

Technical Report

Rastrigin Function: Quality Gain and Progress  
Rate for  $(\mu/\mu_I, \lambda)$ -ES

Amir Omeradzic and Hans-Georg Beyer

March 27, 2023

## Contents

<b>1 Rastrigin Function and Quality Gain</b>	<b>4</b>
1.1 General Introduction . . . . .	4
1.2 Expected Quality Gain and Variance . . . . .	6
1.2.1 Expected Value . . . . .	6
1.2.2 Variance . . . . .	6
1.3 Normal Approximation of Quality Gain Distribution . . . . .	10
1.4 Resulting Quality Gain Distributions . . . . .	11
1.5 $R$ -Dependent Formulation . . . . .	13
1.5.1 Exact Averaging . . . . .	13
1.5.2 Averaging Using Isotropic Random Positions . . . . .	15
1.5.3 Transition Region . . . . .	18
<b>2 Progress Rate</b>	<b>22</b>
2.1 General Derivation . . . . .	22
2.2 Large Population Approximation . . . . .	25
2.3 Defining and Expanding the Distribution Function . . . . .	26
2.3.1 Solving Integral $I_i^0$ . . . . .	28
2.3.2 Solving Integral $I_i^1$ . . . . .	29
2.3.3 Merging Results . . . . .	32
2.4 Large Dimensionality Approximation . . . . .	34
2.4.1 Applying the Approximations . . . . .	35
2.4.2 Discussion . . . . .	36
2.5 Experiments and Numerical Solutions . . . . .	37
2.5.1 Numerical Progress Rate Solution . . . . .	37
2.5.2 Overview of Approximations . . . . .	39
2.5.3 Approximations A1 and A2 . . . . .	39
2.5.4 Approximation A3 . . . . .	43
2.5.5 Approximation A3 compared to C . . . . .	47
2.6 Alternative Progress Rate via Progress Coefficient . . . . .	48

<b>3 Quadratic Progress Rate</b>	<b>52</b>
3.1 Definition . . . . .	52
3.2 Expectations via Noisy Order Statistics . . . . .	53
3.2.1 Definition and Generic Solution . . . . .	54
3.2.2 Evaluating Expectations . . . . .	55
3.3 Expectations via Large Population Approximation . . . . .	61
3.3.1 Expectation of $E^{(2)}$ . . . . .	61
3.3.2 Expectation of $E^{(1,1)}$ . . . . .	65
3.3.3 Collecting results of $E^{(2)}$ and $E^{(1,1)}$ . . . . .	70
3.4 Relation to Residual Distance (Squared) . . . . .	71
3.5 Comparing Experiments with Approximations . . . . .	72
3.6 Investigation of Loss Terms . . . . .	77
<b>4 R-Dependent Progress Rate</b>	<b>85</b>
4.1 Progress Rate Contour Maps . . . . .	88
4.2 Investigating Zero of Progress Rate . . . . .	91
4.2.1 Zero-Progress for Sphere . . . . .	92
4.2.2 Zero-Progress for Vanishing Exponentials . . . . .	92
4.2.3 Transition Point of Vanishing Exponentials . . . . .	96
4.2.4 Intersection point and population scaling . . . . .	96
4.3 Investigation of Local Attraction and Global Convergence . . . . .	102
4.3.1 Local attraction along the diagonal . . . . .	105
4.3.2 Local attraction via a probabilistic approach . . . . .	105
4.3.3 Positive gain condition . . . . .	109
4.3.4 Optimization under constant $\sigma$ . . . . .	117
4.4 Conclusion . . . . .	120
<b>5 Dynamical Systems Approach</b>	<b>122</b>
5.1 Motivation and Introduction . . . . .	122
5.2 Experimental Setup . . . . .	123
5.2.1 Measures for Averaging . . . . .	124
5.3 Comparing Real Simulations with Iteration . . . . .	127
5.3.1 High Success Probability Investigation . . . . .	127
5.3.2 Lower Success Probability Investigation . . . . .	129
5.4 Component Equipartition . . . . .	130
<b>A Expectation Values</b>	<b>136</b>
A.1 Expectation Value of $\cos(\alpha x)$ . . . . .	136
A.2 Expectation Value of $\cos^2(\alpha x)$ and $\sin^2(\alpha x)$ . . . . .	137
A.3 Expectation Values of $x \cos(\alpha x)$ and $x \sin(\alpha x)$ . . . . .	137
A.4 Expectation Value of $x^2 \cos(\alpha x)$ . . . . .	137
A.5 Expectation Value of $x^2 \cos^2(\alpha x)$ and $x^2 \sin^2(\alpha x)$ . . . . .	138
A.6 Sums Over Expected Values and Variances . . . . .	138
<b>B Large Population Identity</b>	<b>140</b>
B.1 Derivation . . . . .	140
B.2 Applications . . . . .	144
<b>C Noisy Order Statistics</b>	<b>146</b>

<b>D Identities</b>	<b>153</b>
<b>E Additional Plots</b>	<b>154</b>

# 1 Rastrigin Function and Quality Gain

## 1.1 General Introduction

For an input vector  $\mathbf{y} = (y_1, y_2, \dots, y_N)$  the Rastrigin fitness  $f$  is defined as

$$f(\mathbf{y}) = \sum_{i=1}^N f_i(y_i) = \sum_{i=1}^N y_i^2 + A - A \cos(\alpha y_i), \quad (1.1)$$

with the number of dimensions  $N$ , the oscillation strength  $A$  and a frequency parameter  $\alpha$ . The local quality gain  $Q_{\mathbf{y}}(\mathbf{x})$  at a position  $\mathbf{y}$  due to the isotropic mutation  $\mathbf{x} \sim N(0, \sigma^2)$  can be written as [5, p. 27]

$$Q_{\mathbf{y}}(\mathbf{x}) = f(\mathbf{y} + \mathbf{x}) - f(\mathbf{y}), \quad (1.2)$$

and yields in the case of fitness minimization a negative value  $Q_{\mathbf{y}} < 0$ . For independent components the fitness gain can be calculated for each  $i$ -th component of  $(Q_{\mathbf{y}}(\mathbf{x}))_i = Q_i$  such that

$$Q_{\mathbf{y}}(\mathbf{x}) = \sum_{i=1}^N Q_i(x_i). \quad (1.3)$$

Thus, the derivation can be done for a single component and we get

$$Q_i(x_i) = f_i(y_i + x_i) - f_i(y_i) \quad (1.4)$$

$$f_i(y_i) = y_i^2 + A - A \cos(\alpha y_i) \quad (1.5)$$

$$f_i(y_i + x_i) = (y_i + x_i)^2 + A - A \cos(\alpha(y_i + x_i)). \quad (1.6)$$

Now we can apply the following trigonometric identity

$$\cos(\alpha(y_i + x_i)) = \cos(\alpha y_i) \cos(\alpha x_i) - \sin(\alpha y_i) \sin(\alpha x_i). \quad (1.7)$$

This will be useful later for the calculation of the expectation value with respect to the mutation  $x_i$ . By defining two position and fitness dependent constants

$$c_i := A \cos(\alpha y_i), \text{ and } s_i := A \sin(\alpha y_i), \quad (1.8)$$

one arrives at the  $i$ -th quality gain as a function of location  $y_i$  and mutation  $x_i$

$$\begin{aligned} Q_i(x_i) &= x_i^2 + 2y_i x_i + A \cos(\alpha y_i) \\ &\quad - A \cos(\alpha y_i) \cos(\alpha x_i) + A \sin(\alpha y_i) \sin(\alpha x_i) \\ &= x_i^2 + 2y_i x_i + c_i(1 - \cos(\alpha x_i)) + s_i \sin(\alpha x_i), \end{aligned} \quad (1.9)$$

For analytic derivations of the progress rate, a linearized version of  $Q_i$  with respect to the mutation  $x_i$  will be necessary. The quality gain can be locally approximated by Taylor expanding the function at location  $y_i$  for small mutations  $x_i$  according to

$$\begin{aligned} f_i(y_i + x_i) &= f_i(y_i) + \frac{\partial f_i}{\partial y_i} x_i + \frac{1}{2} \frac{\partial^2 f_i}{\partial y_i^2} x_i^2 + O(x_i^3) \\ Q_i(x_i) &= f_i(y_i + x_i) - f_i(y_i) = \frac{\partial f_i}{\partial y_i} x_i + \frac{1}{2} \frac{\partial^2 f_i}{\partial y_i^2} x_i^2 + O(x_i^3). \end{aligned} \quad (1.10)$$

**First Order** The first derivative is calculated as

$$\begin{aligned} f'_i &:= \frac{\partial f_i}{\partial y_i} = \frac{\partial}{\partial y_i} (y_i^2 + A - A \cos(\alpha y_i)) \\ &= \frac{\partial}{\partial y_i} y_i^2 + \frac{\partial}{\partial y_i} (A - A \cos(\alpha y_i)) \\ &= 2y_i + \alpha A \sin(\alpha y_i). \end{aligned} \quad (1.11)$$

The following quantities are defined

$$k_i := 2y_i \quad (1.12)$$

$$d_i := \alpha A \sin(\alpha y_i), \quad (1.13)$$

such that

$$f'_i = k_i + d_i. \quad (1.14)$$

This decomposition was introduced to distinguish the quadratic term derivative  $k_i$  from the oscillation term derivative  $d_i$ , see also Fig. 1. These two components will reappear later during the progress rate derivations. The linearized quality gain for the first order approximation yields

$$Q_i \approx f'_i x_i = (k_i + d_i)x_i = (2y_i + \alpha A \sin(\alpha y_i))x_i. \quad (1.15)$$

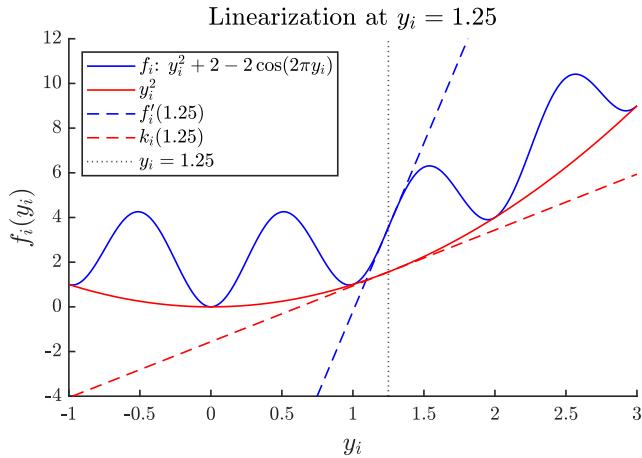


Figure 1: Rastrigin and quadratic function with their respective derivatives at  $y_i = 1.25$ . In general the linearization using  $f'_i$  is highly position dependent due to the oscillation and only useful for very small mutations. For larger mutations the spherical approximation using  $k_i$  is more suitable. This will be important during later derivations.

**Second Order** Referring to Eq. (1.10) the second derivative is evaluated

$$f''_i := \frac{\partial^2 f_i}{\partial y_i^2} = 2 + \alpha^2 A \cos(\alpha y_i), \quad (1.16)$$

such that the second order approximation yields

$$\begin{aligned} Q_i &\approx f'_i x_i + \frac{1}{2} f''_i x_i^2 \\ &\approx (2y_i + \alpha A \sin(\alpha y_i))x_i + \frac{1}{2}(2 + \alpha^2 A \cos(\alpha y_i))x_i^2. \end{aligned} \quad (1.17)$$

This approximation is used deriving Eq. (1.36), which is a sphere model of the Rastrigin function for both small mutations and small residual distance.

## 1.2 Expected Quality Gain and Variance

For the progress rate calculation the expected quality gain and variance due to the mutation strength  $\sigma$  is needed. This can be formulated as

$$E_Q = E[Q_{\mathbf{y}}(\mathbf{x})] = \sum_{i=1}^N E[Q_i] \quad (1.18)$$

$$D_Q^2 = \text{Var}[Q_{\mathbf{y}}(\mathbf{x})] = \sum_{i=1}^N \text{Var}[Q_i] = \sum_{i=1}^N E[Q_i^2] - E[Q_i]^2. \quad (1.19)$$

In order to calculate  $E_Q$  and  $D_Q^2$ , the quantities  $E[Q_i]$ ,  $E[Q_i^2]$  and  $E[Q_i]^2$  need to be evaluated, starting from the results of Eq. (1.9). We will arrive at expressions containing different expectation values of trigonometric functions, which were calculated in Appendix A and are summarized in Eqs. (1.26) and (1.27).

### 1.2.1 Expected Value

Starting the determination of quantity  $E[Q_i]$  using  $Q_i$  from Eq. (1.9) we have

$$\begin{aligned} E[Q_i] &:= E_{Q_i} = E[x_i^2 + 2y_i x_i + c_i(1 - \cos(\alpha x_i)) + s_i \sin(\alpha x_i)] \\ &= \sigma^2 + c_i(1 - E[\cos(\alpha x_i)]), \end{aligned} \quad (1.20)$$

since  $E[\sin(\alpha x_i)]$  and  $E[x_i]$  are both zero (odd parity). After reinserting the definitions for  $c_i$  and  $s_i$  and the result for  $E[\cos(\alpha x)]$  we obtain

Expected value of quality gain

$$E_Q = \sum_{i=1}^N E_{Q_i} = \sum_{i=1}^N \sigma^2 + A \cos(\alpha y_i) \left(1 - e^{-\frac{(\alpha \sigma)^2}{2}}\right) \quad (1.21)$$

### 1.2.2 Variance

From the result above  $E[Q_i]^2$  can be easily obtained

$$E[Q_i]^2 = (\sigma^2 + c_i - c_i E[\cos(\alpha x_i)])^2, \quad (1.22)$$

The more involving term  $E [Q_i^2]$  will be evaluated next. Squaring  $Q_i$  we obtain

$$\begin{aligned} Q_i^2 &= (x_i^2 + 2y_i x_i + c_i(1 - \cos(\alpha x_i)) + s_i \sin(\alpha x_i))^2 \\ &= x_i^4 + 4y_i^2 x_i^2 + c_i^2 + c_i^2 \cos^2(\alpha x_i) + s_i^2 \sin^2(\alpha x_i) + \dots \\ &\quad - 2c_i x_i^2 \cos(\alpha x_i) - 2c_i^2 \cos(\alpha x_i) + 4s_i y_i x_i \sin(\alpha x_i) + 2c_i x_i^2 + \dots \quad (1.23) \\ &\quad - 2c_i s_i \cos(\alpha x_i) \sin(\alpha x_i) - 4c_i y_i x_i \cos(\alpha x_i) + \dots \\ &\quad + 2s_i x_i^2 \sin(\alpha x_i) + 2c_i s_i \sin(\alpha x_i) + 4c_i y_i x_i + 4y_i x_i^3. \end{aligned}$$

We know that  $\cos(x)$  is an even function of  $x$ , while  $\sin(x)$  is odd. This can be easily seen by looking at the corresponding power series. We also know that for even and odd functions following product relations hold

- (even)  $\times$  (even) = (even),
- (even)  $\times$  (odd) = (odd),
- (odd)  $\times$  (odd) = (even).

Since the mutations are  $N(0, \sigma^2)$  normally distributed, all odd moments are zero. Thus one can conclude that all terms of  $Q_i^2$  containing odd powers of  $x_i$  will be zero. The expression of  $Q_i^2$  is already rearranged in a way that all odd powers are given in the last two lines of Eq. (1.23). These terms can be discarded. Thus the expected value reads

$$\begin{aligned} E [Q_i^2] &= 3\sigma^4 + 4y_i^2 \sigma^2 + c_i^2 + c_i^2 E [\cos^2(\alpha x_i)] + s_i^2 E [\sin^2(\alpha x_i)] + \dots \\ &\quad - 2c_i E [x_i^2 \cos(\alpha x_i)] - 2c_i^2 E [\cos(\alpha x_i)] + 4s_i y_i E [x_i \sin(\alpha x_i)] + 2c_i \sigma^2. \quad (1.24) \end{aligned}$$

Given Eqs. (1.22) and (1.24) the difference  $E [Q_i^2] - E [Q_i]^2$  is evaluated. After reinserting the definitions of  $c_i$  and  $s_i$  and collecting terms the result is

$$\begin{aligned} D_Q^2 &= \sum_{i=1}^N E [Q_i^2] - E [Q_i]^2 \\ &= \sum_{i=1}^N 2\sigma^4 + 4y_i^2 \sigma^2 + \dots \\ &\quad + A^2 \sin^2(\alpha y_i) \text{Var}[\sin(\alpha x_i)] \\ &\quad + A^2 \cos^2(\alpha y_i) \text{Var}[\cos(\alpha x_i)] \\ &\quad - 2A \cos(\alpha y_i) E [x^2 \cos(\alpha x_i)] \\ &\quad + 2A \sigma^2 \cos(\alpha y_i) E [\cos(\alpha x_i)] \\ &\quad + 4Ay_i \sin(\alpha y_i) E [x \sin(\alpha x_i)]. \quad (1.25) \end{aligned}$$

The terms containing trigonometric functions are calculated in Appendix A

and yield

$$\begin{aligned}
E[\cos(\alpha x)] &= \exp\left[-\frac{1}{2}(\alpha\sigma)^2\right] \\
E[\cos^2(\alpha x)] &= \frac{1}{2}\left(1 + \exp\left[-\frac{1}{2}(2\alpha\sigma)^2\right]\right) \\
E[\sin^2(\alpha x)] &= \frac{1}{2}\left(1 - \exp\left[-\frac{1}{2}(2\alpha\sigma)^2\right]\right) \\
E[x \sin(\alpha x)] &= \alpha\sigma^2 \exp\left[-\frac{1}{2}(\alpha\sigma)^2\right] \\
E[x^2 \cos(\alpha x)] &= (\sigma^2 - \alpha^2\sigma^4) \exp\left[-\frac{1}{2}(\alpha\sigma)^2\right]
\end{aligned} \tag{1.26}$$

with variances  $\text{Var}[(\cdot)] = E[(\cdot)^2] - E[(\cdot)]^2$

$$\begin{aligned}
\text{Var}[\cos(\alpha x)] &= \frac{1}{2}\left(1 + \exp\left[-\frac{1}{2}(2\alpha\sigma)^2\right]\right) - \exp[-(\alpha\sigma)^2] \\
\text{Var}[\sin(\alpha x)] &= E[\sin^2(\alpha x)] = \frac{1}{2}\left(1 - \exp\left[-\frac{1}{2}(2\alpha\sigma)^2\right]\right)
\end{aligned} \tag{1.27}$$

By inserting results (1.27) and (1.26) into (1.25) one gets

$$\begin{aligned}
D_Q^2 &= \sum_{i=1}^N 2\sigma^4 + 4y_i^2\sigma^2 + \dots \\
&\quad + A^2 \sin^2(\alpha y_i) \frac{1 - e^{-2(\alpha\sigma)^2}}{2} \\
&\quad + A^2 \cos^2(\alpha y_i) \frac{1 + e^{-2(\alpha\sigma)^2} - 2e^{-(\alpha\sigma)^2}}{2} \\
&\quad - 2A \cos(\alpha y_i) (\sigma^2 - \alpha^2\sigma^4) e^{-\frac{(\alpha\sigma)^2}{2}} \\
&\quad + 2A\sigma^2 \cos(\alpha y_i) e^{-\frac{(\alpha\sigma)^2}{2}} \\
&\quad + 4Ay_i \sin(\alpha y_i) \alpha\sigma^2 e^{-\frac{(\alpha\sigma)^2}{2}}.
\end{aligned} \tag{1.28}$$

The trigonometric identities  $\cos^2(x) + \sin^2(x) = 1$  and  $\cos^2(x) = 1/2 + \cos(2x)/2$  can be applied to further simplify the equation. After collecting terms, a simplified expression for the variance is obtained

Variance of quality gain

$$\begin{aligned}
D_Q^2 &= \sum_{i=1}^N 2\sigma^4 + 4y_i^2\sigma^2 + \frac{A^2}{2} \left[1 - e^{-(\alpha\sigma)^2}\right] \left[1 - \cos(2\alpha y_i) e^{-(\alpha\sigma)^2}\right] \\
&\quad + 2A\alpha\sigma^2 e^{-\frac{1}{2}(\alpha\sigma)^2} \left[\alpha\sigma^2 \cos(\alpha y_i) + 2y_i \sin(\alpha y_i)\right].
\end{aligned} \tag{1.29}$$

## Spherical Approximation

For vanishing oscillation amplitude or frequency, i.e.  $A = 0$  or  $\alpha = 0$ , the Rastrigin function degenerates to the spherical function  $f(\mathbf{y}) = \sum_{i=1}^N y_i^2$  with quality gain  $Q_{\text{sph}} = \sum_{i=1}^N x_i^2 + 2y_i x_i$  and we obtain the known variance relation

$$\begin{aligned} D_{\text{sph}}^2 &= \sum_{i=1}^N \text{Var}[Q_i] = \sum_{i=1}^N \mathbb{E}[Q_i^2] - \mathbb{E}[Q_i]^2 \\ &= \sum_{i=1}^N \mathbb{E}[(x_i^2 + 2y_i x_i)^2] - \mathbb{E}[x_i^2 + 2y_i x_i]^2 = \sum_{i=1}^N 2\sigma^4 + 4y_i^2\sigma^2. \\ &= 4R^2\sigma^2 + 2N\sigma^4, \end{aligned} \quad (1.30)$$

using the relation  $\sum_{i=1}^N y_i^2 = \|\mathbf{y}\|^2 = R^2$  and  $x_i \sim \mathcal{N}(0, \sigma^2)$ . Therefore the quality gain variance can be expressed as a function of the residual distance  $R$ .

## Approximation for Large Mutations and large $R$

For large mutations strengths  $\sigma$  given some  $\alpha$  such that  $(\alpha\sigma)^2/2 \gg 1$ , the last two lines of Eq. (1.29) are suppressed exponentially. This means that the trigonometric terms, i.e. the local fitness structure, has vanishing contribution for  $\sigma \rightarrow \infty$

$$\begin{aligned} D_Q^2 &\simeq \sum_{i=1}^N \left[ 2\sigma^4 + 4y_i^2\sigma^2 + \frac{A^2}{2} \right] = 2N\sigma^4 + 4R^2\sigma^2 + \frac{NA^2}{2} \\ &= D_{\text{sph}}^2 + \frac{NA^2}{2} \end{aligned} \quad (1.31)$$

using the result of Eq. (1.30). Therefore the variance for large mutations (or large  $R$ ) consists of a spherical contribution  $D_{\text{sph}}^2$  and a Rastrigin-specific term due to the oscillation defined as

$$\sigma_R^2 = \frac{NA^2}{2}. \quad (1.32)$$

Analogously, the same result is obtained for large oscillation frequency  $\alpha$  given a mutation strength  $\sigma$ .

## Approximation for Small Mutations and Small $R$

In contrast to the derivation of (1.31), a spherical model for the variance can also be derived assuming small mutation and residual distance, which corresponds to a sphere model of the global attractor depending on the fitness parameters  $A$  and  $\alpha$ .

To this end, the second order approximation for small mutations of the quality gain from Eq. (1.17) is needed. Additionally, a Taylor expansion of the derivative terms  $f'_i$  and  $f''_i$  for small  $y_i$  (and therefore small  $R$ ) has to be performed. The goal is to discard higher order terms  $O(y_i^3)$  and to relate second order terms with the radius according to  $\sum_{i=1}^N y_i^2 = R^2$ .

Starting from (1.17) the variance of a second order approximation can be evaluated as

$$\begin{aligned}
\text{Var}[Q_i] &= \text{Var}\left[f'_i x_i + \frac{1}{2} f''_i x_i^2\right] \\
&= \mathbb{E}\left[\left(f'_i x_i + \frac{1}{2} f''_i x_i^2\right)^2\right] - \mathbb{E}\left[f'_i x_i + \frac{1}{2} f''_i x_i^2\right]^2 \\
&= (f'_i)^2 \sigma^2 + \frac{3}{4} (f''_i)^2 \sigma^4 - \frac{1}{4} (f''_i)^2 \sigma^4 \\
&= (f'_i)^2 \sigma^2 + \frac{1}{2} (f''_i)^2 \sigma^4.
\end{aligned} \tag{1.33}$$

Taylor-expansion of  $f'_i(y_i)$  from (1.15) and  $f''_i(y_i)$  from (1.17) and squaring both quantities yields

$$\begin{aligned}
(f'_i(y_i))^2 &= (2y_i + \alpha A [\alpha y_i + O(y_i^3)])^2 \\
(f''_i(y_i))^2 &= (2 + \alpha^2 A [1 - (\alpha y_i)^2/2 + O(y_i^4)])^2.
\end{aligned} \tag{1.34}$$

As terms of at least  $O(y_i^3)$  are neglected, the approximation yields

$$\begin{aligned}
(f'_i(y_i))^2 &\approx (2 + \alpha^2 A)^2 y_i^2 \\
(f''_i(y_i))^2 &\approx (2 + \alpha^2 A)^2 - \alpha^4 A(2 + \alpha^2 A)y_i^2.
\end{aligned} \tag{1.35}$$

Inserting (1.35) into (1.33) and summing over  $N$  terms one gets

$$\begin{aligned}
D_Q^2 &= \sum_i^N \text{Var}[Q_i] \\
&= (2 + \alpha^2 A)^2 \sigma^2 \sum_i^N y_i^2 + \frac{\sigma^4}{2} \sum_i^N (2 + \alpha^2 A)^2 - \alpha^4 A(2 + \alpha^2 A)y_i^2 \\
&= (2 + \alpha^2 A)^2 \sigma^2 R^2 + \frac{\sigma^4}{2} (N(2 + \alpha^2 A)^2 - \alpha^4 A(2 + \alpha^2 A)R^2) \\
&= (2 + \alpha^2 A)^2 \left(\sigma^2 R^2 + \frac{N\sigma^4}{2}\right) + O(\sigma^4 R^2)
\end{aligned} \tag{1.36}$$

which is a spherical model of the Rastrigin variance for small  $R$  and small mutations by neglecting higher orders thereof. The term  $(2 + \alpha^2 A)^2$  serves as a correction of the standard sphere variance from (1.30) and setting  $\alpha = 0$  or  $A = 0$  recovers it.

Assuming constant  $\sigma^* = \sigma N/R$  in Eq. (1.36) one gets

$$D_Q^2(\sigma^*) = (2 + \alpha^2 A)^2 R^4 \left(\frac{\sigma^*}{N}\right)^2 \left(1 + \frac{\sigma^{*2}}{2N}\right) + O\left(\frac{\sigma^{*4} R^6}{N^4}\right), \tag{1.37}$$

which can be applied to both limiting cases  $N \rightarrow \infty$  (constant  $R$ ) and  $R \rightarrow 0$  (constant  $N$ ).

### 1.3 Normal Approximation of Quality Gain Distribution

When the progress of an Evolution Strategy is modeled an assumption for the distribution of the realized quality gain is needed. As described in more detail

in [5, Ch. 4], a common approach is to assume a normally distributed quality gain. This is justified for large search spaces  $N \rightarrow \infty$  by the CLT provided all  $N$  components are independent and identically distributed, which is the case for Rastrigin's function.

Starting at a position  $\mathbf{y}$  and performing one random mutation step  $\mathbf{x} \sim \sigma \mathcal{N}(0, \mathbf{I})$ , a random quality gain  $Q$  is realized. The normal approximation assumes that this mutation induced quality gain is distributed with mean  $E_Q = \mathbb{E}[Q_{\mathbf{y}}(\mathbf{x})]$  and variance  $D_Q^2 = \text{Var}[Q_{\mathbf{y}}(\mathbf{x})]$  given a location  $\mathbf{y}$  and mutation strength  $\sigma$ . Fitness parameters such as  $A$  and  $\alpha$  are also relevant and assumed to be implicitly given. Introducing a standardized random variable  $Z$  we have

$$Z = \frac{Q - E_Q}{D_Q}, \quad (1.38)$$

which translates to

$$\Pr\{Z \leq z\} = \Pr\left\{\frac{Q - E_Q}{D_Q} \leq z\right\} \xrightarrow{N \rightarrow \infty} \Phi(z) = \Phi\left(\frac{q - E_Q}{D_Q}\right). \quad (1.39)$$

given the respective target values  $z$  and  $q$ . Denoting the cumulative distribution function of the quality gain as  $P_Q(q)$  and the corresponding density  $\frac{dP_Q(q)}{dq} = p_Q(q)$  we obtain the following approximations for  $N \rightarrow \infty$

$$P_Q(q) \simeq \Phi\left(\frac{q - E_Q}{D_Q}\right) \quad (1.40)$$

$$p_Q(q) \simeq \frac{1}{D_Q} \phi\left(\frac{q - E_Q}{D_Q}\right) = \frac{1}{\sqrt{2\pi} D_Q} \exp\left[-\frac{1}{2} \left(\frac{q - E_Q}{D_Q}\right)^2\right]. \quad (1.41)$$

Within the normal approximation the inverse  $P_Q^{-1}(p)$  given some probability  $p$  can be easily obtained by using the quantile function  $\Phi^{-1}(p)$  of the normal distribution. This relation will be used later to obtain a quality gain for some given probability using

$$q = E_Q + D_Q \Phi^{-1}(p). \quad (1.42)$$

## 1.4 Resulting Quality Gain Distributions

Using the previously obtained results the quality gain distribution at location  $\mathbf{y}$  due to all  $N$  components being mutated with strength  $\sigma$  is given by Eq. (1.40) and following relations

Quality gain distribution
$P_Q(q) \simeq \Phi\left(\frac{q - E_Q}{D_Q}\right), \text{ with}$ $E_Q = \text{Eq. (1.21)}$ $D_Q^2 = \text{Eq. (1.29)}$

During the progress rate derivation for the  $i$ -th component, a quality gain distribution is needed conditioned on the  $i$ -th mutation component  $x_i$  being fixed, and  $N - 1$  components being mutated. The distribution function now changes to

$$P_Q(q|x_i) = \Phi\left(\frac{q - \mathbb{E}[Q_y|x_i]}{\sqrt{\text{Var}[Q_y|x_i]}}\right), \quad (1.44)$$

for which expected value and variance are evaluated. First the quality gain from Eq. (1.3) is split into two terms

$$Q_y(x_i, (\mathbf{x})_{j \neq i}) = \sum_{i=1}^N Q_i(x_i) = Q_i(x_i) + \sum_{j \neq i} Q_j(x_j). \quad (1.45)$$

Now the expected value and variance with respect to  $x_i$  fixed and  $N - 1$  variable components (denoted by index  $j$ ) can be taken.

**Expected Value** The expected value  $\mathbb{E}[Q_i] = Q_i$  remains constant and we obtain

$$\begin{aligned} \mathbb{E}[Q_y|x_i] &=: E_{Q|x_i} = Q_i + \sum_{j \neq i} \mathbb{E}[Q_j] \\ &= x_i^2 + 2y_i x_i + c_i(1 - \cos(\alpha x_i)) + s_i \sin(\alpha x_i) \\ &\quad + \sum_{j \neq i} \sigma^2 + A \cos(\alpha y_i) \left(1 - e^{-\frac{(\alpha \sigma)^2}{2}}\right) \\ &= x_i^2 + 2y_i x_i + c_i(1 - \cos(\alpha x_i)) + s_i \sin(\alpha x_i) + E_i, \end{aligned} \quad (1.46)$$

with  $E_i$  defined as the expectation over  $N - 1$  terms excluding the  $i$ -th

$$E_i := \sum_{j \neq i} \sigma^2 + A \cos(\alpha y_i) \left(1 - e^{-\frac{(\alpha \sigma)^2}{2}}\right). \quad (1.47)$$

For the progress rate determination an analytic integration will be done over the  $i$ -th mutation component. This will require linearization of Eq. (1.46) with respect to  $x_i$ . Using result (1.15) the approximation yields

$$\begin{aligned} E_{Q|x_i} &= x_i^2 + 2y_i x_i + c_i(1 - \cos(\alpha x_i)) + s_i \sin(\alpha x_i) + E_i \\ &\approx f'_i x_i + E_i. \end{aligned} \quad (1.48)$$

**Variance** Returning to Eq. (1.45), the variance of a constant  $\text{Var}[Q_i] = 0$  and we have

$$\text{Var}[Q_y|x_i] =: D_i^2 = \sum_{j \neq i} \text{Var}[Q_j], \quad (1.49)$$

with  $D_i^2$  denoting the variances of  $N - 1$  variable terms excluding the  $i$ -th, see also Eq. (1.29). The results can be summarized as follows

Quality gain distribution given  $x_i$

$$\begin{aligned}
 P_Q(q|x_i) &= \Phi\left(\frac{q - E_{Q|x_i}}{D_i}\right), \text{ with} \\
 E_{Q|x_i} &= \begin{cases} \text{Non-linear Eq. (1.46),} \\ \text{Linearized Eq. (1.48),} \end{cases} \\
 D_i &= \text{Eq. (1.49).}
 \end{aligned} \tag{1.50}$$

## 1.5 $R$ -Dependent Formulation

### 1.5.1 Exact Averaging

During later derivations, an  $R$ -dependent formulation of the progress rate will be needed. This will be essential for the study of the convergence behavior, where the overall residual distance  $R = \|\mathbf{y}\|$  is more important than the individual position values of  $\mathbf{y}$ , assuming that the global optimizer is approached in an isotropic way.

The fitness function, its mutation induced variance and the later derived progress rate are all quantities depending on the location  $\mathbf{y}$ . Since the Rastrigin fitness contains the cosine of  $y_i$ , the derived quantities also contain trigonometric functions of the position. The approach of formulating the expressions as  $R$ -dependent will be shown on the Rastrigin fitness (1.1), but are also applicable to the variance and later the progress rate.

Given a certain  $R^2 = \sum_{i=1}^N y_i^2$ , the Rastrigin function can be written as

$$\begin{aligned}
 f(\mathbf{y}, R) &= \sum_{i=1}^N y_i^2 + A - A \cos(\alpha y_i) = R^2 + NA - A \sum_{i=1}^N \cos(\alpha y_i) \\
 &= R^2 + NA + T(\mathbf{y}),
 \end{aligned} \tag{1.51}$$

with

$$T(\mathbf{y}) = -A \sum_{i=1}^N \cos(\alpha y_i). \tag{1.52}$$

The main issue of  $T(y_i)$  is that given some  $R$  there are different Cartesian realizations  $\mathbf{y}$  thereof leading to different fitness values. As we are interested in quantities which are expected values, such as the progress rate or the mean value dynamics of Sec. 5, one is interested in the strategy's expected behavior and the function average is a natural solution candidate. Therefore the first approach is averaging the function  $T$  over the (hyper-)sphere with  $R = \|\mathbf{y}\|$ . Another argumentation assuming larger dimensionality  $N$  is that given  $R$  the individual fluctuations of the positions should be negligible, if the sum is taken over many components. Canceling effects should occur, as some components contribute positively and others negatively to the overall average fitness level.

Averaging (1.52) could be achieved for  $N = 1$  and  $N = 2$  in an exact way. For  $N \geq 3$  no closed form solution of the resulting integrals could be found to this point.

Starting with  $N = 1$  and requiring  $R = \|y_1\|$ , the two possible solutions are  $y_1 = \pm R$ . The average of  $T$  (denoted with overline) over two points therefore yields

$$\begin{aligned}\overline{T}(R) &= -\frac{A}{2} \sum_{y_i=\pm R} \cos(\alpha y_i) = -\frac{A}{2} (\cos(\alpha R) + \cos(-\alpha R)) \\ &= -A \cos(\alpha R).\end{aligned}\quad (1.53)$$

The calculation of the average for  $N > 1$  is more involved, as one has to integrate over the  $(N-1)$ -dimensional sphere-surface  $S_N$  (embedded in  $N$  dimensions) using some parametrization  $s$  according to

$$\overline{T}(R) = \frac{1}{S_N} \int_{\|\mathbf{y}\|=R} T(\mathbf{y}) \, ds, \quad (1.54)$$

with the sphere surface area

$$S_N = \frac{2\pi^{N/2} R^{N-1}}{\Gamma(N/2)}. \quad (1.55)$$

For  $N = 2$  the parametrization is defined as  $(y_1, y_2) = (R \cos \phi, R \sin \phi)$  with derivative vector  $\frac{d(y_1, y_2)}{d\phi} = (-R \sin \phi, R \cos \phi)$  on  $\phi \in [0, 2\pi]$ . Additionally one has  $S_2 = 2\pi R$ . Therefore, inserting the parametrization into (1.54) and using path element length  $\left\| \frac{d(y_1, y_2)}{d\phi} \right\| = R$  one has

$$\begin{aligned}\overline{T}(R) &= \frac{-A}{2\pi R} \int_0^{2\pi} T(y_1(R, \phi), y_2(R, \phi)) \left\| \frac{d(y_1, y_2)}{d\phi} \right\| \, d\phi \\ &= \frac{-A}{2\pi} \int_0^{2\pi} [\cos(\alpha R \cos \phi) + \cos(\alpha R \sin \phi)] \, d\phi.\end{aligned}\quad (1.56)$$

The integrals obtained in (1.56) can be solved in terms of Bessel functions of the first kind  $J_n(x)$  with  $n \geq 0$  by applying the integral identity [1, p. 360, 9.1.18]

$$J_0(x) = \frac{1}{\pi} \int_0^\pi \cos(x \sin t) \, dt = \frac{1}{\pi} \int_0^\pi \cos(x \cos t) \, dt. \quad (1.57)$$

Due to the periodicity, integrating  $\cos t$  and  $\sin t$  over  $[0, \pi]$  yields the same contribution as the integration over  $[\pi, 2\pi]$ . Thus, identity (1.57) is reformulated by including the second interval as

$$2J_0(x) = \frac{1}{\pi} \int_0^{2\pi} \cos(x \sin t) \, dt = \frac{1}{\pi} \int_0^{2\pi} \cos(x \cos t) \, dt. \quad (1.58)$$

Comparing (1.56) with (1.58) and setting  $x = \alpha R$ , the expression (1.56) is evaluated as

$$\overline{T}(R) = \frac{-A}{2\pi} [2\pi J_0(\alpha R) + 2\pi J_0(\alpha R)] = -2A J_0(\alpha R). \quad (1.59)$$

The Rastrigin fitness averaged over  $R$  can therefore be summarized as

$$F(R) = R^2 + NA + \overline{T}(R) = \begin{cases} R^2 + A(1 - \cos(\alpha R)) & \text{for } N = 1 \\ R^2 + 2A(1 - J_0(\alpha R)) & \text{for } N = 2. \end{cases} \quad (1.60)$$

An exemplary evaluation of (1.60) is shown in Fig. 2. For  $N \geq 3$  no closed form solution is available at this point. Deriving the progress rate, many approximations rely on the large dimensionality assumption for  $N$ , such that the analytic approach is unfeasible, if no generic solution is available. Furthermore, there are additional terms to be evaluated when averaging other functions such as the variance, e.g.  $T(y_i) = y_i \sin(\alpha y_i)$  in result (1.29). Therefore a different approach is taken for the  $R$ -dependent formulation for large  $N$ .

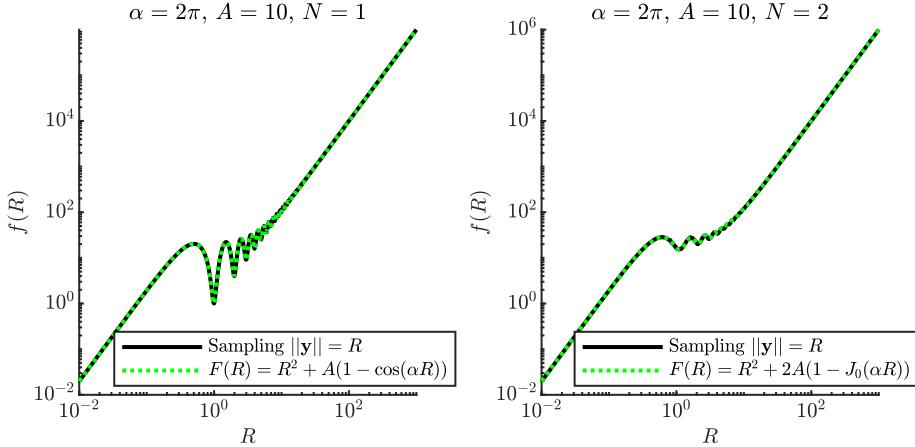


Figure 2: Average Rastrigin fitness value as function of  $R$  for  $N = 1$  (left) and  $N = 2$  (right). The green curves are the analytic results from (1.60). The black curves are experimentally obtained by choosing randomly and isotropically  $y_i \sim \mathcal{N}_i(0, 1)$ , normalizing to be  $\|\mathbf{y}\| = R$  and averaging over 1000 trials.

### 1.5.2 Averaging Using Isotropic Random Positions

For a different averaging approach, one can assume that for any given  $R$  and  $N$  the positions  $y_i$  should be independent and normally distributed around the optimizer with zero mean and standard deviation  $\sigma_y$ , such that

$$y_i \sim \sigma_y \mathcal{N}(0, 1). \quad (1.61)$$

Following the requirement  $R^2 = \sum_i y_i^2$ , the property should hold in expectation

$$R^2 = \mathbb{E} \left[ \sum_{i=1}^N y_i^2 \right] = \sigma_y^2 \mathbb{E} \left[ \sum_{i=1}^N \mathcal{N}_i^2(0, 1) \right] = \sigma_y^2 \mathbb{E} [\chi_N^2] = \sigma_y^2 N. \quad (1.62)$$

It was used that the sum over  $N$  independent standard normally distributed variables squared is equal to the chi-squared distributed variable  $\chi_N^2$  with  $\mathbb{E} [\chi_N^2] = N$ . Solving (1.62) for  $\sigma_y$ , expression (1.61) can be rewritten as

$$y_i \sim \frac{R}{\sqrt{N}} \mathcal{N}(0, 1). \quad (1.63)$$

Having locations distributed according to (1.63) can be justified for an ES operating with large mutation strengths relative to the sphere function. Large

(normalized) mutations are obtained on the sphere for  $\sigma^*$  smaller (but close) to the second zero of the progress rate, see e.g. [5, Eq. (6.54)]. The limit of large mutations is also important on the Rastrigin function to minimize the probability of local convergence.

Having established (1.63) the idea will be to approximate the sum over the trigonometric  $y_i$ -dependent terms of fitness (1.1) and variance (1.29) as random variables. In this way, the oscillation part is replaced by a variable with expected value and fluctuations. Considering the fitness as an example, one can define a new random number  $Y$  for the sum over the cosines

$$Y := \sum_{i=1}^N \cos(\alpha y_i), \quad (1.64)$$

such that the fitness is rewritten containing a random part

$$f(R, Y) = R^2 + NA - AY. \quad (1.65)$$

As the terms within the sum are i.i.d. variates, the CLT can be applied according to

$$\frac{Y - \mathbb{E}[Y]}{\sqrt{\text{Var}[Y]}} \xrightarrow{N \rightarrow \infty} \mathcal{N}(0, 1), \quad (1.66)$$

with asymptotic equality  $Y \sim \mathcal{N}(\mathbb{E}[Y], \text{Var}[Y])$ . Furthermore it will be shown that  $\mathbb{E}[Y]$  scales with  $N$  and the standard deviation  $\sqrt{\text{Var}[Y]}$  only with  $\sqrt{N}$ . Hence, for large  $N$  the fluctuations of  $Y$  are negligible with following limit

$$\frac{\sqrt{\text{Var}[Y]}}{\mathbb{E}[Y]} \xrightarrow{N \rightarrow \infty} 0, \quad (1.67)$$

such that  $Y \sim \mathbb{E}[Y] + \sqrt{\text{Var}[Y]} \mathcal{N}(0, 1)$  can be (deterministically) approximated as

$$Y \approx \mathbb{E}[Y]. \quad (1.68)$$

Applying (1.68) by taking the expected value of (1.65) yields

$$\begin{aligned} f(R) &:= \mathbb{E}[f(R, Y)|R] \\ &= R^2 + NA - A\mathbb{E}[Y]. \end{aligned} \quad (1.69)$$

The expression  $f(R)$  will be a function of  $R$  with removed  $y_i$ -dependency, which is desired. This approach is analogously applicable to the trigonometric terms of variance (1.29), which are summed over  $N$  components.

The relevant sums to be evaluated are over the terms  $\cos(\alpha y_i)$ ,  $\cos(2\alpha y_i)$  and  $y \sin(\alpha y_i)$ , respectively. For each sum the limit behavior (1.67) needs to be investigated.

For the expectations and variances over the sums of trigonometric terms the results from Appendix A.6 with  $y_i \sim \mathcal{N}(0, \sigma_y^2)$  with  $\sigma_y = \frac{R}{\sqrt{N}}$  are applicable.

Given aforementioned results, ratio (1.67) can be evaluated as follows

$$\frac{\text{Var} \left[ \sum_{i=1}^N \cos(\alpha y_i) \right]^{\frac{1}{2}}}{\text{E} \left[ \sum_{i=1}^N \cos(\alpha y_i) \right]} = \frac{\left( \frac{1}{2} + \frac{1}{2} e^{-\frac{1}{2} \frac{(2\alpha R)^2}{N}} - e^{-\frac{(\alpha R)^2}{N}} \right)^{\frac{1}{2}}}{\sqrt{N} e^{-\frac{1}{2} \frac{(\alpha R)^2}{N}}} \xrightarrow{N \rightarrow \infty} 0 \quad (1.70)$$

$$\frac{\text{Var} \left[ \sum_{i=1}^N \cos(2\alpha y_i) \right]^{\frac{1}{2}}}{\text{E} \left[ \sum_{i=1}^N \cos(2\alpha y_i) \right]} = \frac{\left( \frac{1}{2} + \frac{1}{2} e^{-\frac{1}{2} \frac{(4\alpha R)^2}{N}} - e^{-\frac{(2\alpha R)^2}{N}} \right)^{\frac{1}{2}}}{\sqrt{N} e^{-2 \frac{(\alpha R)^2}{N}}} \xrightarrow{N \rightarrow \infty} 0 \quad (1.71)$$

$$\begin{aligned} & \frac{\text{Var} \left[ \sum_{i=1}^N y_i \sin(\alpha y_i) \right]^{\frac{1}{2}}}{\text{E} \left[ \sum_{i=1}^N y_i \sin(\alpha y_i) \right]} \\ &= \frac{\left( \frac{1}{2} - \frac{1}{2} e^{-\frac{1}{2} \frac{(2\alpha R)^2}{N}} + 2\alpha^2 \frac{R^2}{N} e^{-\frac{1}{2} \frac{(2\alpha R)^2}{N}} - \alpha^2 \frac{R^2}{N} e^{-\frac{(\alpha R)^2}{N}} \right)^{\frac{1}{2}}}{\alpha R e^{-\frac{1}{2} \frac{(\alpha R)^2}{N}}} \xrightarrow{N \rightarrow \infty} 0. \end{aligned} \quad (1.72)$$

Note that the limit considerations hold for constant  $R$  and for a scaling relation  $R^2 = N$ , see Eq. (1.62). For constant  $R$  the exponential factors yield “1” in the limit  $N \rightarrow \infty$ , such that the numerators of (1.70), (1.71), and (1.72) vanish. The denominators of (1.70) and (1.71) are also suppressing the ratio with  $O(1/\sqrt{N})$ , while the denominator of (1.72) remains constant. For  $R^2 = N$  the exponentials yield non-zero  $\alpha$ -dependent values and all ratios vanish with  $O(1/\sqrt{N})$ .

Therefore the approximation (1.68) is justified for all corresponding terms of fitness and variance. Applying the expected value (A.32) to expression (1.69) yields the  $R$ -dependent fitness for large  $N$

$$f(R) \simeq R^2 + NA \left( 1 - e^{-\frac{1}{2} \frac{(\alpha R)^2}{N}} \right). \quad (1.73)$$

Analogously, the expected value (A.32) can be applied to the sum over the cosine terms of the  $\mathbf{y}$ -dependent quality gain expectation value (1.21). Therefore one obtains the  $R$ -dependent formulation as

$$E_Q(R, \sigma) = N\sigma^2 + NAe^{-\frac{1}{2} \frac{(\alpha R)^2}{N}} \left( 1 - e^{-\frac{(\alpha \sigma)^2}{2}} \right). \quad (1.74)$$

Furthermore, one can apply (A.32), (A.33), and (A.34) to the sum over the trigonometric terms of variance (1.29), which yields the  $R$ -dependent variance formula

$$\begin{aligned} D_Q^2(R, \sigma) & \simeq 4R^2\sigma^2 + 2N\sigma^4 + \frac{NA^2}{2} \left[ 1 - e^{-(\alpha \sigma)^2} \right] \left[ 1 - e^{-(\alpha \sigma)^2} e^{-2 \frac{(\alpha R)^2}{N}} \right] \\ & + 2A\alpha\sigma^2 e^{-\frac{1}{2}(\alpha \sigma)^2} \left[ N\alpha\sigma^2 e^{-\frac{1}{2} \frac{(\alpha R)^2}{N}} + 2\alpha R^2 e^{-\frac{1}{2} \frac{(\alpha R)^2}{N}} \right] \\ & = 4R^2\sigma^2 + 2N\sigma^4 + \frac{NA^2}{2} \left[ 1 - e^{-(\alpha \sigma)^2} \right] \left[ 1 - e^{-\alpha^2 \left( \sigma^2 + 2 \frac{R^2}{N} \right)} \right] \\ & + 2NA\alpha^2\sigma^2 e^{-\frac{\alpha^2}{2} \left( \sigma^2 + \frac{R^2}{N} \right)} \left[ \sigma^2 + 2 \frac{R^2}{N} \right]. \end{aligned} \quad (1.75)$$

Figures 3 and 4 show the simulated average function value compared to the results (1.73) and (1.75), respectively.

From result (1.75) one can easily see, that the following inequality holds for the sphere variance derived in (1.30)

$$D_{\text{sph}}^2(R, \sigma) \leq D_Q^2(R, \sigma), \quad (1.76)$$

which is valid within the limit  $N \rightarrow \infty$  applying the CLT. Setting  $A = 0$  or  $\alpha = 0$  recovers the sphere variance. Different approximations are compared in Fig. 5.

### 1.5.3 Transition Region

From Figures 3 and 4 one can observe that fitness and quality gain variance exhibit a transitional region between two log-linear regimes. Within the two regimes (small and large  $R$ , respectively) the functions behave sphere-like with different prefactors. For small  $R$  the global basin is dominating and can be modeled by a quadratic function. For large  $R$  the overall quadratic structure of the fitness is dominating. Therefore within the transitional region the influence of the local minima landscape is significant.

Given fitness result of (1.73), for small  $R$  a Taylor expansion can be performed giving

$$\begin{aligned} f(R) &= R^2 + NA \left( 1 - e^{-\frac{1}{2} \frac{(\alpha R)^2}{N}} \right) \\ &= R^2 + NA \left( 1 - \left[ 1 - \frac{(\alpha R)^2}{2N} + O(R^4) \right] \right) \\ &= R^2 + \frac{A\alpha^2 R^2}{2} + O(R^4) \\ &= R^2 \left( 1 + \frac{A\alpha^2}{2} \right) + O(R^4). \end{aligned} \quad (1.77)$$

Conversely, for large  $R$  the exponential can be neglected. Additionally, the term  $NA$  is negligible compared to  $R^2$ . The  $R$ -dependent Rastrigin fitness can therefore be quadratically approximated as

$$f(R) = \begin{cases} R^2 \left( 1 + \frac{A\alpha^2}{2} \right) & \text{for small } R \\ R^2 & \text{for large } R. \end{cases} \quad (1.78)$$

The exponential function can be interpreted as the source term of the transition. One way of defining the transition at residual distance  $R_{tr}$  is to look at the attenuation of the exponential function  $e^{-\delta}$  with some value  $\delta > 0$ , such that

$$\begin{aligned} e^{-\delta} &\stackrel{!}{=} e^{-\frac{1}{2} \frac{(\alpha R_{tr})^2}{N}}, \quad \text{such that} \\ R_{tr} &= \frac{\sqrt{2\delta N}}{\alpha}. \end{aligned} \quad (1.79)$$

The result of (1.79) is shown in Fig. 3. The scaling relation  $R \sim \sqrt{N}$  will be used throughout the report, as it describes how the “interesting” region with high influence of local minima scales as dimensionality  $N$  is increased.

For the variance  $D_Q^2$  in Fig. 4, a transitional region can also be observed for given constant  $\sigma^*$ . The spherical limits were already calculated in (1.31) and (1.36). Inserting  $\sigma = \sigma^* R/N$  into (1.75) yields the variance as a function of  $\sigma^*$  and  $R$  giving

$$\begin{aligned} D_Q^2(R, \sigma^*) &= 4R^2 \left( \frac{\sigma^* R}{N} \right)^2 + 2N \left( \frac{\sigma^* R}{N} \right)^4 \\ &\quad + \frac{NA^2}{2} \left[ 1 - e^{-(\alpha \frac{\sigma^* R}{N})^2} \right] \left[ 1 - e^{-\alpha^2 \left( \left( \frac{\sigma^* R}{N} \right)^2 + 2 \frac{R^2}{N} \right)} \right] \\ &\quad + 2NA\alpha^2 \left( \frac{\sigma^* R}{N} \right)^2 e^{-\frac{\alpha^2}{2} \left( \left( \frac{\sigma^* R}{N} \right)^2 + \frac{R^2}{N} \right)} \left[ \left( \frac{\sigma^* R}{N} \right)^2 + 2 \frac{R^2}{N} \right], \end{aligned} \quad (1.80)$$

such that after simplification one obtains the Rastrigin variance in terms of  $\sigma^*$  as

$$\begin{aligned} D_Q^2(R, \sigma^*) &= 4R^4 \left( \frac{\sigma^*}{N} \right)^2 \left[ 1 + \frac{\sigma^{*2}}{2N} \right] \\ &\quad + \frac{NA^2}{2} \left[ 1 - e^{-(\alpha R \sigma^*)^2} \right] \left[ 1 - e^{-(\alpha R)^2 \left[ \left( \frac{\sigma^*}{N} \right)^2 + \frac{2}{N} \right]} \right] \\ &\quad + 2NA\alpha^2 R^4 \left( \frac{\sigma^*}{N} \right)^2 \left[ \left( \frac{\sigma^*}{N} \right)^2 + \frac{2}{N} \right] e^{-\frac{(\alpha R)^2}{2} \left[ \left( \frac{\sigma^*}{N} \right)^2 + \frac{1}{N} \right]}. \end{aligned} \quad (1.81)$$

The transition point for result (1.81) is also defined in terms of an exponential factor in such a way, that for  $\sigma^* = 0$  result (1.79) is obtained again. Considering the exponential factor of the last term of (1.81) one can define

$$\begin{aligned} e^{-\delta} &\stackrel{!}{=} e^{-\frac{(\alpha R_{tr})^2}{2} \left[ \left( \frac{\sigma^*}{N} \right)^2 + \frac{1}{N} \right]}, \quad \text{such that} \\ R_{tr} &= \frac{\sqrt{2\delta N}}{\alpha} \frac{1}{\sqrt{1 + \sigma^{*2}/N}}. \end{aligned} \quad (1.82)$$

Hence,  $R_{tr}$  of the variance is additionally a function of the (normalized) mutation strength. The exponential from (1.82) will also appear later in the  $R$ -dependent progress rate result (4.5), which backs up its choice. Within the limit  $N \rightarrow \infty$ , assuming constant  $\sigma^*$  and  $\delta = 1$ , the same scaling  $R_{tr} = \sqrt{2N}/\alpha$  is observed as in (1.79). The result of (1.82) is shown in Fig. 4.

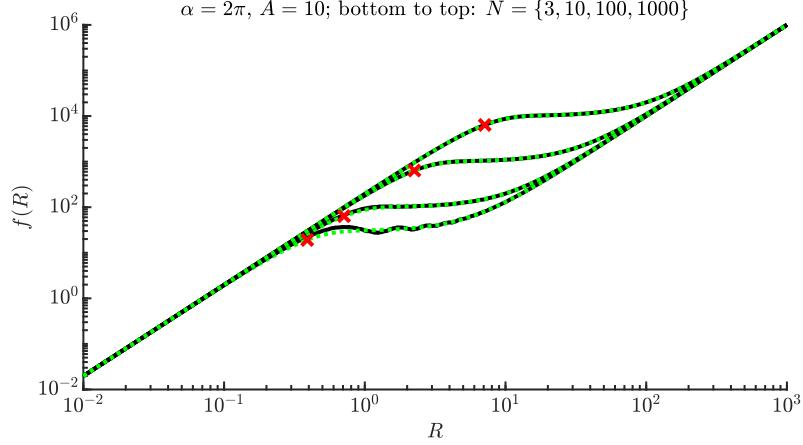


Figure 3: Rastrigin fitness value as function of  $R$  using expected value for  $\sum_i \cos(\alpha y_i)$  for  $N = \{3, 10, 100, 1000\}$ , bottom to top. The green dotted curves are the results from (1.73). The black curves are experimentally obtained by sampling randomly  $y_i \sim \mathcal{N}_i(0, 1)$ , then normalizing by  $R$  and averaging over 1000 trials. The approximation quality is very good even for moderately large values of  $N$ . For small  $R$  only the  $R^2$  term is relevant. For large  $R$  the offset  $NA$  is negligible, such that the curves appear very close to each other, see also (1.78). Applying Eq. (1.79) with  $\delta = 1$ , the red transition points are at  $R_{tr} = \{0.39, 0.71, 2.25, 7.11\}$ .

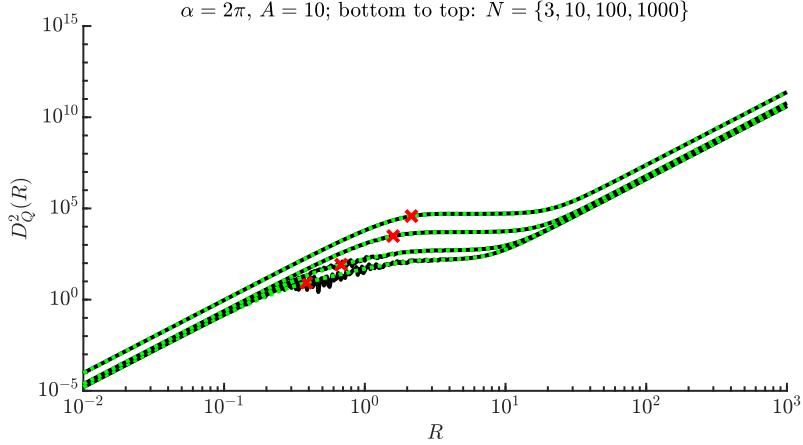


Figure 4: Variance  $D_Q^2$  as function of  $R$  for constant  $\sigma^*$  using expected values for sum over  $y_i$ -dependent trigonometric terms. The value  $\sigma^*/N = 1/10$  is constant for increasing  $N = \{3, 10, 100, 1000\}$ , bottom to top. It was chosen due to the idea that larger  $N$  requires increasing  $\mu$  leading to increased mutation strength. Value  $1/10$  was chosen for displaying purposes. The green dotted curves are the results from (1.75). The black curves are experimentally obtained. As expected, the approximation quality is better for larger  $N$ . Overall, the approximation quality is very good. Applying Eq. (1.82) with  $\delta = 1$ , the red transition points are at  $R = \{0.38, 0.68, 1.59, 2.15\}$

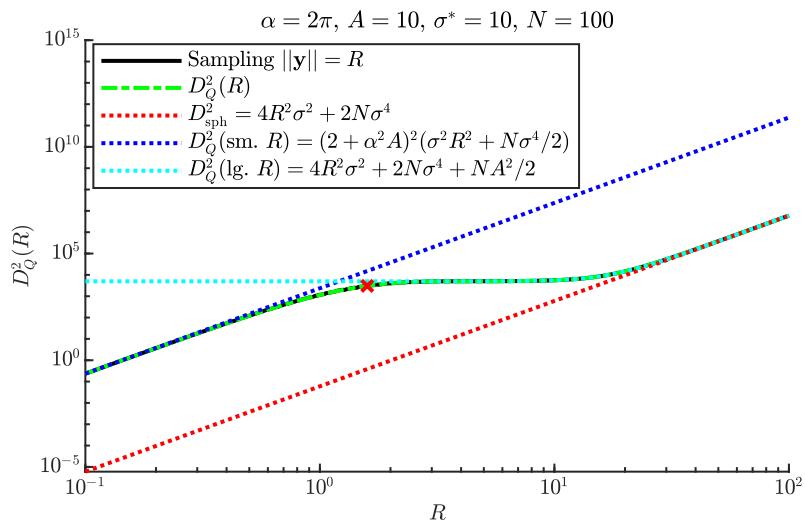


Figure 5: Variance approximations as functions of  $R$  for exemplary values  $\sigma^* = 10$  and  $N = 100$  (see also Fig. 4). Comparing: sampling [black], result (1.75) [green], sphere variance (1.30) [red], variance for small  $R$  (1.36) [blue], and variance for large  $R$  including  $NA^2/2$  using (1.31) [cyan].  $D_{\text{sph}}^2$  and  $D_Q^2(\text{sm. } R)$  are both spherical  $R^2$ -dependent variances with different prefactors. For large and moderate  $R$ , the cyan curve with additional term  $NA^2/2$  (exponentials neglected) captures a large part of the transition region, but fails around the transition point (red cross), where exponential terms are significant.

