The Dynamics of Self-Adaptive Multi-Recombinant Evolution Strategies on the General Ellipsoid Model

Hans-Georg Beyer and Alexander Melkozerov

***Abstract*—The optimization behavior of the self-adaptation (SA) evolution strategy (ES) with intermediate multi-recombination (the $(\mu/\mu_I,\lambda)$-σSA-ES) using isotropic mutations is investigated on convex-quadratic functions (referred to as ellipsoid model). An asymptotically exact quadratic progress rate formula is derived. This is used to model the dynamical ES system by a set of difference equations. The solutions of this system are used to analytically calculate the optimal learning parameter τ. The theoretical results are compared and validated by comparison with real $(\mu/\mu_I,\lambda)$-σSA-ES runs on two ellipsoid test model cases. The theoretical results clearly indicate that using a model-independent learning parameter τ leads to suboptimal performance of the $(\mu/\mu_I,\lambda)$-σSA-ES on objective functions with changing local condition numbers as often encountered in practical problems with complex fitness landscapes.**

***Index Terms*—Evolution strategy, ellipsoid model, progress rate, self-adaptation**

I. Introduction

Theoretical analyses of Evolution Strategies have a long-standing tradition starting with Rechenberg's early work concerning the $(1+1)$-ES on the sphere model published in [22]. While in the last decade of the 20th century parts of more complex ES algorithms such as (μ,λ)- and $(\mu/\mu,\lambda)$-ES have been analyzed, the treatment of the complete algorithm including σ mutation strength control started with the turn of the century [9]. It was continued by different authors such as Arnold [1], Auger [5], and Jägersküpper [16]. Considering test functions beyond the sphere model was the next step. In [17], Jägersküpper considered the $(1+1)$-ES with 1/5-rule on a subset of positive definite quadratic forms (PDQFs). The complementing analysis of the $(\mu/\mu_I,\lambda)$-ES has been done in [10]. Furthermore, the Cigar as a special PDQF [3] and ridge functions [21], [19] have been analyzed so far. However, unlike the acronym PDQF suggests, the general PDQF case has not been treated so far. Since the level set of this general case defines an ellipsoid in the N-dimensional space, we refer to this kind of test function as *general ellipsoid model*.

The analysis of the dynamics of the $(\mu/\mu_I,\lambda)$-ES on ellipsoid models may be regarded as a milestone on the way to a full analysis of covariance matrix adaptation ES (CMA-ES). While these strategies are currently among the best-performing direct search methods [14], their theoretical analysis is still in its infancy. A full analysis that considers the real CMA-ES [15] or the CMSA-ES [12] requires the analysis of the covariance learning *and* the mutation strength adaptation. This paper provides the solution for the second problem in the case of the self-adaptive mutation control as used in CMSA-ES. A similar analysis concerning the cumulative step-size adaptation would solve the respective problem for the CMA-ES. Besides being a step towards the analysis of CMA-like strategies, the analysis to be presented finalizes the chapter of theoretical analyses regarding the dynamical systems approach on quadratic fitness functions started in the 1990s. Furthermore, the analysis approach extends the standard analysis method by utilizing quadratic progress rate measures.

The paper is organized as follows. First, the $(\mu/\mu_I,\lambda)$-σSA-ES algorithm, the ellipsoid model and previous results on the topic are presented in the remaining parts of the introduction. The quadratic progress rate is introduced and derived in Section II. This new progress measure is the basis for the dynamical systems approach in this paper. In Section III, a system of discrete nonlinear difference equations is derived and solved for the steady-state limit. The obtained solutions are compared with real $(\mu/\mu_I,\lambda)$-σSA-ES experiments. Based on these results, in Section IV the problem regarding the optimal choice of the learning parameter τ is tackled yielding an approximate τ formula. The paper concludes with a discussion of the results and their implications for future work.

A. ES Algorithm

The $(\mu/\mu_I,\lambda)$-σSA-ES algorithm investigated in this work is presented in Fig. 1. The parental mutation strength $\sigma^{(0)}$ and the parental parameter vector, or *parental centroid* $\mathbf{y}^{(0)}$, are initialized in Lines 1 and 2. λ offspring individuals are generated from Line 5 to Line 11 in the following way. For each offspring, the mutation of $\sigma^{(g)}$ is performed in Line 6 using the log-normal operator $\mathrm{e}^{\tau\mathcal{N}_l(0,1)}$, where $\mathcal{N}_l(0,1)$ is a standard normally distributed random scalar. The learning parameter τ in the log-normal operator controls the self-adaptation rate. In Line 7, an isotropic mutation direction is generated by means of a random vector $\boldsymbol{\mathcal{N}}_l(\mathbf{0},\mathbf{I})$ the components of which are standard normal variates. This direction vector is scaled with the individual's mutation strength $\tilde{\sigma}_l$ in Line 8 forming the mutation. The offspring vector $\tilde{\mathbf{y}}_l$ is generated in Line 9 and used in the calculation of the objective function value $\tilde{\mathrm{F}}_l$ in Line 10.

H.-G. Beyer is with the Research Center Process and Product Engineering at the Vorarlberg University of Applied Sciences, Dornbirn, Austria, Email: Hans-Georg.Beyer@fhv.at

A. Melkozerov is with the Department of Television and Control, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russia and the Institute of Neural Information Processing, University of Ulm, Ulm, Germany, Email: ame@tu.tusur.ru

1 $\sigma^{(0)} \leftarrow \sigma_{init}$
2 $\mathbf{y}^{(0)} \leftarrow \mathbf{y}_{init}$
3 $g \leftarrow 0$
4 **do**
5 **for** $l = 1, \ldots, \lambda$ **begin**
6 $\tilde{\sigma}_l \leftarrow \sigma^{(g)} \mathrm{e}^{\tau \mathcal{N}_l(0,1)}$
7 $\mathbf{z}_l \leftarrow \boldsymbol{\mathcal{N}}_l(\mathbf{0}, \mathbf{I})$
8 $\mathbf{x}_l \leftarrow \tilde{\sigma}_l \mathbf{z}_l$
9 $\tilde{\mathbf{y}}_l \leftarrow \mathbf{y}^{(g)} + \mathbf{x}_l$
10 $\tilde{F}_l \leftarrow F(\tilde{\mathbf{y}}_l)$
11 **end**
12 $\tilde{\mathbf{F}}_{sort} \leftarrow \mathrm{sort}\left(\tilde{F}_{1\ldots\lambda}\right)$
13 $\sigma^{(g+1)} \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\sigma}_{m;\lambda}$
14 $\mathbf{y}^{(g+1)} \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\mathbf{y}}_{m;\lambda}$
15 $g \leftarrow g + 1$
16 **until** termination criterion fulfilled

Figure 1: The algorithm of the $(\mu/\mu_I, \lambda)$-σSA-ES

After creation, the λ offspring are ranked according to their $\tilde{F}_l$ values in Line 12. The intermediate recombination of offspring mutation strengths and parameter vectors is performed in Lines 13 and 14 to obtain a new parental mutation strength $\sigma^{(g+1)}$ and a new parental vector $\mathbf{y}^{(g+1)}$. The subscript $m;\lambda$ refers to the mth best of λ offspring (i.e., the offspring with the mth smallest F-value in the case of minimization).

After the termination criterion is fulfilled, the current parental parameter vector is considered as an approximation of the optimizer of the objective function $F(\mathbf{y})$.

B. Fitness Environment

The $(\mu/\mu_I, \lambda)$-σSA-ES analysis in this work is performed for the ellipsoid model

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

where N is the search space dimensionality and a_i are the coefficients of the ellipsoid model. Its optimizer $\hat{\mathbf{y}} = \mathbf{0}$ resides at the origin of coordinates. Special cases of the ellipsoid model (1) include cigar function ($a_1 = 1$, $a_i = \xi$ for $i = 2, \ldots, N$, where $\xi > 1$ is the condition number), a subset of PDQF ($a_i = \xi$ for $i = 1, \ldots, \lfloor N\vartheta \rfloor$ and $a_i = 1$ for $i = \lfloor N\vartheta \rfloor + 1, \ldots, N$, where $\vartheta \in [0, 1]$ is the partition parameter), and the sphere model ($a_i = 1$). Note, the model (1) already represents the *general* case of positive definite quadratic forms for the $(\mu/\mu_I, \lambda)$-σSA-ES. This is due to the isotropy of the mutations used in Line 9: The algorithm is invariant w.r.t. arbitrary rotations of the coordinate system.

Applying the $(\mu/\mu_I, \lambda)$-σSA-ES of Fig. 1 to the objective function (1) results in a dynamic behavior approaching the optimizer at $\mathbf{y} = \mathbf{0}$. Figure 2 shows the dynamics of typical runs considering some squared components of the $\mathbf{y}^{(g)}$ vector. As one can see, starting from an initial $\mathbf{y}^{(0)} = (1, \ldots, 1)$, $\sigma^{(0)} = 1$, the y-component belonging to the largest a_i, i.e. y_N, exhibits the sharpest drop whereas y_1 is only slowly decreasing. Remarkably, after a transient phase, all y^2-curves

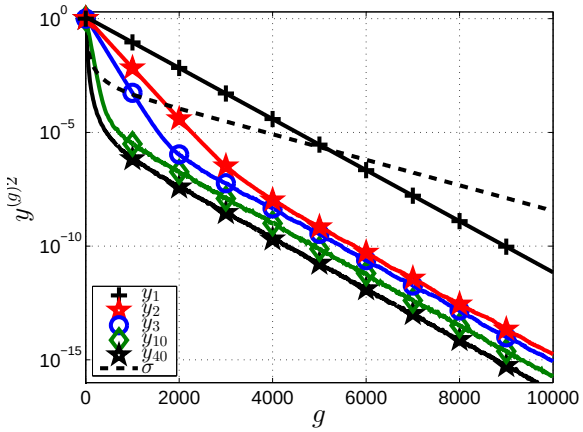


Figure 2: Dynamics of the $(\mu/\mu_I, \lambda)$-σSA-ES on a fitness function (1) with $a_i = i$ and $N = 40$. The quadratic deviation of y_i from the optimizer is displayed for the components $i = 1, 2, 3, 10, 40$. Additionally, the mutation strength σ has been plotted. ES parameters are $\mu = 3$, $\lambda = 10$, $\tau = 1/\sqrt{N}$. Note, the graphs are averages over 1000 independent runs.

exhibit log-linear behavior with the same declination angle. Additionally, the σ-dynamics also approach a log-linear behavior, however, with a different declination rate. The aim of this analysis is to provide formulae that are able to predict this behavior quantitatively. The analysis is based on the dynamical systems approach developed in [9] considering mean value dynamics. As for most of the ES analyses performed, the assumption $N \to \infty$ must be made. Actually, this makes the analysis tractable at all. However, the results obtained can and will be used as approximations for the finite N case, thus providing insights for the real-world case. The analysis to be presented requires an extension of the techniques developed so far: Unlike previous analyses [9], [1], [19], [13] where separate search space dimensions have been lumped together and the objective function has been treated as a function of a single or two state variables, each axis of the ellipsoid model (1) must be considered separately. Due to this distinction, the definition of the measure for the ES progress in the object parameter space of the ellipsoid model – referred to as *progress rate* φ – differs from that of the sphere model. For the parental parameter vector $\mathbf{y}^{(g)} = \left(y_1^{(g)}, y_2^{(g)}, \ldots, y_N^{(g)}\right)^T$ in generation g (the symbol T stands for the transposition of the vector), N progress rates $\varphi_1, \ldots, \varphi_N$ must be calculated. The φ_i formula has been derived previously [18] and is presented in the next section along with other published results.

C. Previous Results

In this section, a summary of results concerning $(\mu/\mu_I, \lambda)$-σSA-ES obtained in [18] are presented.

Definition 1. The *self-adaptation response (SAR) function* of the σSA-ES is the expected relative change of the parental mutation strength from generation g to generation $(g + 1)$

$$\psi\left(\sigma^{(g)}\right) = \mathrm{E}\left[\frac{\sigma^{(g+1)} - \sigma^{(g)}}{\sigma^{(g)}}\right]. \tag{2}$$

Introducing the abbreviation $\Sigma a := \sum_{i=1}^{N} a_i$ and using the mutation strength normalization [18]

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

the SAR function formula for the $(\mu/\mu_I, \lambda)$-σSA-ES on the ellipsoid model reads [18]

$$\psi\left(\sigma^*\right) = \tau^2 \left(\frac{1}{2} + e_{\mu,\lambda}^{1,1} - \sigma^* c_{\mu/\mu,\lambda}\right), \tag{4}$$

where $e_{\mu,\lambda}^{a,b}$ are the generalized progress coefficients [9]

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

$\Phi\left(t\right)$ is the cumulative distribution function of the standard normal variate and the progress coefficient $c_{\mu/\mu,\lambda} := e_{\mu,\lambda}^{1,0}$. Note, surprisingly, Eq. (4) is equivalent to the known SAR function of the $(\mu/\mu_I, \lambda)$-σSA-ES on the Sphere [20] except the different mutation strength normalization.

The second published result of the $(\mu/\mu_I, \lambda)$-σSA-ES analysis on the ellipsoid model is the first-order progress rate:

Definition 2. The progress rate of the $(\mu/\mu_I, \lambda)$-ES along the ith axis of the ellipsoid model (1) is the expected change of the parental parameter vector component y_i from generation g to generation $(g+1)$

$$\varphi_i := \mathrm{E}\left[y_i^{(g)} - y_i^{(g+1)} \mid \mathbf{y}^{(g)}\right]. \tag{6}$$

Note that the progress rate analysis usually neglects the mutation of the mutation strength σ (Line 6 in Fig. 1) since the learning parameter τ is rather small. For example, for the sphere model it was proven in [8] that for optimal ES performance $\tau \propto 1/\sqrt{N}$ must hold. That is, in the asymptotic Al limit $N \to \infty$ the exponential function approaches one in Line 6 of the ES in Fig. 1, thus keeping σ constant.

