On the Use of the "Mutate Large, But Inherit Small Principle" in Global and Noisy Optimization

Lisa Schönenberger
lisa.schoenenberger@uni-ulm.de
Ulm University, Institute of Theoretical Computer Science and Institute of Medical Systems Biology
89069 Ulm, Germany
Vorarlberg University of Applied Sciences, Research Center Business Informatics
6850 Dornbirn, Austria

Hans-Georg Beyer
hans-georg.beyer@evolutionary-computation.com
http://www.evolutionary-computation.com
Ulm University, Institute of Theoretical Computer Science and Institute of Medical Systems Biology
89069 Ulm, Germany

Abstract

Rescaled mutations have been proven to improve the convergence behavior of Evolution Strategies (ES) in noisy or highly multimodal fitness landscapes. The basic principle is "mutate large, but inherit small", meaning that the step toward the new parental centroid is reduced by a factor $1/\kappa$. This paper investigates the impact of κ on the progress rate of a $(\mu/\mu_I, \lambda)$-ES in a highly multimodal and noisy scenario. For the second-order progress rate, the gain part scales inverse linearly and the loss part scales inverse quadratically with κ. This relationship is applied to the Rastrigin function in the limit of large dimensionality. Furthermore, a κ-dependent first-order progress rate for the noisy sphere is derived, yielding the aforementioned scaling behavior. The findings suggest that as κ increases, the interval of the normalized noise strength where positive progress is possible gets larger. Consequently, the search space can be explored with a larger mutation strength, leading to larger success rates in the Rastrigin landscape and to a smaller steady-state distance in the noisy sphere case.

Keywords

Evolution Strategy, rescaled mutations, global optimization, progress rate, noise, multimodality, Rastrigin function

ACM Reference Format:

Lisa Schönenberger and Hans-Georg Beyer. 2026. On the Use of the "Mutate Large, But Inherit Small Principle" in Global and Noisy Optimization. In *Genetic and Evolutionary Computation Conference (GECCO '26), July 13–17, 2026, San Jose, Costa Rica.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3795095.3805063

1 Introduction

In the context of Evolution Strategies (ES) in the N-dimensional search space, the presence of high multimodality or noise poses a significant challenge. In a multimodal landscape, an ES may encounter difficulties in localizing the global optimizer, resulting in convergence to one of several local optima. Furthermore, in landscapes disturbed by substantial noise, the distance to the optimizer can increase due to the selection process being distorted by the noise. Counteracting these behaviors, there are many methods for improving the progress of an ES. Many techniques work by increasing the gain part of the progress rate (see (10)). One such technique is the increase of the offspring population size λ for a $(1, \lambda)$-ES [4]. Conversely, more substantial progress can be achieved by reducing the loss part. For example, recombination [4] has been demonstrated to be a viable approach. In [7], it was shown that, using the CMA-ES with weighted recombination, it is possible to identify the global optimum under high multimodality. However, the required population size must be chosen sufficiently large exceeding the usually recommended default value. As an alternative, rescaled mutations can be advantageous and can lead to positive progress in highly multimodal structures, even for moderate population sizes.

The basic principle of rescaled mutations was initially proposed in [16]. It signifies that the search space is explored with a substantial mutation strength, but the step toward the new parental centroid is reduced by a factor of $1/\kappa$, where $\kappa > 1$ is referred to as the *rescaling parameter*. In [2], rescaled mutations were implemented for the $(1, \lambda)$-ES on the noisy sphere to investigate the impact on the progress rate. The results revealed that the gain part is diminished inverse linearly with the rescaling parameter κ, while the reduction of the loss part is quadratically related. These results were obtained for the asymptotic case, $N \to \infty$. In [3], the N-dependent progress rate was derived. In the case of the noisy sphere, it was observed that an increase in κ leads to an expansion of the region of the normalized mutation strength σ^* where positive progress is feasible, resulting in a decrease in the steady-state distance to the global optimizer as κ increases. In [6] rescaled mutations were applied to comparison-based algorithms on the noisy sphere, resulting in faster convergence rates. In [10], rescaled mutations for the CMA-ES were investigated in several multimodal landscapes. It has been shown that a reduced learning rate[1] has the same effect on the success rate as a larger population size. The learning rate in [10] was regarded as a hyperparameter and was constant for the entire duration of the experiment. An online adaptation of the learning rate was proposed in [8], with the objective of accelerating the process. In [12] and [11] the learning rate is adapted to maintain a constant signal-to-noise ratio. A deeper theoretical understanding of the effect of rescaled mutations can be achieved by the progress rate analysis.

[1]The concept of the learning rate is used in the field of machine learning and in the context of CMA-ES. Rescaled mutations can also be interpreted as a reduced learning rate.

GECCO '26, July 13–17, 2026, San Jose, Costa Rica

ACM ISBN 979-8-4007-2487-9/2026/07
https://doi.org/10.1145/3795095.3805063

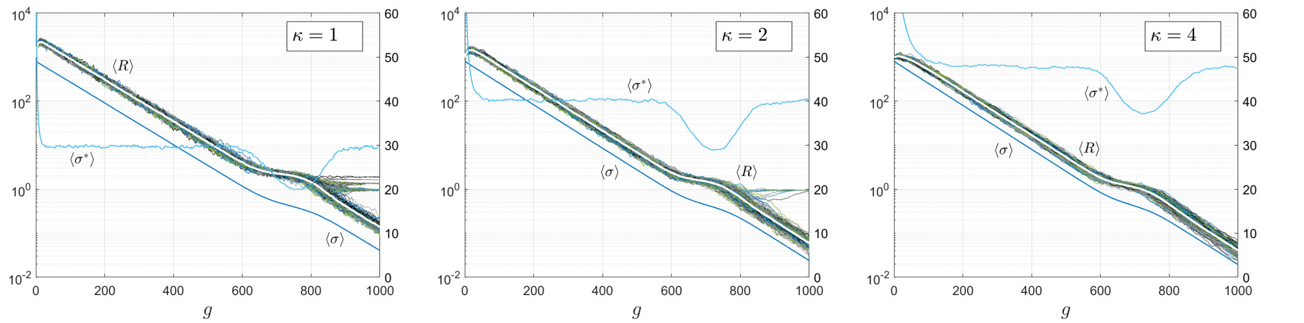

Figure 1: Thin lines: Residual distance dynamics for 200 $(50/50_I, 100)$-σSA-ES runs in the Rastrigin landscape with $\tau = 1/\sqrt{8N}$, $N = 100$, and rescaling parameter κ. For each run, the ES was initialized randomly at distance $R^{(0)} = 1000$ to the global optimizer. The success probabilities are (from left to right) $P_s = 0.55$, $P_s = 0.96$, and $P_s = 1$. Bold lines show the mean value dynamics derived from the successful runs: R-dynamics (left y-axis and white line), mutation strength σ (left y-axis and blue line) and its normalization σ^* (right y-axis and light blue line).

Up until now, a detailed analysis of the progress rate of rescaled mutations was conducted exclusively for the $(1, \lambda)$-ES on the sphere and noisy sphere model. An analysis of the progress rate with rescaled mutations for more complex scenarios is still pending. This paper focuses on the impact of rescaled mutations on the progress rate for a multi-membered $(\mu/\mu_I, \lambda)$-ES, where μ denotes the parent population size. To this end, two scenarios will be selected in which rescaled mutations have already proven themselves: The noisy sphere function and the highly multimodal Rastrigin function. The progress rate of the $(\mu/\mu_I, \lambda)$-ES on the noisy sphere was previously obtained in [1]. A progress rate analysis for the Rastrigin function was recently conducted in [14] and [13]. The present analysis extends this earlier research in the case of rescaled mutations.