## 2 Progress Rate

The progress rate between two generations for the  $i$ -th component  $y_i$  is defined as [5, p. 29]

$$\varphi_i = E \left[ y_i^{(g)} - y_i^{(g+1)} \mid \mathbf{y}^{(g)}, \sigma^{(g)} \right], \quad (2.1)$$

given the position  $\mathbf{y}^{(g)}$  and mutation strength  $\sigma^{(g)}$  at generation  $g$ . The conditional quantities  $\mathbf{y}^{(g)}$  and  $\sigma^{(g)}$  will be dropped for better readability.

### 2.1 General Derivation

After applying a mutation  $\mathbf{x} \sim \sigma \mathcal{N}(0, \mathbf{1})$  to the current parental state  $\mathbf{y}^{(g)}$  and selecting the best individuals, the position update in search space is performed using following relation

$$\mathbf{y}^{(g+1)} = \frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{y}^{(g)} + \mathbf{x}_{m;\lambda}) = \mathbf{y}^{(g)} + \frac{1}{\mu} \sum_{m=1}^{\mu} \mathbf{x}_{m;\lambda}, \quad (2.2)$$

where  $\mathbf{x}_{m;\lambda}$  denotes the mutation vector of the  $m$ -th best offspring after selection. Considering the position component  $y_i$  and abbreviating the  $i$ -th mutation component as

$$x_{m;\lambda} := (\mathbf{x}_{m;\lambda})_i \quad (2.3)$$

the expected value can be taken

$$E \left[ y_i^{(g+1)} \right] = y_i^{(g)} + E \left[ \frac{1}{\mu} \sum_{m=1}^{\mu} x_{m;\lambda} \right] = y_i^{(g)} + \frac{1}{\mu} \sum_{m=1}^{\mu} E [x_{m;\lambda}]. \quad (2.4)$$

Going back to (2.1), the progress rate definition is reformulated using the expected value of the order statistics of  $x$

$$\varphi_i = E \left[ y_i^{(g)} - y_i^{(g+1)} \right] = y_i^{(g)} - E \left[ y_i^{(g+1)} \right] \quad (2.5)$$

$$= -\frac{1}{\mu} \sum_{m=1}^{\mu} E [x_{m;\lambda}]. \quad (2.6)$$

The expected value of a random variable  $X$  with density  $p_X(x)$  can be calculated as

$$E [X] = \int_{-\infty}^{\infty} x p_X(x) dx. \quad (2.7)$$

In our case the probability density incorporates the induced order statistics of the  $\mu$  best offspring given the location  $\mathbf{y}^{(g)} = \mathbf{y}$ . Using the definition of the expected value, the progress rate can be rewritten as

$$\varphi_i = -\frac{1}{\mu} \sum_{m=1}^{\mu} \int_{-\infty}^{\infty} x_i p_{m;\lambda}(x_i | \mathbf{y}) dx_i, \quad (2.8)$$

with  $x_i$  denoting the integration over  $i$ -th component. The density  $p_{m;\lambda}(x_i | \mathbf{y})$  of the  $m$ -th best individual will be established using subsequent arguments.

Mutations are distributed normally with zero mean and variance  $\sigma^2$  according to the normal density

$$p_x(x_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x_i}{\sigma}\right)^2\right]. \quad (2.9)$$

Given a position  $\mathbf{y}$  and mutation  $x_i$ , a random quality gain value  $Q$  is distributed according to a conditional probability density  $p_Q(q|x_i, \mathbf{y})$  given by

$$p_Q(q|x_i, \mathbf{y}) = \frac{dP_Q(q|x_i, \mathbf{y})}{dq}, \quad (2.10)$$

an approximation of which was presented in Eq. (1.50).

Given that the  $m$ -th best individual attains a quality gain within  $[q, q + dq]$ , we must have  $m - 1$  better individuals having a smaller quality value with probability  $[\Pr\{Q \leq q\}]^{m-1} = [P_Q(q)]^{m-1}$ , and  $\lambda - m$  individuals having a larger value with  $[\Pr\{Q > q\}]^{\lambda-m} = [1 - P_Q(q)]^{\lambda-m}$ . To account for all relevant combinations we have  $\frac{\lambda!}{(m-1)!(\lambda-m)!}$ , where  $1/(m-1)!$  and  $1/(\lambda-m)!$  exclude the irrelevant combinations among the two groups of better and worse individuals, respectively.

The conditional density for the  $m$ -th individual as a function of the quality gain  $q$  yields

$$\begin{aligned} p_{Q;m;\lambda}(q|x_i, \mathbf{y}) &= \frac{\lambda!}{(m-1)!(\lambda-m)!} p_Q(q|x_i, \mathbf{y}) \\ &\times P_Q(q|\mathbf{y})^{m-1} [1 - P_Q(q|\mathbf{y})]^{\lambda-m}. \end{aligned} \quad (2.11)$$

By integrating over all attainable quality gain values  $[q_l, q_u]$ , one arrives at the density

$$p_{m;\lambda}(x_i|\mathbf{y}) = p_x(x_i) \int_{q_l}^{q_u} p_{Q;m;\lambda}(q|x_i, \mathbf{y}) dq \quad (2.12)$$

$$\begin{aligned} p_{m;\lambda}(x_i|\mathbf{y}) &= p_x(x_i) \frac{\lambda!}{(m-1)!(\lambda-m)!} \\ &\times \int_{q_l}^{q_u} p_Q(q|x_i, \mathbf{y}) P_Q(q|\mathbf{y})^{m-1} [1 - P_Q(q|\mathbf{y})]^{\lambda-m} dq. \end{aligned} \quad (2.13)$$

Plugging Eq. (2.13) into (2.8) one obtains the progress rate

$$\begin{aligned} \varphi_i &= -\frac{1}{\mu} \sum_{m=1}^{\mu} \int_{-\infty}^{\infty} x_i p_x(x_i) \int_{q_l}^{q_u} p_Q(q|x_i, \mathbf{y}) \\ &\times \frac{\lambda!}{(m-1)!(\lambda-m)!} P_Q(q|\mathbf{y})^{m-1} [1 - P_Q(q|\mathbf{y})]^{\lambda-m} dq dx_i. \end{aligned} \quad (2.14)$$

Moving the sum into the innermost integration, another transformation can be applied using a well known relation between the sum over  $m$  and the regu-

larized incomplete beta function [5, p. 147]

$$\begin{aligned} & \sum_{m=1}^{\mu} \frac{P(q)^{m-1}[1-P(q)]^{\lambda-m}}{(m-1)!(\lambda-m)!} \\ &= \frac{1}{(\lambda-\mu-1)!(\mu-1)!} \int_0^{1-P(q)} t^{\lambda-\mu-1}(1-t)^{\mu-1} dt. \end{aligned} \quad (2.15)$$

Furthermore, we will rewrite the population dependent factor as follows

$$\begin{aligned} \frac{1}{\mu} \frac{\lambda!}{(\lambda-\mu-1)!(\mu-1)!} &= \frac{\lambda}{\mu} \frac{(\lambda-1)!}{(\lambda-\mu-1)!(\mu-1)!} \\ &= \frac{\lambda}{\mu} \frac{\Gamma(\lambda)}{\Gamma(\lambda-\mu) \Gamma(\mu)} \\ &= \frac{\lambda}{\mu} \frac{1}{B(\lambda-\mu, \mu)}, \end{aligned} \quad (2.16)$$

where we have used the property of the Gamma function  $\Gamma(n) = (n-1)!$  (for any integer  $n > 0$ ) and the known relation between Gamma and Beta functions  $\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = B(x, y)$ . These replacements will be useful later to derive approximations for large population sizes. After replacing the sum and refactoring we arrive at the following progress rate integral

$$\begin{aligned} \varphi_i &= -\frac{\lambda}{\mu B(\lambda-\mu, \mu)} \int_{x_i=-\infty}^{x_i=\infty} x_i p_x(x_i) \\ &\times \int_{q=q_l}^{q=q_u} p_Q(q|x_i, \mathbf{y}) \int_{t=0}^{t=1-P_Q(q)} t^{\lambda-\mu-1}(1-t)^{\mu-1} dt dq dx_i \end{aligned} \quad (2.17)$$

Now the integration order of  $t$  and  $q$  will be exchanged. This will enable an analytically closed form for the quality gain integration  $q$ . The current integral consists of following integration ranges

$$q_l \leq q \leq q_u, \text{ and } 0 \leq t \leq 1 - P_Q(q). \quad (2.18)$$

Defining the inverse transformation  $q = P_Q^{-1}(1-t)$  and integrating over  $t$  first, one obtains the new ranges as

$$0 \leq t \leq 1, \text{ and } q_l \leq q \leq P_Q^{-1}(1-t). \quad (2.19)$$

The integral changes to

$$\begin{aligned} \varphi_i &= -\frac{\lambda}{\mu B(\lambda-\mu, \mu)} \int_{x_i=-\infty}^{x_i=\infty} x_i p_x(x_i) \\ &\times \int_{t=0}^{t=1} t^{\lambda-\mu-1}(1-t)^{\mu-1} \int_{q=q_l}^{q=P_Q^{-1}(1-t)} p_Q(q|x_i, \mathbf{y}) dq dt dx_i. \end{aligned} \quad (2.20)$$

Now the innermost integral can be solved

$$\int_{q_l}^{P_Q^{-1}(1-t)} p_Q(q|x_i) dq = [P_Q(q|x_i)]_{q_l}^{P_Q^{-1}(1-t)} \quad (2.21)$$

$$= P_Q(P_Q^{-1}(1-t)|x_i) - P_Q(q_l|x_i) \quad (2.22)$$

$$= P_Q(P_Q^{-1}(1-t)|x_i) \quad (2.23)$$

$$=: f(t, x_i). \quad (2.24)$$

where the probability  $P_Q(q_l|x_i) = \Pr(Q \leq q_l|x_i) = 0$  for any lower bound value  $q_l$ . For better readability the function  $f(t, x_i)$  was introduced and  $\mathbf{y}$  was dropped within the conditional probability. Thus we arrive at the following progress rate integral

$$\begin{aligned} \varphi_i &= -\frac{\lambda}{\mu} \int_{x_i=-\infty}^{x_i=\infty} x_i p_x(x_i) \\ &\times \frac{1}{B(\lambda-\mu, \mu)} \int_{t=0}^{t=1} t^{\lambda-\mu-1} (1-t)^{\mu-1} f(t, x_i) dt dx_i. \end{aligned} \quad (2.25)$$

## 2.2 Large Population Approximation

Unfortunately a closed form solution of (2.25) is not possible due to the factor  $f(t, x_i) = P_Q(P_Q^{-1}(1-t)|x_i)$ .

But within the so-called large-population limit with  $(\mu, \lambda) \rightarrow \infty$  and constant truncation ratio  $\vartheta = \mu/\lambda$  a solution for the  $t$ -integration can be given using the results of Appendix B. Comparing (2.25) with identity (B.1) one can identify integral  $I_{\mu, \lambda}^a[f]$  with parameters  $a = 1$  and  $b = 0$  such that

$$\varphi_i = -\frac{\lambda}{\mu} \int_{x_i=-\infty}^{x_i=\infty} x_i p_x(x_i) I_{\mu, \lambda}^1[f] dx_i. \quad (2.26)$$

Evaluating function  $f(t, x_i)$  at  $t = 1 - \vartheta$  gives

$$f(t, x_i)|_{t=1-\vartheta} = P_Q(P_Q^{-1}(1-t)|x_i)|_{t=1-\vartheta} = P_Q(P_Q^{-1}(\vartheta)|x_i). \quad (2.27)$$

Therefore the progress rate integral in the asymptotic limit of infinitely large population sizes (constant  $\vartheta$ ) yields

Progress rate for large populations and generic  $P_Q(q)$

$$\varphi_i \simeq -\frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i p_x(x_i) P_Q(P_Q^{-1}(\vartheta)|x_i) dx_i, \quad (2.28)$$

which now consists only of a single integration over the  $i$ -th mutation component  $x_i$ . The main next task is to choose  $P_Q(q)$  and  $P_Q^{-1}(p)$  in such a way that the integral is analytically solvable.

### 2.3 Defining and Expanding the Distribution Function

An analytic approximation for  $\varphi_i$  can be derived starting from Eq. (2.28) as follows. For the first step expressions for  $P_Q(q|x_i)$  and  $q = P_Q^{-1}(\vartheta)$  are needed. The results from Eqs. (1.44) and (1.42) assuming a normally distributed quality gain can be used with

$$P_Q(q|x_i) = \Phi\left(\frac{q - E_{Q|x_i}}{D_i}\right) \quad (2.29)$$

$$q = P_Q^{-1}(\vartheta) = E_Q + D_Q \Phi^{-1}(\vartheta), \quad (2.30)$$

giving the result

$$P_Q(P_Q^{-1}(\vartheta)|x_i) = \Phi\left(\frac{E_Q + D_Q \Phi^{-1}(\vartheta) - E_{Q|x_i}}{D_i}\right). \quad (2.31)$$

Now the results can be collected. Using the large population approximation (2.28) with result (2.31) and inserting Eq. (1.46) for  $E_{Q|x_i}$  the progress rate yields

$$\begin{aligned} \varphi_i &= -\frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i p_x(x_i) \\ &\times \Phi\left(\frac{E_Q + D_Q \Phi^{-1}(\vartheta) - (x_i^2 + 2y_i x_i + c_i(1 - \cos(\alpha x_i)) + s_i \sin(\alpha x_i)) - E_i}{D_i}\right) dx_i. \end{aligned} \quad (2.32)$$

A closed form solution of the given integral cannot be obtained, since the argument of  $\Phi(\cdot)$  contains nonlinear terms in  $x_i$ . The subsequent derivations will tackle this problem. The idea will be to provide an approximate solution by splitting the argument of  $\Phi(\cdot)$  into a linear function  $g(x_i)$  and a nonlinear function  $h(x_i)$ . After splitting, a Taylor expansion of  $\Phi$  can be done for a small nonlinear perturbation  $h(x_i)$ . Keeping only the first two terms of the expansion will result in two analytically solvable integrals for the progress rate.

**Decomposing the quality gain** Considering an arbitrary position in search space  $\mathbf{y}$  and a comparably large mutation strength  $\sigma$ , the Rastrigin quality gain will be dominated by the global, i.e. spherical, structure. The fitness oscillations with strength  $A$  in Eq. (1.1) are superimposed on the spherical function. Therefore a resulting quality gain can be regarded as globally sphere-like with deviations due to nonlinear local perturbations, cf. Fig. 1.

The linearized quality gain component of the sphere function  $f_{i,\text{sph}}(y_i) = y_i^2$  is given by following expression, cf. Eq. (1.15),

$$\begin{aligned} Q_{i,\text{sph}} &= f_{i,\text{sph}}(y_i + x_i) - f_{i,\text{sph}}(y_i) \\ &= \frac{df_{i,\text{sph}}}{dy_i} x_i + O(x_i^2) \\ &\approx 2y_i x_i =: k_i x_i. \end{aligned} \quad (2.33)$$

The nonlinear Rastrigin quality gain from Eq. (1.9) is given by

$$Q_i = x_i^2 + 2y_i x_i + c_i(1 - \cos(\alpha x_i)) + s_i \sin(\alpha x_i), \quad (2.34)$$

from which we can identify the same linear term  $2y_i x_i = k_i x_i$ . Reordering the expression and defining the nonlinear terms of  $x_i$  as perturbation  $\delta(x_i)$  we get

$$\begin{aligned} Q_i &= 2y_i x_i + x_i^2 + c_i(1 - \cos(\alpha x_i)) + s_i \sin(\alpha x_i) \\ &= k_i x_i + \delta(x_i). \end{aligned} \quad (2.35)$$

with

$$\begin{aligned} k_i &= 2y_i \\ \delta(x_i) &= x_i^2 + c_i(1 - \cos(\alpha x_i)) + s_i \sin(\alpha x_i). \end{aligned} \quad (2.36)$$

Plugging the obtained relation back into the argument  $\Phi(\cdot)$  of Eq. (2.31), the addends can be rearranged in terms of a linear and nonlinear part

$$\begin{aligned} &\Phi\left(\frac{E_Q + D_Q \Phi^{-1}(\vartheta) - (x_i^2 + 2y_i x_i + c_i(1 - \cos(\alpha x_i)) + s_i \sin(\alpha x_i)) - E_i}{D_i}\right) \\ &= \Phi\left(\frac{-k_i x_i + E_Q - E_i + D_Q \Phi^{-1}(\vartheta)}{D_i} - \frac{\delta(x_i)}{D_i}\right) \\ &= \Phi(g(x_i) + h(x_i)), \end{aligned} \quad (2.37)$$

The linear function in  $x_i$  is obtained as

$$g(x_i) = -\frac{k_i}{D_i} x_i + \frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_i}, \quad (2.38)$$

with the following abbreviation from Eq. (1.21)

$$E_Q - E_i = E_{Q_i} = \sigma^2 + A \cos(\alpha y_i) \left(1 - e^{-\frac{(\alpha \sigma)^2}{2}}\right). \quad (2.39)$$

The nonlinear function in  $x_i$  yields

$$h(x_i) = -\frac{\delta(x_i)}{D_i} = -\frac{x_i^2 + c_i(1 - \cos(\alpha x_i)) + s_i \sin(\alpha x_i)}{D_i}. \quad (2.40)$$

**Expanding the CDF** At this point an expansion of the distribution can be done at  $g$  assuming small perturbations  $h$  (argument  $x_i$  is dropped for brevity), analogous to [5, p. 337, B.4],

$$\begin{aligned} \Phi(g + h) &= \sum_{k=0}^{\infty} \frac{1}{k!} \frac{d^k \Phi}{d g^k} h^k \\ &= \Phi(g) + \frac{d \Phi}{d g} h + \sum_{k=2}^{\infty} \frac{1}{k!} \frac{d^k \Phi}{d g^k} h^k \\ &= \Phi(g) + \frac{e^{-\frac{1}{2}g^2}}{\sqrt{2\pi}} h + O(h^2). \end{aligned} \quad (2.41)$$

Applying the decomposition (2.37) to Eq. (2.32) and considering only the first two terms of the Taylor expansion (2.41) we obtain a new progress rate approximation

$$\begin{aligned}\varphi_i &\approx -\frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i p_x(x_i) \Phi(g(x_i)) dx_i \\ &\quad - \frac{1}{\sqrt{2\pi\vartheta}} \int_{-\infty}^{\infty} x_i p_x(x_i) h(x_i) e^{-\frac{1}{2}g(x_i)^2} dx_i \\ &=: I_i^0 + I_i^1.\end{aligned}\tag{2.42}$$

The integrals  $I_i^0$  and  $I_i^1$  will be analytically solved in the next sections. The superscript denotes the 0th and 1st order term of the expansion, respectively.

### 2.3.1 Solving Integral $I_i^0$

Starting from the first line of Eq. (2.42) and using definition (2.38), the equation can be rewritten as

$$I_i^0 = -\frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i p_x(x_i) \Phi \left( -\frac{k_i}{D_i} x_i + \frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_i} \right) dx_i. \tag{2.43}$$

Inserting the mutation density  $p_x(x_i)$  from Eq. (2.9) and applying the substitution  $z = x_i/\sigma$  one gets

$$I_i^0 = -\frac{\sigma}{\sqrt{2\pi\vartheta}} \int_{-\infty}^{\infty} z e^{-\frac{1}{2}z^2} \Phi \left( -\frac{k_i \sigma}{D_i} z + \frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_i} \right) dz. \tag{2.44}$$

At this point the following integral identity [5, p. 330, A.12] can be applied

$$\int_{-\infty}^{\infty} t e^{-\frac{1}{2}t^2} \Phi(at + b) dt = \frac{a}{\sqrt{1+a^2}} \exp \left[ -\frac{1}{2} \frac{b^2}{1+a^2} \right]. \tag{2.45}$$

The corresponding coefficients can be identified as

$$a = -\frac{k_i \sigma}{D_i} \tag{2.46}$$

$$b = \frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_i}. \tag{2.47}$$

Evaluating the factor  $a/\sqrt{1+a^2}$  gives

$$\frac{a}{\sqrt{1+a^2}} = -\frac{\frac{k_i \sigma}{D_i}}{\sqrt{\frac{D_i^2}{D_i^2} + \frac{(k_i \sigma)^2}{D_i^2}}} = -\frac{k_i \sigma}{\sqrt{D_i^2 + (k_i \sigma)^2}} =: -\frac{k_i \sigma}{D_+}, \tag{2.48}$$

where following definition was introduced

$$D_+^2 := D_i^2 + (k_i \sigma)^2. \tag{2.49}$$

The factor  $\exp[-b^2/2(1+a^2)]$  yields

$$\begin{aligned}\exp \left[ -\frac{1}{2} \frac{b^2}{1+a^2} \right] &= \exp \left[ -\frac{1}{2} \left( \frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_i} \right)^2 \frac{1}{\frac{D_i^2}{D_i^2} + \frac{(k_i \sigma)^2}{D_i^2}} \right] \\ &= \exp \left[ -\frac{1}{2} \left( \frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_+} \right)^2 \right].\end{aligned}\tag{2.50}$$

Inserting results (2.48), (2.50) into identity relation (2.45), the integral (2.44) yields

$$I_i^0 = \frac{1}{\sqrt{2\pi}\vartheta} \frac{1}{\vartheta} \exp \left[ -\frac{1}{2} \left( \frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_+} \right)^2 \right] \frac{k_i \sigma^2}{D_+}. \quad (2.51)$$

### 2.3.2 Solving Integral $I_i^1$

The second summand of the progress rate from Eq. (2.42) is given by

$$I_i^1 = -\frac{1}{\sqrt{2\pi}\vartheta} \int_{-\infty}^{\infty} x_i h(x_i) p_x(x_i) e^{-\frac{1}{2}g(x_i)^2} dx_i \quad (2.52)$$

$$= -\frac{1}{\sqrt{2\pi}\vartheta} \int_{-\infty}^{\infty} x_i h(x_i) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i}{\sigma}\right)^2} e^{-\frac{1}{2}g(x_i)^2} dx_i. \quad (2.53)$$

The product of two Gaussian functions can be rewritten as a single Gaussian with a scaling factor  $C$  and resulting mean  $m$  and variance  $s^2$ , such that

$$I_i^1 = -\frac{C}{2\pi\vartheta\sigma} \int_{-\infty}^{\infty} x_i h(x_i) e^{-\frac{1}{2}\left(\frac{x_i-m}{s}\right)^2} dx_i. \quad (2.54)$$

Keeping this in mind and recalling definition (2.40) we have

$$h(x_i) = -\frac{\delta(x_i)}{D_i} = -\frac{x_i^2 + s_i \sin(\alpha x_i) + c_i(1 - \cos(\alpha x_i))}{D_i}. \quad (2.55)$$

Using this relation, the integral (2.53) will be reformulated later as an expected value of the function  $x_i h(x_i)$  over the normal density  $\mathcal{N}(m, s^2)$ .

We start with quadratic completion of the exponential functions and by using Eq. (2.38) with

$$g(x_i) := c_1 x_i + c_0, \text{ with} \quad (2.56)$$

$$c_1 = -\frac{k_i}{D_i} \quad (2.57)$$

$$c_0 = \frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_i}. \quad (2.58)$$

Quadratic completion shall give

$$e^{-\frac{1}{2}\frac{x_i^2}{\sigma^2}} e^{-\frac{1}{2}g(x_i)^2} = e^{-\frac{1}{2}\frac{x_i^2}{\sigma^2}} e^{-\frac{1}{2}(c_1 x_i + c_0)^2} \stackrel{!}{=} C e^{-\frac{1}{2}\frac{(x_i - m)^2}{s^2}}. \quad (2.59)$$

We have

$$\begin{aligned} \frac{x_i^2}{\sigma^2} + c_1^2 x_i^2 + 2c_1 x_i c_0 + c_0^2 &= \left( \frac{1}{\sigma^2} + c_1^2 \right) x_i^2 + 2c_0 c_1 x_i + c_0^2 \\ &= \beta \left( x_i^2 + \frac{2c_0 c_1}{\beta} x_i + \frac{c_0^2}{\beta} \right) \\ &= \beta \left( \left[ x_i + \frac{c_0 c_1}{\beta} \right]^2 - \left[ \frac{c_0 c_1}{\beta} \right]^2 + \frac{c_0^2}{\beta} \right), \end{aligned} \quad (2.60)$$

with a temporary variable

$$\beta = \frac{1}{\sigma^2} + c_1^2 = \frac{1 + (c_1\sigma)^2}{\sigma^2} = \frac{\frac{D_i^2}{D_i^2} + \left(\frac{k_i}{D_i}\sigma\right)^2}{\sigma^2} = \left(\frac{D_+}{D_i\sigma}\right)^2, \quad (2.61)$$

using again the definition (2.49) for  $D_+$ . Therefore the exponential terms read

$$e^{-\frac{1}{2}\frac{x_i^2}{\sigma^2}} e^{-\frac{1}{2}g(x_i)^2} = e^{\frac{\beta}{2}\left[\frac{c_0c_1}{\beta}\right]^2 - \frac{c_0^2}{2}} e^{-\frac{\beta}{2}\left[x_i + \frac{c_0c_1}{\beta}\right]^2} \stackrel{!}{=} C e^{-\frac{1}{2}\frac{(x_i - m)^2}{s^2}}. \quad (2.62)$$

The mean value  $m$  after quadratic completion and inserting the definitions for  $c_0$ ,  $c_1$  and  $\beta$  reads

$$\begin{aligned} m &= -\frac{c_0c_1}{\beta} = \frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_i} \frac{k_i}{D_i} \left(\frac{D_i\sigma}{D_+}\right)^2 \\ &= \frac{[E_{Q_i} + D_Q\Phi^{-1}(\vartheta)]k_i\sigma^2}{D_+^2}. \end{aligned} \quad (2.63)$$

The standard deviation  $s$  is given by

$$s = \frac{1}{\sqrt{\beta}} = \frac{D_i\sigma}{D_+}. \quad (2.64)$$

The factor  $C$  is evaluated by resolving variables  $\beta$ ,  $c_0$ ,  $c_1$  and using the relation  $D_+^2 = D_i^2 + (k_i\sigma)^2$ , which gives

$$\begin{aligned} C &= \exp\left[\frac{\beta}{2}\left(\frac{c_0c_1}{\beta}\right)^2 - \frac{c_0^2}{2}\right] = \exp\left[\frac{c_0^2}{2}\left(\frac{c_1^2}{\beta} - 1\right)\right] \\ &= \exp\left[\frac{1}{2}\left(\frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_i}\right)^2 \left(\left[\frac{k_i\sigma}{D_+}\right]^2 - 1\right)\right] \\ &= \exp\left[\frac{1}{2}\left(\frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_i}\right)^2 \left(\frac{(k_i\sigma)^2 + D_i^2 - D_+^2}{D_+^2} - 1\right)\right] \\ &= \exp\left[\frac{1}{2}\left(\frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_i}\right)^2 \left(\frac{D_+^2}{D_+^2} - \frac{D_i^2}{D_+^2} - 1\right)\right] \\ &= \exp\left[-\frac{1}{2}\left(\frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_+}\right)^2\right]. \end{aligned} \quad (2.65)$$

The result is the same exponential factor as in  $I_i^0$  from Eq. (2.51). Now all results for the quadratic completion can be applied to Eq. (2.54) and relation

(2.55) for  $h(x_i)$  is inserted. This yields

$$\begin{aligned}
I_i^1 &= -\frac{1}{2\pi\vartheta\sigma} \exp\left[-\frac{1}{2}\left(\frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_+}\right)^2\right] \int_{-\infty}^{\infty} x_i h(x_i) e^{-\frac{1}{2}\left(\frac{x_i-m}{s}\right)^2} dx_i \\
&= -\frac{s}{\sqrt{2\pi}\vartheta\sigma} \exp\left[-\frac{1}{2}\left(\frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_+}\right)^2\right] \frac{1}{\sqrt{2\pi}s} \int_{-\infty}^{\infty} x_i h(x_i) e^{-\frac{1}{2}\left(\frac{x_i-m}{s}\right)^2} dx_i \\
&= \frac{s}{\sqrt{2\pi}\vartheta\sigma D_i} \exp\left[-\frac{1}{2}\left(\frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_+}\right)^2\right] \\
&\quad \times \frac{1}{\sqrt{2\pi}s} \int_{-\infty}^{\infty} [x_i^3 + s_i x_i \sin(\alpha x_i) + c_i(x_i - x_i \cos(\alpha x_i))] e^{-\frac{1}{2}\left(\frac{x_i-m}{s}\right)^2} dx_i.
\end{aligned} \tag{2.66}$$

Having the newly obtained form, one can identify expected values of the terms in  $[\cdot]$  with respect to the density  $\frac{1}{\sqrt{2\pi}s} e^{-\frac{1}{2}\left(\frac{x_i-m}{s}\right)^2}$ . To avoid confusion with the actual random mutation variable  $x_i \sim \mathcal{N}(0, \sigma^2)$ , the integration variable is renamed  $t = x_i$  and  $t \sim \mathcal{N}(m, s^2)$ . Additionally replacing the prefactor  $s = \frac{D_i\sigma}{D_+}$ , the integral becomes

$$\begin{aligned}
I_i^1 &= \frac{1}{\sqrt{2\pi}\vartheta D_+} \exp\left[-\frac{1}{2}\left(\frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_+}\right)^2\right] \\
&\quad \times \{E[t^3] + s_i E[t \sin(\alpha t)] + c_i (E[t] - E[t \cos(\alpha t)])\}.
\end{aligned} \tag{2.67}$$

The occurring expected values are

$$\begin{aligned}
E[t] &= m \\
E[t^3] &= m^3 + 3ms^2 \\
E[t \sin(\alpha t)] &= e^{-\frac{1}{2}(\alpha s)^2} [m \sin(\alpha m) + \alpha s^2 \cos(\alpha m)] \\
E[t \cos(\alpha t)] &= e^{-\frac{1}{2}(\alpha s)^2} [m \cos(\alpha m) - \alpha s^2 \sin(\alpha m)],
\end{aligned} \tag{2.68}$$

with the trigonometric expectations given in Appendix (A.17) and (A.19). Inserting the expected values one gets

$$\begin{aligned}
I_i^1 &= \frac{1}{\sqrt{2\pi}\vartheta D_+} \exp\left[-\frac{1}{2}\left(\frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_+}\right)^2\right] \\
&\quad \times \{m^3 + 3ms^2 \\
&\quad + s_i e^{-\frac{1}{2}(\alpha s)^2} [m \sin(\alpha m) + \alpha s^2 \cos(\alpha m)] \\
&\quad + c_i (m - e^{-\frac{1}{2}(\alpha s)^2} [m \cos(\alpha m) - \alpha s^2 \sin(\alpha m)])\}.
\end{aligned} \tag{2.69}$$

Reinserting the definitions of  $m$  (2.63),  $s$  (2.64) gives

$$\begin{aligned}
I_i^1 &= \frac{1}{\sqrt{2\pi}} \frac{1}{\vartheta} \exp \left[ -\frac{1}{2} \left( \frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_+} \right)^2 \right] \frac{1}{D_+} \\
&\times \left\{ \left( \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \right)^3 + 3 \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \left( \frac{D_i \sigma}{D_+} \right)^2 \right. \\
&+ s_i e^{-\frac{1}{2} \left( \alpha \frac{D_i \sigma}{D_+} \right)^2} \left[ \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \sin \left( \alpha \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \right) \right. \\
&+ \alpha \left( \frac{D_i \sigma}{D_+} \right)^2 \cos \left( \alpha \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \right) \left. \right] \\
&+ c_i \left( \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \right. \\
&- e^{-\frac{1}{2} \left( \alpha \frac{D_i \sigma}{D_+} \right)^2} \left[ \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \cos \left( \alpha \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \right) \right. \\
&- \alpha \left( \frac{D_i \sigma}{D_+} \right)^2 \sin \left( \alpha \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \right) \left. \right] \left. \right\}. \tag{2.70}
\end{aligned}$$

### 2.3.3 Merging Results

At this point the results for  $I_i^0$  (2.51) and  $I_i^1$  (2.70) will be merged. Their sum is the exact solution of Eq. (2.42) incorporating the first two terms of the Taylor expansion.

The variables  $k_i, c_i, s_i, E_{Q_i}, D_Q, D_i$  and  $D_+$  will not be inserted into the expression due to very poor readability. Instead the variables are summarized below. The solution will be calculated and validated computationally in Sec. 2.5.

Progress rate for large  $\lambda$  and expanded CDF

$$\begin{aligned}
\varphi_i &= I_i^0 + I_i^1 \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\vartheta} \exp \left[ -\frac{1}{2} \left( \frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_+} \right)^2 \right] \\
&\quad \times \left\{ \frac{k_i \sigma^2}{D_+} \right. \\
&\quad + \frac{1}{D_+} \left[ \left( \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \right)^3 + 3 \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \left( \frac{D_i \sigma}{D_+} \right)^2 \right. \\
&\quad + s_i e^{-\frac{1}{2} (\alpha \frac{D_i \sigma}{D_+})^2} \left[ \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \sin \left( \alpha \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \right) \right. \\
&\quad \left. \left. + \alpha \left( \frac{D_i \sigma}{D_+} \right)^2 \cos \left( \alpha \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \right) \right] \right. \\
&\quad + c_i \left( \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \right. \\
&\quad \left. - e^{-\frac{1}{2} (\alpha \frac{D_i \sigma}{D_+})^2} \left[ \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \cos \left( \alpha \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \right) \right. \right. \\
&\quad \left. \left. - \alpha \left( \frac{D_i \sigma}{D_+} \right)^2 \sin \left( \alpha \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)] k_i \sigma^2}{D_+^2} \right) \right] \right) \right\} \\
&\tag{2.71}
\end{aligned}$$

with following variable definitions

$$\begin{aligned}
k_i &= 2y_i, \quad \text{Eq. (2.33)} \\
c_i &= A \cos(\alpha y_i) \text{ and } s_i = A \sin(\alpha y_i), \quad \text{Eq. (1.8)} \\
E_{Q_i} &= \sigma^2 + A \cos(\alpha y_i) \left( 1 - e^{-\frac{(\alpha \sigma)^2}{2}} \right), \quad \text{Eq. (1.21)} \\
D_Q^2 &= \sum_{i=1}^N \text{Var}[Q_i] = \sum_{i=1}^N 2\sigma^4 + 4y_i^2 \sigma^2 + \dots \\
&\quad + \frac{A^2}{2} \left[ 1 - e^{-(\alpha \sigma)^2} \right] \left[ 1 - \cos(2\alpha y_i) e^{-(\alpha \sigma)^2} \right] \\
&\quad + 2A\alpha\sigma^2 e^{-\frac{1}{2}(\alpha \sigma)^2} \left[ \alpha\sigma^2 \cos(\alpha y_i) + 2y_i \sin(\alpha y_i) \right], \quad \text{Eq. (1.29)} \\
D_i^2 &= \sum_{j \neq i} \text{Var}[Q_j], \quad \text{Eq. (1.49)} \\
D_+^2 &= D_i^2 + (k_i \sigma)^2 = \sum_{j \neq i} \text{Var}[Q_j] + (k_i \sigma)^2, \quad \text{Eq. (2.49)} \\
&\tag{2.72}
\end{aligned}$$

## 2.4 Large Dimensionality Approximation

The results of Eqs. (2.71) and (2.72) are the exact solution of Eq. (2.42) and can be used if highest precision is needed. At this point a result which is shorter and easier to grasp is desired, since it will be used for upcoming studies. A significantly simpler result can be obtained by analyzing  $m$  from Eq. (2.63) and  $s$  from Eq. (2.64) in terms of their  $N$  scaling behavior.

Assuming large dimensionality  $N$ , the variance quantities are approximately equal, since they only differ by a single component, the  $i$ -th component. Neglecting the contribution of one component given large  $N$  we have

$$\begin{aligned} D_Q^2 &= \sum_{i=1}^N \text{Var}[Q_i] \approx \sum_{j \neq i} \text{Var}[Q_j] \approx \sum_{j \neq i} \text{Var}[Q_j] + (k_i \sigma)^2 \\ &\approx D_i^2 \approx D_+^2. \end{aligned} \quad (2.73)$$

Within the approximation we define the variances to be equal to  $D_Q^2$ , namely

$$D_i^2 = D_Q^2, \text{ and } D_+^2 = D_Q^2. \quad (2.74)$$

This approximation changes  $m$  from Eq. (2.63) and  $s$  from Eq. (2.64) as follows

$$m \approx \frac{(E_{Q_i} + D_Q \Phi^{-1}(\vartheta)) k_i \sigma^2}{D_Q^2} = \left( \frac{E_{Q_i}}{D_Q^2} + \frac{\Phi^{-1}(\vartheta)}{D_Q} \right) k_i \sigma^2 \quad (2.75)$$

$$s \approx \frac{D_Q}{D_Q} \sigma = \sigma. \quad (2.76)$$

Since Eq. (2.75) contains  $D_Q$  and  $D_Q^2$  in its denominator, it can further be simplified for large  $N$ . From Eq. (2.73) we observe that  $D_Q^2$  scales with  $N$  and  $D_Q$  with  $\sqrt{N}$ .

The first term  $E_{Q_i}$  is just the quality gain expectation of a single component. The second term  $\Phi^{-1}(\vartheta)$  diverges only for  $\vartheta = 0$  and  $\vartheta = 1$ , which are not useful truncation values. Both terms are suppressed by  $N$  and  $\sqrt{N}$ , respectively, and the infinite dimension limit can be evaluated as

$$\lim_{N \rightarrow \infty} m(N) = \lim_{N \rightarrow \infty} \left( \frac{E_{Q_i}}{D_Q^2} + \frac{\Phi^{-1}(\vartheta)}{D_Q} \right) k_i \sigma^2 = 0, \quad (2.77)$$

which is valid for any finite  $\sigma$ . The obtained results  $s \approx \sigma$  and  $m \approx 0$  change the density of  $t \sim \mathcal{N}(m, s^2)$  to the mutation density  $t \sim \mathcal{N}(0, \sigma^2)$  in Eq. (2.67). Reevaluating the expected values of Eq. (2.68) only a single term remains, such that

$$\mathbb{E}[t \sin(\alpha t)] \simeq \alpha \sigma^2 \exp\left[-\frac{1}{2}(\alpha \sigma)^2\right]. \quad (2.78)$$

The approximation (2.74) also changes the common exponential factor of  $I_i^0$

and  $I_i^0$  as follows

$$\begin{aligned} & \frac{1}{\sqrt{2\pi}} \frac{1}{\vartheta} \exp \left[ -\frac{1}{2} \left( \frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_+} \right)^2 \right] \\ & \approx \frac{1}{\sqrt{2\pi}} \frac{1}{\vartheta} \exp \left[ -\frac{1}{2} \left( \frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_Q} \right)^2 \right] \\ & \simeq \frac{1}{\sqrt{2\pi}} \frac{1}{\vartheta} \exp \left[ -\frac{1}{2} (\Phi^{-1}(\vartheta))^2 \right] =: c_\vartheta, \end{aligned} \quad (2.79)$$

where the ratio  $E_{Q_i}/D_Q$  vanishes with  $O(1/\sqrt{N})$  giving the last line.

The newly defined coefficient  $c_\vartheta$  can be identified as  $e^{1,0}$  from Eq. (B.30) yielding the asymptotic generalized progress coefficients. The expression for  $c_\vartheta$  is also in accordance with the asymptotic expansion of the progress coefficient  $c_{\mu/\mu,\lambda}$  by Beyer [5, p. 249]. The difference in the  $\Phi^{-1}$ -argument ( $\vartheta$  vs.  $1-\vartheta$ ) is due to a differently chosen substitution that exchanges the integrand powers of Eq. (B.2) and its maximizer location. Since  $\Phi^{-1}(\vartheta) = -\Phi^{-1}(1-\vartheta)$  the two results are equal after squaring  $\Phi^{-1}$ .

As  $c_\vartheta$  is obtained for infinitely large populations, it poses an upper bound of  $c_{\mu/\mu,\lambda}$  and overestimates the progress, see also Fig. 14.

The approximation quality of the subsequent result is expected to be better for  $\vartheta$  not close to 0 or 1, since the Taylor expansion point in Eq. (2.38) depends on  $\Phi^{-1}(\vartheta)$ , and  $\Phi^{-1}(0) = -\infty$  and  $\Phi^{-1}(1) = \infty$ .

#### 2.4.1 Applying the Approximations

Inserting the result of Eq. (2.79) into Eq. (2.51) and setting  $D_+ = D_Q$  yields

$$I_i^0 \approx c_\vartheta \frac{\sigma^2}{D_Q} k_i. \quad (2.80)$$

Analogously the results of Eqs. (2.78) and (2.79) are applied to Eq. (2.69). Only a single term is left within the brackets  $\{\cdot\}$  giving

$$\begin{aligned} I_i^1 & \approx c_\vartheta \frac{1}{D_Q} s_i e^{-\frac{1}{2}(\alpha\sigma)^2} \alpha\sigma^2 = c_\vartheta \frac{1}{D_Q} A \sin(\alpha y_i) e^{-\frac{1}{2}(\alpha\sigma)^2} \alpha\sigma^2 \\ & = c_\vartheta \frac{\sigma^2}{D_Q} e^{-\frac{1}{2}(\alpha\sigma)^2} [\alpha A \sin(\alpha y_i)] \\ & = c_\vartheta \frac{\sigma^2}{D_Q} e^{-\frac{1}{2}(\alpha\sigma)^2} d_i, \end{aligned} \quad (2.81)$$

where the definition  $s_i = A \sin(\alpha y_i)$  was used. Additionally, the derivative of the Rastrigin oscillation term  $d_i = \alpha A \sin(\alpha y_i)$  was identified, as introduced in Eq. (1.11). Finally the progress rate is simplified and yields

Progress rate for large  $\lambda$  and  $N$

$$\begin{aligned}
\varphi_i &= I_i^0 + I_i^1 \\
&= c_\vartheta \frac{\sigma^2}{D_Q} \left( k_i + e^{-\frac{1}{2}(\alpha\sigma)^2} d_i \right) \\
&= c_\vartheta \frac{\sigma^2}{D_Q} \left( 2y_i + e^{-\frac{1}{2}(\alpha\sigma)^2} \alpha A \sin(\alpha y_i) \right).
\end{aligned} \tag{2.82}$$

with the variance  $D_Q^2$  given in Eq. (1.29).

#### 2.4.2 Discussion

The obtained result of Eq. (2.82) is very interesting, especially compared to the alternative progress rate (2.109) via  $c_{\mu/\mu,\lambda}$ . A numeric comparison is given in Fig. 14. Displaying both equations gives

$$\varphi_{i,c_\vartheta} = c_\vartheta \frac{\sigma^2}{D_Q} \left( 2y_i + e^{-\frac{1}{2}(\alpha\sigma)^2} \alpha A \sin(\alpha y_i) \right) \tag{2.83}$$

$$\varphi_{i,c_{\mu/\mu,\lambda}} = c_{\mu/\mu,\lambda} \frac{\sigma^2}{D_Q} \left( 2y_i + \alpha A \sin(\alpha y_i) \right). \tag{2.84}$$

The progress coefficients are asymptotically equal  $c_\vartheta \simeq c_{\mu/\mu,\lambda}$  within the infinite population limit.

The first term  $k_i = 2y_i$  is the same for both and determines the progress on the spherical function. The second term shows a remarkable difference.

In Eq. (2.83) the oscillation derivative  $d_i = \alpha A \sin(\alpha y_i)$  is suppressed exponentially for large mutations  $\sigma$  (or high oscillation frequency  $\alpha$ ). It correctly accounts for the fact that local perturbations of the spherical fitness have only a local effect on the expected progress, i.e. they are only relevant for small  $\sigma$ .

This difference shows also the shortcomings of Eq. (2.84). The perturbation derivative  $d_i$  is not suppressed and has a global effect on the progress independent of  $\sigma$ . Therefore Eq. (2.84) may show very large deviations from experimental values for larger  $\sigma$ , especially if the oscillation is prominent with large  $A$  or  $\alpha$ .

Setting  $A = 0$  or  $\alpha = 0$  the Rastrigin function degenerates to the sphere and both expressions yield asymptotically the same result.

## 2.5 Experiments and Numerical Solutions

### 2.5.1 Numerical Progress Rate Solution

During the previous derivation of the progress rate multiple integrals were obtained analytically as intermediate steps. In this section exact integral expressions of the progress rate are solved numerically. The only approximation is the choice of a CDF, see below. The results are compared to experiments. We have the integrals

- I1: Eq. (2.14)
- I2: Eq. (2.17), triple integral and numerically most involving
- I3: Eq. (2.25)

To solve the equations numerically, a cumulative distribution function has to be chosen. The conditioned distribution function ( $x_i = \text{const.}$ ) is chosen according to Eq. (1.50), non-linear version, and yields

$$P_Q(q|x_i) = \Phi\left(\frac{q - E_{Q|x_i}}{D_i}\right). \quad (2.85)$$

The distribution function due to  $N$  mutated components is given by (1.43) and inverted according to (1.42) giving

$$P_Q^{-1}(p) = q = E_Q + D_Q \Phi^{-1}(p). \quad (2.86)$$

Combining the equations and evaluating at  $p = \vartheta$  we obtain

$$P_Q(P_Q^{-1}(\vartheta)|x_i) = \Phi\left(\frac{E_Q + D_Q \Phi^{-1}(\vartheta) - E_{Q|x_i}}{D_i}\right). \quad (2.87)$$

Given the distribution functions (2.85) and (2.87), integrals I1, I2 and I3 are all equivalent and should yield the same results, see Fig. 6.

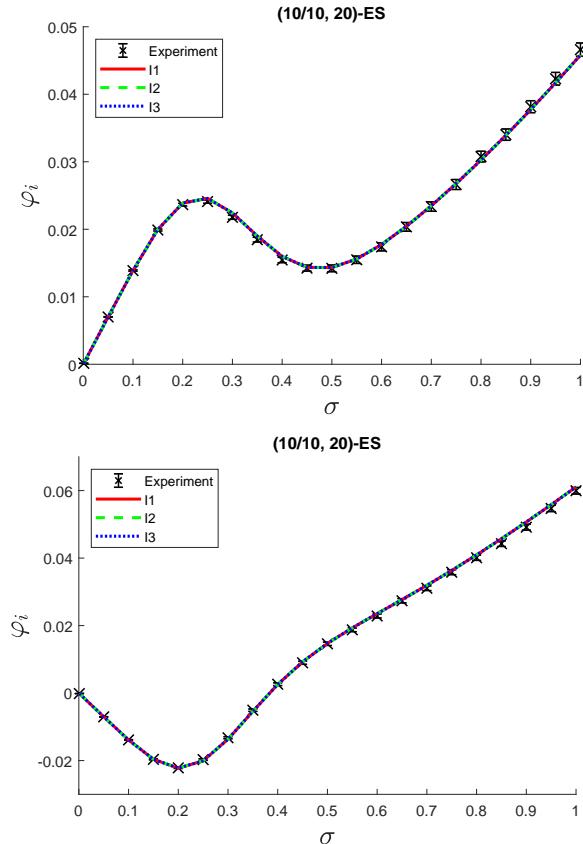


Figure 6: For the experimental results, one-generation experiments were done and the progress was obtained according to Eq. (2.1) by averaging over 100000 runs with dimensionality  $N = 30$  and frequency  $\alpha = 2\pi$ . The starting position was chosen to be  $\mathbf{y} = 1.25$  (top) and  $\mathbf{y} = 1.75$  (bottom). The three integrals yield the same result, as expected. Very good agreement between experiment and numerical solution is observed. The only approximation is the assumption of a normally distributed quality gain.

### 2.5.2 Overview of Approximations

The integrals from Fig. 6 give the most precise, i.e. least approximated, numeric result of the progress rate and will be referred to as only “NUM” for the subsequent plots. Approximations are denoted by “A” and a number. They are compared to the experimental and numeric results. Following data will be shown

- Experiment: average over 100000 trials
- NUM: Eq. (2.25); normally distributed quality gain; only numerically solvable
- A1-NUM: Eq. (2.28); additionally large population  $\lambda$ ; only numerically solvable
- A2: additionally Taylor-expanding CDF and keeping 0th and 1st order
  - A2-NUM: numeric solution of Eq. (2.42)
  - A2: analytic solution according to Eqs. (2.71) and (2.72)
- A3: Eq. (2.82); additionally large dimensionality  $N$ 
  - A3: spherical progress term  $k_i$  with  $d_i = 0$  in Eq. (2.82)
  - A3: perturbation progress term  $d_i \exp[-...]$  with  $k_i = 0$  in Eq. (2.82)
- C: Eq. (2.109); alternative progress rate using  $c_{\mu/\mu,\lambda}$ ; limited applicability for Rastrigin fitness
  - C: spherical progress using only  $k_i$
  - C: progress using full derivative  $f'_i = k_i + d_i$

**Frequency Value** For all following experiments the oscillation frequency was set to  $\alpha = 2\pi$  and remains as such, if not explicitly stated otherwise.

### 2.5.3 Approximations A1 and A2

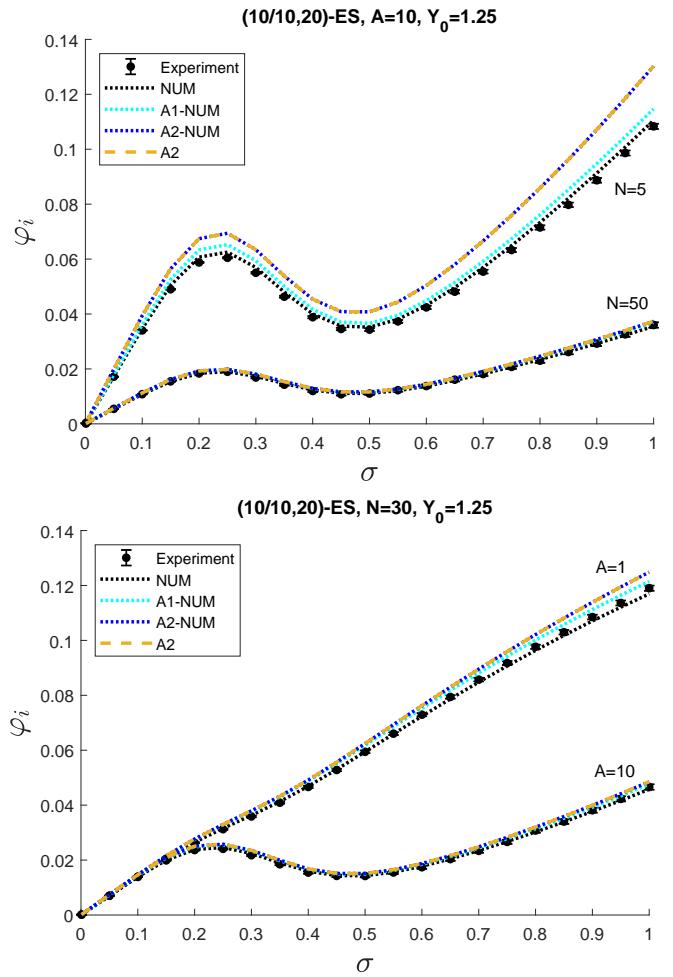


Figure 7: Variation of  $N$  (top) and  $A$  (bottom). We have better agreement for larger dimensionality  $N$ , which can be often observed due to the CLT. For  $N = 5$ , the Taylor expansion of  $A2$  introduces a moderately larger deviation. Very good agreement is observed for  $A = 1$  and  $A = 10$ , i.e. for a broader range of oscillation strengths. Numeric and analytic calculation of  $A2$  match perfectly.

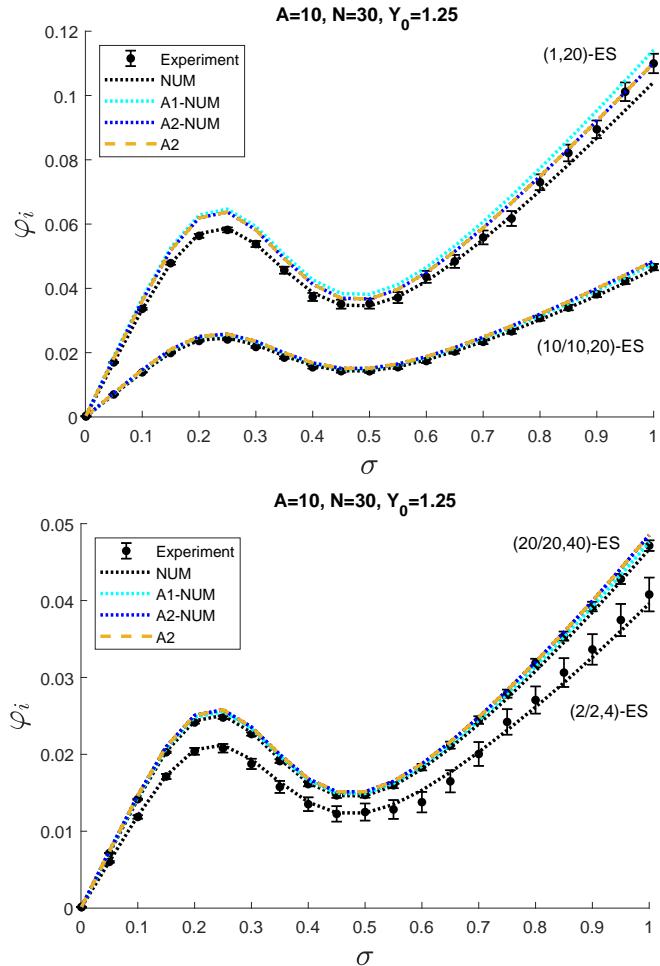


Figure 8: Variation of the truncation ratio  $\vartheta = \{0.05, 0.5\}$  (top) and  $\lambda$  (bottom,  $\vartheta=\text{const.}$ ). Decreasing the ratio increases the selection pressure and thus the progress. The approximation quality tends to decrease for more extreme truncation values, here for  $\vartheta = 0.05$ . For smaller populations (constant  $\vartheta$ ) the progress decreases and deviation increases, as expected. The approximations depend only on the truncation ratio  $\vartheta$ , assuming a large  $\lambda$ . Numeric and analytic calculation of A2 match perfectly.

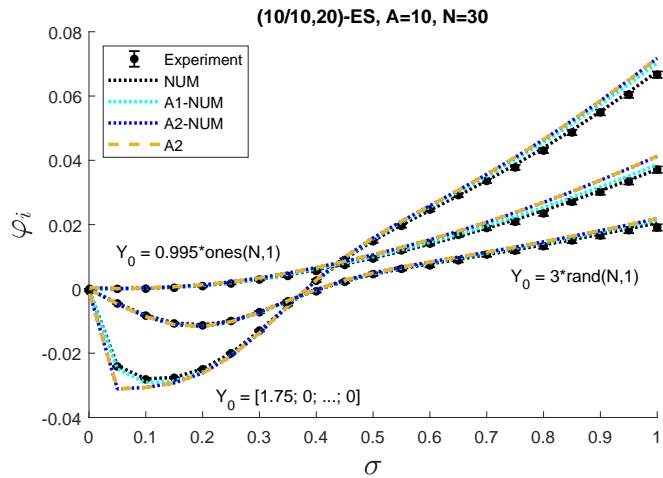


Figure 9: Variation of position  $\mathbf{y}$ . The vector  $\mathbf{y} = 0.995\text{ones}(N,1)$  corresponds to all components at the same local minimum. The vector  $\mathbf{y} = 3\text{rand}(N,1)$  has uniformly distributed random values  $\mathbf{y} \sim \mathcal{U}([0, 3])$  for all components. The location  $\mathbf{y} = [1.75, 0, \dots, 0]$  has only one non-zero component. One can see moderately higher deviations for small mutations. This is investigated later within approximation A3.

#### 2.5.4 Approximation A3

From last section the parameters  $N$  and  $\vartheta$  were identified to have a larger influence. They are varied in Figs. 10 and 11 over a larger  $\sigma$ -range. Figure 12 shows the progress for large normalized  $\sigma^*$  values, motivated by dynamic experiments. Figures 13 and 14 show a more detailed analysis for a specific position.

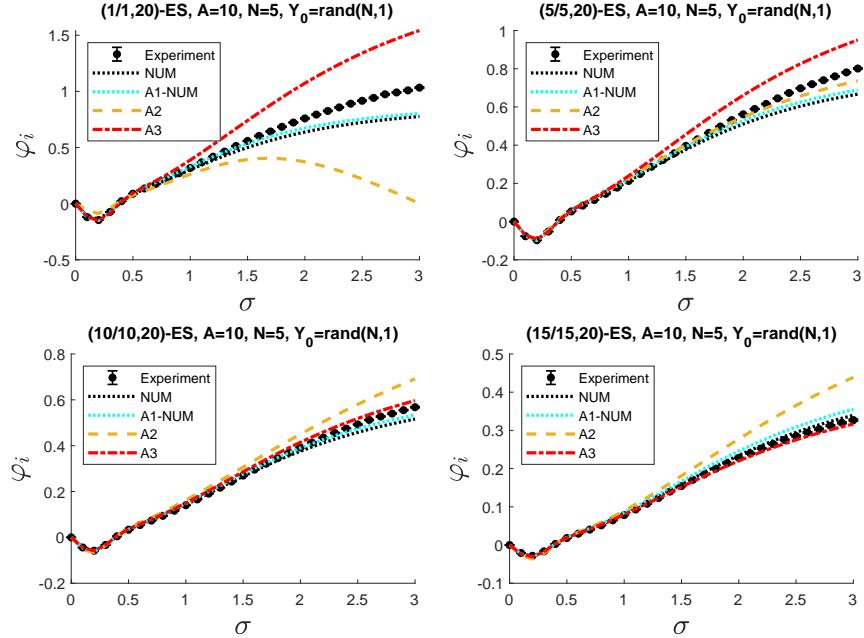


Figure 10: Variation of  $\vartheta = \{0.05, 0.25, 0.5, 0.75\}$  with  $N = 5$  and  $y \sim \mathcal{U}([0, 3])$  with constant seed and  $R_0 \approx 4.2$ . For small  $N$  the applicability of the CLT for the CDF is very limited, such that even NUM and A1-NUM may show larger deviations. Approximation A3 yields relatively good results even for small  $N$ , if  $\vartheta$  is not close to 0.

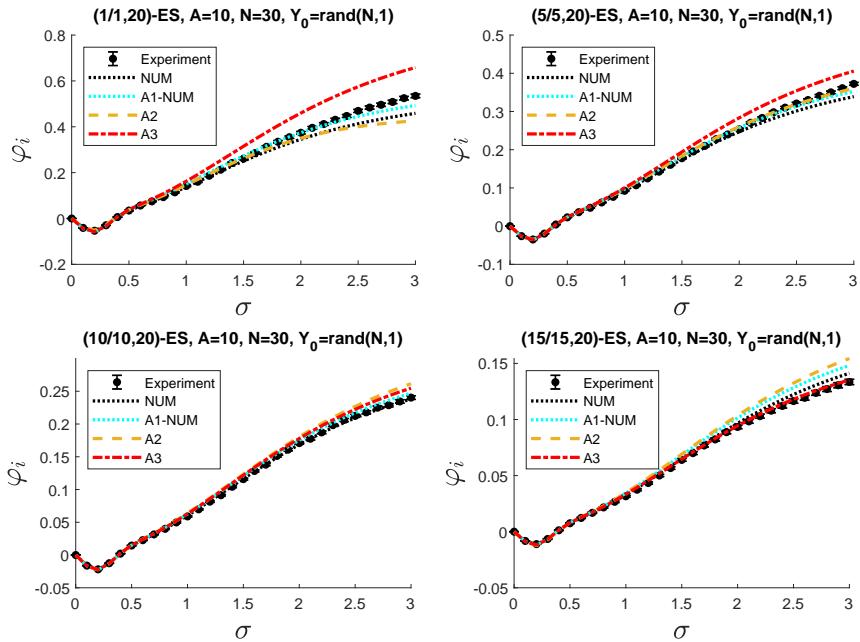


Figure 11: Variation of  $\vartheta = \{0.05, 0.25, 0.5, 0.75\}$  with  $N = 30$  and  $\mathbf{y} \sim \mathcal{U}([0, 3])$  with constant seed and  $R_0 \approx 9$ . As expected the approximation quality increases for larger  $N$  and  $\vartheta$  not close to 0. Within the range  $0.25 \lesssim \vartheta \lesssim 0.75$  approximations A2 and A3 yield similarly good results.