Taking into account that $y_i^{(g+1)} = \frac{1}{\mu}\sum_{m=1}^{\mu}\left(\tilde{y}_i^{(g)}\right)_{m;\lambda}$ is the mean value of the parameter vector components $\tilde{y}_i^{(g)}$ of the μ best offspring in generation g, Eq. (6) transforms into

$$\varphi_i = \frac{1}{\mu}\sum_{m=1}^{\mu} \mathrm{E}\left[y_i^{(g)} - \left(\tilde{y}_i^{(g)}\right)_{m;\lambda} |\mathbf{y}^{(g)}\right]. \tag{7}$$

Introducing the progress rate normalization [18]

$$\varphi_i^* := \varphi_i \Sigma a \tag{8}$$

the normalized progress rate formula reads [18]

$$\varphi_i^*\left(\sigma^*\right) = \sigma^* c_{\mu/\mu,\lambda} a_i y_i. \tag{9}$$

Eq. (9) is linear in the normalized mutation strength σ^*. This is an approximation for small mutation strengths typically observed in the steady state regime of the evolution process. The progress rate (9) can be used to describe the expected approach to the optimizer for each component of the parental centroid as long as the distance to the optimizer is sufficiently large compared to the respective progress rate values. If this condition is not fulfilled, the mean value dynamics are significantly overlaid by the fluctuations of the evolutionary process. As a result, the predictive quality deteriorates when approaching the optimizer. This is the reason why a new, more stable mean value quantity is needed. It turned out that the appropriate progress measure is the quadratic progress rate which is introduced in the next section.

II. Quadratic Progress Rate φ^{II}

Definition 3. The quadratic progress rate of the $(\mu/\mu_I, \lambda)$-ES along the ith axis of the ellipsoid model (1) is the expected change of the squared component y_i^2 of the parental parameter vector from generation g to generation $(g+1)$

$$\varphi_i^{II} := \mathrm{E}\left[\left(y_i^{(g)}\right)^2 - \left(y_i^{(g+1)}\right)^2 \mid \mathbf{y}^{(g)}\right]. \tag{10}$$

As one will see below, this progress measure shares the typical properties of well-defined progress measures: it contains gain as well as loss terms which depend on the mutation strength. Therefore, there exists an optimal mutation strength maximizing the progress towards the optimizer. Furthermore, it seems a natural measure because it also allows for the direct calculation of the quality gain $\overline{Q}$ [9]. The latter is defined as the expected parental fitness change $\overline{Q} := \mathrm{E}\left[F(\mathbf{y}^{(g+1)}) - F(\mathbf{y}^{(g)})\right]$. Taking (1) and (10) into account, this leads to

$$\overline{Q} = \mathrm{E}\left[\sum_{i=1}^{N} a_i (y_i^{(g+1)})^2 - \sum_{i=1}^{N} a_i (y_i^{(g)})^2\right] = -\sum_{i=1}^{N} a_i \varphi_i^{II}. \tag{11}$$

A. On the Derivation of φ_i^{II}

To derive a formula for φ_i^{II}, the $(\mu/\mu_I, \lambda)$-ES recombination step $\mathbf{y}^{(g+1)} = \frac{1}{\mu}\sum_{m=1}^{\mu}\tilde{\mathbf{y}}_{m;\lambda}$ is considered. The ith component of the parameter vector $\mathbf{y}^{(g+1)}$ is calculated as follows (cf. Lines 9 and 14 in Fig. 1)

$$y_i^{(g+1)} = \frac{1}{\mu}\sum_{m=1}^{\mu}\left(\tilde{y}_i\right)_{m;\lambda} = y_i + \frac{1}{\mu}\sum_{m=1}^{\mu}\left(x_i\right)_{m;\lambda}, \tag{12}$$

where indices (g) are omitted for brevity. Inserting (12) into the φ_i^{II} definition yields

$$\varphi_i^{II} = \mathrm{E}\left[-2y_i\frac{1}{\mu}\sum_{m=1}^{\mu}\left(x_i\right)_{m;\lambda} - \frac{1}{\mu^2}\left(\sum_{m=1}^{\mu}\left(x_i\right)_{m;\lambda}\right)^2 |\mathbf{y}^{(g)}\right], \tag{13}$$

which is further transformed using the equalities $\left(x_i\right)_{m;\lambda} = \left(\tilde{y}_i\right)_{m;\lambda} - y_i$ and $\left(\sum_{m=1}^{\mu} a_{m;\lambda}\right)^2 = 2\sum_{l=2}^{\mu}\sum_{k=1}^{l-1} a_{k;\lambda}a_{l;\lambda} + \sum_{m=1}^{\mu} a_{m;\lambda}^2$ into

$$\begin{aligned}\varphi_i^{II} = &-2y_i\frac{1}{\mu}\sum_{m=1}^{\mu}\mathrm{E}\left[\left(\tilde{y}_i\right)_{m;\lambda} - y_i|\mathbf{y}^{(g)}\right] \\ &- \frac{2}{\mu^2}\mathrm{E}\left[\sum_{l=2}^{\mu}\sum_{k=1}^{l-1}\left(x_i\right)_{k;\lambda}\left(x_i\right)_{l;\lambda}|\mathbf{y}^{(g)}\right] \\ &- \frac{1}{\mu^2}\mathrm{E}\left[\sum_{m=1}^{\mu}\left(x_i\right)_{m;\lambda}^2|\mathbf{y}^{(g)}\right].\end{aligned} \tag{14}$$

Comparing the first line in (14) with Eq. (7), the progress rate φ_i can be recognized. With expectations in the second and third lines denoted by E_1 and E_2, Eq. (14) reads

$$\varphi_i^{II} = 2y_i\varphi_i - \frac{2}{\mu^2}E_1 - \frac{1}{\mu^2}E_2. \tag{15}$$

The sums of product moments E_1 and E_2 are calculated in the Appendix, Eqs. (96) and (100). Inserting those results into (15) leads after normalization using (8) to the normalized quadratic progress rate formula[1]

$$\varphi_i^{II*}(\sigma^*) = 2y_i\varphi_i^*(\sigma^*) - \frac{(\sigma^*)^2}{\mu\Sigma a} \times \left[\sum_{j=1}^{N} a_j^2 y_j^2 + \left((\mu-1)\, e_{\mu,\lambda}^{2,0} + e_{\mu,\lambda}^{1,1}\right) a_i^2 y_i^2\right], \tag{16}$$

where $\varphi_i^*(\sigma^*)$ is given by (9).

The quadratic progress rate formula (16) depends on the first-order φ_i^* as well as on a negative higher-order term which corresponds to the progress rate loss. Due to the coefficient proportional to y_i, the influence of φ_i^* depends on how far from the optimizer ($y_i = 0$ for all i) the ES works. The loss term is proportional to the squared mutation strength. That is, the $(\mu/\mu_I, \lambda)$-ES progress rate grows for small σ^*, reaches a maximum and decreases after that. The loss term is also inversely proportional to the parent number μ. That is, the *genetic repair effect* of recombination [9], first found for the sphere model, does also hold for the ellipsoid: recombining the $\mu > 1$ best offspring reduces the loss part of φ_i^{II*}.

Taking into account the complexity of Eq. (16), a simpler φ_i^{II*} formula will be used for the dynamical analysis to be performed in Section III. Provided that there is not a dominating a_i coefficient, i.e., the condition $\forall i : \sum_{j\neq i} a_j^2 \gg a_i^2$ holds, and $N \gg \mu$, the expression $\left((\mu-1)\, e_{\mu,\lambda}^{2,0} + e_{\mu,\lambda}^{1,1}\right) a_i^2 y_i^2$ can be neglected in the loss term.[2] Thus, taking Eq. (9) into account, one obtains asymptotically

$$\varphi_i^{II*}(\sigma^*) = 2\sigma^* c_{\mu/\mu,\lambda} a_i y_i^2 - \frac{(\sigma^*)^2}{\mu\Sigma a}\sum_{j=1}^{N} a_j^2 y_j^2. \tag{17}$$

The renormalized version of (17), obtained by applying (8) and (3),

$$\boxed{\varphi_i^{II}\left(\sigma^{(g)}\right) = \frac{2\sigma^{(g)} c_{\mu/\mu,\lambda} a_i y_i^{(g)2}}{\sqrt{\sum_{j=1}^{N} a_j^2 y_j^{(g)2}}} - \frac{\left(\sigma^{(g)}\right)^2}{\mu}} \tag{18}$$

will be used to derive the evolution equations of the ES in Section III.

The result (17) is in accordance with former findings including the sphere model quality gain $\overline{Q}_{\mathrm{sp}}$ introduced in [7]. This can be shown easily using (11) together with the normalization (8) in (17) taking $F_{\mathrm{sp}}(\mathbf{y}) = \sum_{i=1}^{N} y_i^2$ (i.e. $a_i = 1$) into account

$$\overline{Q}_{\mathrm{sp}} = -\sum_{i=1}^{N}\varphi_i^{II} = -\frac{1}{N}\sum_{i=1}^{N}\varphi_i^{II*} = -\frac{2F_{\mathrm{sp}}}{N}\varphi_{\mathrm{sp}}^* \tag{19}$$

with

$$\varphi_{\mathrm{sp}}^* = c_{\mu/\mu,\lambda}\sigma^* - \frac{(\sigma^*)^2}{2\mu}. \tag{20}$$

As one can see, this calculation also recovered the normalized progress rate φ_{sp}^* for the sphere model [9].

B. One-Generation Experiments

In this section, ES experiments are performed to check the validity of the progress rate formulae (16) and (17). To gather experimental data, so-called one-generation experiments [9] are used which consist of the following operations:

1) One iteration of the $(\mu/\mu_I, \lambda)$-ES algorithm is executed for a given σ^* value and initial parameter vector $\mathbf{y}^{(0)}$.
2) The newly generated parameter vector $\mathbf{y}^{(1)}$ is registered. Its squared components $\left(y_i^{(1)}\right)^2$ are subtracted from the squared components $\left(y_i^{(0)}\right)^2$ of the initial parameter vector $\mathbf{y}^{(0)}$ resulting in N quadratic progress samples.
3) Steps 1–2 are repeated $g_{\max}$ times, gathered quadratic progress samples are averaged and finally normalized according to (8).

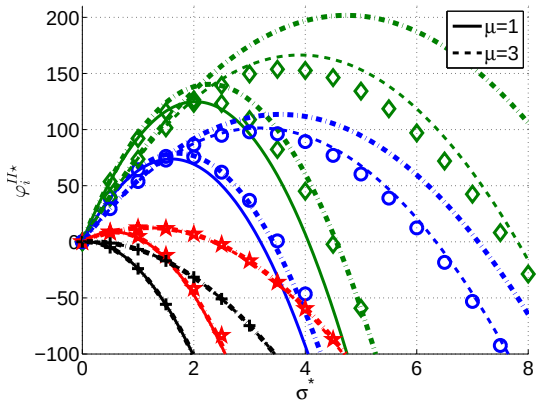
The one-generation experiments produce N experimental φ_i^{II*} values, where each φ_i^{II*} is a normalized mean of $g_{\max}$ randomly generated quadratic progress samples. In order to obtain significant results, $g_{\max} = 10^8$ has been chosen to perform experimental validation of Eqs. (16) and (17). The results of the $(\mu/\mu_I, 10)$-ES one-generation experiments for $a_i = i$ with initial parameter vector $\mathbf{y}^{(0)} = \mathbf{1}$ are shown in Fig. 3 for $\mu = 1$ (solid curves) and $\mu = 3$ (dashed curves). Comparing Figs. 3a and 3b, one can observe that $N = 400$ theoretical curves match the experimental points for larger σ^* values better than in the $N = 40$ case. This is in accordance with the assumptions made in the φ^{II*} derivation: It is to be expected that the approximation error of Eq. (16) vanishes for $N \to \infty$.

Dot-dash curves in Fig. 3 represent the outcome of the simplified formula (17). These curves can be regarded as a satisfactory approximation of the more complex Eq. (16) for sufficiently small σ^* in the $N = 40$ case (Fig. 3a) and for most σ^* considered in the $N = 400$ case (Fig. 3b). Note that Eq. (17) curves reproduce the behavior of Eq. (16) even for $N = 40$: φ_i^{II*} grows until a maximum is reached, then φ_i^{II*} constantly decreases. Thus Eq. (17) can be used instead of Eq. (16) as an upper bound estimate to study the maximal attainable performance of the $(\mu/\mu_I, \lambda)$-ES as well as to select optimal σ^* values.

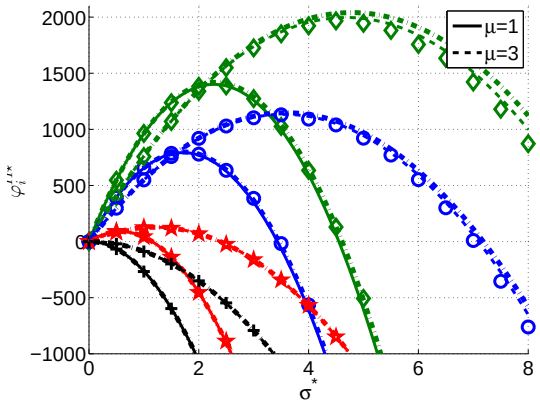
One can further infer from Fig. 3 that the quadratic progress rate results correctly show the effect of the multi-recombination: Since the loss term of Eq. (16) is inversely proportional to μ, the single-parent $(1, 10)$-ES ($\mu = 1$, solid curves) reaches smaller maximal φ_i^{II*} values than the multirecombinant $(3/3_I, 10)$-ES (dashed curves). In contrast

[1] In order to obtain (16) from (15), the $\sigma^2/2$ terms in the denominators of (96) and (100) have been dropped. This is admissible as long as $(\sigma^*)^2\left(\sum_k a_k^2\right)/\left(2\left(\sum_k a_k\right)^2\right) \ll 1$. For the cases $a_i = i, i^2$, this is fulfilled if $(\sigma^*)^2/N \ll 1$, as can be easily checked.