The remainder of this paper is organized as follows: After a brief explanation of the ES algorithm and test functions in Section 2, Section 3 introduces ES with rescaled mutations. Section 4 will explore the impact of rescaled mutations on the progress rate. These results will then be applied to the Rastrigin function in Section 4.1 and to the noisy sphere in Section 4.2. Finally, Section 5 will provide a summary of the results and an outlook on future research.

2 Test Functions and Evolution Strategy

The Rastrigin test function F for an N-dimensional search vector $\mathbf{y} = (y_1, \ldots, y_N)$ is given by

$$F_{\mathcal{R}}(\mathbf{y}) = \sum_{i=1}^{N} \left[y_i^2 + A\left(1 - \cos(\alpha y_i)\right) \right], \qquad (1)$$

where the parameter A denotes the oscillation amplitude and α denotes the frequency. Unless otherwise stated, the parameters A and α will have the values $A = 1$ and $\alpha = 2\pi$ in all experiments. The global minimizer of the Rastrigin function is located at $\hat{\mathbf{y}} = \mathbf{0}$.

The noisy sphere function with normally distributed noise is given by

$$F_{\mathcal{S}}(\mathbf{y}) = \sum_{i=1}^{N} y_i^2 + \sigma_{\text{ES}} \mathcal{N}(0, 1), \qquad (2)$$

where σ_{ES} denotes the *noise strength*.

The basic $(\mu/\mu_I, \lambda)$-ES investigated here consist of μ parents and λ offspring with *truncation ratio* $\vartheta := \mu/\lambda$. The subscript "$m;\lambda$" denotes the selection of the $m = 1, \ldots, \mu$ best individuals out of λ. The control of the strength σ of the isotropic Gaussian mutations used is done by σ *self-adaptation* (σSA), where the offspring mutation strengths are sampled from a log-normal distribution with *learning parameter* τ. The standard choice of the learning parameter is $\tau = 1/\sqrt{2N}$, which guarantees (nearly) optimal performance on the sphere model [9]. A smaller τ yields a slower adaptation, which is advantageous in the case of multimodal problems as it increases the *success probability* P_s [18], therefore $\tau = 1/\sqrt{8N}$ will be used in the experiments.

3 Rescaled Mutations

The basic principle of *rescaled mutations* is to perform large mutations while avoiding traversing the entire distance to the new parental centroid. Instead, only a part of the distance is realized. The direction remains constant, but the step size is reduced by

Algorithm 1 The $(\mu/\mu_I, \lambda)$-σSA-ES with Rescaled Mutations

Initialize $\left(\mathbf{y}^{(0)}, \sigma^{(0)}, \sigma_{\text{stop}}, g = 0, \kappa\right)$
repeat
  for $l = 1$ **to** λ **do**
    $\tilde{\sigma}_l = \sigma^{(g)} \mathrm{e}^{\tau \mathcal{N}(0,1)}$ ▷ mutate parental σ
    $\tilde{\mathbf{z}}_l = (\mathcal{N}(0,1), \ldots, \mathcal{N}(0,1))$ ▷ search direction
    $\tilde{\mathbf{y}}_l = \mathbf{y}^{(g)} + \tilde{\sigma}_l \tilde{\mathbf{z}}_l$ ▷ mutate $\mathbf{y}$
    $\tilde{F}_l = F\left(\tilde{\mathbf{y}}_l\right)$ ▷ evaluate offspring
  end for
  $\left(\tilde{\mathbf{y}}_{1;\lambda}, \ldots, \tilde{\mathbf{y}}_{\mu;\lambda}\right) \leftarrow$ *sort* $\left(\tilde{\mathbf{y}} \text{ w.r.t. ascending } \tilde{F}\right)$
  $\mathbf{x}^{(g)} = \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\sigma}_{m;\lambda} \tilde{\mathbf{z}}_{m;\lambda}$ ▷ recombine the μ best $\tilde{\sigma}\tilde{\mathbf{z}}$
  $g = g + 1$
  $\mathbf{y}^{(g)} = \mathbf{y}^{(g-1)} + \frac{1}{\kappa} \mathbf{x}^{(g-1)}$ ▷ new centroid
  $\sigma^{(g)} = \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\sigma}_{m;\lambda}$ ▷ recombine the μ best $\tilde{\sigma}$
until $\sigma^{(g)} < \sigma_{\text{stop}}$

the factor $1/\kappa$, where κ is designated as the *rescaling parameter*. In the following, the notation $\mathbf{x}$ is used to denote the step from the parental centroid of generation g to the centroid of the μ best offspring individuals of generation g, i.e.,

$$\mathbf{x}^{(g)} = \frac{1}{\mu}\sum_{m=1}^{\mu}\tilde{\mathbf{y}}_{m;\lambda} - \mathbf{y}^{(g)} = \frac{1}{\mu}\sum_{m=1}^{\mu}\tilde{\sigma}_{m;\lambda}\tilde{\mathbf{z}}_{m;\lambda}, \tag{3}$$

where the single mutations $\tilde{\sigma}_l\tilde{\mathbf{z}}_l$ from Line 6 in Alg. 1 were inserted. If the step toward the new centroid is reduced by the factor $1/\kappa$ the parental centroid of the subsequent generation is

$$\mathbf{y}^{(g+1)} = \mathbf{y}^{(g)} + \frac{1}{\kappa}\mathbf{x}^{(g)}. \tag{4}$$

This results in the σSA-ES with rescaled mutations, as outlined in Alg. 1.

The influence of the rescaling parameter κ on the performance and *success probability* P_s is investigated in Fig. 1 which shows the R-dynamics of the σSA-ES in the Rastrigin landscape, where $R = \|\mathbf{y}\|$ is the *residual distance* to the global minimizer. For larger values of κ, the ES becomes slightly faster and the success probability gets significantly larger. The bold white line shows the mean value dynamics of only the successful runs. The blue line shows the corresponding mean value of σ, and the light blue line the mean value of the normalized mutation strength

$$\sigma^* = \sigma\frac{N}{R}, \tag{5}$$

which is represented by the right y-axis. After just a few generations σ^* has reached its steady-state value. This value depends strongly on κ and is considerably larger the larger κ is.

The success probability P_s and the number of function evaluations F_E for an ES with rescaled mutations and different values of κ is investigated in Fig. 2. Each data point was generated based on 1000 independent σSA-ES. The dynamics in Fig. 1 show that the σSA-ES becomes slightly faster with larger κ. This observation is consistent with the results from Fig. 2, where F_E decreases slightly with increasing κ and a fixed value of μ. The figure shows that a larger value of κ is beneficial. The number of function evaluations to reach a 90% success rate is approximately $17 \cdot 10^4$ for $\kappa = 1$, approximately $10 \cdot 10^4$ for $\kappa = 2$, and approximately $6 \cdot 10^4$ for $\kappa = 4$.

4 Progress Rate

In order to analyze the observed behavior, the standard approach is to perform the dynamical systems approach. To this end, the first step is to calculate the different progress rates. The expected difference in the distance to the global optimizer from one generation g to the subsequent generation $g+1$ defines the *progress rate*. It is utilized to measure the expected positional difference in search space and the speed of convergence. The progress rate can be expressed as[2]

$$\varphi := \mathrm{E}\left[R^{(g)} - R^{(g+1)} \mid \mathbf{y}^{(g)}, \sigma^{(g)}\right], \tag{6}$$

where the parental centroid $\mathbf{y}^{(g)}$ and mutation strength $\sigma^{(g)}$ from the previous generation are fixed values.

