

Towards an Analysis of Self-Adaptive Evolution Strategies on the Noisy Ellipsoid Model: Progress Rate and Self-Adaptation Response

Alexander Melkozerov

Department of Television and Control
Tomsk State University
of Control Systems
and Radioelectronics
Lenin ave. 40, 634050 Tomsk, Russia
ame@tu.tusur.ru

Hans-Georg Beyer

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

ABSTRACT

This paper analyzes the multi-recombinant self-adaptive evolution strategy (ES), denoted as $(\mu/\mu_I, \lambda)\text{-}\sigma\text{SA-ES}$ on the convex-quadratic function class under the influence of noise, which is referred to as noisy ellipsoid model. Asymptotically exact progress rate and self-adaptation response measures are derived (i.e., for $N \rightarrow \infty$, N – search space dimensionality) for the considered objective function model and verified using experimental ES runs.

Categories and Subject Descriptors

I.2.6 [Artificial Intelligence]: Learning—parameter learning; G.1.6 [Numerical Analysis]: Optimization

Keywords

Evolution strategy, ellipsoid model, noise, self-adaptation, progress rate, mutation strength

1. INTRODUCTION

Theoretical analysis of evolutions strategies (ES) is an area of active research, where much of the recent effort has been devoted to the extension of the range of the test functions analyzed, including particular cases of the positive definite quadratic forms (PDQFs) [4]. The general case of the PDQF, referred to as general ellipsoid model, has been treated in [7, 5], where asymptotically exact quadratic progress rate and self-adaptation rate formulae have been obtained for the self-adaptation evolutionary strategy with intermediate recombination (the $(\mu/\mu_I, \lambda)\text{-}\sigma\text{SA-ES}$). Based on these results, the expected value dynamics of the ES have been derived and in turn the optimal learning parameter,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](http://permissions.acm.org).

GECCO '15 July 11–15, 2015, Madrid, Spain

© 2015 ACM. ISBN 978-1-4503-3472-3/15/07...\$15.00

DOI: <http://dx.doi.org/10.1145/2739480.2754800>

which provides for the highest possible convergence rate, has been calculated analytically.

Still, the aforementioned analysis has been done for the noise-free ellipsoid model which does not allow to account for noisy distortions of the objective function values often arising in practical optimization tasks. The aim of this paper is to extend the analysis of the $(\mu/\mu_I, \lambda)\text{-}\sigma\text{SA-ES}$ to the case of the noisy ellipsoid model and derive the corresponding local progress measures.

The paper is organized as follows. First, the noisy ellipsoid model to be analyzed is presented followed by the description of the the $(\mu/\mu_I, \lambda)\text{-}\sigma\text{SA-ES}$ algorithm. The noisy progress measures are introduced in Section 2, where their outcome is also compared with results of one-generation experiments. Section 3 is devoted to the self-adaptation response function, which is derived for the noisy ellipsoid model and tested for validity by comparison with the $(\mu/\mu_I, \lambda)\text{-}\sigma\text{SA-ES}$ experimental runs. Finally, the obtained results are discussed in the conclusion.

1.1 Noisy Ellipsoid Model

The consideration of noisy fitness environments is of particular importance because noise often arises in practical optimization tasks. For example, the objective function may depend on noisy physical measurements or computer simulations with simplified models. In both cases, the ES observes a noise-disturbed $F_{\text{noisy}}(\mathbf{y})$ value. The *perceived* fitness [3] $F_{\text{noisy}}(\mathbf{y})$ consists of an *ideal* noise-free fitness influenced by the noise term. Therefore, the noisy ellipsoid model is defined as

$$F_{\text{noisy}}(\mathbf{y}) = \sum_{i=1}^N a_i y_i^2 + \sigma_\epsilon(\mathbf{y}) Z, \quad \mathbf{y} \in \mathbb{R}^N, \quad a_i > 0, \quad (1)$$

where $\sigma_\epsilon(\mathbf{y})$ is the *noise strength*, or standard deviation of the noise term, and Z is a random variate. Eq. (1) is a noise-disturbed model where the noise term represents random influence of such factors as measurement errors, approximations, or calculations with limited accuracy.

From a number of noise models considered for the ES analysis [1], it is assumed in this work that the distribution of Z is Gaussian, i.e., $Z \sim \mathcal{N}(0, 1)$ and $\mathcal{N}(0, 1)$ is a standard normally distributed random variate.

1.2 ES Algorithm

The $(\mu/\mu_I, \lambda)$ - σ SA-ES algorithm is presented in Alg. 1. Note that a symbol with superscript (g) is used to refer to a value in a given generation g , $g = 0, 1, \dots, g_{max}$. The subscript l after a symbol refers to a value associated with the l th offspring, $l = 1, 2, \dots, \lambda$.

Algorithm 1 The algorithm of the $(\mu/\mu_I, \lambda)$ - σ SA-ES

```

1  $\sigma^{(0)} \leftarrow \sigma_{init}$ ,  $\mathbf{y}^{(0)} \leftarrow \mathbf{y}_{init}$ ,  $g \leftarrow 0$ 
2 do
3   for  $l = 1, \dots, \lambda$  begin
4      $\tilde{\sigma}_l \leftarrow \sigma^{(g)} e^{\tau \mathcal{N}_I(0,1)}$ 
5      $\tilde{\mathbf{z}}_l \leftarrow \mathcal{N}_I(\mathbf{0}, \mathbf{I})$ 
6      $\tilde{\mathbf{y}}_l \leftarrow \mathbf{y}^{(g)} + \tilde{\sigma}_l \tilde{\mathbf{z}}_l$ 
7      $\tilde{F}_l \leftarrow F(\tilde{\mathbf{y}}_l)$ 
8   end
9    $\tilde{\mathbf{F}}_{sort} \leftarrow \text{sort}(\tilde{F}_{1\dots\lambda})$ 
10   $\sigma^{(g+1)} \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\sigma}_{m;\lambda}$ 
11   $\mathbf{y}^{(g+1)} \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\mathbf{y}}_{m;\lambda}$ 
12   $g \leftarrow g + 1$ 
13 until termination criterion fulfilled

```

The parental mutation strength $\sigma^{(0)}$ and the parental parameter vector, or *parental centroid* $\mathbf{y}^{(0)}$ are initialized in line 1. λ offspring are generated from line 3 to line 8 in the following way. For each offspring, the mutation of $\sigma^{(g)}$ is performed in line 4 using the log-normal operator $e^{\tau \mathcal{N}_I(0,1)}$, where $\mathcal{N}_I(0, 1)$ is a $(0, 1)$ normally distributed random scalar. The learning parameter τ in the log-normal operator controls the self-adaptation rate. In line 5, direction of the mutation vector $\tilde{\sigma}_l \tilde{\mathbf{z}}_l$ is determined by means of a $(0, 1)$ normally distributed random vector $\mathcal{N}_I(\mathbf{0}, \mathbf{I})$. The offspring parameter vector $\tilde{\mathbf{y}}_l$ is generated in line 6 and used in the calculation of the objective function value \tilde{F}_l in line 7.

After creation, λ offspring are ranked according to their \tilde{F}_l values in line 9. The recombination of offspring mutation strengths and parameter vectors is performed in lines 10–11 in order to obtain a new parental mutation strength $\sigma^{(g+1)}$ and a new parental parameter vector $\mathbf{y}^{(g+1)}$. The subscript $m; \lambda$ refers to the m th-best of λ offspring (the m th-smallest for minimization).