[2] The validity of this assumption also requires that the y_i^2 dynamics behave "nicely". This can be checked by reinserting the final y_i^2 results confirming the consistency of the approach.



(a) $a_i = i$, $N = 40$



(b) $a_i = i$, $N = 400$

Figure 3: One-generation experiments for the $(\mu/\mu_I, 10)$-ES. The solid lines and dashed curves depict theoretical predictions of Eq. (16) for $\mu = 1$ and $\mu = 3$, respectively, while points represent experimental results for $N = 40$ and $N = 400$: $+$ φ_1^{II*}, $\star$ $\varphi_{N/4}^{II*}$, $\circ$ $\varphi_{N/2}^{II*}$ and $\Diamond$ φ_N^{II*}. Dot-dashed curves show the results of the simplified formula (17).

to the first-order progress rate results in [18], where φ_i^* exhibit saturation behavior for $\sigma^* \to \infty$, φ_i^{II*} is in accordance with the known results obtained for the sphere model [9].

III. Evolution Equations

The progress rate (18) and the SAR function (4) describe the expected change between two consecutive generations, i.e., the short-term ES behavior. The aim of this section is to derive analytic formulae which predict the long-term $(\mu/\mu_I, \lambda)$-σSA-ES behavior.

A. Deriving the Evolution Equations

In the framework of the dynamical systems approach [9], the stochastic mapping of the ES state at (g) to that at $(g+1)$ can be described in the case of the general quadratic fitness model by

$$\begin{cases} (y_i^{(g+1)})^2 = (y_i^{(g)})^2 - \varphi_i^{II}\left(\sigma^{(g)}, \mathbf{y}^{(g)}\right) + \epsilon_i\left(\sigma^{(g)}, \mathbf{y}^{(g)}\right), \\ \sigma^{(g+1)} \quad\quad = \sigma^{(g)}\left(1 + \psi\left(\sigma^{(g)}, \mathbf{y}^{(g)}\right)\right) + \epsilon_\sigma\left(\sigma^{(g)}, \mathbf{y}^{(g)}\right). \end{cases} \tag{21}$$

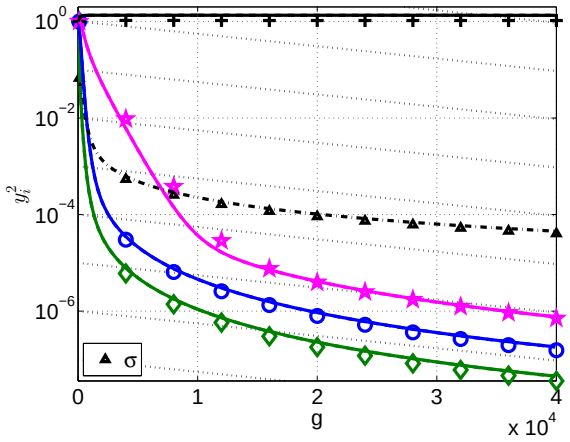


Figure 4: Iterative experiments for the $(3/3_I, 10)$-σSA-ES ($N = 400$, $a_i = i^2$, $\tau = 1/\sqrt{N}$). The solid lines depict predictions of Eqs. (22) and (23), while points represent experimental results averaged over 10^5 runs: $+$ y_1^2, $\star$ $y_{N/4}^2$, $\circ$ $y_{N/2}^2$ and $\Diamond$ y_N^2. Dot-dash curve and $\triangle$ show theoretical and experimental σ, respectively.

This modelling subdivides the stochastic process into mean-value parts and the fluctuation terms ϵ_i and ϵ_σ. The mean-value parts can be directly derived from the definitions of the quadratic progress rate (10) and the SAR function (2). In order to keep the analysis tractable, the fluctuation terms in (21) are disregarded in the following. Using (18) and (4) one obtains the iterative scheme

$$\left(y_i^{(g+1)}\right)^2 = \left(y_i^{(g)}\right)^2 \left(1 - \frac{2\sigma^{(g)} c_{\mu/\mu,\lambda} a_i}{\sqrt{\sum_{j=1}^{N} a_j^2 \left(y_j^{(g)}\right)^2}}\right) + \frac{\left(\sigma^{(g)}\right)^2}{\mu}, \tag{22}$$

$$\sigma^{(g+1)} = \sigma^{(g)} \left[1 + \tau^2 \left(\frac{1}{2} + e_{\mu,\lambda}^{1,1} - \sigma^{(g)} \frac{c_{\mu/\mu,\lambda} \Sigma a}{\sqrt{\sum_{i=1}^{N} a_i^2 \left(y_i^{(g)}\right)^2}}\right)\right]. \tag{23}$$

One can use (22) and (23) to check whether the modelling approach yields meaningful results by iterating the system and comparing with real ES runs. Figure 4 shows a typical example of the $(\mu/\mu_I, \lambda)$-σSA-ES long-term dynamics for $a_i = i^2$ obtained by iterating (22) and (23) starting from $\mathbf{y}^{(0)} = \mathbf{1}$, $\sigma = 1$.

As one can see, there is a good agreement with the data points obtained by running the real ES algorithm. Two phases of the $(\mu/\mu_I, \lambda)$-σSA-ES dynamics can be distinguished in Fig. 4: A transient period after the start of the optimization is followed by a steady state behavior. The transient period is characterized by a rapid decrease of $y_{N/4}^2$, $y_{N/2}^2$, and y_N^2 curves and σ values (the y_1^2 curve decreases as well, albeit at a much smaller rate). In the steady state, $y_{N/4}^2$, $y_{N/2}^2$, and y_N^2 curves diminish slower with the same rate and obey a

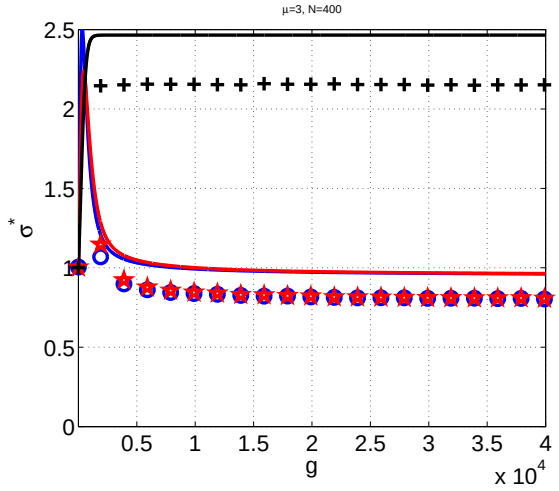


Figure 5: Iterative experiments for the $(3/3_I, 10)$-σSA-ES ($N = 400$). The solid lines depict theoretical predictions of Eqs. (22) and (23), while points represent experimental σ^* values averaged over 10^5 runs: $+\ a_i = 1$, $\star\ a_i = i$ and $\circ\ a_i = i^2$.

log-linear law (compare also Fig. 2).

Both phases are also observed in the normalized σ^* plot, Fig. 5, where the end of the transition phase is clearly visible because σ^* values cease to change. Note that experimental σ^* points for $a_i = i$ (stars) and $a_i = i^2$ (circles) coincide due to the mutation strength normalization. Note, there is a certain deviation of the theoretical results from the experimental ones, more pronounced for the sphere model case. This was also observed for the $(1, \lambda)$-ES in [9] and is to be attributed to (a) basically the neglection of the σ^* fluctuations (only mean value dynamics are considered) and (b) to a certain extend to the approximation error made due to finite search space dimensionality N.

After having motivated qualitatively the validity of the modelling approach, trying to get closed-form solutions to the system (22), (23) appears as a hard task given the fact that this is a system of $N+1$ nonlinear difference equations. Switching to the corresponding differential equations does not improve the situation. Yet, one can draw conclusions from (22), (23) regarding general convergence conditions (also referred to as evolution criteria). Since convergence in expectation necessarily requires $\left(y_i^{(g+1)}\right)^2 \leq \left(y_i^{(g)}\right)^2$, it follows from Eq. (22)

$$\sigma^{(g)} \leq 2\mu c_{\mu/\mu,\lambda} a_i \left(y_i^{(g)}\right)^2 / \sqrt{\textstyle\sum_{j=1}^{N} a_j^2 (y_j^{(g)})^2}. \tag{24}$$

Normalizing (24) using (3), multiplying with a_i and finally taking the sum from 1 to N on both sides of the inequality yields the surprisingly simple convergence criterion

$$\boxed{\sigma^* \leq 2\mu c_{\mu/\mu,\lambda}.} \tag{25}$$

For example, Eq. (25) gives $\sigma^* \leq 6.39$ for $\mu = 3$, $\lambda = 10$. As one can check in Fig. 5, the ES is converging for the given set of parameters. Parenthetically, it is to be mentioned that (25) is identical to the evolution criterion of the $(\mu/\mu_I, \lambda)$-ES on the sphere model (for $N \to \infty$) as one can easily infer from (20) demanding $\varphi^*_{\mathrm{sp}} \geq 0$.

The prediction quality of Eqs. (22), (23) has been investigated in Fig. 4 where a satisfactory agreement between theoretical and experimental $y_i^{(g)}$ results has been shown. After an initial transient phase, y_i^2 curves in Fig. 4 exhibit a log-linear behavior and have the same slopes[3]. Looking at Fig. 2 one also sees that the σ dynamics approaches a log-linear behavior, however, with a different slope. Using this observation, a closed form solution of (22) and (23) in terms of exponential functions for sufficiently large g comes into mind. That is, the system might reach a linear systems behavior in the asymptotic limit ($g \to \infty$). Therefore, the following *Ansatz* is used to solve (22), (23) in the steady state

$$\left(y_i^{(g)}\right)^2 = b_i \mathrm{e}^{-\nu g},\ b_i > 0,\ \nu > 0 \tag{26}$$

$$\sigma^{(g)} = \sigma_0 \mathrm{e}^{-\frac{\nu}{2} g},\ \sigma_0 > 0. \tag{27}$$

This Ansatz takes already the peculiarity of the observed different slopes of σ and y_i^2 correctly into account (cf. Fig. 2). As a consequence, plugging (27) and (26) into the mutation strength normalization formula (3), one obtains a *constant* normalized mutation strength

$$\sigma^* = \sigma_0 \Sigma a / \sqrt{\textstyle\sum_{j=1}^{N} a_j^2 b_j} =: \sigma^*_{\mathrm{ss}}. \tag{28}$$

This σ^* is the normalized steady state mutation strength σ^*_{ss} observed in the right side of Fig. 5.

As a next step, the system (22), (23) will be solved for the steady state using the Ansatz (26), (27). To this end, an eigenvalue problem will be derived in the next section. Special cases will be discussed in subsequent sections.

B. Eigenvalue Problem

The Ansatz (26), (27) allows for a direct connection of the $y_i^{(g+1)}$ and $\sigma^{(g+1)}$ states to those at (g). For example, $(y_i^{(g+1)})^2 = b_i \mathrm{e}^{-\nu g} \mathrm{e}^{-\nu} = (y_i^{(g)})^2 \mathrm{e}^{-\nu}$. As one can infer from the y_i^2 slopes in Figs. 2 and 4, ν is rather small.[4] Therefore, the $\mathrm{e}^{-\nu}$ can be further simplified using Taylor expansion $\mathrm{e}^{-\nu} = 1 - \nu + \mathcal{O}(\nu^2)$. Thus, one obtains for (26) and (27)

$$\left(y_i^{(g+1)}\right)^2 = (1-\nu)\, b_i \mathrm{e}^{-\nu g} + \mathcal{O}\left(\nu^2\right), \tag{29}$$

$$\sigma^{(g+1)} = \left(1 - \frac{\nu}{2}\right) \sigma_0 \mathrm{e}^{-\frac{\nu}{2} g} + \mathcal{O}\left(\nu^2\right). \tag{30}$$

Plugging (29) and (30) into (22) and (23) leads after simplification to

$$\nu b_i = 2\sigma_0 c_{\mu/\mu,\lambda} \frac{a_i}{\sqrt{\sum_{j=1}^{N} a_j^2 b_j}} b_i - \frac{\sigma_0^2}{\mu} + \mathcal{O}\left(\nu^2\right), \tag{31}$$

$$\nu = \tau^2 \left(2\sigma_0 \frac{c_{\mu/\mu,\lambda} \Sigma a}{\sqrt{\sum_{j=1}^{N} a_j^2 b_j}} - 2e^{1,1}_{\mu,\lambda} - 1 \right) + \mathcal{O}\left(\nu^2\right). \tag{32}$$

[3]Note, the transition period for y_1^2 is much longer than the transition period for $y_{N/4}^2$, $y_{N/2}^2$, and y_N^2.

[4]Actually, it decreases with increasing N and $\nu \overset{N\to\infty}{\longrightarrow} 0$.

Substituting σ_0 in (31) and (32) by means of Eq. (28) results in a nonlinear system of $(N+1)$ equations (neglecting higher-order ν terms)

$$\nu b_i = 2\sigma^*_{\rm ss} c_{\mu/\mu,\lambda} \frac{a_i}{\Sigma a} b_i - \frac{\left(\sigma^*_{\rm ss}\right)^2 \sum_{j=1}^{N} a_j^2 b_j}{\mu \left(\Sigma a\right)^2}, \tag{33}$$

$$\nu = \tau^2 \left(2\sigma^*_{\rm ss} c_{\mu/\mu,\lambda} - 2e^{1,1}_{\mu,\lambda} - 1\right), \tag{34}$$

where ν, b_i, and $\sigma^*_{\rm ss}$ are unknowns. Rewriting Eq. (33) in matrix form reveals that this set of equations builds an eigenvalue problem

$$\mathbf{A} \cdot \mathbf{b} = \nu \mathbf{b}, \tag{35}$$

where $\mathbf{b} = \left(b_1, b_2, \ldots, b_N\right)^T$,

$$(\mathbf{A})_{ii} = 2\sigma^*_{\rm ss} c_{\mu/\mu,\lambda} \frac{a_i}{\Sigma a} - \frac{\left(\sigma^*_{\rm ss}\right)^2 a_i^2}{\mu \left(\Sigma a\right)^2}, \tag{36}$$

$$(\mathbf{A})_{ij} = -\left(\sigma^*_{\rm ss}\right)^2 a_j^2 / \left(\mu \left(\Sigma a\right)^2\right),\ i \neq j, \tag{37}$$

and $\sigma^*_{\rm ss} = \text{const}$ is the steady state mutation strength.