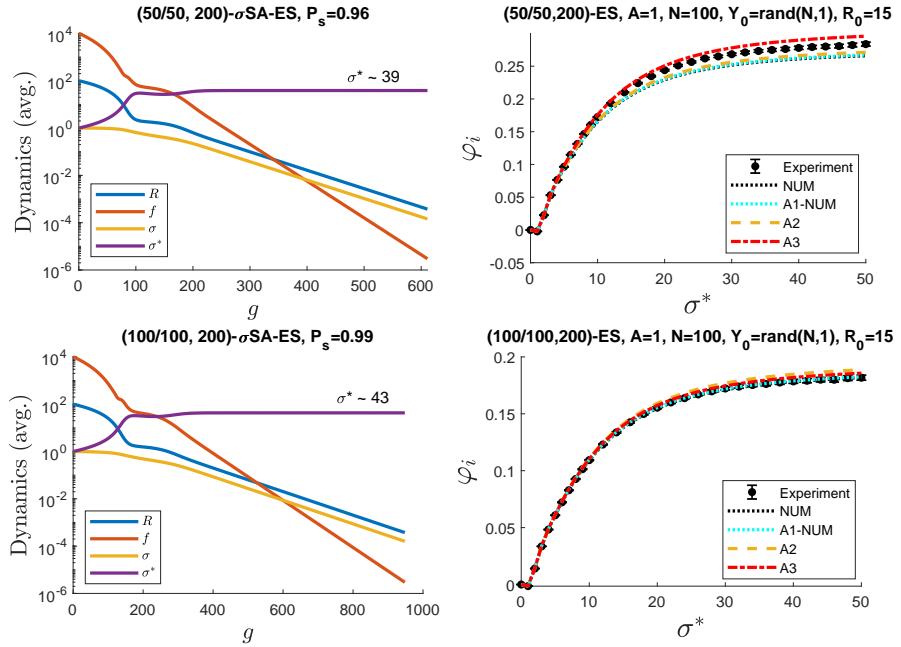


Figure 12: Approximation quality for large  $\sigma^* = \sigma N/R$  and  $\vartheta = \{0.25, 0.5\}$ . Plots on the left-hand side show the averaged dynamics for Rastrigin problem  $N = 100, A = 1, \mathbf{y} = 10$  and two truncation ratios. The average dynamics yield large levels of  $\sigma^* \sim 40$  which are relatively constant after the initial phase has passed. Therefore the right plots show the progress rate for  $0 \leq \sigma^* \leq 50$ . The approximation quality of A2 and A3 is good even for large  $\sigma^*$ . The dynamic experiments were repeated  $10^4$  times using  $\tau = 1/\sqrt{8N}$  and the success probability is denoted by  $P_s$ . The initial position for the one-generation experiments was again chosen as  $\mathbf{y} \sim \mathcal{U}([0, 3])$ . Important to note is that the dynamic  $\sigma^*$  changes if  $\mu$  or  $\vartheta$  are modified. Larger populations also enable larger mutation strengths.

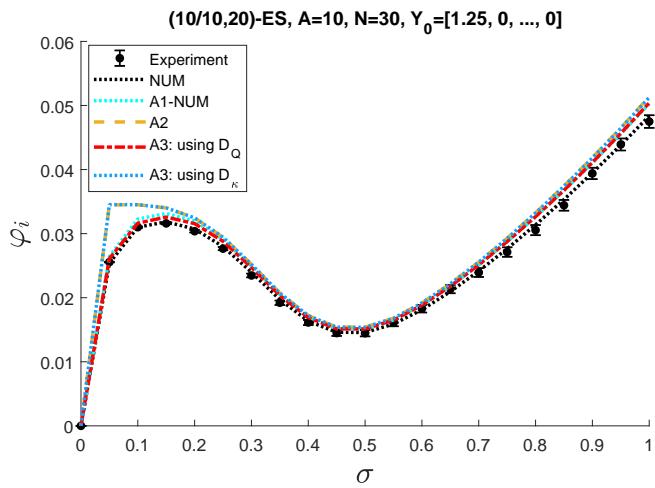


Figure 13: Approximation A3 from Eq. (2.82) with  $D_Q$  compared to A3 using  $D_\kappa$ , since for large  $N$  we should have  $D_Q \approx D_\kappa$ . One can see that A3 with  $D_\kappa$  to some degree reproduces the characteristics of A2. This indicates that in the case of a single dominating component, e.g.  $\mathbf{y} = [1.25, 0, \dots, 0]$ , the standard deviation differences become more pronounced. The term  $D_Q$  contains precise information on the  $i$ -th component variation, whereas  $D_\kappa$  approximates it and  $D_i$  neglects it, see also Eq. (2.72). Still, the primary source of error is the Taylor expansion of Eq. (2.41) applied in A2.

### 2.5.5 Approximation A3 compared to C

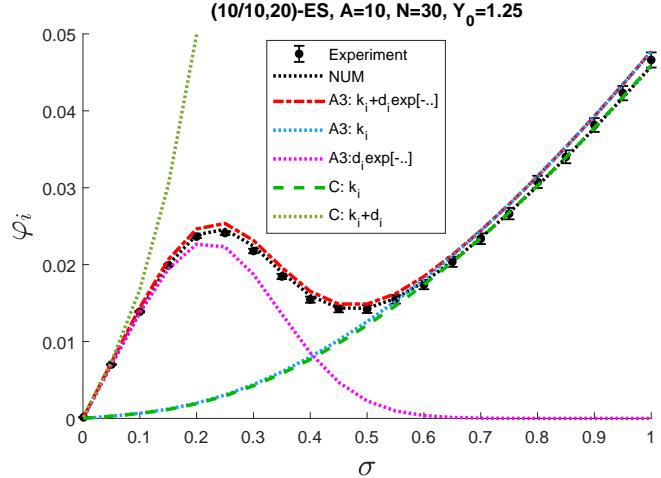


Figure 14: Approximation A3 decomposed into its two progress terms and compared to the progress rate using  $c_{\mu/\mu,\lambda}$  (denoted by C). The term  $k_i = 2y_i$  yields the spherical progress for both A3 and C for large  $\sigma$ , where C is closer to the black numeric solution since no large populations are assumed. Thus, A3 using  $c_\theta$  overestimates the true progress. For A3 the term  $d_i = \alpha A \sin(\alpha y_i)$  is only relevant for small  $\sigma$  and gets exponentially suppressed. The superposition of both terms (red line) yields very good results. The progress C with  $k_i + d_i$  shows very large deviations for larger  $\sigma$  since the derivative  $d_i$  is globally relevant, i.e. not suppressed.

## 2.6 Alternative Progress Rate via Progress Coefficient

Starting from Eq. (2.14) an alternative progress rate can be derived using a so-called progress coefficient  $c_{\mu/\mu,\lambda}$ , already established in [5, p. 216]. However, in contrast to the progress rate in Eq. (2.42), the derivation is limited to using only a linearized quality gain which in turn leads to sub-optimal results for the Rastrigin fitness.

As opposed to Sec. 2.1 and 2.2, the mutation integral over  $x_i$  is solved first and Eq. (2.14) will be restructured by exchanging  $x$  and  $q$ . We have

$$\begin{aligned}\varphi_i &= -\frac{1}{\mu} \sum_{m=1}^{\mu} E[x_{m;\lambda}] \\ \varphi_i &= -\frac{1}{\mu} \sum_{m=1}^{\mu} \int_{q_l}^{q_u} \left[ \int_{-\infty}^{\infty} x_i p_x(x_i) p_Q(q|x_i, \mathbf{y}) dx_i \right] \\ &\quad \times \frac{\lambda!}{(m-1)!(\lambda-m)!} P_Q(q|\mathbf{y})^{m-1} [1 - P_Q(q|\mathbf{y})]^{\lambda-m} dq \\ &= -\frac{1}{\mu} \sum_{m=1}^{\mu} \int_{q_l}^{q_u} I_{x_i}(q) \frac{\lambda!}{(m-1)!(\lambda-m)!} P_Q(q|\mathbf{y})^{m-1} [1 - P_Q(q|\mathbf{y})]^{\lambda-m} dq\end{aligned}\tag{2.88}$$

with the mutation integral  $I_{x_i}(q)$  that will be solved approximately. Assuming that the conditional quality gain density  $p_Q(q|x_i, \mathbf{y})$  is normally distributed with *linearized* mutation component  $x_i$ , see also Eqs. (1.48) and (1.49), we get

$$\begin{aligned}p_Q(q|x_i, \mathbf{y}) &= \frac{1}{\sqrt{2\pi}D_i} \exp \left[ -\frac{1}{2} \left( \frac{q - E_{Q|x_i}}{D_i} \right)^2 \right] \\ &= \frac{1}{\sqrt{2\pi}D_i} \exp \left[ -\frac{1}{2} \left( \frac{q - f'_i x_i - E_i}{D_i} \right)^2 \right].\end{aligned}\tag{2.89}$$

Inserting the two densities into the mutation integral we have

$$I_{x_i}(q) = \int_{-\infty}^{\infty} x_i \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x_i}{\sigma})^2} \frac{1}{\sqrt{2\pi}D_i} e^{-\frac{1}{2}\left(\frac{q-f'_i x_i - E_i}{D_i}\right)^2} dx_i.\tag{2.90}$$

Substituting  $t = x_i/\sigma$  and reordering terms gives

$$I_{x_i}(q) = \frac{\sigma}{\sqrt{2\pi}D_i} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{1}{2}t^2} e^{-\frac{1}{2}\left(\frac{q-f'_i \sigma t - E_i}{D_i}\right)^2} dt.\tag{2.91}$$

Now following identity can be applied [5, p. 330, A.8] to Eq. (2.91)

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{1}{2}t^2} e^{-\frac{1}{2}(at+b)^2} dt = \frac{-ab}{(1+a^2)^{3/2}} \exp \left[ -\frac{1}{2} \frac{b^2}{1+a^2} \right].\tag{2.92}$$

We can identify the coefficients  $a$  and  $b$  as follows

$$\begin{aligned}a &= -\frac{f'_i \sigma}{D_i} \\ b &= \frac{q - E_i}{D_i}.\end{aligned}\tag{2.93}$$

The first factor of the identity is calculated as

$$\begin{aligned} \frac{-ab}{(1+a^2)^{3/2}} &= \frac{f'_i\sigma(q-E_i)}{D_i^2} \frac{1}{(1+(f'_i\sigma/D_i)^2)^{3/2}} \\ &= \frac{f'_i\sigma(q-E_i)D_i}{(D_i^2+(f'_i\sigma)^2)^{3/2}} \\ &= \frac{f'_i\sigma D_i}{D_f^2} \frac{q-E_i}{D_f}, \end{aligned} \quad (2.94)$$

with the newly defined quantity.

$$D_f^2 := D_i^2 + (f'_i\sigma)^2. \quad (2.95)$$

The second factor of yields

$$\begin{aligned} \exp\left[-\frac{1}{2}\frac{b^2}{1+a^2}\right] &= \exp\left[-\frac{1}{2}\frac{(q-E_i)^2}{D_i^2}\frac{1}{1+(f'_i\sigma/D_i)^2}\right] \\ &= \exp\left[-\frac{1}{2}\frac{(q-E_i)^2}{D_i^2+(f'_i\sigma)^2}\right] = \exp\left[-\frac{1}{2}\frac{(q-E_i)^2}{D_f^2}\right]. \end{aligned} \quad (2.96)$$

Using the results above we get for integral (2.91)

$$\begin{aligned} I_{x_i}(q) &= \frac{\sigma}{\sqrt{2\pi}D_i} \frac{f'_i\sigma D_i}{D_f^2} \left(\frac{q-E_i}{D_f}\right) \exp\left[-\frac{1}{2}\left(\frac{q-E_i}{D_f}\right)^2\right] \\ &= \frac{1}{\sqrt{2\pi}} \frac{f'_i\sigma^2}{D_f^2} \left(\frac{q-E_i}{D_f}\right) \exp\left[-\frac{1}{2}\left(\frac{q-E_i}{D_f}\right)^2\right]. \end{aligned} \quad (2.97)$$

Inserting the result for  $I_{x_i}(q)$  into progress rate integral (2.88) and assuming a normal cumulative distribution function  $P_Q(q) = \Phi\left(\frac{q-E_Q}{D_Q}\right)$ , see Eq. (1.43), we obtain the following

$$\begin{aligned} \varphi_i &= -\frac{\lambda}{\mu} \frac{1}{\sqrt{2\pi}} \frac{f'_i\sigma^2}{D_f^2} \int_{-f(\mathbf{y})}^{\infty} \left(\frac{q-E_i}{D_f}\right) e^{-\frac{1}{2}\left(\frac{q-E_i}{D_f}\right)^2} \\ &\quad \times \sum_{m=1}^{\mu} \frac{(\lambda-1)!}{(m-1)!(\lambda-m)!} \left[\Phi\left(\frac{q-E_Q}{D_Q}\right)\right]^{m-1} \left[1-\Phi\left(\frac{q-E_Q}{D_Q}\right)\right]^{\lambda-m} dq. \end{aligned} \quad (2.98)$$

The integration range of  $q$  was set to  $[-f(\mathbf{y}), \infty)$ , corresponding to the attainable fitness changes  $Q_{\mathbf{y}}(\mathbf{x})$  of Rastrigin's function.

The obtained integral will only be solvable in terms of a progress coefficient  $c_{\mu/\mu,\lambda}$ , if an approximation for large dimensionality  $N \rightarrow \infty$  is applied to the arguments  $\frac{q-E_i}{D_f}$  and  $\frac{q-E_Q}{D_Q}$ , namely that

$$\begin{aligned} E_i &= \sum_{j \neq i} \text{E}[Q_j] \approx \sum_{j=1}^N \text{E}[Q_j] = E_Q, \\ D_f^2 &= \sum_{j \neq i} \text{Var}[Q_j] + (f'_i\sigma)^2 \approx \sum_{j=1}^N \text{Var}[Q_j] = D_Q^2. \end{aligned} \quad (2.99)$$

The assumption is that expectation value and variance are dominated by  $N - 1$  terms, such that the contribution of a single  $i$ -th component is negligible.

Setting  $E_i = E_Q$ ,  $D_f = D_Q$  in Eq. (2.98) and defining the substitution  $z = \frac{q-E_Q}{D_Q}$  the integral becomes

$$\begin{aligned} \varphi_i &= -\frac{\lambda}{\mu} \frac{1}{\sqrt{2\pi}} \frac{f'_i \sigma^2}{D_Q} \\ &\times \int_{-\frac{f(\mathbf{y})-E_Q}{D_Q}}^{\infty} z e^{-\frac{1}{2}z^2} \sum_{m=1}^{\mu} \frac{(\lambda-1)!}{(m-1)!(\lambda-m)!} [\Phi(z)]^{m-1} [1-\Phi(z)]^{\lambda-m} dz. \end{aligned} \quad (2.100)$$

Now the lower boundary  $z_l = \frac{-f(\mathbf{y})-E_Q}{D_Q}$  and its scaling behavior with respect to  $N$  are investigated. Assuming (w.l.o.g.) that  $\mathbf{y} = \mathbf{1}$  and  $A = 0$ , the fitness term  $f(\mathbf{y}) = \|\mathbf{y}\|^2 = N$ . Noting that  $E_Q$  scales with  $N$  and  $D_Q$  with  $\sqrt{N}$ , the value for  $z_l$  scales with  $\sqrt{N}$  and we have

$$\lim_{N \rightarrow \infty} \frac{-f(\mathbf{y}) - E_Q}{D_Q} = -\infty. \quad (2.101)$$

After extending the integration range, identity (2.15) can be applied again to transform the sum into an integral

$$\begin{aligned} \varphi_i &= -\frac{\lambda}{\mu} \frac{1}{\sqrt{2\pi}} \frac{f'_i \sigma^2}{D_Q} \int_{z=-\infty}^{z=\infty} z e^{-\frac{1}{2}z^2} \\ &\times \frac{(\lambda-1)!}{(\lambda-\mu-1)!(\mu-1)!} \int_{t=0}^{t=1-\Phi(z)} t^{\lambda-\mu-1} (1-t)^{\mu-1} dt dz. \end{aligned} \quad (2.102)$$

Now the integration variables  $t$  and  $z$  can be exchanged giving the new ranges

$$0 \leq t \leq 1, \quad -\infty \leq z \leq \Phi^{-1}(1-t), \quad (2.103)$$

which changes the progress rate after reordering prefactors to

$$\begin{aligned} \varphi_i &= -\frac{f'_i \sigma^2}{D_Q} \frac{1}{\sqrt{2\pi}} \frac{\lambda}{\mu} \frac{(\lambda-1)!}{(\lambda-\mu-1)!(\mu-1)!} \\ &\times \int_{t=0}^{t=1} t^{\lambda-\mu-1} (1-t)^{\mu-1} \int_{z=-\infty}^{z=\Phi^{-1}(1-t)} z e^{-\frac{1}{2}z^2} dz dt. \end{aligned} \quad (2.104)$$

At this point another substitution is introduced, which will enable to identify the progress coefficient definition. Setting  $t = \Phi(y)$ ,  $y = \Phi^{-1}(t)$ ,  $\frac{dt}{dy} = \phi(y)$ , and evaluating the upper bound  $z_u = \Phi^{-1}(1-t) = \Phi^{-1}(\Phi(-y)) = -y$  we get

$$\begin{aligned} \varphi_i &= -\frac{f'_i \sigma^2}{D_Q} \\ &\times \frac{1}{\sqrt{2\pi}} \frac{\lambda}{\mu} \frac{(\lambda-1)!}{(\lambda-\mu-1)!(\mu-1)!} \int_{y=-\infty}^{y=\infty} \phi(y) [\Phi(y)]^{\lambda-\mu-1} [1-\Phi(y)]^{\mu-1} \\ &\times \int_{z=-\infty}^{z=-y} z e^{-\frac{1}{2}z^2} dz dy. \end{aligned} \quad (2.105)$$

The inner integration over  $z$  yields

$$\int_{-\infty}^{-y} z e^{-\frac{1}{2}z^2} dz = -e^{-\frac{1}{2}y^2}. \quad (2.106)$$

Using  $\phi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$  and  $\frac{\lambda}{\mu} \frac{(\lambda-1)!}{(\lambda-\mu-1)!(\mu-1)!} = (\lambda-\mu)\binom{\lambda}{\mu}$  we arrive at

$$\begin{aligned} \varphi_i &= \frac{f'_i \sigma^2}{D_Q} \frac{\lambda - \mu}{(\sqrt{2\pi})^2} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} e^{-y^2} [\Phi(y)]^{\lambda-\mu-1} [1 - \Phi(y)]^{\mu-1} dy \\ &= \frac{f'_i \sigma^2}{D_Q} c_{\mu/\mu, \lambda}, \end{aligned} \quad (2.107)$$

with the definition of  $c_{\mu/\mu, \lambda} = e_{\mu, \lambda}^{1,0}$  given by [5, p. 172]

$$e_{\mu, \lambda}^{a,b} = \frac{\lambda - \mu}{(2\pi)^{\frac{a+1}{2}}} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} x^b e^{-\frac{a+1}{2}x^2} [\Phi(x)]^{\lambda-\mu-1} [1 - \Phi(x)]^{\mu-a} dx. \quad (2.108)$$

The progress rate finally yields

Progress rate via  $c_{\mu/\mu, \lambda}$

$$\begin{aligned} \varphi_i &= c_{\mu/\mu, \lambda} \frac{\sigma^2}{D_Q} f'_i \\ &= c_{\mu/\mu, \lambda} \frac{\sigma^2}{D_Q} (k_i + d_i) \\ &= c_{\mu/\mu, \lambda} \frac{\sigma^2}{D_Q} (2y_i + \alpha A \sin(\alpha y_i)). \end{aligned} \quad (2.109)$$

with  $f'_i$  being the linearized quality gain given in Eq. (1.15) and  $D_Q^2$  the variance from Eq. (1.29).

### 3 Quadratic Progress Rate

The first order progress rate definition from Eq. (2.1) has a major disadvantage when approaching the optimizer  $\hat{y}_i = 0$ . It does not correctly model the progress if zero-crossings occur, namely if  $y_i^{(g)} > 0$  and  $y_i^{(g+1)} < 0$ , or along the negative axis for both  $y_i^{(g)} < 0$  and  $y_i^{(g+1)} < 0$ . In these cases the difference term  $y_i^{(g)} - y_i^{(g+1)} > 0$  yields positive progress even though the strategy is moving away from  $\hat{y}_i = 0$ .

Given this behavior, the convergence towards  $\hat{y}_i$  cannot be modeled correctly. At this point one can identify that squaring the individual components resolves this issue, which was already recognized in [6] studying the Ellipsoid model. Introducing the difference term  $(y_i^{(g)})^2 - (y_i^{(g+1)})^2$  models the progress consistently approaching zero from positive and negative axes.

Therefore the new quadratic progress measure for the  $i$ -th component reads

$$\varphi_i^{\text{II}} = E \left[ \left( y_i^{(g)} \right)^2 - \left( y_i^{(g+1)} \right)^2 \mid \mathbf{y}^{(g)}, \sigma^{(g)} \right]. \quad (3.1)$$

During the  $\varphi_i^{\text{II}}$  derivation the first order progress  $\varphi_i$  from previous section will reappear such that the obtained results will be reused.

#### 3.1 Definition

Starting with Eq. (2.2) again, the position vector at  $(g + 1)$  yields

$$\mathbf{y}^{(g+1)} = \mathbf{y}^{(g)} + \frac{1}{\mu} \sum_{m=1}^{\mu} \mathbf{x}_{m;\lambda}. \quad (3.2)$$

Referring to the  $i$ -th components  $y_i^{(g)}$ ,  $y_i^{(g+1)}$ ,  $x_{m;\lambda} = (\mathbf{x}_{m;\lambda})_i$  and squaring both sides gives

$$\begin{aligned} \left( y_i^{(g+1)} \right)^2 &= \left( y_i^{(g)} + \frac{1}{\mu} \sum_{m=1}^{\mu} x_{m;\lambda} \right)^2 \\ &= \left( y_i^{(g)} \right)^2 + 2y_i^{(g)} \frac{1}{\mu} \sum_{m=1}^{\mu} x_{m;\lambda} + \frac{1}{\mu^2} \left( \sum_{m=1}^{\mu} x_{m;\lambda} \right)^2. \end{aligned} \quad (3.3)$$

Squaring the last term needs additional treatment

$$\begin{aligned} \left( \sum_{m=1}^{\mu} x_{m;\lambda} \right)^2 &= \left( \sum_{k=1}^{\mu} x_{k;\lambda} \right) \left( \sum_{l=1}^{\mu} x_{l;\lambda} \right) = \sum_{m=1}^{\mu} (x_{m;\lambda})^2 + \sum_{k \neq l} x_{k;\lambda} x_{l;\lambda} \\ &= \sum_{m=1}^{\mu} (x_{m;\lambda})^2 + \sum_{k < l} x_{k;\lambda} x_{l;\lambda} + \sum_{k > l} x_{k;\lambda} x_{l;\lambda} \\ &= \sum_{m=1}^{\mu} (x_{m;\lambda})^2 + 2 \sum_{k < l} x_{k;\lambda} x_{l;\lambda} \\ &= \sum_{m=1}^{\mu} (x_{m;\lambda})^2 + 2 \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} x_{k;\lambda} x_{l;\lambda}. \end{aligned} \quad (3.4)$$

The idea was to split the squared sum into one part with  $\mu$  terms of equal indices and a double sum of  $\mu(\mu - 1)/2$  mixed terms.

Reordering Eq. (3.3) and inserting the result of Eq. (3.4) we obtain

$$\begin{aligned} & \left( y_i^{(g)} \right)^2 - \left( y_i^{(g+1)} \right)^2 \\ &= -2y_i^{(g)} \frac{1}{\mu} \sum_{m=1}^{\mu} x_{m;\lambda} - \frac{1}{\mu^2} \sum_{m=1}^{\mu} (x_{m;\lambda})^2 - \frac{2}{\mu^2} \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} x_{k;\lambda} x_{l;\lambda}. \end{aligned} \quad (3.5)$$

Now the expectation value can be taken with respect to mutation  $x$

$$\begin{aligned} & E \left[ \left( y_i^{(g)} \right)^2 - \left( y_i^{(g+1)} \right)^2 \right] \\ &= -2y_i^{(g)} \frac{1}{\mu} E \left[ \sum_{m=1}^{\mu} x_{m;\lambda} \right] - \frac{1}{\mu^2} E \left[ \sum_{m=1}^{\mu} (x_{m;\lambda})^2 \right] - \frac{2}{\mu^2} E \left[ \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} x_{k;\lambda} x_{l;\lambda} \right]. \end{aligned} \quad (3.6)$$

Identifying  $\varphi_i^{\text{II}} = E \left[ (y_i^{(g)})^2 - (y_i^{(g+1)})^2 \right]$  from Eq. (3.1) and  $\varphi_i = -\frac{1}{\mu} E \left[ \sum_{m=1}^{\mu} x_{m;\lambda} \right]$  from Eq. (2.6) we get the important intermediate result

$$\varphi_i^{\text{II}} = 2y_i^{(g)} \varphi_i - \frac{1}{\mu^2} E \left[ \sum_{m=1}^{\mu} (x_{m;\lambda})^2 \right] - \frac{2}{\mu^2} E \left[ \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} x_{k;\lambda} x_{l;\lambda} \right], \quad (3.7)$$

with the second order progress being a function of the first order progress, and two expectations yet to be determined. Therefore the following quantities are defined

$$E^{(2)} := \sigma^2 E \left[ \sum_{m=1}^{\mu} (z_{m;\lambda})^2 \right] \quad (3.8)$$

$$E^{(1,1)} := \sigma^2 E \left[ \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} z_{k;\lambda} z_{l;\lambda} \right], \quad (3.9)$$

with  $x = \sigma z$ , and the superscript  $(\cdot)$  denoting the power of occurring mutation terms. For the sake of completeness, the expectation yielding  $\varphi_i$  from Eq. (2.6) is also given

$$E^{(1)} := \sigma E \left[ \sum_{m=1}^{\mu} z_{m;\lambda} \right] = -\mu \varphi_i. \quad (3.10)$$

Solutions for Eq. (3.7) will be derived in Sec. 3.2 and 3.3 using two different approaches.

### 3.2 Expectations via Noisy Order Statistics

In this section, an approach by Arnold [2] is introduced and applied to calculate expectations from Eqs. (3.8), (3.9) and (3.10). Since the quantity  $\varphi_i$  of Eq. (3.10)

was already determined in Sec. 2.6, its result will be checked for correctness by applying the new approach. The following derivations give a summary of the more involved calculations by Arnold.

First the method will be presented in a more abstract form before being applied to the problem at hand. A slight complication arises due to two cases for the sign relation between quality gain and mutation, which influences the ordering of mutation components  $z_{m;\lambda}$ . This issue will result in slightly modified equations compared to the derivation by Arnold.

### 3.2.1 Definition and Generic Solution

Let  $z$  be a random variable with density  $p_z(z)$  and zero mean. The density is approximated using a finite number of cumulants using a Gram-Charlier series with variance  $\kappa_2$ , skewness  $\gamma_1$  and excess  $\gamma_2$ , see also (C.1). Furthermore, let  $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$  be a normally distributed random variable modeling additive noise disturbance. Resulting measured values  $v$  are obtained as

$$v = z + \mathcal{N}(0, \sigma_\epsilon^2), \quad (3.11)$$

of which the realizations are independent and identically distributed with density  $p_v(v)$  in (C.10). By performing selection of  $m = 1, \dots, \mu$  largest elements over  $\lambda$  realizations of  $v$ , the selected values are given by

$$v_{m;\lambda} = (z + \mathcal{N}(0, \sigma_\epsilon^2))_{m;\lambda}. \quad (3.12)$$

Therefore  $v_{m;\lambda}$  and its source term  $z_{m;\lambda}$  are being governed by order statistics, c.f. density  $p_{m;\lambda}(z)$  in Eq. (2.13). The variates  $z_{m;\lambda}$  are referred to as *noisy order statistics* due to the added noise term  $\mathcal{N}(0, \sigma_\epsilon^2)$  and are linearly related to measured (selected) values  $v_{m;\lambda}$ . After selection, the elements  $z_{m;\lambda}$  among the group  $m = 1, \dots, \mu$  now depend on each other with joint density denoted in (C.24).

Assuming we are interested in the expected value of a *sum*  $S_P$  of noisy order statistics over  $\nu$  factors with corresponding powers  $P = (p_1, \dots, p_\nu)$  and indices  $\{n_1, \dots, n_\nu\}$ , we can formulate the problem as

$$\mathbb{E}[S_P] = \mathbb{E}\left[\sum z_{n_1;\lambda}^{p_1} \cdots z_{n_\nu;\lambda}^{p_\nu}\right], \quad (3.13)$$

which is explained in (C.20) in more detail. Given the relation of Eq. (3.12) and the problem formulation of (3.13), Arnold has provided a generic solution for the expected value in [2].

As already mentioned, two cases for the sign relation between  $v$  and  $z$  may occur which will require minor modifications of Arnold's derivation. Denoting the sign as  $s$  we have

$$v = sz + \mathcal{N}(0, \sigma_\epsilon^2) \quad \text{with} \quad s \in \{+1, -1\}. \quad (3.14)$$

The details concerning the modified derivations are shown in Appendix C. At this point only a summary is provided.

**Summary of solution** Arnold aims to find a closed form expression for the  $(2\nu + 1)$ -fold integration of Eq. (C.26). Therefore he successively solves the  $2\nu$ -fold integration for  $I_P(x)$  over  $z$  and  $v$ -variables, respectively, presented in (C.27) and (C.28). To this end, the noise coefficient  $a > 0$  from Eq. (C.11) is defined and special coefficients  $\zeta_{i,j}^{(P)}(k)$  are introduced, such that

$$a = \sqrt{\frac{\kappa_2}{\kappa_2 + \sigma_\epsilon^2}}, \quad \text{with} \quad (3.15)$$

$$\zeta_{i,j}^{(P)}(k) = \text{Polynomial}(a). \quad (3.16)$$

Integral  $I_P(x)$  is expressed in terms of the coefficients  $\zeta_{i,j}^{(P)}(k)$ , and they in turn depend on the exponent vector  $P$ , on the number of terms  $\nu$  via index  $i$  and on the expansion order of Eq. (C.1) via  $j$ . Given  $\{P, i, j\}$ , only certain  $k \geq 0$  yield non-zero coefficient values. The results were obtained using *Mathematica* and are tabulated by Arnold [2, p. 141]. In the end we are left with a single integration, which is incorporated within a new coefficient  $h_{\mu,\lambda}^{\nu-n,k}$ , see (3.18).

The final result for the expected value of sum (3.13) in terms of  $z$  defined in (3.14) with  $s \in \{+1, -1\}$  reads

$$\begin{aligned} \mathbb{E}[S_P] &= (s\sqrt{\kappa_2})^{\|P\|_1} \frac{\mu!}{(\mu - \nu)!} \\ &\times \sum_{n=0}^{\nu} \sum_{k \geq 0} \left( \zeta_{n,0}^{(P)}(k) + s \frac{\gamma_1}{6} \zeta_{n,1}^{(P)}(k) + \frac{\gamma_2}{24} \zeta_{n,2}^{(P)}(k) + \dots \right) h_{\mu,\lambda}^{\nu-n,k}. \end{aligned} \quad (3.17)$$

The relation  $v \sim -z + \mathcal{N}(0, \sigma_\epsilon^2)$  from Eq. (3.14) results in a switched sign for skewness  $\gamma_1$  (neglecting cumulants of higher order than four) and exchanged permutation ordering with  $(-1)^{\|P\|_1}$  compared to Arnold's result in [2, p. 142, D.28].

The introduced coefficients  $h_{\mu,\lambda}^{i,k}$  are numerically obtainable solving

$$h_{\mu,\lambda}^{i,k} = \frac{\lambda - \mu}{\sqrt{2\pi}} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} \text{He}_k(x) e^{-\frac{1}{2}x^2} [\phi(x)]^i [\Phi(x)]^{\lambda - \mu - 1} [1 - \Phi(x)]^{\mu - i} dx. \quad (3.18)$$

They are closely related to the generalized progress coefficient by Beyer [5, p. 172], with the definition here using Hermite Polynomials. As an example,  $h_{\mu,\lambda}^{1,0} = c_{\mu/\mu,\lambda}$  with  $\text{He}_0(x) = 1$ .

Given  $z \sim \mathcal{N}(0, \sigma^2)$  expansion (C.1) yields a normal distribution with  $\gamma_1 = \gamma_2 = 0$  and only coefficient  $\zeta_{n,0}^{(P)}(k)$  remains within brackets  $(\cdot)$  of Eq. (3.17).

### 3.2.2 Evaluating Expectations

In this section the result of Eq. (3.17) is applied to the sums of Eqs. (3.8) and (3.9) for the determination of  $\varphi_i^{\text{II}}$ . But first Eq. (3.10) is evaluated to validate the correct application of the method.

The first step is to redefine the fitness and quality gain in order to achieve maximization instead of minimization. Applying  $(-1)$  to the fitness leads to a maximization problem (denoted by overset  $\sim$ ). Thus, the maximized quality gain of Eq. (1.2) is

$$\tilde{Q}_y(\mathbf{x}) = -f(\mathbf{y} + \mathbf{x}) - (-f(\mathbf{y})) = f(\mathbf{y}) - f(\mathbf{y} + \mathbf{x}) \quad (3.19)$$

$$= -Q_y(\mathbf{x}) = -\sum_{i=1}^N Q_i(x_i). \quad (3.20)$$

Therefore the  $i$ -th component of Eq. (1.9) changes sign

$$\tilde{Q}_i(x_i) = -(x_i^2 + 2y_i x_i + c_i(1 - \cos(\alpha x_i)) + s_i \sin(\alpha x_i)). \quad (3.21)$$

Evaluating the expectation value and variance yields

$$E[\tilde{Q}_y(\mathbf{x})] = E\left[\sum_{i=1}^N \tilde{Q}_i\right] = -E\left[\sum_{i=1}^N Q_i\right] = -E_Q \quad (3.22)$$

$$\text{Var}[\tilde{Q}_y(\mathbf{x})] = \text{Var}\left[\sum_{i=1}^N \tilde{Q}_i\right] = (-1)^2 \text{Var}\left[\sum_{i=1}^N Q_i\right] = D_Q^2. \quad (3.23)$$

We are interested in the  $i$ -th component expectation  $E[\sum_{m=1}^\mu x_{m;\lambda}]$  with  $x_{m;\lambda} := (\mathbf{x}_{m;\lambda})_i$  due to quality gain selection. Now the main idea is to isolate the  $i$ -th component of the quality gain, linearize it and to model the effect of  $N - 1$  remaining components as a noisy disturbance. The goal is obtaining the form of Eq. (3.14). Starting with

$$\begin{aligned} \tilde{Q}_y(\mathbf{x}) &= \sum_{j=1}^N \tilde{Q}_j = \tilde{Q}_i + \sum_{j \neq i} \tilde{Q}_j \\ &= -x_i^2 - 2y_i x_i - c_i(1 - \cos(\alpha x_i)) - s_i \sin(\alpha x_i) + \sum_{j \neq i} \tilde{Q}_j \quad (3.24) \\ &= -k_i x_i - \delta(x_i) + \sum_{j \neq i} \tilde{Q}_j, \end{aligned}$$

where in the last line Eqs. (2.35) and (2.36) were used. As a linear relation between  $\tilde{Q}_y(\mathbf{x})$  and  $x_i$  is required, the perturbation is neglected assuming  $\delta(x_i) \approx 0$  for small mutations.

For large dimensionality the sum  $\sum_{j \neq i} \tilde{Q}_j$  asymptotically approaches a normal distribution, cf. Eq. (1.39), and we can define a substitute random variate  $w$

$$\begin{aligned} w &:= \sum_{j \neq i} \tilde{Q}_j, \quad w \sim \mathcal{N}(E_w, D_w^2), \text{ with} \\ E_w &= E\left[\sum_{j \neq i} \tilde{Q}_j\right] = -E_i, \quad D_w^2 = \text{Var}\left[\sum_{j \neq i} \tilde{Q}_j\right] = D_i^2, \end{aligned} \quad (3.25)$$

where Eqs. (1.47) and (1.49) were used with sign relation (3.22) applied to  $E_i$ .

Setting  $x_i = \sigma z_i$ , neglecting  $\delta(x_i)$  and using substitute  $w$ , Eq. (3.24) now becomes

$$\begin{aligned}\tilde{Q}_y(\mathbf{x}) &\approx -k_i \sigma z_i + \mathcal{N}(-E_i, D_i^2) \\ \tilde{Q}_y(\mathbf{x}) &\approx -k_i \sigma z_i - E_i + \mathcal{N}(0, D_i^2) \\ \tilde{Q}_y(\mathbf{x}) + E_i &\approx -\operatorname{sgn}(k_i) |k_i| \sigma z_i + \mathcal{N}(0, D_i^2) \\ \frac{\tilde{Q}_y(\mathbf{x}) + E_i}{|k_i| \sigma} &\approx \operatorname{sgn}(-k_i) z_i + \mathcal{N}\left(0, \left(\frac{D_i}{k_i \sigma}\right)^2\right).\end{aligned}\quad (3.26)$$

The sign function  $\operatorname{sgn}(\cdot)$  and the decomposition  $k_i = \operatorname{sgn}(k_i) |k_i|$  were introduced, as  $k_i = 2y_i$  may be positive or negative depending on the position. In order to ensure constant order of selected values on the left-hand side, the equation is divided by  $|k_i|$  keeping the sign relation at  $z_i$ . As  $|k_i|^2 = k^2$  the absolute value is dropped when squared. By defining

$$\begin{aligned}v_i &:= (\tilde{Q}_y(\mathbf{x}) + E_i) / (|k_i| \sigma) \\ \sigma_\epsilon^2 &:= (D_i / k_i \sigma)^2,\end{aligned}\quad (3.27)$$

we arrive at the desired form introduced in Eq. (3.14) with  $s = \operatorname{sgn}(-k_i)$  giving

$$v_i = \operatorname{sgn}(-k_i) z_i + \mathcal{N}(0, \sigma_\epsilon^2). \quad (3.28)$$

We have  $z_i \sim \mathcal{N}(0, 1)$  such that  $\kappa_2 = 1$ ,  $\kappa_i = 0$  for  $i \neq 2$ . The ordering of maximized  $\tilde{Q}_y(\mathbf{x})$  is not affected by the linear transformation  $v_i(\tilde{Q}_y(\mathbf{x}))$ .

Now the obtained results are applied. Given the sum (3.10)

$$E^{(1)} = \sigma \operatorname{E} \left[ \sum_{m=1}^{\mu} z_{m;\lambda} \right], \quad (3.29)$$

relevant parameters for applying Eq. (3.17) are  $\kappa_2 = 1$ ,  $\gamma_1 = \gamma_2 = 0$ ,  $P = (1)$ ,  $\|P\|_1 = 1$  and  $\nu = 1$ , such that

$$\begin{aligned}E^{(1)} &= \sigma \operatorname{sgn}(-k_i) \frac{\mu!}{(\mu-1)!} \sum_{n=0}^1 \sum_{k \geq 0} \zeta_{n,0}^{(1)}(k) h_{\mu,\lambda}^{1-n,k} \\ &= \sigma \operatorname{sgn}(-k_i) \mu \sum_{k \geq 0} \left( \zeta_{0,0}^{(1)}(k) h_{\mu,\lambda}^{1,k} + \zeta_{1,0}^{(1)}(k) h_{\mu,\lambda}^{0,k} \right) \\ &= \sigma \operatorname{sgn}(-k_i) \mu \zeta_{0,0}^{(1)}(0) h_{\mu,\lambda}^{1,0} \\ &= \sigma \operatorname{sgn}(-k_i) \mu a h_{\mu,\lambda}^{1,0}.\end{aligned}\quad (3.30)$$

The result was obtained applying Table [2, p. 141] by using  $\zeta_{1,0}^{(1)}(k) = 0$  for any  $k$ , and  $\zeta_{0,0}^{(1)}(k) \neq 0$  only for  $k = 0$  giving the noise coefficient  $a$ . The result (3.30)

is further evaluated using  $a$  from (3.15) and  $h_{\mu,\lambda}^{1,0}$  from (3.18)

$$\begin{aligned}
E^{(1)} &= \sigma \operatorname{sgn}(-k_i) \mu a h_{\mu,\lambda}^{1,0} \\
&= \sigma \operatorname{sgn}(-k_i) \mu \frac{c_{\mu/\mu,\lambda}}{\sqrt{1 + \sigma_\epsilon^2}} \\
&= \frac{\sigma \operatorname{sgn}(-k_i) \mu c_{\mu/\mu,\lambda}}{\sqrt{1 + \left(\frac{D_i}{k_i \sigma}\right)^2}} \\
&= \frac{\sigma \operatorname{sgn}(-k_i) \mu c_{\mu/\mu,\lambda}}{\sqrt{\left(\frac{k_i \sigma}{k_i \sigma}\right)^2 + \left(\frac{D_i}{k_i \sigma}\right)^2}} \\
&= -\frac{\mu c_{\mu/\mu,\lambda} \operatorname{sgn}(k_i) |k_i| \sigma^2}{\sqrt{(k_i \sigma)^2 + D_i^2}} \\
&= -\mu \frac{c_{\mu/\mu,\lambda} k_i \sigma^2}{D_+}.
\end{aligned} \tag{3.31}$$

From first to second line it was used that  $h_{\mu,\lambda}^{1,0} = c_{\mu/\mu,\lambda}$  and  $a = \sqrt{1/(1 + \sigma_\epsilon^2)}$ . In the third line  $\sigma_\epsilon^2 = (D_i/k_i \sigma)^2$ . From fourth to fifth line in order to ensure the noise coefficient  $a > 0$ , taking  $\sqrt{(k_i \sigma)^2}$  was set to  $+|k_i| \sigma$  and the corresponding positive solution is chosen. For the last line the definition  $k_i = \operatorname{sgn}(k_i) |k_i|$  was applied again, as well as the definition of  $D_+$  from Eq. (2.49).

Relating the result of (3.31) to progress rate  $\varphi_i$  of Eq. (3.10) and applying the large  $N$  approximation  $D_+ \approx D_Q$ , see Eq. (2.73), we finally arrive at the progress rate by means of noisy order statistics

$$\begin{aligned}
\varphi_i &= -\frac{1}{\mu} E^{(1)} \\
&= c_{\mu/\mu,\lambda} \frac{\sigma^2}{D_Q} k_i.
\end{aligned} \tag{3.32}$$

The result is equivalent to Eq. (2.109) showing the progress of the spherical part  $k_i$  due to neglected perturbations. If the linearized quality gain from Eq. (1.15) was chosen in Eq. (3.24), the coefficient  $k_i$  would be replaced by  $f'_i$ .

Therefore the method has been validated using the already obtained result. Now the remaining expectations are evaluated.

**Expectations of  $E^{(2)}$  and  $E^{(1,1)}$**  For the sum defined in Eq. (3.8)

$$E^{(2)} = \sigma^2 \mathbb{E} \left[ \sum_{m=1}^{\mu} (z_{m;\lambda})^2 \right], \tag{3.33}$$

we now have  $P = (2)$ ,  $\|P\|_1 = 2$  and  $\nu = 1$ . Distribution parameters  $\kappa_2 = 1$ ,  $\gamma_1 = \gamma_2 = 0$  remain the same. As  $\|P\|_1 = 2$ , the sign yields  $s^2 = [\operatorname{sgn}(-k_i)]^2 = 1$

and can be dropped. The result of Eq. (3.17) yields

$$\begin{aligned}
E^{(2)} &= \sigma^2 \frac{\mu!}{(\mu-1)!} \sum_{n=0}^1 \sum_{k \geq 0} \zeta_{n,0}^{(2)}(k) h_{\mu,\lambda}^{1-n,k} \\
&= \sigma^2 \mu \sum_{k \geq 0} \left( \zeta_{0,0}^{(2)}(k) h_{\mu,\lambda}^{1,k} + \zeta_{1,0}^{(2)}(k) h_{\mu,\lambda}^{0,k} \right) \\
&= \sigma^2 \mu \left( \zeta_{0,0}^{(2)}(1) h_{\mu,\lambda}^{1,1} + \zeta_{1,0}^{(2)}(0) h_{\mu,\lambda}^{0,0} \right) \\
&= \sigma^2 \mu \left( a^2 h_{\mu,\lambda}^{1,1} + h_{\mu,\lambda}^{0,0} \right) \\
&= \sigma^2 \mu \left( a^2 e_{\mu,\lambda}^{1,1} + 1 \right) \\
&= \sigma^2 \mu \left( \frac{e_{\mu,\lambda}^{1,1}}{1 + \left( \frac{D_i}{k_i \sigma} \right)^2} + 1 \right) \\
&= \mu \left( \frac{e_{\mu,\lambda}^{1,1} k_i^2 \sigma^4}{D_+^2} + \sigma^2 \right).
\end{aligned} \tag{3.34}$$

For  $\sum_{k \geq 0}$  it was used that  $\zeta_{n,0}^{(2)}(k) = 0$  except  $\zeta_{0,0}^{(2)}(1) = a^2$  and  $\zeta_{1,0}^{(2)}(0) = 1$ . The coefficient  $h_{\mu,\lambda}^{1,1} = e_{\mu,\lambda}^{1,1}$  given  $\text{He}_1(x) = x$  and comparing Eqs. (3.18) and (2.108). Additionally  $h_{\mu,\lambda}^{0,0} = 1$  with  $\text{He}_0(x) = 1$ , see below. The value of  $a^2 = 1/(1 + (D_i/k_i \sigma)^2)$ , and finally the definition  $D_+^2 = (k_i \sigma)^2 + D_i^2$  was used to obtain the last line.

Coefficient  $h_{\mu,\lambda}^{0,0}$  can be evaluated as

$$\begin{aligned}
h_{\mu,\lambda}^{0,0} &= \frac{\lambda - \mu}{\sqrt{2\pi}} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} [\Phi(x)]^{\lambda-\mu-1} [1 - \Phi(x)]^{\mu} dx \\
&= (\lambda - \mu) \binom{\lambda}{\mu} \int_0^1 t^{\lambda-\mu-1} (1-t)^{\mu} dt \\
&= (\lambda - \mu) \binom{\lambda}{\mu} B(\lambda - \mu, \mu + 1) \\
&= \frac{\lambda!}{(\lambda - \mu - 1)! \mu!} \frac{(\lambda - \mu - 1)! \mu!}{\lambda!} \\
&= 1,
\end{aligned} \tag{3.35}$$

using the substitution  $t = \Phi(x)$ , transformation  $(\lambda - \mu) \binom{\lambda}{\mu} = \frac{\lambda!}{(\lambda - \mu - 1)! \mu!}$ , the relation between beta and gamma functions  $B(x, y) = \frac{\Gamma(x) \Gamma(y)}{\Gamma(x+y)}$  and  $\Gamma(n) = (n-1)!$ .

Given the sum defined in Eq. (3.9)

$$E^{(1,1)} := \sigma^2 E \left[ \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} z_{k;\lambda} z_{l;\lambda} \right], \tag{3.36}$$

we have  $P = (1, 1)$ ,  $\|P\|_1 = 2$ ,  $\nu = 2$  and  $s^2 = 1$ . Distribution parameters are

$\kappa_2 = 1$  and  $\gamma_1 = \gamma_2 = 0$ . Equation (3.17) yields

$$\begin{aligned}
E^{(1,1)} &= \sigma^2 \frac{\mu!}{(\mu-2)!} \sum_{n=0}^2 \sum_{k \geq 0} \zeta_{n,0}^{(1,1)}(k) h_{\mu,\lambda}^{2-n,k} \\
&= \sigma^2 \mu(\mu-1) \sum_{k \geq 0} \left( \zeta_{0,0}^{(1,1)}(k) h_{\mu,\lambda}^{2,k} + \zeta_{1,0}^{(1,1)}(k) h_{\mu,\lambda}^{1,k} + \zeta_{2,0}^{(1,1)}(k) h_{\mu,\lambda}^{0,k} \right) \\
&= \sigma^2 \mu(\mu-1) \zeta_{0,0}^{(1,1)}(0) h_{\mu,\lambda}^{2,0} \\
&= \sigma^2 \mu(\mu-1) \frac{a^2}{2} e_{\mu,\lambda}^{2,0} \\
&= \frac{\sigma^2 \mu(\mu-1)}{2} \frac{e_{\mu,\lambda}^{2,0}}{1 + \left( \frac{D_i}{k_i \sigma} \right)^2} \\
&= \frac{\sigma^2 \mu(\mu-1)}{2} \frac{e_{\mu,\lambda}^{2,0} (k_i \sigma)^2}{D_+^2} \\
&= \frac{\mu(\mu-1)}{2} \frac{e_{\mu,\lambda}^{2,0} k_i^2 \sigma^4}{D_+^2}.
\end{aligned} \tag{3.37}$$

It was used that  $\zeta_{n,0}^{(1,1)}(k) = 0$  except  $\zeta_{0,0}^{(1,1)}(0) = a^2/2$  and  $h_{\mu,\lambda}^{2,0} = e_{\mu,\lambda}^{2,0}$  comparing Eqs. (3.18) and (2.108). Again,  $a^2 = 1/(1 + (D_i/k_i \sigma)^2)$  and  $D_+^2 = (k_i \sigma)^2 + D_i^2$  was used.

**Collecting results** At this point the results of Eqs. (3.34) and (3.37) are plugged back into Eq. (3.7) giving

$$\begin{aligned}
\varphi_i^{\text{II}} &= 2y_i \varphi_i - \frac{1}{\mu^2} E^{(2)} - \frac{2}{\mu^2} E^{(1,1)} \\
&= 2y_i \varphi_i - \frac{1}{\mu} \left( \frac{e_{\mu,\lambda}^{1,1} k_i^2 \sigma^4}{D_+^2} + \sigma^2 \right) - \frac{\mu-1}{\mu} \frac{e_{\mu,\lambda}^{2,0} k_i^2 \sigma^4}{D_+^2} \\
&= 2y_i \varphi_i - \frac{\sigma^2}{\mu} - \frac{1}{\mu} \frac{k_i^2 \sigma^4}{D_+^2} \left( e_{\mu,\lambda}^{1,1} + (\mu-1) e_{\mu,\lambda}^{2,0} \right).
\end{aligned} \tag{3.38}$$

At this point the expression (2.82) for  $\varphi_i$  is inserted into  $\varphi_i^{\text{II}}$ . Additionally, using definition  $k_i = 2y_i$  from Eq. (1.12), applying variance approximation  $D_+ \approx D_Q^2$  for large  $N$  from (2.74) and collecting the factor  $\sigma^2/\mu$  one gets the quadratic progress rate result

$$\begin{aligned}
\varphi_i^{\text{II}} &= c_\vartheta \frac{2y_i \sigma^2}{D_Q} \left( 2y_i + e^{-\frac{1}{2}(\alpha \sigma)^2} \alpha A \sin(\alpha y_i) \right) \\
&\quad - \frac{\sigma^2}{\mu} \left[ 1 + \frac{(2y_i)^2 \sigma^2}{D_Q^2} \left( e_{\mu,\lambda}^{1,1} + (\mu-1) e_{\mu,\lambda}^{2,0} \right) \right].
\end{aligned} \tag{3.39}$$

### 3.3 Expectations via Large Population Approximation

In this section an alternative derivation to Sec. 3.2 is presented for the expectations (3.8) and (3.9) for the derivation of  $\varphi_i^{\text{II}}$  in Eq. (3.7). The large population approximation will be applied for the expectations analogous to Sec. 2.2.

#### 3.3.1 Expectation of $E^{(2)}$

Starting from (3.8) and referring to the corresponding term in (3.7) one has

$$\begin{aligned} \frac{1}{\mu^2} E^{(2)} &= \frac{1}{\mu^2} \sum_{m=1}^{\mu} \mathbb{E} [x_{m;\lambda}^2] \\ &= \frac{1}{\mu^2} \sum_{m=1}^{\mu} \int_{-\infty}^{\infty} x_i^2 p_{m;\lambda}(x_i | \mathbf{y}) dx_i \end{aligned} \quad (3.40)$$

analogous to the definition in Eq. (2.8), but with squared quantity  $x_i^2$ . Both (2.8) and (3.40) have the same structure after inserting the order statistic density  $p_{m;\lambda}(x_i | \mathbf{y})$  from (2.13) and the integration over the changed mutation component is performed at last. The results of Sec. 2.1 and 2.2 can be applied to Eq. (3.40) by including the large population approximation up to result (2.42), for which the mutation integration has to be performed.

Therefore (3.40) is rewritten as

$$\begin{aligned} \frac{1}{\mu^2} E^{(2)} &= \frac{1}{\mu} \left[ \frac{1}{\mu} \sum_{m=1}^{\mu} \int_{-\infty}^{\infty} x_i^2 p_{m;\lambda}(x_i | \mathbf{y}) dx_i \right] \\ &= \frac{1}{\mu} \left[ \frac{\lambda}{\mu} \int_{x_i=-\infty}^{x_i=\infty} x_i^2 p_x(x_i) \frac{1}{B(\lambda - \mu, \mu)} \right. \\ &\quad \times \left. \int_{t=0}^{t=1} t^{\lambda - \mu - 1} (1-t)^{\mu-1} P_Q(P_Q^{-1}(1-t) | x_i) dt dx_i \right], \end{aligned} \quad (3.41)$$

and an equation analogous to (2.25) is obtained. Solving the  $t$ -integration in Eq. (3.41), the large population approximation of (B.1) is applied with  $a = 1$  and the integrand  $P_Q(P_Q^{-1}(1-t) | x_i)$  evaluated at  $\hat{t} = 1 - \vartheta$ . This yields

$$\frac{1}{\mu^2} E^{(2)} \simeq \frac{1}{\mu} \frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i^2 p_x(x_i) P_Q(P_Q^{-1}(\vartheta) | x_i) dx_i, \quad (3.42)$$

which is analogous to (2.28). Inserting the normal approximation of the quality gain distribution (2.31) into (3.42) leads to an analytically not solvable integration due to non-linear terms in  $x_i$  within  $\Phi(\cdot)$ . Given  $\Phi(\cdot)$ , see (2.37), with  $g(x_i)$  from (2.38) being a linear function in  $x_i$  and  $h(x_i)$  from (2.40) a non-linear function, one can write

$$\frac{1}{\mu^2} E^{(2)} = \frac{1}{\mu} \frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i^2 p_x(x_i) \Phi(g(x_i) + h(x_i)) dx_i. \quad (3.43)$$

At this point the resulting distribution function is again expanded according to (2.41) considering only the first two terms of the Taylor series. One obtains an

approximation analogous to (2.42) giving

$$\begin{aligned} \frac{1}{\mu^2} E^{(2)} &\approx \frac{1}{\mu} \left[ \frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i^2 p_x(x_i) \Phi(g(x_i)) dx_i \right. \\ &\quad \left. + \frac{1}{\sqrt{2\pi}\vartheta} \int_{-\infty}^{\infty} x_i^2 p_x(x_i) h(x_i) e^{-\frac{1}{2}g(x_i)^2} dx_i \right] \\ &=: I_i^0 + I_i^1, \end{aligned} \quad (3.44)$$

with the two integrations abbreviated as  $I_i^0$  and  $I_i^1$ , which are evaluated now. As a remark, the sign difference compared to (2.42) results from  $\varphi_i = -\frac{1}{\mu} \sum_{m=1}^{\mu} E[x_{m;\lambda}]$  calculated in (2.42), while for (3.44) the expression  $\frac{1}{\mu^2} \sum_{m=1}^{\mu} E[x_{m;\lambda}^2]$  is evaluated.

Starting with the first integration  $I_i^0$ , it is rewritten analogously to (2.44) using  $g(x_i)$  from (2.38) and the substitution  $z = x_i/\sigma$  giving

$$\begin{aligned} I_i^0 &= \frac{1}{\mu\vartheta} \int_{-\infty}^{\infty} x_i^2 p_x(x_i) \Phi(g(x_i)) dx_i \\ &= \frac{\sigma^2}{\sqrt{2\pi}\mu\vartheta} \int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}z^2} \Phi\left(-\frac{k_i\sigma}{D_i}z + \frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_i}\right) dz. \end{aligned} \quad (3.45)$$

At this point the result of Identity (D.1) can be applied to (3.45). Defining the coefficients

$$a = -\frac{k_i\sigma}{D_i}, \quad b = \frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_i}, \quad (3.46)$$

expressions needed for (D.1) are evaluated as

$$\begin{aligned} (1+a^2)^{1/2} &= \sqrt{\frac{D_i^2}{D_i^2} + \left(\frac{k_i\sigma}{D_i}\right)^2} = \sqrt{\frac{D_+^2}{D_i^2}} = \frac{D_k}{D_i} \\ \frac{a^2 b}{(1+a^2)^{3/2}} &= \frac{(k_i\sigma)^2}{D_+^2} \frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_+} \\ e^{-\frac{1}{2}\frac{b^2}{1+a^2}} &= \exp\left[-\frac{1}{2} \left(\frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_+}\right)^2\right], \end{aligned} \quad (3.47)$$

using  $D_+^2 := D_i^2 + (k_i\sigma)^2$  from (2.49). Therefore the first integration yields

$$\begin{aligned} I_i^0 &= \frac{\sigma^2}{\mu\vartheta} \left\{ \Phi\left(\frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_+}\right) - \right. \\ &\quad \left. \frac{1}{\sqrt{2\pi}} \frac{(k_i\sigma)^2}{D_+^2} \frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_+} \exp\left[-\frac{1}{2} \left(\frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_+}\right)^2\right] \right\}. \end{aligned} \quad (3.48)$$

Applying the large dimensionality approximation from (2.73), one neglects  $E_{Q_i} \approx$

0 and sets  $D_Q \approx D_+$ . This significantly simplifies the result of (3.48) giving

$$\begin{aligned} I_i^0 &\approx \frac{\sigma^2}{\mu\vartheta} \left[ \Phi(\Phi^{-1}(\vartheta)) - \frac{1}{\sqrt{2\pi}} \frac{(k_i\sigma)^2}{D_Q^2} \Phi^{-1}(\vartheta) \exp \left[ -\frac{1}{2} [\Phi^{-1}(\vartheta)]^2 \right] \right] \\ &\approx \frac{\sigma^2}{\mu} \left[ 1 - \Phi^{-1}(\vartheta) \left[ \frac{e^{-\frac{1}{2}[\Phi^{-1}(\vartheta)]^2}}{\sqrt{2\pi}\vartheta} \right] \frac{(k_i\sigma)^2}{D_Q^2} \right] \end{aligned} \quad (3.49)$$

Given (3.49) the asymptotic generalized progress coefficient  $e_\vartheta^{1,1}$  from (B.30) can be recognized with parameters  $a = 1$  and  $b = 1$ , such that

$$e_\vartheta^{1,1} = [-\Phi^{-1}(\vartheta)] \left[ \frac{e^{-\frac{1}{2}[\Phi^{-1}(\vartheta)]^2}}{\sqrt{2\pi}\vartheta} \right]. \quad (3.50)$$

This leads to following result for the first integral

$$I_i^0 = \frac{\sigma^2}{\mu} \left[ 1 + e_\vartheta^{1,1} \frac{(k_i\sigma)^2}{D_Q^2} \right]. \quad (3.51)$$

Note that the large dimensionality approximation can also be applied earlier, namely to coefficients (3.47), which results in simpler expressions

$$\begin{aligned} (1+a^2)^{1/2} &\approx 1 \\ \frac{a^2 b}{(1+a^2)^{3/2}} &\approx \frac{(k_i\sigma)^2}{D_Q^2} \Phi^{-1}(\vartheta) \\ e^{-\frac{1}{2}\frac{b^2}{1+a^2}} &\approx \exp \left[ -\frac{1}{2} [\Phi^{-1}(\vartheta)]^2 \right], \end{aligned} \quad (3.52)$$

when evaluating Identity (D.1).

Second integration  $I_i^1$  from (3.44) is defined as

$$I_i^1 = \frac{1}{\sqrt{2\pi}\mu\vartheta} \int_{-\infty}^{\infty} x_i^2 p_x(x_i) h(x_i) e^{-\frac{1}{2}g(x_i)^2} dx_i \quad (3.53)$$

with  $g(x_i)$  and  $h(x_i)$  defined in (2.38) and (2.55), respectively, giving

$$\begin{aligned} g(x_i) &= -\frac{k_i}{D_i} x_i + \frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_i} \\ h(x_i) &= -\frac{x_i^2 + s_i \sin(\alpha x_i) + c_i(1 - \cos(\alpha x_i))}{D_i}. \end{aligned} \quad (3.54)$$

Quadratic completion for the Gaussians of (3.53) was already evaluated in Eq. (2.54) with parameters  $C$ ,  $m$  and  $s$  given in (2.65), (2.63) and (2.64), respectively. Again, the large dimensionality approximation is applied to simplify the lengthy expressions and the results of Sec. 2.4 are applicable giving

$$\begin{aligned} m &\approx 0 \\ s &\approx \sigma \\ C &\approx \exp \left[ -\frac{1}{2} [\Phi^{-1}(\vartheta)]^2 \right]. \end{aligned} \quad (3.55)$$

Therefore integral (3.53) with quadratic completion (2.54) assuming large  $N$  yields

$$\begin{aligned} I_i^1 &= \frac{1}{\mu} \frac{C}{2\pi\vartheta\sigma} \int_{-\infty}^{\infty} x_i^2 h(x_i) e^{-\frac{1}{2}(\frac{x_i-m}{\sigma})^2} dx_i \\ &\approx \frac{1}{\mu} \frac{e^{-\frac{1}{2}[\Phi^{-1}(\vartheta)]^2}}{\sqrt{2\pi}\vartheta} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x_i^2 h(x_i) e^{-\frac{1}{2}\frac{x_i^2}{\sigma^2}} dx_i \end{aligned} \quad (3.56)$$

Given last line of (3.56), one can compare coefficients with the asymptotic generalized progress coefficient from (B.30) and identify following using  $a = 1$  and  $b = 0$

$$\frac{e^{-\frac{1}{2}[\Phi^{-1}(\vartheta)]^2}}{\sqrt{2\pi}\vartheta} = e_{\vartheta}^{1,0} = c_{\vartheta}, \quad (3.57)$$

see also Eq. (2.79). Additionally in (3.56), the definition of the expected value of  $x_i^2 h(x_i)$  w.r.t.  $x_i \sim \mathcal{N}(0, \sigma^2)$  can be applied. Inserting  $h(x_i)$  from (3.54) with  $D_i \approx D_Q$ , expression (3.56) is reformulated

$$I_i^1 = -\frac{c_{\vartheta}}{\mu D_Q} [\mathbb{E}[x_i^4] + s_i \mathbb{E}[x_i^2 \sin(\alpha x_i)] + c_i \mathbb{E}[x_i^2] - c_i \mathbb{E}[x_i^2 \cos(\alpha x_i)]]. \quad (3.58)$$

with  $c_i = A \cos(\alpha y_i)$  and  $s_i = A \sin(\alpha y_i)$ . Using results from Appendix A the expected values read

$$\begin{aligned} \mathbb{E}[x_i^4] &= 3\sigma^4 \\ \mathbb{E}[x_i^2] &= \sigma^2 \\ \mathbb{E}[x_i^2 \sin(\alpha x_i)] &= 0 \\ \mathbb{E}[x_i^2 \cos(\alpha x_i)] &= (\sigma^2 - \alpha^2 \sigma^4) e^{-\frac{1}{2}(\alpha\sigma)^2}. \end{aligned} \quad (3.59)$$

Therefore one gets

$$I_i^1 = -\frac{c_{\vartheta}\sigma^2}{\mu D_Q} [3\sigma^2 + A \cos(\alpha y_i) \left(1 - e^{-\frac{1}{2}(\alpha\sigma)^2} + \alpha^2 \sigma^2 e^{-\frac{1}{2}(\alpha\sigma)^2}\right)]. \quad (3.60)$$

Collecting the results (3.51) and (3.60) with  $k_i = 2y_i$  and inserting them back into (3.44) the expectation value reads

$$\begin{aligned} \frac{1}{\mu^2} E^{(2)} &= \frac{\sigma^2}{\mu} \left\{ 1 + e_{\vartheta}^{1,1} \frac{(2y_i)^2 \sigma^2}{D_Q^2} - \frac{c_{\vartheta}}{D_Q} \left[ 3\sigma^2 \right. \right. \\ &\quad \left. \left. + A \cos(\alpha y_i) \left(1 - e^{-\frac{1}{2}(\alpha\sigma)^2} + \alpha^2 \sigma^2 e^{-\frac{1}{2}(\alpha\sigma)^2}\right) \right] \right\}. \end{aligned} \quad (3.61)$$

### 3.3.2 Expectation of $E^{(1,1)}$

The derivation of  $E^{(1,1)}$  is tackled in this section. Referring to (3.9) and its application in (3.7), the expected value to be calculated is

$$\frac{1}{\mu^2} E^{(1,1)} = \frac{1}{\mu^2} \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} \mathbb{E}[x_{k;\lambda} x_{l;\lambda}], \quad (3.62)$$

The double sum includes mixed contributions from the  $k$ -th and  $l$ -th best elements of the  $i$ -th mutation component. To avoid confusion with the summation indices  $k$  and  $l$ , the integration variables associated with  $k$ -th element will be denoted as  $x_1$  (mutation) and  $q_1$  (quality), while the  $l$ -th element is integrated over  $x_2$  and  $q_2$ .