In line 13, the termination criterion is checked. If it is fulfilled, the current parental parameter vector is considered an approximation of the optimizer of the objective function $F(\mathbf{y})$. Otherwise, the algorithm returns to line 2.

2. NOISY PROGRESS RATE

First, a noisy version of the $(\mu/\mu_I, \lambda)$ - σ SA-ES progress rate, defined as

$$\varphi_i = E \left[y_i^{(g)} - y_i^{(g+1)} | \mathbf{y}^{(g)} \right], \quad (2)$$

is derived which is used in the noisy quadratic progress rate formula obtained next.

Noise-disturbed objective function evaluations influence the $(\mu/\mu_I, \lambda)$ - σ SA-ES selection process because the offspring ranking depends on the perceived fitness $F_{\text{noisy}}(\mathbf{y})$. The influence of the noise on the selection is modeled by the noisy local quality change introduced in the following definition. Note that the noise-free local quality change definition

$$Q_{\mathbf{y}}(\mathbf{x}) := F(\mathbf{y} + \mathbf{x}) - F(\mathbf{y}) \quad (3)$$

can not be adapted to the noisy case by substitution of $F(\mathbf{y})$ with $F_{\text{noisy}}(\mathbf{y})$ because it leads to an incorrect formula with the noise term included twice. The $(\mu/\mu_I, \lambda)$ - σ SA-ES evaluates the objective function of each individual only once (as it uses the “comma”-selection which discards individuals from the previous generation), therefore, $Q_{\text{noisy}}(\mathbf{x})$ must contain one noise term.

Let \mathbf{y} be the ES parental parameter vector and \mathbf{x} be a mutation vector applied to it (cf. line 6 in Alg. 1). Then the noisy local quality change is defined as [6]

$$Q_{\text{noisy}}(\mathbf{x}, \mathbf{y}) := Q_{\mathbf{y}}(\mathbf{x}) + \sigma_{\epsilon}(\mathbf{y} + \mathbf{x}) \mathcal{N}(0, 1), \quad (4)$$

where $Q_{\mathbf{y}}(\mathbf{x})$ is the noise-free local quality change given by Eq. (3) and $\sigma_{\epsilon}(\mathbf{y} + \mathbf{x})$ is the noise strength.

The noise strength $\sigma_{\epsilon}(\mathbf{y})$ in (4) may represent different noise models including the constant non-normalized noise model as well as the constant normalized noise model. However, for the latter it will be assumed that $\sigma_{\epsilon}(\mathbf{y}^{(g)}) \simeq \sigma_{\epsilon}(\tilde{\mathbf{y}}_l)$, which states that the noise strength values of the parental individual and offspring are sufficiently close to each other (this holds exactly for $N \rightarrow \infty$ or sufficiently small mutation steps). Using Eq. (4), the progress rate formula for the $(\mu/\mu_I, \lambda)$ - σ SA-ES on the noisy ellipsoid model is obtained in the next section.

2.1 Progress Rate φ

The derivation steps for the noisy progress rate formula are analogous to the noise-free case presented in [7] taking into account that $Q_{\text{noisy}}(\mathbf{x}, \mathbf{y})$ is used in place of $Q_{\mathbf{y}}(\mathbf{x})$. The resulting formula for the progress rate of the $(\mu/\mu_I, \lambda)$ -ES along the i th axis of the noisy ellipsoid model (1) without dominating¹ coefficients a_i is

$$\varphi_i(\sigma) \simeq \frac{2\sigma c_{\mu/\mu, \lambda} y_i a_i}{\sqrt{\sigma_{\epsilon}^2/\sigma^2 + \sum_{j=1}^N 2a_j^2 (2y_j^2 + \sigma^2)}}, \quad (5)$$

where the progress coefficient $c_{\mu/\mu, \lambda} := e_{\mu, \lambda}^{1,0}$ is a special case of generalized progress coefficients

$$e_{\mu, \lambda}^{a, b} = \frac{\lambda - \mu}{\sqrt{2\pi}^{a+1}} \binom{\lambda}{\mu} \int_{-\infty}^{+\infty} (-t)^b e^{-\frac{a+1}{2}t^2} \times (1 - \Phi(t))^{\lambda - \mu - 1} \Phi(t)^{\mu - a} dt, \quad (6)$$

The derivation of Eq. (5) is sketched in Appendix A.

Since the noise strength $\sigma_{\epsilon}(\mathbf{y})$ in general case depends on the parameter vector \mathbf{y} of an individual, it is normalized in order to obtain formula invariant to the position in the search space. Departing from the normalization used for the sphere model [6] $\sigma_{\epsilon}^* = \sigma_{\epsilon} N / (2(R^{(g)})^2)$ (where $R^{(g)}$ is the distance to the optimizer at generation g), the following generalization is used for the noise strength normalization on the ellipsoid model

$$\sigma_{\epsilon}^* = \sigma_{\epsilon} \sum_{i=1}^N a_i / \left(2 \sum_{j=1}^N a_j^2 y_j^2 \right). \quad (7)$$

¹This assumption guarantees that the Lindeberg condition is fulfilled and the central limit theorem can be used in the limit $N \rightarrow \infty$.

Applying it together with the mutation strength normalization

$$\sigma^{*(g)} = \sigma^{(g)} \sum_{i=1}^N a_i / \sqrt{\sum_{i=1}^N a_i^2 y_i^{(g)2}} \quad (8)$$

to the noisy progress rate (5) yields

$$\varphi_i(\sigma^*) \simeq \sigma^* c_{\mu/\mu, \lambda} a_i y_i / \left(\sqrt{1 + (\sigma_\epsilon^*/\sigma^*)^2} \sum_{j=1}^N a_j \right), \quad (9)$$

where the assumption

$$\frac{(\sigma^*)^2}{2} \cdot \frac{\sum_{j=1}^N a_j^2}{\left(\sum_{j=1}^N a_j \right)^2} \ll 1 \quad (10)$$

has been used. The assumption (10) is valid for sufficiently small σ^* values. For the cases $a_i = i, i^2$, it is fulfilled if $(\sigma^*)^2/N \ll 1$. The term $\sigma_\epsilon^*/\sigma^*$ in Eq. (9) is referred to as *noise-to-signal ratio*

$$\vartheta = \sigma_\epsilon^*/\sigma^*. \quad (11)$$

After the progress rate normalization

$$\varphi_i^* := \varphi_i \sum_{j=1}^N a_j \quad (12)$$

and substitution of ϑ , the normalized progress rate of the $(\mu/\mu_I, \lambda)$ -ES on the noisy ellipsoid model reads

$$\varphi_i^*(\sigma^*) \simeq \sigma^* c_{\mu/\mu, \lambda} a_i y_i \frac{1}{\sqrt{1 + \vartheta^2}}. \quad (13)$$

Eq. (13) shows that the noisy objective function evaluations negatively influence the $(\mu/\mu_I, \lambda)$ -ES progress rate: The larger is the noise-to-signal ratio ϑ , the smaller is the progress rate. Note that Eq. (13) has the same deficiencies as its noise-free counterpart [7] $\varphi_i^*(\sigma^*) = \sigma^* c_{\mu/\mu, \lambda} a_i y_i$, in that it does not measure the approach toward the optimizer. To this end, the second-order y_i^2 term must be taken into account. Due to these reasons, Eq. (13) is used exclusively as a part of the noisy quadratic progress rate formula obtained in the next section.