In addition to the general progress in the search space also the progress in one coordinate direction can be considered. The

[2]It is assumed, without loss of generality, that the global optimizer is $\hat{\mathbf{y}} = \mathbf{0}$.

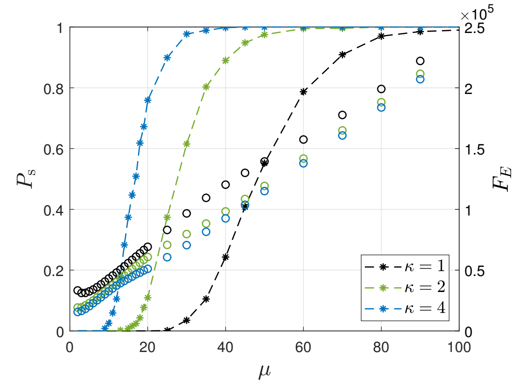

Figure 2: Success probability (*, left y-axis) and number of function evaluations (○, right y-axis) of an $(\mu/\mu_I, 2\mu)$-σSA-ES with $\tau = 1/\sqrt{8N}$ and rescaling parameter κ in the Rastrigin landscape with $N = 100$.

component-wise first-order progress rate in coordinate direction i for a given parental centroid $\mathbf{y}^{(g)}$ and mutation strength $\sigma^{(g)}$ is defined as

$$\varphi_i = \mathrm{E}\left[y_i^{(g)} - y_i^{(g+1)} \mid \mathbf{y}^{(g)}, \sigma^{(g)}\right]. \tag{7}$$

The component-wise first-order progress rate (7) with rescaled mutations as a function of κ is[3]

$$\varphi_i^\kappa := \mathrm{E}\left[y_i^{(g)} - \left(y_i^{(g)} + \frac{1}{\kappa}x_i^{(g)}\right)\right] = -\frac{1}{\kappa}\mathrm{E}\left[x_i^{(g)}\right]. \tag{8}$$

This implies that the component-wise first-order progress rate scales inverse linearly with κ. Formally, it holds independent of the underlying landscape that

$$\varphi_i^\kappa = \frac{1}{\kappa}\varphi_i. \tag{9}$$

This arises because κ has no influence on the selection process. The inverse linear relationship between κ and φ_i will be investigated for the Rastrigin function in Section 4.1.

In instances where the ith component is negative, the component-wise first-order progress rate yields a negative value when the progress is positive. In instances where the parental centroid is close to the origin, the sign of the ith component can change from one generation to the next. In such cases, the component-wise first-order progress rate can attain a negative value despite the distance to the optimizer in the ith direction decreases. Therefore, this measure is not well-suited as a measure that quantifies the progress toward the global optimizer directly. Instead the *component-wise second-order progress rate* has been introduced in [5], which involves comparing the squared coordinates. It is defined as

$$\begin{aligned}\varphi_i^{II} &:= \mathrm{E}\left[\left(y_i^{(g)}\right)^2 - \left(y_i^{(g+1)}\right)^2\right] = \mathrm{E}\left[\left(y_i^{(g)}\right)^2 - \left(y_i^{(g)} + x_i^{(g)}\right)^2\right] \\ &= -2y_i^{(g)}\mathrm{E}\left[x_i^{(g)}\right] - \mathrm{E}\left[\left(x_i^{(g)}\right)^2\right]. \end{aligned}\tag{10}$$

[3]For better readability, the dependency of the expected values on $\mathbf{y}^{(g)}$, $\sigma^{(g)}$ is no longer explicitly stated in the following.

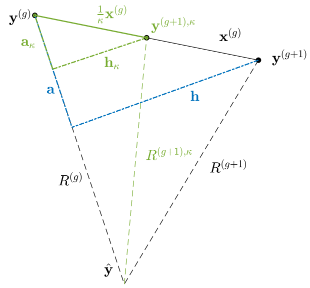

Figure 3: Decomposition of vector x into two components a and h.

The component-wise second-order progress rate delivers always a positive value if the distance to the optimizer in coordinate direction i is reduced from one generation to the next. The sign of the first expression is positive if the progress in coordinate direction i is positive, otherwise, the first expression is negative. This expression will be referred to as the *gain part*. The second part of (10) is always negative and will be referred to as the *loss part*. According to (10), the component-wise second-order progress rate for rescaled mutations depending on κ is given by

$$\varphi_i^{II,\kappa} := -\frac{2}{\kappa} y_i^{(g)} \mathrm{E}\left[x_i^{(g)}\right] - \frac{1}{\kappa^2} \mathrm{E}\left[\left(x_i^{(g)}\right)^2\right]. \tag{11}$$

This indicates that the gain part of the component-wise second-order progress rate scales inverse linearly with κ, whereas the loss part scales inverse quadratically with κ. The validity of (11) in the Rastrigin landscape is investigated in Section 4.1.

In comparison to the component-wise progress rate, the decomposition of the R-dependent progress rate (6), into the gain and the loss part is less straightforward. The fundamental concept is the decomposition of vector $\mathbf{x}$ into two components, as illustrated by the blue vectors in Fig. 3. The first component, denoted by $\mathbf{a}$, is directed toward the global optimizer. The second component, designated as $\mathbf{h}$, is perpendicular to $\mathbf{a}$. Let $a := \mathrm{sign}(R^{(g)} - R^{(g+1)})\|\mathbf{a}\|$ be the signed length of $\mathbf{a}$. Consequently, the progress rate can be expressed as

$$\varphi := \mathrm{E}\left[R^{(g)} - R^{(g+1)}\right] = \mathrm{E}\left[R^{(g)} - \sqrt{\left(R^{(g)} - a\right)^2 + \|\mathbf{h}\|^2}\right]. \tag{12}$$

Similar to $\mathbf{x}$, the vector $\mathbf{x}/\kappa$ can also be decomposed into a vector $\mathbf{a}_\kappa$ directed towards the global optimizer and a vector $\mathbf{h}_\kappa$ perpendicular to it. In this case, $\mathbf{a}$ and $\mathbf{h}$ scale also in a $1/\kappa$ ratio resulting in

$$\varphi^\kappa := \mathrm{E}\left[R^{(g)} - \sqrt{\left(R^{(g)} - \frac{1}{\kappa}a\right)^2 + \left(\frac{1}{\kappa}\|\mathbf{h}\|\right)^2}\right]. \tag{13}$$

This is visualized in Fig. 3 by the green dashed dotted lines. In (13), it is not immediately evident how the progress can be decomposed into its gain and loss components. This will be further investigated for the noisy sphere in Section 4.2.

The *R-dependent second-order progress rate* is defined as

$$\varphi_R^{II} := \mathrm{E}\left[\left(R^{(g)}\right)^2 - \left(R^{(g+1)}\right)^2\right]. \tag{14}$$

Applying the aforementioned decomposition for the rescaled case, yields

$$\varphi_R^{II,\kappa} := \mathrm{E}\left[\left(R^{(g)}\right)^2 - \left(R^{(g)} - \frac{1}{\kappa}a\right)^2 - \left(\frac{1}{\kappa}\|\mathbf{h}\|\right)^2\right] \tag{15}$$

$$= \mathrm{E}\left[\frac{2R^{(g)}a}{\kappa} - \frac{a^2}{\kappa^2} - \frac{\|\mathbf{h}\|^2}{\kappa^2}\right] = \frac{2R^{(g)}}{\kappa}\mathrm{E}\left[a\right] - \frac{1}{\kappa^2}\mathrm{E}\left[a^2 + \|\mathbf{h}\|^2\right].$$

This indicates that the gain part is reduced linearly with κ, whereas the loss part is reduced quadratically with κ. This will be further investigated for the Rastrigin function in Section 4.1.