The ordering  $1 \leq k < l \leq \lambda$  is assumed with  $k$  yielding a smaller (better) quality value  $q_1 < q_2$ . Calculating (3.62) the joint probability density  $p_{k,l;\lambda}(x_1, x_2)$  is needed, such that the expected value can be formulated as

$$\frac{1}{\mu^2} E^{(1,1)} = \frac{1}{\mu^2} \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p_{k,l;\lambda}(x_1, x_2) dx_2 dx_1. \quad (3.63)$$

The mutation densities are independent and denoted by  $p_x(x_1)$  and  $p_x(x_2)$ , respectively. Given mutation components  $x_1$  and  $x_2$ , the conditional density obtaining the quality values  $q_1$  and  $q_2$  is  $p_Q(q_1|x_1)$  and  $p_Q(q_2|x_2)$ , respectively. Given  $q_1$  and  $q_2$ , one has  $k-1$  values are smaller than  $q_1$ ,  $l-k-1$  values between  $q_1$  and  $q_2$  and  $\lambda-l$  values larger than  $q_2$  with probabilities

$$\begin{aligned} \Pr\{Q \leq q_1\}^{k-1} &= P_Q(q_1)^{k-1} \\ \Pr\{q_1 \leq Q \leq q_2\}^{l-k-1} &= [P_Q(q_2) - P_Q(q_1)]^{l-k-1} \\ \Pr\{Q > q_2\}^{\lambda-l} &= [1 - P_Q(q_2)]^{\lambda-l} \end{aligned} \quad (3.64)$$

with  $P_Q(q)$  denoting the distribution function. The joint probability density can therefore be written as

$$\begin{aligned} p_{k,l;\lambda}(x_1, x_2) &= p_x(x_1) p_x(x_2) \int_{q_{\min}}^{\infty} p_Q(q_1|x_1) \int_{q_1}^{\infty} p_Q(q_2|x_2) \\ &\times \lambda! \frac{P_Q(q_1)^{k-1} [P_Q(q_2) - P_Q(q_1)]^{l-k-1} [1 - P_Q(q_2)]^{\lambda-l}}{(k-1)! (l-k-1)! (\lambda-l)!} dq_2 dq_1, \end{aligned} \quad (3.65)$$

with integration ranges  $q_{\min} \leq q_1 < \infty$  and  $q_1 < q_2 < \infty$  as  $k < l$ . Lower bound  $q_{\min}$  denotes the smallest possible quality value, which is resolved later. The factorials exclude the irrelevant combinations among the three groups given in (3.64). Plugging (3.65) into (3.63) and moving the sum into the integration one gets

$$\begin{aligned} \frac{1}{\mu^2} E^{(1,1)} &= \frac{\lambda!}{\mu^2} \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) \int_{q_{\min}}^{\infty} p_Q(q_1|x_1) \int_{q_1}^{\infty} p_Q(q_2|x_2) \\ &\times \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} \frac{P_Q(q_1)^{k-1} [P_Q(q_2) - P_Q(q_1)]^{l-k-1} [1 - P_Q(q_2)]^{\lambda-l}}{(k-1)! (l-k-1)! (\lambda-l)!} \\ &\times dq_2 dq_1 dx_2 dx_1. \end{aligned} \quad (3.66)$$

The double sum of (3.66) over the  $P_Q$ -values will be exchanged by an integration. This can be done using an identity from [2, p. 113]. Setting  $\nu = 2$  and identifying the indices as  $i_1 = l$  and  $i_2 = k$ , the evaluated identity yields

$$\begin{aligned} & \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} \frac{Q_1^{\lambda-l}[Q_2-Q_1]^{l-k-1}[1-Q_2]^{k-1}}{(\lambda-l)!(l-k-1)!(k-1)!} \\ &= \frac{1}{(\lambda-\mu-1)!(\mu-2)!} \int_0^{Q_1} t^{\lambda-\mu-1}(1-t)^{\mu-2} dt \end{aligned} \quad (3.67)$$

for real values  $Q_1$  and  $Q_2$  and integers  $\nu \leq \mu < \lambda$ . Now the substitution  $Q_1 = 1 - P_Q(q_2)$ ,  $Q_2 = 1 - P_Q(q_1)$  can be performed and the double sum of (3.66) can be recognized by comparing with (3.67). Applying the identity therefore yields

$$\begin{aligned} & \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} \frac{[1-P_Q(q_2)]^{\lambda-l}[P_Q(q_2)-P_Q(q_1)]^{l-k-1}[P_Q(q_1)]^{k-1}}{(\lambda-l)!(l-k-1)!(k-1)!} \\ &= \frac{1}{(\lambda-\mu-1)!(\mu-2)!} \int_0^{1-P_Q(q_2)} t^{\lambda-\mu-1}(1-t)^{\mu-2} dt, \end{aligned} \quad (3.68)$$

and Eq. (3.66) changes to

$$\begin{aligned} \frac{1}{\mu^2} E^{(1,1)} &= \frac{\lambda!}{\mu^2} \frac{1}{(\lambda-\mu-1)!(\mu-2)!} \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) \\ &\times \int_{q_{\min}}^{\infty} p_Q(q_1|x_1) \int_{q_1}^{\infty} p_Q(q_2|x_2) \\ &\times \int_0^{1-P_Q(q_2)} t^{\lambda-\mu-1}(1-t)^{\mu-2} dt dq_2 dq_1 dx_2 dx_1. \end{aligned} \quad (3.69)$$

The prefactor of Eq. (3.69) can be evaluated as

$$\begin{aligned} \frac{\lambda!}{\mu^2} \frac{1}{(\lambda-\mu-1)!(\mu-2)!} &= \frac{\lambda(\lambda-1)!(\mu-1)}{\mu^2(\lambda-\mu-1)!(\mu-1)!} \\ &= \frac{\lambda}{\mu} \frac{\mu-1}{\mu} \frac{(\lambda-1)!}{(\lambda-\mu-1)!(\mu-1)!} \\ &= \frac{1}{\vartheta} \frac{\mu-1}{\mu} \frac{1}{B(\lambda-\mu, \mu)} \end{aligned} \quad (3.70)$$

which will be useful during subsequent calculations.

Now the integration order will be exchanged twice in (3.69). First the order between  $t$  and  $q_2$  is exchanged. Then the order between  $t$  and  $q_1$  is exchanged, such that both  $q$ -integrations are performed before the  $t$ -integration enabling the application of the large population identity of Appendix B. Starting with integration bounds

$$\begin{aligned} q_1 &\leq q_2 < \infty \\ 0 &\leq t \leq 1 - P_Q(q_2) \end{aligned} \quad (3.71)$$

and using the inverse function  $P_Q^{-1}$  with  $q_2 = P_Q^{-1}(1-t)$  the exchanged bounds between  $t$  and  $q_2$  are given by

$$\begin{aligned} 0 &\leq t \leq 1 - P_Q(q_1) \\ q_1 &\leq q_2 \leq P_Q^{-1}(1-t). \end{aligned} \quad (3.72)$$

Using factor (3.70) and exchanged bounds (3.72) the expression (3.69) is reformulated as

$$\begin{aligned} \frac{1}{\mu^2} E^{(1,1)} &= \frac{1}{\vartheta} \frac{\mu-1}{\mu} \frac{1}{B(\lambda-\mu, \mu)} \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) \\ &\times \int_{q_{\min}}^{\infty} p_Q(q_1|x_1) \int_0^{1-P_Q(q_1)} t^{\lambda-\mu-1} (1-t)^{\mu-2} \\ &\times \int_{q_1}^{P_Q^{-1}(1-t)} p_Q(q_2|x_2) dq_2 dt dq_1 dx_2 dx_1. \end{aligned} \quad (3.73)$$

Now the integration order between  $t$  and  $q_1$  is exchanged starting from

$$\begin{aligned} q_{\min} &\leq q_1 < \infty \\ 0 &\leq t \leq 1 - P_Q(q_1), \end{aligned} \quad (3.74)$$

yielding

$$\begin{aligned} 0 &\leq t \leq 1 \\ q_{\min} &\leq q_1 \leq P_Q^{-1}(1-t). \end{aligned} \quad (3.75)$$

Therefore, we arrive at the following integration to be solved (beta function moved inside as it will be evaluated during the  $t$ -integration)

$$\begin{aligned} \frac{1}{\mu^2} E^{(1,1)} &= \frac{1}{\vartheta} \frac{\mu-1}{\mu} \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) \\ &\times \left( \frac{1}{B(\lambda-\mu, \mu)} \int_0^1 t^{\lambda-\mu-1} (1-t)^{\mu-2} \right. \\ &\times \left. \left[ \int_{q_{\min}}^{P_Q^{-1}(1-t)} p_Q(q_1|x_1) \left\{ \int_{q_1}^{P_Q^{-1}(1-t)} p_Q(q_2|x_2) dq_2 \right\} dq_1 \right] dt \right) dx_2 dx_1. \end{aligned} \quad (3.76)$$

Now the integrations in (3.76) will be successively solved. Starting with integration  $\{\cdot\}$  over  $q_2$  one has

$$\begin{aligned} \int_{q_1}^{P_Q^{-1}(1-t)} p_Q(q_2|x_2) dq_2 &= \left[ P_Q(q_2|x_2) \right]_{q_1}^{P_Q^{-1}(1-t)} \\ &= P_Q(P_Q^{-1}(1-t)|x_2) - P_Q(q_1|x_2). \end{aligned} \quad (3.77)$$

The  $q_1$ -integration within  $[\cdot]$  using (3.77) yields

$$\int_{q_{\min}}^{P_Q^{-1}(1-t)} p_Q(q_1|x_1) \left( P_Q(P_Q^{-1}(1-t)|x_2) - P_Q(q_1|x_2) \right) dq_1 \quad (3.78)$$

$$= P_Q(P_Q^{-1}(1-t)|x_2) \int_{q_{\min}}^{P_Q^{-1}(1-t)} p_Q(q_1|x_1) dq_1 \quad (3.79)$$

$$- \int_{q_{\min}}^{P_Q^{-1}(1-t)} p_Q(q_1|x_1) P_Q(q_1|x_2) dq_1 \quad (3.80)$$

First integration (3.79) is easily evaluated, as the conditional density is integrated over its support giving

$$\begin{aligned}
P_Q(P_Q^{-1}(1-t)|x_2) & \int_{q_{\min}}^{P_Q^{-1}(1-t)} p_Q(q_1|x_1) dq_1 \\
& = P_Q(P_Q^{-1}(1-t)|x_2) \left[ P_Q(q_1|x_1) \right]_{q_{\min}}^{P_Q^{-1}(1-t)} \quad (3.81) \\
& = P_Q(P_Q^{-1}(1-t)|x_2) P_Q(P_Q^{-1}(1-t)|x_1)
\end{aligned}$$

with  $P_Q(q_{\min}|x_1) = \Pr\{Q \leq q_{\min}|x_1\} = 0$ . Note that the resulting factors are equal up to the conditional variables  $x_1$  and  $x_2$ .

The second integral (3.80) will be simplified using integration by parts. Thereafter, one can exchange the  $x_1$  and  $x_2$  variables to find a significantly simpler expression for the original integral. Integration by parts yields

$$\begin{aligned}
& \int_{q_{\min}}^{P_Q^{-1}(1-t)} p_Q(q_1|x_1) P_Q(q_1|x_2) dq_1 \\
& = P_Q(P_Q^{-1}(1-t)|x_1) P_Q(P_Q^{-1}(1-t)|x_2) - \int_{q_{\min}}^{P_Q^{-1}(1-t)} P_Q(q_1|x_1) p_Q(q_1|x_2) dq_1. \quad (3.82)
\end{aligned}$$

Equation (3.82) inserted into (3.76) has to be integrated over  $x_1$  and  $x_2$ , of which the order can be exchanged. For the following argument the  $t$ -integration and the prefactors of (3.76) have no influence, such that they are dropped for better readability. Integrating both sides of (3.82) yields

$$\begin{aligned}
& \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) \int_{q_{\min}}^{P_Q^{-1}(1-t)} p_Q(q_1|x_1) P_Q(q_1|x_2) dq_1 dx_2 dx_1 \\
& = \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) P_Q(P_Q^{-1}(1-t)|x_1) P_Q(P_Q^{-1}(1-t)|x_2) dx_2 dx_1 \\
& \quad - \int_{-\infty}^{\infty} x_2 p_x(x_2) \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{q_{\min}}^{P_Q^{-1}(1-t)} P_Q(q_1|x_2) p_Q(q_1|x_1) dq_1 dx_1 dx_2, \quad (3.83)
\end{aligned}$$

where in the last line the integration order of  $x_1$  and  $x_2$  was exchanged, such that an expression equivalent to the *lhs* of (3.83) is obtained with given arguments for  $p_Q$  and  $P_Q$ . Collecting the terms, Eq. (3.83) can be formulated as

$$\begin{aligned}
& \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) \int_{q_{\min}}^{P_Q^{-1}(1-t)} p_Q(q_1|x_1) P_Q(q_1|x_2) dq_1 dx_2 dx_1 \\
& = \frac{1}{2} \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) P_Q(P_Q^{-1}(1-t)|x_1) P_Q(P_Q^{-1}(1-t)|x_2) dx_2 dx_1. \quad (3.84)
\end{aligned}$$

Noting that the *rhs* of result (3.84) is one half of the first integration result (3.81) after  $x$ -integration and noting the minus sign in (3.80), one gets for (3.78) the

expression

$$\begin{aligned}
& \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) \int_{q_{\min}}^{P_Q^{-1}(1-t)} p_Q(q_1|x_1) \\
& \quad \times \left( P_Q(P_Q^{-1}(1-t)|x_2) - P_Q(q_1|x_2) \right) dq_1 dx_2 dx_1 \\
& = \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) \\
& \quad \times \left( 1 - \frac{1}{2} \right) P_Q(P_Q^{-1}(1-t)|x_1) P_Q(P_Q^{-1}(1-t)|x_2) dx_2 dx_1.
\end{aligned} \tag{3.85}$$

Including prefactors and integration over  $t$  again, the result within  $[ \cdot ]$  of (3.76) simplifies significantly giving

$$\begin{aligned}
\frac{1}{\mu^2} E^{(1,1)} &= \frac{1}{2} \frac{1}{\vartheta} \frac{\mu-1}{\mu} \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) \\
& \quad \times \left( \frac{1}{B(\lambda-\mu, \mu)} \int_0^1 t^{\lambda-\mu-1} (1-t)^{\mu-2} \right. \\
& \quad \left. \times P_Q(P_Q^{-1}(1-t)|x_1) P_Q(P_Q^{-1}(1-t)|x_2) dt \right) dx_2 dx_1.
\end{aligned} \tag{3.86}$$

Given (3.86) and the integral in  $( \cdot )$ , the large population identity of (B.1) can be applied for  $(\mu, \lambda) \rightarrow \infty$  with constant  $\vartheta$ . Identifying  $a = 2$  and evaluating  $P_Q(P_Q^{-1}(1-t)|x_1) P_Q(P_Q^{-1}(1-t)|x_2)$  at the integrand's maximum location  $\hat{t} = 1 - \vartheta$  yields

$$\begin{aligned}
& \frac{1}{B(\lambda-\mu, \mu)} \int_0^1 t^{\lambda-\mu-1} (1-t)^{\mu-2} P_Q(P_Q^{-1}(1-t)|x_1) P_Q(P_Q^{-1}(1-t)|x_2) dt \\
& \simeq \frac{1}{\vartheta} P_Q(P_Q^{-1}(\vartheta)|x_1) P_Q(P_Q^{-1}(\vartheta)|x_2).
\end{aligned} \tag{3.87}$$

Using asymptotic equality (3.87) and noting that the terms containing  $x_1$  and  $x_2$  can be separated accordingly, Eq. (3.86) becomes

$$\begin{aligned}
\frac{1}{\mu^2} E^{(1,1)} &\simeq \frac{1}{2} \frac{1}{\vartheta^2} \frac{\mu-1}{\mu} \int_{-\infty}^{\infty} x_1 p_x(x_1) P_Q(P_Q^{-1}(\vartheta)|x_1) dx_1 \\
& \quad \times \int_{-\infty}^{\infty} x_2 p_x(x_2) P_Q(P_Q^{-1}(\vartheta)|x_2) dx_2 \\
& = \frac{1}{2} \frac{\mu-1}{\mu} \left[ \frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i p_x(x_i) P_Q(P_Q^{-1}(\vartheta)|x_i) dx_i \right]^2
\end{aligned} \tag{3.88}$$

where the integration variable is now denoted as  $x_i$  referring to the  $i$ -th component defined in Sec. 2. Additionally, the factor  $1/\vartheta$  was moved into  $[ \cdot ]$ .

Interestingly, the expression within  $[ \cdot ]$  can now be identified as the (negative) first order progress rate  $-\varphi_i$  within the large population limit derived in Eq. (2.28). The result of (3.88) can therefore be expressed as

$$\frac{1}{\mu^2} E^{(1,1)} \simeq \frac{1}{2} \frac{\mu-1}{\mu} \varphi_i^2. \quad (3.89)$$

The derivation of an analytic progress rate formula for  $\varphi_i$  starting from Eq. (2.28) was performed in Secs. 2.3 and 2.4 by expanding the distribution function up to first order and applying a large dimensionality approximation. The main result is given in (2.82) and inserted into (3.89), such that

$$\begin{aligned} \frac{1}{\mu^2} E^{(1,1)} &\simeq \frac{1}{2} \frac{\mu-1}{\mu} \varphi_i^2 \\ &\approx \frac{1}{2} \frac{\mu-1}{\mu} \left( c_\vartheta \frac{\sigma^2}{D_Q} \left( k_i + e^{-\frac{1}{2}(\alpha\sigma)^2} d_i \right) \right)^2 \\ &= \frac{1}{2} \frac{\mu-1}{\mu} e^{2,0} \frac{\sigma^4}{D_Q^2} \left( 2y_i + e^{-\frac{1}{2}(\alpha\sigma)^2} \alpha A \sin(\alpha y_i) \right)^2, \end{aligned} \quad (3.90)$$

using  $k_i = 2y_i$  and  $d_i = \alpha A \sin(\alpha y_i)$  to obtain the last line. Additionally, squaring the asymptotic progress coefficient yields  $c_\vartheta^2 = e_\vartheta^{2,0}$  using result (B.30) according to

$$c_\vartheta^2 = \left( e_\vartheta^{1,0} \right)^2 = \left[ \frac{e^{-\frac{1}{2}[\Phi^{-1}(\vartheta)]^2}}{\sqrt{2\pi}\vartheta} \right]^2 = e_\vartheta^{2,0}. \quad (3.91)$$

The final result for the expected value of  $E^{(1,1)}$  (for large populations and dimensionality) is

$$\frac{1}{\mu^2} E^{(1,1)} = \frac{1}{2} \frac{\sigma^2}{\mu} (\mu-1) e_\vartheta^{2,0} \frac{\sigma^2}{D_Q^2} \left( 2y_i + e^{-\frac{1}{2}(\alpha\sigma)^2} \alpha A \sin(\alpha y_i) \right)^2. \quad (3.92)$$

### 3.3.3 Collecting results of $E^{(2)}$ and $E^{(1,1)}$

By inserting the results (3.61) and (3.92) into the expression (3.7) and result (2.82) for  $\varphi_i$ , one gets for the quadratic progress rate

$$\begin{aligned} \varphi_i^{\text{II}} &= 2y_i \varphi_i - \frac{1}{\mu^2} E^{(2)} - \frac{2}{\mu^2} E^{(1,1)} \\ &= c_\vartheta \frac{\sigma^2}{D_Q} \left( 4y_i^2 + e^{-\frac{1}{2}(\alpha\sigma)^2} 2\alpha A y_i \sin(\alpha y_i) \right) \\ &\quad - \frac{\sigma^2}{\mu} \left\{ 1 + e_\vartheta^{1,1} \frac{(2y_i)^2 \sigma^2}{D_Q^2} - \frac{c_\vartheta}{D_Q} \left[ 3\sigma^2 \right. \right. \\ &\quad \left. \left. + A \cos(\alpha y_i) \left( 1 - e^{-\frac{1}{2}(\alpha\sigma)^2} + \alpha^2 \sigma^2 e^{-\frac{1}{2}(\alpha\sigma)^2} \right) \right] \right. \\ &\quad \left. + (\mu-1) e_\vartheta^{2,0} \frac{\sigma^2}{D_Q^2} \left( 2y_i + e^{-\frac{1}{2}(\alpha\sigma)^2} \alpha A \sin(\alpha y_i) \right)^2 \right\}. \end{aligned} \quad (3.93)$$

Comparing the terms within  $\{\cdot\}$  to the noisy order statistic result (3.39) interesting similarities can be observed. One can recognize that (3.93) contains corresponding terms in the large population limit with asymptotic progress coefficients

$$\begin{aligned} 1 + e_{\mu,\lambda}^{1,1} \frac{(2y_i)^2 \sigma^2}{D_Q^2} &\longleftrightarrow 1 + e_{\vartheta}^{1,1} \frac{(2y_i)^2 \sigma^2}{D_Q^2} \\ (\mu - 1)e_{\mu,\lambda}^{2,0} \frac{(2y_i)^2 \sigma^2}{D_Q^2} &\longleftrightarrow (\mu - 1)e_{\vartheta}^{2,0} \frac{(2y_i)^2 \sigma^2}{D_Q^2}. \end{aligned} \quad (3.94)$$

However, due to the different approach obtaining (3.93) by expanding the distribution function according to (2.41), additional correction terms are obtained which are not present for the noisy order statistic result (3.39).

### 3.4 Relation to Residual Distance (Squared)

Starting from definition (3.1) an important relation between the  $i$ -th component quadratic progress rate and the corresponding residual distance can be made. Summing (3.1) over all  $N$  components and noting that  $(R^{(g)})^2 = \sum_{i=1}^N (y_i^{(g)})^2$  one has

$$\begin{aligned} \sum_{i=1}^N \varphi_i^{\text{II}} &= \sum_{i=1}^N \mathbb{E} \left[ \left( y_i^{(g)} \right)^2 - \left( y_i^{(g+1)} \right)^2 \right] \\ &= \mathbb{E} \left[ \sum_{i=1}^N \left( y_i^{(g)} \right)^2 - \sum_{i=1}^N \left( y_i^{(g+1)} \right)^2 \right] \\ &= \mathbb{E} \left[ \left( R^{(g)} \right)^2 - \left( R^{(g+1)} \right)^2 \right] \\ &=: \varphi_R^{\text{II}} \end{aligned} \quad (3.95)$$

such that the progress rate w.r.t. the residual distance squared can be defined as

$$\varphi_R^{\text{II}} = \mathbb{E} \left[ \left( R^{(g)} \right)^2 - \left( R^{(g+1)} \right)^2 \mid R^{(g)}, \sigma^{(g)} \right]. \quad (3.96)$$

The obtained relation (3.95) is very important, as it relates the component-wise  $y_i$ -dependent progress to the progress of the residual distance. This is very useful for the investigation of the approximation quality of expressions derived for  $\varphi_i^{\text{II}}$ , as summing over all components can be used as a cumulative measure. It is also useful to study the convergence behavior later, for which an  $R$ -dependent formula is needed.

The radial progress rate (3.96) can also be interpreted as the (negative) expected quality gain of the spherical function  $f_{\text{sph}}(R) = R^2$  according to

$$\mathbb{E} \left[ f_{\text{sph}}(R^{(g+1)}) - f_{\text{sph}}(R^{(g)}) \right] = \mathbb{E} \left[ \left( R^{(g+1)} \right)^2 - \left( R^{(g)} \right)^2 \right] = -\varphi_R^{\text{II}}. \quad (3.97)$$

Following a relation provided by Beyer [4, p. 173, Eq. (34)], the progress of the sphere  $\varphi_{\text{sph}} = \mathbb{E} [R^{(g)} - R^{(g+1)}]$  after normalization can be related to the

quality gain of  $f_{\text{sph}}(R)$  and therefore to  $\varphi_R^{\text{II}}$  via

$$\varphi_R^{\text{II}} = \frac{2R^2}{N} \varphi_{\text{sph}}^*. \quad (3.98)$$

This yields the normalization for  $\varphi_R^{\text{II}}$  denoted by  $\varphi_R^{\text{II},*}$  according to

$$\varphi_R^{\text{II},*} = \frac{N}{2R^2} \varphi_R^{\text{II}}. \quad (3.99)$$

The quantity  $\varphi_R^{\text{II},*}$  will be evaluated for different approximations of  $\varphi_i^{\text{II}}$  in Sec. 3.5 by summing over all components and normalizing.

### 3.5 Comparing Experiments with Approximations

Having obtained the quadratic progress rate via two methods, experiments are performed and compared to the following progress rate approximations

- SIM: average over repeated experiments
- B1:  $\varphi_i^{\text{II}}$  using noisy order statistic derivation (3.39)
- B2: simplified Eq. (3.39) considering only the loss term  $-\frac{\sigma^2}{\mu}$ . Theoretical derivation of the formula assuming large  $N$  will be presented in (3.116).
- L1:  $\varphi_i^{\text{II}}$  via large population approximation (3.93)

Approximations are listed as obtained in chronological order. All  $\varphi_i^{\text{II}}$  solutions use approximation A3 for  $\varphi_i$ , see also Sec. 2.5.2 and Eq. (2.82). The labels “B” and “L” were chosen to distinguish the different approaches. For L1, neglecting all loss-terms in (3.93) except  $-\sigma^2/\mu$ , the approximations L1 and B2 are equal.

**Initialization of Position** For the following experiments the initial locations were chosen randomly to be on the sphere surface for given radius  $R$ . This is done by choosing  $y_i \sim \mathcal{N}(0, 1)$  independently for all components  $i$ , normalizing  $\mathbf{y}$  to unit length and rescaling by  $R$ . For each one-generation experiment the initial random location is fixed for all trials, as the progress rate is evaluated (averaged) component-wise for a given location  $y_i$ .

**Frequency Value** For all following experiments the oscillation frequency was set to  $\alpha = 2\pi$  and remains as such, if not explicitly stated otherwise.

**Sphere Function for  $A = 0$**  For the first experiment in Fig. 15, which is a plausibility check, the oscillation strength is set to zero, such that the sphere function is recovered. Quantity (3.95) is evaluated by summing over all components and normalizing using (3.99). Additionally, the sphere progress rate formula by Beyer [5, p. 216, Eq. (6.54)] is plotted to check the relation given in (3.98).

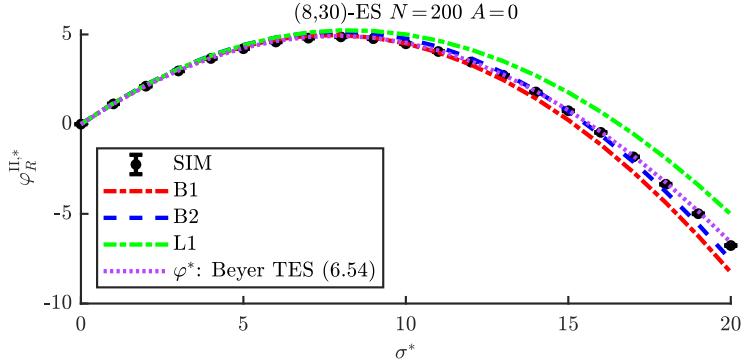


Figure 15: The sphere function is recovered by setting  $A = 0$  ( $N = 200$ ), such that  $f(\mathbf{y}) = \sum_i y_i^2 = R^2$ . The normalized progress  $\varphi_R^{II,*}$  was measured and averaged over  $10^5$  trials for  $(8/8,30)$ -ES. Setting  $A = 0$  the approximations B1, B2, and L1 yield a component-wise progress rate formula for the sphere. The results were obtained by summing over all  $N$  components, see (3.95), and then normalizing using (3.99). Setting  $A > 0$  deviations from the sphere will be observed depending on  $R$ . Beyer's formula [5, Eq. (6.54)] is plotted as an additional reference.

**Rastrigin Function for  $A > 0$**  For the following experiments, the progress rate is evaluated over  $10^6$  trials at four values  $R = \sqrt{N} \cdot \{10, 1, 0.1, 0.01\}$  including rescaling by  $\sqrt{N}$  to be within the relevant range of local minima as  $N$  is changed. For each experiment, three progress rate plots are shown, namely  $\varphi_i$  in Fig. 16,  $\varphi_i^{II}$  in Fig. 17 (both for the first component  $i = 1$ ) and  $\varphi_R^{II,*}$  as a normalized cumulative measure in Fig. 18. First order  $\varphi_i$  is shown as some of the  $\varphi_i^{II}$ -deviations can be attributed to  $\varphi_i$ . The range for the normalized mutation  $\sigma^*$  was set accordingly to display the transition between positive and negative progress.

In Fig. 16 the approximation A3 of Eq. (2.82) yields good results for the first order progress. Slightly larger deviations occur for the upper experiment with smaller  $N$  and a smaller truncation ratio (with smaller  $\mu$ ). The lower experiment shows very good results as  $N$  and  $\mu$  are larger. The approximation quality is good over different magnitudes of  $R$ . Possible deviations within  $\varphi_i$  apply to all approximations of  $\varphi_i^{II}$ , as we have  $\varphi_i^{II}(\varphi_i)$  from Eq. (3.7).

In Fig. 17 the three approximations for  $\varphi_i^{II}$  are shown compared to simulations. For all approximations larger deviations are observed in the upper experiment, as  $N$  and  $\mu$  are relatively small. The deviations are consistent over different  $R$ -values with L1 overestimating progress for large  $\sigma^*$ , while B1 and B2 are underestimating. The overall approximation quality of the significantly simpler formula of B2 is surprisingly good. The effect of the different terms within the approximation is studied in more detail in Sec. 3.6.

In Fig. 18, the  $R$ -dependent progress  $\varphi_R^{II,*}$  was calculated by summing over all components according to (3.95). Furthermore, normalization (3.99) is applied. The radial progress and its approximation quality is comparable to results of Fig. 17. Somewhat harder to see, the approximation quality of L1 is slightly better for small  $\sigma^*$ . Note that the normalization yields very similar values of

$\varphi_R^{\Pi,*}$  independent of  $R$ . However, the shape of the curve shows slight changes with varying  $R$  due to  $A > 0$  compared to the sphere.

Additional experiments for  $(10/10, 40)$ -ES with  $N = 100$  are shown in the Appendix in Figs. 63 and 64.

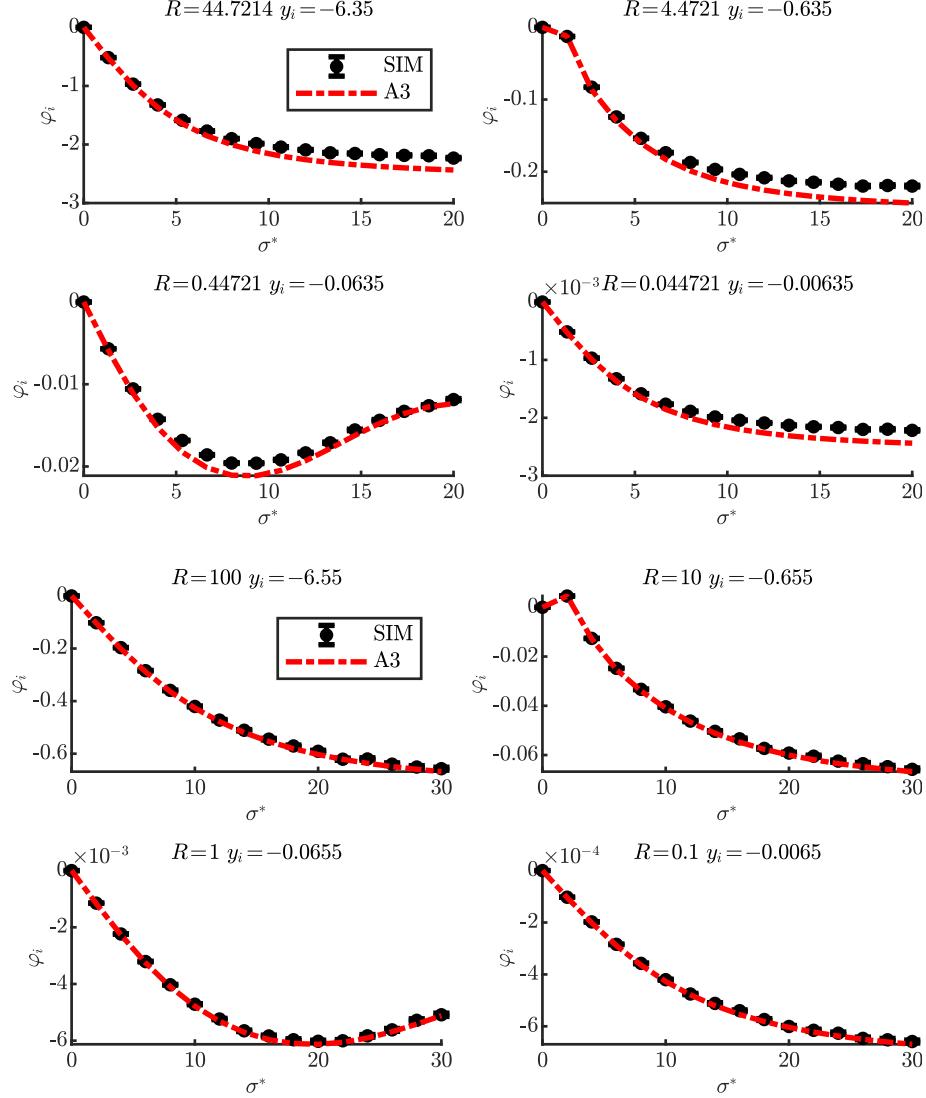


Figure 16: First order progress rate  $\dot{\varphi}_i$  for component  $i = 1$ . Upper four plots show  $(10/10, 40)$ -ES with  $N = 20$  and  $A = 1$  and lower plots show  $(25/25, 50)$ -ES with  $N = 100$  and  $A = 1$ . The progress has a negative value since  $y_i < 0$ .

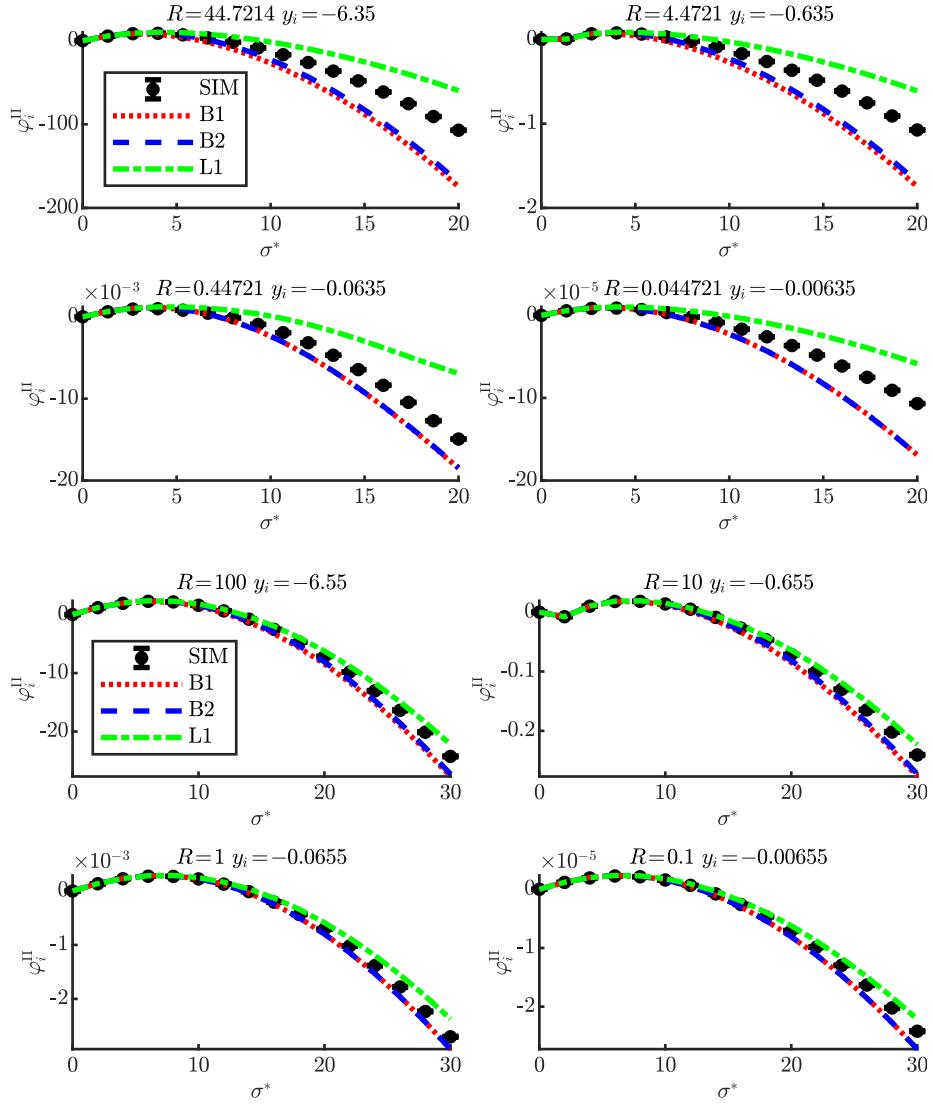


Figure 17: Second order progress rate  $\varphi_i^{II}$  for component  $i = 1$ . Upper four plots show  $(10/10, 40)$ -ES with  $N = 20$  and  $A = 1$  and lower plots show  $(25/25, 50)$ -ES with  $N = 100$  and  $A = 1$ .

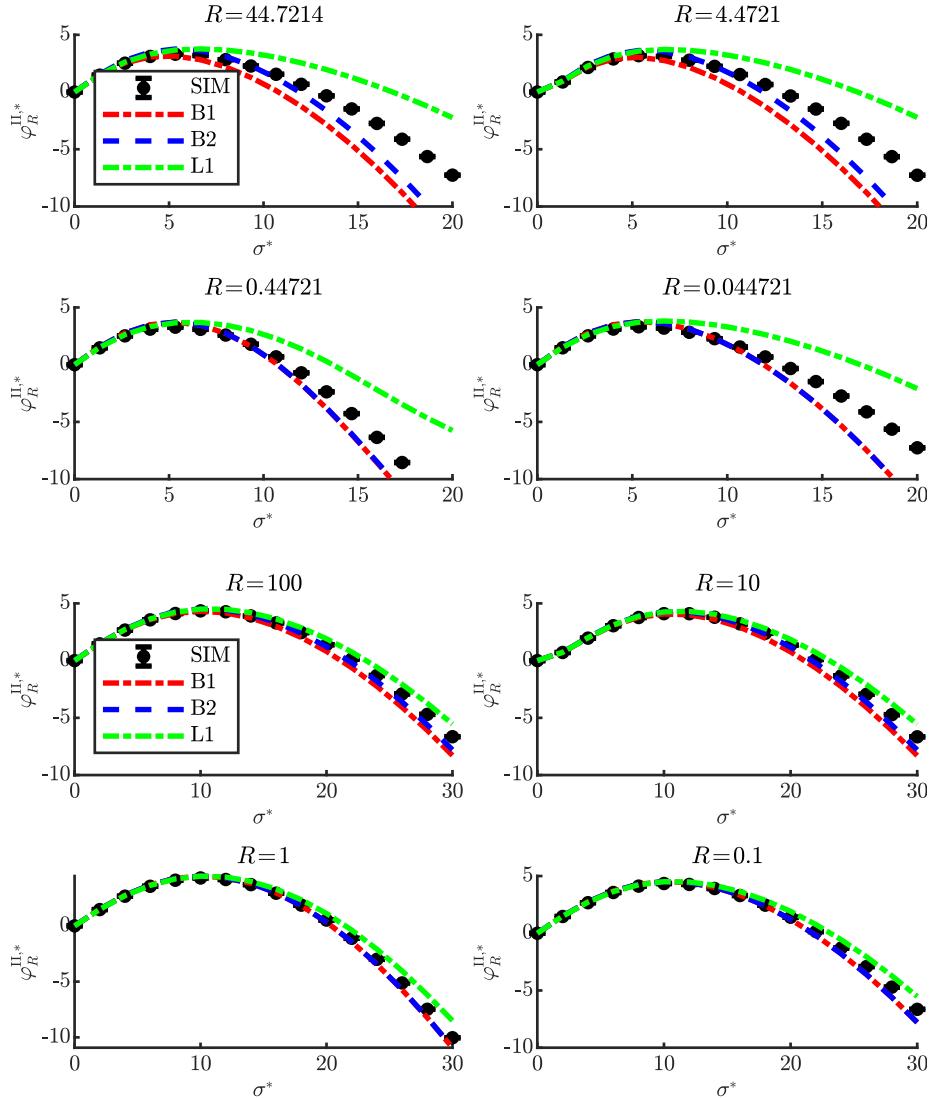


Figure 18: Normalized  $R$ -dependent progress rate as a cumulative measure (summed over all components). Upper four plots show  $(10/10, 40)$ -ES with  $N = 20$  and  $A = 1$  and lower plots show  $(25/25, 50)$ -ES with  $N = 100$  and  $A = 1$ . The approximation quality of B2 is surprisingly good considering its simpler expression. The deviations are comparable to the deviations seen in Fig. 17 for a single component.

### 3.6 Investigation of Loss Terms

In order to understand the deviations observed in Figs. 17 and 18 the loss terms within  $\{\cdot\}$  of Eq. (3.93) are investigated in more detail. The terms are abbreviated according to their respective factors as  $e_\vartheta^{1,1}$ ,  $c_\vartheta/D_Q$  and  $e_\vartheta^{2,0}$ . First, experimental tests are done followed by a theoretical analysis.

Figure 19 shows the terms relative to the value “1” for (10/10, 40)-ES and  $A = 1$  with varying dimensionality  $N$ . A single exemplary component is displayed for a given  $R$ . All the terms are suppressed for increasing  $N$ . The term  $e_\vartheta^{1,1}$  (blue) is negligible for any parameter set independent of  $\sigma^*$ . Term  $c_\vartheta/D_Q$  (red) is relevant for small  $N$  and shows an approximately constant contribution for  $\sigma^*$  large enough. This is in accordance with the top experiment of Fig. 18 showing that L1 systematically overestimates the simulation. This is largely due to the  $-c_\vartheta/D_Q$  term reducing the loss. It is obtained from Eq. (3.61), which in turn depends on the expansion in (3.44). This suggests that higher order terms of the expansion are needed, if higher accuracy is desired. Term  $e_\vartheta^{2,0}$  contributes mainly for small  $\sigma^*$  and is suppressed for large  $\sigma^*$ . This behavior is also in accordance with L1 of Fig. 18, as the approximation yields slightly better results for the progress at small  $\sigma^*$ .

In Figure 20 the dimensionality is varied together with  $\mu$  for constant  $\vartheta = 0.5$ . Therefore the relevant  $\sigma^*$ -range is also adapted, such that unnecessary large values are omitted for experiments with small  $\mu$ . Note, that the only  $\mu$ -dependency of Eq. (3.93) is the prefactor  $(\mu-1)$  of  $e_\vartheta^{2,0}$ , as all progress coefficients are given within the asymptotic limit of infinite populations. For Fig. 20, the coefficient  $e_\vartheta^{1,1} = 0$  due to  $\Phi^{-1}(0.5) = 0$ , such that its corresponding term yields zero contribution. Term  $e_\vartheta^{2,0}$  scales with  $\mu$ , which was expected, but it is damped for increasing  $\sigma^*$ . Similar to Fig. 19, term  $-c_\vartheta/D_Q$  is approximately constant for large  $\sigma^*$  and suppressed for large  $N$  independent of  $\mu$ .

As a conclusion, Eq. (3.93) will be investigated under the asymptotic limit  $N \rightarrow \infty$  in order to further simplify the expression for the progress rate  $\varphi_i^{\text{II}}$ . Attention has to be paid w.r.t. the scaling  $\mu(N)$ , for which no theoretical investigations exist at this point. In practical applications  $\mu$  needs to be increased for increasing  $N$  to achieve a higher success probability. Figure 21 will show preliminary experiments of the scaling behavior of the population size  $\mu(N)$  needed to achieve high success probability.

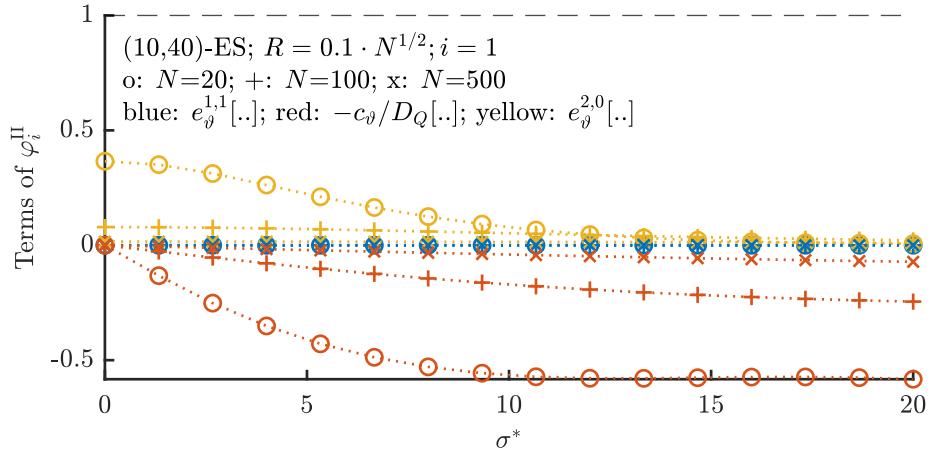


Figure 19: Investigating loss terms within  $\{\cdot\}$  of Eq. (3.93) for increasing  $N$ . For constant population size and increasing  $N$ , all terms are suppressed compared to the value “1”.

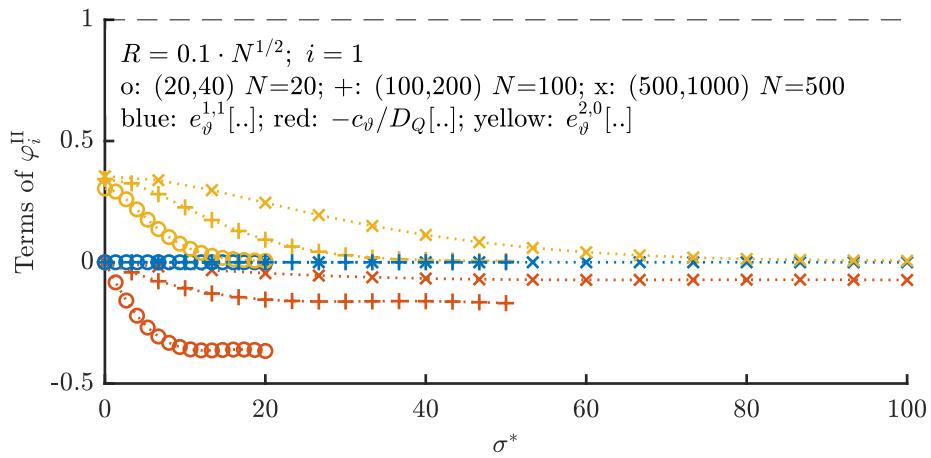


Figure 20: Investigating loss terms within  $\{\cdot\}$  of Eq. (3.93) for increasing  $N$  and  $\mu$ . The maximum value of  $\sigma^*$  was adapted to capture the relevant range. Term  $e_\vartheta^{2,0}$  scales linearly with  $\mu$ .

Having experimentally investigated the  $N$ -scaling of the loss terms in Figs. 19 and 20, a theoretical investigation of the behavior is tackled now.

Starting from (3.93), a first approach was trying to find a lower bound for the variance  $D_Q^2$ , such that terms containing  $1/D_Q^2$  and  $1/D_Q$  are maximized. The idea was to find upper bounds for the terms relative to the value “1”. Due to the trigonometric terms of second line in (1.29) no useful (sufficiently tight) bound of  $D_Q^2$  could be established at this point without yielding negative variance. Therefore a Taylor expansion of the relevant terms will be performed for large dimensionality  $N$ .

As the  $\varphi_i^{\text{II}}$  approximation shall be valid for constant normalized mutations  $\sigma^*$  given some residual distance  $R$ , the transformed mutation is given by

$$\sigma = \frac{\sigma^* R}{N}, \quad (3.100)$$

and will be expanded within the exponential function for large  $N$ . Within the limit  $N \rightarrow \infty$  attention must be paid considering the relation  $R(N)$ , as the (interesting)  $R$ -range with high density of local minima grows as well, see also Fig. 3. Assuming w.l.o.g. that the current location  $\mathbf{y}$  is at a local minimum (of arbitrary order  $j$ ) denoted by  $\hat{y}_j$ , one has  $\mathbf{y} = [\hat{y}_j]$  for  $N = 1$  and therefore  $R^2 = \hat{y}_j^2$ . Having  $N$  components at the same order local minimum yields  $\mathbf{y} = [\hat{y}_j, \hat{y}_j, \dots, \hat{y}_j]$ , such that  $R^2 = N\hat{y}_j^2$ . This motivates the scaling relation  $R \sim \sqrt{N}$  for large  $N$ . Therefore the expansion orders will be displayed as functions of  $R/N$  for the following derivations.

First the variance (1.29) is analyzed. The exponentials will be expanded using

$$e^{-c(\alpha\sigma^*\frac{R}{N})^2} = 1 - c \left( \alpha\sigma^* \frac{R}{N} \right)^2 + O\left(\frac{R^4}{N^4}\right), \quad (3.101)$$

with  $c \in \{1/2, 1\}$ . Using (3.100) the variance (1.29) yields

$$\begin{aligned} D_Q^2 &= \sum_{i=1}^N 2 \left( \sigma^* \frac{R}{N} \right)^4 + 4y_i^2 \left( \sigma^* \frac{R}{N} \right)^2 \\ &\quad + \frac{A^2}{2} \left( 1 - e^{-(\alpha\sigma^*\frac{R}{N})^2} \right) \left( 1 - \cos(2\alpha y_i) e^{-(\alpha\sigma^*\frac{R}{N})^2} \right) \\ &\quad + 2\alpha A \left( \sigma^* \frac{R}{N} \right)^2 e^{-\frac{1}{2}(\alpha\sigma^*\frac{R}{N})^2} \left( \alpha \left( \sigma^* \frac{R}{N} \right)^2 \cos(\alpha y_i) + 2y_i \sin(\alpha y_i) \right). \end{aligned} \quad (3.102)$$

Applying expansion (3.101) and collecting higher order terms one gets

$$\begin{aligned}
D_Q^2 &= \sum_{i=1}^N 4y_i^2 \left( \sigma^* \frac{R}{N} \right)^2 + O \left( \frac{R^4}{N^4} \right) \\
&\quad + \frac{A^2}{2} \left( 1 - \left[ 1 - \left( \alpha \sigma^* \frac{R}{N} \right)^2 + O \left( \frac{R^4}{N^4} \right) \right] \right) \\
&\quad \times \left( 1 - \cos(2\alpha y_i) \left[ 1 - \left( \alpha \sigma^* \frac{R}{N} \right)^2 + O \left( \frac{R^4}{N^4} \right) \right] \right) \\
&\quad + 2\alpha A \left( \sigma^* \frac{R}{N} \right)^2 \left[ 1 - \frac{1}{2} \left( \alpha \sigma^* \frac{R}{N} \right)^2 + O \left( \frac{R^4}{N^4} \right) \right] \\
&\quad \times \left( \alpha \left( \sigma^* \frac{R}{N} \right)^2 \cos(\alpha y_i) + 2y_i \sin(\alpha y_i) \right). \tag{3.103}
\end{aligned}$$

The summand with prefactor  $A^2/2$  yields

$$\begin{aligned}
&\frac{A^2}{2} \left( 1 - \left[ 1 - \left( \alpha \sigma^* \frac{R}{N} \right)^2 + O \left( \frac{R^4}{N^4} \right) \right] \right) \\
&\quad \times \left( 1 - \cos(2\alpha y_i) \left[ 1 - \left( \alpha \sigma^* \frac{R}{N} \right)^2 + O \left( \frac{R^4}{N^4} \right) \right] \right) \\
&= \frac{A^2}{2} \left( \left( \alpha \sigma^* \frac{R}{N} \right)^2 + O \left( \frac{R^4}{N^4} \right) \right) \\
&\quad \times \left( 1 - \cos(2\alpha y_i) + \cos(2\alpha y_i) \left( \alpha \sigma^* \frac{R}{N} \right)^2 + O \left( \frac{R^4}{N^4} \right) \right) \tag{3.104} \\
&= \frac{A^2}{2} \left( \alpha \sigma^* \frac{R}{N} \right)^2 (1 - \cos(2\alpha y_i)) + O \left( \frac{R^4}{N^4} \right) \\
&= \left( A \alpha \sigma^* \frac{R}{N} \right)^2 \sin^2(\alpha y_i) + O \left( \frac{R^4}{N^4} \right).
\end{aligned}$$

using  $1 - \cos 2x = 2 \sin^2 x$  for the last line. The last summand of (3.103) yields

$$\begin{aligned}
&2\alpha A \left( \sigma^* \frac{R}{N} \right)^2 \left[ 1 - \frac{1}{2} \left( \alpha \sigma^* \frac{R}{N} \right)^2 + O \left( \frac{R^4}{N^4} \right) \right] \left( \alpha \left( \sigma^* \frac{R}{N} \right)^2 \cos(\alpha y_i) + 2y_i \sin(\alpha y_i) \right) \\
&= 2\alpha A \left( \sigma^* \frac{R}{N} \right)^2 2y_i \sin(\alpha y_i) + O \left( \frac{R^4}{N^4} \right). \tag{3.105}
\end{aligned}$$

Collecting results (3.104) and (3.105) the variance simplifies

$$\begin{aligned}
D_Q^2 &= \sum_{i=1}^N 4y_i^2 \left( \sigma^* \frac{R}{N} \right)^2 + 2\alpha A \left( \sigma^* \frac{R}{N} \right)^2 2y_i \sin(\alpha y_i) \\
&\quad + \left( \sigma^* \frac{R}{N} \right)^2 (\alpha A \sin(\alpha y_i))^2 + O\left(\frac{R^4}{N^4}\right) \\
&= \left( \sigma^* \frac{R}{N} \right)^2 \sum_{i=1}^N (2y_i + \alpha A \sin(\alpha y_i))^2 + \sum_{i=1}^N O\left(\frac{R^4}{N^4}\right) \\
&= \left( \sigma^* \frac{R}{N} \right)^2 \sum_{i=1}^N (f'_i)^2 + O\left(\frac{R^4}{N^3}\right).
\end{aligned} \tag{3.106}$$

using definition (1.11) for the derivative  $f'_i$ . Note that result (3.106) is contained in the last line of Eq. (1.33), as it is an expansion for small  $\sigma$  and therefore equivalent to an expansion for large  $N$  with  $\sigma = \sigma^* R/N$  (constant  $\sigma^*$  and  $R$ ).

Given (3.106), the scaling of  $\sum_{i=1}^N (f'_i)^2$  w.r.t.  $N$  and  $R$  can be deduced applying the triangle inequality to the corresponding vectors. Considering the positional vector  $\mathbf{y}$  and definition  $\sin(\alpha \mathbf{y}) := \sin(\alpha y_1)\mathbf{e}_1 + \sin(\alpha y_2)\mathbf{e}_2 + \dots + \sin(\alpha y_N)\mathbf{e}_N$  with  $\mathbf{e}_i$  being the  $i$ -th unit vector, one has

$$\sum_{i=1}^N (f'_i)^2 = \|2\mathbf{y} + \alpha A \sin(\alpha \mathbf{y})\|^2. \tag{3.107}$$

Using inequality  $\|\mathbf{a} + \mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\|$  and therefore  $\|\mathbf{a} + \mathbf{b}\|^2 \leq (\|\mathbf{a}\| + \|\mathbf{b}\|)^2$ , and using  $\|\mathbf{y}\|^2 = R^2$ , an upper bound for expression (3.107) can be given as

$$\begin{aligned}
\|2\mathbf{y} + \alpha A \sin(\alpha \mathbf{y})\|^2 &\leq 4\|\mathbf{y}\|^2 + 4\alpha A\|\mathbf{y}\|\|\sin(\alpha \mathbf{y})\| + (\alpha A)^2\|\sin(\alpha \mathbf{y})\|^2 \\
&= 4R^2 + 4\alpha A R \sqrt{\sum_{i=1}^N \sin^2(\alpha y_i)} + (\alpha A)^2 \sum_{i=1}^N \sin^2(\alpha y_i) \\
&\leq 4R^2 + 4\alpha A R \sqrt{N} + (\alpha A)^2 N = (2R + \alpha A \sqrt{N})^2.
\end{aligned} \tag{3.108}$$

From (3.108) one can deduce the (upper bound) scaling  $\sum_{i=1}^N (f'_i)^2 \sim N$ , which is valid for both constant  $R$  and  $R \sim \sqrt{N}$ . Therefore we conclude the scaling relation of the variance for large  $N$  as

$$\begin{aligned}
D_Q^2 &= \left( \sigma^* \frac{R}{N} \right)^2 \sum_{i=1}^N (f'_i)^2 + O\left(\frac{R^4}{N^3}\right) \\
&\sim \frac{(\sigma^* R)^2}{N}.
\end{aligned} \tag{3.109}$$

Having obtained  $D_Q^2$ , now the terms within  $\{\cdot\}$  of (3.93) are investigated. The first term  $e_\vartheta^{1,1}$  is easily evaluated. Inserting scaling relation (3.109) for  $D_Q^2$  one gets

$$e_\vartheta^{1,1} \frac{\sigma^2}{D_Q^2} (2y_i)^2 = e_\vartheta^{1,1} \frac{\left(\sigma^* \frac{R}{N}\right)^2}{\frac{(\sigma^* R)^2}{N}} (2y_i)^2 = e_\vartheta^{1,1} \frac{(2y_i)^2}{N} = O\left(\frac{1}{N}\right). \tag{3.110}$$

The second term  $c_\vartheta/D_Q$  is evaluated using normalization (3.100) and expansion (3.101) as

$$\begin{aligned}
& \frac{c_\vartheta}{D_Q} \left\{ 3 \left( \sigma^* \frac{R}{N} \right)^2 + A \cos(\alpha y_i) \left( 1 - \left[ 1 - \frac{1}{2} \left( \alpha \sigma^* \frac{R}{N} \right)^2 + O\left(\frac{R^4}{N^4}\right) \right] \right. \right. \\
& \quad \left. \left. + \left( \alpha \sigma^* \frac{R}{N} \right)^2 \left[ 1 - \frac{1}{2} \left( \alpha \sigma^* \frac{R}{N} \right)^2 + O\left(\frac{R^4}{N^4}\right) \right] \right) \right\} \\
& = \frac{c_\vartheta}{D_Q} \left\{ 3 \left( \sigma^* \frac{R}{N} \right)^2 + A \cos(\alpha y_i) \left( \frac{1}{2} \left( \alpha \sigma^* \frac{R}{N} \right)^2 + \left( \alpha \sigma^* \frac{R}{N} \right)^2 + O\left(\frac{R^4}{N^4}\right) \right) \right\} \\
& = c_\vartheta \frac{\left( \sigma^* \frac{R}{N} \right)^2}{D_Q} \left\{ 3 + \frac{3}{2} \alpha^2 A \cos(\alpha y_i) + O\left(\frac{R^2}{N^2}\right) \right\}. 
\end{aligned} \tag{3.111}$$

Inserting  $D_Q = \sigma^* R / \sqrt{N}$  from (3.109) into (3.111) yields

$$\begin{aligned}
& c_\vartheta \frac{\left( \sigma^* \frac{R}{N} \right)^2}{\frac{\sigma^* R}{\sqrt{N}}} \left\{ 3 + \frac{3}{2} \alpha^2 A \cos(\alpha y_i) + O\left(\frac{R^2}{N^2}\right) \right\} \\
& = O\left(\frac{R}{N^{3/2}}\right) = \begin{cases} O\left(\frac{1}{N^{3/2}}\right) & \text{if } R = \text{const.} \\ O\left(\frac{1}{N}\right) & \text{if } R \sim \sqrt{N}. \end{cases}
\end{aligned} \tag{3.112}$$

The last term  $e_\vartheta^{2,0}$  yields after expansion

$$\begin{aligned}
& (\mu - 1) e_\vartheta^{2,0} \frac{\left( \sigma^* \frac{R}{N} \right)^2}{D_Q^2} \left( 2y_i + \alpha A \sin(\alpha y_i) \left[ 1 - \frac{1}{2} \left( \alpha \sigma^* \frac{R}{N} \right)^2 + O\left(\frac{R^4}{N^4}\right) \right] \right)^2 \\
& = (\mu - 1) e_\vartheta^{2,0} \frac{\left( \sigma^* \frac{R}{N} \right)^2}{D_Q^2} \left( 2y_i + \alpha A \sin(\alpha y_i) + O\left(\frac{R^2}{N^2}\right) \right)^2.
\end{aligned} \tag{3.113}$$

Using scaling (3.109) for  $D_Q^2$  and writing  $\mu(N)$  to denote the (unknown) population dependency on  $N$  one gets

$$\begin{aligned}
& (\mu - 1) e_\vartheta^{2,0} \frac{\left( \sigma^* \frac{R}{N} \right)^2}{\frac{(\sigma^* R)^2}{N}} \left( 2y_i + \alpha A \sin(\alpha y_i) + O\left(\frac{R^2}{N^2}\right) \right)^2 \\
& = \mu(N) O\left(\frac{1}{N}\right) = \begin{cases} O\left(\frac{1}{N}\right) & \text{if } \mu(N) = \text{const.} \\ O\left(\frac{\mu(N)}{N}\right) & \text{else.} \end{cases}
\end{aligned} \tag{3.114}$$

The problem of population sizing, namely choosing  $\mu(N)$  to achieve high global convergence probability will be investigated in the future in more detail. Preliminary simulations are shown in Fig. 21 and yield a sub-linear relation.