Matrix $\mathbf{A}$ in (35) has N eigenvalues ν and N eigenvectors $\mathbf{b}$ of which only the solutions $\forall i:\ b_i > 0$ and $\nu > 0$ are admissible due to the conditions in the Ansatz (26), (27). Moreover, it follows from the Ansatz that larger ν values lead to a faster decay of $\left(y_i^{(g)}\right)^2$ and $\sigma^{(g)}$. That is, for $g \to \infty$ these ν values will have a neglectable impact in comparison to the smallest ν. The second smallest ν determines the rate at which the slowest mode (corresponding to the smallest ν) is reached. Consequently, the reciprocal of the second smallest ν determines the transient time.

Therefore, the smallest positive eigenvalue ν should be found such that the condition $\forall i:\ b_i > 0$ is satisfied by the corresponding eigenvector $\mathbf{b}$. The solution of the eigenvalue problem (35) for the particular case $a_i = 1$ will be presented in the next section and thereafter the ellipsoidal case will be tackled.

C. Solution of the Eigenvalue Problem for the Sphere Model

Since the approach to the ES dynamics presented in this paper is new, it will be first applied to the sphere model $a_i = 1$ in order to a) compare with the classical sphere model results and b) to prepare for the general ellipsoidal case. Equation (33) yields with $a_i = 1$ for the sphere model

$$\left(2\sigma^*_{\rm ss} \frac{c_{\mu/\mu,\lambda}}{N} - \nu\right) b_i - \left(\sigma^*_{\rm ss}\right)^2 \sum_{j=1}^{N} b_j / \left(\mu N^2\right) = 0. \tag{38}$$

Since (38) holds for any i, one can subtract the equation for the kth component of $\mathbf{b}$ from those of the ith component leading to

$$\left(2\sigma^*_{\rm ss} \frac{c_{\mu/\mu,\lambda}}{N} - \nu\right) b_i - \left(2\sigma^*_{\rm ss} \frac{c_{\mu/\mu,\lambda}}{N} - \nu\right) b_k = 0. \tag{39}$$

It follows from (39) that $b_i = b_k = b$. Note that there exist other eigenvectors of (35) for $a_i = 1$, but since these must be orthogonal to the eigenvector $\mathbf{b} = (b, b, \ldots, b)^T$, $b > 0$, under consideration, they necessarily have components $b_i < 0$ and thus do not satisfy the condition $\forall i: b_i > 0$.

With the solution $b_i = b$, Eq. (38) can be solved for the eigenvalue ν yielding

$$\nu\left(\sigma^*_{\rm ss}\right) = \frac{2}{N}\left(c_{\mu/\mu,\lambda}\sigma^*_{\rm ss} - \frac{\left(\sigma^*_{\rm ss}\right)^2}{2\mu}\right). \tag{40}$$

This eigenvalue (40) is proportional to the normalized progress rate (20), $\nu\left(\sigma^*_{\rm ss}\right) = \frac{2}{N}\varphi^*_{\rm sp}\left(\sigma^*_{\rm ss}\right)$, and connects the dynamic quantities with the local performance measures. Its maximum is reached at $\sigma^*_{\rm ss} = \mu c_{\mu/\mu,\lambda} =: \sigma^*_{\rm opt}$ and is equal to

$$\nu_{\max} = \nu\left(\sigma^*_{\rm opt}\right) = \mu c^2_{\mu/\mu,\lambda}/N. \tag{41}$$

Inserting $\sigma^*_{\rm opt}$ and (41) into (34) yields the optimal learning parameter for the sphere model

$$\tau_{\rm opt_{sp}} = \sqrt{\frac{\mu c^2_{\mu/\mu,\lambda}}{2N\left(\mu c^2_{\mu/\mu,\lambda} - e^{1,1}_{\mu,\lambda} - 1/2\right)}}. \tag{42}$$

Eq. (42) agrees with the known $\tau_{\rm opt}$ formula derived in [19] for the $(\mu/\mu_I, \lambda)$-σSA-ES on the sphere model. Therefore, the solution (40) substantiates the appropriateness of the Ansatz (26), (27). Actually, Eq. (40) is in fact another expression of the sphere model steady state condition [20]

$$\varphi^*_{\rm sp}\left(\sigma^*_{\rm ss}\right)/N = -\psi\left(\sigma^*_{\rm ss}\right). \tag{43}$$

Indeed, it follows from comparison of Eqs. (34) and (4) that

$$\nu = -2\psi\left(\sigma^*_{\rm ss}\right). \tag{44}$$

Finally, replacing $\varphi^*_{\rm sp}$, Eq. (20), for the sphere model progress expression in the rhs of Eq. (40) yields the steady state condition (43).

D. Solutions of the Eigenvalue Problem for the Ellipsoid Model

A straightforward approach to the eigenvalue problem (35) for arbitrary a_i is to find its solutions numerically. An example of ν values obtained numerically for the $(3/3_I, 10)$-σSA-ES is shown in Fig. 6 for $N = 40$ considering the three models $a_i = 1, i, i^2$.

Interestingly, as for the cases $a_i = i$ (stars) and $a_i = i^2$ (circles), the numerically obtained data points grow linearly with the normalized mutation strength σ^* over a wide range of σ^* values before they exhibit a sudden sharp drop. This observation paves the way for an analytical calculation of $\nu(\sigma^*)$ for sufficiently small σ^* values and later on for the estimation of the optimal τ parameter.

In order to get the linear part of the $\nu(\sigma^*)$ function one has to neglect the quadratic σ^* terms in (35). Considering (36) and (37) one finds $(\mathbf{A})_{ii} = 2\sigma^*_{\rm ss} c_{\mu/\mu,\lambda} a_i/\Sigma a$ and $\mathbf{A}_{ij} = 0, \forall i \neq j$. As a result, the problem is diagonalized and one can directly read off the eigenvalues $\nu_i = 2\sigma^*_{\rm ss} c_{\mu/\mu,\lambda} a_i/\Sigma a$. Taking into account that the steady state dynamics are governed by the smallest positive eigenvalue, one gets the linear part for that ν that belongs to the smallest a_i

$$\boxed{\nu_{lin}(\sigma^*_{\rm ss}) = 2\sigma^*_{\rm ss} c_{\mu/\mu,\lambda} \min\left(a_i\right) / \sum_{k=1}^{N} a_k.} \tag{45}$$

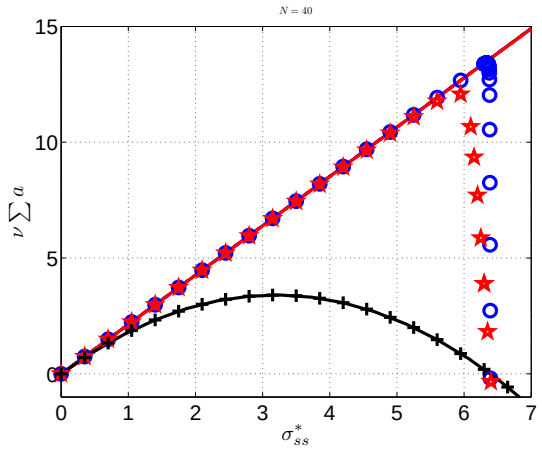


Figure 6: Numerical solutions ν ($N = 40$) of the eigenvalue problem (35) multiplied by $\Sigma a = \sum_{i=1}^{N} a_i$ (points) as a function of the normalized mutation strength compared with analytical solutions obtained using Eqs. (40) for the sphere (parabolic arc, $+$ numerical data) and (45) (straight ascending line and $\star$ $a_i = i$ and $\circ$ $a_i = i^2$ for the numerical data.

While (45) offers an approximation for the steady state mode eigenvalue that agrees well for sufficiently small σ^* values (see Fig. 6), the strengths b_i of the different y_i^2 modes in Ansatz (26) remain to be determined. To this end, eigenvalue perturbation technique will be used noting that (35) can be written in terms of

$$\left(\mathbf{A}_1 + \mathbf{A}_2\right)\mathbf{b}_\beta = \nu_\beta \mathbf{b}_\beta, \tag{46}$$

where $\left(\mathbf{A}_1\right)_{ij} = 0$ for any $i \neq j$ and

$$\left(\mathbf{A}_1\right)_{ii} = 2\sigma^*_{\mathrm{ss}} c_{\mu/\mu,\lambda}\frac{a_i}{\Sigma a}, \quad \left(\mathbf{A}_2\right)_{ij} = -\frac{\left(\sigma^*_{\mathrm{ss}}\right)^2 a_j^2}{\mu(\Sigma a)^2}. \tag{47}$$

Since $\mathbf{A}_1$ is diagonal dominating compared to $\mathbf{A}_2$, the solution to the subproblem

$$\mathbf{A}_1\mathbf{h}_\beta = \gamma_\beta \mathbf{h}_\beta \tag{48}$$

can be used for the eigenvalue perturbation. One immediately obtains for the solution of the eigenvalue problem (48)

$$\gamma_\beta = 2\sigma^*_{\mathrm{ss}} c_{\mu/\mu,\lambda}\frac{a_\beta}{\Sigma a} = \left(\mathbf{A}_1\right)_{\beta\beta}, \tag{49}$$

$$\mathbf{h}_\beta = \mathbf{e}_\beta = (0,\ldots,1,\ldots,0)^T \tag{50}$$

where $\mathbf{e}_\beta$ is the unit vector with 1 at position β. An approximate solution to (46) can be constructed by adding perturbations δ_β and $\mathbf{d}_\beta$ to the solution (49), (50), i.e.,

$$\nu_\beta = \gamma_\beta + \delta_\beta, \tag{51}$$

$$\mathbf{b}_\beta = \mathbf{h}_\beta + \mathbf{d}_\beta. \tag{52}$$

Inserting (52) and (51) into (46) yields

$$\mathbf{A}_1\mathbf{h}_\beta + \mathbf{A}_2\mathbf{h}_\beta + \mathbf{A}_1\mathbf{d}_\beta + \mathbf{A}_2\mathbf{d}_\beta = \gamma_\beta\mathbf{h}_\beta + \delta_\beta\mathbf{h}_\beta + \gamma_\beta\mathbf{d}_\beta + \delta_\beta\mathbf{d}_\beta. \tag{53}$$

Taking (48) into account and assuming that $\mathbf{A}_2\mathbf{d}_\beta$ and $\delta_\beta\mathbf{d}_\beta$ are small compared to the other terms, Eq. (53) simplifies to

$$\mathbf{A}_1\mathbf{d}_\beta + \mathbf{A}_2\mathbf{h}_\beta = \gamma_\beta\mathbf{d}_\beta + \delta_\beta\mathbf{h}_\beta + \mathcal{O}\left(\left(\sigma^*_{\mathrm{ss}}\right)^2\right). \tag{54}$$

It follows further from the solutions (49) and (50) that

$$\left(\mathbf{A}_1\right)_{ii}\left(\mathbf{d}_\beta\right)_i + \left(\mathbf{A}_2\right)_{i\beta} = \left(\mathbf{A}_1\right)_{\beta\beta}\left(\mathbf{d}_\beta\right)_i + \delta_\beta\delta_{i\beta} + \mathcal{O}\left(\left(\sigma^*_{\mathrm{ss}}\right)^2\right), \tag{55}$$

where $\delta_{i\beta} = \begin{cases} 1, & i = \beta \\ 0, & i \neq \beta \end{cases}$ is Kronecker's delta. Setting $i = \beta$ in Eq. (55) leads to $\delta_\beta = \left(\mathbf{A}_2\right)_{\beta\beta} + \mathcal{O}\left(\left(\sigma^*_{\mathrm{ss}}\right)^2\right)$. This equation yields with Eqs. (51) and (49) the eigenvalue formula ν_β

$$\begin{aligned}\nu_\beta &= \left(\mathbf{A}_1\right)_{\beta\beta} + \left(\mathbf{A}_2\right)_{\beta\beta} + \mathcal{O}\left(\left(\sigma^*_{\mathrm{ss}}\right)^2\right) \\ &= 2\sigma^*_{\mathrm{ss}} c_{\mu/\mu,\lambda}\frac{a_\beta}{\Sigma a} - \frac{\left(\sigma^*_{\mathrm{ss}}\right)^2 a_\beta^2}{\mu(\Sigma a)^2} + \mathcal{O}\left(\left(\sigma^*_{\mathrm{ss}}\right)^2\right).\end{aligned} \tag{56}$$

Having obtained an analytical approximation for the eigenvalues, one can calculate the corresponding eigenvectors. As have been explained, the solution to the smallest eigenvalue is of interest for the steady state behavior of the ES. The eigenvector $\mathbf{b} = (b_1,\ldots,b_N)^T$ is determined up to a scalar factor. That is, it suffices to consider the b_k/b_i ratio depending on ν_β. To this end, Eq. (33) is used in its original form for b_i and in a second form replacing b_i by b_k. Subtracting both equations from each other and resolving for the b_k/b_i ratio yields

$$\frac{b_k}{b_i} = \frac{2\sigma^*_{\mathrm{ss}} c_{\mu/\mu,\lambda} a_i - \nu_\beta \Sigma a}{2\sigma^*_{\mathrm{ss}} c_{\mu/\mu,\lambda} a_k - \nu_\beta \Sigma a}. \tag{57}$$