4.1 The Influence of κ on the Rastrigin Progress Rate

The component-wise first-order progress rate φ_i in the Rastrigin landscape (1) has previously been derived in [14]. Assuming an infinitely large N and μ, and a constant $\vartheta = \mu/\lambda$, the following asymptotic behavior holds

$$\varphi_i \simeq c_\vartheta \frac{\sigma^2}{D_Q}\left(2y_i + \mathrm{e}^{-\frac{1}{2}(\alpha\sigma)^2}\alpha A \sin(\alpha y_i)\right), \tag{16}$$

where c_ϑ is the progress coefficient, which remains constant for a given truncation ratio and sufficiently large population size [4][4]. D_Q^2 denotes the variance of the quality gain[5]. In [14] it was shown that the expected value of the quality gain can be written as

$$E_Q = \sum_{i=1}^{N}\left[\sigma^2 + A\cos\left(\alpha y_i\right)\left(1 - \mathrm{e}^{-\frac{1}{2}(\alpha\sigma)^2}\right)\right] \tag{17}$$

and the variance is

$$D_Q^2 = \sum_{i=1}^{N}\left[\frac{A^2}{2}\left(1 - \mathrm{e}^{-(\alpha\sigma)^2}\right)\left(1 - \cos\left(2\alpha y_i\right)\mathrm{e}^{-(\alpha\sigma)^2}\right)\right. \tag{18}$$

$$\left. + 2\sigma^4 + 4\sigma^2 y_i^2 + 2A\alpha\sigma^2 \mathrm{e}^{-\frac{1}{2}(\alpha\sigma)^2}\left(\alpha\sigma^2\cos\left(\alpha y_i\right) + 2y_i\sin\left(\alpha y_i\right)\right)\right].$$

Combining (9) and (16) yields

$$\varphi_i^\kappa \simeq c_\vartheta \frac{\sigma^2}{\kappa D_Q}\left(2y_i + \mathrm{e}^{-\frac{1}{2}(\alpha\sigma)^2}\alpha A \sin(\alpha y_i)\right), \tag{19}$$

which is represented by the dashed lines in the two left plots of Fig. 4 as a function of σ^*. The theoretical values from (19) are compared with experimental values. For $\mathbf{y}$ a random point on the sphere with radius R was chosen. The same $\mathbf{y}$ was used for all experiments in the same figure.

The left plot in Fig. 4 shows the results for the $(10/10_I, 20)$-ES with $N = 10$. For small values of σ^*, the experimental values of φ_i exhibits a strong agreement with its theoretical prediction. For larger σ^* some deviations emerge, yet the general tendency is that the theoretical and experimental values remain in agreement. The second plot shows the results for the $(100/100_I, 200)$-ES with $N = 100$. The discrepancies between the theoretical values and their predictions are diminished in the right plot, where the dimension is larger. This is as expected, since (19) describes the asymptotic behavior for $N, \mu \to \infty$.

[4] The progress coefficient is in the range of roughly $[0.8, 1.2]$ for $\vartheta \in [1/4, 1/2]$.
[5] The quality gain [3] measures the fitness change before selection and is defined by $Q = F(\tilde{\mathbf{y}}_l) - F(\mathbf{y})$, where $\mathbf{y}$ is the parental centroid and $\tilde{\mathbf{y}}_l$ is an offspring individual.

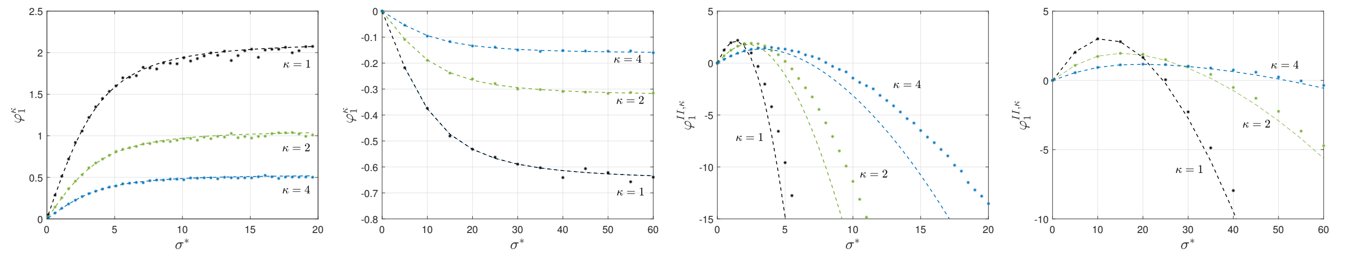

Figure 4: Component-wise progress rate in coordinate direction 1 for the Rastrigin function at residual distance $R = 10\sqrt{N}$. Stars show experimental results obtained with one-generation experiments described in Alg. 2 with $r_{\max} = 10^5$. Dashed lines show (19) (left plots) and (22) (right plots), respectively. From left to right: first-order, $(10/10_I, 20)$-ES with $N = 10$ and $y_1 = 6.20$; first-order, $(100/100_I, 200)$-ES with $N = 100$ and $y_1 = -5.80$; second-order, $(10/10_I, 20)$-ES with $N = 10$ and $y_1 = -4.47$; second-order, $(100/100_I, 200)$-ES with $N = 100$ and $y_1 = 5.55$.

In (11), the component-wise second-order progress rate was divided into a gain and a loss part. Provided that $\mathbf{y}^{(g)}$ is constant and utilizing (8), the gain part can be rewritten as

$$-\frac{2y_i^{(g)}}{\kappa}\mathrm{E}\left[x_i^{(g)} \mid \mathbf{y}^{(g)}, \sigma^{(g)}\right] = 2y_i\varphi_i^{\kappa}. \tag{20}$$

It was demonstrated in [14] that in the Rastrigin landscape the asymptotic behavior for the loss part

$$\mathrm{E}\left[\left(x_i^{(g)}\right)^2 \mid \mathbf{y}^{(g)}, \sigma^{(g)}\right] \simeq \frac{\sigma^2}{\mu} \tag{21}$$

is valid for $N, \mu \to \infty$, assuming a constant truncation ratio. Applying rescaled mutations, it can be deduced from (11) and (19) that the component-wise second-order progress rate in the Rastrigin landscape with rescaling parameter κ is asymptotically

$$\varphi_i^{II,\kappa} \simeq 2y_i\varphi_i^{\kappa} - \frac{\sigma^2}{\kappa^2\mu}, \tag{22}$$