2.2 Quadratic Progress Rate φ_i^{II}

The quadratic progress rate φ_i^{II} is defined as

$$\varphi_i^{II} = E \left[\left(y_i^{(g)} \right)^2 - \left(y_i^{(g+1)} \right)^2 | \mathbf{y}^{(g)} \right]. \quad (14)$$

The expression for φ_i^{II} on the noisy ellipsoid model (1) is derived by considering expectations of the respective product moments. Using the derivation steps for the noise-free quadratic progress rate from [5], one obtains for the noisy case the same equation

$$\varphi_i^{II} = 2y_i \varphi_i - \frac{2}{\mu^2} E_1 - \frac{1}{\mu^2} E_2, \quad (15)$$

which includes the noisy progress rate φ_i given by Eq. (5) and the product moments E_1 and E_2 . The sums of product moments E_1 and E_2 for the noisy case are provided in Appendix B. Inserting Eq. (49) and (50) into (15) yields the quadratic progress rate formula

$$\begin{aligned} \varphi_i^{II}(\sigma) &\simeq 2y_i \varphi_i(\sigma) \\ &- \frac{\sigma^2}{\mu} \left[1 + ((\mu - 1) e_{\mu, \lambda}^{2,0} + e_{\mu, \lambda}^{1,1}) \frac{a_i^2 y_i^2}{\frac{\sigma_\epsilon^2}{4\sigma^2} + \sum_{j=1}^N a_j^2 (y_j^2 + \frac{\sigma^2}{2})} \right], \end{aligned} \quad (16)$$

where the progress rate $\varphi_i(\sigma)$ is given by (5) and the progress coefficients $e_{\mu, \lambda}^{2,0}$ and $e_{\mu, \lambda}^{1,1}$ are calculated using (6).

Applying to Eq. (16) the mutation strength normalization (8), taking into account the assumption (10) and the progress rate normalization (12) leads to

$$\begin{aligned} \varphi_i^{II*}(\sigma^*) &\simeq 2y_i \varphi_i^*(\sigma^*) \\ &- \frac{(\sigma^*)^2}{\mu \sum_{j=1}^N a_j} \left[\sum_{j=1}^N a_j^2 y_j^2 + ((\mu - 1) e_{\mu, \lambda}^{2,0} + e_{\mu, \lambda}^{1,1}) \frac{a_i^2 y_i^2}{1 + \vartheta^2} \right], \end{aligned} \quad (17)$$

where $\varphi_i^*(\sigma^*)$ is given by (13).

As expected, the noisy quadratic progress rate formula (17) yields for $\vartheta = 0$ the corresponding noise-free equation obtained in [5]. For $\vartheta > 0$, the gain term $2y_i \varphi_i^*(\sigma^*)$ is decreased by a factor of $1/\sqrt{1 + \vartheta^2}$ (cf. Eq. (13)). Interestingly, the second part in the loss term in Eq. (17), i.e., the last factor in the second line of (17) also gets smaller with increasing ϑ .

A simplified φ_i^{II*} formula can be obtained from Eq. (17) by discarding the expression $((\mu - 1) e_{\mu, \lambda}^{2,0} + e_{\mu, \lambda}^{1,1}) a_i^2 y_i^2$ in the loss term

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

Denormalization of Eq. (18) yields

$$\varphi_i^{II}(\sigma) \simeq \frac{2\sigma c_{\mu/\mu, \lambda} a_i y_i^2}{\sqrt{(1 + \vartheta^2) \sum_{i=1}^N a_i^2 y_i^2}} - \frac{\sigma^2}{\mu}, \quad (19)$$

where $\vartheta^2 = \sigma_\epsilon^2 / \left(4\sigma^2 \sum_{i=1}^N a_i^2 y_i^2 \right)$. Eq. (19) is a rough approximation of Eq. (17), and its validity is checked in the next section by means of one-generation experiments.

2.3 One-Generation Experiments

In this section, the procedure for one-generation experiments is employed to compare the theoretical predictions of Eqs. (13), (17) and (18) with experimental results. In the experimental code, an important implementation detail is the noise strength denormalization: Due to the assumption $\sigma_\epsilon(\mathbf{y}^{(g)}) \approx \sigma_\epsilon(\hat{\mathbf{y}}_l)$, the parental parameter vector $\mathbf{y}^{(g)}$ is used to denormalize σ_ϵ^* and the resulting single σ_ϵ participates in the \tilde{F}_l calculation for all offspring.

First, the progress rate φ_i formulae (13) and (5) are checked in Fig. 1 using the $(3/3_I, 10)$ -ES one-generation experiments for $a_i = 1$ and $a_i = i$. The experimental settings are $\sigma_\epsilon^* = 10$, $G = 10^6$ and $\mathbf{y}^{(0)} = \mathbf{1}$. Each point in Fig. 1 represents

the mean of 100 one-generation experiments, while bars depict their standard deviations. φ_i^*/a_i points for $N = 400$, $a_i = i$ are not shown due to large deviations.

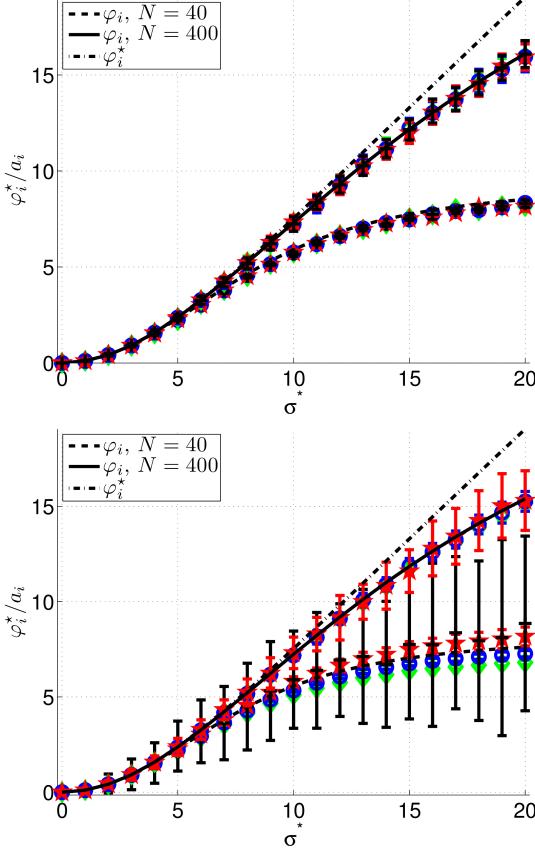


Figure 1: One-generation experiments for the $(3/3_I, 10)$ -ES for $a_i = 1$ (upper figure) and $a_i = i$ (lower figure), $\sigma_e^* = 10$. Curves depict theoretical predictions of Eq. (13) (dash-dot lines) and (5) (dashed curves for $N = 40$ and solid curves for $N = 400$, $y_i = 1$), while points represent experimental results: $+ \varphi_1^*/a_1$, $* \varphi_{N/4}^*/a_{N/4}$, $\circ \varphi_{N/2}^*/a_{N/2}$ and $\diamond \varphi_N^*/a_N$.

In Fig. 1, the first-order φ_i^* results (dash-dot lines) obtained using Eq. (13) are depicted. Note that φ_i^* lines for $N = 40$ and $N = 400$ coincide due to the φ_i^*/a_i normalization. In comparison to the noise-free case, φ_i^* values are smaller since noisy objective function evaluations reduce the progress rate of the $(\mu/\mu_I, 10)$ -ES. For small σ^* , the reduction is particularly large because the ES generates new offspring which lie close to the parental individuals and the difference between the ideal fitness values is dominated by the noise term (cf. Eq. (1)). Experimental φ_i^* mean values match theoretical curves for the same ranges of σ^* values as in the noise-free case. That is, errors due to assumptions used in the derivation of φ_i^* formulae do not increase for the noisy ellipsoid model and both Eqs. (13) and (5) remain asymptotically exact.