Since $\forall k : b_k > 0$, Eq. (57) should be positive for all k, ν_β. This requirement is satisfied only if ν_β is chosen as small as possible. As $\sigma^*_{\mathrm{ss}} = \mathrm{const}$ in Eq. (56), the smallest eigenvalue ν_β is determined by the smallest a_β value. That is, $a_\beta = \min(a_i) =: \breve{a}$ yields the steady state eigenvalue approximation[5]

$$\nu\left(\sigma^*_{\mathrm{ss}}\right) = \frac{2}{\Sigma a}\left(\breve{a} c_{\mu/\mu,\lambda}\sigma^*_{\mathrm{ss}} - \frac{\left(\sigma^*_{\mathrm{ss}}\right)^2 \breve{a}^2}{2\mu\Sigma a}\right). \tag{58}$$

Inserting Eq. (58) into (57) yields the steady state b_k/b_i ratio

$$\frac{b_k}{b_i} = \frac{a_i - \breve{a} + \frac{\sigma^*_{\mathrm{ss}}\breve{a}^2}{2\mu c_{\mu/\mu,\lambda}\Sigma a}}{a_k - \breve{a} + \frac{\sigma^*_{\mathrm{ss}}\breve{a}^2}{2\mu c_{\mu/\mu,\lambda}\Sigma a}}. \tag{59}$$

Specifying $i = 1$ and $\breve{a} = 1$ in (57), one obtains an approximation for the b_k values of the steady state eigenvector

$$\boxed{b_k = b_1 \frac{\frac{\sigma^*_{\mathrm{ss}} a_1^2}{2\mu c_{\mu/\mu,\lambda}\Sigma a}}{a_k - a_1\left(1 - \frac{\sigma^*_{\mathrm{ss}} a_1}{2\mu c_{\mu/\mu,\lambda}\Sigma a}\right)}.} \tag{60}$$

Using (58) and (60), the (small σ^*) approximation for the example case $a_i = i$ ($\breve{a} = 1$) reads

$$\nu\left(\sigma^*_{\mathrm{ss}}\right) = \frac{4}{N(N+1)}\left(c_{\mu/\mu,\lambda}\sigma^*_{\mathrm{ss}} - \frac{\left(\sigma^*_{\mathrm{ss}}\right)^2}{\mu N(N+1)}\right), \tag{61}$$

$$b_k = b_1 \frac{\sigma^*_{\mathrm{ss}}}{\sigma^*_{\mathrm{ss}} + \mu c_{\mu/\mu,\lambda}(a_k - 1)N(N+1)}. \tag{62}$$

[5]Note that the term with negative sign in (58) cannot be neglected, otherwise the numerator of Eq. (57) would be equal to zero for the case $k : a_k = \breve{a}$.

As the length of the eigenvector $\mathbf{b}$ can be chosen arbitrarily, Eq. (62) completely describes $\mathbf{b}$. For $a_i = i^2$, one obtains

$$\nu\left(\sigma_{\mathrm{ss}}^*\right) = \frac{12\left(c_{\mu/\mu,\lambda}\sigma_{\mathrm{ss}}^* - \frac{3(\sigma_{\mathrm{ss}}^*)^2}{\mu N(N+1)(2N+1)}\right)}{N\left(N+1\right)\left(2N+1\right)}, \tag{63}$$

$$b_k = b_1 \frac{3\sigma_{\mathrm{ss}}^*}{3\sigma_{\mathrm{ss}}^* + \mu c_{\mu/\mu,\lambda}\left(a_k - 1\right) N\left(N+1\right)\left(2N+1\right)}. \tag{64}$$

Equations (61) and (63) can be further simplified by neglecting the $\left(\sigma_{\mathrm{ss}}^*\right)^2$ term assuming that σ_{ss}^* is sufficiently small. This linear approximation has already been obtained in Eq. (45). It reads for the two cases $a_i = i$ and $a_i = i^2$

$$a_i = i \;\; : \;\; \nu_{lin}\left(\sigma_{\mathrm{ss}}^*\right) = 4\sigma_{\mathrm{ss}}^* \frac{c_{\mu/\mu,\lambda}}{N\left(N+1\right)} \tag{65}$$

$$a_i = i^2 : \; \nu_{lin}\left(\sigma_{\mathrm{ss}}^*\right) = 12\sigma_{\mathrm{ss}}^* \frac{c_{\mu/\mu,\lambda}}{N\left(N+1\right)\left(2N+1\right)}. \tag{66}$$

To check the correctness of the analytic solutions of the eigenvalue problem (35), its numerical solution for the $(3/3_I, 10)$-σSA-ES is compared with the results of Eqs. (40) and (58) in the next section.

E. Experiments and Discussion

Figure 6 presents the comparison of the numerical and analytical solutions of the eigenvalue problem (35) for different normalized mutation strength σ_{ss}^* values. As expected, the exact solution (40) for the sphere model (parabolic arc in Fig. 6)[6] coincides with the numerically calculated solutions (crosses) for all σ_{ss}^* considered.

The general ν approximation (58) yields predictions ($a_i = i$: "$\star$" and $a_i = i^2$: "$\circ$" in Fig. 6) which coincide with the predictions of the linear approximations (65) and (66). These predictions describe the real behavior of the ES quite well as long as the mutation strength is not too large. If σ_{ss}^* gets larger, one observes a sharp drop of the real ν values and ν finally changes its sign at $\sigma_{\mathrm{ss}}^* \approx 6.4$. That is, the $(3/3_I, 10)$-σSA-ES does not converge anymore. This is in full accordance with the evolution criterion (25).

Analogously to ν, the correctness of the eigenvectors formula (60) is checked in Fig. 7 by comparison with the numerical solution of (35) for the $(3/3_I, 10)$-σSA-ES ($b_1 = 1$, $a_i = i$). In Fig. 7 ($N = 40$), theoretical curves for b_2, $b_{N/2}$, b_N are close to the numerically obtained points (stars, circles, and diamonds) for sufficiently small $\sigma_{\mathrm{ss}}^* < 2$, while the trivial case $b_1 = 1$ (crosses) coincides with the numerical solution up to $\sigma_{\mathrm{ss}}^* = 6$. For the $\sigma_{\mathrm{ss}}^* > 6.4$ region, where there is no convergence and $\nu < 0$, the behavior of numerical solutions qualitatively changes for all b_i. The same behavior can be observed for $N = 400$ (not displayed due to space restrictions). The question is whether these deviations are relevant for the real $(\mu/\mu_I, \lambda)$-σSA-ES. To answer this question, the actual steady state σ_{ss}^* realized by the ES must be calculated.

[6]Note that Eq. (58) describes the behavior of the sphere model for small σ_{ss}^* values only.

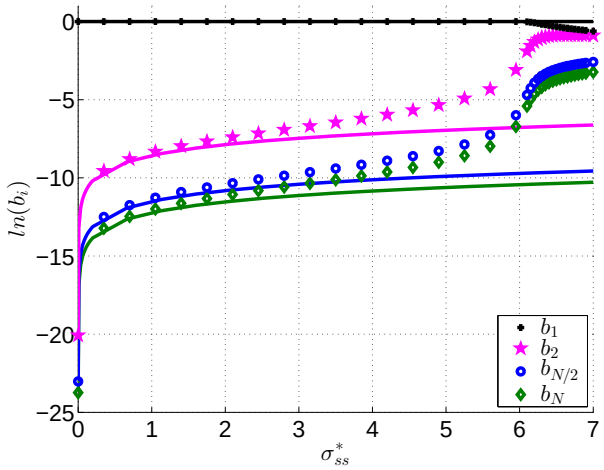


Figure 7: Numerical b_i solutions (points) as a function of σ_{ss}^* compared with analytical solutions Eq. (60) (solid curves) for $N = 40$: $+$ b_1, $\star$ b_2, $\circ$ $b_{N/2}$ and $\Diamond$ b_N ($b_1 = 1$, $a_i = i$).

To this end, the linear approximation (45) with $\breve{a} = \min(a_i)$ is inserted into (34). After rearrangement one obtains

$$\boxed{\sigma_{\mathrm{ss}}^* = \frac{1/2 + e_{\mu,\lambda}^{1,1}}{c_{\mu/\mu,\lambda}} \cdot \frac{1}{1 - \breve{a}/\left(\tau^2 \Sigma a\right)}.} \tag{67}$$

Taking experimental settings $a_i = i^2$ and $\tau = 1/\sqrt{N}$, used to produce Fig. 4, one obtains

$$\sigma_{\mathrm{ss}}^*|_{a_i=i} \approx \frac{1/2 + e_{\mu,\lambda}^{1,1}}{c_{\mu/\mu,\lambda}}. \tag{68}$$

Analogously, Eq. (67) leads for $a_i = i$ to (68) under assumption that $N \to \infty$. For the $(3/3_I, 10)$-σSA-ES considered in Fig. 7, $\sigma_{\mathrm{ss}}^* = 0.95$. b_i values calculated using Eq. (60) for $\sigma_{\mathrm{ss}}^* = 0.95$ agree comparatively well with the numerical solutions of (35) in Fig. 7. Therefore, Eq. (60) can be compared with the $(\mu/\mu_I, \lambda)$-σSA-ES experimental runs.

In order to obtain the experimental eigenvector components b_i, the $(3/3_I, 10)$-σSA-ES with the same settings as in Fig. 4 has been run for 10^5 generations for $N = 40$ and 10^9 generations for $N = 400$. The y_i^2 values of the last 25% of generations have been averaged over 10^5 independent runs. After that, a linear polynomial $\ln y_i^2 = -\nu g + \ln b_i$ has been fitted to the experimental y_i^2 data yielding b_i which are compared in Fig. 8 with the predictions of Eq. (60) for $a_i = i^2$ (σ_{ss}^* is given by Eq. (68)).

The experimental points in Fig. 8 ($a_i = i^2$) are located in the vicinity of the theoretical curves depicting the results of Eq. (60) both for $N = 40$ ("$+$") and $N = 400$ ("$\circ$"). The same observation is valid for $a_i = i$ (not shown due to space restrictions).

The analytic solution of the eigenvalue problem (35) for ν – the eigenvalue formula (58) – should be verified experimentally as well. To this end, the $(3/3_I, 10)$-σSA-ES with the same settings as used in Fig. 8 has been run for fixed σ_{ss}^* (σ_{ss}^* has been renormalized to $\sigma^{(g)}$ in each generation). The gathered data points have been used to obtain N experimental ν values – one for each y_i^2 curve. The ν values corresponding

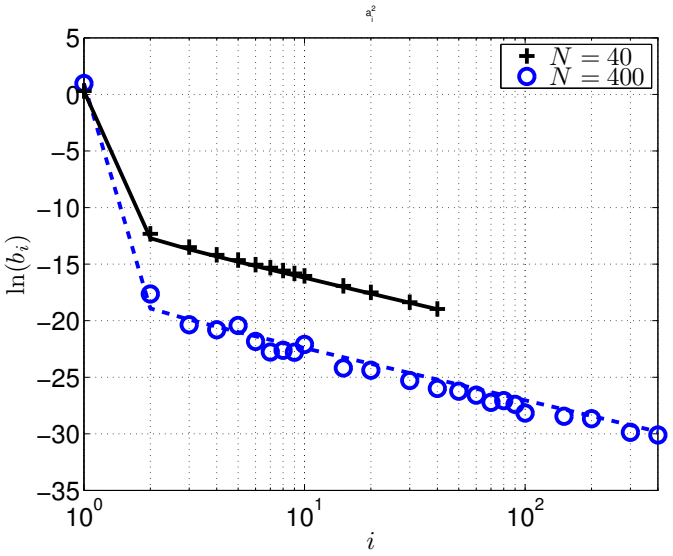


Figure 8: Experimental b_i values (points) for the $(3/3_I, 10)$-σSA-ES (points) compared with analytical solution Eq. (60) (curves) for $a_i = i^2$: $+$ $N = 40$, $\circ$ $N = 400$.

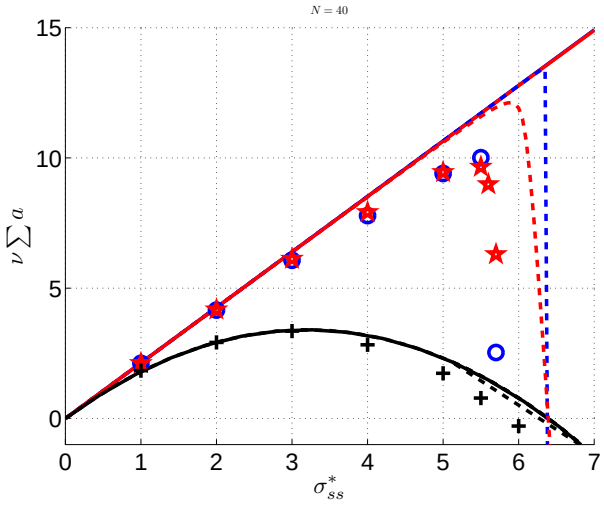
to $\left(y_1^{(g)}\right)^2$ are plotted in Fig. 9 because the deviations of ν for other y_i^2 from the data shown in Fig. 9 are small. For example, the maximal deviation for $a_i = i$ and $N = 40$ is 2%.

In Fig. 9a, the experimental ν values (points) are compared with the outcome of Eq. (58) (solid lines) for $N = 40$. The theoretical predictions match the experimental points for small values of $\sigma^*_{\rm ss} < 4$ for $a_i = 1$ (black crosses) and $\sigma^*_{\rm ss} < 3$ for $a_i = i, i^2$ ("$\star$" and "$\circ$", respectively). The reason is that Eq. (58) is the solution of the eigenvalue problem (35) based on the system of equations (22). Equation (22) contains the asymptotically exact quadratic progress rate formula (18) which is an approximation for $N < \infty$. The quality of (18) decreases with increasing σ^*. Since the quality of (18) increases for larger N, the experimental points in Fig. 9b ($N = 400$) match the theoretical curves for $\sigma^*_{\rm ss} < 6$ for $a_i = 1$ and $\sigma^*_{\rm ss} < 5$ for $a_i = i, i^2$. Thus, the analytical solution (58) is increasingly correct in the limit $N \to \infty$.