Algorithm 2 One-Generation Progress Rate

Initialize $(\mathbf{y}, \sigma, r = 0, r_{\max})$
repeat
  for $l = 1$ **to** λ **do**
    $\tilde{\mathbf{z}}_l = (\mathcal{N}(0,1), \ldots, \mathcal{N}(0,1))$ ▷ search direction
    $\tilde{\mathbf{y}}_l = \mathbf{y} + \sigma\tilde{\mathbf{z}}_l$ ▷ mutate $\mathbf{y}$
    $\tilde{F}_l = F(\tilde{\mathbf{y}}_l)$ ▷ evaluate offspring
  end for
  $\left(\tilde{\mathbf{y}}_{1;\lambda}, \ldots, \tilde{\mathbf{y}}_{\mu;\lambda}\right) \leftarrow sort\left(\tilde{\mathbf{y}} \text{ w.r.t. ascending } \tilde{F}\right)$
  $\mathbf{x} = \frac{1}{\mu}\sum_{m=1}^{\mu}\sigma\tilde{\mathbf{z}}_{m;\lambda}$ ▷ recombine the μ best
  $\tilde{\mathbf{y}} = \mathbf{y} + \frac{1}{\kappa}\mathbf{x}$ ▷ mutation step with reduced step size
  $\varphi_i^{\kappa}(r) = y_i - \tilde{y}_i$
  $\varphi_i^{II,\kappa}(r) = (y_i)^2 - (\tilde{y}_i)^2$
  $\varphi_R^{\kappa}(r) = \|\mathbf{y}\|^2 - \|\tilde{\mathbf{y}}\|^2$ ▷ calculate progress
  $r = r + 1$
until $r \geq r_{\max}$
$\varphi_i^{\kappa} = mean\left[\varphi_i^{\kappa}(r)\right]$
$\varphi_i^{II,\kappa} = mean\left[\varphi_i^{II,\kappa}(r)\right]$
$\varphi_R^{\kappa} = mean\left[\varphi_R^{\kappa}(r)\right]$ ▷ mean over all rounds

where φ_i^{κ} is given by (19). The validity of the aforementioned equation is investigated in the right plots of Fig. 4 where (22) is compared with experimental values.

The third plot presents the outcomes for the $(10/10_I, 20)$-ES with $N = 10$. A substantial disparity exists between the theoretical and experimental values, which amplifies with increasing values of σ^*. The observed deviations for small N were expected, as (22) represents the asymptotic behavior for $N, \mu \to \infty$. As illustrated in the right plot, which depicts the results of the $(100/100_I, 200)$-ES with $N = 100$, the deviations between the experimental and predicted values are significantly smaller. The general tendency is that the theoretical and experimental values agree.

As illustrated in Fig. 4, the maximum progress decreases with larger values of κ. However, the region of σ^* where positive progress is possible is larger for larger κ. Consequently, if κ is larger, achieving positive success is possible with a larger normalized mutation strength.

The component-wise second-order progress rate can be utilized to derive an approximation for the R-dependent progress rate. For the Rastrigin function, this approach was done in [13] by aggregating the component-wise progress rate across all components. Applying this to rescaled mutations, yields

$$\begin{aligned}\varphi_R^{II,\kappa} &= \mathrm{E}\left[\sum_{i=1}^{N}\left(y_i^{(g)}\right)^2 - \sum_{i=1}^{N}\left(y_i^{(g)} + \frac{1}{\kappa}x_i^{(g)}\right)^2\right] \\ &= \sum_{i=1}^{N}\mathrm{E}\left[\left(y_i^{(g)}\right)^2 - \left(y_i^{(g)} + \frac{1}{\kappa}x_i^{(g)}\right)^2\right] = \sum_{i=1}^{N}\varphi_i^{II,\kappa}.\end{aligned} \tag{23}$$

The variance of the quality gain (18) also depends on the single components y_i. It was shown in [13, 15] that E_q and D_q can be rewritten as

$$E_Q(R) = N\sigma^2 + NA\mathrm{e}^{-\frac{1}{2}\frac{(\alpha R)^2}{N}}\left(1 - \mathrm{e}^{-\frac{1}{2}(\alpha\sigma)^2}\right) \tag{24}$$

$$\begin{aligned}\left(D_Q\right)^2(R) &= \frac{NA^2}{2}\left(1 - \mathrm{e}^{-(\alpha\sigma)^2}\right)\left(1 - \mathrm{e}^{-\alpha^2\left(\sigma^2 + 2\frac{R^2}{N}\right)}\right) \\ &\quad + 2N\sigma^4 + 4\sigma^2R^2 + 2NA\alpha^2\sigma^2\mathrm{e}^{-\frac{\alpha^2}{2}\left(\sigma^2 + \frac{R^2}{N}\right)}\left(\sigma^2 + 2\frac{R^2}{N}\right),\end{aligned} \tag{25}$$

which are expressions that only depend on R. The further derivation is contingent upon the assumption of normally distributed

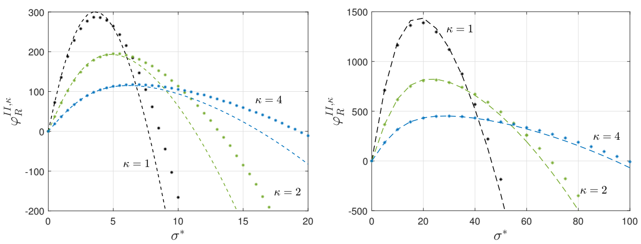

Figure 5: R-dependent progress rate for the Rastrigin function at residual distance $R = 10\sqrt{N}$. The theoretical equation (27) is represented by the dashed lines, whereas the stars represent the experimental results obtained with one-generation experiments (see Alg. 2). Left plot: $(10/10_I, 20)$-ES with $N = 10$. Right plot: $(100/100_I, 200)$-ES with $N = 100$.

$y_i = \mathcal{N}(0, R^2/N)$. This property has already been confirmed in [17] indicating that each component contributes approximately equally to the residual distance R. Using this assumption, the sum of the y_i-dependent terms in (23) can be estimated by their expected values, yielding for the sum of sine terms and $N \to \infty$

$$\sum_{i=1}^{N} y_i \sin(\alpha y_i) \simeq \alpha R^2 \mathrm{e}^{-\frac{1}{2}\frac{(\alpha R)^2}{N}}, \tag{26}$$

as shown in [13]. Subsequent insertion of (22), (19), (25), and (26) into (23) yields the asymptotic second-order progress rate

$$\begin{aligned}
\varphi_R^{II,\kappa} &\simeq \sum_{i=1}^{N} \left(c_\vartheta \frac{\sigma^2}{\kappa D_Q} \left(4y_i^2 + \mathrm{e}^{-\frac{1}{2}(\alpha\sigma)^2} 2y_i \alpha A \sin(\alpha y_i) \right) - \frac{\sigma^2}{\kappa^2 \mu} \right) \\
&\simeq c_\vartheta \frac{\sigma^2}{\kappa D_Q(R)} \left(4R^2 + \mathrm{e}^{-\frac{1}{2}(\alpha\sigma)^2} 2A\alpha^2 R^2 \mathrm{e}^{-\frac{1}{2}\frac{(\alpha R)^2}{N}} \right) - \frac{\sigma^2 N}{\kappa^2 \mu} \\
&= c_\vartheta \frac{2R^2\sigma^2}{\kappa D_Q(R)} \left(2 + A\alpha^2 \mathrm{e}^{-\frac{\alpha^2}{2}\left(\sigma^2 + \frac{R^2}{N}\right)} \right) - \frac{\sigma^2 N}{\kappa^2 \mu}.
\end{aligned} \tag{27}$$

In Fig. 5 the theoretical predictions derived in (27) are compared with experimental values. 10^5 trials were employed, each with a new random parent located on the R-sphere. The results are similar to that of the second-order progress rate in Fig. 4. The left figure, where $N = 10$ and $\mu = 10$, exhibits significant deviations between the theoretical and experimental values which increase for larger σ^*. In the right figure, with $N = \mu = 100$, the deviations are much smaller. The maximal progress diminishes with increasing κ. However, the range of σ^* where positive progress is feasible expands with increasing κ.