**Collecting Results** Inserting the scaling results for the three terms (3.110), (3.112) and (3.114) back into  $\{\cdot\}$  of the quadratic progress rate (3.93), one gets

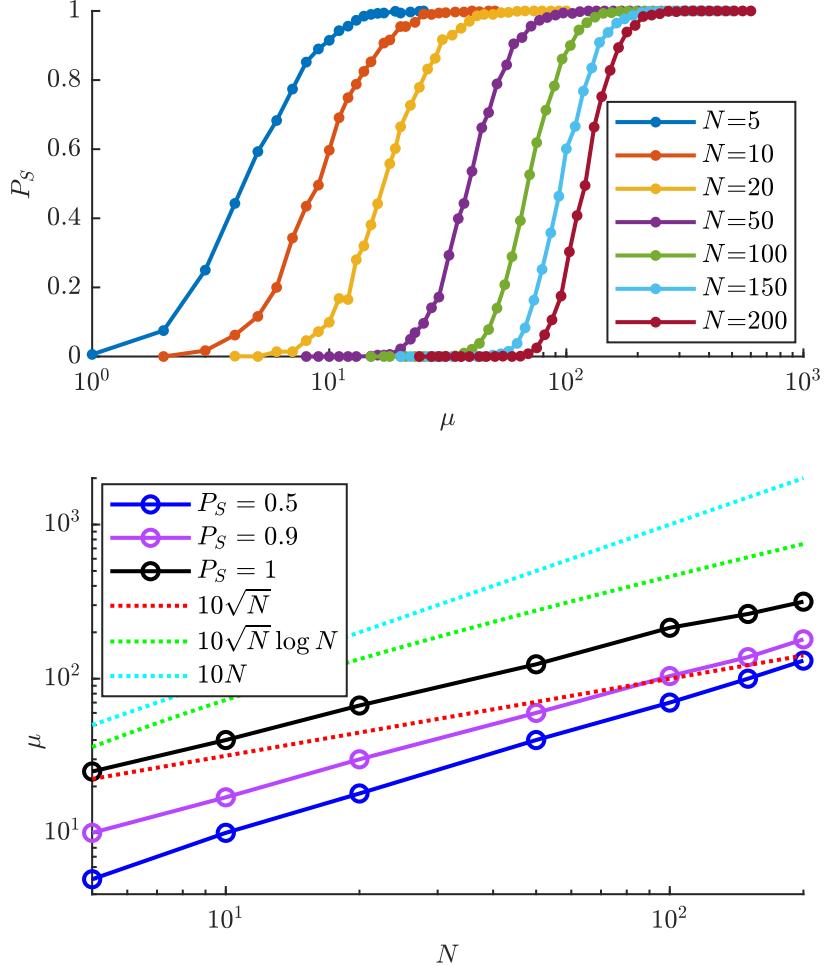


Figure 21: Measured success probability  $P_S$  as a function of  $\mu$  for varying  $N$  (top) and scaling relation  $\mu(N)$  to achieve high success probability (bottom). The experiments were performed using  $(\mu/\mu_I, \lambda)$ - $\sigma$ SA-ES with learning parameter  $\tau = 1/\sqrt{2N}$  for  $\alpha = 2\pi$ ,  $A = 1$ , and  $\vartheta = 0.5$ . Each data point was averaged over 1000 trials. A sub-linear increase of  $\mu(N)$  can be experimentally observed.

for large dimensionality  $N \rightarrow \infty$  and residual distance scaling  $R \sim \sqrt{N}$  the relation

$$\begin{aligned} \varphi_i^{\text{II}} &= c_\vartheta \frac{\sigma^2}{D_Q} \left( 4y_i^2 + e^{-\frac{1}{2}(\alpha\sigma)^2} 2\alpha A y_i \sin(\alpha y_i) \right) \\ &\quad - \frac{\sigma^2}{\mu} \left\{ 1 + O\left(\frac{1}{N}\right) + O\left(\frac{\mu(N)}{N}\right) \right\}. \end{aligned} \quad (3.115)$$

Provided that the population size  $\mu = o(N)$ , i.e. increasing sub-linearly with  $N$ , all terms except ‘‘1’’ can be neglected for large dimensionality and the result

yields

Quadratic progress rate for large  $N$  and sub-linear  $\mu(N)$

$$\varphi_i^{\text{II}} = c_\vartheta \frac{\sigma^2}{D_Q} \left( 4y_i^2 + e^{-\frac{1}{2}(\alpha\sigma)^2} 2\alpha A y_i \sin(\alpha y_i) \right) - \frac{\sigma^2}{\mu}. \quad (3.116)$$

Result (3.116) is also denoted as approximation B2 in Sec. 3.5. It was obtained using the large population approximation for first order progress  $\varphi_i$  in Sec. 2.2 and second order terms calculated for  $\varphi_i^{\text{II}}$  in Sec. 3.3. Furthermore, non-linear terms within  $Q_i(x_i)$  were treated as a perturbation of the linear term giving rise to an expansion of the distribution function, see Eq. (2.41). This expansion introduces additional terms for both  $\varphi_i$  and  $\varphi_i^{\text{II}}$ . To further simplify the lengthy results, the large dimensionality approximation was applied multiple times throughout the derivation. Comparing the different variance terms for large  $N$ , see Sec. 2.4,  $\varphi_i$  could be simplified considerably. Additionally, the loss part of  $\varphi_i^{\text{II}}$  could also be simplified significantly by looking at the scaling for large  $N$ . This is important, as (3.116) will be the starting point for further theoretical and experimental investigations.

## 4 R-Dependent Progress Rate

In order to investigate the convergence behavior of progress rate result (3.116) using the dynamical systems approach in Sec. 5, an  $R$ -dependent formulation of the progress rate is needed. Being a function of the residual distance (instead of single components) also has the advantage, that a (spherical) normalization can be applied. Additionally, relations to the spherical progress rate can be easier identified.

The tools for transforming (3.116) into an  $R$ -dependent formula have already been established and only need to be applied. In Sec. 1.5.2 the averaging method was presented and the  $R$ -dependent variance  $D_Q^2(R, \sigma)$  was given in (1.75). In (3.95) the relation between the component-wise progress rate and  $\varphi_R^{\text{II}}$  was already introduced. Summing result (3.116) over all  $N$  components yields

$$\begin{aligned}\varphi_R^{\text{II}} &= \sum_{i=1}^N \varphi_i^{\text{II}} \\ &= c_\vartheta \frac{\sigma^2}{D_Q(R, \sigma)} \left( 4 \sum_{i=1}^N y_i^2 + e^{-\frac{1}{2}(\alpha\sigma)^2} 2\alpha A \sum_{i=1}^N y_i \sin(\alpha y_i) \right) - \sum_{i=1}^N \frac{\sigma^2}{\mu} \quad (4.1) \\ &= c_\vartheta \frac{\sigma^2}{D_Q(R, \sigma)} \left( 4R^2 + e^{-\frac{1}{2}(\alpha\sigma)^2} 2\alpha A \sum_{i=1}^N y_i \sin(\alpha y_i) \right) - N \frac{\sigma^2}{\mu}.\end{aligned}$$

The expression for  $D_Q(R, \sigma)$  in (1.75) is not inserted at this point for better visibility. Aiming at an  $R$ -dependent expression of (4.1), a new random variable is introduced, analogous to the method in (1.64), for the sum over the trigonometric terms

$$Y := \sum_{i=1}^N y_i \sin(\alpha y_i). \quad (4.2)$$

Assuming i.i.d. locations  $y_i \sim \frac{R}{\sqrt{N}} \mathcal{N}(0, 1)$ , see more detailed discussion in Sec. 1.5.2, and using the CLT for  $N \rightarrow \infty$ , see (1.66), the random variable is rewritten as

$$Y \sim \mathcal{N}(\mathbb{E}[Y], \text{Var}[Y]). \quad (4.3)$$

The ratio  $\sqrt{\text{Var}[Y]} / \mathbb{E}[Y] \rightarrow 0$  was already investigated for  $N \rightarrow \infty$  in Eq. (1.72) to find an  $R$ -dependent variance expression. The same limit behavior is applied here to neglect the fluctuations of the random variable.

By writing  $Y \sim \mathbb{E}[Y] + \epsilon_Y$ , expression (4.1) is split into an expected value and a fluctuation term denoted by  $\epsilon_Y$ , which is then neglected in the second step. Setting  $\mathbb{E}[Y] = \alpha R^2 e^{-\frac{1}{2} \frac{(\alpha R)^2}{N}}$ , see (A.34), one has

$$\begin{aligned}\varphi_R^{\text{II}} &\stackrel{N \rightarrow \infty}{=} c_\vartheta \frac{\sigma^2}{D_Q(R, \sigma)} \left( 4R^2 + e^{-\frac{1}{2}(\alpha\sigma)^2} 2\alpha A [\mathbb{E}[Y] + \epsilon_Y] \right) - N \frac{\sigma^2}{\mu} \\ &\approx c_\vartheta \frac{\sigma^2}{D_Q(R, \sigma)} \left( 4R^2 + 2\alpha^2 A R^2 e^{-\frac{1}{2}(\alpha\sigma)^2} e^{-\frac{1}{2} \frac{(\alpha R)^2}{N}} \right) - N \frac{\sigma^2}{\mu},\end{aligned} \quad (4.4)$$

such that after simplifying we obtain the result

### Progress rate as a function of $R$

$$\varphi_R^{\text{II}} = c_\vartheta \frac{2R^2\sigma^2}{D_Q(R, \sigma)} \left( 2 + \alpha^2 A e^{-\frac{\alpha^2}{2}(\sigma^2 + \frac{R^2}{N})} \right) - N \frac{\sigma^2}{\mu}. \quad (4.5)$$

In Figure 22 simulations are shown with result (4.5) compared to the approximation (3.116) it is based on.

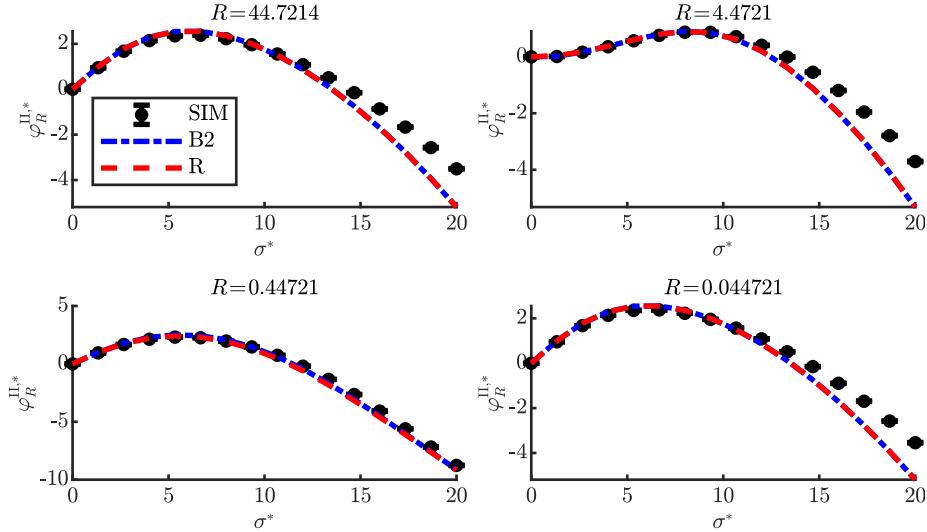


Figure 22: Comparing approximations B2, i.e. (3.116) summed over  $i = 1, \dots, N$ , and Eq. (4.5), denoted by “R”, after normalization for (20/20,40)-ES,  $\alpha = 2\pi$ ,  $A = 10$ , and  $N = 20$ . Four values with  $R = \sqrt{N}\{10, 1, 0.1, 0.01\}$  are chosen. The initial positions were chosen randomly on surface  $R$  and are constant for all trials. This initialization is needed to compare the component-wise progress B2 (constant coordinates) with R. The two formulas yield very similar results, even for moderate  $N$ . One reason is that only the trigonometric terms are approximated by the expected value, the term  $R^2 = \sum_i y_i^2$  is exact and the loss term remains unchanged. Furthermore, for sufficiently large or small  $R$ , the Rastrigin function becomes sphere-like and both approximations yield the same result (top-left and bottom-right progress are practically identical).

**Sphere progress rate** A component-wise progress rate for the sphere function can be derived starting from (3.116) and using variance from (1.29) by applying  $A = 0$  (or  $\alpha = 0$ ). This simplifies the formulae considerably giving

$$\varphi_{i,\text{sph}}^{\text{II}} = c_\vartheta \frac{4\sigma^2 y_i^2}{\sqrt{4R^2\sigma^2 + 2N\sigma^4}} - \frac{\sigma^2}{\mu}. \quad (4.6)$$

Now one can sum (4.6) over all components, set  $\sigma = \sigma^* R/N$  and afterwards apply the normalization  $\varphi_R^{\text{II},*} = \frac{N}{2R^2} \varphi_R^{\text{II}}$  from (3.99), such that

$$\begin{aligned}\varphi_R^{\text{II}} &= \sum_{i=1}^N \varphi_{i,\text{sph}}^{\text{II}} = c_\vartheta \frac{4\sigma^2}{\sqrt{4R^2\sigma^2 + 2N\sigma^4}} \sum_{i=1}^N y_i^2 - \frac{N\sigma^2}{\mu} \\ &= c_\vartheta \frac{4(\sigma^* R/N)^2}{\sqrt{4R^2(\sigma^* R/N)^2 + 2N(\sigma^* R/N)^4}} R^2 - \frac{N(\sigma^* R/N)^2}{\mu} \\ \varphi_R^{\text{II},*} &= c_\vartheta \frac{4(\sigma^*/N)^2 R^4}{2R^2(\sigma^*/N)\sqrt{1 + \sigma^*/2N}} \frac{N}{2R^2} - \frac{N(\sigma^* R/N)^2}{\mu} \frac{N}{2R^2}.\end{aligned}\quad (4.7)$$

After simplifying (4.7) the result for the sphere progress rate is obtained

$$\varphi_R^{\text{II},*} = c_\vartheta \frac{\sigma^*}{\sqrt{1 + \sigma^{*2}/2N}} - \frac{\sigma^{*2}}{2\mu}, \quad (4.8)$$

which for the limit  $N \rightarrow \infty$  yields the well-known progress rate formula [5, p. 217, (6.56)]

$$\varphi_R^{\text{II},*} = \varphi_{\text{sph}}^* = c_\vartheta \sigma^* - \frac{\sigma^{*2}}{2\mu}, \quad (4.9)$$

by setting  $c_{\mu/\mu,\lambda} \simeq c_\vartheta$ . Now the Rastrigin function is investigated with respect to its spherical properties within the limits of small and large  $R$ , respectively.

**Sphere limit for large  $R$**  Given (4.5) the result of the sphere progress rate (4.8) can be deduced for large  $R$ . The variance approximation (1.30) for large  $R$  and  $\sigma = \sigma^* R/N$  yields (neglecting  $NA^2/2$  term)

$$\begin{aligned}D_{\text{sph}}^2 &= 4R^2\sigma^2 + 2N\sigma^4 = 4R^2 \left( \frac{\sigma^* R}{N} \right)^2 + 2N \left( \frac{\sigma^* R}{N} \right)^4 \\ &= 4R^4 \left( \frac{\sigma^*}{N} \right)^2 \left( 1 + \frac{\sigma^{*2}}{2N} \right).\end{aligned}\quad (4.10)$$

Additionally, the exponential within (4.5) can be dropped, which has the same effect as setting  $A = 0$ . This immediately yields the same form as already shown in Eqs. (4.7) and (4.8), such that the sphere progress rate is recovered.

**Sphere limit for small  $R$**  The sphere progress rate can also be deduced for small residual distances. Using variance (1.36) yields (neglecting higher orders)

$$\begin{aligned}D_Q^2 &= (2 + \alpha^2 A)^2 \left( \sigma^2 R^2 + \frac{N\sigma^4}{2} \right) \\ &= (2 + \alpha^2 A)^2 \left( (\sigma^* R/N)^2 R^2 + \frac{N(\sigma^* R/N)^4}{2} \right) \\ &= (2 + \alpha^2 A)^2 R^4 \left( \frac{\sigma^*}{N} \right)^2 \left( 1 + \frac{\sigma^{*2}}{2N} \right).\end{aligned}\quad (4.11)$$

Given the exponential from (4.5) and setting  $\sigma = \sigma^* R/N$ , the exponential can be expanded for small  $R$  as

$$e^{-\frac{\alpha^2 R^2}{2} \left( \left(\frac{\sigma^*}{N}\right)^2 + \frac{1}{N} \right)} = 1 + O(R^2). \quad (4.12)$$

Setting the exponential to one and inserting (4.11), the progress rate yields

$$\begin{aligned} \varphi_R^{\text{II}} &= c_\vartheta \frac{2R^2(\sigma^* R/N)^2 (2 + \alpha^2 A)}{(2 + \alpha^2 A)R^2 (\sigma^*/N) \sqrt{1 + \sigma^{*2}/2N}} - N \frac{(\sigma^* R/N)^2}{\mu} \\ \varphi_R^{\text{II},*} &= c_\vartheta \frac{2R^2(\sigma^* R/N)^2}{R^2 (\sigma^*/N) \sqrt{1 + \sigma^{*2}/2N}} \frac{N}{2R^2} - N \frac{(\sigma^* R/N)^2}{\mu} \frac{N}{2R^2} \\ &= c_\vartheta \frac{\sigma^*}{\sqrt{1 + \sigma^{*2}/2N}} - \frac{\sigma^{*2}}{2\mu} \end{aligned} \quad (4.13)$$

Both limits  $R \rightarrow \infty$  and  $R \rightarrow 0$  of the Rastrigin progress rate (4.5) therefore yield for  $A > 0$  the expected spherical progress rate from Eq. (4.8). For  $A = 0$  or  $\alpha = 0$  the exponential terms vanish and the result holds for all  $R$ . For  $A > 0$  and  $\alpha > 0$  the exponentials within progress rate and variance can be interpreted as the transition terms between the spherical limits, see also Fig. 5.

## 4.1 Progress Rate Contour Maps

As the  $R$ -dependent progress rate formula is a function of only two parameters  $\sigma^*$  and  $R$  (given fitness parameters  $N, \alpha, A$ ), it can be visualized using a contour plot, see Figs. 23, 24, 25, and 26 for different parameter sets.

High positive progress is shown in red, while low negative progress is shown in blue. The line of zero progress is shown in bold white. For the sphere function only vertical lines can be observed, as the normalized progress rate is independent of  $R$ . For the Rastrigin function deviations from the sphere can be observed depending on  $A$  and  $\alpha$ .

A characteristic horizontal progress dip can be observed with significantly lower progress compared to the sphere. It has a certain extension in  $R$ - and  $\sigma^*$ -space, and vanishes for both large and small  $R$ -values.

From the contour plots one can deduce the behavior of the deterministic iteration (5.6) shown in Sec. 5. Setting a constant  $\sigma^*$ -value the iteration moves vertically through the landscape. A slow-down is observed when regions of lower progress are entered and vice-versa. If the white zero-progress line is reached, a stationary  $R$ -value occurs.

Important to note is that the maps do not show negative progress rates for small mutations, which one might expect considering real experiments, where small mutations increase the probability of local convergence. This issue is related to the  $R$ -averaging and it is discussed in Sec. 5 in Fig. 58.

Additionally, averaged real optimization runs of the  $(\mu/\mu_I, \lambda)$ - $\sigma$ SA-ES are overlaid as vertical lines with  $\tau = 1/\sqrt{2N}$  (black) and  $\tau = 1/\sqrt{8N}$  (magenta). The median of both  $R(g)$  and  $\sigma^*(g)$  was applied as the measure of central tendency, see upcoming discussion in Sec. 5.2.1. The reduction of  $\tau$  leads to a higher  $\sigma^*$ -level (closer to zero-progress line) and therefore higher success probabilities for all configurations. However, within the spherical limits the mutations are too large.

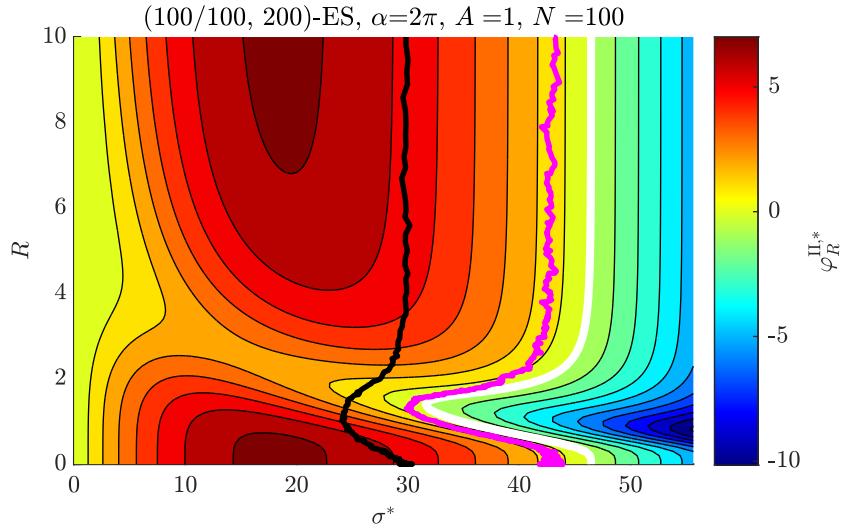


Figure 23: Progress contours with parameters given in title. For small and large  $R$  the sphere progress rate (4.8) is recovered. A characteristic “valley” of decreased progress can be observed. The lines of equal progress are separated by one unit value. The median over 200  $\sigma$ SA-runs is overlaid with  $\tau = 1/\sqrt{2N}$  (black,  $P_S = 0.91$ ) and  $\tau = 1/\sqrt{8N}$  (magenta,  $P_S = 0.99$ ).

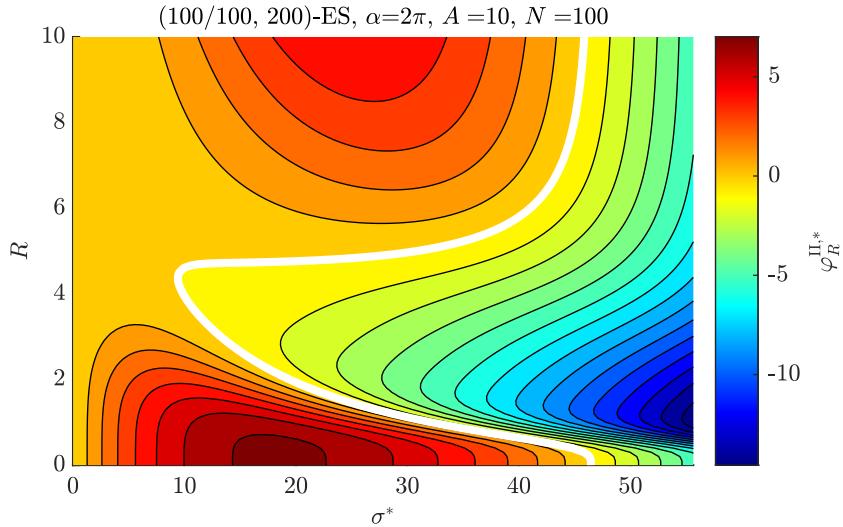


Figure 24: Progress contours with parameters given in title. High value  $A = 10$  is chosen compared to Fig. 23. The region of decreased progress is significantly larger, as expected. Real  $\sigma$ SA-runs are not converging for given parameter sets.

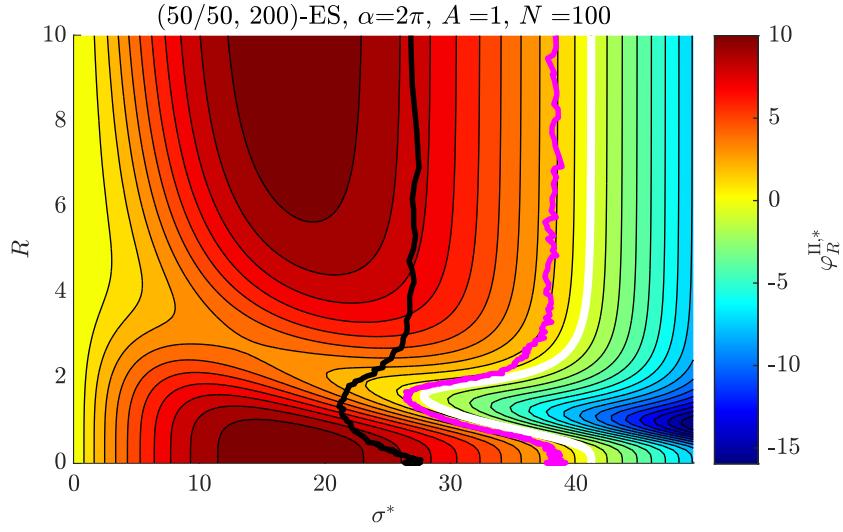


Figure 25: Lower truncation value  $\vartheta = 0.25$  is chosen compared to Fig. 23. The zero-progress line shifts slightly to the left (compared to Fig. 23) giving smaller favorable mutations. The overall characteristic is very similar. The median over 200  $\sigma$ SA-runs is taken with  $\tau = 1/\sqrt{2N}$  (black,  $P_S = 0.65$ ) and  $\tau = 1/\sqrt{8N}$  (magenta,  $P_S = 0.94$ ).

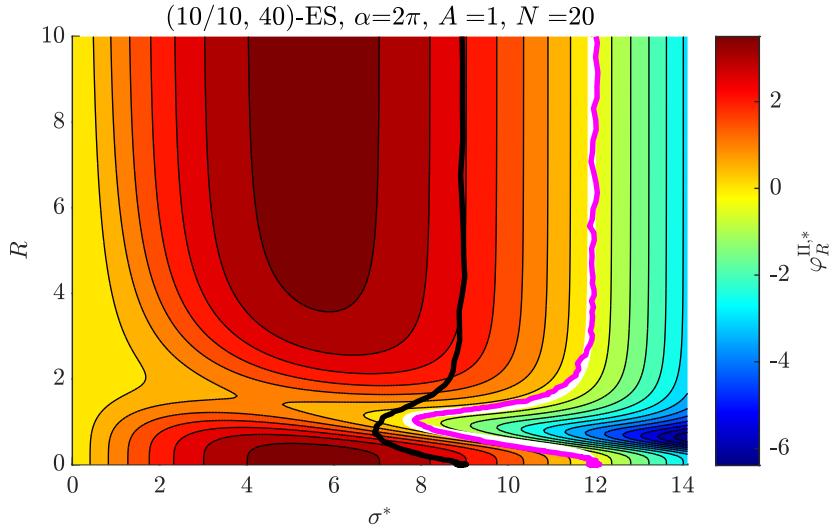


Figure 26: Reduced dimensionality and population compared to previous plots. Contour step-size was set to 0.5 for better resolution. The overall characteristic does not change considerably. Now the median over 2000  $\sigma$ SA-runs is shown to reduce fluctuations effects (black:  $P_S = 0.42$ ) and (magenta:  $P_S = 0.69$ ).

## 4.2 Investigating Zero of Progress Rate

In order to study the convergence properties of result (4.5), we are interested in a function describing the zero-progress line, see Figures of previous Sec. 4.1. Here, the *second* zero of the progress rate is referred to as “the zero”, as the first zero corresponds to the trivial solution  $\sigma = 0$ .

To find the root of the equation, the relevant terms of (4.5) will be split into a part containing “spherical” terms and a Rastrigin-specific part. Then, the equation will be restructured and its solvability in terms of the involved expressions is discussed. Afterwards, a possible analytic solution will be provided by neglecting the exponential terms.

Starting with variance  $D_Q^2(\sigma^*, R)$  from (1.81), the expression is restructured and the Rastrigin-specific part is denoted by function  $h$  according to

$$\begin{aligned} D_Q^2(\sigma^*, R) &= 4R^4 \left( \frac{\sigma^*}{N} \right)^2 \left[ 1 + \frac{\sigma^{*2}}{2N} + h(\sigma^*, R) \right], \quad \text{with} \\ h(\sigma^*, R) &:= \frac{N^2}{4R^4\sigma^{*2}} \left\{ \frac{NA^2}{2} \left[ 1 - e^{-\left(\frac{\alpha R \sigma^*}{N}\right)^2} \right] \left[ 1 - e^{-\left(\alpha R\right)^2 \left[\left(\frac{\sigma^*}{N}\right)^2 + \frac{2}{N}\right]} \right] \right. \\ &\quad \left. + 2NA\alpha^2 R^4 \left( \frac{\sigma^*}{N} \right)^2 \left[ \left( \frac{\sigma^*}{N} \right)^2 + \frac{2}{N} \right] e^{-\frac{(\alpha R)^2}{2} \left[\left(\frac{\sigma^*}{N}\right)^2 + \frac{1}{N}\right]} \right\}, \end{aligned} \quad (4.14)$$

which will be useful later when recovering the solution for the sphere. Now the exponential term within (4.5) is expressed using a new function  $g$  and setting  $\sigma = \sigma^* R/N$  according to

$$g(\sigma^*, R) := \alpha^2 A e^{-\frac{(\alpha R)^2}{2} \left(\frac{\sigma^{*2}}{N^2} + \frac{2}{N}\right)}. \quad (4.15)$$

Using (4.14) and (4.15) the normalized progress rate can be written as

$$\begin{aligned} \varphi_R^{\text{II},*} &= \varphi_R^{\text{II}} \frac{N}{2R^2} \\ &= c_\vartheta \frac{2R^2(\sigma^* R/N)^2 [2 + g]}{2R^2 \frac{\sigma^*}{N} \sqrt{1 + \frac{\sigma^{*2}}{2N} + h}} \frac{N}{2R^2} - \frac{N(\sigma^* R/N)^2}{\mu} \frac{N}{2R^2} \\ &= \frac{c_\vartheta \sigma^*}{2} \frac{2 + g}{\sqrt{1 + \frac{\sigma^{*2}}{2N} + h}} - \frac{\sigma^{*2}}{2\mu}. \end{aligned} \quad (4.16)$$

Expression (4.16) was defined in such a way that for  $g = 0$  and  $h = 0$  (e.g. by setting  $A = 0$ ) sphere equation (4.8) is recovered. Now we are looking for solutions of  $\varphi_R^{\text{II},*} = 0$  for  $\sigma^* > 0$  and  $R > 0$ , which yields

$$\begin{aligned} \frac{2 + g(\sigma^*, R)}{\sqrt{1 + \frac{\sigma^{*2}}{2N} + h(\sigma^*, R)}} &= \frac{\sigma^*}{c_\vartheta \mu} \\ (2 + g(\sigma^*, R))^2 &= \frac{\sigma^{*2}}{c_\vartheta^2 \mu^2} \left[ 1 + \frac{\sigma^{*2}}{2N} + h(\sigma^*, R) \right], \end{aligned} \quad (4.17)$$

such that after reordering the following expression is obtained

$$\frac{\sigma^{*4}}{2Nc_\vartheta^2\mu^2} + \frac{\sigma^{*2}}{c_\vartheta^2\mu^2} - 4 = g^2(\sigma^*, R) + 2g(\sigma^*, R) - \frac{\sigma^{*2}}{c_\vartheta^2\mu^2}h(\sigma^*, R). \quad (4.18)$$

The functional dependencies of  $g$  and  $h$  are explicitly written in (4.18) to illustrate the problem of solvability. We want to solve for  $\sigma^*(R)$ , which would give a relation describing the boundary  $\varphi_R^{\text{II},*} = 0$ .

One can immediately see that with  $g(\sigma^*, R)$  in (4.14) and  $h(\sigma^*, R)$  in (4.15) containing arguments  $\sigma^*$  and  $R$  within exponential functions, no closed form solution of Eq. (4.18) can be given at this point.

Expanding the functions  $g(\sigma^*, R)$  and  $h(\sigma^*, R)$  in a Taylor series around some point  $\tilde{R}$  or  $\tilde{\sigma}^*$  would be technically possible, but the results would have only very limited applicability for the investigation of global convergence even for a solution using higher order polynomials. Expanding the exponentials around  $\tilde{R} = 0$  or  $\tilde{\sigma}^* = 0$  assuming small changes could also be done, but its applications are again limited as small  $R$  or  $\sigma^*$  can only model the global or local attraction basins, respectively, but not the larger scale dynamics.

Therefore, the only analytically tractable solution at this point will be the limit of vanishing exponential factors, discussed below.

#### 4.2.1 Zero-Progress for Sphere

Setting  $g = h = 0$  in (4.17), the right side vanishes and the sphere function is recovered. A solution can be easily given by solving the fourth-order polynomial

$$\sigma^{*4} + 2N\sigma^{*2} - 8Nc_\vartheta^2\mu^2 = 0. \quad (4.19)$$

An equation of the form (4.19) with different coefficients will reappear later. Hence, in the general case one can write for a real coefficient  $c$

$$\sigma^{*4} + 2N\sigma^{*2} + c = 0. \quad (4.20)$$

The only positive non-complex solution of the fourth order Eq. (4.20) can be identified as

$$\sigma^* = \left[ [N^2 - c]^{1/2} - N \right]^{1/2}. \quad (4.21)$$

Applying (4.21) to (4.19) the sphere formula for zero progress (subscript “ $\varphi_0$ ”) is given by

$$\sigma_{\varphi_0, \text{sph}}^* = \left[ [N^2 + 8Nc_\vartheta^2\mu^2]^{1/2} - N \right]^{1/2}. \quad (4.22)$$

#### 4.2.2 Zero-Progress for Vanishing Exponentials

Neglecting the exponential factors is needed for an analytic solution of (4.18), but it can also be justified under following conditions.

The exponential terms are not just suppressed by very large values of  $R$ , but also for moderate  $R$  values. This is due to the fact that the argument is  $R^2$ , and it is being multiplied by frequency  $\alpha^2$  and larger mutations  $\sigma^{*2}$  (favorable to decrease the probability of local convergence). As an example in

Fig. 5, the variance result (1.31) without exponentials models large parts of the transitional region and not only the limit  $R \rightarrow \infty$ . The idea is to describe the dynamics for large and moderately sized  $R$  at the start of the transitional region (characterized by a slow-down) up to some extent, which is discussed later. The descent into the global attractor basin cannot be modeled by this approach, as can be seen from the large deviations of the cyan line in Fig. 5.

Important to note is the limit  $N \rightarrow \infty$  with constant  $R$ , for which the exponentials approach the value 1. Considering the scaling  $R_{tr} \sim \sqrt{N}$  of the transition point, see Eq. (1.79) and Fig. 3, the limit  $N \rightarrow \infty$  for any constant  $R$  automatically leads to progress rate equations of the global attractor basin, since the transition point is also diverging. Interestingly, the limits  $R \rightarrow 0$  (constant  $N$  and  $\sigma^*$ ) and  $N \rightarrow \infty$  (constant  $R$  and  $\sigma^*$ ) yield the same terms of the Taylor expansion of the exponentials. Therefore one has to be careful applying this limit, although the progress rate approximations do require larger  $N$  values for better agreement.

The limit of vanishing exponentials corresponds to  $g = 0$  and  $h = \frac{N^3 A^2}{8R^4 \sigma^{*2}}$ , such that (4.18) yields after rearranging the polynomial

$$\sigma^{*4} + 2N\sigma^{*2} + \frac{N^4 A^2}{4R^4} - 8Nc_\vartheta^2 \mu^2 = 0. \quad (4.23)$$

Solution (4.21) can therefore be applied again with coefficient  $c = N^4 A^2 / (4R^4) - 8Nc_\vartheta^2 \mu^2$ , which yields

$$\sigma_{\varphi_0}^* = \left[ \left[ N^2 - \frac{N^4 A^2}{4R^4} + 8Nc_\vartheta^2 \mu^2 \right]^{1/2} - N \right]^{1/2}. \quad (4.24)$$

One can see that for  $R \rightarrow \infty$  we have  $\sigma_{\varphi_0}^* \simeq \sigma_{\varphi_0, sph}^*$  from Eq. (4.22). For  $R < \infty$  a condition for (4.24) is required to guarantee a non-complex result for  $\sigma^*$ . Requiring the inner square-root of (4.24) to be real, one has

$$\begin{aligned} N^2 + 8Nc_\vartheta^2 \mu^2 &\geq \frac{N^4 A^2}{4R^4} \\ R^4 &\geq \frac{N^4 A^2}{4N^2 + 32Nc_\vartheta^2 \mu^2}. \end{aligned} \quad (4.25)$$

Analogously, requiring the outer square-root of (4.24) to be real, one has

$$\begin{aligned} \left[ N^2 - \frac{N^4 A^2}{4R^4} + 8Nc_\vartheta^2 \mu^2 \right]^{1/2} &\geq N \\ N^2 - \frac{N^4 A^2}{4R^4} + 8Nc_\vartheta^2 \mu^2 &\geq N^2 \\ R^4 &\geq \frac{N^4 A^2}{32Nc_\vartheta^2 \mu^2}. \end{aligned} \quad (4.26)$$

As bound (4.25) is smaller than bound (4.26) the latter is chosen. The minimum distance to have a real solution for  $\sigma_{\varphi_0}^*$  is therefore

$$R_{\varphi_0, \min} \geq \left( \frac{N^3 A^2}{32c_\vartheta^2 \mu^2} \right)^{1/4}, \quad (4.27)$$

concluding that  $\sigma_{\varphi_0}^*(R)$  is defined for  $R \in [R_{\varphi_0,\min}, \infty)$ .

As iterations and simulations are performed for a constant given  $\sigma^*$ , it is more convenient to reformulate (4.23) as  $R(\sigma^*)$  in order to have a “distance of zero progress”  $R_{\varphi_0}$ . The result can be easily given as

$$R_{\varphi_0} = \left[ \frac{1}{4} \frac{N^4 A^2}{N^2 + 8Nc_\vartheta^2 \mu^2 - (\sigma^{*2} + N)^2} \right]^{1/4} \quad (4.28)$$

with  $R_{\varphi_0}(\sigma^*)$  defined for  $\sigma^* \in [0, \sigma_{\varphi_0,\text{sph}}^*)$ .

It remains to show how the sign of  $\varphi_R^{\text{II},*}$  behaves w.r.t. zero-progress boundary (4.28). Referring to (4.16) with  $g = 0$  and  $h = N^3 A^2 / 8R^2 \sigma^{*2}$  one has

$$\begin{aligned} \varphi_R^{\text{II},*} &= \frac{c_\vartheta \sigma^*}{2} \frac{2}{\sqrt{1 + \frac{\sigma^{*2}}{2N} + \frac{N^3 A^2}{8R^4 \sigma^{*2}}}} - \frac{\sigma^{*2}}{2\mu} \stackrel{!}{>} 0, \quad \text{such that} \\ 2c_\vartheta \mu &> \sigma^* \sqrt{1 + \frac{\sigma^{*2}}{2N} + \frac{N^3 A^2}{8R^4 \sigma^{*2}}} \\ 4c_\vartheta^2 \mu^2 &> \sigma^{*2} + \frac{\sigma^{*4}}{2N} + \frac{N^3 A^2}{8R^4} \\ R^4 &> \frac{N^3 A^2}{8(4c_\vartheta^2 \mu^2 - \sigma^{*2} - \sigma^{*4}/2N)} = \frac{1}{4} \frac{N^4 A^2}{8Nc_\vartheta^2 \mu^2 - 2N\sigma^{*2} - \sigma^{*4}}. \end{aligned} \quad (4.29)$$

Positive progress requires inequality (4.29) to hold. Comparing (4.29) with zero-progress condition (4.28) yields

$$R^4 > \frac{1}{4} \frac{N^4 A^2}{8Nc_\vartheta^2 \mu^2 - 2N\sigma^{*2} - \sigma^{*4}} = R_{\varphi_0}^4, \quad (4.30)$$

such that  $\varphi_R^{\text{II},*} > 0$  is guaranteed for any  $R > R_{\varphi_0}$  given  $\sigma^*$  within the approximation of neglected exponentials. The opposite condition  $\varphi_R^{\text{II},*} < 0$  for  $R < R_{\varphi_0}$  also follows from the results above.

Figures 27 and 28 show the progress rate map from Figs. 23 and 24 with overlaid newly obtained approximation results for the progress rate zero. The white line depicts  $\varphi_R^{\text{II},*} = 0$  from (4.5) including all exponential terms. The black dashed line shows the result (4.28) for  $R_{\varphi_0}(\sigma^*)$  with neglected exponential terms.

**Relation to noisy sphere model** An important relation to the noisy sphere model can be made. In [3] the residual location error  $R_\infty$  was derived for the  $(\mu/\mu_I, \lambda)$ -ES assuming a constant noise strength  $\sigma_\epsilon$  in the limit  $\sigma^* \rightarrow 0$  as

$$R_\infty \simeq \sqrt{\frac{\sigma_\epsilon N}{4c_{\mu/\mu,\lambda}\mu}}. \quad (4.31)$$

Applying the limit  $\sigma^* \rightarrow 0$  to Eq. (4.30), identifying the constant noise strength of the Rastrigin function (for sufficiently large  $R$ ) as  $\sigma_{\text{Ras}} = \sqrt{NA^2/2}$ , see also (1.31) and (1.32), and taking the fourth root one gets

$$R_{\varphi_0} = \left( \frac{N^3 A^2}{32c_\vartheta^2 \mu^2} \right)^{1/4} = \left( \frac{NA^2}{2} \frac{N^2}{16c_\vartheta^2 \mu^2} \right)^{1/4} = \sqrt{\frac{\sigma_{\text{Ras}} N}{4c_\vartheta \mu}}, \quad (4.32)$$

which corresponds to result (4.31) with  $c_{\mu/\mu,\lambda} \rightarrow c_\vartheta$  and  $\sigma_\epsilon \rightarrow \sigma_{\text{Ras}}$ .

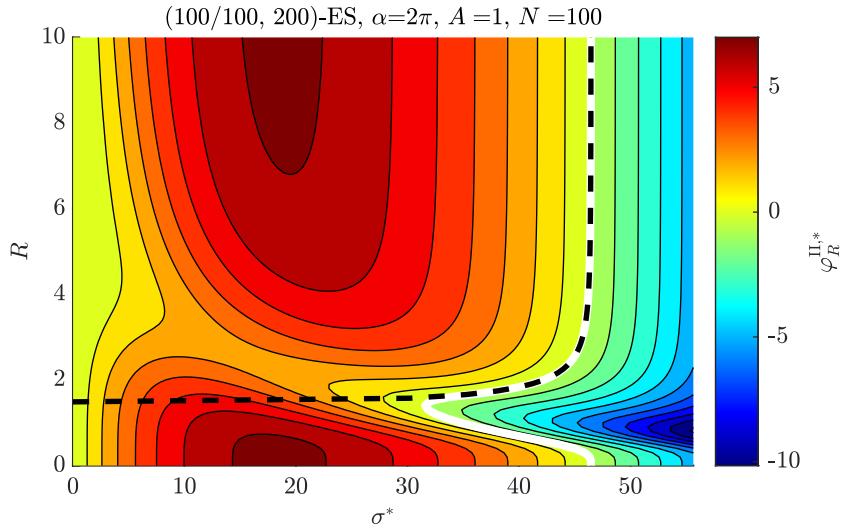


Figure 27: Zero-progress approximation using (4.28) with exponential factors set to zero (dashed line) compared to full  $R$ -dependent solution (4.5). The model is accurate for larger  $R$  and  $\sigma^*$  values, as both factors suppress the exponential terms. For smaller  $R$  the sphere function is slowly regained and the approximation fails at this “turning point”.

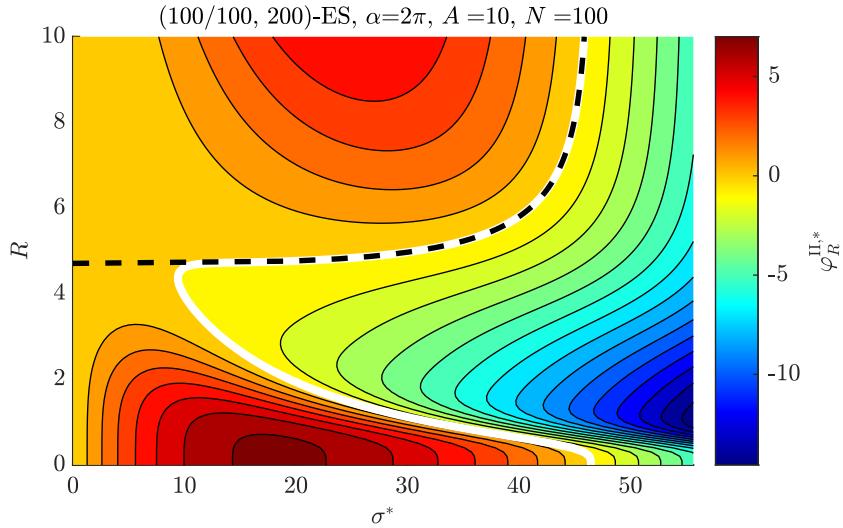


Figure 28: Zero-progress approximation for larger  $A = 10$  compared to Fig. 27. Again, the boundary is modeled for larger  $R$  and  $\sigma^*$  up to the turning point. This problem is significantly more difficult compared to  $A = 1$ , as the region of decreased progress is larger horizontally (for a broad mutation range) and vertically as a function of distance.

### 4.2.3 Transition Point of Vanishing Exponentials

After deriving the progress rate in the limit of vanishing exponentials, one has to address the question at which point the exponential terms are negligible or significant.

Therefore, the results for the transition points from (1.79) and (1.82) will be shown in the progress contour plots, see Fig. 29 with smaller attenuation factor  $\delta = 1$  and Fig. 30 with significantly higher factor  $\delta = 5$ . Result (1.79) is expected to yield worse results as it describes only the transition of the fitness function without any  $\sigma^*$ -dependence, while (1.82) is defined using a characteristic term of the progress gain part and its variance.

Given a transition relation, the further approach is to relate the distance of zero progress  $R_{\varphi_0}(\sigma^*)$  to the transition relation  $R_{tr}(\sigma^*)$ . The idea behind it is that for sufficiently small exponential terms, defined by  $R_{tr}(\sigma^*)$  via attenuation factor  $\delta$ , the zero-progress formula  $R_{\varphi_0}(\sigma^*)$  should be a valid approximation. The relation of both expressions should characterize strategy's behavior at the start of the transitional phase.

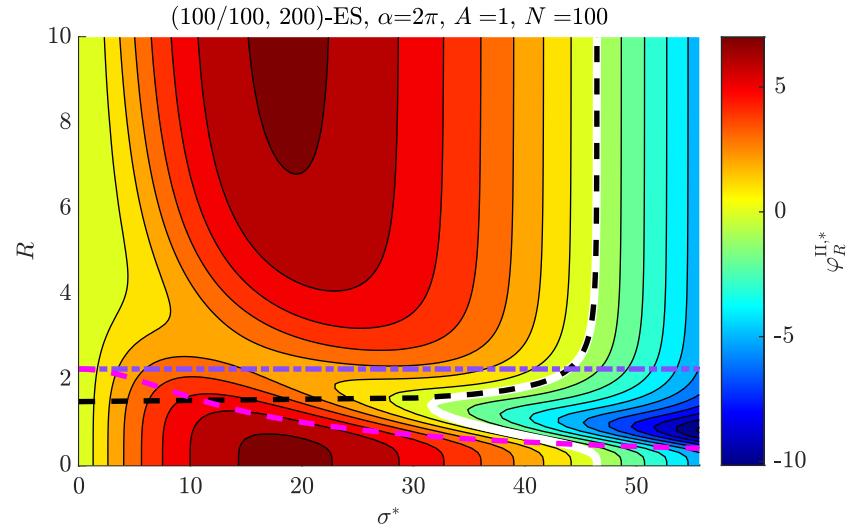


Figure 29: Transition formulae (1.79) [dash-dotted violet], and (1.82) [dashed magenta] for attenuation factor  $\delta = 1$ . Expression (1.79) has no  $\sigma^*$ -dependence and is therefore constant. Attenuation by  $e^{-1}$  means that the exponential factors are relatively large and one can see, that the sphere function is almost recovered for  $R$  below the magenta line.

### 4.2.4 Intersection point and population scaling

Now the intersection point  $\sigma_{sec}^*$  between the zero progress boundary  $R_{\varphi_0}(\sigma^*)$  and the transition relation  $R_{tr}(\sigma^*)$  is derived, see intersection between magenta and black dashed lines in Figs. 29 and 30. After deriving the relation one can deduce a population scaling relation  $\mu(N)$ . The transition point derived in

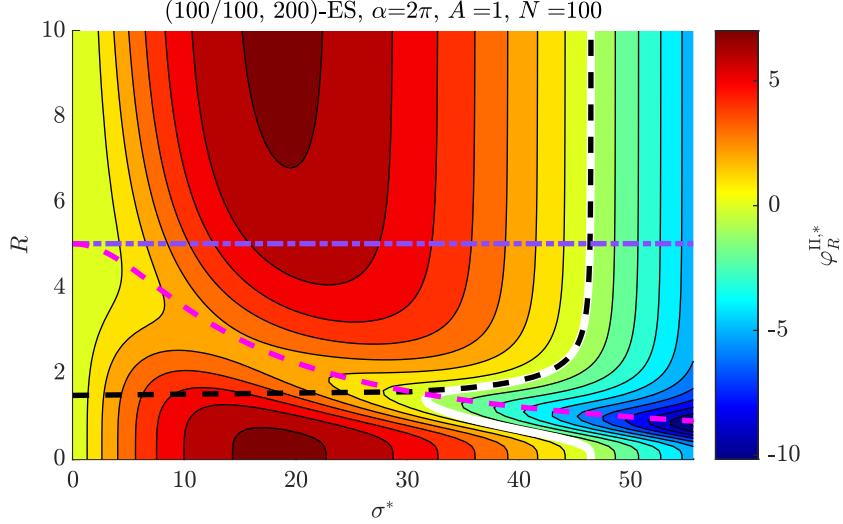


Figure 30: Transition formulae (1.79) [dash-dotted violet] and (1.82) [dashed magenta] for higher attenuation factor  $\delta = 5$ . The value was chosen exemplary for displaying purposes as the magenta line follows the valley of lower progress, which is interesting to note. Compared to Fig. 29, the attenuation (of the corresponding exponential term) is  $e^{-1}/e^{-5} \approx 55$  times larger.

(1.82) is recalled as

$$R_{tr} = \frac{\sqrt{2\delta N}}{\alpha} \frac{1}{\sqrt{1 + \sigma^{*2}/N}}. \quad (4.33)$$

Setting  $R_{\varphi_0}^4 = R_{tr}^4$  by using (4.30) and (4.33) yields after rearranging the terms

$$-\frac{1}{4} \frac{N^4 A^2}{\sigma^{*4} + 2N\sigma^{*2} - 8Nc_\vartheta^2\mu^2} = \frac{4\delta^2 N^4}{\alpha^4} \frac{1}{\sigma^{*4} + 2N\sigma^{*2} + N^2}. \quad (4.34)$$

Collecting the corresponding factors of  $\sigma^*$  yields

$$\begin{aligned} \left(1 + \frac{16\delta^2}{\alpha^4 A^2}\right) \sigma^{*4} + 2N\sigma^{*2} \left(1 + \frac{16\delta^2}{\alpha^4 A^2}\right) + N^2 - \frac{128\delta^2 N c_\vartheta^2 \mu^2}{\alpha^4 A^2} &= 0 \\ \sigma^{*4} + 2N\sigma^{*2} + \frac{N^2 (\alpha^4 A^2 - 128\delta^2 c_\vartheta^2 \mu^2 / N)}{\alpha^4 A^2 + 16\delta^2} &= 0. \end{aligned} \quad (4.35)$$

An equation of the form (4.20) was obtained and now solution (4.21) can be applied

$$\begin{aligned} \sigma_{\text{sec}}^* &= \left[ \left( N^2 - N^2 \frac{\alpha^4 A^2 - 128\delta^2 c_\vartheta^2 \mu^2 / N}{\alpha^4 A^2 + 16\delta^2} \right)^{1/2} - N \right]^{1/2} \\ &= \left[ \left( \frac{16\delta^2 N^2 + 16\delta^2 N^2 \frac{8c_\vartheta^2 \mu^2}{N}}{16\delta^2 (1 + \frac{\alpha^4 A^2}{16\delta^2})} \right)^{1/2} - N \right]^{1/2}, \end{aligned} \quad (4.36)$$

such that the intersection point is obtained as

$$\sigma_{\text{sec}}^* = \left[ N \frac{\left(1 + \frac{8c_\vartheta^2 \mu^2}{N}\right)^{1/2}}{\left(1 + \frac{\alpha^4 A^2}{16\delta^2}\right)^{1/2}} - N \right]^{1/2}. \quad (4.37)$$

Two examples for result (4.37) are shown in Fig. 31. The transition parameter was set to  $\delta = 5$  analogous to Fig. 30. The intersection point  $\sigma_{\text{sec}}^*$  serves as an approximation of the characteristic progress dip location. The expression (4.37) is further investigated now. In order to have a converging normalized mutation strength (on the sphere) it must hold

$$0 < \sigma_{\text{sec}}^* < \sigma_{\varphi_0, \text{sph}}^*, \quad (4.38)$$

with the sphere-zero  $\sigma_{\varphi_0, \text{sph}}^*$  given in Eq. (4.22). The relation  $\sigma_{\text{sec}}^* < \sigma_{\varphi_0, \text{sph}}^*$  follows immediately for any  $A, \alpha, \delta > 0$ . Setting  $A = 0$  or  $\alpha = 0$  recovers Eq. (4.22). Requiring a positive mutation strength  $\sigma_{\text{sec}}^* > 0$  one must have

$$\begin{aligned} \left(1 + \frac{8c_\vartheta^2 \mu^2}{N}\right)^{1/2} &> \left(1 + \frac{\alpha^4 A^2}{16\delta^2}\right)^{1/2} \\ \frac{8c_\vartheta^2 \mu^2}{N} &> \frac{\alpha^4 A^2}{16\delta^2}. \end{aligned} \quad (4.39)$$

Now one can solve (4.39) for  $\mu$  to fulfill the inequality as

$$\mu^2 > \frac{N \alpha^4 A^2}{128 c_\vartheta^2 \delta^2}. \quad (4.40)$$

Taking the square-root of (4.40) one arrives at the important result

$$\mu > \sqrt{\frac{N}{2} \frac{\alpha^2 A}{8c_\vartheta \delta}}. \quad (4.41)$$

Result (4.41) can be regarded as a population sizing relation  $\mu(N, A, \alpha)$  for given (fixed) truncation ratio  $\vartheta$  and transition parameter  $\delta$ . It models the scaling of the characteristic progress dip within the  $R$ -dependent progress rate formulation. Note that expression (4.41) can also be rewritten using the Rastrigin-specific variance term  $\sigma_{\text{Ras}}$  from (1.32) as

$$\mu > \frac{\sigma_{\text{Ras}} \alpha^2}{8c_\vartheta \delta}. \quad (4.42)$$

**Population scaling** In this section the population scaling  $\mu(N)$  is investigated for constant  $\sigma^*$ . Experimental results of real optimization runs with  $\sigma^* = \hat{\sigma}_{\text{sph}}^*$  are shown in Fig. 32 for  $\vartheta = 0.25$  and  $A = 1$ . The maximum number of generations was set to 5000, as it is necessary to terminate non-converging runs. Furthermore, for each given  $\mu$  and  $N$  the sphere-optimal value  $\hat{\sigma}_{\text{sph}}^*$  was

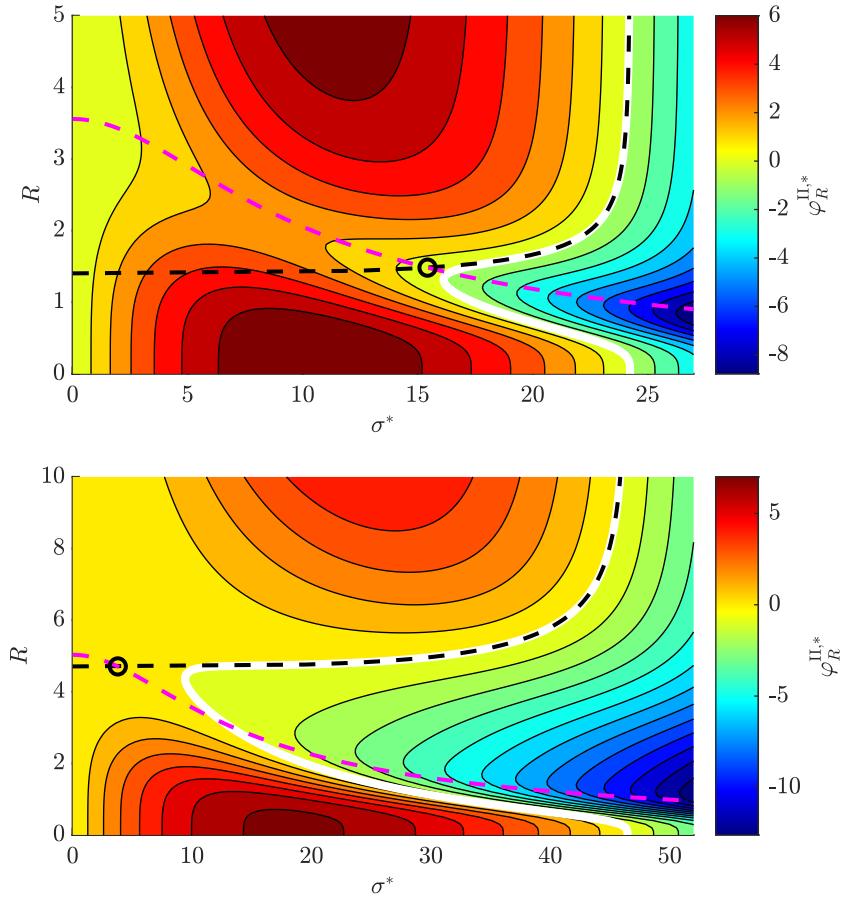


Figure 31: Plots showing intersection point (4.37) with transition parameter  $\delta = 5$  for  $(25/25, 100)$ -ES with  $N = 50$  and  $A = 1$  (top), and  $(100/100, 200)$ -ES with  $N = 100$  and  $A = 10$  (bottom).

calculated numerically using [5, Eq. (6.54)]. Random initial  $\mathbf{y}$ -values were chosen with  $\|\mathbf{y}\| = 20\sqrt{N}$  outside the local attraction region. The results for small  $N$  and therefore small  $\mu$  need to be interpreted carefully, as fluctuation effects tend to be pronounced. Given a maximum of 5000 generations (fixed number for all  $N$ ) and constant  $\sigma^*$ , the dimensionality  $N = 30$  is significantly easier to optimize than  $N = 50$  due to fluctuations and the ES ‘‘spontaneously’’ descending into the global attractor. This effect is more pronounced for smaller  $N$  and  $\mu$  than for large values thereof. For larger  $N$  a different behavior is observed showing a sub-linear relation  $\mu(N)$ .