In Fig. 2, the quadratic progress rate φ_i^{II*} formula (17) and its approximation (18) are compared with the results of one-generation experiments. The experimental settings

$G = 10^6$ and $\mathbf{y}^{(0)} = \mathbf{1}$ are the same as in the φ_i experiments. The results of the $(3/3_I, 10)$ -ES one-generation experiments for $a_i = 1$ and $a_i = i$ are shown in Fig. 2.

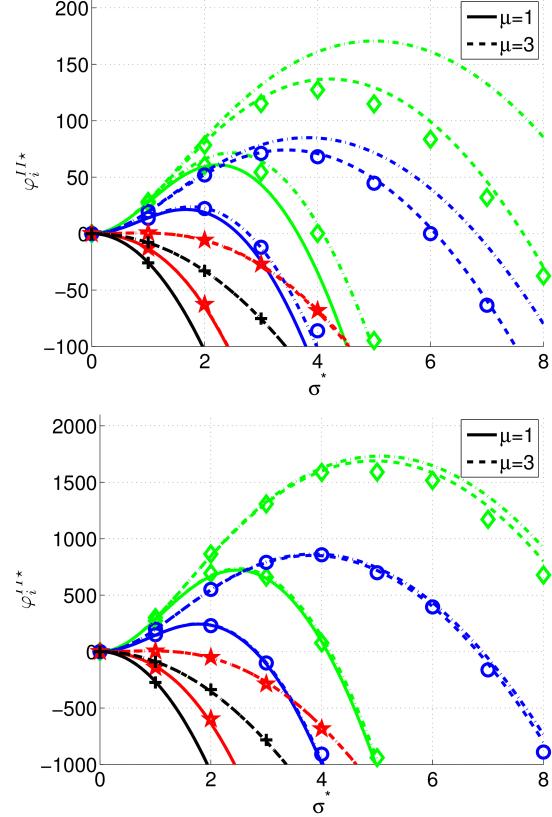


Figure 2: One-generation experiments for the $(3/3_I, 10)$ -ES for $N = 40$ (upper figure) and $N = 400$ (lower figure), $\sigma_e^* = 2$. The solid curves and dashed curves depict theoretical predictions of Eq. (17) for $\mu = 1$ and $\mu = 3$, respectively, while points represent experimental results for $N = 40$ and $N = 400$ ($y_i = 1$): $+ \varphi_1^{II*}$, $* \varphi_{N/4}^{II*}$, $\circ \varphi_{N/2}^{II*}$ and $\diamond \varphi_N^{II*}$. Dot-dash curves show the results of the simplified formula (18).

According to the results depicted in Fig. 2, the quadratic progress rate approximation using Eq. (17) improves for $N = 400$ (lower figure) in comparison with $N = 40$ (upper figure) as solid and dashed curves corresponding to Eq. (17) are situated closer to the experimental points in the former case. This observation is in accordance to the fact that simplifications based on the assumption $N \rightarrow \infty$ have been made during the derivation of Eq. (17). Still, the experimental behavior is correctly reproduced in the $N = 40$ case as well: the $\varphi_i^{II*}(\sigma^*)$ dependency has the characteristic form with a maximum, after which the negative loss terms in φ_i^{II*} prevail over the positive gain term. The results of the simplified Eq. (18) (dot-dash curves) approach the curves corresponding to Eq. (17) for sufficiently small σ^* and replicate the functional dependency of Eq. (17) for larger σ^* values. Thus, Eq. (18) can be regarded as an upper bound estimate of the quadratic progress rate of the $(\mu/\mu_I, \lambda)$ -ES on the noisy ellipsoid model.

3. THE SELF-ADAPTATION RESPONSE

The self-adaptation response (SAR) function is defined as the expected relative mutation strength change from the generation g to $(g+1)$

$$\psi = E \left[\left(\sigma^{(g+1)} - \sigma^{(g)} \right) / \sigma^{(g)} \mid \mathbf{y}^{(g)}, \sigma^{(g)} \right]. \quad (20)$$

The derivation steps for the SAR formula are similar to the noise-free case presented in [7]. One obtains for the $(\mu/\mu_I, \lambda)$ - σ SA-ES on the noisy ellipsoid model (1)

$$\begin{aligned} \psi(\sigma) \simeq \tau^2 & \left(\frac{1}{2} + e_{\mu,\lambda}^{1,1} \frac{\sum_{i=1}^N 2a_i^2 (2y_i^2 + \sigma^2)}{\frac{\sigma_e^2}{\sigma^2} + \sum_{i=1}^N 2a_i^2 (2y_i^2 + \sigma^2)} \right. \\ & \left. - c_{\mu/\mu_I, \lambda} \sigma \frac{2 \sum_{i=1}^N a_i}{\sqrt{\sigma_e^2/\sigma^2 + \sum_{i=1}^N 2a_i^2 (2y_i^2 + \sigma^2)}} \right), \quad (21) \end{aligned}$$

where the progress coefficients $c_{\mu/\mu_I, \lambda} = e_{\mu,\lambda}^{1,0}$ and $e_{\mu,\lambda}^{1,1}$ are special cases of generalized progress coefficients (6) and τ is the learning parameter (cf. line 4 in Alg. 1). A very short sketch of the derivation of Eq. (21) is presented in App. C.

Applying to Eq. (21) the normalization (8) and taking into account the assumption (10) yields

$$\psi(\sigma^*) \simeq \tau^2 \left(\frac{1}{2} + e_{\mu,\lambda}^{1,1} \frac{1}{1 + \vartheta^2} - c_{\mu/\mu_I, \lambda} \sigma^* \frac{1}{\sqrt{1 + \vartheta^2}} \right). \quad (22)$$

Similar to the case of the noise-free SAR function [7], the normalized noisy SAR formula (22) has the same form as the normalized SAR function obtained in [9] for the $(\mu/\mu_I, \lambda)$ - σ SA-ES on the noisy sphere model. That is, Eq. (22) generalizes the previously published result to the class of noisy ellipsoid models (1) by introducing an appropriate mutation strength normalization.

3.1 One-Generation Experiments

In Fig. 3, the results of one-generation experiments are shown for the $(3/3_I, 10)$ - σ SA-ES with initial parameter vector $\mathbf{y}^{(0)} = \mathbf{1}$, learning parameter $\tau = 1/\sqrt{N}$ on the sphere model $a_i = 1$ and ellipsoid model with coefficients $a_i = i$ and $a_i = i^2$. Each point represents the mean of 10 one-generation experiments, while the corresponding standard deviations are smaller than the size of the data points and not shown. Note that the ψ values on the vertical axis are multiplied by N . Due to the mutation strength normalization (8), experimental points for $a_i = 1$, $a_i = i$ and $a_i = i^2$ coincide for the $N = 400$ case.