Evolution Equations for the Rastrigin Landscape. Given the initial values $R^{(0)}$ and $\sigma^{(0)}$ one can calculate the expected residual distance in the first generation. The expected position in generation 1 is

$$R^{(1)} = \sqrt{\left(R^{(0)}\right)^2 - \varphi_R^{II,\kappa}\left(R^{(0)}, \sigma^{(0)}\right)}. \tag{28}$$

To calculate subsequent steps, the expected σ-change is required. The self-adaptation response (SAR) is a measure of the relative

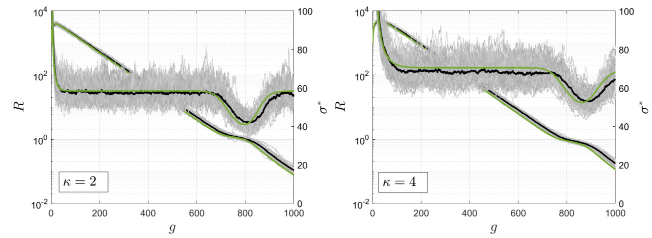

Figure 6: Dynamics for 100 $(100/100_I, 200)$-σSA-ES runs in the Rastrigin landscape with $\tau = 1/\sqrt{8N}$, $N = 100$ and rescaling parameter κ. For each run, the ES was initialized randomly at distance $R^{(0)} = 1000$ to the global optimizer. The R-dynamics are represented by the left y-axis and the σ^*-dynamics by the right y-axis, respectively. The black lines show the median of the experimental runs. The green lines show (31) and (32).

σ-change and is defined as

$$\psi\left(\mathbf{y}^{(g)}, \sigma^{(g)}\right) = \mathrm{E}\left[\frac{\sigma^{(g+1)} - \sigma^{(g)}}{\sigma^{(g)}} \mid \mathbf{y}^{(g)}, \sigma^{(g)} \right]. \tag{29}$$

The SAR-function remains unaffected by rescaled mutations. In [15, (33)], an R-dependent equation for the SAR-function in the Rastrigin landscape was derived. The SAR-function can be written as

$$\psi(R, \sigma) \simeq \tau^2 \left(\frac{1}{2} - c_\vartheta \sigma \frac{E_Q'(R)}{D_Q(R)} + e_\vartheta^{1,1} \sigma \frac{D_Q'(R)}{D_Q(R)} \right), \tag{30}$$

where $E_Q(R)$ (24) is the expected quality gain, $D_Q^2(R)$ (25) is the variance of the quality gain, and $E_Q'(R)$, $D_Q'(R)$ are the corresponding derivatives w.r.t. σ. $e_\vartheta^{a,b}$ denotes the asymptotic generalized progress coefficient [14, (45)]. When (27) and (30) are combined the expected step from one generation to the next can be expressed by the evolution equations

$$R^{(g+1)} = \sqrt{\left(R^{(g)}\right)^2 - \varphi_R^{II,\kappa}\left(R^{(g)}, \sigma^{(g)}\right)} \tag{31}$$

$$\sigma^{(g+1)} = \sigma^{(g)} \left(1 + \psi\left(R^{(g)}, \sigma^{(g)}\right) \right). \tag{32}$$

The evolution equations (31) and (32) are represented in Fig. 6 by the green lines. They were compared with 100 ES runs, whose dynamics are represented by the gray lines. The median of the experimental runs is represented by the black line. There is a strong agreement between the theoretical values and their predictions. Nonetheless, substantial deviations can occur for smaller N and μ values because (31) and (32) were derived for the asymptotic case $N, \mu \to \infty$. A more thorough theoretical analysis of the evolution equations is a matter for future research.

4.2 The Influence of κ on the Noisy Sphere Progress Rate

Rescaled mutations have been demonstrated to reduce the steady-state distance in noisy scenarios for the $(1, \lambda)$-ES so far. In order to demonstrate this aspect analytically for the $(\mu/\mu_I, \lambda)$-ES on the noisy sphere (2), this section investigates the first order progress

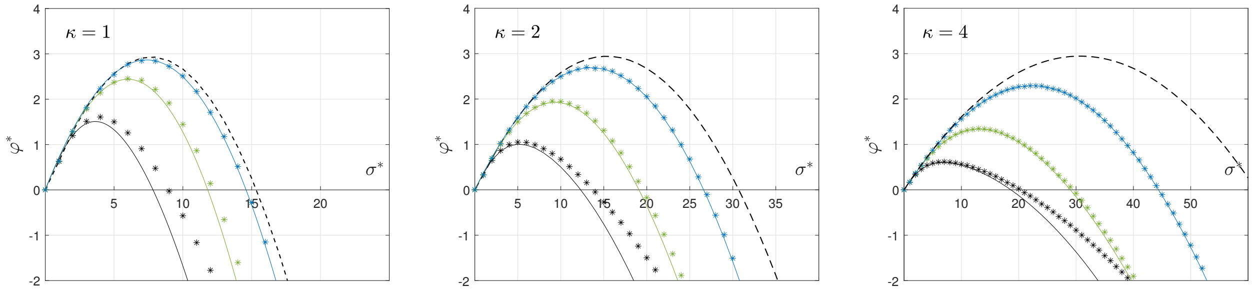

Figure 7: R-dependent progress rate for the noisy sphere with $\sigma_{\text{ES}} = 1$, $\mu = 10$, and $\vartheta = 0.5$ at residual distance $R = \sqrt{N}$. The theoretical equation (40) is represented by the solid lines, whereas the stars represent the experimental results obtained by one-generation experiments using Alg. 2 with $r_{\max} = 10^5$. Black for $N = 10$, green for $N = 100$, and blue for $N = 1000$. The black dashed line shows the asymptotic progress rate (41).

rate. Introduce

$$\varphi^* = \varphi\frac{N}{R},\ \sigma^* = \sigma\frac{N}{R},\ \sigma^*_{\text{ES}} = \sigma_{\text{ES}}\frac{N}{2R^2} \tag{33}$$

the normalized progress, the normalized mutation strength, and the normalized noise strength. In [1] it was demonstrated for the $(\mu/\mu_I, \lambda)$-ES on the noisy (quadratic) sphere model, that

$$\mathrm{E}\left[a \mid \mathbf{y}^{(g)}, \sigma^{(g)}\right] := \overline{a} \simeq \frac{c_{\mu/\mu_I,\lambda}\sigma}{\sqrt{1+\eta^2+\frac{(\sigma^*)^2}{2N}}}, \tag{34}$$

where $\eta = \sigma^*_{\text{ES}}/\sigma^*$ denotes the noise-to-signal ratio and the overline bar indicates the expected value operation. Furthermore, the expected squared length of the vector $\mathbf{h}$ for the $(\mu/\mu_I, \lambda)$-ES in the noisy sphere model was shown to be

$$\begin{aligned}\mathrm{E}\left[\|\mathbf{h}\|^2\right] := \overline{\|\mathbf{h}\|^2} &\simeq \frac{\sigma^2}{\mu}\left(N - \frac{c_{\mu/\mu_I,\lambda}\sigma^*}{\sqrt{1+\eta^2+\frac{(\sigma^*)^2}{2N}}}\right)\\ &= \frac{\sigma^2}{\mu}\left(N - \frac{N}{R}\overline{a}\right)\end{aligned} \tag{35}$$

as outlined in [1]. The application of the normalization (33) to (13) results in the following expression for the normalized progress rate:

$$\varphi^{*,\kappa} = \frac{N}{R}\mathrm{E}\left[R - \sqrt{R^2 - \frac{2R}{\kappa}a + \frac{1}{\kappa^2}a^2 + \frac{1}{\kappa^2}\|\mathbf{h}\|^2}\right]. \tag{36}$$