Now the results (4.41) for the population scaling  $\mu = O(\sqrt{N})$  is investigated numerically in Fig. 33 by assuming constant  $\sigma^*$ . To this end, three  $\sigma^*$ -values were chosen relative to the (numerically) obtained sphere optimal value  $\hat{\sigma}_{\text{sph}}^*$ . The values are, from top to bottom,  $\sigma^* = \{\hat{\sigma}_{\text{sph}}^*, \hat{\sigma}_{\text{sph}}^*/2, (\hat{\sigma}_{\text{sph}}^* + \sigma_{\varphi_0, \text{sph}}^*)/2\}$ . Given fixed  $\vartheta$ ,  $N$ ,  $A$ , and  $\alpha$ , the population size  $\mu$  was successively increased and

the  $R$ -dependent progress rate (4.5) was evaluated given  $\sigma^*$ . Global convergence occurs at some  $\mu$ , if  $\varphi_R^{\text{II},*}(\sigma^*) > 0$  for all  $R$ . If  $\varphi_R^{\text{II},*}(\sigma^*) \leq 0$  occurs the corresponding  $\mu$  is marked unsuccessful. The numerical experiments were done by discretizing  $\sigma^* \in [0, \text{ceil}(\sigma_{\varphi_0, \text{sph}}^*)]$  and  $R \in [0, \text{ceil}(3\sqrt{N})]$  to cover the range of possible negative progress. Local attraction effects are not modeled by  $\varphi_R^{\text{II}}$ , as already mentioned.

In Fig. 33 one can see that the scaling  $\mu = O(\sqrt{N})$  is recovered in all plots. Important to note is that for the central plot ( $\sigma^* = \hat{\sigma}_{\text{sph}}^*/2$ ) oscillation strength  $A$  had to be increased to  $A = 10$  compared to top and bottom plots ( $A = 1$ ). The reason is that for  $A = 1$  global convergence occurred for all  $N$  for any  $\mu \geq 1$  giving a scaling  $O(1)$ . This can be attributed to small  $\sigma^* = \hat{\sigma}_{\text{sph}}^*/2$  relative to a comparably small  $A = 1$ , such that the progress dip has no significant influence. For the bottom plot the  $\sqrt{N}$ -scaling is recovered for larger  $N$ . For smaller  $N$  similar population sizes  $\mu(N)$  are needed for global convergence, in contrast to Fig. 32. This can be attributed to the non-linearity of Eq. (4.5) and the particular choice  $\sigma^* = (\hat{\sigma}_{\text{sph}}^* + \sigma_{\varphi_0, \text{sph}}^*)/2$ , such that a small increase in  $N$  does not necessarily require larger  $\mu$ -values.

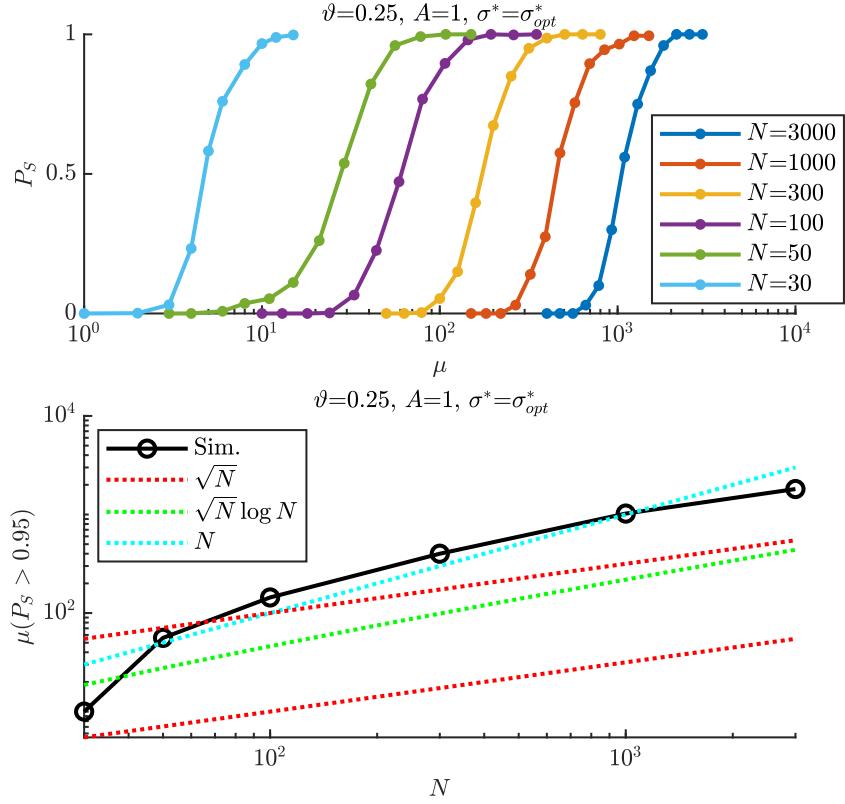


Figure 32: Population scaling experiments for  $\vartheta = 1/4$ ,  $\alpha = 2\pi$ ,  $A = 1$ , and  $\sigma^* = \hat{\sigma}_{\text{sph}}^*$ . The maximum number of generations was set to 5000 and for each data point 300 trials were evaluated. Fluctuation effects are more pronounced for smaller  $N$  and  $\mu$ .

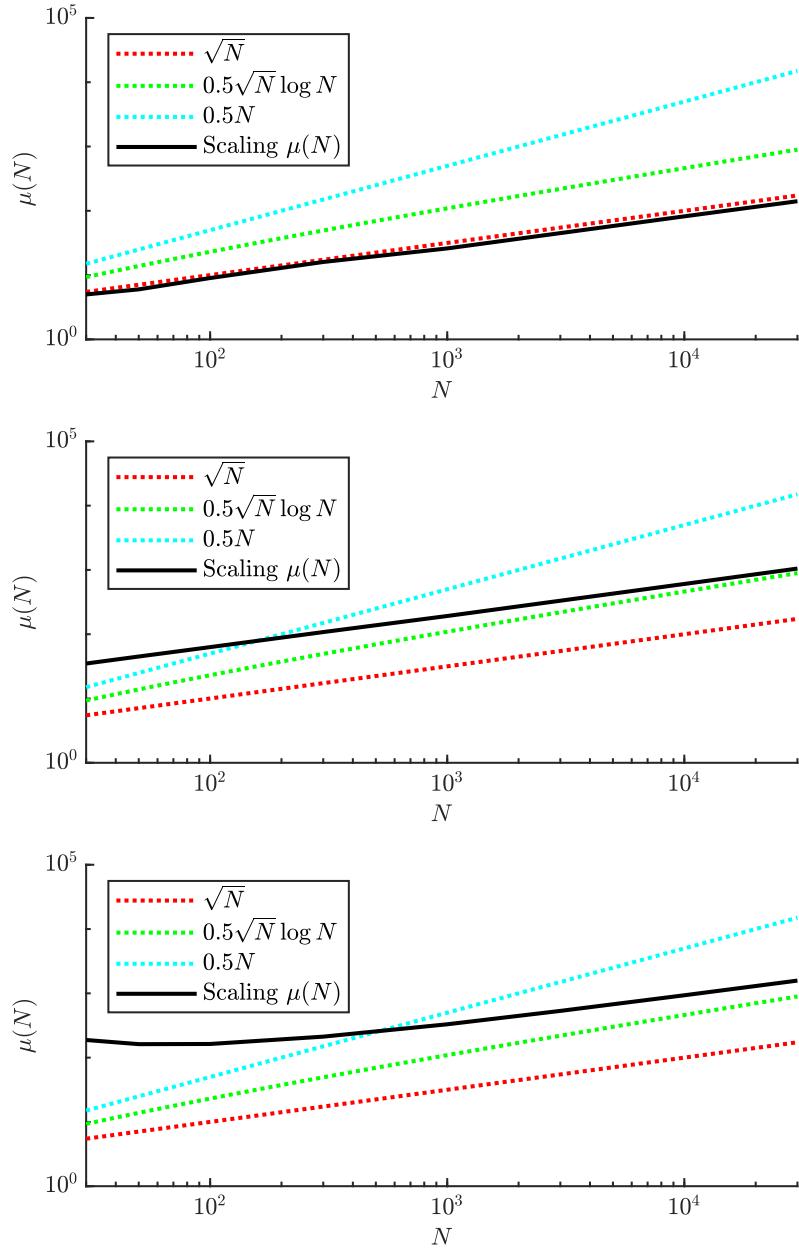


Figure 33: Numerically obtained scaling relation  $\mu(N)$  for  $\vartheta = 1/4$  and  $\alpha = 2\pi$  by evaluating the  $R$ -dependent progress rate (4.5). Top plot shows  $A = 1$  with  $\sigma^* = \hat{\sigma}_{\text{sph}}^*$ , central plot  $A = 10$  with  $\sigma^* = \hat{\sigma}_{\text{sph}}^*/2$ , and bottom plot  $A = 1$  with  $\sigma^* = (\hat{\sigma}_{\text{sph}}^* + \sigma_{\varphi_0, \text{sph}}^*)/2$ .

### 4.3 Investigation of Local Attraction and Global Convergence

In this section different ideas are documented considering the modeling of local attraction. As it was already explained in the previous sections, the  $R$ -dependent formula (4.5) is not able to model local attraction phenomena for small mutation strengths. This is due to the assumption of independent normally distributed positions  $y_i$  around the global optimum. This assumption holds well for large mutations, see Fig. 35, as they tend to spread out the candidate solutions more evenly in search space. For smaller mutations local effects tend to dominate and the resulting  $y_i$ -components are distributed mostly around the local minima locations resembling a discrete distribution.

As a first step, see Fig. 34, dynamic experiments are performed for an exemplary configuration with different values  $\sigma^* \in (0, \sigma_{\varphi_0, \text{sph}}^*)$  and given two different values  $A = 1$  (top) and  $A = 10$  (bottom). For each  $\sigma^*$ , 100 runs are conducted and the residual distance  $R_{st}$  of the unsuccessful runs is measured within the steady state after many generations. The maximum number of generations was fixed to 5000, as there is no local convergence in the sense of  $\sigma^* \rightarrow 0$ . For each trial a new random initialization for  $\mathbf{y}$  was chosen with  $\|\mathbf{y}\| = R_0 = 50\sqrt{N}$ , as for  $A = 10$  ( $N = 1$ ) the last local minimum is at  $y_i \approx 31$ . This is necessary to start outside the local attraction region and to prevent the ES reaching the same local attractor for each trial if very small  $\sigma^*$ -values are given. For each trial the average  $R_{st}$  is measured within the last 1000 generations. Then, the average  $R_{st}$  is then taken over all 100 trials to obtain the values shown in Fig. 34. The success probability  $P_S$  is shown on the right axis (not shown for  $A = 10$  as there are no globally converging runs).

One can observe a characteristic shape of  $R_{st}(\sigma^*)$ . For large  $\sigma^*$ , this behavior was predicted by the zero-progress lines derived in the previous section, see illustration of Fig. 36. The zero-progress lines were obtained by including the (constant) noise term  $NA^2/2$  within the quality gain variance. The residual distance  $R_{st}$  for large  $\sigma^*$  can therefore be attributed to the overall noise effect of the sum over all cosine terms. The effect of local convergence for small  $\sigma^*$ , and global convergence for certain intermediate values, e.g.  $\sigma^* \approx 30$  in Fig. 34 ( $A = 1$ ), needs further investigation.

To gain some knowledge and insight about local attraction, a hybrid model is used starting from Eq. (4.1). In order to keep the equations tractable the variance is taken as  $R$ -dependent, assuming that the effect of the trigonometric terms for large dimensionality has an averaging effect, such that it can be expressed as  $R$ -dependent. On the other side, the gain term  $\sum_i y_i \sin(\alpha y_i)$  will be kept  $y_i$ -dependent, such that following expression will be investigated

$$\varphi_R^{\text{II}}(R, \mathbf{y}) = c_\vartheta \frac{\sigma^2}{D_Q(R, \sigma)} \left( 4R^2 + e^{-\frac{1}{2}(\alpha\sigma)^2} 2\alpha A \sum_{i=1}^N y_i \sin(\alpha y_i) \right) - N \frac{\sigma^2}{\mu}. \quad (4.43)$$

As an example for the application of “hybrid” formula (4.43), denoted by “HYB”, a one generation experiment is shown in Fig. 35. The hybrid formula is able to regenerate the negative progress dip for small  $\sigma^*$  despite using an  $R$ -dependent variance, similar to purely  $\mathbf{y}$ -dependent formula B2. Progress rate  $R$  on the other hand does not model local attraction.

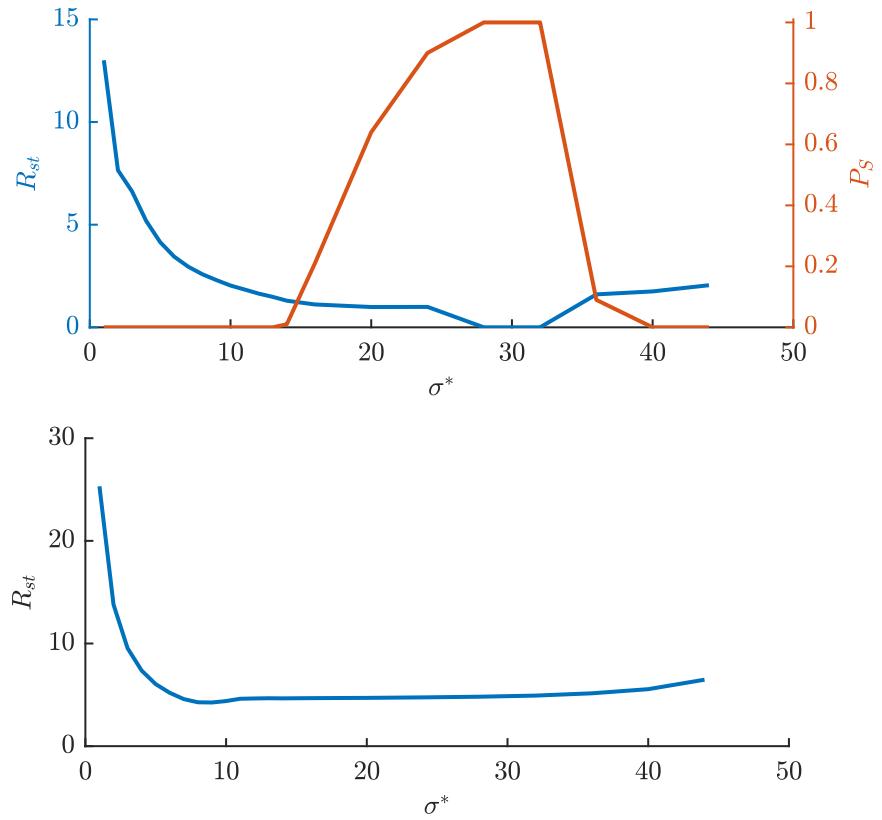


Figure 34: Dynamic experiments with  $(100/100, 200)$ -ES,  $\alpha = 2\pi$ ,  $N = 100$ ,  $A = 1$  (top), and  $A = 10$  (bottom). The average residual distance  $R_{st}$  is evaluated over 100 repetitions. The convergence probability is shown along the right axis (top plot, bottom plot has  $P_S = 0$  for all trials).

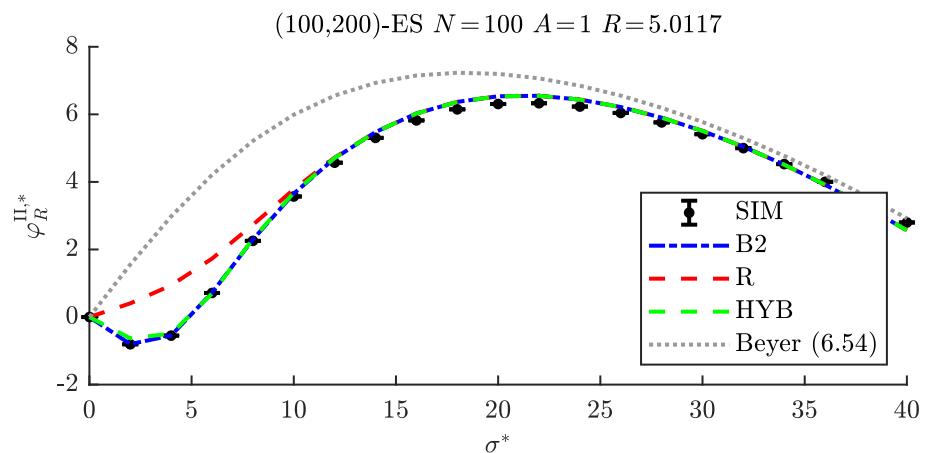


Figure 35: The progress rate approximations are compared at a certain  $y$  ( $R \approx 5$ ), where local attraction effects are observed for small  $\sigma^*$ , see also Fig. 58 in the following section.

### 4.3.1 Local attraction along the diagonal

Each attractor region can be considered differently depending on the actual  $y_i$ -coordinates. To illustrate the effects of attraction on the progress contour map, one can set  $y_i = R/\sqrt{N}$  for all  $i = 1, \dots, N$  given some  $R$  in hybrid Eq. (4.43). This way one moves along the diagonal of strictly positive  $y_i$  through search space, see Fig. 36 with overlaid  $R_{st}$ -curve from Fig. 34.

For small mutations local attraction regions are present, where progress is entirely negative, e.g. for  $R \in [5, 10]$ . The characteristic of the  $R_{st}$ -curve is not modeled correctly. However, for large mutations the spherical structure is recovered, where the  $R_{st}$ -curve agrees very well with the progress rate result. One can observe that the strategy faces negative progress regions for a broad range of mutations strengths. This effect can of course be counteracted by increasing the population size of the strategy. The population in Fig. 36 was deliberately chosen too small (for  $A = 10$  and  $N = 100$ ) to illustrate the problem.

This approach illustrates the problem of local attraction, it is not a complete model. However, within the effect of component equipartition during dynamic iterations, see Sec. 5.4, components are reaching values  $y_i = R/\sqrt{N}$  (after many generations) and the progress rate along the diagonal will be relevant.

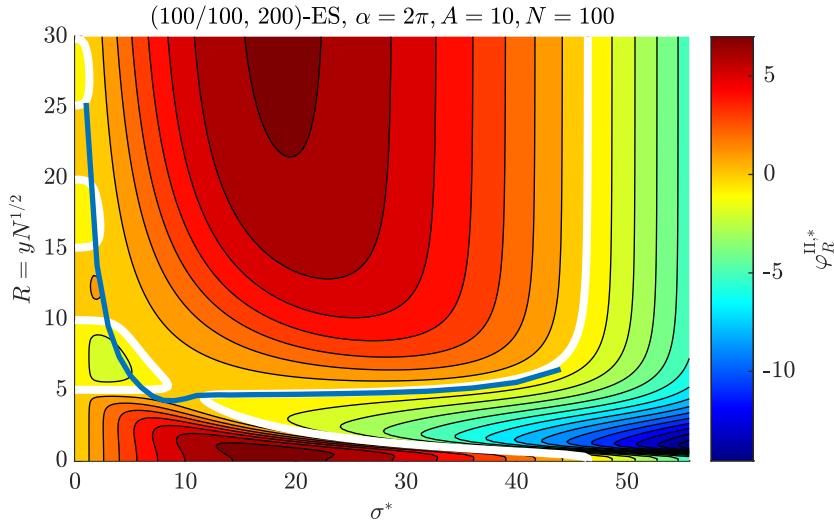


Figure 36: Progress rate contour map using hybrid Eq. (4.43) with  $R$ -dependent variance and locations  $y_i = R/\sqrt{N}$  for all  $i = 1, \dots, N$  given  $R$ . The blue curve shows the result of Fig. 34. The plot illustrates the effects of local attraction (small  $\sigma^*$ ) on the progress rate.

### 4.3.2 Local attraction via a probabilistic approach

For the second approach the sum in Eq. (4.43) is modeled as a normally distributed random variable  $Y$  using the CLT, which was already introduced in (4.3). However, now the fluctuations will not be neglected and the progress will

be treated as a random variable. We define

$$E_Y := \text{E}[Y] \quad \text{and} \quad D_Y^2 := \text{Var}[Y], \quad (4.44)$$

with  $E_Y$  evaluated in (A.34), and  $D_Y^2$  in (A.38). Then, one can demand positive progress

$$\varphi_R^{\text{II}} = c_\vartheta \frac{\sigma^2}{D_Q(R, \sigma)} \left( 4R^2 + e^{-\frac{1}{2}(\alpha\sigma)^2} 2\alpha A [E_Y + D_Y \mathcal{N}(0, 1)] \right) - N \frac{\sigma^2}{\mu} > 0, \quad (4.45)$$

and rearrange the equation to isolate the random variate according to

$$\mathcal{N}(0, 1) > \frac{e^{\frac{1}{2}(\alpha\sigma)^2}}{2\alpha A D_Y} \left( \frac{ND_Q}{c_\vartheta \mu} - 4R^2 \right) - \frac{E_Y}{D_Y}. \quad (4.46)$$

From (4.46) one can deduce the probability  $P_\varphi$  to have positive progress using the CDF of the normal distribution for given  $R$  and  $\sigma$  according to

$$\begin{aligned} P_\varphi &:= \Pr\{\varphi_R^{\text{II}} > 0 | R, \sigma\} = 1 - \Pr\{\varphi_R^{\text{II}} \leq 0 | R, \sigma\} \\ &= 1 - \Phi\left(\frac{e^{\frac{1}{2}(\alpha\sigma)^2}}{2\alpha A D_Y} \left( \frac{ND_Q}{c_\vartheta \mu} - 4R^2 \right) - \frac{E_Y}{D_Y}\right). \end{aligned} \quad (4.47)$$

Figure 37 displays the results of (4.47) applied to two parameter configurations. Dark red regions correspond to  $P_\varphi = 1$ , while dark blue regions show  $P_\varphi = 0$ . A sharp boundary can be observed for large  $\sigma^*$  and  $R$  corresponding to the spherical limit including the characteristic dip of low progress (probability). Within these limits the Rastrigin noise variance term  $NA^2/2$  dominates and the progress probability yields a sharp transition. This can also be explained by looking at (4.45), where the random variable vanishes with prefactor  $e^{-\frac{1}{2}(\alpha\sigma)^2}$  yielding a deterministic equation.

For small mutations an interesting (orange-yellow) region of decreased progress probability emerges. This can be attributed to local attraction effects giving negative (or zero) realizations  $\varphi_R^{\text{II}} \leq 0$ . This is in contrast to the sphere function, where strictly positive progress  $\varphi^* > 0$  for  $\sigma^* > 0$  in the limit of  $\sigma^* \rightarrow 0$  is expected.

One can see that a continuous region of decreased  $P_\varphi$  is present w.r.t.  $\sigma^*$ . This matches the experimental observations of Fig. 34, especially for  $A = 10$ . However, result (4.47) is only an indicator of local convergence effects and not a complete model. Local attraction can also occur with  $\varphi_R^{\text{II}} > 0$ , if local and global attractor are aligned in such a way.

The dynamics of Eq. (4.45) is investigated in Fig. 38 for small  $\sigma^* = 1$  and large  $A = 10$  using three different random number generator seeds compared to its deterministic expression setting the fluctuation  $D_Y = 0$ . This is done using the iteration scheme of Eq. (5.6). The overall effect of the fluctuations of Eq. (4.45) are very small, as there are only minor differences in the overall  $R$ -dynamics. The progress probability  $P_\varphi$  was checked during the iteration and yields consistent results as shown in Fig. 37. Negative progress occurs, but with comparably small contributions. All runs are converging globally and  $R_{st}$  of Fig. 34 is not recovered. Therefore, fluctuations using hybrid Eq. (4.45) are not sufficient to recover experimental results of local convergence.

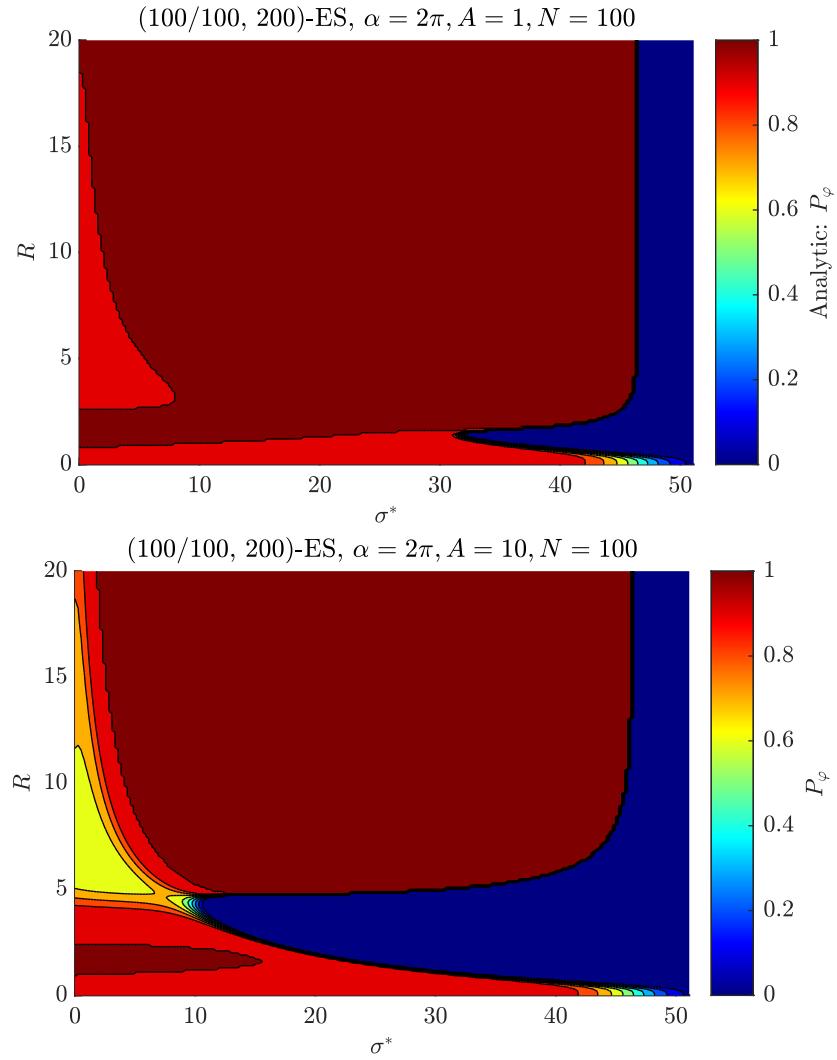


Figure 37: Probability of positive progress  $P_\varphi$  from result (4.47) modeling the sum over  $y_i$ -dependent terms from (4.43) as a normally distributed variate. Decreased  $P_\varphi$ -values for small  $\sigma^*$  indicate local convergence effects. Top plot shows  $A = 1$  and bottom plot  $A = 10$  with (100/100, 200)-ES,  $\alpha = 2\pi$ , and  $N = 100$ .

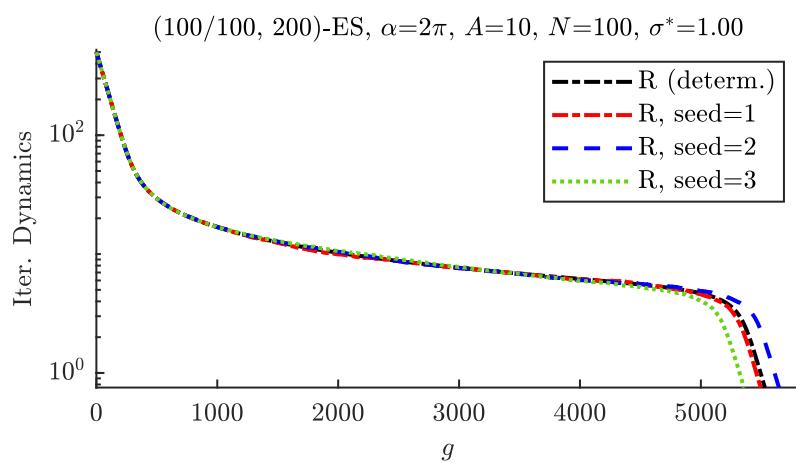


Figure 38: Dynamics of Eq. (4.45) for (100/100, 200)-ES,  $\alpha = 2\pi$ ,  $N = 100$ ,  $\sigma^* = 1$ , and  $A = 10$ .

### 4.3.3 Positive gain condition

The goal of this section is to derive a condition on the mutation strength, such that positive progress is ensured for all component-wise progress rates  $\varphi_i^{\text{II}}$ . Starting from Eq. (3.116) one has

$$\varphi_i^{\text{II}} = c_\vartheta \frac{\sigma^2}{D_Q} \left( 4y_i^2 + e^{-\frac{1}{2}(\alpha\sigma)^2} 2\alpha A y_i \sin(\alpha y_i) \right) - \frac{\sigma^2}{\mu}. \quad (4.48)$$

The gain function  $G$  is defined as

$$G(y_i, \sigma) := 4y_i^2 + e^{-\frac{1}{2}(\alpha\sigma)^2} 2\alpha A y_i \sin(\alpha y_i). \quad (4.49)$$

Requiring positive progress  $\varphi_i^{\text{II}} > 0$  for Eq. (4.48) yields

$$\frac{c_\vartheta}{D_Q} G(y_i, \sigma) > \frac{1}{\mu}. \quad (4.50)$$

The only term of (4.50) which is not necessarily positive is the gain function. Additionally, even for  $G > 0$ , depending on the actual  $y_i$ - and  $\sigma$ -values the inequality  $\varphi_i^{\text{II}}(\mathbf{y}) > 0$  may not be fulfilled. However, one can consider the infinite population size limit  $\mu \rightarrow \infty$ , as the only  $\mu$ -dependent term is on the rhs of (4.50). Within this limit it suffices to show that  $G > 0$  to satisfy inequality (4.50). An exemplary numerical evaluation of (4.49) for  $A = 10$ , and  $\alpha = 2\pi$  is shown in Fig. 39 for  $\sigma \in [0, 1]$  and  $y_i \in [0, 5]$ . Negative gain values are shown in darker blue colors and the  $G = 0$  boundary is displayed in bold white. A derivation of the dashed black line (constant  $\sigma$ ) ensuring positive gain for any  $y_i$  is given now. The  $\sigma$ -value ensuring  $G > 0$  will be referred to as  $\sigma_{\text{esc}}$  (escaping the local attractor).

Besides the condition  $G = 0$  an additional condition is needed. From Fig. 39 it can be inferred that  $\frac{\partial G}{\partial y_i} = 0$  must hold at the point of vanishing gain. Due to the periodicity there will be multiple solutions. However, the contribution of a negative sine term will be largest for small  $\|y_i\| < 1$ , where  $y_i^2 < \|y_i\|$ .

Requiring  $G \stackrel{!}{=} 0$  and assuming  $y_i \neq 0$ , Eq. (4.49) can be formulated as

$$G = 2y_i \tilde{G}, \quad (4.51)$$

with  $\tilde{G}$  defined as

$$\tilde{G} := 2y_i + e^{-\frac{1}{2}(\alpha\sigma)^2} \alpha A \sin(\alpha y_i) \stackrel{!}{=} 0. \quad (4.52)$$

The second condition  $\frac{\partial G}{\partial y_i} = 0$  yields for (4.51)

$$\frac{\partial G}{\partial y_i} = 2\tilde{G} + 2y_i \frac{\partial \tilde{G}}{\partial y_i} \stackrel{!}{=} 0, \quad (4.53)$$

As  $\tilde{G} = 0$  and  $y_i \neq 0$ ,  $\frac{\partial G}{\partial y_i} = 0$  is equivalent to  $\frac{\partial \tilde{G}}{\partial y_i} = 0$ , see (4.53). Therefore one gets

$$\frac{\partial \tilde{G}}{\partial y_i} = 2 + e^{-\frac{1}{2}(\alpha\sigma)^2} \alpha^2 A \cos(\alpha y_i) = 0, \quad (4.54)$$

such that the following condition is obtained

$$e^{-\frac{1}{2}(\alpha\sigma)^2}\alpha A = -\frac{2}{\alpha \cos(\alpha y_i)}. \quad (4.55)$$

Inserting condition (4.55) into (4.52), it follows

$$2y_i - \frac{2 \sin(\alpha y_i)}{\alpha \cos(\alpha y_i)} = 0. \quad (4.56)$$

Introducing the substitution  $x = \alpha y_i$  and applying  $\sin x / \cos x = \tan x$  yields

$$\frac{2}{\alpha}(x - \tan x) = 0. \quad (4.57)$$

The first non-trivial solution of (4.57) satisfying  $x = \tan x$  is

$$x_0 \approx 4.493. \quad (4.58)$$

Multiplying (4.52) by  $\alpha$ , identifying  $x_0 = \alpha y_i$  and  $\sigma = \sigma_{esc}$  (point of vanishing gain) results in

$$2x_0 + e^{-\frac{1}{2}(\alpha\sigma_{esc})^2}\alpha^2 A \sin x_0 = 0 \\ e^{\frac{1}{2}(\alpha\sigma_{esc})^2} = -\frac{\alpha^2 A \sin x_0}{2x_0}. \quad (4.59)$$

Resolving (4.59) for  $\sigma_{esc}$  yields the final result

$$\sigma_{esc} = \frac{1}{\alpha} \sqrt{2 \ln \left( -\frac{\alpha^2 A \sin x_0}{2x_0} \right)} \\ \sigma_{esc} \approx \frac{1}{\alpha} \sqrt{2 \ln (0.1086 \alpha^2 A)}. \quad (4.60)$$

Numerical evaluations of (4.60) for  $\alpha = 2\pi$  (default) and different  $A$ -values is shown in Table 1.

$A$	1	2	5	10
$\sigma_{esc}$	0.272	0.330	0.394	0.436

Table 1: Evaluation of (4.60) for constant  $\alpha = 2\pi$  and different values of  $A$ .

The result of (4.60) can be interpreted as a stability criterion of the ES to achieve positive component wise-progress. It represents a worst-case scenario considering the last local attractor and the position requiring the largest  $\sigma$  at  $y_i \approx 0.75$ , see Fig. 39. This criterion is valid for all  $y_i$  until the global attractor basin is reached, where  $\sigma < \sigma_{esc}$  must be chosen to have convergence. An exemplary evaluation of (4.60) is shown in Fig. 39, see black dashed line.

Figures 40 and 41 show  $\sigma_{esc}$  as a black dashed line in a  $\sigma^*-R$ -plot with the characteristic declining function  $R = \sigma N/\sigma^*$ . The oscillation amplitude  $A = 10$  is chosen relatively large and tackled using large ES-populations, which is in accordance with the assumptions made above. Figure 41 shows different configurations (by varying  $\tau$ ) of the median of unsuccessful  $\sigma$ SA-ES runs. Configurations using larger  $\tau$ -values tend to converge locally when  $\sigma$  falls below  $\sigma_{esc}$ .

Runs with smaller  $\tau$  (showing the highest  $P_S$ ) tend to follow the path between progress rate zero and above  $\sigma_{esc}$  to minimize the residual distance (and maximize the success rate). However, falling below  $\sigma_{esc}$  does not necessarily mean that local convergence must occur. Due to fluctuations the global attractor can also be reached ( $P_S > 0$  for all  $\tau$ -values).

Figure 41 shows single ES-runs with constant  $\sigma^*$  (same parameters as in Fig. 40). None of the runs are converging globally. Therefore the residual distance is evaluated for the last 1000 generations and shown as blue dots. Both the Rastrigin noise (larger  $\sigma^*$ ) and  $\sigma_{esc}$  (smaller  $\sigma^*$ ) pose a challenge for the ES to find the global attractor. The results are similar to previous results of Fig. 34.

From these observations one can conclude, that a strategy with constant  $\sigma = \sigma_{esc}$  should be able to find the global attractor. Figure 42 shows the same parameter set as Figs. 40 and 41, with three configurations  $\sigma = \sigma_{esc} \approx 0.43$ ,  $\sigma = 0.4$ , and  $\sigma = 0.35$  and fixed maximum number of generations  $N_G = 10^4$ . The black line follows the characteristic path of  $\sigma_{esc} \approx 0.43$  up until the global attractor is reached, while the two smaller  $\sigma$ -values show stagnating behavior at larger  $y_i$ -values, which can be inferred from the distribution of  $y_i$ -values in Fig. 43. A second example is shown in Fig. 44, where at  $A = 5$  one has  $\sigma_{esc} \approx 0.39$ . The distribution indicates that the global attractor is reached. However, due to smaller  $\mu = 200$  the fluctuations tend to be larger and some  $y_i$  (over 1000 generations) are found outside  $[-0.5, 0.5]$ , see axis limits (bars barely visible, but present). This underlines the importance of the limit  $\mu \rightarrow \infty$ .

The condition for  $\sigma_{esc}$  from Eq. (4.60) contains no dependency on  $N$  or  $\mu$ . However, increasing  $A$  results in an increase of the Rastrigin noise term  $\sigma_R = \sqrt{N}/2A$ . To reduce the expected residual distance at which the noise term starts to dominate,  $\mu$  has to be increased appropriately, see also (4.32). If  $\mu$  is too small compared to  $A$ , intersection of the zero-progress line and  $\sigma_{esc}$  may occur at  $R \approx R_{\varphi_0}$ , see Fig. 45. This is the worst case, in which no further progress is expected to occur for any  $\sigma^*$ .

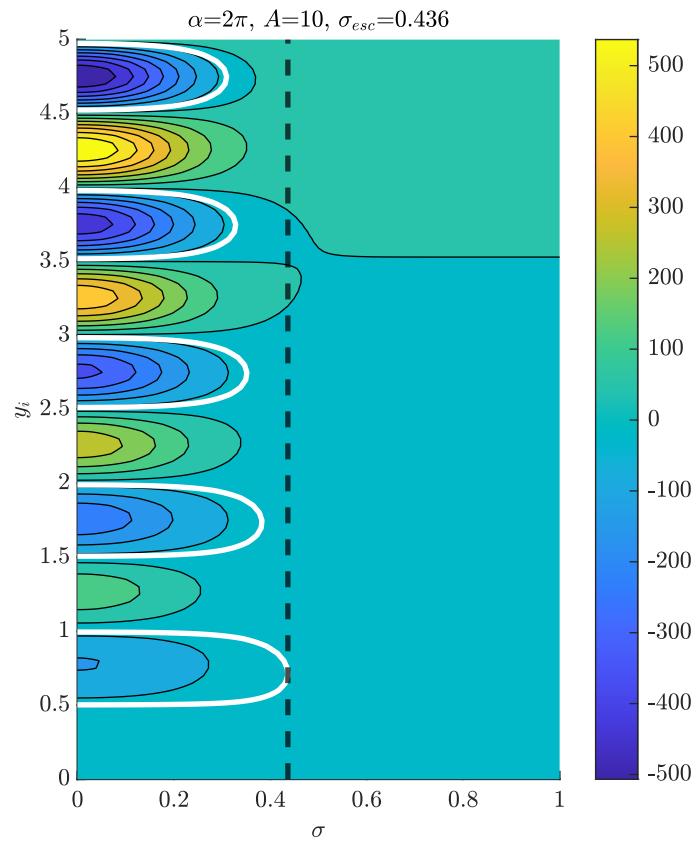


Figure 39: Gain function  $G(y_i)$  for  $\alpha = 2\pi$  and  $A = 10$ , see Eq. (4.49). The  $G = 0$  boundary is shown in bold white. The value of  $\sigma_{esc} \approx 0.436$  is shown in dashed black.

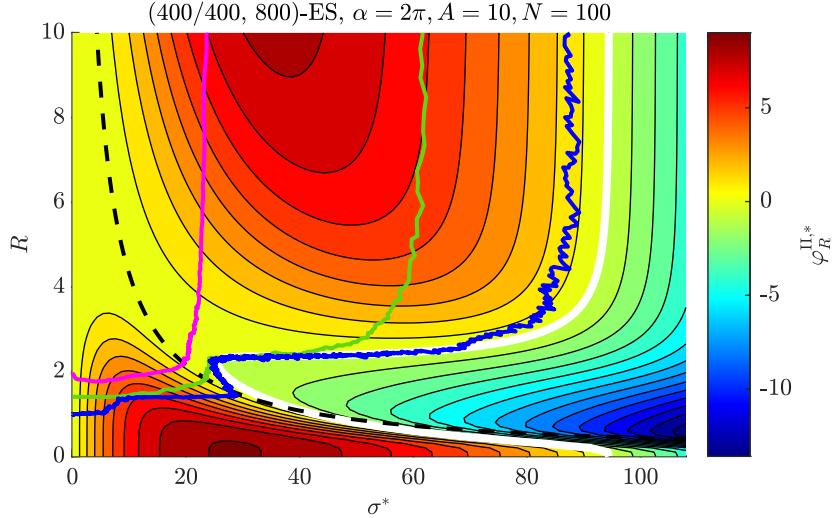


Figure 40: Dynamics for  $(400/400, 800)$ - $\sigma$ SA-ES with  $N = 100$ ,  $A = 10$ , and  $\alpha = 2\pi$ , showing the median of unsuccessful runs (out of 100 trials). The  $\tau$ -values are varied:  $\tau = 1/\sqrt{N}$  (magenta,  $P_S = 0.02$ ),  $\tau = 1/\sqrt{2N}$  (green,  $P_S = 0.06$ ), and  $\tau = 1/\sqrt{8N}$  (blue,  $P_S = 0.33$ ). The dashed black line marks the constant  $\sigma_{esc} \approx 0.436$  according to Table 1.

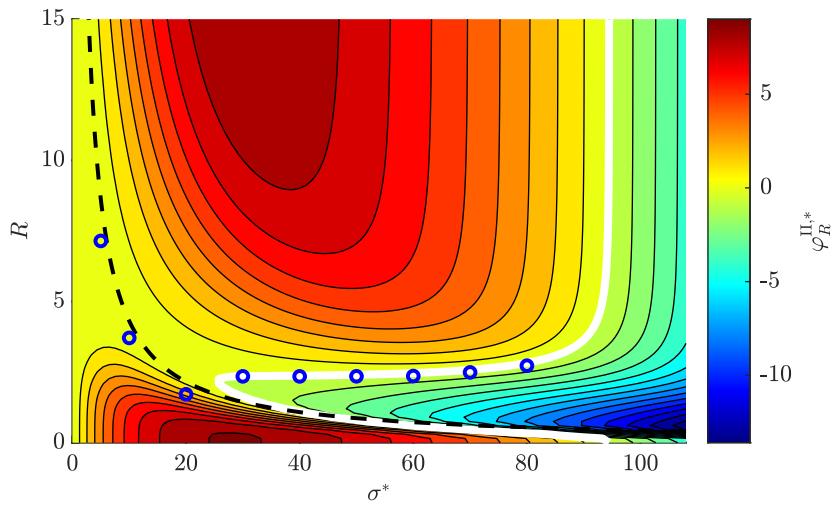


Figure 41: Dynamics for  $(400/400, 800)$ -ES (constant  $\sigma^*$ ) with  $N = 100$ ,  $A = 10$ , and  $\alpha = 2\pi$ . The maximum number of generations was set to  $N_G = 10^4$  and different  $\sigma^*$ -values are tested. The blue dots show the mean of the residual distance of the last 1000 generations, as no local convergence can occur. The dashed black line marks the constant  $\sigma_{esc} \approx 0.436$  according to Table 1.

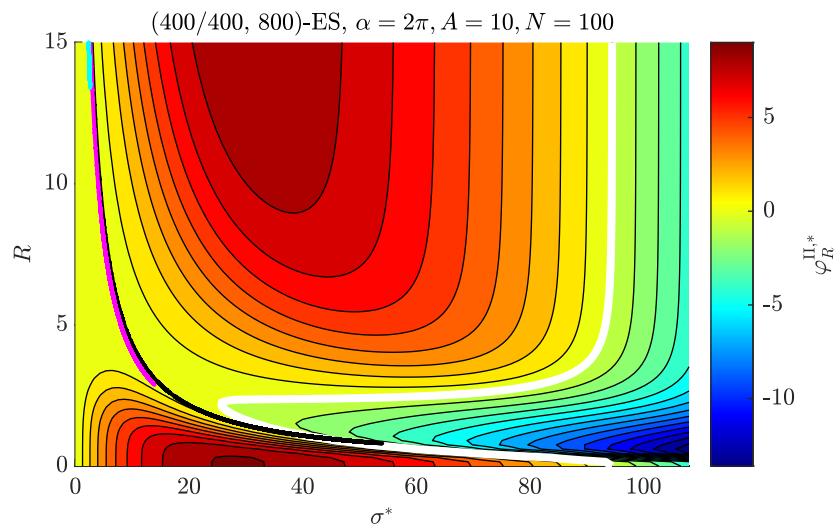


Figure 42: Dynamics for  $(400/400, 800)$ -ES (constant  $\sigma$ ) with  $N = 100$ ,  $A = 10$ , and  $\alpha = 2\pi$ . The maximum number of generations was set to  $N_G = 10^4$  and different constant  $\sigma$ -values are tested. The black line shows  $\sigma = \sigma_{esc} \approx 0.436$ , the magenta line  $\sigma = 0.4$ , and the cyan line  $\sigma = 0.35$ . The distribution of the  $y_i$ -values of the three configurations is shown in Fig. 43.

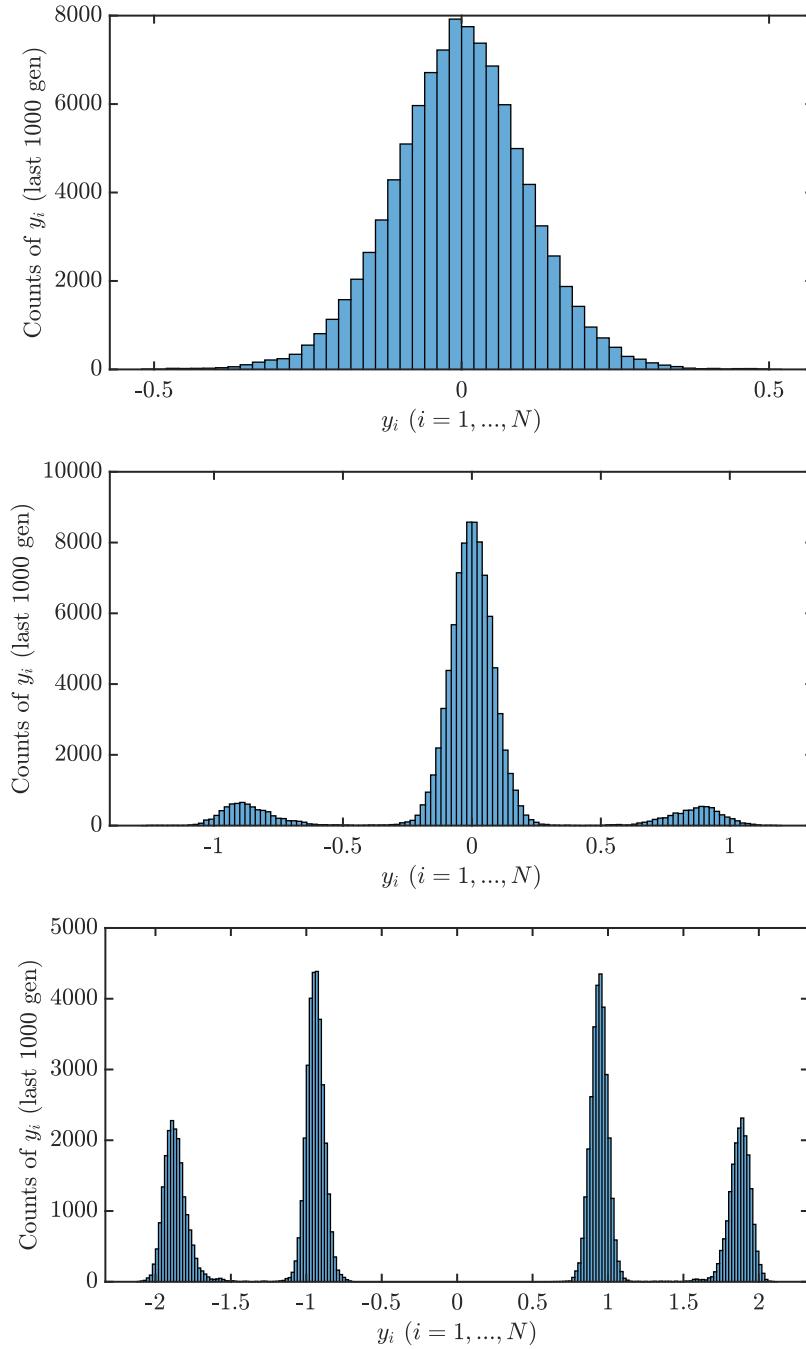


Figure 43: The distribution of  $y_i$ -values (for all  $i$ , last 1000 generations) of the three configurations of Fig. 42 is shown with (top to bottom)  $\sigma = \sigma_{esc} \approx 0.43$ ,  $\sigma = 0.4$ , and  $\sigma = 0.35$ .

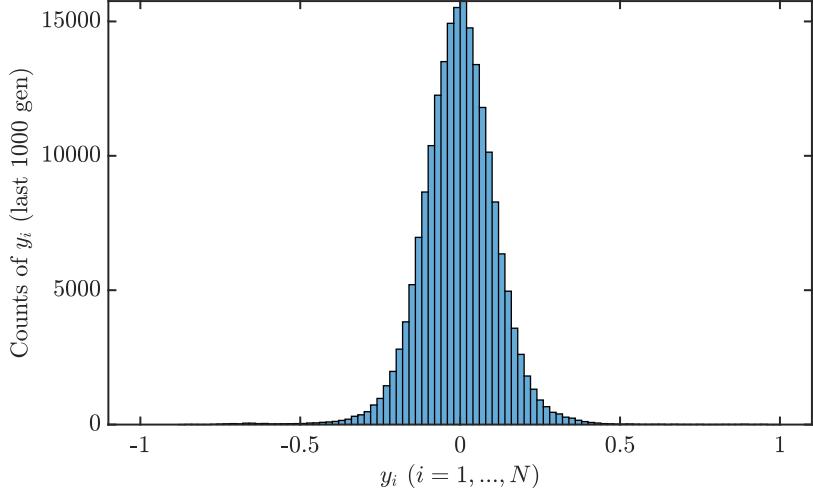


Figure 44: The distribution of  $y_i$ -values (for all  $i$ , last 1000 generations) for (200/200, 600)-ES (constant  $\sigma_{esc} \approx 0.39$ ) with  $N = 200$ ,  $A = 5$ , and  $\alpha = 2\pi$ . The maximum number of generations was set to  $N_G = 10^4$  and the global attractor is reached. A few values (although rare) can be found outside  $[-0.5, 0.5]$ .

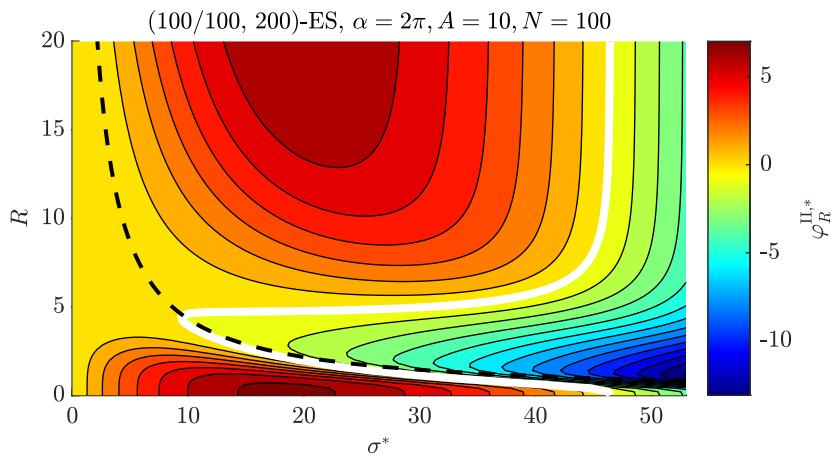


Figure 45: Progress rate zero and stability condition via  $\sigma_{esc}$  from (4.60) intersecting at comparably large  $R \approx R_{\varphi_0}$ . Due to  $\sigma_{esc}$  and noise no further progress is expected to occur for  $R$ -values below the intersection point.

#### 4.3.4 Optimization under constant $\sigma$

In Fig. 42 it was observed that by optimizing the Rastrigin function under constant  $\sigma$  one can approach the global optimizer for roughly  $\sigma \approx \sigma_{esc}$ . For comparably small  $\sigma$ -values however, the ES is jumping between local attractors, see Fig. 43, such that it gets stuck in the fitness landscape at a large residual distance.

This effect is studied by comparing real ES-dynamics (constant  $\sigma$ ) with dynamics obtained by iterating the progress rates  $\varphi_i$  and  $\varphi_i^{II}$  (see also (5.4)). For both real and iterated dynamics, the ES is initialized randomly and far away from the optimizer ( $R = 100 \cdot \sqrt{N}$ ). All runs start at the same initial  $\mathbf{y}$ -coordinates.

Figure 46 shows the results of a (200/200, 600)-ES,  $N = 100$ ,  $\alpha = 2\pi$ , and  $A = 10$  at varying  $\sigma$ -values. Solid lines show the real dynamics (single trial), while the dotted lines show the progress rate  $\varphi_i$ -dynamics. Figure 47 is obtained using the same parameters as Fig. 46, but via progress rate  $\varphi_i^{II}$ . The number of generations is chosen such that a steady state occurs.

A few observations can be made regarding Figs. 46 and 47. For small  $\sigma$  the real dynamics shows practically a steady state behavior, or more precisely, a very slowly converging dynamics. Due to  $\sigma$  being fixed, some progress is observed over time. For large  $\sigma$  the initial descent is comparably fast, following by a noise-type steady state (see fluctuations). The progress rates on the other hand do show a steady state (horizontal line) with vanishing progress. Significant deviations between real dynamics and iteration occur especially for moderate  $\sigma \approx \sigma_{esc} = 0.436$ . In this limit the fluctuations support the escape process, as the  $R$  of real dynamics lies usually below the steady state  $R$  of the iterations.

These observations can be explained by looking at Figs. 48 and 49. Figure 48 shows the progress landscape of  $\varphi_i^{II}(\sigma)$  assuming equipartition for the variance term  $D_Q(R)$  setting  $R = y_i \sqrt{N}$ . Figure 49 shows the same as Fig. 48, but displaying all local minima. The progress rate iterations reach a steady state if (starting from large  $y_i$ ) at any point the zero-progress line is crossed. Finding the expected steady state therefore corresponds to finding the largest  $y_i$  at which  $\varphi_i = 0$  (or  $\varphi_i^{II} = 0$ ). Choosing  $\sigma > \sigma_{esc}$ , see Fig. 48, the iteration reaches the Rastrigin noise floor and stagnates. This was already observed in Fig. 42. The prediction of zero-progress due to the noise floor was already investigated before. An analytic solution can only be obtained for relatively large  $R$  (or  $y_i = R/N$ ) neglecting the exponential functions. For small  $R$  no closed-form solution is available.

Furthermore, the results of  $\varphi_i$  and  $\varphi_i^{II}$  are practically the same for small  $\sigma$ , as the iteration remains far away from the optimizer (negligible loss term). Interestingly,  $\varphi_i$  shows global convergence when  $\sigma > \sigma_{esc}$ . This can be explained by looking at the positive gain conditions of (4.49) and (4.52). Requiring  $G = 0$  is equivalent to  $\tilde{G} = 0$  and  $\tilde{G}$  is the  $y_i$ -dependent gain prefactor of the progress rate  $\varphi_i$ , see (2.82). Furthermore, assuming that the variance  $D_Q(\mathbf{y})$  does not change significantly as a single component is varied, the condition  $\frac{\partial \tilde{G}}{\partial y_i} = 0$  translates to  $\frac{\partial \varphi_i}{\partial y_i} = 0$ . Choosing  $\sigma > \sigma_{esc}$  therefore ensures  $\varphi_i > 0$  for any  $y_i$  and global convergence occurs (as there is no loss term).

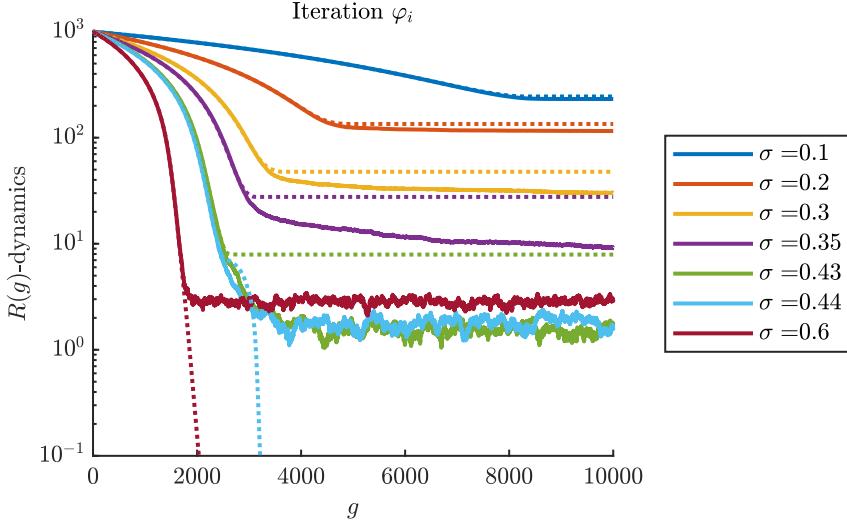


Figure 46: Real dynamics (solid lines showing a single trial, respectively) and iterated  $\varphi_i$ -dynamics (dotted lines) for the (200/200, 600)-ES,  $N = 100$ ,  $\alpha = 2\pi$ , and  $A = 10$  at varying  $\sigma$ -values.

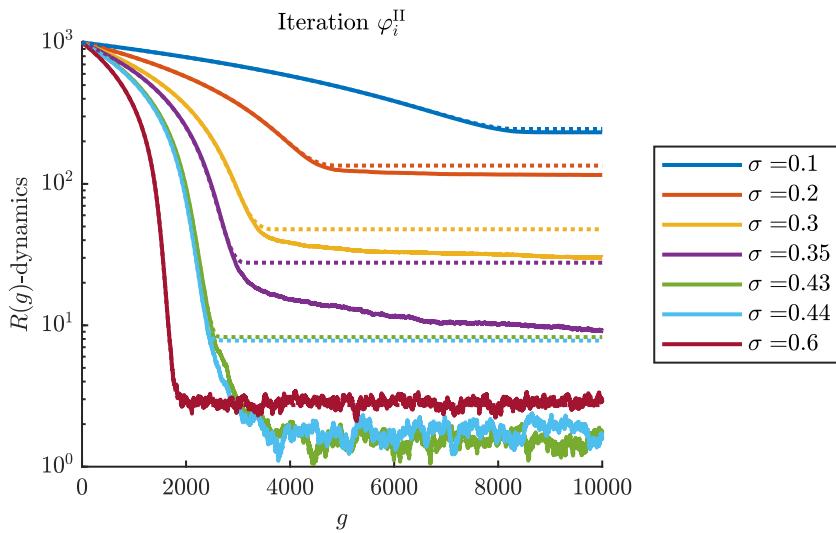


Figure 47: Same configuration as Fig. 46, but with  $\varphi_i^{\text{II}}$ -dynamics shown by the dotted lines.

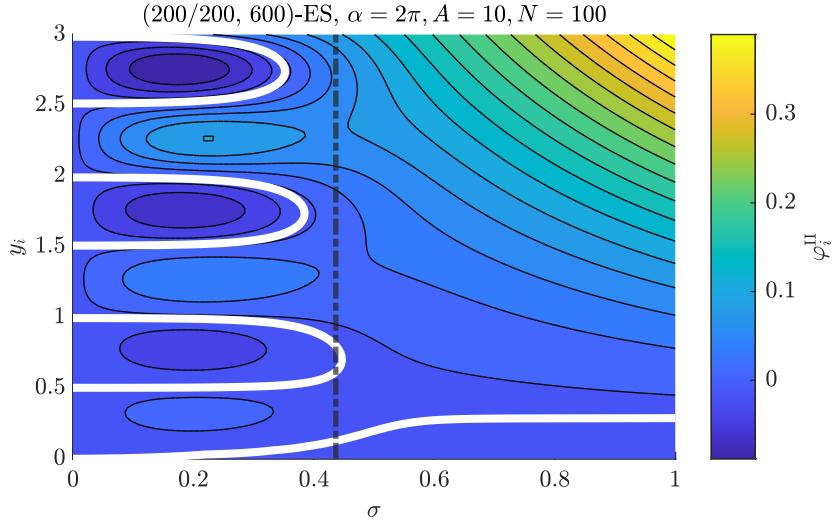


Figure 48: Progress landscape  $\varphi_i^{\text{II}}(\sigma)$  for  $(200/200, 600)$ -ES,  $N = 100$ ,  $\alpha = 2\pi$ , and  $A = 10$ . The zero-progress boundary is shown in bold white. The Rastrigin noise floor is the lowermost white s-shaped line. Due to  $\sigma$  relatively small, and large population  $\mu$ ,  $\sigma_{esc}$  (black dash-dotted line) from Fig. 39 is a good approximation of the escape condition for  $\varphi_i^{\text{II}}$ .