The experimental results in Fig. 3 are presented for 4 noise strength values (from the bottom group of curves to the top: solid curves for $\sigma_e^* = 1$, curves with longer dashes for $\sigma_e^* = 5$, solid curves for $\sigma_e^* = 10$, and curves with shorter dashes for $\sigma_e^* = 20$, respectively). The experimental points in Fig. 3 match the theoretical curves based on Eqs. (22) and (21) (groups of 3 curves with different colors) for sufficiently small σ^* values. These borderline σ^* values increase for larger N (cf. the upper figure for $N = 40$ and the lower figure for $N = 400$) due to the assumption $N \rightarrow \infty$ used in the analysis. As expected, the curves corresponding to Eq. (21) lie closer to the experimental points than the curves depicting the simplified Eq. (22) with the difference

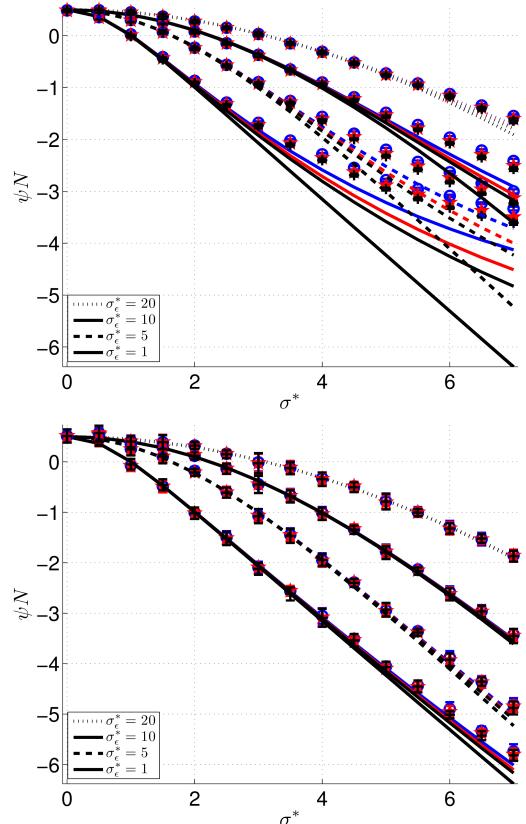


Figure 3: One-generation experiments for the $(3/3_I, 10)$ - σ SA-ES for $N = 40$ (upper figure) and $N = 400$ (lower figure), $y_i = 1$, $\tau = 1/\sqrt{N}$, $\sigma_e^* = 1, 5, 10, 20$ from the bottom group of curves to the top. Curves depict theoretical predictions of Eqs. (22) and (21), respectively, multiplied by N , while points represent experimental results for $N = 40$ and $N = 400$: $+ a_i = 1$, $\star a_i = i$, $\circ a_i = i^2$.

more pronounced for the smaller $N = 40$. The approximation quality appears to improve with increasing noise (compare, for example, the lower curves for $\sigma_e^* = 1$ with the upper curves for $\sigma_e^* = 20$). This issue does not stem from the better approximation for the larger noise strength, but is due to smaller SAR values for higher noise strengths and can be resolved with appropriate normalization.

4. SUMMARY AND CONCLUSIONS

The analysis of the $(\mu/\mu_I, \lambda)$ - σ SA-ES on the noisy ellipsoid model (1) between two consecutive generations has been performed. To this end, an asymptotically exact noisy progress rate (9) and noisy quadratic progress rate (17) have been derived along with corresponding simplified formulae (13) and (18), respectively. Comparison with experiments showed that the obtained formula predicts the noisy progress rate of the $(\mu/\mu_I, \lambda)$ -ES satisfactorily even for the $N = 40$ case despite of the $N \rightarrow \infty$ assumption used in the derivations.

As one can conclude from the experimental results, the SAR function is satisfactorily approximated by Eq. (22) for small mutation strength values. Since the $(\mu/\mu_I, \lambda)$ - σ SA-

ES in the steady state usually yields $\sigma_{\text{st}}^* = 0, \dots, 2.5$, this behavior should not pose a problem even for $N = 40$. For smaller N , a more precise Eq. (21) can be used, although one should keep in mind that the assumption $N \rightarrow \infty$ has been used in the derivation of the SAR function.

Following the steps of the noise-free analysis further, the continuation of the work will be the formulation of a time discrete system describing the evolutionary dynamics of the $(\mu/\mu_I, \lambda)$ - σ SA-ES. Solutions of this system will lead to an analytical formula for the optimal learning parameter of the self-adaptive ES on the noisy ellipsoid model.

Acknowledgments

This work was supported by the Austrian Science Fund FWF under grants P19069-N18 and P22649-N23. The work of A. Melkozerov was supported by the state contract 8.1802.2014/K of the Russian Ministry of Education and Science (software development), the RFBR grant 14-29-0925 (analytic derivations), and the RSF grant 14-19-01232 (numerical experiments) in the TUSUR University.

5. REFERENCES

- [1] Arnold, D.V. and Beyer, H.-G. A general noise model and its effects on evolution strategy performance. *IEEE Trans. Evol. Comp.*, 10(4) (2006), 380–391.
- [2] Arnold, D.V. *Noisy Optimization with Evolution Strategies*. Kluwer, Dordrecht, 2002.
- [3] Arnold, D.V. and Beyer, H.-G. Performance analysis of evolution strategies with multi-recombination in high-dimensional search spaces disturbed by noise. *Theoretical Computer Science*, 289 (1) (2002), 629–647.
- [4] Beyer, H.-G. and Finck, S. Performance of the $(\mu/\mu_I, \lambda)$ - σ SA-ES on a class of PDQFs. *IEEE Trans. Evol. Comp.*, 14 (3) (2010), 400–418.
- [5] Beyer, H.-G. and Melkozerov, A. The dynamics of self-adaptive multi-recombinant evolution strategies on the general ellipsoid model. *IEEE Trans. Evol. Comp.*, 18(5) (2014), 764–778.
- [6] Beyer, H.-G. *The Theory of Evolution Strategies*. Natural Computing Series. Springer, Heidelberg, 2001.
- [7] Melkozerov, A. and Beyer, H.-G. On the analysis of self-adaptive evolution strategies on elliptic model: first results. In *GECCO'10*, pp. 369–376, ACM, 2010.
- [8] Melkozerov, A. and Beyer, H.-G. *On the Derivation of the Progress Rate and Self-Adaptation Response for the $(\mu/\mu_I, \lambda)$ - σ SA-ES on the Noisy Ellipsoid Model* (Technical Report TR 2015/01). Vorarlberg University of Applied Sciences, PPE, Dornbirn, Austria, 2015.
- [9] Meyer-Nieberg, S. *Self-Adaptation in Evolution Strategies*. PhD thesis, University of Dortmund, CS Department, Dortmund, Germany, 2007.

APPENDIX

A brief outline of steps necessary to obtain the progress rate and the SAR formulae is presented in the following. For a detailed treatment it is referred to [8].

A. NOISY PROGRESS RATE

The goal of this section is to determine the $(\mu/\mu_I, \lambda)$ -ES progress rate along the i th axis of the noisy ellipsoid model.