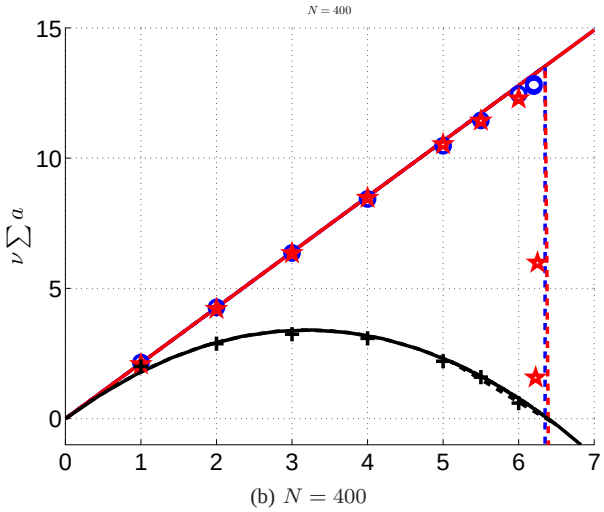
For comparison, Fig. 9 includes the numerical solutions ν (dashed curves) which move for $N = 400$ closer to the solid lines. They show that the errors due to assumptions in the derivation of Eq. (58) diminish for $N \to \infty$. Still, Eq. (58) reproduces the linear part of the numerical ν curve only. Consequently, it can not yield a formula for the optimal $\sigma^*_{\rm ss}$ value which is required to obtain an analytical expression for the optimal learning parameter $\tau_{\rm opt}$.

To illustrate this problem, real $(3/3_I, 10)$-σSA-ES runs with different τ values have been performed for 10^6 generations and $N = 40, 400$ in order to obtain an experimental $\nu = f(\tau)$ data set. This set is compared with the analytical solution (58) in Fig. 10.

The solid curves represent the analytical solution (58). These curves for $N = 40, 400$ coincide due to the normalization of the abscissa. They reproduce the behavior of the experimental points for $\tau > 1.5/N$ well. For smaller τ values, the (58) curves go to infinity. That is, there is not an optimum point which could be used to determine the ν maximum. Note



(a) $N = 40$



(b) $N = 400$

Figure 9: Experimental ν values (points) for the $(3/3_I, 10)$-σSA-ES compared with analytical solution (58) (solid lines) for fixed $\sigma^*_{\rm ss}$: $+$ $a_i = 1$, $\star$ $a_i = i$ and $\circ$ $a_i = i^2$. Dashed curves depict numerical solutions of the eigenvalue problem (35).

that the numerical ν solutions of the eigenvalue problem (35) (depicted by the dashed curves which also coincide due to the normalization) have a maximum and decrease to zero for $\tau \to 0$ showing that Eq. (35) correctly represents the behavior of the $(\mu/\mu_I, \lambda)$-σSA-ES on the ellipsoid model. To bracket the τ optimum analytically, an alternative method will be developed in the next section.

IV. Optimal Learning Parameter

Looking at the $\mu = 1$ and $\mu = 3$ curves in Fig. 3, one observes that the σ^* value leading to the maximal quadratic progress rate grows with μ. However, Eq. (68) yields $\sigma^*|_{a_i=i} \approx 1$ for $\tau = 1/\sqrt{N}$ in the limit $N \to \infty$. While it has been shown that the choice of $\tau = 1/\sqrt{2N}$ is asymptotically optimal for the sphere model [19], using $\tau \propto 1/\sqrt{N}$ can seriously hinder the $(\mu/\mu_I, \lambda)$-σSA-ES performance on non-spherical problems. Furthermore, as Fig. 10 suggests, the

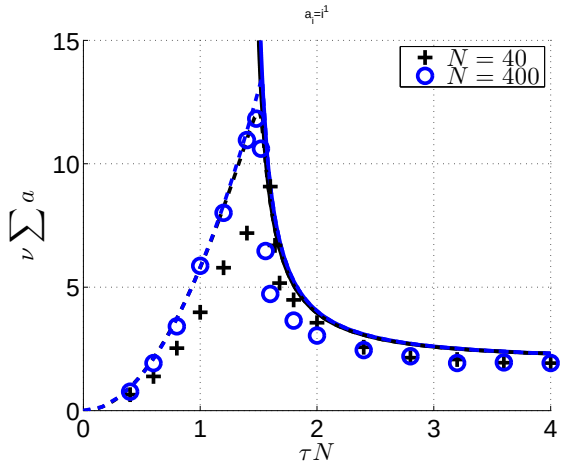


Figure 10: Experimental ν values (points) for the $(3/3_I, 10)$-σSA-ES ($a_i = i$) as a function of τN compared with the analytical solution (58) (solid lines) for $+\ N = 40$ and $\circ\ N = 400$. Dashed curves depict numerical solutions of the eigenvalue problem (35).

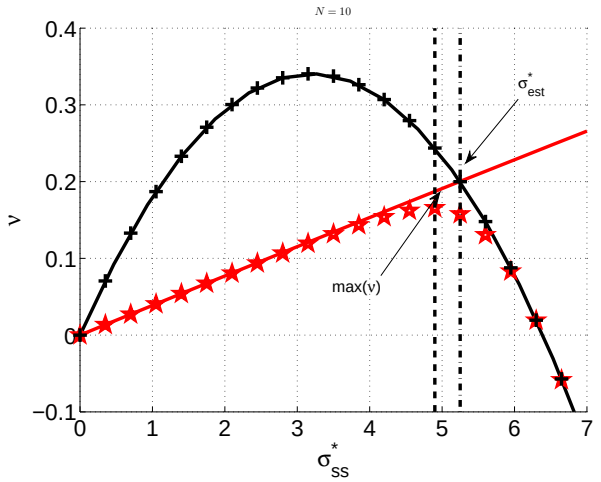


Figure 11: Numerical (points) and analytical (solid curves) solutions ν for the $(3/3_I, 10)$-σSA-ES as a function of σ^*_{ss} ($N = 10$): $+\ a_i = 1$ and $\star\ a_i = i$.

maximal ν can sensitively depend on the learning parameter τ.

Choosing τ correctly leads to the highest possible convergence rate, i.e., the negative slope of the progress rate determined by ν in the exponent of Eq. (26) is maximized for the optimal τ_{opt}. As shown in Fig. 11, Eq. (58) correctly predicts the linear region of ν only and can not be used to determine the maximizer $\hat{\sigma}^* : \nu(\hat{\sigma}^*) = \max(\nu(\sigma^*))$ (indicated by the arrow $\max(\nu)$ in Fig. 11) in the general case of arbitrary a_i.

As a workaround, an upper bound heuristic estimate $\hat{\sigma}^*_{\mathrm{est}} > \hat{\sigma}^*$ is calculated in the following by looking at the intersection of the sphere model curve (black curve in Fig. 11) with ellipsoid model line (indicated by the arrow σ^*_{est} in Fig. 11). This approach is justified by the observation that the $\nu(\sigma^*)$ curves of non-spherical models are below that of the sphere model (40). Actually, according to (40) the slope of the sphere model ν is $2c_{\mu/\mu,\lambda}/N$ at $\sigma^* = 0$ (recall $a_i = 1$) while that of the ellipsoid model is given by (45) $2c_{\mu/\mu,\lambda}/\Sigma a$. Therefore, considering the cases $a_i = i, i^2$, it always holds $1/N > \breve{a}/\Sigma a$ for $N > 1$. That is, the real ν curve must be below of that of the linear approximation (45). Furthermore, according to (25) all ν curves (including the sphere model case) must pass the horizontal axis at $\sigma^* = 2\mu c_{\mu/\mu,\lambda}$. Looking again at Fig. 11, it becomes clear that the intersection of the linear ν curve with that of the sphere model (black curve) yields an estimate $\hat{\sigma}^*_{\mathrm{est}}$ for the optimal σ^*. Actually, since the linear slope of the non-spherical model drops faster than that of the sphere model for increasing N, this estimate improves with increasing N. Numerical investigations considering the relative error $|\nu(\hat{\sigma}^*) - \nu(\sigma^*_{\mathrm{est}})|/\nu(\hat{\sigma}^*)$ using (72) as estimate also confirm this statement (not shown here).

A. Approximate τ_{opt} Formulae

In this section, $\hat{\sigma}^*_{\mathrm{est}}$ is calculated. First, Eq. (45) is equated to the ν formula of the sphere model (40) to calculate the intersection point

$$\hat{\sigma}^*_{\mathrm{est}} = 2c_{\mu/\mu,\lambda}\mu\left(1 - N\breve{a}/\Sigma a\right). \tag{69}$$

Since $\hat{\sigma}^*_{\mathrm{est}}$ is an upper bound approximation of $\hat{\sigma}^*$, they differ from each other by an unknown positive value $(\hat{\sigma}^*_{\mathrm{est}} - \hat{\sigma}^*)$. To account for this difference, a coefficient $0 < \alpha_\sigma < 1$ is introduced such that

$$\hat{\sigma}^* = \alpha_\sigma \hat{\sigma}^*_{\mathrm{est}} \tag{70}$$

and $(\hat{\sigma}^*_{\mathrm{est}} - \hat{\sigma}^*) = (1 - \alpha_\sigma)\,\hat{\sigma}^*_{\mathrm{est}}$. With (70), Eq. (69) transforms into an exact formula for *non-spherical* models (keeping in mind that α_σ is close to 1)

$$\boxed{\hat{\sigma}^* = 2\alpha_\sigma c_{\mu/\mu,\lambda}\mu\left(1 - N\breve{a}/\Sigma a\right).} \tag{71}$$

One obtains the corresponding ν value by inserting (71) into (45) (note, $\min(a_i) = \breve{a}$)

$$\boxed{\nu\left(\hat{\sigma}^*\right) = \frac{4\alpha_\sigma c^2_{\mu/\mu,\lambda}\mu\breve{a}}{\Sigma a}\left(1 - \frac{N\breve{a}}{\Sigma a}\right).} \tag{72}$$

Finally, inserting Eqs. (71) and (72) into (34) yields

$$\boxed{\tau_{\mathrm{opt}} = \sqrt{\frac{\breve{a}}{\Sigma a}\cdot\frac{1}{1 - \frac{1+2e^{1,1}_{\mu,\lambda}}{4\alpha_\sigma c^2_{\mu/\mu,\lambda}\mu(1-N\breve{a}/\Sigma a)}}}.} \tag{73}$$

To apply Eq. (73), α_σ must be chosen. Having a look at Fig. 11, $\alpha_\sigma = 1$ seems to be a reasonable choice. It can be additionally verified by comparison with a known τ_{opt} formula for the special PDQF model [10] ($\breve{a} = 1$ and $\Sigma a = N\left(\vartheta\left(\xi - 1\right) + 1\right)$). In the limit $N \to \infty$, $\xi \gg 1$ and large μ, Eq. (73) simplifies to

$$\tau_{\mathrm{opt}}|_{\mathrm{PDQF}} \simeq \frac{1}{\sqrt{N}}\frac{1}{\sqrt{\vartheta\left(\xi - 1\right)}} \tag{74}$$

which matches the τ_{opt} formula obtained in [10].

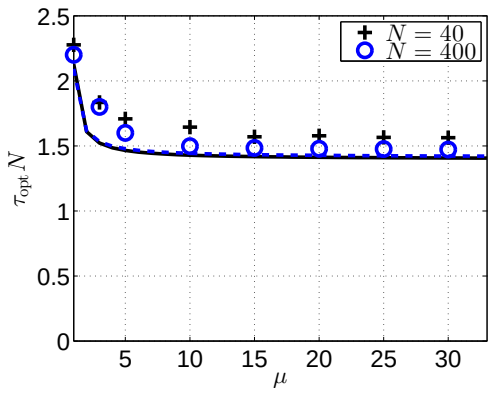


Figure 12: Optimal learning parameter $\tau_{\rm opt}$ of the $(\mu/\mu_I, 3\mu)$-σSA-ES for $a_i = i$ as a function of the number of parents μ. Curves depict the theoretical predictions of Eq. (73), while points represent experimental results for $+\ N = 40$ and $\circ\ N = 400$.

As for the ellipsoid examples considered, the approximate $\tau_{\rm opt}$ formula ($\alpha_\sigma = 1$, $N \to \infty$, $\mu \to \infty$) reads for $a_i = i$

$$\tau_{\rm opt}|_{a_i=i} \simeq \sqrt{2}/N \tag{75}$$

and for $a_i = i^2$

$$\tau_{\rm opt}|_{a_i=i^2} \simeq \sqrt{3/N^3}. \tag{76}$$

Note, the learning parameter formulae (75) and (76) deviate from the known sphere model result $\tau_{\rm opt}|_{\rm sp} \simeq \sqrt{1/2N}$.

B. Experiments and Discussion

In order to evaluate the prediction quality of (73), real $(\mu/\mu_I, 10)$-σSA-ES runs have been performed over periods of $G_{\max}$ generations using different τ values. The $F\left(\mathbf{y}^{(g)}\right)$ values for generations $g = G_{\max}/2$ to $g = G_{\max}$ have been recorded in order to empirically estimate the normalized local quality gain [9, p.132] using

$$\overline{Q}^*_{\mathbf{y}} \approx -\frac{\Sigma a}{G_{\max}/2} \sum_{g=G_{\max}/2}^{G_{\max}} \frac{F\left(\mathbf{y}^{(g)}\right) - F\left(\mathbf{y}^{(g-1)}\right)}{2\sum_{i=1}^N a_i^2 \left(y_i^{(g-1)}\right)^2}. \tag{77}$$

The obtained $\overline{Q}^*_{\mathbf{y}}$ data have been averaged over 30 independent runs. These runs have been performed for a set of equidistantly chosen τ values. The τ value producing the maximum $\overline{Q}^*_{\mathbf{y}}$ has been considered optimal.