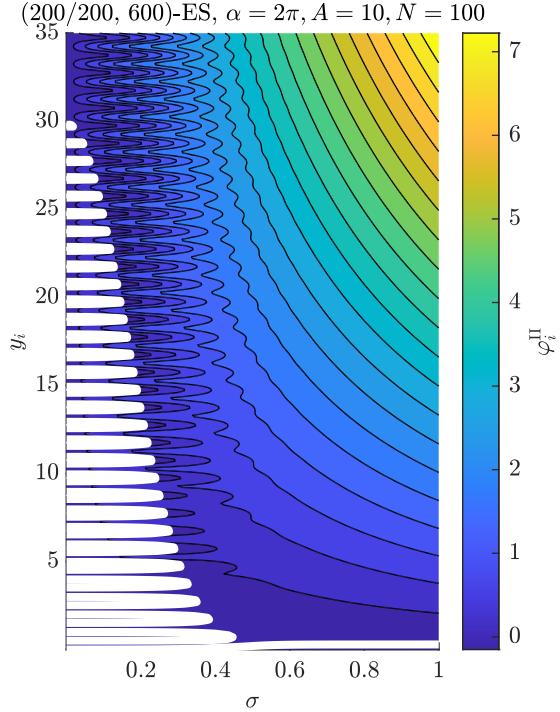


Figure 49: Same configuration as Fig. 48, but showing all local minima attraction basins.

## 4.4 Conclusion

In this section the  $R$ -dependent progress rate  $\varphi_R^{\text{II}}$  of Eq. (4.5) was derived based on the result of the  $N$ -dimensional progress rate  $\varphi_i^{\text{II}}$  from (3.116). The  $R$ -formulation was derived by assuming normally distributed  $y_i$ -values around the optimizer. This approximation is justified for large normalized mutation strengths  $\sigma^*$  (compared to the sphere progress rate), but deteriorates if  $\sigma^*$  becomes too small where local attraction is dominating.

The  $R$ -formulation enables further analytical analysis, as result (4.5) is a two dimensional function of  $R$  and  $\sigma^*$  given the fitness parameters. Furthermore, within this formulation relations to the sphere progress rate can be easily established by setting  $A = 0$  or  $\alpha = 0$ . As a result, an alternative sphere progress rate was derived  $y_i$ -dependent in (4.6), and  $\sigma^*$ -dependent in (4.8). It is more accurate at  $N < \infty$  compared to  $\varphi^* = c_{\mu/\mu,\lambda} \sigma^* - \sigma^{*2}/(2\mu)$  as it includes additional correction terms in the variance.

Having investigated the progress rate landscape in Sec. 4.1, the Rastrigin function shows a large  $A$ - and  $\alpha$ -dependent progress dip (region of negative progress relative to the sphere function). Real  $\sigma$ SA-runs at small  $\tau$ -values, shown in Fig. 23, tend to move close to the negative progress boundary by operating at large  $\sigma^*$ -values. This also justifies the assumptions made for the distribution of  $y_i$ -values. For small  $\sigma^*$ -values on the other hand, no negative progress  $\varphi_R^{\text{II}} < 0$  can be observed independent of fitness or ES-parameters, which indicates the limits of  $\varphi_R^{\text{II}}$  not modeling local attraction (see also example of Fig. 35).

The progress boundary  $\varphi_R^{\text{II}} = 0$  is of most interest, especially the characteristic progress dip, as the ES has to move “around” it in  $R$ - $\sigma^*$ -space. This zero-progress curve can be interpreted as an  $R$ -dependent noise due to the cosine terms in all  $N$ -dimensions. It is therefore a cumulative effect of all oscillation terms present at large mutation strengths. This is in contrast to the local attraction of a single attractor occurring for  $\sigma \rightarrow 0$ . Both effects combined therefore pose a major challenge for the ES, as  $\sigma^*$  has to be reduced when the noise terms start to dominate, which then in turn increases the risk of local convergence. In the limit of constant maximum Rastrigin noise  $\sigma_R = \sqrt{N/2A}$ , the relation to the noisy sphere model was established in (4.32).

A closed form expression  $R(\sigma^*)$  under the condition that  $\varphi_R^{\text{II}} = 0$  can not be obtained due to the terms containing exponential functions, especially in the variance  $D_Q(R)$  from (1.75). An analytic solution of the dip location is therefore not available. However, assuming a constant (maximum value) Rastrigin noise strength and defining a characteristic transition of the exponential function, see (4.33), a solution for the intersection point (4.37) could be obtained as a function of an (unknown) transition parameter  $\delta$ . From this result, a population sizing relation (4.41) with  $\mu = O(\sqrt{N})$  could be deduced. It can be thought of as a “lower bound” scaling relation emerging due to the characteristic progress dip (noise) at a certain transition level parameterized by  $\delta$  (significance of exponential terms). Modeling the actual descent into the global attraction basin likely requires a probabilistic approach, as the progress rate is (only) an expected value without considering fluctuations in  $y_i$ -space.

While the large  $\sigma^*$ -limit of  $\varphi_R^{\text{II}}$  is comparably well understood, as it is dominated by  $R$ -dependent noise and the zero of the sphere progress rate, the effects of local attraction need more investigation. An open question is if there are any

characteristic mutation strengths that enable (statically within one generation, or dynamically) to escape one local attractor and enter a neighboring attractor. A first approach by requiring positive gain for the component-wise  $\varphi_i^{\text{II}}$  was presented in 4.3.3. First results indicate that it is possible to reach the global attractor with an ES operating under constant  $\sigma$ .

## 5 Dynamical Systems Approach

### 5.1 Motivation and Introduction

In order to assess the quality of the quadratic progress rate, one-generation experiments were performed in the previous section given a certain location  $\mathbf{y}$ . Having obtained the progress rate, the next goal is to predict the dynamic behavior (in expectation) of the ES and analyze its convergence properties. To this end, the dynamical systems approach introduced by Beyer [5] is applied. It was also used iterating the dynamics of the Ellipsoid model [6].

Given definition (3.1),  $\sigma^{(g)}$  and  $y_i^{(g)}$  at current generation  $g$ , and assuming  $\varphi_i^{\text{II}}$  is known, one can formulate an iteration  $g \rightarrow g + 1$  according to

$$\left(y_i^{(g+1)}\right)^2 = \left(y_i^{(g)}\right)^2 - \varphi_i^{\text{II}}(\sigma^{(g)}, \mathbf{y}^{(g)}) + \epsilon_i(\sigma^{(g)}, y_i^{(g)}), \quad (5.1)$$

where  $\varphi_i^{\text{II}}$  yields the expected positional difference (by definition) and  $\epsilon_i$  modeling positional fluctuations, which are unknown at this point.

Furthermore, an iteration for the mutations strength  $\sigma^{(g)}$  is needed. As a first approach, one can introduce the spherical normalization  $\sigma^*$  of the mutation strength  $\sigma$  for a residual distance  $R = \|\mathbf{y}\|$ , such that for given  $\sigma^* > 0$  one has

$$\sigma^* = \frac{\sigma}{R} N = \text{const.} \quad (5.2)$$

This ensures that  $\sigma$  decreases simultaneously with  $R$  and global convergence can be modeled. However, the quantity  $\sigma^*$  is only known for theoretical models and not in real world applications. Still, valuable information can be extracted from these models. The control scheme of  $\sigma^{(g)}$  can therefore be given by

$$\sigma^{(g)} = \sigma^* \|\mathbf{y}^{(g)}\| / N = \sigma^* R^{(g)} / N. \quad (5.3)$$

The investigation of step-size adaption schemes such as self-adaption or cumulative step-size adaption is part of future research.

By setting  $\epsilon_i = 0$  in (5.1) as a first step neglecting fluctuations, one arrives at the *deterministic* approximation of the dynamic iteration for the  $i$ -th component

$$\left(y_i^{(g+1)}\right)^2 = \left(y_i^{(g)}\right)^2 - \varphi_i^{\text{II}}(\sigma^{(g)}, \mathbf{y}^{(g)}), \quad (5.4)$$

which is shown in Alg. 1. Considering relation (3.95), the result (5.4) can be expressed as a function of  $R$  by summing over  $i$  according to

$$\sum_{i=1}^N \left(y_i^{(g+1)}\right)^2 = \sum_{i=1}^N \left(y_i^{(g)}\right)^2 - \sum_{i=1}^N \varphi_i^{\text{II}}(\sigma^{(g)}, \mathbf{y}^{(g)}) \quad (5.5)$$

$$\left(R^{(g+1)}\right)^2 = \left(R^{(g)}\right)^2 - \varphi_R^{\text{II}}(\sigma^{(g)}, R^{(g)}), \quad (5.6)$$

where Sec. 4 was used to derive the  $R$ -dependent progress rate  $\varphi_R^{\text{II}}$ .

The iteration schemes will be used for the investigation of the (deterministic) dynamics obtained from different progress rate approximations. The iterated dynamics will also be compared to real simulations with constant  $\sigma^*$ , where fluctuations are of course present.

Two important issues need to be discussed before iterating. Firstly, the scheme (5.4) describes the iteration of a single component  $i$ . For large  $N$  it is not feasible to analyze each component individually and global convergence is achieved for all components vanishing at the same time. While the components will be iterated separately, the dynamics will be presented as a function of the residual distance  $R$ , see also relation (3.95).

Secondly, for the evaluation of  $\varphi_i^{\text{II}}$  being a function of  $\mathbf{y}$ , the square root of the components  $(y_i^{(g)})^2$  has to be taken after iteration giving two solutions  $\pm y_i^{(g)}$ . Interestingly, the result of (3.39) shows that corresponding terms of  $\varphi_i^{\text{II}}$  are even in  $y_i^{(g)}$ . This also holds for the Rastrigin quality variance  $D_Q^2(\mathbf{y})$  in (1.29). This effect is attributed to the progress being second order (quadratic) and the Rastrigin function being symmetric. Therefore the iteration is equivalent for both  $\pm y_i^{(g)}$ , see pseudocode of iteration in Alg. 1.

---

**Algorithm 1** Deterministic component-wise iteration using  $\varphi_i^{\text{II}}$  with  $\sigma^*=\text{const.}$ 


---

```

1:  $g \leftarrow 0$ 
2:  $\mathbf{y}^{(0)} \leftarrow \mathbf{y}^{(\text{init})}$ 
3:  $\sigma^{(0)} \leftarrow \sigma^* \|\mathbf{y}^{(0)}\|/N$ 
4: repeat
5:   for  $i = 1, \dots, N$  do
6:      $(y_i^{(g+1)})^2 \leftarrow (y_i^{(g)})^2 - \varphi_i^{\text{II}}(\sigma^{(g)}, \mathbf{y}^{(g)})$ 
7:      $y_i^{(g+1)} \leftarrow +\sqrt{(y_i^{(g+1)})^2}$            both solutions  $\pm\sqrt{(\cdot)}$  equivalent
8:   end for
9:    $\sigma^{(g+1)} \leftarrow \sigma^* \|\mathbf{y}^{(g+1)}\|/N$ 
10:   $g \leftarrow g + 1$ 
11: until termination criterion

```

---

## 5.2 Experimental Setup

The optimization runs are performed using constant normalized mutation strength  $\sigma^*$  according to Alg. 2. As we are interested in the expected convergence behavior of the algorithm, the residual distance  $R^{(g)}$  is measured and averaged over multiple runs.

**Position initialization** For the initialization,  $\mathbf{y}^{(0)}$  is chosen randomly such that  $\|\mathbf{y}^{(0)}\| = R^{(0)}$  for a given residual distance  $R^{(0)}$ . The starting position is kept constant for consecutive runs of the same experiment. For the magnitude of  $R^{(0)}$  it must be ensured, that the strategy starts outside the local minima landscape. For  $A = 1$  and  $\alpha = 2\pi$ , the default configuration of the following experiments, the last local minimum is located at  $y_i \approx 3$ . A larger reference value  $y_i = 20$  is chosen to be farther away. Including the dimensionality scaling  $R \sim \sqrt{N}$  from (1.79), we therefore set  $R^{(0)} = 20\sqrt{N}$  as the default initialization distance, if not explicitly stated otherwise.

**Mutation initialization** For the choice of  $\sigma^*$  the sphere progress rate by Beyer [5, p. 216, Eq. (6.54)] can be used as a reference by numerically solving to obtain  $\hat{\sigma}_{\text{sph}}^* = \text{argmax}(\varphi_{\text{sph}}^*(\sigma^*))$ . This serves a guideline as the Rastrigin

---

**Algorithm 2**  $(\mu/\mu_I, \lambda)$ -ES with constant  $\sigma^*$ 


---

```

1:  $g \leftarrow 0$ 
2:  $\mathbf{y}^{(0)} \leftarrow \mathbf{y}^{(\text{init})}$ 
3:  $\sigma^{(0)} \leftarrow \sigma^* \|\mathbf{y}^{(0)}\|/N$ 
4: repeat
5:   for  $l = 1, \dots, \lambda$  do
6:      $\tilde{\mathbf{x}}_l \leftarrow \sigma^{(g)} \mathcal{N}_l(0, \mathbf{1})$ 
7:      $\tilde{\mathbf{y}}_l \leftarrow \mathbf{y}^{(g)} + \tilde{\mathbf{x}}_l$ 
8:      $f_l \leftarrow f(\tilde{\mathbf{y}}_l)$ 
9:   end for
10:   $(\tilde{\mathbf{y}}_{1;\lambda}, \dots, \tilde{\mathbf{y}}_{\mu;\lambda}) \leftarrow \text{sort}(\tilde{\mathbf{y}} \text{ w.r.t. ascending } f)$ 
11:   $\mathbf{y}^{(g+1)} \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\mathbf{y}}_{m;\lambda}$ 
12:   $\sigma^{(g+1)} \leftarrow \sigma^* \|\mathbf{y}^{(g+1)}\|/N$ 
13:   $g \leftarrow g + 1$ 
14: until termination criterion

```

---

function is sphere-like for large  $R$  and within the global attractor. However, for a higher  $P_S$  in the experiments a larger value than  $\hat{\sigma}_{\text{sph}}^*$  is needed as larger mutations decrease the probability for local convergence. Therefore  $\sigma^* > \hat{\sigma}_{\text{sph}}^*$  is empirically chosen, such that a high value  $P_S \approx 1$  is obtained (more details in Sec. 5.3).

### 5.2.1 Measures for Averaging

The first issue to discuss is the averaging of the dynamic simulations. More precisely, a measure of central tendency is needed. Experiments have shown that due to fluctuations large run-time differences among converging runs may occur for the same parameter set, see as an example Fig. 50. Multiple averaging approaches are compared against each other.

To simplify the notation, the following definitions are used. The experiment repetitions are denoted with  $j = 1, \dots, M$  and the generation counter with  $g$ , such that abbreviated notation is used as

$$\text{mean}(R) = \underset{g=\text{const}}{\text{mean}}(R_j^{(g)}) \quad (5.7)$$

$$\text{median}(R) = \underset{g=\text{const}}{\text{median}}(R_j^{(g)}). \quad (5.8)$$

**Measure of central tendency for constant  $g$**  One group of measures (denoted by lines in Fig. 50) evaluates the data for fixed  $g$ , such that the measures  $\text{mean}(R)$  and  $\text{median}(R)$  can be easily evaluated. Additionally, the mean of  $\log(R)$  is plotted which corresponds to a “visual averaging” in log-space of the  $R$ -values.

In order to discuss the difference between the standard  $\text{mean}(R)$  and the other measures, see Fig. 50, the distribution of  $R$ -values is plotted in Fig. 51 for  $g = 10$  and in Fig. 52 for  $g = 75$ , respectively. The results indicate that the median is more robust as a statistical measure, if large differences for  $R$  by orders of magnitude are present.

Evaluating  $\text{median}(R)$  has another useful property. We have  $R > 0$  and  $\log(R)$  is a monotonic transformation thereof. For an odd number of elements the median is the central value from the corresponding sorted list. Therefore  $\log(\text{median}(R))$  yields the same as  $\text{median}(\log(R))$ , which makes it suitable for evaluation in linear and logarithmic space. For an even number of elements the median is the average of the two central values, such that the relation above is not exact. Still, for a large number of runs the central values are expected to be very close to each other such that the deviations are negligible.

**Measure of central tendency for constant  $R$**  The second group of measures (denoted by dots in Fig. 50) is an approach to average for fixed  $R$  values. Given a certain range  $[R, R \cdot \delta R]$ , the respective generation values are collected for all trials and evaluated using mean and median, respectively. The constant factor  $\delta R$  is used to obtain evenly spaced points in log-space.

The results are displayed at the interval center point and show good agreement with averaging methods for constant  $g$ . The larger deviations at the descent point are due to high influence of present fluctuations.

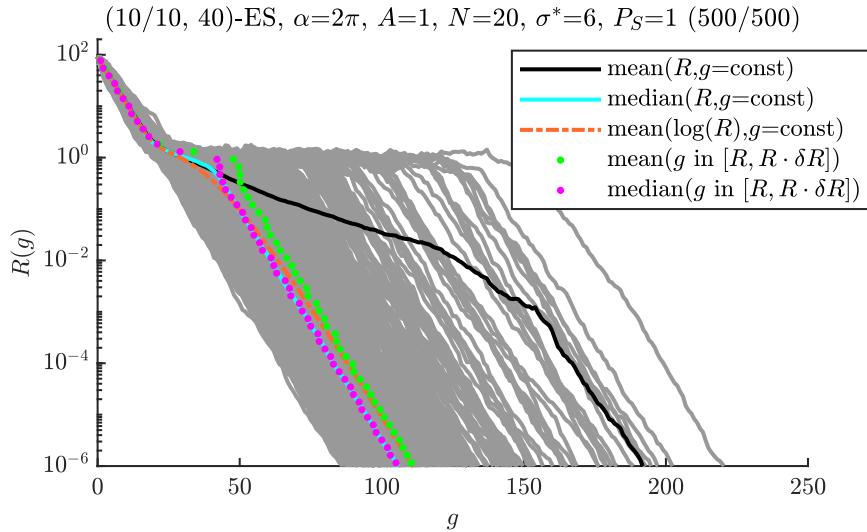


Figure 50: Different averaging approaches for the dynamics of 500 runs with success probability  $P_S = 1$ . The simulation parameters are in the title. Mean( $R$ ) with linear  $R$ -values is heavily skewed for moderate  $g$ -values due to different orders of magnitudes of  $R$ . All other measures yield comparably similar results. Slightly larger deviations occur before descending into the global attractor basin, where fluctuation effects are strong.

**Concluding averaging discussion** From Figures 50, 51, and 52 it was established that standard linear averaging of  $R$ -values is not suitable for dynamical simulations with moderate to large run-time differences, which occurs on the Rastrigin function. This is due to the fact that large outliers create a skewed

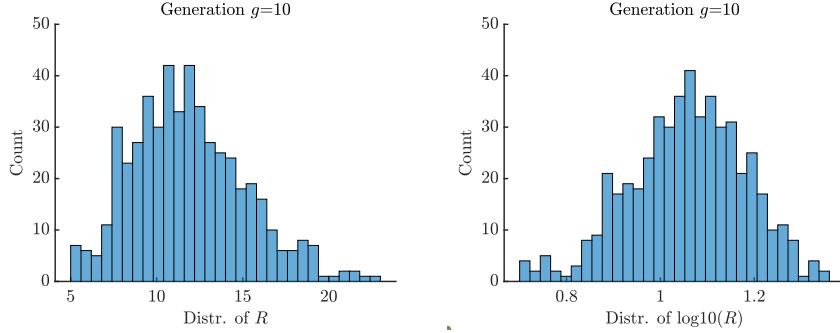


Figure 51: At an early stage with  $g = 10$  the distributions of linear and logarithmic  $R$ -values are relatively symmetric and not heavily skewed. The measures  $\text{mean}(R)$ ,  $\text{median}(R)$ , and  $\text{mean}(\log(R))$  therefore yield comparable results.

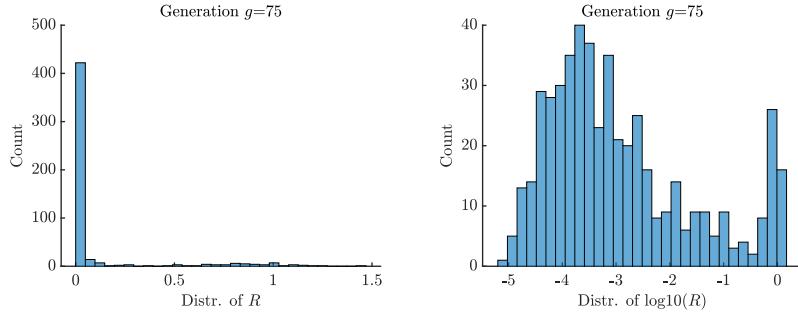


Figure 52: For moderately large  $g = 75$  the distributions of linear and logarithmic  $R$ -values are completely different, as the simulations are at different stages having  $R$ -values that are deviating in orders of magnitude. The mean calculated for the left distribution will therefore be highly influenced by very few large values (in relation), which is observed in the dynamics of Fig. 50, while the median as a central value is more robust w.r.t. outliers. The right plot shows the distribution of exponents of  $R$  by taking the logarithm.

distribution, such that large deviations between mean and median values are observed.

The median is significantly more robust with respect to outliers, as it represents the central value. The property of the median being invariant under the monotonic transformation  $R \mapsto \log(R)$  is also useful. Therefore the median will be displayed when comparing the real simulation runs with iterated (expected) dynamics.

The averaging using constant intervals  $[R, R + \delta R]$  was also shown in Fig. 50, but it will not be displayed for the upcoming dynamics, as no additional value is provided by the method at this point. It also requires an additional  $R$ -discretization, which is not needed for averaging via constant  $g$ .

### 5.3 Comparing Real Simulations with Iteration

In this section the median of real simulation runs will be compared to iterated dynamics. Two deterministic iteration schemes are used with following underlying progress rate evaluations

- Iteration SIM:  $\varphi_i^{\text{II}}$  by simulating (3.1) over  $10^4$  runs for each  $g$  and iterating (5.4); serves as a reference deterministic iteration
- Iteration Y:  $\varphi_i^{\text{II}}$  using (3.116) with iteration (5.4)
- Iteration R:  $\varphi_R^{\text{II}}$  (4.5) with iteration (5.6)

The investigations are done for high success probability runs close to one and lower success probabilities to test the limitations of the approach.

#### 5.3.1 High Success Probability Investigation

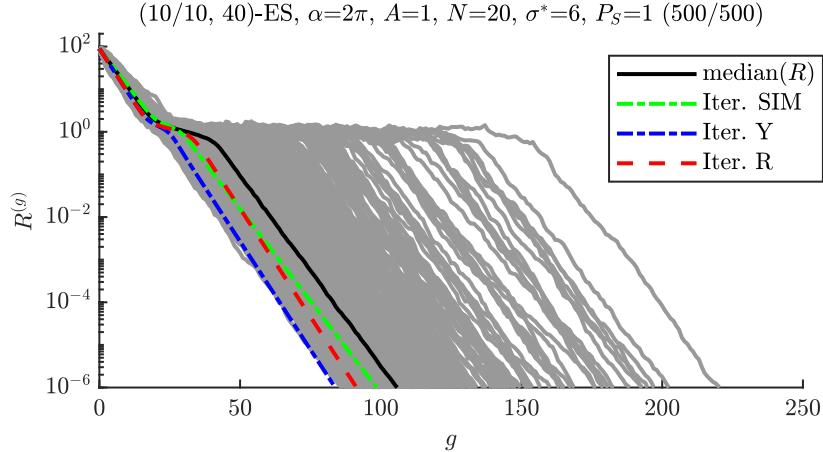


Figure 53: Median dynamics compared to iterated dynamics using three progress rate calculations. Parameters are given in the title. Approximations Y and R show the same convergence speed for large and small  $R$  (parallel lines). The offset emerges during the transition at  $R \approx 1$ . Iterations Y and R slightly overestimate the convergence speed compared to the median and SIM. The offset between median and SIM is due to fluctuations, as the accuracy of simulated  $\varphi_i^{\text{II}}$  is expected to be very high. Differences between SIM and iterations Y and R, respectively, are due to the underlying approximations. The agreement is relatively good considering small  $N$  and small populations with lower truncation ratio  $\vartheta = 0.25$ .

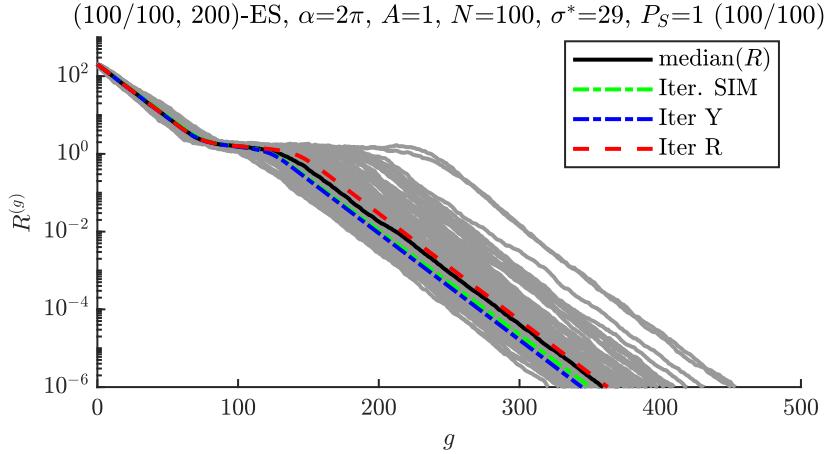


Figure 54: Median dynamics compared to iterated dynamics for  $N = 100$  with population and  $\sigma^*$  set accordingly. For larger  $N$  and population size better agreement is observed between the median and the iterations. Furthermore, the difference between the iterations tends to decrease, as the approximation quality increases for larger  $N$ . This is also shown in Fig. 55. Again, an offset between Y and R emerges before the descent at  $R \approx 1$  due to the underlying approximations.

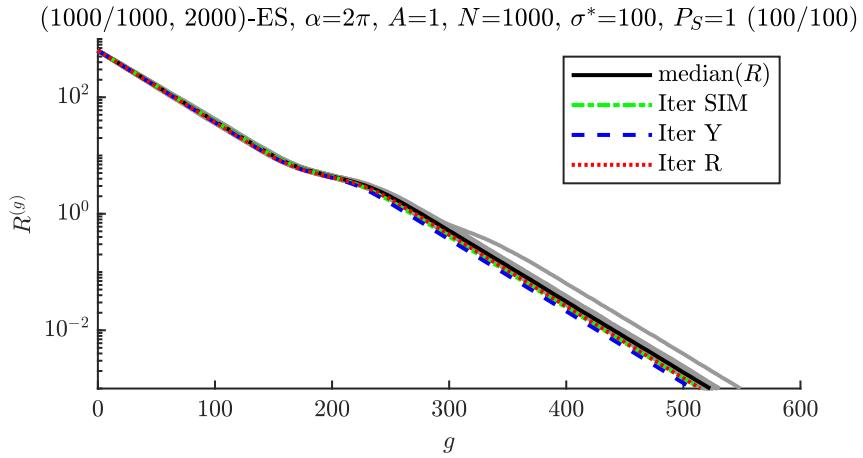


Figure 55: The dynamics is shown for large dimensionality  $N = 1000$  with population and  $\sigma^*$  set accordingly. Within this limit fluctuations are decreasing (loss term scaling as  $1/\mu$ ) and different runs tend to follow the same path. Deviations between iterations R and Y are decreasing compared to Fig. 54 as the approximation quality increases.

### 5.3.2 Lower Success Probability Investigation

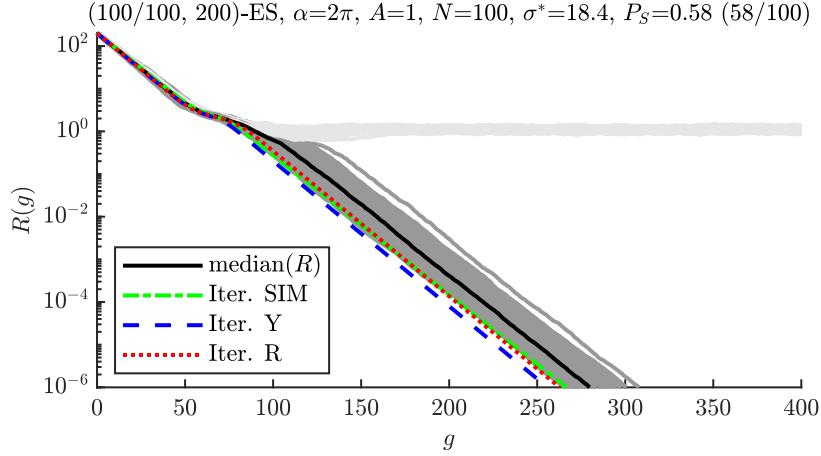


Figure 56: The experiment of Fig. 54 is repeated with decreased  $\sigma^*$ . Now  $\sigma^* \approx 18.4$  is chosen to be very close to the sphere optimal value by numerically solving [5, p. 216, Eq. (6.54)]. All iterations now lie below the median curve and are overestimating the overall progress, as fluctuations are not modeled. The median is taken over the globally converging runs. A relatively sharp splitting point occurs between locally (light gray) and globally (gray) converging runs. This example illustrates the problem of smaller mutation strengths on the Rastrigin function and why  $\sigma^* = 29$  was chosen in Fig. 54.

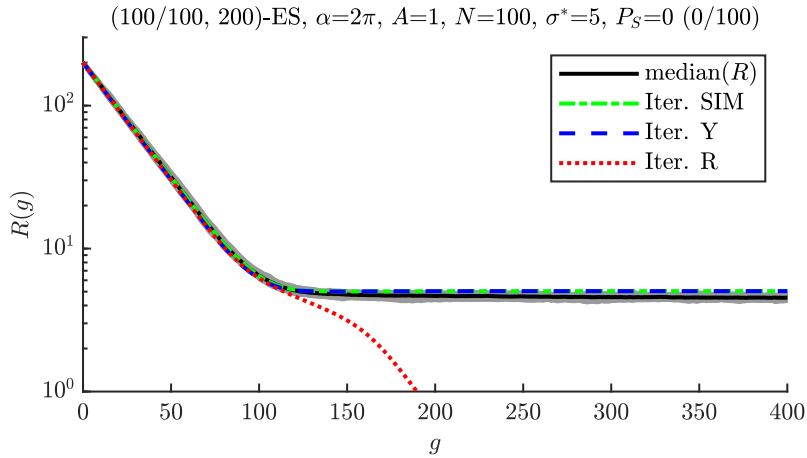


Figure 57: The value of  $\sigma^*$  is further reduced compared to Fig. 56 and no global convergence occurs. Iterations SIM and Y correctly predict a stagnation point around  $R \approx 5$ , while R predicts global convergence. This is due to the  $R$ -dependent model being an average over all locations with  $\|\mathbf{y}\| = R$ , while SIM and Y are actual  $\mathbf{y}$ -dependent models. The issue is explained in Fig. 58.

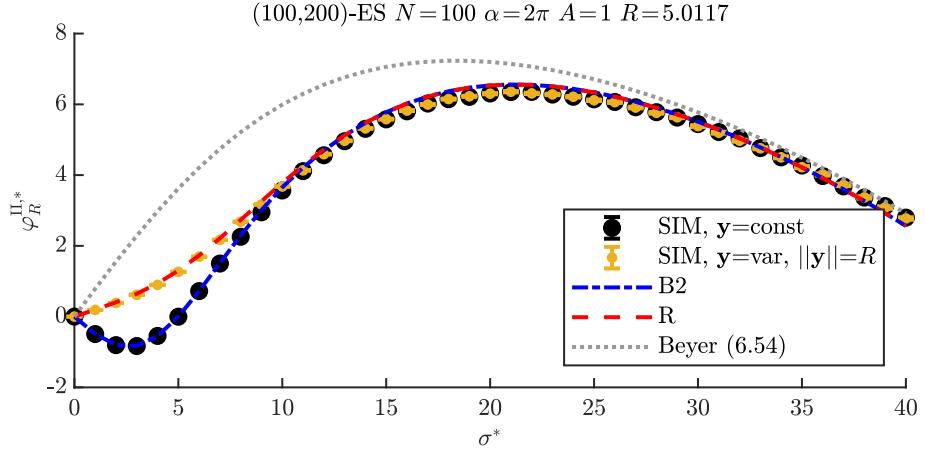


Figure 58: The difference between Y and R from Fig. 57 is explained by looking at  $\varphi_R^{II,*}(\sigma^* = 5)$ . Iteration Y in Fig. 57 was stopped the first time negative progress  $\varphi_R^{\text{II}} = \sum_i \varphi_i^{\text{II}} < 0$  is observed, which triggers at  $g = 146$ . At this point the Cartesian coordinates  $\mathbf{y}$  are saved and  $\varphi_R^{II,*}$  is evaluated to study the difference (approximation B2 corresponds to iteration Y). Between red (R) and blue (B2) there is a gap at  $\sigma^* = 5$  with B2 showing (small) negative progress and R positive progress. This can be attributed to the positional dependence of the progress rates. The black dots show  $\varphi_R^{II,*}$  for one-generation experiments with constant  $\mathbf{y}$  (trigger point), where good agreement with B2 is observed. The orange dots show  $\varphi_R^{II,*}$  for the  $R$ -value of the trigger point, but for each trial  $\mathbf{y}$  newly chosen with  $\|\mathbf{y}\| = R$ , where good agreement with R is seen. Therefore R is not able to reproduce the local convergence behavior observed with Y, where the actual  $\mathbf{y}$ -values are iterated. But it is a good approximation when evaluating the progress over  $\|\mathbf{y}\| = R$ . For small  $\sigma^*$  significant deviations occur compared to Beyer (6.54), while for larger  $\sigma^*$  the sphere-like structure is recovered. This also explains the better agreement between SIM, Y, and R in Fig. 54 with larger  $\sigma^* = 29$ .

#### 5.4 Component Equipartition

During the iteration of squared locations via (5.4) it could be experimentally observed that all the components converge to a value  $(y_i^{(g)})^2 = (R^{(g)})^2/N$  for  $g$  and  $\sigma^*$  sufficiently large. This was observed independent of the initial value provided that there is linear convergence of the overall residual distance. This observation, which is referred to as ‘‘component equipartition’’, is investigated here in more detail. The effect happens on both Rastrigin and sphere function, such that the analysis here is performed on the sphere by investigating the difference equations.

For better readability of the subsequent derivations the following abbreviations are introduced

$$\begin{aligned} y_g &:= (y_i^{(g)})^2 \\ r_g &:= (R^{(g)})^2 \\ \varphi^* &:= \varphi_R^{II,*}. \end{aligned} \tag{5.9}$$

Recalling the deterministic iteration (5.4) of a single component one has

$$y_{g+1} = y_g - \varphi_i^{\text{II},(g)}. \quad (5.10)$$

The  $i$ -th component progress rate of the sphere function for constant  $\sigma^*$  at generation  $g$  was given in (4.6). By applying  $\sigma^2 = \sigma^{*2}r_g/N^2$  one gets

$$\varphi_i^{\text{II},(g)} = \frac{2c_\vartheta\sigma^*}{N\sqrt{1+\sigma^{*2}/2N}}y_g - \frac{\sigma^{*2}}{N^2\mu}r_g. \quad (5.11)$$

Now gain and loss prefactors (constant during iteration) are defined as follows

$$\beta := \frac{2c_\vartheta\sigma^*}{N\sqrt{1+\sigma^{*2}/2N}} \quad (5.12)$$

$$\gamma := \frac{\sigma^{*2}}{N^2\mu}, \quad (5.13)$$

such that

$$\varphi_i^{\text{II},(g)} = \beta y_g - \gamma r_g. \quad (5.14)$$

Hence, Eq. (5.10) becomes

$$y_{g+1} = y_g - \beta y_g + \gamma r_g. \quad (5.15)$$

Summing (5.11) over all  $N$  components, multiplying by  $1/r_g$ , and using normalization  $\frac{\varphi_R^{\text{II}}}{r_g} = \frac{2}{N}\varphi^*$  yields the  $r_g$ -dependent progress rate of the sphere in terms of  $\beta$  and  $\gamma$

$$\begin{aligned} \sum_{i=1}^N \varphi_i^{\text{II},(g)} &= \varphi_R^{\text{II},(g)} = \frac{2c_\vartheta\sigma^*}{N\sqrt{1+\sigma^{*2}/2N}}r_g - \frac{\sigma^{*2}}{N\mu}r_g \\ \frac{2}{N}\varphi^* &= \beta - N\gamma, \end{aligned} \quad (5.16)$$

which will be needed later for solving the difference equations.

Returning to (5.15) one has

$$y_{g+1} = (1 - \beta)y_g + \gamma r_g. \quad (5.17)$$

An expression for the residual distance  $r_g$  is needed. The analysis of the  $y_g$ -dynamics is performed under the condition that there is convergence of the  $r_g$ -dynamics. Using iteration (5.6), normalization  $\varphi^* = \varphi_R^{\text{II}} \frac{N}{2r_g}$ , and abbreviation  $\varphi_R^{\text{II},*} = \varphi^*$ , one can write

$$\begin{aligned} r_{g+1} &= r_g - \varphi_R^{\text{II},(g)} \\ &= r_g - \frac{2r_g}{N}\varphi^* \\ &= r_g \left(1 - \frac{2}{N}\varphi^*\right). \end{aligned} \quad (5.18)$$

Assuming constant  $\varphi^* > 0$ , the dynamics of Eq. (5.18) given initial value  $r_0$  can be easily evaluated as

$$r_g = r_0 \left( 1 - \frac{2}{N} \varphi^* \right)^g. \quad (5.19)$$

Inserting (5.19) into Eq. (5.17), one has a first order linear difference equation in variable  $y_g$  with constant coefficient  $a := (1 - \beta)$  and a generation-dependent term  $b_g := \gamma r_g$ , which couples to the  $y_g$ -evolution. The linear difference equation is now written as

$$y_{g+1} = ay_g + b_g. \quad (5.20)$$

Starting from  $y_0$  and repeatedly evaluating expression (5.20), the value at  $g$  can be given as

$$y_g = a^g y_0 + \sum_{k=1}^g a^{g-k} b_{k-1}. \quad (5.21)$$

This can be verified by inserting (5.21) into (5.20) giving

$$\begin{aligned} ay_g + b_g &= a \left( a^g y_0 + \sum_{k=1}^g a^{g-k} b_{k-1} \right) + b_g \\ &= a^{g+1} y_0 + \sum_{k=1}^g a^{g-k+1} b_{k-1} + a^0 b_g \\ &= a^{g+1} y_0 + \sum_{k=1}^{g+1} a^{g+1-k} b_{k-1} \\ &= y_{g+1}. \end{aligned} \quad (5.22)$$

Therefore (5.21) is the generic solution to iteration scheme (5.17), which yields

$$y_g = (1 - \beta)^g y_0 + \gamma \sum_{k=1}^g (1 - \beta)^{g-k} r_{k-1}. \quad (5.23)$$

The term  $r_{k-1}$  needs to be rewritten using (5.19) according to

$$\begin{aligned} r_g &= r_k \left( 1 - \frac{2}{N} \varphi^* \right)^{g-k}, \quad \text{such that} \\ r_{k-1} &= r_g \left( 1 - \frac{2}{N} \varphi^* \right)^{k-1-g} = r_0 \left( 1 - \frac{2}{N} \varphi^* \right)^{k-1}. \end{aligned} \quad (5.24)$$

Inserting (5.24) into (5.23) and moving all  $k$ -independent terms out of the sum yields

$$y_g = (1 - \beta)^g y_0 + \gamma r_0 \frac{(1 - \beta)^g}{\left( 1 - \frac{2}{N} \varphi^* \right)} \sum_{k=1}^g \left( \frac{1 - \frac{2}{N} \varphi^*}{1 - \beta} \right)^k. \quad (5.25)$$

The sum in Eq. (5.25) can be evaluated using the formula for the  $n$  first terms of a geometric series with  $q := \frac{1-\frac{2}{N}\varphi^*}{1-\beta}$  according to

$$\begin{aligned} \sum_{k=1}^n q^{k-1} &= 1 + q^1 + \dots + q^{n-1} = \frac{1 - q^n}{1 - q}, \quad \text{such that} \\ q^1 + \dots + q^g &= \frac{1 - q^{g+1}}{1 - q} - 1 \\ &= \frac{q - q^{g+1}}{1 - q}, \end{aligned} \quad (5.26)$$

under the condition that  $q \neq 1$ , see denominator in Eq. (5.28).

First, evaluating the numerator  $q - q^{g+1}$  of (5.26) yields

$$q - q^{g+1} = q(1 - q^g) = \frac{1 - \frac{2}{N}\varphi^*}{1 - \beta} \left[ 1 - \left( \frac{1 - \frac{2}{N}\varphi^*}{1 - \beta} \right)^g \right]. \quad (5.27)$$

Evaluating the denominator  $1 - q$  of (5.26) using  $\frac{2}{N}\varphi^* = \beta - N\gamma$  from (5.16) gives

$$1 - q = 1 - \frac{1 - \frac{2}{N}\varphi^*}{1 - \beta} = \frac{1 - \beta - (1 - \frac{2}{N}\varphi^*)}{1 - \beta} = -\frac{N\gamma}{1 - \beta}, \quad (5.28)$$

which is well defined for  $\beta \neq 1$ , see also Eq. (5.32).

The result of the sum can be given as

$$\frac{q - q^{g+1}}{1 - q} = -\frac{1 - \frac{2}{N}\varphi^*}{N\gamma} \left[ 1 - \left( \frac{1 - \frac{2}{N}\varphi^*}{1 - \beta} \right)^g \right]. \quad (5.29)$$

Substituting the sum in (5.25) with result (5.29) one gets

$$\begin{aligned} y_g &= (1 - \beta)^g y_0 - \frac{r_0}{N}(1 - \beta)^g \left[ 1 - \left( \frac{1 - \frac{2}{N}\varphi^*}{1 - \beta} \right)^g \right] \\ &= (1 - \beta)^g \left( y_0 - \frac{r_0}{N} \right) + \frac{r_0 (1 - \frac{2}{N}\varphi^*)^g}{N}. \end{aligned} \quad (5.30)$$

Using  $r_g = r_0(1 - 2\varphi^*/N)^g$  and resolving abbreviations defined in (5.9), the final result yields

$$(y_i^{(g)})^2 = (1 - \beta)^g \left( (y_i^{(0)})^2 - \frac{(R^{(0)})^2}{N} \right) + \frac{(R^{(g)})^2}{N}. \quad (5.31)$$

One can immediately see from result (5.31) that the initial values  $(y_i^{(0)})^2$  and  $(R^{(0)})^2$  are decaying for  $g > 0$  as long as  $0 < \beta < 1$ . For  $g$  large enough the initial values are suppressed and the term  $(R^{(g)})^2/N$  is dominating. Therefore within the deterministic iteration any (squared) component  $(y_i^{(g)})^2$  will approach the value  $(R^{(g)})^2/N$  for  $g$  large enough. At  $g = 0$  the initial value  $(y_i^{(0)})^2$  is recovered.

Figure 59 shows an example where the iteration is compared to analytic solutions of the difference equations. As  $\sigma^*$  is relatively small, the approaching

of components to the value  $(R^{(g)})^2/N$  is relatively slow (see generation axis). There is no (visible) difference between the two calculations, as the equations are solved analytically.

Figure 60 shows the same experimental setup of Fig. 59 but with increased  $\sigma^* = 20$ . Now the adaption time is much faster (see generation axis) due to larger  $\beta$ , see (5.12), and both components quickly approach the target value  $(R^{(g)})^2/N$ .

Additional investigation is needed concerning the gain parameter  $\beta$  occurring in solution (5.31). In order to have convergence it must hold

$$0 < \beta < 1. \quad (5.32)$$

The lower bound is satisfied for any  $\sigma^* > 0$  with

$$0 < \frac{2c_\vartheta\sigma^*}{N\sqrt{1+\sigma^{*2}/2N}}. \quad (5.33)$$

The upper bound is given by (assuming  $\sigma^* \neq 0$ )

$$\begin{aligned} \frac{2c_\vartheta\sigma^*}{N\sqrt{1+\sigma^{*2}/2N}} &< 1 \\ 4c_\vartheta^2\sigma^{*2} &< N^2(1 + \sigma^{*2}/2N) \\ 4c_\vartheta^2\sigma^{*2} - \frac{N\sigma^{*2}}{2} &< N^2 \\ 4c_\vartheta^2 - \frac{N}{2} &< \frac{N^2}{\sigma^{*2}}. \end{aligned} \quad (5.34)$$

Equation (5.34) is satisfied for most practical cases with moderate to large  $N$ . The right side is strictly positive. Additionally, the (asymptotic) progress coefficient is usually a relatively small number for moderate  $\vartheta$ , such that  $4c_\vartheta^2 < \frac{N}{2}$  yields negative values on the left side.

**Component equipartition on Rastrigin** The argumentation made on the sphere function can also be transferred to the Rastrigin function, as long as one has negligible influence of the local attractors. As the iteration experiments are usually initialized far away from the optimizer, equipartition is established by the time the iteration reaches the local attractors.

The argumentation can also be illustrated by looking at component-wise progress  $\varphi_i^{\text{II}}$  (3.116) and comparing it with Eq. (5.16). As long as the gain factor  $\beta > 0$  (the loss terms are equal) and  $R^{(g)}$ -convergence is present over many generations, equipartition can be observed, which is the case for Rastrigin within the spherical limits of negligible local attractors.

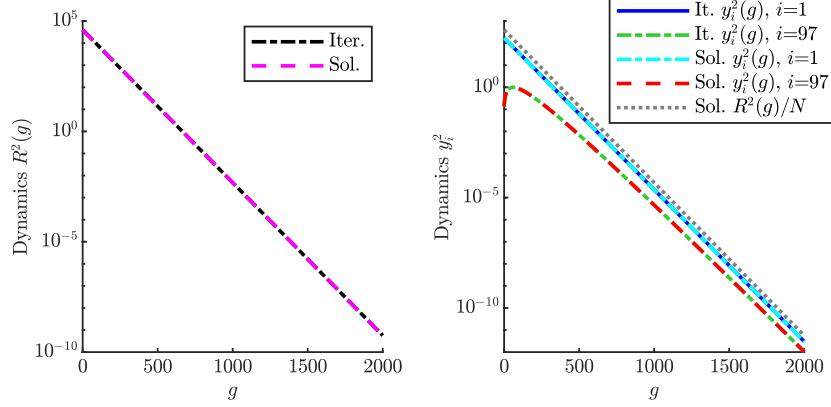


Figure 59: Iteration compared to analytic solution of difference equations for (100/100, 200)-ES on the sphere function with  $N = 100$  at  $\sigma^* = 1$ . The left plot shows the  $(R^{(g)})^2$ -iteration and solution (5.19). The right plot shows the  $(y_i^{(g)})^2$ -iteration and solution (5.31) for two components  $i = 1$  and  $i = 97$ . The latter component was chosen as its initial value is relatively small, such that one can observe the initialization dynamics. The value  $\sigma^* = 1$  yields relatively small value  $\beta \approx 0.016$  giving slow adaption compared to Fig. 60.

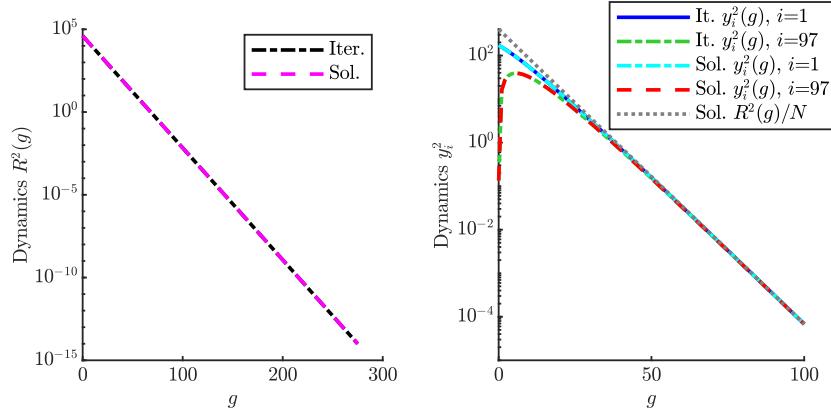


Figure 60: Iteration compared to analytic solution of difference equations for (100/100, 200)-ES on the sphere function with  $N = 100$  at larger  $\sigma^* = 20$ . The left plot shows the  $(R^{(g)})^2$ -iteration and solution (5.19). The right plot shows the  $(y_i^{(g)})^2$ -iteration and solution (5.31). The larger mutation strength yields  $\beta \approx 0.18$  and therefore faster adaption.

## Appendix A Expectation Values

### A.1 Expectation Value of $\cos(\alpha x)$

For a normally distributed random variable  $x \sim \mathcal{N}(\mu, \sigma^2)$ , the expectation value of a function  $g(x)$  is given by

$$E[g(x)] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} g(x) e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx. \quad (\text{A.1})$$

The characteristic function  $\varphi_Y(t)$  of a random variable  $Y$  with density  $p_Y(y)$  is defined as

$$\varphi_Y(t) = E[e^{ity}] = \int_{-\infty}^{\infty} e^{ity} p_Y(y) dy. \quad (\text{A.2})$$

The expectation value of  $\sin(\alpha x)$  and  $\cos(\alpha x)$  can be obtained using the definition of the complex exponential  $e^{i\alpha x} = \cos(\alpha x) + i \sin(\alpha x)$  according to

$$E[e^{i\alpha x}] = E[\cos(\alpha x)] + i E[\sin(\alpha x)]. \quad (\text{A.3})$$

Calculating the expectation of  $e^{i\alpha x}$  using definition (A.1) one obtains

$$E[e^{i\alpha x}] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{i\alpha x} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx. \quad (\text{A.4})$$

Given integral can be identified as the definition of the characteristic function of a normally distributed variable [7, p. 160]

$$E[e^{i\alpha x}] = \varphi_{N(\mu, \sigma^2)}(\alpha) = e^{i\alpha\mu - \frac{1}{2}\alpha^2\sigma^2} \quad (\text{A.5})$$

$$= e^{-\frac{1}{2}\alpha^2\sigma^2} [\cos(\alpha\mu) + i \sin(\alpha\mu)]. \quad (\text{A.6})$$

Using (A.3) the expectation values read

$$E[\cos(\alpha x)] = e^{-\frac{1}{2}\alpha^2\sigma^2} \cos(\alpha\mu) \quad (\text{A.7})$$

$$E[\sin(\alpha x)] = e^{-\frac{1}{2}\alpha^2\sigma^2} \sin(\alpha\mu). \quad (\text{A.8})$$

Since mutations of the ES are distributed according to  $x \sim \mathcal{N}(0, \sigma^2)$ , we obtain simplified equations

$$E[\cos(\alpha x)] \stackrel{\mu=0}{=} \exp\left[-\frac{1}{2}(\alpha\sigma)^2\right] \quad (\text{A.9})$$

$$E[\sin(\alpha x)] \stackrel{\mu=0}{=} 0. \quad (\text{A.10})$$

## A.2 Expectation Value of $\cos^2(\alpha x)$ and $\sin^2(\alpha x)$

These expectation values can be easily simplified by applying trigonometric identities and  $\cos^2(x) + \sin^2(x) = 1$  and using the previous results for  $E[\cos(\alpha x)]$  with a scaled constant. We obtain

$$E[\cos^2(\alpha x)] = E\left[\frac{1}{2}[1 + \cos(2\alpha x)]\right] = \frac{1}{2}\left(1 + e^{\frac{(2\alpha\sigma)^2}{2}} \cos(2\alpha\mu)\right) \quad (\text{A.11})$$

$$\stackrel{\mu=0}{=} \frac{1}{2}\left(1 + \exp\left[-\frac{1}{2}(2\alpha\sigma)^2\right]\right) \quad (\text{A.12})$$

and analogously

$$E[\sin^2(\alpha x)] = E\left[\frac{1}{2}[1 - \cos(2\alpha x)]\right] = \frac{1}{2}\left(1 - e^{\frac{(2\alpha\sigma)^2}{2}} \cos(2\alpha\mu)\right) \quad (\text{A.13})$$

$$\stackrel{\mu=0}{=} \frac{1}{2}\left(1 - \exp\left[-\frac{1}{2}(2\alpha\sigma)^2\right]\right). \quad (\text{A.14})$$

## A.3 Expectation Values of $x \cos(\alpha x)$ and $x \sin(\alpha x)$

The quantities  $E[x \cos(\alpha x)]$  and  $E[x \sin(\alpha x)]$  can be deduced by differentiating the expression (A.6) on both sides with respect to  $\alpha$  and identifying real and imaginary parts. The left-hand side yields

$$\begin{aligned} \frac{d}{d\alpha} E[e^{i\alpha x}] &= E[ixe^{i\alpha x}] \\ &= E[ix(\cos(\alpha x) + i \sin(\alpha x))] \\ &= i E[x \cos(\alpha x)] - E[x \sin(\alpha x)]. \end{aligned} \quad (\text{A.15})$$

Evaluating the derivative of the right-hand side gives

$$\begin{aligned} \frac{d}{d\alpha} \left[ e^{-\frac{1}{2}\alpha^2\sigma^2} [\cos(\alpha\mu) + i \sin(\alpha\mu)] \right] &= \dots \\ &= -\mu e^{-\frac{1}{2}(\alpha\sigma)^2} \sin(\alpha\mu) - \alpha\sigma^2 e^{-\frac{1}{2}(\alpha\sigma)^2} \cos(\alpha\mu) \\ &\quad + i \left[ -\alpha\sigma^2 e^{-\frac{1}{2}(\alpha\sigma)^2} \sin(\alpha\mu) + \mu e^{-\frac{1}{2}(\alpha\sigma)^2} \cos(\alpha\mu) \right]. \end{aligned} \quad (\text{A.16})$$

By identifying the real and imaginary part one gets

$$E[x \cos(\alpha x)] = e^{-\frac{1}{2}(\alpha\sigma)^2} [\mu \cos(\alpha\mu) - \alpha\sigma^2 \sin(\alpha\mu)] \quad (\text{A.17})$$

$$\stackrel{\mu=0}{=} 0 \quad (\text{A.18})$$

$$E[x \sin(\alpha x)] = e^{-\frac{1}{2}(\alpha\sigma)^2} [\mu \sin(\alpha\mu) + \alpha\sigma^2 \cos(\alpha\mu)] \quad (\text{A.19})$$

$$\stackrel{\mu=0}{=} \alpha\sigma^2 \exp\left[-\frac{1}{2}(\alpha\sigma)^2\right]. \quad (\text{A.20})$$

## A.4 Expectation Value of $x^2 \cos(\alpha x)$

The expected value can be derived by differentiating the result from Eqs. (A.15) and (A.16) a second time with respect to  $\alpha$ . One obtains

$$\frac{d^2}{d\alpha^2} E[e^{i\alpha x}] = \frac{d}{d\alpha} E[ixe^{i\alpha x}] = E[-x^2 e^{i\alpha x}] \quad (\text{A.21})$$

$$= -E[x^2 \cos(\alpha x)] - i E[x^2 \sin(\alpha x)]. \quad (\text{A.22})$$

for the left-hand side. Evaluating the right-hand side gives

$$\begin{aligned} \frac{d^2}{d\alpha^2} \left[ e^{-\frac{1}{2}\alpha^2\sigma^2} [\cos(\alpha\mu) + i\sin(\alpha\mu)] \right] &= \dots \\ &- \mu^2 e^{-\frac{1}{2}(\alpha\sigma)^2} \cos(\alpha\mu) + 2\alpha\mu\sigma^2 e^{-\frac{1}{2}(\alpha\sigma)^2} \sin(\alpha\mu) \\ &- \sigma^2 e^{-\frac{1}{2}(\alpha\sigma)^2} \cos(\alpha\mu) + \alpha^2\sigma^4 e^{-\frac{1}{2}(\alpha\sigma)^2} \cos(\alpha\mu) \\ &+ i \left[ -\mu^2 e^{-\frac{1}{2}(\alpha\sigma)^2} \sin(\alpha\mu) - \sigma^2 e^{-\frac{1}{2}(\alpha\sigma)^2} \sin(\alpha\mu) \right. \\ &\left. - 2\alpha\mu\sigma^2 e^{-\frac{1}{2}(\alpha\sigma)^2} \cos(\alpha\mu) + \alpha^2\sigma^4 e^{-\frac{1}{2}(\alpha\sigma)^2} \sin(\alpha\mu) \right]. \end{aligned} \quad (\text{A.23})$$

Using the real part one obtains the result

$$E[x^2 \cos(\alpha x)] = e^{-\frac{1}{2}(\alpha\sigma)^2} [\mu^2 \cos(\alpha\mu) + \sigma^2 \cos(\alpha\mu) - 2\alpha\mu\sigma^2 \sin(\alpha\mu) - \alpha^2\sigma^4 \cos(\alpha\mu)] \quad (\text{A.24})$$

$$\stackrel{\mu=0}{=} (\sigma^2 - \alpha^2\sigma^4) \exp\left[-\frac{1}{2}(\alpha\sigma)^2\right]. \quad (\text{A.25})$$

### A.5 Expectation Value of $x^2 \cos^2(\alpha x)$ and $x^2 \sin^2(\alpha x)$

First the identities  $\cos^2(x) = \frac{1}{2}[1 + \cos(2x)]$  and  $\sin^2(x) = \frac{1}{2}[1 - \cos(2x)]$  are applied, respectively. Afterwards, the result (A.24) can be applied with rescaled frequency  $\alpha$ . One has

$$E[x^2 \cos^2(\alpha x)] = E\left[\frac{x^2}{2} + \frac{x^2 \cos(2\alpha x)}{2}\right] \quad (\text{A.26})$$

$$\begin{aligned} &= \frac{1}{2}(\mu^2 + \sigma^2) + \frac{1}{2}e^{-\frac{1}{2}(2\alpha\sigma)^2} [\mu^2 \cos(2\alpha\mu) + \sigma^2 \cos(2\alpha\mu) \\ &\quad - 4\alpha\mu\sigma^2 \sin(2\alpha\mu) - (2\alpha)^2\sigma^4 \cos(2\alpha\mu)] \end{aligned} \quad (\text{A.27})$$

$$\stackrel{\mu=0}{=} \frac{1}{2}\sigma^2 + \frac{1}{2}(\sigma^2 - (2\alpha)^2\sigma^4) e^{-\frac{1}{2}(2\alpha\sigma)^2}. \quad (\text{A.28})$$

Analogously one gets

$$E[x^2 \sin^2(\alpha x)] = E\left[\frac{x^2}{2} - \frac{x^2 \cos(2\alpha x)}{2}\right] \quad (\text{A.29})$$

$$\begin{aligned} &= \frac{1}{2}(\mu^2 + \sigma^2) - \frac{1}{2}e^{-\frac{1}{2}(2\alpha\sigma)^2} [\mu^2 \cos(2\alpha\mu) + \sigma^2 \cos(2\alpha\mu) \\ &\quad - 4\alpha\mu\sigma^2 \sin(2\alpha\mu) - (2\alpha)^2\sigma^4 \cos(2\alpha\mu)] \end{aligned} \quad (\text{A.30})$$

$$\stackrel{\mu=0}{=} \frac{1}{2}\sigma^2 - \frac{1}{2}(\sigma^2 - (2\alpha)^2\sigma^4) e^{-\frac{1}{2}(2\alpha\sigma)^2}. \quad (\text{A.31})$$

### A.6 Sums Over Expected Values and Variances

In Sec. 1.5.2 the expected values and variances over the sums of  $i = 1, \dots, N$  trigonometric terms with random variable  $y_i \sim \mathcal{N}(0, \sigma_y^2)$  and  $\sigma_y = \frac{R}{\sqrt{N}}$  is needed assuming independent components. The expected values and variances are taken over sums of the terms  $\cos(\alpha y_i)$ ,  $\cos(2\alpha y_i)$  and  $y \sin(\alpha y_i)$ , respectively.

First, the expected values are evaluated according to

$$E \left[ \sum_{i=1}^N \cos(\alpha y_i) \right] = N e^{-\frac{1}{2}(\alpha \sigma_y)^2} = N e^{-\frac{1}{2} \frac{(\alpha R)^2}{N}} \quad (\text{A.32})$$

$$E \left[ \sum_{i=1}^N \cos(2\alpha y_i) \right] = N e^{-\frac{1}{2}(2\alpha \sigma_y)^2} = N e^{-2 \frac{(\alpha R)^2}{N}} \quad (\text{A.33})$$

$$E \left[ \sum_{i=1}^N y_i \sin(\alpha y_i) \right] = N \alpha \sigma_y^2 e^{-\frac{1}{2}(\alpha \sigma_y)^2} = \alpha R^2 e^{-\frac{1}{2} \frac{(\alpha R)^2}{N}}. \quad (\text{A.34})$$

Now the corresponding variances are calculated. Applying  $\text{Var} [\sum_i (\cdot)] = \sum_i E [(\cdot)^2] - E [(\cdot)]^2$  and using previously obtained results of Appendix A with  $\mu_y = 0$  and  $\sigma_y = R/\sqrt{N}$  yields for the variances

$$\begin{aligned} \text{Var} \left[ \sum_{i=1}^N \cos(\alpha y_i) \right] &= \sum_{i=1}^N E [\cos^2(\alpha y_i)] - E [\cos(\alpha y_i)]^2 \\ &= N \left( \frac{1}{2} + \frac{1}{2} e^{-\frac{1}{2} \frac{(2\alpha R)^2}{N}} - e^{-\frac{(\alpha R)^2}{N}} \right) \end{aligned} \quad (\text{A.35})$$

$$\begin{aligned} \text{Var} \left[ \sum_{i=1}^N \cos(2\alpha y_i) \right] &= \sum_{i=1}^N E [\cos^2(2\alpha y_i)] - E [\cos(2\alpha y_i)]^2 \\ &= N \left( \frac{1}{2} + \frac{1}{2} e^{-\frac{1}{2} \frac{(4\alpha R)^2}{N}} - e^{-\frac{(2\alpha R)^2}{N}} \right) \end{aligned} \quad (\text{A.36})$$

$$\begin{aligned} \text{Var} \left[ \sum_{i=1}^N y_i \sin(\alpha y_i) \right] &= \sum_{i=1}^N E [y_i^2 \sin^2(\alpha y_i)] - E [y_i \sin(\alpha y_i)]^2 \\ &= N \left( \frac{1}{2} \frac{R^2}{N} - \frac{1}{2} \left[ \frac{R^2}{N} - (2\alpha)^2 \frac{R^4}{N^2} \right] e^{-\frac{1}{2} \frac{(2\alpha R)^2}{N}} - \alpha^2 \frac{R^4}{N^2} e^{-\frac{(\alpha R)^2}{N}} \right) \end{aligned} \quad (\text{A.37})$$

$$= R^2 \left( \frac{1}{2} - \frac{1}{2} e^{-\frac{1}{2} \frac{(2\alpha R)^2}{N}} + 2\alpha^2 \frac{R^2}{N} e^{-\frac{1}{2} \frac{(2\alpha R)^2}{N}} - \alpha^2 \frac{R^2}{N} e^{-\frac{(\alpha R)^2}{N}} \right). \quad (\text{A.38})$$

The obtained results are used for the evaluation of ratio (1.67) in Sec. 1.5.2.