In [1] it was assumed that the progress rate can be approximated as the progress associated with the expected progress vector. Applying this assumption to (36), the normalized progress can be written as

$$\begin{aligned}\varphi^{*,\kappa} &\simeq N\left(1 - \frac{1}{R}\sqrt{R^2 - 2R\frac{1}{\kappa}\overline{a} + \frac{1}{\kappa^2}\overline{a^2} + \frac{1}{\kappa^2}\overline{\|\mathbf{h}\|^2}}\right)\\ &= N\left(1 - \sqrt{1 - \frac{2}{R\kappa}\overline{a} + \frac{1}{(R\kappa)^2}\overline{a^2} + \frac{1}{(R\kappa)^2}\overline{\|\mathbf{h}\|^2}}\right).\end{aligned} \tag{37}$$

For sufficiently large N the length of the step $\mathbf{a}$ toward the global optimizer is negligible compared to R, thus the term $\overline{a^2}/(R\kappa)^2$ can be neglected. Inserting (34) and (35) into (37) implies

$$\begin{aligned}\varphi^{*,\kappa} &\simeq N\left(1 - \sqrt{1 - \frac{2}{R\kappa}\overline{a} + \frac{\sigma^2}{\mu(R\kappa)^2}\left(N - \frac{N}{R}\overline{a}\right)}\right)\\ &= N\left(1 - \sqrt{1 - \frac{2}{R\kappa}\overline{a}\left(1 + \frac{N\sigma^2}{2\mu R^2\kappa}\right) + \frac{N\sigma^2}{\mu(R\kappa)^2}}\right)\\ &= N\left(1 - \sqrt{1 - \frac{2}{N\kappa}\frac{c_{\mu/\mu_I,\lambda}\sigma^*}{\sqrt{1+\eta^2+\frac{(\sigma^*)^2}{2N}}}\left(1 + \frac{(\sigma^*)^2}{2N\mu\kappa}\right) + \frac{(\sigma^*)^2}{N\mu\kappa^2}}\right).\end{aligned} \tag{38}$$

Substituting $C = (\sigma^*)^2/(N\mu\kappa^2)$ yields

$$\begin{aligned}\varphi^{*,\kappa} &\simeq N\left(1 - \sqrt{1 + C - \frac{2}{N\kappa}\frac{c_{\mu/\mu_I,\lambda}\sigma^*}{\sqrt{1+\eta^2+\frac{(\sigma^*)^2}{2N}}}\left(1 + \frac{\kappa}{2}C\right)}\right)\\ &= N\left(1 - \sqrt{1+C}\sqrt{1 - \frac{2}{N\kappa}\frac{c_{\mu/\mu_I,\lambda}\sigma^*}{\sqrt{1+\eta^2+\frac{(\sigma^*)^2}{2N}}}\frac{1+\frac{\kappa}{2}C}{1+C}}\right).\end{aligned} \tag{39}$$

Applying the Taylor approximation $\sqrt{1+2x} \approx 1 + x$ for $x \approx 0$ to the second square root results in

$$\begin{aligned}\varphi^{*,\kappa} &\simeq N\left(1 - \sqrt{1+C}\left(1 - \frac{c_{\mu/\mu_I,\lambda}\sigma^*}{N\kappa\sqrt{1+\eta^2+\frac{(\sigma^*)^2}{2N}}}\frac{1+\frac{\kappa}{2}C}{1+C}\right)\right)\\ &= N\sqrt{1+C}\frac{c_{\mu/\mu_I,\lambda}\sigma^*}{N\kappa\sqrt{1+\eta^2+\frac{(\sigma^*)^2}{2N}}}\frac{1+\frac{\kappa}{2}C}{1+C} - N\left(\sqrt{1+C} - 1\right)\\ &\simeq \frac{c_{\mu/\mu_I,\lambda}\sigma^*}{\kappa\sqrt{1+\eta^2+\frac{(\sigma^*)^2}{2N}}}\frac{1+\frac{\kappa}{2}C}{\sqrt{1+C}} - N\left(1 + \frac{C}{2} - 1\right),\end{aligned} \tag{40}$$

where the Taylor approximation was applied to the second expression in the last step. In the limit as $N \to \infty$ it holds that $C \to 0$. Consequently, the asymptotic normalized progress rate for the $(\mu/\mu_I, \lambda)$-ES in the noisy sphere landscape is

$$\varphi^{*,\kappa} \simeq \frac{c_{\mu/\mu_I,\lambda}\sigma^*}{\kappa\sqrt{1+\eta^2}} - \frac{(\sigma^*)^2}{2\mu\kappa^2}. \tag{41}$$

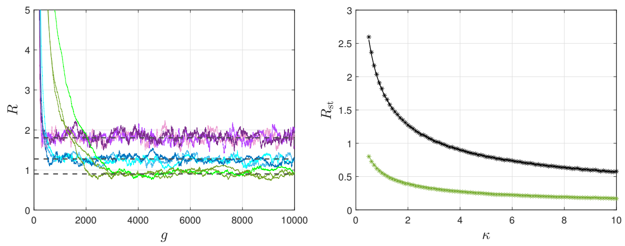

Figure 8: Left plot: R-dynamics on the noisy sphere for the $(10/10, 20)$-σSA-ES with $\tau = 1/\sqrt{2N}$, $\sigma_{\text{ES}} = 1$, and $N = 100$. Purple lines use $\kappa = 1$, blue lines $\kappa = 2$, and green lines $\kappa = 4$, respectively. Black dashed line show (45) (from top to bottom for $\kappa = 1, 2, 4$). Right plot: R_{st} depending on κ. Solid lines show (45) (green for $N = 10$ and black for $N = 100$), stars in corresponding color show the experimental values.

The gain part is reduced linearly with κ, whereas the loss part is reduced quadratically with κ.

In Fig. 7 the theoretical normalized progress rate (41) is compared with experimental values. The results are displayed as a function of the normalized mutation strength. For $N = 10$ there are clear deviations between (40) and the experimental results. As N increases, the discrepancy between the predictions (40) and the experimental values decreases. The figure shows that the range of σ^* for which positive progress is possible increases with κ. Conversely, the maximal progress is larger for smaller κ. This effect gets smaller with increasing N. For the asymptotic case, the maximum progress appears independent of κ.

In the presence of a constant noise variance an ES reaches a global or local optimum only up to a steady-state distance R_{st}. If this distance is reached, the expected progress is zero. Zeroing the normalized progress rate of the noisy sphere model (41) results in

$$\varphi^{*,\kappa} = 0 \Leftrightarrow 2\kappa\mu c_{\mu/\mu_I,\lambda} = \sqrt{(\sigma^*)^2 + \left(\sigma^*_{\text{ES}}\right)^2}. \tag{42}$$

Inserting the normalization (33) yields

$$2\kappa\mu c_{\mu/\mu_I,\lambda} = \sqrt{\sigma^2 \frac{N^2}{R^2} + \sigma^2_{\text{ES}} \frac{N^2}{4R^4}}, \tag{43}$$

which necessarily implies

$$2\kappa\mu c_{\mu/\mu_I,\lambda} \geq \sqrt{\sigma^2_{\text{ES}} \frac{N^2}{4R^4}} = \sigma_{\text{ES}} \frac{N}{2R^2} \Leftrightarrow R^2 \geq \frac{\sigma_{\text{ES}} N}{4\kappa\mu c_{\mu/\mu_I,\lambda}}, \tag{44}$$

yielding for the steady-state

$$R_{\text{st}} := \sqrt{\frac{\sigma_{\text{ES}} N}{4\kappa\mu c_{\mu/\mu_I,\lambda}}} \tag{45}$$

for the quadratic sphere provided that $\sigma \to 0$. That is, if σ is sufficiently small the parental centroid $\mathbf{y}$ approaches the expected distance R_{st} to the global optimizer when the steady-state is reached.