Following the approach introduced in [7], the ES progress along each axis is considered separately. That is, the noisy progress rate along the i th axis of the ellipsoid model (1) is defined as (2) leading to

$$\varphi_i = \frac{1}{\mu} \sum_{m=1}^{\mu} E \left[y_i^{(g)} - \left(\tilde{y}_i^{(g)} \right)_{m;\lambda} | \mathbf{y}^{(g)} \right], \quad (23)$$

where the subscript $m;\lambda$ refers to the m th-best of the λ offspring. To simplify notation, indices (g) are omitted in the following derivations. Introducing the mutation vector $\mathbf{x}_l = \tilde{\sigma}_l \tilde{\mathbf{z}}_l$ (note that σ is fixed and $\mathbf{x}_l = \sigma \tilde{\mathbf{z}}_l$ in the progress rate analysis due to the assumption $\tau \xrightarrow{N \rightarrow \infty} 0$) yields [7]

$$\varphi_i = -\frac{1}{\mu} \sum_{m=1}^{\mu} \int_{-\infty}^{\infty} x p_{m;\lambda}(x|\mathbf{y}) dx. \quad (24)$$

The density of induced order statistics $p_{m;\lambda}(x|\mathbf{y})$ in (24) has been obtained in [7] for $N \rightarrow \infty$ and reads

$$p_{m;\lambda}(x|\mathbf{y}) = \frac{\lambda!}{(m-1)! (\lambda-m)!} p_x(x) \int_{-\infty}^{+\infty} p_Q(q|x, \mathbf{y}) \\ \times P_Q(q|\mathbf{y})^{m-1} [1 - P_Q(q|\mathbf{y})]^{\lambda-m} dq, \quad (25)$$

where $p_Q(q|x, \mathbf{y})$ is the conditional density and $P_Q(q|\mathbf{y})$ is the cumulative distribution function to be determined below. Plugging (25) into (24), changing the order of integration and denoting the inner integral by

$$I_i(q|\mathbf{y}) := \int_{-\infty}^{+\infty} x p_x(x) p_Q(q|x, \mathbf{y}) dx, \quad (26)$$

the progress rate formula (24) reads

$$\varphi_i = -\frac{1}{\mu} \sum_{m=1}^{\mu} \frac{\lambda!}{(m-1)! (\lambda-m)!} \int_{-\infty}^{+\infty} I_i(q|\mathbf{y}) \\ \times P_Q(q|\mathbf{y})^{m-1} [1 - P_Q(q|\mathbf{y})]^{\lambda-m} dq. \quad (27)$$

To determine $P_Q(q|\mathbf{y})$, the normal approximation is used

$$P_Q(q|\mathbf{y}) \simeq \Phi \left(\frac{q - E[Q_{\text{noisy}}(\mathbf{x}, \mathbf{y})]}{D[Q_{\text{noisy}}(\mathbf{x}, \mathbf{y})]} \right), \quad (28)$$

where $E[Q_{\text{noisy}}(\mathbf{x}, \mathbf{y})]$ is the expectation and $D[Q_{\text{noisy}}(\mathbf{x}, \mathbf{y})]$ is the standard deviation of the noisy local quality change (4). Inserting the expansion of $Q_{\mathbf{y}}(\mathbf{x})$ into Eq. (4) yields

$$Q_{\text{noisy}}(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^N a_j (2y_j x_j + x_j^2) + \sigma_{\epsilon} \mathcal{N}(0, 1), \quad (29)$$

where x_j are the components of the mutation vector \mathbf{x} . The expectation of $Q_{\text{noisy}}(\mathbf{x}, \mathbf{y})$ is equal to the expectation of the noise free $Q_{\mathbf{y}}(\mathbf{x})$

$$E[Q_{\text{noisy}}(\mathbf{x}, \mathbf{y})] = \sigma^2 A_0, \quad A_n := \sum_{j \neq n}^N a_j. \quad (30)$$

To determine $D[Q_{\text{noisy}}(\mathbf{x}, \mathbf{y})]$, $Q_{\text{noisy}}(\mathbf{x}, \mathbf{y})$ is written down as a sum of noise-free components and the noisy term $Q_{\text{noisy}} = \sum_{j=1}^N (Q_{\mathbf{y}})_j + \sigma_{\epsilon} \mathcal{N}(0, 1)$, where each component $(Q_{\mathbf{y}})_j$ is defined as $(Q_{\mathbf{y}})_j := a_j (2y_j x_j + x_j^2)$. $D[Q_{\text{noisy}}(\mathbf{x}, \mathbf{y})]$ is calcu-

lated then by means of the variances

$$D^2 [Q_{\text{noisy}}(\mathbf{x}, \mathbf{y})] = \sum_{j=1}^N D^2 [(Q_{\mathbf{y}})_j] + \sigma_\epsilon^2, \quad (31)$$

where $D^2 [(Q_{\mathbf{y}})_j]$ has been obtained in [7] $D^2 [(Q_{\mathbf{y}})_j] = 2a_j^2 \sigma^2 (2y_j^2 + \sigma^2)$. The standard deviation of Q_{noisy} is

$$D [Q_{\text{noisy}}(\mathbf{x}, \mathbf{y})] = \sigma \sqrt{B_0 + \sigma_\epsilon^2 / \sigma^2}, \quad B_n := \sum_{j \neq n} 2a_j^2 (2y_j^2 + \sigma^2). \quad (32)$$

After inserting (30) and (32) into (28) the conditional probability distribution reads

$$P_Q(q|\mathbf{y}) \simeq \Phi \left(\left(q - \sigma^2 A_0 \right) / \left(\sigma \sqrt{B_0 + \sigma_\epsilon^2 / \sigma^2} \right) \right). \quad (33)$$

The conditional density in Eq. (25) is calculated analogously to the noise free case [7]

$$p_Q(q|x, \mathbf{y}) \simeq \frac{1}{\sqrt{2\pi} D [Q_{\text{noisy}}]} \exp \left[-\frac{1}{2} \left(\frac{q - E [Q_{\text{noisy}}]}{D [Q_{\text{noisy}}]} \right)^2 \right], \quad (34)$$

except that the noisy local quality change Q_{noisy} is used. First, the i th summand of (29) is taken out and the substitution $x_i = \sigma z_i$ is used. Next, under the assumption that $|\sigma z_i| \ll |2y_i|$ for $\sigma \rightarrow 0$, a rough $Q_{\text{noisy}}(\mathbf{x}, \mathbf{y})$ approximation is introduced

$$Q_{\text{noisy}}(\mathbf{x}, \mathbf{y}) \approx 2a_i y_i x_i + \sum_{j \neq i}^N a_j (2y_j x_j + x_j^2) + \sigma_\epsilon \mathcal{N}(0, 1). \quad (35)$$

The validity of (35) is verified experimentally. Keeping $x_i = x$ fixed (since this is the condition, $D^2[x] = 0$) yields

$$E [Q_{\text{noisy}}(\mathbf{x}, \mathbf{y}) | x] = 2a_i y_i x + \sigma^2 A_i. \quad (36)$$

and

$$D^2 [Q_{\text{noisy}}(\mathbf{x}, \mathbf{y}) | x] = \sigma^2 \sum_{j \neq i}^N 2a_j^2 (2y_j^2 + \sigma^2) + \sigma_\epsilon^2, \quad (37)$$

which leads to $D [Q_{\text{noisy}}(\mathbf{x}, \mathbf{y}) | x] = \sigma \sqrt{B_i + \sigma_\epsilon^2 / \sigma^2}$. Inserting (36) and $D [Q_{\text{noisy}}(\mathbf{x}, \mathbf{y}) | x]$ into (34) results in

$$\begin{aligned} p_Q(q|x, \mathbf{y}) &\simeq \frac{1}{\sqrt{2\pi} \sigma \sqrt{B_i + \sigma_\epsilon^2 / \sigma^2}} \\ &\times \exp \left[-\frac{1}{2} \left(\frac{q - 2a_i y_i x - \sigma^2 A_i}{\sigma \sqrt{B_i + \sigma_\epsilon^2 / \sigma^2}} \right)^2 \right]. \end{aligned} \quad (38)$$

Inserting (33) and (38) into (26) and substituting $t = x/\sigma$ yields

$$\begin{aligned} I_i(q|\mathbf{y}) &\simeq \frac{1}{2\pi \sqrt{B_i + \sigma_\epsilon^2 / \sigma^2}} \int_{-\infty}^{\infty} t e^{-\frac{1}{2}t^2} \\ &\times \exp \left[-\frac{1}{2} \left(\frac{q - 2y_i a_i \sigma t - \sigma^2 A_i}{\sigma \sqrt{B_i + \sigma_\epsilon^2 / \sigma^2}} \right)^2 \right] dt. \end{aligned} \quad (39)$$

Further an integral formula from [6] is applied to (39) and the result is simplified by neglecting $\sigma^2 a_i$ and σ^2 terms under assumption that the resulting error is negligible for $N \rightarrow \infty$.