## Appendix B Large Population Identity

Let  $\lambda > \mu + 1$  and  $\mu > a$  with  $a \geq 1$  and  $\vartheta = \mu/\lambda$ , such that  $t^{\lambda-\mu-1}(1-t)^{\mu-a}$  exhibits its maximum on  $(0, 1)$  and vanishes at  $t \in \{0, 1\}$ . Furthermore, let  $f(t)$  be function defined and differentiable on  $(0, 1)$ , and let  $B(\cdot, \cdot)$  be the beta function. For infinitely large populations  $(\mu, \lambda) \rightarrow \infty$  (constant  $\vartheta$ ) the asymptotic equality holds

$$\begin{aligned} I_{\mu, \lambda}^a[f] &= \frac{1}{B(\lambda - \mu, \mu)} \int_0^1 f(t) t^{\lambda-\mu-1} (1-t)^{\mu-a} dt \\ &\simeq \frac{f(1-\vartheta)}{\vartheta^{a-1}}, \end{aligned} \quad (\text{B.1})$$

with higher order terms vanishing with  $O(1/\mu)$  and  $O(1/\lambda)$ .

### B.1 Derivation

**Properties of the integrand** Given the definition above it can be observed that  $t^{\lambda-\mu-1}(1-t)^{\mu-a}$  exhibits a single increasingly sharp maximum on the interval  $(0, 1)$  given a fixed truncation ratio as the population size tends to infinity, see also Fig. 61.

This observation suggests performing a Taylor series expansion of the function  $f(t)$  around the sharp peak located at  $\hat{t}$  defined by

$$\hat{t} = \operatorname{argmax}_{t \in [0, 1]} [t^{\lambda-\mu-1}(1-t)^{\mu-a}]. \quad (\text{B.2})$$

The first terms of the series should already yield a good approximation for large populations. The maximum of the sharp peak can be obtained by setting the first derivative to zero and deriving the corresponding  $\hat{t}$  according to

$$\begin{aligned} \frac{d}{dt} [t^{\lambda-\mu-1}(1-t)^{\mu-a}] &= 0 \\ (\lambda - \mu - 1)t^{\lambda-\mu-2}(1-t)^{\mu-a} - t^{\lambda-\mu-1}(\mu - a)(1-t)^{\mu-a-1} &= 0 \\ (\lambda - \mu - 1)(1-t) - (\mu - a)t &= 0 \\ (\lambda - \mu - 1) - (\lambda - \mu - 1)t - (\mu - a)t &= 0 \\ (\lambda - \mu - 1) &= (\lambda - a - 1)t \\ \Rightarrow \hat{t} &= \frac{\lambda - \mu - 1}{\lambda - a - 1} = 1 - \frac{\mu - a}{\lambda - a - 1} = 1 - \frac{\mu(1 - a/\mu)}{\lambda(1 - a/\lambda - 1/\lambda)}. \end{aligned} \quad (\text{B.3})$$

Looking at the limit of infinitely large populations it can be observed that the maximizer approaches a constant value. Setting  $\mu/\lambda = \vartheta$  one gets

$$\lim_{\substack{(\mu, \lambda) \rightarrow \infty \\ \vartheta = \text{const.}}} \hat{t} = 1 - \vartheta. \quad (\text{B.5})$$

**Taylor expansion** Taylor-expanding  $f(t)$  around  $\hat{t}$  yields

$$f(t) = \sum_{k=0}^{\infty} \frac{1}{k!} \left. \frac{\partial^k f}{\partial t^k} \right|_{t=\hat{t}} (t - \hat{t})^k, \quad (\text{B.6})$$

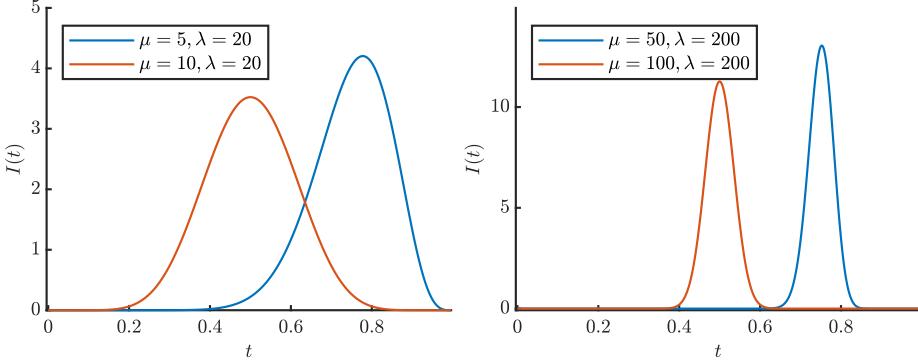


Figure 61: Integrand  $\frac{1}{B(\lambda-\mu,\mu)}t^{\lambda-\mu-1}(1-t)^{\mu-1}$  plotted for  $a = 1$ ,  $\lambda = 20$  (left) and  $\lambda = 200$  (right) for two different truncation ratios  $\vartheta = 1/4$  and  $\vartheta = 1/2$ . The peak sharpness increases with growing  $\lambda$  and the factor  $1/B(\lambda-\mu,\mu)$  rescales the peak heights.

such that integral (B.1) is expressed as

$$\begin{aligned} I_{\mu,\lambda}^a[f] &= \frac{1}{B(\lambda-\mu,\mu)} \int_0^1 \sum_{k=0}^{\infty} \frac{1}{k!} \left. \frac{\partial^k f}{\partial t^k} \right|_{t=\hat{t}} (t-\hat{t})^k t^{\lambda-\mu-1} (1-t)^{\mu-a} dt \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} \left. \frac{\partial^k f}{\partial t^k} \right|_{t=\hat{t}} \frac{1}{B(\lambda-\mu,\mu)} \int_0^1 t^{\lambda-\mu-1} (1-t)^{\mu-a} (t-\hat{t})^k dt \quad (\text{B.7}) \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} \left. \frac{\partial^k f}{\partial t^k} \right|_{t=\hat{t}} C^{(k)}. \end{aligned}$$

The introduced coefficients  $C^{(k)}$  are defined as

$$C^{(k)} := \frac{1}{B(\lambda-\mu,\mu)} \int_0^1 t^{\lambda-\mu-1} (1-t)^{\mu-a} (t-\hat{t})^k dt. \quad (\text{B.8})$$

It will be shown that only the 0-th order coefficient  $C^{(0)}$  will yield significant contributions and all higher orders  $k \geq 1$  will vanish with  $O(1/\lambda)$  for large populations.

Starting with  $k = 0$  the coefficient can be evaluated as

$$\begin{aligned} C^{(0)} &= \frac{1}{B(\lambda-\mu,\mu)} \int_0^1 t^{\lambda-\mu-1} (1-t)^{\mu-a} dt \\ &= \frac{B(\lambda-\mu,\mu-a+1)}{B(\lambda-\mu,\mu)} = \frac{\Gamma(\lambda-\mu)\Gamma(\mu-a+1)}{\Gamma(\lambda-a+1)} \frac{\Gamma(\lambda)}{\Gamma(\lambda-\mu)\Gamma(\mu)} \\ &= \frac{(\lambda-\mu-1)!(\mu-a)!}{(\lambda-a)!} \frac{(\lambda-1)!}{(\lambda-\mu-1)!(\mu-1)!} \\ &= \frac{(\lambda-1)!(\mu-a)!}{(\lambda-a)!(\mu-1)!} = \frac{\prod_{n=1}^{a-1} \lambda-n}{\prod_{n=1}^{a-1} \mu-n} = \prod_{n=1}^{a-1} \frac{\lambda}{\mu} \frac{1-n/\lambda}{1-n/\mu} \\ &= \begin{cases} 1 & \text{for } a = 1, \\ \frac{1}{\vartheta^{a-1}} \prod_{n=1}^{a-1} \frac{1-n/\lambda}{1-n/\mu}. & \text{for } a > 1. \end{cases} \quad (\text{B.9}) \end{aligned}$$

It was used that for  $a > 1$  one has

$$\frac{(\lambda - 1)!}{(\lambda - a)!} = \prod_{n=1}^{a-1} \lambda - n \quad \text{and} \quad \frac{(\mu - a)!}{(\mu - 1)!} = \frac{1}{\prod_{n=1}^{a-1} \mu - n}. \quad (\text{B.10})$$

Therefore the limit yields for any  $a \geq 1$

$$\lim_{\substack{(\mu, \lambda) \rightarrow \infty \\ \vartheta = \text{const.}}} C^{(0)} = \frac{1}{\vartheta^{a-1}}. \quad (\text{B.11})$$

with  $O(1/\mu)$  and  $O(1/\lambda)$ . The analysis of  $C^{(k)}$  with  $k \geq 1$  is slightly more involved and is tackled now. Noting that  $(t - \hat{t})^k = (-\hat{t})^k (1 - t/\hat{t})^k$  one has

$$C^{(k)} = \frac{(-\hat{t})^k}{B(\lambda - \mu, \mu)} \int_0^1 t^{\lambda - \mu - 1} (1 - t)^{\mu - a} \left(1 - \frac{t}{\hat{t}}\right)^k dt. \quad (\text{B.12})$$

By applying the binomial theorem the expression can be reformulated

$$\begin{aligned} C^{(k)} &= \frac{(-\hat{t})^k}{B(\lambda - \mu, \mu)} \int_0^1 t^{\lambda - \mu - 1} (1 - t)^{\mu - a} \sum_{i=0}^k \binom{k}{i} 1^{k-i} \left(-\frac{t}{\hat{t}}\right)^i dt \\ &= (-\hat{t})^k \sum_{i=0}^k \binom{k}{i} (-1)^i \left(\frac{1}{\hat{t}}\right)^i \frac{1}{B(\lambda - \mu, \mu)} \int_0^1 t^i t^{\lambda - \mu - 1} (1 - t)^{\mu - a} dt \\ &= (-\hat{t})^k \sum_{i=0}^k \binom{k}{i} (-1)^i F_1^{i,a} F_2^{i,a}, \end{aligned} \quad (\text{B.13})$$

with additional treatment needed for the factors  $F_1^{i,a}$  and  $F_2^{i,a}$

$$F_1^{i,a} := \left(\frac{1}{\hat{t}}\right)^i \quad (\text{B.14})$$

$$F_2^{i,a} := \frac{1}{B(\lambda - \mu, \mu)} \int_0^1 t^{\lambda - \mu - 1 + i} (1 - t)^{\mu - a} dt. \quad (\text{B.15})$$

Factor  $F_1^{i,a}$  is easily evaluated using (B.5) and yields

$$F_1^{i,a} = \left(\frac{1}{\hat{t}}\right)^i = \left(\frac{\lambda - a - 1}{\lambda - \mu - 1}\right)^i = \left(\frac{1 - a/\lambda - 1/\lambda}{1 - \vartheta - 1/\lambda}\right)^i. \quad (\text{B.16})$$

Factor  $F_2^{i,a}$  yields

$$\begin{aligned} F_2^{i,a} &= \frac{B(\lambda - \mu + i, \mu - a + 1)}{B(\lambda - \mu, \mu)} \\ &= \frac{(\lambda - \mu - 1 + i)!(\mu - a)!}{(\lambda - a + i)!} \frac{(\lambda - 1)!}{(\lambda - \mu - 1)!(\mu - 1)!} \\ &= \frac{(\lambda - \mu - 1 + i)!}{(\lambda - \mu - 1)!} \frac{(\lambda - 1)!(\mu - a)!}{(\lambda - a + i)!(\mu - 1)!}. \end{aligned} \quad (\text{B.17})$$

The first ratio of result (B.17) yields

$$\frac{(\lambda - \mu - 1 + i)!}{(\lambda - \mu - 1)!} = \prod_{j=1}^i (\lambda - \mu - 1 + j) = \lambda^i \prod_{j=1}^i (1 - \vartheta - 1/\lambda + j/\lambda). \quad (\text{B.18})$$

For the second ratio of result (B.17) one can use (B.10) and write  $(\lambda - 1)! = (\lambda - a)! \prod_{n=1}^{a-1} \lambda - n$ , such that

$$\begin{aligned} \frac{(\lambda - 1)!(\mu - a)!}{(\lambda - a + i)!(\mu - 1)!} &\stackrel{a \geq 1}{=} \frac{(\lambda - a)!}{(\lambda - a + i)!} \frac{\prod_{n=1}^{a-1} \lambda - n}{\prod_{n=1}^{a-1} \mu - n} \\ &= \frac{1}{\prod_{j=1}^i (\lambda - a + j)} \frac{\prod_{n=1}^{a-1} \lambda - n}{\prod_{n=1}^{a-1} \mu - n} \\ &= \frac{1}{\lambda^i \prod_{j=1}^i (1 - a/\lambda + j/\lambda)} \frac{\lambda^{a-1} \prod_{n=1}^{a-1} 1 - n/\lambda}{\mu^{a-1} \prod_{n=1}^{a-1} 1 - n/\mu} \quad (\text{B.19}) \\ \frac{(\lambda - 1)!(\mu - a)!}{(\lambda - a + i)!(\mu - 1)!} &\stackrel{a=1}{=} \frac{1}{\lambda^i \prod_{j=1}^i (1 - 1/\lambda + j/\lambda)}. \end{aligned}$$

The result of (B.19) for  $a > 1$  is also valid for  $a = 1$  when defining the product over  $n$  with no elements as  $\prod_{n=1}^0 (\cdot) = 1$ , which is assumed for the following derivations.

Using (B.18) and (B.19) factor  $F_2^{i,a}$  therefore yields for  $a \geq 1$

$$\begin{aligned} F_2^{i,a} &= \frac{\lambda^i \prod_{j=1}^i (1 - \vartheta - 1/\lambda + j/\lambda)}{\lambda^i \prod_{j=1}^i (1 - a/\lambda + j/\lambda)} \frac{\lambda^{a-1} \prod_{n=1}^{a-1} (1 - n/\lambda)}{\mu^{a-1} \prod_{n=1}^{a-1} (1 - n/\mu)} \\ &= \frac{1}{\vartheta^{a-1}} \prod_{j=1}^i \frac{(1 - \vartheta - 1/\lambda + j/\lambda)}{(1 - a/\lambda + j/\lambda)} \prod_{n=1}^{a-1} \frac{(1 - n/\lambda)}{(1 - n/\mu)}. \end{aligned} \quad (\text{B.20})$$

Finally the result for  $C^{(k)}$  from (B.13) can be evaluated using (B.16) and (B.20)

$$\begin{aligned} C^{(k)} &= (-\hat{t})^k \sum_{i=0}^k \binom{k}{i} (-1)^i F_1^{i,a} F_2^{i,a} \\ &= (-\hat{t})^k \sum_{i=0}^k \binom{k}{i} (-1)^i \frac{(1 - a/\lambda - 1/\lambda)^i}{(1 - \vartheta - 1/\lambda)^i} \\ &\quad \times \frac{1}{\vartheta^{a-1}} \prod_{j=1}^i \frac{(1 - \vartheta - 1/\lambda + j/\lambda)}{(1 - a/\lambda + j/\lambda)} \prod_{n=1}^{a-1} \frac{(1 - n/\lambda)}{(1 - n/\mu)} \\ &= \frac{(-\hat{t})^k}{\vartheta^{a-1}} \prod_{n=1}^{a-1} \frac{(1 - n/\lambda)}{(1 - n/\mu)} \\ &\quad \times \sum_{i=0}^k \binom{k}{i} (-1)^i \prod_{j=1}^i \frac{(1 - \vartheta - 1/\lambda + j/\lambda)(1 - a/\lambda - 1/\lambda)}{(1 - \vartheta - 1/\lambda)(1 - a/\lambda + j/\lambda)}. \end{aligned} \quad (\text{B.21})$$

In the second line of (B.21) factors  $F_1^{i,a}$  and  $F_2^{i,a}$  were inserted and in the last line the product over  $n$  was moved out of the sum as no  $i$  dependency

is given. Additionally the factors  $(1 - a/\lambda - 1/\lambda)^i$  and  $1/(1 - \vartheta - 1/\lambda)^i$  were moved into the product over  $j = 1, \dots, i$  which is important for the following limit consideration.

Applying the limit  $(\mu, \lambda) \rightarrow \infty$  significantly simplifies (B.21), as the population dependent terms vanish with  $O(1/\mu)$  and  $O(1/\lambda)$ , respectively. The two products yield asymptotically one. Using the property that the sum of alternating binomial coefficients yields zero for any  $k \geq 1$ , one obtains the limit

$$\lim_{\substack{(\mu, \lambda) \rightarrow \infty \\ \vartheta = \text{const.}}} C^{(k \geq 1)} = \frac{(-\hat{t})^k}{\vartheta^{a-1}} \sum_{i=0}^k \binom{k}{i} (-1)^i = 0. \quad (\text{B.22})$$

**Collecting results** Having established the large population limit of  $C^{(k)}$  in (B.11) and (B.22) one can return to the Taylor expansion of (B.7) and evaluate corresponding expressions. Noting that  $\hat{t} = 1 - \vartheta$  from (B.5) the result is

$$\begin{aligned} \lim_{\substack{(\mu, \lambda) \rightarrow \infty \\ \vartheta = \text{const.}}} I_{\mu, \lambda}^a[f] &= \lim_{\substack{(\mu, \lambda) \rightarrow \infty \\ \vartheta = \text{const.}}} \left[ \sum_{k=0}^{\infty} \frac{1}{k!} \left. \frac{\partial^k f}{\partial t^k} \right|_{t=\hat{t}} C^{(k)} \right] \\ &= \frac{1}{\vartheta^{a-1}} f(1 - \vartheta), \end{aligned} \quad (\text{B.23})$$

with higher order terms vanishing as  $O(1/\mu)$  and  $O(1/\lambda)$ . Therefore within the large population limit it is sufficient to consider only the 0-th order term of the Taylor expansion evaluated at the integrand maximum  $\hat{t}$ . All these considerations hold provided that the derivatives of  $f(t)$  are well defined at  $\hat{t}$ .

## B.2 Applications

**Progress Rate** Identity (B.1) is used to solve the progress rate integral of (2.25) for large populations with parameter  $a = 1$  and  $f(t)$  defined in Eq. (2.24). The result is given in Eq. (2.28).

**Generalized Progress Coefficient** A second application emerges investigating the generalized progress coefficients introduced by Beyer [5, p. 172]

$$e_{\mu, \lambda}^{a, b} = \frac{\lambda - \mu}{(2\pi)^{\frac{a+1}{2}}} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} x^b e^{-\frac{a+1}{2}x^2} [\Phi(x)]^{\lambda - \mu - 1} [1 - \Phi(x)]^{\mu - a} dx, \quad (\text{B.24})$$

for which asymptotic properties can be derived assuming large populations. The population depend prefactors are rewritten as

$$(\lambda - \mu) \binom{\lambda}{\mu} = \frac{\lambda}{\mu} \frac{(\lambda - 1)!}{(\lambda - \mu - 1)!(\mu - 1)!} = \frac{1}{\vartheta} \frac{1}{B(\lambda - \mu, \mu)}. \quad (\text{B.25})$$

Introducing the substitution  $t = \Phi(x)$  with  $x = \Phi^{-1}(t)$ ,  $dx = \sqrt{2\pi} e^{x^2/2} dt$  and changing the bounds  $0 \leq t \leq 1$  the progress coefficients yields

$$e_{\mu, \lambda}^{a, b} = \frac{1}{\vartheta} \frac{1}{B(\lambda - \mu, \mu)} \frac{1}{(2\pi)^{a/2}} \int_0^1 [\Phi^{-1}(t)]^b e^{-\frac{a}{2}[\Phi^{-1}(t)]^2} t^{\lambda - \mu - 1} (1 - t)^{\mu - a} dt. \quad (\text{B.26})$$

Comparing (B.26) with identity (B.1) the function  $f^{a,b}(t)$  (with  $a$  and  $b$  in superscript emphasizing the parameter dependence) can be identified as

$$\begin{aligned} f^{a,b}(t)|_{t=1-\vartheta} &= [\Phi^{-1}(t)]^b e^{-\frac{a}{2}[\Phi^{-1}(t)]^2} \Big|_{t=1-\vartheta} \\ &= [\Phi^{-1}(1-\vartheta)]^b e^{-\frac{a}{2}[\Phi^{-1}(1-\vartheta)]^2}. \end{aligned} \quad (\text{B.27})$$

Therefore the coefficients can be expressed as

$$\begin{aligned} e_{\mu,\lambda}^{a,b} &= \frac{1}{\vartheta} \frac{1}{(2\pi)^{a/2}} I_{\mu,\lambda}^a [f^{a,b}] \\ &\simeq \frac{1}{\vartheta} \frac{1}{(2\pi)^{a/2}} \frac{f(1-\vartheta)}{\vartheta^{a-1}} \\ &= \frac{1}{(2\pi)^{a/2}} \frac{1}{\vartheta^a} [\Phi^{-1}(1-\vartheta)]^b e^{-\frac{a}{2}[\Phi^{-1}(1-\vartheta)]^2} \\ &= \left[ \frac{e^{-\frac{1}{2}[\Phi^{-1}(\vartheta)]^2}}{\sqrt{2\pi}\vartheta} \right]^a [-\Phi^{-1}(\vartheta)]^b. \end{aligned} \quad (\text{B.28})$$

In the second line the asymptotic equality is used. For the last line the properties  $\Phi^{-1}(1-\vartheta) = -\Phi^{-1}(\vartheta)$ ,  $[\Phi^{-1}(1-\vartheta)]^2 = [\Phi^{-1}(\vartheta)]^2$  are applied and all factors being powers of  $a$  and  $b$  are collected. Defining the asymptotic generalized progress coefficient as

$$e_{\vartheta}^{a,b} := \lim_{\substack{(\mu,\lambda) \rightarrow \infty \\ \vartheta = \text{const.}}} e_{\mu,\lambda}^{a,b}, \quad (\text{B.29})$$

the final result yields

#### Asymptotic Generalized Progress Coefficient

$$e_{\vartheta}^{a,b} = \left[ \frac{e^{-\frac{1}{2}[\Phi^{-1}(\vartheta)]^2}}{\sqrt{2\pi}\vartheta} \right]^a [-\Phi^{-1}(\vartheta)]^b. \quad (\text{B.30})$$

The result (B.30) is numerically validated in Fig. 62 showing  $e_{\mu,\lambda}^{a,b}$  as a function of  $\mu$  (constant  $\vartheta$ ) and the corresponding asymptote  $e_{\vartheta}^{a,b}$ .

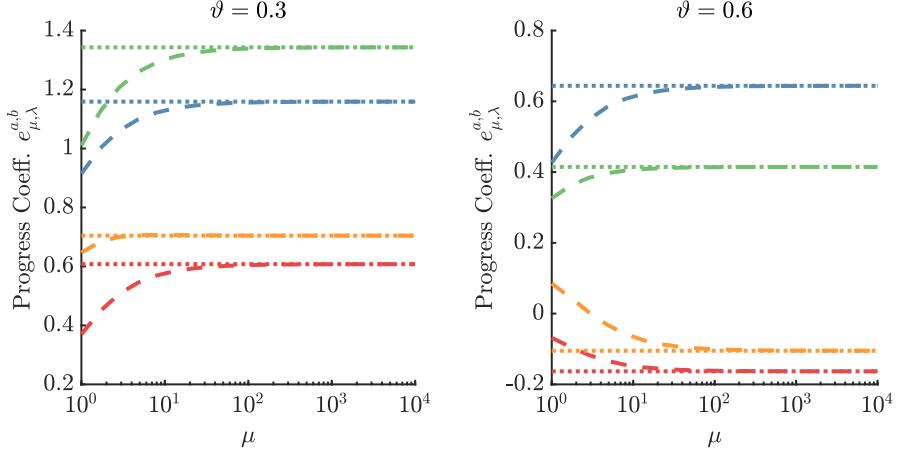


Figure 62: Generalized progress coefficient  $e_{\mu,\lambda}^{a,b}$  numerically calculated (dashed) with the corresponding asymptotic limit  $e_{\vartheta}^{a,b}$  (dotted) from Eq. (B.30) for  $\vartheta \in \{0.3, 0.6\}$ . The coefficient orders are  $e^{1,0}$  (blue),  $e^{1,1}$  (red),  $e^{2,0}$  (green) and  $e^{2,1}$  (orange).

## Appendix C Noisy Order Statistics

**Modeling the distribution of  $z$**  The underlying distribution from which the not selected  $z$  in (3.14) are drawn is modeled in terms of its cumulants  $\kappa_i$  by expanding a normally distributed density function using a Gram-Charlier series

$$p_z(z) = \frac{1}{\sqrt{2\pi\kappa_2}} \exp\left[-\frac{z^2}{2\kappa_2}\right] \left(1 + \frac{\gamma_1}{6} \text{He}_3\left(\frac{z}{\sqrt{\kappa_2}}\right) + \frac{\gamma_2}{24} \text{He}_4\left(\frac{z}{\sqrt{\kappa_2}}\right) + \dots\right), \quad (\text{C.1})$$

with skewness  $\gamma_1$  and excess  $\gamma_2$

$$\begin{aligned} \gamma_1 &= \kappa_3/\kappa_2^{3/2} \\ \gamma_2 &= \kappa_4/\kappa_2^2. \end{aligned} \quad (\text{C.2})$$

The probabilist's Hermite polynomials are defined by

$$\frac{d^k \phi(x)}{dx^k} = (-1)^k \text{He}_k(x) \phi(x). \quad (\text{C.3})$$

For  $z$  being the mutation variable with  $z \sim \mathcal{N}(0, \sigma^2)$ , as in our case, the second cumulant  $\kappa_2 = \sigma^2$  and  $\kappa_i = 0$  for  $i = 1$  and  $i \geq 3$ . In this case the expansion is exact.

If the series is truncated with some cumulants  $\kappa_i \neq 0$  ( $i \geq 3$ ), the expression (C.1) is only an *approximation* of a probability density and may yield  $p_z(z) < 0$  for some values of  $z$ . However the integral  $\int_{-\infty}^{\infty} p_z(z) dz = 1$  as the Hermite terms have vanishing contribution. As higher order approximations are usually not required, Arnold has neglected cumulants of higher order than four.

Considering the case  $s = -1$  of (3.14), an expression for the density is needed evaluated at  $(-z)$  giving

$$\begin{aligned} p_z(-z) &= \frac{1}{\sqrt{2\pi\kappa_2}} \exp\left[-\frac{(-z)^2}{2\kappa_2}\right] \left(1 + \frac{\gamma_1}{6} \text{He}_3\left(\frac{-z}{\sqrt{\kappa_2}}\right) + \frac{\gamma_2}{24} \text{He}_4\left(\frac{-z}{\sqrt{\kappa_2}}\right) + \dots\right), \\ &= \frac{1}{\sqrt{2\pi\kappa_2}} \exp\left[-\frac{z^2}{2\kappa_2}\right] \left(1 + \frac{(-\gamma_1)}{6} \text{He}_3\left(\frac{z}{\sqrt{\kappa_2}}\right) + \frac{\gamma_2}{24} \text{He}_4\left(\frac{z}{\sqrt{\kappa_2}}\right) + \dots\right), \\ &=: p_{-z}(z) \end{aligned} \quad (\text{C.4})$$

where we have used that  $\text{He}_k(-x) = \text{He}_k(x)$  for even  $k$  and  $\text{He}_k(-x) = -\text{He}_k(x)$  for odd  $k$ . The transformed density yields  $-\gamma_1$  instead of  $\gamma_1$ , which is equivalent to the third cumulant being  $-\kappa_3$  instead of  $\kappa_3$ . In general, all odd Hermite terms yield a negative sign. Due to truncation they are not shown here.

**Transition density** Starting at Eq. (3.14) and assuming  $z, s$  are given, the transition density  $p(v|z, s)$  is obtained by reformulating the expression to

$$\frac{v - sz}{\sigma_\epsilon} = \mathcal{N}(0, 1), \quad (\text{C.5})$$

such that

$$p(v|z, s) = \frac{1}{\sigma_\epsilon} \phi\left(\frac{v - sz}{\sigma_\epsilon}\right). \quad (\text{C.6})$$

**Modeling density and distribution of  $v$**  The density of measured values  $p_v(v)$  can be obtained by integrating  $p_z(z)$  from series (C.1) over transition density (C.6) according to

$$\begin{aligned} p_v(v) &= \int_{-\infty}^{\infty} p(v|z, s) p_z(z) dz \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma_\epsilon} \phi\left(\frac{v - sz}{\sigma_\epsilon}\right) p_z(z) dz. \end{aligned} \quad (\text{C.7})$$

Having two cases integral (C.7) yields

$$\int_{-\infty}^{\infty} \frac{1}{\sigma_\epsilon} \phi\left(\frac{v - sz}{\sigma_\epsilon}\right) p_z(z) dz = \begin{cases} \int_{-\infty}^{\infty} \frac{1}{\sigma_\epsilon} \phi\left(\frac{v - z}{\sigma_\epsilon}\right) p_z(z) dz & \text{for } s = 1, \\ \int_{-\infty}^{\infty} \frac{1}{\sigma_\epsilon} \phi\left(\frac{v - z}{\sigma_\epsilon}\right) p_z(-z) dz & \text{for } s = -1, \end{cases} \quad (\text{C.8})$$

using the substitution  $z = -z$  for the second line, which moves the sign from  $\frac{1}{\sigma_\epsilon} \phi\left(\frac{v+z}{\sigma_\epsilon}\right)$  into  $p_z(-z)$ . One can see that given the case  $s = -1$  the density  $p_v(v)$  can be described in terms of  $p_z(-z)$  which in turn only changes the sign of  $\kappa_3$  (and therefore  $\gamma_1$ ) compared to  $s = 1$ . This property is used throughout the subsequent derivations.

In principal, the integrals (C.8) can be solved in a straightforward manner inserting the densities and applying Identity 2 [2, p. 115] with Hermite terms

integrated over the given exponential factors. However, the resulting density can also be established in an easier way using following argumentation.

As  $z$  and  $\epsilon$  are independent variates and  $v = sz + \epsilon$ , the cumulants of  $p_v(v)$  denoted by  $\tilde{\kappa}_i$  can be obtained by addition of respective cumulants

$$\tilde{\kappa}_1 = 0, \quad \tilde{\kappa}_2 = \kappa_2 + \sigma_\epsilon^2, \quad \tilde{\kappa}_3 = s\kappa_3, \quad \tilde{\kappa}_4 = \kappa_4, \quad \dots \quad (\text{C.9})$$

Note that sign  $s$  appears for  $\tilde{\kappa}_3$  to include both cases. Due to truncation higher orders are not shown. Density  $p_v(v)$  must therefore have the same form as expansion (C.1) with corresponding skewness and excess due to added cumulants, such that one gets

$$p_v(v) = \frac{1}{\sqrt{\tilde{\kappa}_2}} \phi\left(\frac{v}{\sqrt{\tilde{\kappa}_2}}\right) \left[ 1 + \frac{\tilde{\gamma}_1(s)}{6} \text{He}_3\left(\frac{v}{\sqrt{\tilde{\kappa}_2}}\right) + \frac{\tilde{\gamma}_2}{24} \text{He}_4\left(\frac{v}{\sqrt{\tilde{\kappa}_2}}\right) + \dots \right]. \quad (\text{C.10})$$

By defining a noise coefficient  $a$  according to [2, p. 139]

$$a = \sqrt{\frac{\kappa_2}{\kappa_2 + \sigma_\epsilon^2}}, \quad (\text{C.11})$$

resulting skewness and excess quantities for  $p_v(v)$  can be expressed as

$$\begin{aligned} \tilde{\gamma}_1 &= \frac{\tilde{\kappa}_3}{\tilde{\kappa}_2^{3/2}} = \frac{s\kappa_3}{(\kappa_2 + \sigma_\epsilon^2)^{3/2}} = s\gamma_1 a^3 \\ \tilde{\gamma}_2 &= \frac{\tilde{\kappa}_4}{\tilde{\kappa}_2^2} = \frac{\kappa_4}{(\kappa_2 + \sigma_\epsilon^2)^2} = \gamma_2 a^4. \end{aligned} \quad (\text{C.12})$$

Density  $p_v(v)$  can therefore be expressed in terms of the original cumulants of expansion (C.1) using the noise coefficient  $a$  giving

$$\begin{aligned} p_v(v) &= \frac{1}{\sqrt{\kappa_2 + \sigma_\epsilon^2}} \phi\left(\frac{v}{\sqrt{\kappa_2 + \sigma_\epsilon^2}}\right) \\ &\quad \left[ 1 + s \frac{\gamma_1 a^3}{6} \text{He}_3\left(\frac{v}{\sqrt{\kappa_2 + \sigma_\epsilon^2}}\right) + \frac{\gamma_2 a^4}{24} \text{He}_4\left(\frac{v}{\sqrt{\kappa_2 + \sigma_\epsilon^2}}\right) + \dots \right]. \end{aligned} \quad (\text{C.13})$$

The resulting distribution function can be obtained by integration

$$P_v(v) = \int_{-\infty}^v p_v(v') dv'. \quad (\text{C.14})$$

The terms of  $p_v(v)$  are easily integrated using (C.3) with  $\phi(x) \text{He}_k(x) = (-1)^k \frac{d^k \phi(x)}{dx^k}$ , then performing integration on the  $k$ -th differential yielding order  $k - 1$ , and then back-transforming again using (C.3).

Integrating the first term of (C.13) yields

$$\begin{aligned} \int_{-\infty}^v \frac{1}{\sqrt{\kappa_2 + \sigma_\epsilon^2}} \phi\left(\frac{v'}{\sqrt{\kappa_2 + \sigma_\epsilon^2}}\right) dv' &= \int_{x=-\infty}^{x=v/\sqrt{\kappa_2 + \sigma_\epsilon^2}} \phi(x) dx \\ &= \Phi\left(\frac{v}{\sqrt{\kappa_2 + \sigma_\epsilon^2}}\right), \end{aligned} \quad (\text{C.15})$$

using the substitution  $x = v'/\sqrt{\kappa_2 + \sigma_\epsilon^2}$ . For the coefficient  $\gamma_1$  of (C.13) one has

$$\begin{aligned}
& s \frac{\gamma_1 a^3}{6} \int_{-\infty}^v \frac{1}{\sqrt{\kappa_2 + \sigma_\epsilon^2}} \phi \left( \frac{v'}{\sqrt{\kappa_2 + \sigma_\epsilon^2}} \right) \text{He}_3 \left( \frac{v'}{\sqrt{\kappa_2 + \sigma_\epsilon^2}} \right) dv' \\
&= s \frac{\gamma_1 a^3}{6} \int_{-\infty}^{v/\sqrt{\kappa_2 + \sigma_\epsilon^2}} \phi(x) \text{He}_3(x) dx \\
&= s \frac{\gamma_1 a^3}{6} \int_{-\infty}^{v/\sqrt{\kappa_2 + \sigma_\epsilon^2}} \left[ (-1)^3 \frac{d^3 \phi(x)}{dx^3} \right] dx \\
&= s \frac{\gamma_1 a^3}{6} \left. (-1) \frac{d^2 \phi(x)}{dx^2} \right|_{x=v/\sqrt{\kappa_2 + \sigma_\epsilon^2}} \\
&= -s \frac{\gamma_1 a^3}{6} \phi \left( \frac{v}{\sqrt{\kappa_2 + \sigma_\epsilon^2}} \right) \text{He}_2 \left( \frac{v}{\sqrt{\kappa_2 + \sigma_\epsilon^2}} \right).
\end{aligned} \tag{C.16}$$

Analogously, for the last term with coefficient  $\gamma_2$  one gets

$$\begin{aligned}
& \frac{\gamma_2 a^4}{24} \int_{-\infty}^v \frac{1}{\sqrt{\kappa_2 + \sigma_\epsilon^2}} \phi \left( \frac{v'}{\sqrt{\kappa_2 + \sigma_\epsilon^2}} \right) \text{He}_4 \left( \frac{v'}{\sqrt{\kappa_2 + \sigma_\epsilon^2}} \right) dv' \\
&= -\frac{\gamma_2 a^4}{24} \phi \left( \frac{v}{\sqrt{\kappa_2 + \sigma_\epsilon^2}} \right) \text{He}_3 \left( \frac{v}{\sqrt{\kappa_2 + \sigma_\epsilon^2}} \right).
\end{aligned} \tag{C.17}$$

Therefore the distribution function of (C.14) yields

$$\begin{aligned}
P_v(v) &= \Phi \left( \frac{v}{\sqrt{\kappa_2 + \sigma_\epsilon^2}} \right) - \phi \left( \frac{v}{\sqrt{\kappa_2 + \sigma_\epsilon^2}} \right) \\
&\quad \left[ s \frac{\gamma_1 a^3}{6} \text{He}_2 \left( \frac{v}{\sqrt{\kappa_2 + \sigma_\epsilon^2}} \right) + \frac{\gamma_2 a^4}{24} \text{He}_3 \left( \frac{v}{\sqrt{\kappa_2 + \sigma_\epsilon^2}} \right) + \dots \right].
\end{aligned} \tag{C.18}$$

During later derivations, the sum of (C.25) is converted into an integral, such that after substitution the quantile function  $P_v^{-1}(\cdot)$  evaluated at  $\Phi(x)$  is needed, see also [2, p. 138, D.14].

As  $P_v(v)$  is given by an expansion w.r.t. a normal distribution, its inverse can be approximated using a Cornish-Fisher series assuming small deviations from a normal distribution. The idea is given the case  $v \sim \mathcal{N}(0, 1)$  with  $\tilde{\kappa}_2 = 1$  and  $\kappa_i = 0$  for  $i \geq 3$ , the series expansion for  $P_v^{-1}$  should be exact giving  $P_v^{-1}(\Phi(x)) = x$ . For  $v$  defined by cumulants (C.9) the series needs correction terms and rescaling, such that using expansion [2, p. 111, A.14] one gets

$$P_v^{-1}(\Phi(x)) = \sqrt{\kappa_2 + \sigma_\epsilon^2} \left( x + s \frac{\gamma_1 a^3}{6} \text{He}_2(x) + \frac{\gamma_2 a^4}{24} \text{He}_3(x) + \dots \right). \tag{C.19}$$

Having defined the densities and distribution functions, the problem can be formulated as follows.

**Problem formulation** A sum  $S_P$  of products of noisy ordered variates containing  $\nu$  terms per summand can be expressed as

$$S_P = \sum_{\{n_1, \dots, n_\nu\}} z_{n_1; \lambda}^{p_1} \cdots z_{n_\nu; \lambda}^{p_\nu}, \tag{C.20}$$

where the vector  $P = (p_1, \dots, p_\nu)$  denotes the positive exponents. Distinct summation indices are denoted by the set  $\{n_1, \dots, n_\nu\}$ . The goal is to provide a generic solution to  $E[S_P]$  for arbitrary  $\nu$  and  $P$  (up to some order).

In order to apply certain integral identities during the derivation, the sum has to be reordered. Defining  $\pi_P$  as the permutation of  $z$  variates, such that for each summand the ordering  $n_1 < \dots < n_i < \dots < n_\nu$  is always maintained with the rightmost element being the largest. The sum (C.20) is rewritten

$$S_P = \sum_{n_1=\nu}^{\mu} \sum_{n_2=\nu-1}^{n_1-1} \cdots \sum_{n_\nu=1}^{n_{\nu-1}-1} \pi_P(z_{n_1;\lambda}, \dots, z_{n_\nu;\lambda}). \quad (\text{C.21})$$

Evaluating the expectation  $E[S_P]$  requires modeling the joint probability density of  $\nu$  ordered elements  $p_{n_1, \dots, n_\nu; \lambda}(z_1, \dots, z_\nu)$ . As an example, the joint density for  $\nu = 2$  and  $\{n_1, n_2\} = \{k, l\}$  will be constructed, which corresponds to the sum of Eq. (3.9).

The variates  $z_k$  and  $z_l$  are distributed with  $p_z(z_k)$  and  $p_z(z_l)$  using (C.1), respectively, and each variate is independently disturbed by noise  $\sigma_\epsilon^2$ . Therefore measured value  $v$  is realized with conditional density of Eq. (C.6) depending on sign  $s$ .

For the measured values ordering  $1 \leq k < l \leq \lambda$  is assumed with  $k$  denoting the  $k$ -th best (largest) value. Given measured values  $v_k$  and  $v_l$  with the distribution function  $P_v(v)$  of Eq. (C.18), the selection of the  $k$ -th and  $l$ -th largest values gives the probabilities

$$\begin{aligned} \Pr\{V > v_l\} &= [1 - P_v(v_l)]^{l-1} \\ \Pr\{v_k \leq V \leq v_l\} &= [P_v(v_l) - P_v(v_k)]^{k-l-1} \\ \Pr\{V < v_k\} &= [P_v(v_k)]^{\lambda-k}, \end{aligned} \quad (\text{C.22})$$

with  $l-1$  denoting larger values than  $v_l$  and  $\lambda-k$  smaller values than  $v_k$ . The number of intermediate values is obtained by  $\lambda - (\lambda - k) - (l - 1) - 2 = k - l - 1$ .

Finally the joint density  $p_{k,l;\lambda}(z_k, z_l)$  can be given using  $p_z(z_k)$ ,  $p_z(z_l)$ , the transition densities  $p(v_k|z_k, s)$ ,  $p(v_l|z_l, s)$  and probabilities from (C.22) by integrating over all  $v_k$  and  $v_l$  according to

$$\begin{aligned} p_{k,l;\lambda}(z_k, z_l) &= p_z(z_k)p_z(z_l) \int_{-\infty}^{\infty} \int_{v_k}^{\infty} \frac{1}{\sigma_\epsilon} \phi\left(\frac{v_k - sz_k}{\sigma_\epsilon}\right) \frac{1}{\sigma_\epsilon} \phi\left(\frac{v_l - sz_l}{\sigma_\epsilon}\right) \\ &\times \lambda! \frac{[1 - P_v(v_l)]^{l-1} [P_v(v_l) - P_v(v_k)]^{k-l-1} [P_v(v_k)]^{\lambda-k}}{(\lambda - k)! (k - l - 1)! (l - 1)!} dv_l dv_k. \end{aligned} \quad (\text{C.23})$$

Note that the integration range  $v_k \leq v_l < \infty$  and integration order was set due to the ordering  $k < l$ . The factor  $\lambda!/[(\lambda - k)!(k - l - 1)!(l - 1)!]$  incorporates all combinations, but excluding the irrelevant ones among the three groups (larger, smaller and values in between). The result of Eq. (C.23) can be generalized to the  $\nu$ -fold density

$$\begin{aligned}
p_{n_1, \dots, n_\nu; \lambda}(z_1, \dots, z_\nu) = & \frac{\lambda!}{\sigma_\epsilon^\nu} \int_{-\infty}^{\infty} \int_{v_1}^{\infty} \cdots \int_{v_{\nu-1}}^{\infty} \left[ \prod_{k=1}^{\nu} p_z(z_k) \phi\left(\frac{v_k - sz_k}{\sigma_\epsilon}\right) \right] \\
& \times \left[ \prod_{k=0}^{\nu} \frac{[P_v(v_{k+1}) - P_v(v_k)]^{n_k - n_{k+1} - 1}}{(n_k - n_{k+1} - 1)!} \right] dv_\nu \cdots dv_1.
\end{aligned} \tag{C.24}$$

Setting  $s = 1$  yields the density derived by Arnold [2, p. 136, D.12], while  $s = -1$  will require some additional steps in order to apply Arnold's equations. The expectation value of sum (C.21) has to be calculated over density (C.24) by integrating over all contributions of  $z$

$$\begin{aligned}
E[S_P] = & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sum_{n_1=\nu}^{\mu} \sum_{n_2=\nu-1}^{n_1-1} \cdots \sum_{n_\nu=1}^{n_{\nu-1}-1} \pi_P(z_{n_1; \lambda}, \dots, z_{n_\nu; \lambda}) \\
& \times p_{n_1, \dots, n_\nu; \lambda}(z_1, \dots, z_\nu) dz_\nu \cdots dz_1.
\end{aligned} \tag{C.25}$$

Now one can insert density (C.24) and replace the  $\nu$ -fold sum by an additional integral, similar to Eq. (2.15). The problem is reformulated as

$$E[S_P] = \frac{\mu!}{(\mu - \nu)!} \frac{\lambda - \mu}{\sqrt{2\pi}} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} [\Phi(x)]^{\lambda - \mu - 1} [1 - \Phi(x)]^{\mu - \nu} I_P(x) dx, \tag{C.26}$$

with  $I_P(x)$  and  $s = \pm 1$  giving

$$\begin{aligned}
I_P(x) \stackrel{s=1}{=} & \frac{1}{\sigma_\epsilon^\nu} \int_{P_v^{-1}(\Phi(x))}^{\infty} \int_{v_1}^{\infty} \cdots \int_{v_{\nu-1}}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \pi_P(z_1, \dots, z_\nu) \\
& \left[ \prod_{k=1}^{\nu} p_z(z_k) \phi\left(\frac{v_k - z_k}{\sigma_\epsilon}\right) \right] dz_\nu \cdots dz_1 dv_\nu \cdots dv_1
\end{aligned} \tag{C.27}$$

$$\begin{aligned}
I_P(x) \stackrel{s=-1}{=} & \frac{1}{\sigma_\epsilon^\nu} \int_{P_v^{-1}(\Phi(x))}^{\infty} \int_{v_1}^{\infty} \cdots \int_{v_{\nu-1}}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \pi_P(z_1, \dots, z_\nu) \\
& \left[ \prod_{k=1}^{\nu} p_z(z_k) \phi\left(\frac{v_k + z_k}{\sigma_\epsilon}\right) \right] dz_\nu \cdots dz_1 dv_\nu \cdots dv_1.
\end{aligned} \tag{C.28}$$

By exchanging the sums with an integral, an intermediate result in [2, p. 138] yields the upper integration bound  $P_v(v_1)$ , see also Identity 1 [2, p.113] explaining the transformation. After substitution and exchanging orders of integration, the expression  $P_v^{-1}(\Phi(x))$  is the lower integration bound of  $v_1$  in  $I_P(x)$ , which is the reason for deriving Eq. (C.19).

The subsequent derivations by Arnold aim to find an approximate solution to the integration (C.27). As Eq. (C.28) differs only by the sign within the transition density, Arnold's results can be reused by transforming (C.28) into the form of (C.27).

Starting with (C.28) the substitution  $z = -z$  is introduced for all  $\nu$  terms and the switched integration bounds and minus sign of the differential cancel

each other. The function arguments change according to

$$I_P(x) \stackrel{s=-1}{=} \frac{1}{\sigma_\epsilon^\nu} \int_{P_v^{-1}(\Phi(x))}^{\infty} \int_{v_1}^{\infty} \cdots \int_{v_{\nu-1}}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \pi_P(-z_1, \dots, -z_\nu) \\ \left[ \prod_{k=1}^{\nu} p_z(-z_k) \phi\left(\frac{v_k - z_k}{\sigma_\epsilon}\right) \right] dz_\nu \cdots dz_1 dv_\nu \cdots dv_1. \quad (\text{C.29})$$

The density  $p_z(-z_k)$  was derived in Eq. (C.4) and yields a negative sign for the resulting skewness  $\gamma_1 = -\kappa_3/\kappa_2^{3/2}$ . The resulting distribution function  $P_v(v)$  due to density (C.7) is given in (C.18) and its quantile function  $P_v^{-1}$  in (C.19). Both yield a negative sign for the resulting skewness  $\tilde{\gamma}_1 = -\kappa_3/(\kappa_2 + \sigma_\epsilon^2)^{3/2}$ , see relations (C.12).

Considering the permutation  $\pi_P(-z_1, \dots, -z_\nu)$  with negative arguments one can construct a simple example. Assuming  $P = (1, 1, 1)$  the permutation with positive and negative arguments, respectively, yields

$$\begin{aligned} \pi_{(1,1,1)}(z_1, z_2, z_3) &= z_1 z_2 z_3, \quad \text{with } z_3 \text{ largest,} \\ \pi_{(1,1,1)}(-z_1, -z_2, -z_3) &= (-z_1)(-z_2)(-z_3), \quad \text{with } (-z_3) \text{ smallest,} \\ &= (-1)^3 z_1 z_2 z_3 \end{aligned} \quad (\text{C.30})$$

Therefore, for a generic vector  $P = (\cdot)$  we have

$$\pi_P(-z_1, \dots, -z_\nu) = (-1)^{\|P\|_1} \pi_P(z_1, \dots, z_\nu). \quad (\text{C.31})$$

Using the results (C.4) and (C.31) expression (C.29) changes to

$$I_P(x) \stackrel{s=-1}{=} \frac{1}{\sigma_\epsilon^\nu} \int_{P_v^{-1}(\Phi(x))}^{\infty} \int_{v_1}^{\infty} \cdots \int_{v_{\nu-1}}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (-1)^{\|P\|_1} \pi_P(z_1, \dots, z_\nu) \\ \left[ \prod_{k=1}^{\nu} p_{-z}(z_k) \phi\left(\frac{v_k - z_k}{\sigma_\epsilon}\right) \right] dz_\nu \cdots dz_1 dv_\nu \cdots dv_1. \quad (\text{C.32})$$

As a conclusion, the obtained result (C.32) has the form of (C.27), where the quantities  $p_{-z}(z_k)$  and  $P_v^{-1}(\Phi(x))$  differ only by the sign of the skewness value  $\kappa_3$  (and therefore  $\gamma_1$ ), neglecting higher orders. Depending on  $P$  the sign of the permutation operator may also change. Still, the same solution method can be applied to both equations with  $s = \pm 1$ .

The final result of Arnold's derivation including the modification due to sign of  $z$  is summarized in Eq. (3.17).

## Appendix D Identities

**Identity 1** For real parameters  $a$  and  $b$  we have

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{1}{2}t^2} \Phi(at+b) dt \\ = \Phi\left(\frac{b}{(1+a^2)^{1/2}}\right) - \frac{1}{\sqrt{2\pi}} \frac{a^2 b}{(1+a^2)^{3/2}} e^{-\frac{1}{2}\frac{b^2}{1+a^2}}. \end{aligned} \quad (\text{D.1})$$

**Proof** It is shown starting from the known identity [5, p.330, A.9]

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{1}{2}t^2} e^{-\frac{1}{2}(at+b)^2} dt = \frac{1+a^2 + a^2 b^2}{(1+a^2)^{5/2}} e^{-\frac{1}{2}\frac{b^2}{1+a^2}}. \quad (\text{D.2})$$

Both sides can be integrated with respect to  $b$ , such that

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{1}{2}t^2} \int_{-\infty}^{b'} e^{-\frac{1}{2}(at+b)^2} db dt \\ = \int_{-\infty}^{b'} \left[ \frac{1+a^2}{(1+a^2)^{5/2}} + \frac{a^2 b^2}{(1+a^2)^{5/2}} \right] e^{-\frac{1}{2}\frac{b^2}{1+a^2}} db \end{aligned} \quad (\text{D.3})$$

Integration of left-hand side yields simply

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{1}{2}t^2} \int_{-\infty}^{b'} e^{-\frac{1}{2}(at+b)^2} db dt = \int_{-\infty}^{\infty} t^2 e^{-\frac{1}{2}t^2} \Phi(at+b') dt, \quad (\text{D.4})$$

which is the left side of (D.1) by renaming  $b = b'$ , up to constant  $1/\sqrt{2\pi}$ . Considering the right-hand side the first term yields

$$\frac{1}{(1+a^2)^{3/2}} \int_{-\infty}^{b'} e^{-\frac{1}{2}\frac{b^2}{1+a^2}} db = \frac{\sqrt{2\pi}}{1+a^2} \Phi\left(\frac{b'}{(1+a^2)^{1/2}}\right). \quad (\text{D.5})$$

For the second term of *rhs* following integral is used

$$\int x e^{-\frac{1}{2}\frac{x^2}{s^2}} dx = \int x \frac{s^2}{x} e^{-y} dy = -s^2 e^{-y} = -s^2 e^{-\frac{1}{2}\frac{x^2}{s^2}}, \quad (\text{D.6})$$

using the substitution  $y = \frac{x^2}{2s^2}$  with  $dx = s^2 dy/x$ . The second term of *rhs* of (D.3) is partially integrated using (D.6), such that

$$\begin{aligned} & \frac{a^2}{(1+a^2)^{5/2}} \int_{-\infty}^{b'} b \left[ b e^{-\frac{1}{2}\frac{b^2}{1+a^2}} \right] db \\ &= \frac{a^2}{(1+a^2)^{5/2}} \left\{ \left[ -b(1+a^2) e^{-\frac{1}{2}\frac{b^2}{1+a^2}} \right]_{-\infty}^{b'} + \int_{-\infty}^{b'} (1+a^2) e^{-\frac{1}{2}\frac{b^2}{1+a^2}} db \right\} \quad (\text{D.7}) \\ &= -\frac{a^2 b'}{(1+a^2)^{3/2}} e^{-\frac{1}{2}\frac{b'^2}{1+a^2}} + a^2 \frac{\sqrt{2\pi}}{1+a^2} \Phi\left(\frac{b'}{(1+a^2)^{1/2}}\right). \end{aligned}$$

Adding results (D.5) and (D.7) for the *rhs*, renaming  $b = b'$  and dividing by  $\sqrt{2\pi}$  yields the result (D.1). The results can be verified by differentiating (D.1) with respect to  $b$  and showing that (D.2) is obtained again.

## Appendix E Additional Plots

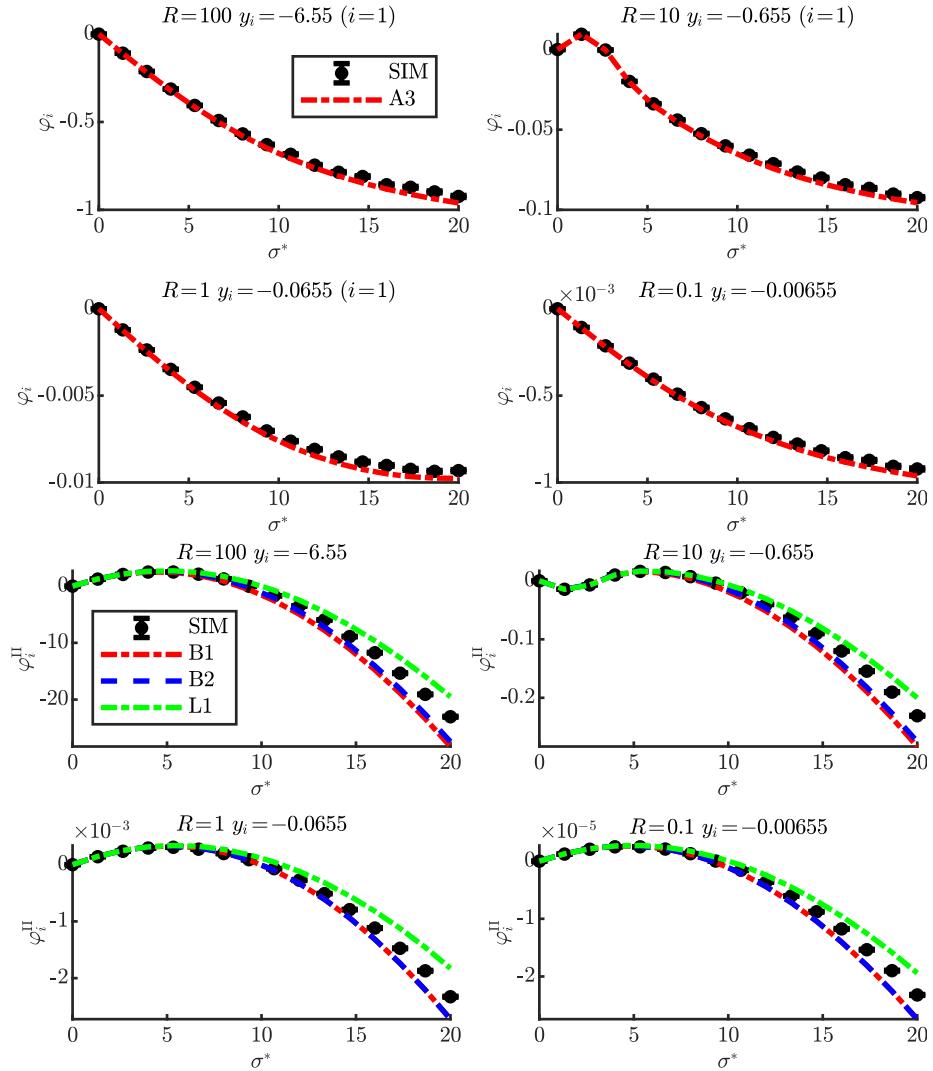


Figure 63: Additional plot of Sec. 3.5 for (10/10, 40)-ES with  $N = 100$ ,  $A = 1$  and  $\alpha = 2\pi$ .

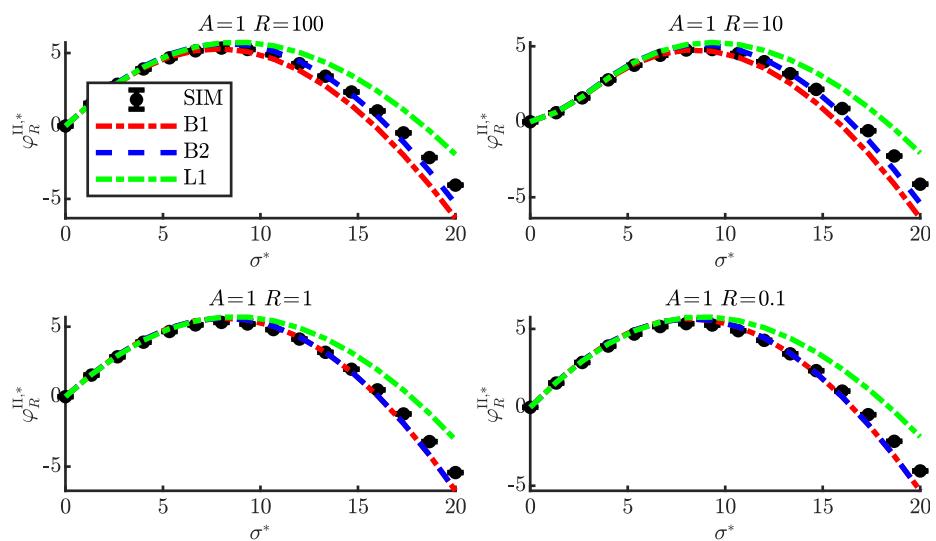


Figure 64: Additional plot of Sec. 3.5 for  $(10/10, 40)$ -ES with  $N = 100$ ,  $A = 1$  and  $\alpha = 2\pi$ .

## References

- [1] M. Abramowitz and I. A. Stegun. *Pocketbook of Mathematical Functions*. Verlag Harri Deutsch, Thun, 1984.
- [2] D.V. Arnold. *Noisy Optimization with Evolution Strategies*. Kluwer Academic Publishers, Dordrecht, 2002.
- [3] D.V. Arnold and H.-G. Beyer. Performance Analysis of Evolution Strategies with Multi-Recombination in High-Dimensional  $\mathbb{R}^N$ -Search Spaces Disturbed by Noise. *Theoretical Computer Science*, 289:629–647, 2002.
- [4] H.-G. Beyer. Toward a Theory of Evolution Strategies: Some Asymptotical Results from the  $(1, + \lambda)$ -Theory. *Evolutionary Computation*, 1(2):165–188, 1993.
- [5] H.-G. Beyer. *The Theory of Evolution Strategies*. Natural Computing Series. Springer, Heidelberg, 2001. DOI: 10.1007/978-3-662-04378-3.
- [6] H.-G. Beyer and A. Melkozerov. The Dynamics of Self-Adaptive Multi-Recombinant Evolution Strategies on the General Ellipsoid Model. *IEEE Transactions on Evolutionary Computation*, 18(5):764–778, 2014. DOI: 10.1109/TEVC.2013.2283968.
- [7] J.A. Gubner. *Probability and Random Processes for Electrical and Computer Engineers*. Cambridge University Press, 2006.