The $\tau_{\rm opt}$ dependency on the population size λ is compared with experiments on the $a_i = i$ ellipsoid in Fig. 12, where the $(\mu/\mu_I, \lambda)$-σSA-ES used a truncation ratio of $1/3$, i.e. $\lambda = 3\mu$. Theoretical curves obtained using Eq. (73) are located close to each other and follow the same rule: A relatively large $\tau_{\rm opt}$ value for small μ decreases down to $\tau_{\rm opt} \approx 1.4/N$ and approaches a constant value for $\mu > 10$. As expected, the empirically determined values of the $N = 400$ case (circles) are closer to the theoretical curves than the $N = 40$ points (crosses).

According to Fig. 12, the $(\mu/\mu_I, \lambda)$-σSA-ES is insensitive to the choice of the population size parameter for sufficiently large $\mu \geq 10$. This property is useful for global optimization of objective functions with multiple local optima: To increase the chance of global convergence on such functions, ES-restarts in conjunction with population size increase are often employed [6]. Only a weak $\tau_{\rm opt}$ dependency on the population size allows for the usage of a fixed $\tau_{\rm opt}$ when increasing the population size. Moreover, the corresponding steady state mutation strength (71) increases in proportion to μ. This also helps in global search. However, unlike the population size parameter, according to Eq. (73) the local landscape of the objective function has strong influence on the $\tau_{\rm opt}$ value. This also holds for the often used test functions Cigar: $F_{\rm c}(\mathbf{y}) := y_1^2 + \xi \sum_{i=2}^N y_i^2$ and Hansen's ellipsoid: $F_{\rm H}(\mathbf{y}) := \sum_{i=1}^N a_i y_i^2$ with $a_i := 10^{\alpha \frac{i-1}{N-1}}$ [15], [14]. However, in these cases the N-scaling behavior is similar to the Sphere. Using (73) and $\Sigma a = 1 + (N-1)\xi$, one easily finds the asymptotic expression $\tau_{\rm opt}|_{\rm Cigar} \simeq \frac{1}{\sqrt{N\xi}}$. For $F_{\rm H}$ a somewhat longer calculation[7] yields $\tau_{\rm opt}|_{\rm H} \simeq \sqrt{\frac{\alpha \ln 10}{10^\alpha N}}$.

V. Summary and Conclusions

The behavior of the self-adaptation evolution strategy with intermediate multirecombination, the $(\mu/\mu_I, \lambda)$-σSA-ES, on the ellipsoid model (1) has been investigated using the dynamical systems approach. To this end, a novel progress quantity measuring the expected quadratic progress of single parent vector components – the quadratic progress rate φ_i^{II} – has been introduced in this paper. The derivation of the asymptotically exact φ_i^{II} formula (16) has been sketched. Being based on φ_i^{II} and the self-adaptation response function ψ (4), a system of $N+1$ nonlinear evolution equations (22), (23) has been derived that governs the mean value dynamics of the $(\mu/\mu_I, \lambda)$-σSA-ES. Due to the nonlinearity of the system (22), (23), closed-form solutions of the dynamics do not exist. However, considering the steady state that is reached in the asymptotic generation limit $g \to \infty$, the system can be solved using a special Ansatz. Having used the Ansatz (26), the steady state problem turned into the eigenvalue problem (35). While such eigenvalue problems can be solved numerically, the primary goal of the paper was to provide closed form expressions for the smallest eigenvalue and the corresponding eigenvector. An approximate solution has been found that describes the steady state behavior of the ES well. In turn, this solution allowed the determination of the optimal learning parameter τ in terms of a closed form expression (73).

The steady state mean value dynamics derived rest on a set of approximations. These (a) neglect σ^* fluctuations and (b) express the progress rates and self-adaptation response function by asymptotically exact expressions the quality of which improves for larger N and smaller σ^* provided that the mutation induced fitness is normally distributed. This is guaranteed through the Central Limit Theorem of Statistics which in turn requires Lyapunov's condition to be fulfilled. This basically means that there is no dominating component in the sum of random variates contributing to the fitness fluctuations. In order to ensure non-dominating contributions,

[7] Here we have used $\Sigma a = \sum_{i=1}^N a_i = \frac{10^{\alpha N/(N-1)} - 1}{10^{\alpha/(N-1)} - 1} \simeq \frac{10^\alpha N}{\ln(10^\alpha)}$.

$a_i/\Sigma a$ $(\forall i = 1, \ldots, N)$ should be small and vanish asymptotically as $N \to \infty$. For the examples considered in detail: Sphere, $a_i = i$, $a_i = i^2$ as well as for well-known models like Cigar, Discus, and Hansen's ellipsoid this is fulfilled. Yet, one can construct ellipsoids, e.g. a discus where the dominating eigenvalue scales with the search space dimensionality N, e.g., $\max(a_i) = N^2$. In that case one never reaches normality and the formulae derived remain approximations even for $N \to \infty$.

The steady state dynamics are governed by exponentially decreasing y_i components given by (26) where the inverse time constant ν is determined by (45). Having a closer look at (45) and the corresponding fitness model (1), it becomes clear that the result does also hold for the general fitness model $F(\mathbf{y}) = \mathbf{y}^T\mathbf{Q}\mathbf{y}$ with $\mathbf{Q}$ as positive definite matrix (minimization considered). The parameter $\breve{a} = \min(a_i)$ is simply the smallest eigenvalue κ of the corresponding eigenvalue problem $\mathbf{Q}\mathbf{u} = \kappa\mathbf{u}$. Since the sum of the eigenvalues of $\mathbf{Q}$ is the trace of $\mathbf{Q}$, $\Sigma a = \mathrm{Tr}[\mathbf{Q}]$, (45) can be expressed in terms of

$$\boxed{\nu = 2\sigma^* c_{\mu/\mu,\lambda} \min(\kappa_i)/\mathrm{Tr}[\mathbf{Q}].} \tag{78}$$

The steady state fitness dynamics can be determined using (1) and (26) starting from generation g_0 for an evolution interval g

$$F\big(\mathbf{y}^{(g_0+g)}\big) = \textstyle\sum_{i=1}^N a_i b_i \mathrm{e}^{-\nu(g_0+g)} = F\big(\mathbf{y}^{(g_0)}\big)\,\mathrm{e}^{-\nu g}. \tag{79}$$

That is, the objective function drops exponentially fast with time constant $1/\nu$. Equation (79) can be used to estimate the expected running time G needed to improve the result by a factor of $2^{-\beta}$. Considering $F\big(\mathbf{y}^{(g_0+G)}\big)/F\big(\mathbf{y}^{(g_0)}\big)$, one immediately obtains from (79) $\mathrm{e}^{-\nu G} = 2^{-\beta}$. Resolving for G, one gets $G = \beta\ln(2)/\nu$ and with (45)

$$\boxed{G = \frac{\beta\ln(2)}{2\sigma^*_{\mathrm{ss}} c_{\mu/\mu,\lambda}} \cdot \frac{\sum_{i=1}^N a_i}{\min(a_i)}.} \tag{80}$$

That is, G is asymptotically proportional to the quotient of the trace of $\mathbf{Q}$ and its smallest eigenvalue $\breve{\kappa}$: $G \propto \mathrm{Tr}[\mathbf{Q}]/\min(\kappa_i)$. Using (71), the minimal expected running time becomes

$$\breve{G} = \frac{\beta\ln(2)}{4\alpha_\sigma \mu c^2_{\mu/\mu,\lambda}} \cdot \frac{\sum_{i=1}^N a_i}{\min(a_i)} \cdot \frac{1}{1 - N\breve{a}/\sum_{i=1}^N a_i} \tag{81}$$

for non-spherical ellipsoid models provided that the optimal learning parameter τ, Eq. (73), is used.

Considering (80), one finds that the expected running time increases with order N^2 for the ellipsoid model $a_i = i$, with N^3 for $a_i = i^2$, and with N for the Cigar and Hansen's ellipsoid. The latter results might come as a surprise: From viewpoint of asymptotic runtime complexity, Cigar and Hansen's ellipsoid yield the same complexity as the Sphere model, i.e. $\mathcal{O}(N)$ (w.r.t. function evaluations). However, unlike the Sphere, these two ellipsoid models have usually large factors, ξ and $10^\alpha/\ln(10^\alpha)$, respectively, obscured by the order notation. Since the runtime predictions made by (80) might be somewhat astonishing, experiments have been conducted to check its validity for real ES runs. Figure 13 shows the N scaling behavior of the $(3/3_I, 10)$-σSA-ES, $\tau = 1/\sqrt{N}$, on the ellipsoids $a_i = i, i^2$, and Hansen's with

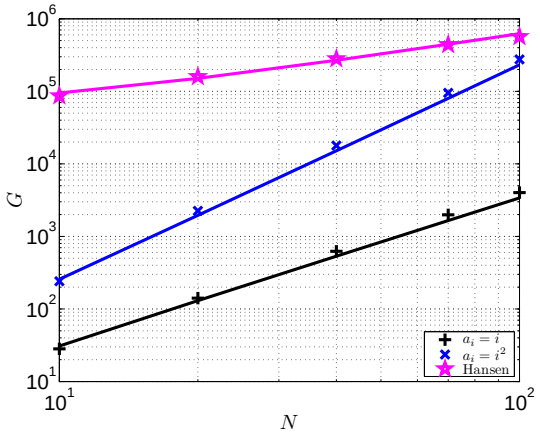


Figure 13: Expected runtime experiments for the ellipsoid models $a_i = i$, i^2, and Hansen with $\alpha = 5$. The predictions of (80) are displayed by curves.

$\alpha = 5$ for $\beta = 2$. There is a good agreement between theory and experiments. While Hansen's ellipsoid requires the largest number of generations for the small N cases (even for the condition number 10^α, $\alpha = 5$, considered here), for sufficiently large N the ellipsoids with $a_i = i$ and i^2 are harder to optimize for the σSA-ES.

The runtime results are also in agreement with findings of Jägersküpper [17] for the $(1+1)$-ES with $1/5$-rule. While his approach provided a rigorous proof of runtime bounds, the analysis presented here yields results for multirecombinant ES including the quantitative influence of the strategy parameters such as the learning parameter τ and the truncation ratio μ/λ.

Comparing Eqs. (80) and (81), one can assess the influence of the choice of τ on the expected running time. The learning parameter controls the steady state σ^*, Eq. (67). Using τ_{opt}, Eq. (73), one can gain approximately a factor of μ compared to the choice $\tau = \mathrm{const}$. Note, even the standard recommendation $\tau \propto 1/\sqrt{N}$, that is optimal for the sphere model, does not provide a runtime reduction. For example, considering the ellipsoids with $a_i = i$ and i^2, τ must be chosen according to Eqs. (65) and (66), respectively. This reveals a dilemma for real world applications: Since the local structure of the real fitness landscape is not known, there is a priori no way to fix τ for optimal ES performance. Therefore, any choice of τ will be a compromise.

Having learned that the $(\mu/\mu_I, \lambda)$-σSA-ES performs suboptimally, the question arises how alternative σ control techniques do behave. There the cumulative step-size adaptation (CSA) of the CMA-ES [15] comes into mind. Its analysis on the ellipsoid model is still pending. Another alternative would be Meta-ES where theoretical treatment has just begun for simple fitness models [4], [11].

The analysis presented can be regarded as a first step towards the analysis of ES with covariance matrix adaptation, especially for the CMSA-ES. While the covariance matrix learning has not been analyzed so far, the current work provides the modeling approach for the general step-size adaptation: Any standard covariance matrix adaptation ES can

be regarded as an ES operating with isotropic mutations on a composite function comprising a linear transformation (controlled by the evolving covariance matrix) and the objective function. That is, if the objective function is a quadratic form, the ES simply "sees" another, but also quadratic form. That is why, the approach presented can also be applied to such cases.

REFERENCES

[1] D. V. Arnold. *Noisy Optimization with Evolution Strategies*. Kluwer Academic Publishers, Dordrecht, 2002.

[2] D. V. Arnold and H.-G. Beyer. Expected Sample Moments of Concomitants of Selected Order Statistics. *Statistics and Computing*, 15:250–241, 2005.

[3] D.V. Arnold, H.-G. Beyer, and A. Melkozerov. On the behaviour of weighted multi-recombination evolution strategies optimising noisy cigar functions. In *GECCO'2009*, pages 483–490, 2009.

[4] D.V. Arnold and A. MacLeod. Hierarchically organised evolution strategies on the parabolic ridge. In *Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation (GECCO'2006)*, pages 437–444. ACM, 2006.

[5] A. Auger. Convergence Results for $(1, \lambda)$-SA-ES using the Theory of φ-irreducible Markov Chains. *Theoretical Computer Science*, 334(1–3):35–69, 2005.

[6] A. Auger and N. Hansen. A Restart CMA Evolution Strategy with Increasing Population Size. In *Congress on Evolutionary Computation*, volume 2, pages 1769–1776. IEEE, 2005.

[7] H.-G. Beyer. Toward a Theory of Evolution Strategies: Some Asymptotical Results from the $(1,^+\lambda)$-Theory. *Evolutionary Computation*, 1(2):165–188, 1993.

[8] H.-G. Beyer. Toward a Theory of Evolution Strategies: Self-Adaptation. *Evolutionary Computation*, 3(3):311–347, 1995.

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

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

[11] H.-G. Beyer and M. Hellwig. Mutation Strength Control by Meta-ES on the Sharp Ridge. In *GECCO'2012*, pages 305–312, 2012.

[12] H.-G. Beyer and B. Sendhoff. Covariance Matrix Adaptation Revisited–The CMSA Evolution Strategy. *Parallel Problem Solving from Nature–PPSN X*, pages 123–132, 2008.

[13] S. Finck. *Analysis of Evolution Strategies on a Subset of Quadratic Functions and Methods for Comparing Optimization Strategies*. PhD thesis, University of Stuttgart, Stuttgart, Germany, 2011.