Inserting the result and (33) into Eq. (27), setting $s = \frac{q - \sigma^2 A_0}{\sigma \sqrt{B_0 + \sigma_\epsilon^2 / \sigma^2}}$ and rearranging the resulting equation leads to an equation, where the sum is substituted by an integral [6]. Using the substitution $v = 1 - \Phi(t)$ and exchanging the order of integration leads to

$$\varphi_i \simeq \frac{2\sigma y_i a_i}{\sqrt{B_0 + \sigma_\epsilon^2 / \sigma^2}} \frac{\lambda - \mu}{2\pi} \binom{\lambda}{\mu} \int_{t=-\infty}^{t=\infty} e^{-t^2} \times (1 - \Phi(t))^{\lambda - \mu - 1} (\Phi(t))^{\mu - 1} dt. \quad (40)$$

Comparing the resulting integral with (6), the coefficient $c_{\mu/\mu, \lambda} = e_{\mu, \lambda}^{1,0}$ is recognized leading to the final formula (5).

B. NOISY QUADRATIC PROGRESS RATE

The product moments E_1 and E_2 in the noisy quadratic progress (15) are given by [5]

$$E_1 = \sigma^2 E \left[\sum_{l=2}^{\mu} \sum_{k=1}^{l-1} z_{k;\lambda} z_{l;\lambda} |\mathbf{y} \right], \quad E_2 = \sigma^2 E \left[\sum_{m=1}^{\mu} z_{m;\lambda}^2 |\mathbf{y} \right], \quad (41)$$

where the $z_{k;\lambda}$ noisy order statistics correspond to the j -components of the mutation vector $\mathbf{x}_{k;\lambda}$ producing the k th best offspring $\tilde{\mathbf{y}}_{k;\lambda} = \mathbf{y} + \mathbf{x}_{k;\lambda} = \mathbf{y} + \tilde{\sigma}_{k;\lambda} \mathbf{z}_{k;\lambda}$. The k th best offspring is ranked according to its objective function value $\tilde{F}(\tilde{\mathbf{y}}_{k;\lambda})$ which depends on the random vector $\mathbf{z}_{k;\lambda} = \mathcal{N}(\mathbf{0}, \mathbf{I})$. Since the progress rate analysis is performed for small τ ($\tau \rightarrow \infty$), it follows that $\tilde{\mathbf{y}}_{k;\lambda} = \mathbf{y} + \sigma \mathbf{z}_{k;\lambda}$.

To compute E_1 and E_2 , the local quality change is considered first. Its expansion leads to

$$\begin{aligned} Q_{\text{noisy}}(\mathbf{x}, \mathbf{y}) &= 2\sigma a_j y_j (z_j)_{k;\lambda} + 2\sigma \sum_{i \neq j}^N a_i y_i (z_i)_{k;\lambda} \\ &+ \sigma^2 \sum_{i=1}^N a_i (z_i)_{k;\lambda}^2 + \sigma_\epsilon \mathcal{N}(0, 1). \end{aligned} \quad (42)$$

Dividing both sides by $2\sigma a_j y_j$ and introducing the quotient $Q_{\text{noisy}}(\mathbf{x}, \mathbf{y}) / 2\sigma a_j y_j =: v_{k;\lambda}$ yields

$$\begin{aligned} v_{k;\lambda} &= (z_j)_{k;\lambda} + \sum_{i \neq j}^N \frac{a_i y_i}{a_j y_j} (z_i)_{k;\lambda} \\ &+ \frac{\sigma}{2} \sum_{i=1}^N \frac{a_i}{a_j y_j} (z_i)_{k;\lambda}^2 + \frac{\sigma_\epsilon}{2\sigma a_j y_j} \mathcal{N}(0, 1). \end{aligned} \quad (43)$$

Equation (43) is a sum of the random variate $(z_j)_{k;\lambda}$, two sum expressions and a noise term. For $N \rightarrow \infty$, the central limit theorem can be applied to the second and third term in (43) yielding an approximate normal distribution

$$\begin{aligned} v_{k;\lambda} &= (z_j)_{k;\lambda} \\ &+ \mathcal{N} \left(\frac{\sigma \sum_{i=1}^N a_i}{2a_j y_j}, \frac{1}{a_j^2 y_j^2} \left(\sum_{i \neq j}^N a_i^2 y_i^2 + \frac{\sigma^2}{2} \sum_{i=1}^N a_i^2 + \frac{\sigma_\epsilon^2}{4\sigma^2} \right) \right). \end{aligned} \quad (44)$$

The k th random variate $(z_j)_{k;\lambda}$ in Eq. (44) corresponds to the k th best $Q_{\text{noisy}}(\mathbf{x}, \mathbf{y})$ value which is proportional to $v_{k;\lambda}$. Considering the second term in Eq. (44) as a noise term, the variates $(z_j)_{k;\lambda}$ can be identified as noisy order statistics or concomitants of $v_{k;\lambda}$.

The sums of product moments of $(z_j)_{k;\lambda}$ have been calculated in [5]

$$E_1 = \mu(\mu-1) \frac{\sigma^2}{2} \rho^2 e_{\mu,\lambda}^{2,0}, \quad E_2 = \mu\sigma^2 (1 + \rho^2 e_{\mu,\lambda}^{1,1}), \quad (45)$$

where ρ is the correlation coefficient [2]

$$\rho = 1/\sqrt{1+\beta^2}. \quad (46)$$

As follows from Eq. (44), the variance β^2 in (46) is expressed as

$$\beta^2 = \frac{1}{a_j^2 y_j^2} \left(\sum_{i \neq j}^N a_i^2 y_i^2 + \frac{\sigma^2}{2} \sum_{i=1}^N a_i^2 + \frac{\sigma_\epsilon^2}{4\sigma^2} \right). \quad (47)$$

Thus the correlation coefficient ρ for the noisy case reads

$$\rho = |a_j y_j| / \sqrt{\frac{\sigma_\epsilon^2}{4\sigma^2} + \sum_{i=1}^N a_i^2 \left(y_i^2 + \frac{\sigma^2}{2} \right)}. \quad (48)$$

Inserting Eq. (48) into Eqs. (45), the final formulae for the expectations of product moments are obtained

$$E_1 \simeq \mu(\mu-1) \frac{\sigma^2}{2} \frac{a_j^2 y_j^2 e_{\mu,\lambda}^{2,0}}{\frac{\sigma_\epsilon^2}{4\sigma^2} + \sum_{i=1}^N a_i^2 \left(y_i^2 + \frac{\sigma^2}{2} \right)}, \quad (49)$$

$$E_2 \simeq \mu\sigma^2 \left(1 + \frac{a_j^2 y_j^2 e_{\mu,\lambda}^{1,1}}{\frac{\sigma_\epsilon^2}{4\sigma^2} + \sum_{i=1}^N a_i^2 \left(y_i^2 + \frac{\sigma^2}{2} \right)} \right). \quad (50)$$

