Technical Report

Progress Analysis of a Multi-Recombinative Evolution Strategy on the Highly Multimodal Rastrigin Function

Amir Omeradzic¹ and Hans-Georg Beyer¹

amir.omeradzic@fhv.at (🖂), hans-georg.beyer@fhv.at

¹Vorarlberg University of Applied Sciences, Research Center Business Informatics, Hochschulstraße 1, 6850 Dornbirn, Austria

Abstract

A first and second order progress rate analysis was conducted for the intermediate multi-recombinative Evolution Strategy $(\mu/\mu_I, \lambda)$ -ES with isotropic scale-invariant mutations on the highly multimodal Rastrigin test function. Closed-form analytic solutions for the progress rates are obtained in the limit of large dimensionality and large populations. The first order results are able to model the one-generation progress including local attraction phenomena. Furthermore, a second order progress rate is derived yielding additional correction terms and further improving the progress model. The obtained results are compared to simulations and show good agreement, even for moderately large populations and dimensionality. The progress rates are applied within a dynamical systems approach, which models the evolution using difference equations. The obtained dynamics are compared to real averaged optimizations runs and yield good agreement. The results improve further when dimensionality and population size are increased. Local and global convergence is investigated within given model showing that large mutations are needed to maximize the probability of global convergence, which comes at the expense of efficiency. An outlook regarding future research goals is provided.

Keywords: Evolution strategy, Progress rate, Global optimization, Rastrigin function

1 Introduction

The theoretical analysis of the performance of Evolution Strategies (ES) [5] optimizing functions $f(\mathbf{y})$ in real-valued N-dimensional search spaces $\mathbf{y} \in \mathbb{R}^N$ is a challenge. This is due to the probabilistic nature of these algorithms allowing up to now the dynamic progress analysis only on simple test functions such as the sphere model, the ridge function class [9], and the ellipsoid model [4]. These test functions are simple w.r.t. their optimization landscape (also referred to as fitness landscape) in that they have at most one optimizer (i.e. the location \mathbf{y} of the optimum). Analyzing the dynamical behavior of ES on more complex and multimodal test functions appears to be even more demanding. However, ES and other evolutionary algorithms are especially designated to optimize such problems. There is empirical evidence that ES are able to globally optimize highly multimodal optimization problems [7] with in N exponential number of local optima. The question arises how and when these ES are able to locate the global optimizer. It is the long term goal to find conditions the ES must fulfill to not get trapped in the vast amount of local optimizers. Ideally, a theoretical analysis should provide the answers regarding the success probability P_S (of locating the global optimum) depending on the ES parameters such as the population size λ and the test function to be optimized. Furthermore, one is interested in the computational complexity of the optimization process.

One approach successfully applied to the analysis of the ES-performance on simple unimodal test functions mentioned above is the dynamical systems approach [3] which is based on progress rate analysis. The progress rate is a measure of expected positional change in search space between two generations depending on location, strategy and test function parameters. It will be shown in this paper that this approach can be extended to the highly multimodal Rastrigin test function

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

where $\mathbf{y} \in \mathbb{R}^N$, with oscillation amplitude A and frequency parameter α . Depending on A and α a finite number of local minima M can be observed for each component *i*. Therefore the overall number of local

minima is scaling as M^N posing a highly multimodal minimization problem with the global optimizer located at $\hat{\mathbf{y}} = \mathbf{0}$. An exemplary optimization landscape of the Rastrigin function is shown in Fig. 1.



Figure 1: The heat map shows the optimization landscape for A = 1, $\alpha = 2\pi$, and N = 2. The global minimizer located at the origin (dark blue) is surrounded by multiple local minima. On the right side the same parameter set is shown for N=1. For increasing y the oscillation contribution is decreasing.

The remarkable observation is that ES – unlike classical nonlinear optimization algorithms (e.g. BFGS) – do not follow the local gradient or Hessian ending in one of the $M^N - 1$ local optimizers. That is, ES perform a rather global search. A deeper understanding of this behavior is still missing. Recently, attempts have been made to analyze the problem from the viewpoint of relaxation using kernel smoothing [10]. However, the sampling process needed to transform the original problem into a convex optimization problem is still lacking a link to the ES.