[14] N. Hansen, A. Auger, R. Ros, S. Finck, and P. Posík. Comparing results of 31 algorithms from the black-box optimization benchmarking BBOB-2009. In *GECCO'2010 (Companion)*, pages 1689–1696, 2010.

[15] N. Hansen, S.D. Müller, and P. Koumoutsakos. Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES). *Evolutionary Computation*, 11(1):1–18, 2003.

[16] J. Jägersküpper. Rigorous Runtime Analysis of the (1+1)-ES: 1/5-Rule and Ellipsoidal Fitness Landscapes. In A.H. Wright et al., editor, *Foundations of Genetic Algorithms, 8*, pages 260–281, Berlin, 2005. Springer-Verlag.

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

[18] A. Melkozerov and H.-G. Beyer. On the Analysis of Self-Adaptative Evolution Strategies on Elliptic Model: First Results. In J. Branke et al., editor, *GECCO'2010: Proceedings of the Genetic and Evolutionary Computation Conference*, pages 369–376, New York, 2010. ACM.

[19] S. Meyer-Nieberg. *Self-Adaptation in Evolution Strategies*. PhD thesis, University of Dortmund, CS Department, Dortmund, Germany, 2007.

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

[21] S. Meyer-Nieberg and H.-G. Beyer. Mutative Self-Adaptation on the Sharp and Parabolic Ridge. In C. Stephens et al., editor, *Foundations of Genetic Algorithms, 9*, pages 70–96, Berlin, 2007. Springer-Verlag.

[22] I. Rechenberg. *Evolutionsstrategie–Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Frommann-Holzboog, 1973.

ACKNOWLEDGEMENTS

Support by the Austrian Science Fund (FWF) under grant P22649-N23 is gratefully acknowledged. The work of A. Melkozerov was supported by the Russian Foundation for Basic Research (grant 12-01-31110) and under the contract no. 96/12 of November 16, 2012 between TUSUR and joint-stock company Academician M. F. Reshetnev "Information Satellite Systems" with the aim of implementing the Resolution of the Russian Federation Government no. 218 of April 9, 2010, contract no. 02.G25.31.0042 of March 12, 2013.

APPENDIX

DERIVATION OF THE QUADRATIC PROGRESS RATE φ_i^{II}

In this section, the product moments in Eq. (15), E_1 and E_2, are calculated. For sake of simplicity, the coordinate index i labeling the ith component of the mutation vector $\mathbf{x}$ (note, only i and j are used to mark vector components in this appendix) and the corresponding progress rate φ^{II} is omitted as long as there is no danger of a mix-up. That is, $x_{k;\lambda}$ refers to the ith component of the mutation vector $\mathbf{x}$ producing the kth best offspring $\tilde{\mathbf{y}}_{k;\lambda} = \mathbf{y} + \mathbf{x}_{k;\lambda} = \mathbf{y} + \tilde{\sigma}_{k;\lambda}\mathbf{z}_{k;\lambda}$. The offspring are ranked according to the offspring objective function values $\tilde{F}(\tilde{\mathbf{y}})$ and $\mathbf{z} \sim \boldsymbol{\mathcal{N}}(\mathbf{0}, \mathbf{I})$ (cf. Fig. 1, Line 7). Since the $N \to \infty$ case is considered, it holds $\tilde{\mathbf{y}}_{k;\lambda} = \mathbf{y} + \sigma\mathbf{z}_{k;\lambda}$. Thus, E_1 and E_2 in (14) can be expressed in terms of

$$E_1 = \sigma^2 \mathrm{E}\left[\sum_{l=2}^{\mu}\sum_{k=1}^{l-1} z_{k;\lambda} z_{l;\lambda} | \mathbf{y}\right], \tag{82}$$

$$E_2 = \sigma^2 \mathrm{E}\left[\sum_{m=1}^{\mu} (z_{m;\lambda})^2 | \mathbf{y}\right]. \tag{83}$$

In order to calculate E_1 and E_2, the $z_{k;\lambda}$ *noisy* order statistics must be considered for each component of the $\mathbf{z}_{k;\lambda}$ vector. In a first step, z_i must be related to the local quality change $Q_{\mathbf{y}}(\mathbf{x}) = F(\tilde{\mathbf{y}}) - F(\mathbf{y})$ of the mutation $\mathbf{x}$.

A. Local Quality Change

Using (1), the quality change of the offspring reads

$$\begin{aligned} Q_{\mathbf{y}}(\mathbf{x}) = & \sum_{j=1}^{N} a_j (y_j + \sigma z_j)^2 - \sum_{j=1}^{N} a_j y_j^2 \\ = & \ 2\sigma a_i y_i z_i + 2\sigma \sum_{j\neq i}^{N} a_j y_j z_j + \sigma^2 \sum_{j=1}^{N} a_j z_j^2, \end{aligned} \tag{84}$$

where in the 2nd line terms are rearranged in such a manner that the ith component (z_i) of the $\mathbf{z}$ vector is separated from the rest. Dividing both sides by $2\sigma a_i y_i$ already isolates the z_i variate. Introducing v_i for the quotient

$$\frac{Q_{\mathbf{y}}(\mathbf{x})}{2\sigma a_i y_i} =: v_i, \tag{85}$$

one obtains

$$v_i = z_i + \sum_{j\neq i}^{N} \frac{a_j y_j}{a_i y_i} z_j + \frac{\sigma}{2} \sum_{j=1}^{N} \frac{a_j}{a_i y_i} z_j^2. \tag{86}$$

This is a sum of an $\mathcal{N}(0,1)$ standard normal variate z_i and two additional sums having $N-1$ and N addends, respectively.

The latter is non-centrally χ^2 distributed. Since the case $N \to \infty$ is considered, the central limit theorem can be applied and the two sums can be approximated by a normal distribution. The parameters of that distribution can be easily obtained. One immediately reads from (86) that the expected value is given by $\sigma(\sum_{j=1}^{N} a_j)/2a_i y_i$. Due to the stochastic independence of the mutation components, the variance can be calculated as the sum of the variances of the individual z components. Using the simple formula[8] $\mathrm{Var}[Az + Bz^2] = A^2 + 2B^2$ (for $z \sim \mathcal{N}(0,1)$), one obtains for the variance of the two sum expressions in (86)

$$\vartheta_i^2 = \frac{1}{a_i^2 y_i^2}\left(\sum_{j\neq i}^{N} a_j^2 y_j^2 + \frac{\sigma^2}{2}\sum_{j=1}^{N} a_j^2\right). \tag{87}$$

As a result, (86) can be expressed as

$$v_i \sim z_i + \mathcal{N}\left(\frac{\sigma\Sigma a}{2a_i y_i}, \frac{1}{a_i^2 y_i^2}\left(\sum_{j\neq i}^{N} a_j^2 y_j^2 + \frac{\sigma^2}{2}\sum_{j=1}^{N} a_j^2\right)\right) \tag{88}$$

Let us consider the distribution of $z_{k;\lambda}$ (note, the index i has been dropped here) belonging to the kth best Q value, i.e. $v_{k;\lambda}$. The variates $z_{k;\lambda}$ are noisy order statistics (also referred to as concomitants of $v_{k;\lambda}$) due to the $\mathcal{N}$ term in (88). Calculating sums of product moments of these statistics, such as (82) and (83), is a technically involved task. However, that has already been solved in [2] for the general case $\mathrm{E}[S_A]$, where

$$S_A := \sum \cdots \sum z_{n_1;\lambda}^{\alpha_1} \cdots z_{n_\nu;\lambda}^{\alpha_\nu} \tag{89}$$

is a ν-fold sum and $A = (\alpha_1, \ldots, \alpha_\nu)$ is the vector of exponents α_n. Under the condition that $z \sim \mathcal{N}(0,1)$, it has been proven in [2] that

$$\mathrm{E}[S_A] = \frac{\mu!}{(\mu-\nu)!}\sum_{n=0}^{\nu}\sum_{k\geq 0}\Big[\zeta_{n,0}^{(A)}(k) + \frac{\gamma_1}{6}\zeta_{n,1}^{(A)}(k) + \frac{\gamma_2}{24}\zeta_{n,2}^{(A)}(k) + \ldots\Big] h_{\mu,\lambda}^{\nu-n,k}. \tag{90}$$

Here γ_1 and γ_2 are the coefficients of skewness and kurtosis of the noise. The $\zeta_{n,l}^{(A)}(k)$ are special polynomials of the correlation coefficient[9]

$$\rho = 1/\sqrt{1+\vartheta^2} \tag{91}$$

derived in [2]. The h coefficients are defined as

$$h_{\mu,\lambda}^{m,k} := (\lambda-\mu)\binom{\lambda}{\mu}\int_{-\infty}^{\infty} \mathrm{He}_k(x)\,[\phi(x)]^{m+1} \times [\Phi(x)]^{\lambda-\mu-1}\,[1-\Phi(x)]^{\mu-m}\,\mathrm{d}x, \tag{92}$$

where $\mathrm{He}_k(x)$ is the kth Hermite polynomial and $\phi(x)$ and $\Phi(x)$ are the pdf and cdf, respectively, of the standard normal distribution.

In the following, E_1 and E_2 will be calculated using (90). Since the noise is approximated by a normal distribution in (88), the coefficients of skewness and kurtosis, γ_1 and γ_2, are zero in (90). That is, only the $\zeta_{n,0}^{(A)}(k)$ functions must be considered in the next sections.

[8]Note that for $z \sim \mathcal{N}(0,1)$, it holds $\mathrm{E}[z] = 0$, $\mathrm{E}[z^2] = 1$, $\mathrm{E}[z^3] = 0$, and $\mathrm{E}[z^4] = 3$.

[9]Recall that $\rho[v,z] = \mathrm{Cov}[v,z]/\sqrt{\mathrm{Var}[v]\mathrm{Var}[z]}$.

B. Expectation E_1

Comparing the double sum in (82) with (89), one sees that $\nu = 2$, $A = (\alpha_1, \alpha_2) = (1,1)$ and thus,

$$E_1 = \sigma^2 \mathrm{E}\left[S_{(1,1)}\right]. \tag{93}$$

Applying (90), one gets

$$\mathrm{E}\left[S_{(1,1)}\right] = \frac{\mu!}{(\mu-2)!}\sum_{n=0}^{2}\sum_{k\geq 0}\zeta_{n,0}^{(1,1)}(k) h_{\mu,\lambda}^{2-n,k} = \mu(\mu-1)\frac{\rho^2}{2}h_{\mu,\lambda}^{2,0}, \tag{94}$$

because, according to Table 1 in [2], all $\zeta_{n,0}^{(1,1)}(k)$ are identically zero except $\zeta_{0,0}^{(1,1)}(0) = \rho^2/2$ and $\gamma_1 = \gamma_2 = 0$. Taking $\mathrm{He}_0(x) = 1$ into account and the pdf $\phi(x) = \mathrm{e}^{-\frac{1}{2}x^2}/\sqrt{2\pi}$ of the standard normal variate, a comparison of (92) with (5) reveals that $h_{\mu,\lambda}^{2,0} = e_{\mu,\lambda}^{2,0}$. Taking (91) into account, one obtains $E_1 = \mu(\mu-1)\sigma^2\rho^2 e_{\mu,\lambda}^{2,0}/2$. As a final step, ρ^2 for the ith coordinate is calculated using (87) and (91). This yields

$$\rho^2 = \frac{a_i^2 y_i^2}{a_i^2 y_i^2 + \sum_{j\neq i}^{N} a_j^2 y_j^2 + \frac{\sigma^2}{2}\sum_{j=1}^{N} a_j^2} = \frac{a_i^2 y_i^2}{\sum_{j=1}^{N} a_j^2\left(y_j^2 + \frac{\sigma^2}{2}\right)} \tag{95}$$

and finally

$$E_1 = \mu(\mu-1)\frac{\sigma^2}{2}\frac{a_i^2 y_i^2 e_{\mu,\lambda}^{2,0}}{\sum_{j=1}^{N} a_j^2\left(y_j^2 + \frac{\sigma^2}{2}\right)} \tag{96}$$

C. Expectation E_2

The sum in the E_2 formula (83) can be expressed using (89) by $\nu = 1$ and $A = (\alpha_1) = (2)$

$$E_2 = \sigma^2 \mathrm{E}\left[S_{(2)}\right]. \tag{97}$$

Applying (90), one obtains by means of Table 1 in [2]

$$\begin{aligned}\mathrm{E}\left[S_{(2)}\right] &= \frac{\mu!}{(\mu-1)!}\sum_{n=0}^{1}\sum_{k\geq 0}\zeta_{n,0}^{(2)}(k) h_{\mu,\lambda}^{1-n,k} \\ &= \mu\left(\sum_{k\geq 0}\zeta_{0,0}^{(2)}(k) h_{\mu,\lambda}^{1,k} + \sum_{k\geq 0}\zeta_{1,0}^{(2)}(k) h_{\mu,\lambda}^{0,k}\right) \\ &= \mu\left(\rho^2 h_{\mu,\lambda}^{1,1} + h_{\mu,\lambda}^{0,0}\right)\end{aligned} \tag{98}$$

since all $\zeta_{n,0}^{(2)}(k) = 0$ except $\zeta_{0,0}^{(2)}(1) = \rho^2$ and $\zeta_{1,0}^{(2)}(0) = 1$. Noting that $\mathrm{He}_1(x) = x$, one easily finds using (92) and (5) that $h_{\mu,\lambda}^{1,1} = e_{\mu,\lambda}^{1,1}$. Furthermore, $h_{\mu,\lambda}^{0,0} = 1$, thus one obtains

$$E_2 = \mu\sigma^2\left(1 + \rho^2 e_{\mu,\lambda}^{1,1}\right). \tag{99}$$

Finally plugging (95) into (99), one gets

$$E_2 = \mu\sigma^2\left(1 + \frac{a_i^2 y_i^2 e_{\mu,\lambda}^{1,1}}{\sum_{j=1}^{N} a_j^2\left(y_j^2 + \frac{\sigma^2}{2}\right)}\right). \tag{100}$$