The left plot of Fig. 8 shows the R-dynamics of σSA-ES runs on the noisy sphere with rescaled mutations. The experimental results demonstrate that the steady-state distance decreases with increasing κ. The black dashed lines represent the theoretical prediction (45). The prediction and experimental values exhibit a strong agreement.

In the right plot of Fig. 8 R_{st} is represented as a function of κ. For each data point 100 ES runs on the noisy sphere model were executed. The experimental value of R_{st} was calculated as the mean value of R over all generations starting from a generation where the steady-state was already reached. Subsequently, the mean value was calculated over the total of 100 runs. For all values of κ the experimental values align closely with the predictions from (45).

5 Summary and Conclusion

Rescaled mutations aim to maintain a large mutation strength while reducing the step toward the new parental centroid by a factor $1/\kappa$. This work analyzed the effect of rescaled mutations on the progress rate of evolution strategies in noisy and highly multimodal scenarios.

The component-wise second-order progress rate was determined as a function of the rescaling parameter κ. The result was that the gain part scales inverse linearly whereas the loss part scales inverse quadratically with κ. The same result was also obtained for the R-dependent second-order progress rate. These findings, which are independent of the underlying fitness landscape, were applied to the progress rate of the Rastrigin function for $N \to \infty$. The results were confirmed by one-generation experiments.

The R-dependent first-order progress rate was derived for the $(\mu/\mu_I, \lambda)$-σSA-ES on the noisy sphere model with the result that the gain part scales inverse linearly and the loss part scales inverse quadratically with κ. This implies that with increasing κ the interval for the normalized noise strength, for which positive progress is possible, gets larger. On the other hand, the maximum progress gets smaller. In the limit $N \to \infty$, the second effect diminishes. By using the progress rate of the noise sphere the expected steady-state distance to the global optimizer was calculated with the result that the steady-state distance decreases with the square root of κ.

With the steady-state distance a population sizing equation can be derived that describes the scaling behavior for the population size of an ES required to achieve a positive outcome in the Rastrigin landscape, depending on the dimension and rescaling parameter. This will be the topic of a subsequent paper.

For the $(\mu/\mu_I, \lambda)$-σSA-ES, increasing κ requires an adjustment of the learning parameter τ to maintain stable behavior, a topic that will be addressed in a forthcoming paper.

Throughout the analysis, it was assumed that κ is constant. An adaptation of κ to the current phase of the search process can enhance efficiency. This has already been confirmed in experiments. A detailed analysis of this, employing the progress rates derived in this paper, is a subject for further consideration in future research.

Although rescaled mutations have demonstrated promising performance in many multimodal and noisy settings, it is unclear whether they consistently offer advantages across all problem classes, which is also a topic for future research.

Acknowledgments

This work was partially supported by the Austrian Science Fund (FWF) under grant P33702-N.

References

[1] D.V. Arnold and H.-G. Beyer. 2002. Performance Analysis of Evolution Strategies with Multi-Recombination in High-Dimensional R^N-Search Spaces Disturbed by Noise. *Theoretical Computer Science* 289 (2002), 629–647.

[2] H.-G. Beyer. 1998. Mutate large, but inherit small! On the analysis of rescaled mutations in $(\tilde{1}, \tilde{\lambda})$-ES with noisy fitness data. In *International Conference on Parallel Problem Solving from Nature*. Springer, 109–118.

[3] H.-G. Beyer. 2000. Evolutionary algorithms in noisy environments: Theoretical issues and guidelines for practice. *Computer methods in applied mechanics and engineering* 186, 2-4 (2000), 239–267.

[4] H.-G. Beyer. 2001. *The Theory of Evolution Strategies*. Springer, Heidelberg. DOI: 10.1007/978-3-662-04378-3.

[5] H.-G. Beyer and A. Melkozerov. 2014. The Dynamics of Self-Adaptive Multi-Recombinant Evolution Strategies on the General Ellipsoid Model. *IEEE Transactions on Evolutionary Computation* 18, 5 (2014), 764–778. DOI: 10.1109/TEVC.2013.2283968.

[6] M.-L. Cauwet and O. Teytaud. 2016. Noisy optimization: Fast convergence rates with comparison-based algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference 2016*. ACM, New York, 1101–1106.

[7] N. Hansen and S. Kern. 2004. Evaluating the CMA Evolution Strategy on Multimodal Test Functions. *Parallel Problem Solving from Nature* 8 (2004), 282–291.

[8] I. Loshchilov, M. Schoenauer, M. Sebag, and N. Hansen. 2014. Maximum likelihood-based online adaptation of hyper-parameters in CMA-ES. In *International Conference on Parallel Problem Solving from Nature*. Springer, 70–79.

[9] S. Meyer-Nieberg. 2007. *Self-Adaptation in Evolution Strategies*. Ph.D. Dissertation. University of Dortmund, CS Department.

[10] H. Miyazawa and Y. Akimoto. 2017. Effect of the mean vector learning rate in CMA-ES. In *Proceedings of the Genetic and Evolutionary Computation Conference*. 721–728.

[11] M. Nomura, Y. Akimoto, and I. Ono. 2024. CMA-ES with Learning Rate Adaptation. *ACM Transactions on Evolutionary Learning* (2024).

[12] M. Nomura, Akimoto Y., and Ono I. 2023. CMA-ES with Learning Rate Adaptation: Can CMA-ES with Default Population Size Solve Multimodal and Noisy Problems?. In *Proceedings of the Genetic and Evolutionary Computation Conference*. ACM.

[13] A. Omeradzic and H.-G. Beyer. 2023. Convergence Properties of the $(\mu/\mu_I, \lambda)$-ES on the Rastrigin Function. In *FOGA XVII*. ACM.

[14] A. Omeradzic and H.-G. Beyer. 2023. Progress analysis of a multi-recombinative evolution strategy on the highly multimodal Rastrigin function. *Theoretical Computer Science* 978 (2023), 114179.

[15] A. Omeradzic and H.-G. Beyer. 2025. Self-Adaptation of Multi-Recombinant Evolution Strategies on the Highly Multimodal Rastrigin Function. *IEEE Transactions on Evolutionary Computation* 29, 4 (2025), 880–890. https://doi.org/10.1109/TEVC.2024.3400857

[16] I. Rechenberg. 1994. *Evolutionsstrategie '94*. Frommann-Holzboog.

[17] L. Schönenberger and H.-G. Beyer. 2023. On a Population Sizing Model for Evolution Strategies Optimizing the Highly Multimodal Rastrigin Function. In *Proceedings of the Genetic and Evolutionary Computation Conference* (Lisbon, Portugal) *(GECCO '23)*. Association for Computing Machinery, New York, NY, USA, 848–855.

[18] L. Schönenberger and H.-G. Beyer. 2025. On a Population Sizing Model for Evolution Strategies in Multimodal Landscapes. *IEEE Transactions on Evolutionary Computation* 29, 5 (2025), 1807–1819. https://doi.org/10.1109/TEVC.2024.3419931