldsymbol{C}$-update if strong noise is detected. Consequently, under strong noise the $\boldsymbol{C}$-update appears not only useless, but counter-productive.

ACKNOWLEDGEMENTS

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

REFERENCES

[1] Dirk V. Arnold. 2002. *Noisy optimization with evolution strategies.* Kluwer Academic Publishers.
[2] S. Astete-Morales, M.-L. Cauwet, and O. Teytaud. 2015. Evolution Strategies with Additive Noise: A Convergence Rate Lower Bound. In *Proceedings of the 2015 ACM Conference on Foundations of Genetic Algorithms XIII (FOGA '15).* ACM, New York, NY, USA, 76–84.
[3] H.-G. Beyer. 2001. *The Theory of Evolution Strategies.* Springer.
[4] H.-G. Beyer, Dirk V. Arnold, and S. Meyer-Nieberg. 2005. A New Approach for Predicting the Final Outcome of Evolution Strategy Optimization Under Noise. *Genetic Programming and Evolvable Machines* 6, 1 (2005), 7–24.
[5] H.-G. Beyer and A. Melkozerov. 2014. The Dynamics of Self-Adaptive Multi-Recombinant Evolution Strategies on the General Ellipsoid Model. *IEEE Transactions on Evolutionary Computation* 18, 5 (2014), 764–778.
[6] H.-G. Beyer and B. Sendhoff. 2008. Covariance Matrix Adaptation Revisited – The CMSA Evolution Strategy. In *Parallel Problem Solving from Nature – PPSN X*, G. Rudolph *et al.* (Eds.). *LNCS 5199*, Springer, 123–132.
[7] N. Hansen. 2006. An analysis of mutative σ-self-adaptation on linear fitness functions. *Evolutionary Computation* 14, 3 (2006), 255–275.
[8] M. Hellwig and H.-G. Beyer. 2016. Evolution under Strong Noise: A Self-Adaptive Evolution Strategy Can Reach the Lower Performance Bound - the pcCMSA-ES. In *Parallel Problems Solving from Nature - PPSN XIV*, J. Handl *et al.* (Eds.). *LNCS 9921*, Springer, 26–37.
[9] Y. Jin and J. Branke. 2005. Evolutionary optimization in uncertain environments – a survey. *IEEE Trans. Evolutionary Computation* 9, 3 (2005), 303–317.
[10] A. Melkozerov and H.-G. Beyer. 2015. Towards an Analysis of Self-Adaptive Evolution Strategies on the Noisy Ellipsoid Model: Progress Rate and Self-Adaptation Response. In *GECCO 2015, Madrid, Spain.* ACM, 297–304.
[11] O. Shamir. 2013. On the Complexity of Bandit and Derivative-Free Stochastic Convex Optimization.. In *COLT - The 26th Annual Conference on Learning Theory, NJ, USA.* 3–24.