C. THE NOISY SAR FUNCTION

The derivation of the SAR formula for the $(\mu/\mu_I, \lambda)$ - σ SASES on the noisy ellipsoid model follows the analysis steps for its noise-free counterpart described in [7]. The starting point is the integral representation of the SAR function [7]

$$\psi(\sigma) = \frac{1}{\mu} \sum_{m=1}^{\mu} \int_0^{\infty} \left(\frac{\tilde{\sigma} - \sigma}{\sigma} \right) p_{m;\lambda}(\tilde{\sigma}|\sigma) d\tilde{\sigma}, \quad (51)$$

where $p_{m;\lambda}(\tilde{\sigma}|\sigma)$ is the density of induced order statistics. The $p_{m;\lambda}(\tilde{\sigma}|\sigma)$ formula reads [7]

$$p_{m;\lambda}(\tilde{\sigma}|\sigma) = \frac{\lambda!}{(m-1)!(\lambda-m)!} p_{\sigma}(\tilde{\sigma}|\sigma) \int_{-\infty}^{\infty} p_Q(q|\tilde{\sigma}) \times P_Q(q|\sigma)^{m-1} (1 - P_Q(q|\sigma))^{\lambda-m} dq, \quad (52)$$

where $p_{\sigma}(\tilde{\sigma}|\sigma)$ is the distribution density [6]

$$p_{\sigma}(\tilde{\sigma}|\sigma) = \frac{1}{\sqrt{2\pi}\tau\tilde{\sigma}} \exp \left[-\frac{1}{2} \left(\frac{\ln(\tilde{\sigma}/\sigma)}{\tau} \right)^2 \right]. \quad (53)$$

The conditional probability distribution $P_Q(q|\sigma)$ is calculated using the formula

$$P_Q(q|\sigma) \simeq \int_0^{\infty} \Phi \left(\frac{q - \tilde{\sigma}^2 A_0}{\tilde{\sigma} \sqrt{\tilde{B}_0 + \sigma_\epsilon^2/\tilde{\sigma}^2}} \right) p_{\sigma}(\tilde{\sigma}|\sigma) d\tilde{\sigma}, \quad (54)$$

where $\tilde{B}_0 = \sum_{j=0}^N 2a_j^2 (2y_j^2 + \tilde{\sigma}^2)$. For sufficiently small τ , the integral in (54) can be approximated by [6] $P_Q(q|\sigma) \simeq$

$\Phi \left(\frac{q - \text{E}[\sigma]}{\text{D}[\sigma]} \right)$, where the expectation $\text{E}[\sigma] = \sigma^2 A_0$ and the standard deviation $\text{D}[\sigma] = \sigma \sqrt{B_0 + \sigma_\epsilon^2/\sigma^2}$ have been calculated in Section A. Inserting these results yields

$$P_Q(q|\sigma) \simeq \Phi \left(\frac{q - \sigma^2 A_0}{\sigma \sqrt{B_0 + \sigma_\epsilon^2/\sigma^2}} \right) \quad (55)$$

Approximating $p_Q(q|\tilde{\sigma})$ as normal distribution yields

$$p_Q(q|\tilde{\sigma}) \simeq \frac{1}{\sqrt{2\pi}\text{D}[Q_{\text{noisy}}|\tilde{\sigma}]} \exp \left[-\frac{1}{2} \left(\frac{q - \text{E}[Q_{\text{noisy}}|\tilde{\sigma}]}{\text{D}[Q_{\text{noisy}}|\tilde{\sigma}]} \right)^2 \right]. \quad (56)$$

The expectation $\text{E}[Q_{\text{noisy}}(\mathbf{x}, \mathbf{y})|\tilde{\sigma}]$ and standard deviation $\text{D}[Q_{\text{noisy}}(\mathbf{x}, \mathbf{y})|\tilde{\sigma}]$ can be obtained similarly to results in Section A. Taking into account these results, one obtains

$$p_Q(q|\tilde{\sigma}) \simeq \frac{1}{\sqrt{2\pi}\tilde{\sigma}\sqrt{\tilde{B}_0 + \sigma_\epsilon^2/\tilde{\sigma}^2}} \exp \left[-\frac{1}{2} \left(\frac{q - \tilde{\sigma}^2 A_0}{\tilde{\sigma}\sqrt{\tilde{B}_0 + \sigma_\epsilon^2/\tilde{\sigma}^2}} \right)^2 \right]. \quad (57)$$

Inserting Eqs. (57) and (55) into (52) and Eq. (52) into (51) yields

$$\begin{aligned} \psi &\simeq \frac{1}{\mu} \sum_{m=1}^{\mu} \int_0^{\infty} \left(\frac{\tilde{\sigma} - \sigma}{\sigma} \right) \frac{\lambda!}{(m-1)!(\lambda-m)!} p_{\sigma}(\tilde{\sigma}|\sigma) \\ &\times \frac{\sigma}{\tilde{\sigma}} \sqrt{\frac{B_0 + \sigma_\epsilon^2/\sigma^2}{\tilde{B}_0 + \sigma_\epsilon^2/\tilde{\sigma}^2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{\sigma\sqrt{B_0 + \sigma_\epsilon^2/\sigma^2}s - A_0(\tilde{\sigma}^2 - \sigma^2)}{\tilde{\sigma}\sqrt{\tilde{B}_0 + \sigma_\epsilon^2/\tilde{\sigma}^2}} \right)^2} \\ &\times \Phi(s)^{m-1} (1 - \Phi(s))^{\lambda-m} ds d\tilde{\sigma}, \end{aligned} \quad (58)$$

where the substitution $s = (q - \sigma^2 A_0) / (\sigma\sqrt{B_0 + \sigma_\epsilon^2/\sigma^2})$ has been employed.

The sum in (58) can be expressed by an integral representation of the incomplete regularized beta function. After a further transformation and an exchange of the integration order one ends up with a triple integral

$$\begin{aligned} \psi &\simeq \frac{\lambda - \mu}{\sqrt{2\pi}} \binom{\lambda}{\mu} \int_0^{\infty} \left(\frac{\tilde{\sigma} - \sigma}{\sigma} \right) p_{\sigma}(\tilde{\sigma}|\sigma) \\ &\times \frac{1}{\sqrt{2\pi}} \int_{p=-\infty}^{p=\infty} e^{-\frac{1}{2}p^2} (1 - \Phi(p))^{\lambda - \mu - 1} \Phi(p)^{\mu - 1} \\ &\times \frac{\sigma\sqrt{B_0 + \sigma_\epsilon^2/\sigma^2}}{\tilde{\sigma}\sqrt{\tilde{B}_0 + \sigma_\epsilon^2/\tilde{\sigma}^2}} \int_{s=-\infty}^{s=p} e^{-\frac{1}{2} \left(\frac{\sigma\sqrt{B_0 + \sigma_\epsilon^2/\sigma^2}s - A_0(\tilde{\sigma}^2 - \sigma^2)}{\tilde{\sigma}\sqrt{\tilde{B}_0 + \sigma_\epsilon^2/\tilde{\sigma}^2}} \right)^2} ds dp d\tilde{\sigma}, \end{aligned} \quad (59)$$

where the innermost integral can be expressed by the CDF of the normal distribution. This CDF is expanded into a Taylor series about σ including the linear term. After that step, the remaining integral can be determined using formulae obtained in [9]. Thus, one finally obtains the SAR function (21)

$$\begin{aligned} \psi(\sigma) &\simeq \frac{\tau^2}{2} + e_{\mu,\lambda}^{1,1} \frac{B_0}{B_0 + \sigma_\epsilon^2/\sigma^2} \tau^2 \\ &- c_{\mu/\mu,\lambda} \frac{2A_0\sigma}{\sqrt{B_0 + \sigma_\epsilon^2/\sigma^2}} \tau^2 + \mathcal{O}(\tau^4). \end{aligned} \quad (60)$$