In this paper the scale-invariant $(\mu/\mu_I, \lambda)$ -ES is analyzed, see Alg. 1. It consists of a population of μ parents and λ offspring $(\mu < \lambda)$. The selection (truncation) ratio is denoted by $\vartheta = \mu/\lambda$ and will be an essential quantity for the progress rate results in the limit of infinite populations. For each generation increment $g \to g + 1$, isotropic Gaussian mutations $\mathbf{x} \sim \sigma \mathcal{N}(0, \mathbf{1})$ with mutation strength σ are applied to the parent recombinant $\mathbf{y}^{(g)}$ and λ candidate solutions are obtained. The best $m = 1, ..., \mu$ individuals are selected as the new parent generation. Then, using intermediate recombination with equal weights the update $\mathbf{y}^{(g+1)}$ is obtained. In the following, subscript " $m; \lambda$ " can be read as the m-th best solution out of λ candidate solutions. Algorithm 1 operates with given constant normalized mutation σ^* using the spherical normalization with $\|\mathbf{y}^{(g)}\| = R^{(g)}$ as

$$\sigma^* = \frac{\sigma^{(g)}N}{\|\mathbf{y}^{(g)}\|} = \frac{\sigma^{(g)}N}{R^{(g)}}.$$
(2)

This property ensures scale invariance and therefore global convergence of the algorithm, as the mutation strength $\sigma^{(g)}$ decreases if and only if the residual distance $R^{(g)}$ decreases. The quantity σ^* is unknown during black-box optimizations, but it is very useful for theoretical investigations to obtain scale-invariant mutations strengths.

Algorithm 1 ($\mu/\mu_I, \lambda$)-ES with constant σ^*

```
1: q \leftarrow \overline{0}
   2: \mathbf{y}^{(0)} \leftarrow \mathbf{y}^{(\text{init})}
   3: \sigma^{(0)} \leftarrow \sigma^* \| \mathbf{y}^{(0)} \| / N
   4: repeat
                          for l = 1, ..., \lambda do
   5:
                        \begin{split} \tilde{\mathbf{x}}_l &\leftarrow \sigma^{(g)} \mathcal{N}_l(0, \mathbf{1}) \\ \tilde{\mathbf{y}}_l &\leftarrow \mathbf{y}^{(g)} + \tilde{\mathbf{x}}_l \\ \tilde{f}_l &\leftarrow f(\tilde{\mathbf{y}}_l) \\ \end{split}  end for
   6:
   7:
   8:
   9:
                        (\tilde{\mathbf{y}}_{1;\lambda},\ldots,\tilde{\mathbf{y}}_{\mu;\lambda}) \leftarrow \operatorname{sort}\left(\tilde{\mathbf{y}} \text{ w.r.t. ascending } \tilde{f}\right)
10:
                        \mathbf{y}^{(g+1)} \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\mathbf{y}}_{m;\lambda}\sigma^{(g+1)} \leftarrow \sigma^* \| \mathbf{y}^{(g+1)} \| / N
11:
12:
                          g \leftarrow g + 1
13:
14: until termination criterion
```

The remainder of this paper is organized as follows. In the next Section the local performance measures

will be introduced being the basis for both the progress rate analysis and the dynamical systems approach. Section 3 is devoted to the determination and evaluation of the first order progress rate. Section 4 sketches the derivation of the second order progress rate. The details of the lengthy derivations are provided in the Appendices A, B, C, and D. Section 5 uses the local performance measures to establish the evolution equations that govern the dynamical behavior of the ES. Experiments will be presented to show the usefulness of the approach. In the final Section 6 conclusions will be drawn and being based on open problems the further research direction will be outlined.

2 Local Performance Measures and Quality Gain Distribution

The performance of an ES between two generations can be evaluated in both fitness and search space. The quality gain $Q_{\mathbf{y}}(\mathbf{x})$ of fitness f at a position $\mathbf{y}^{(g)}$ due to an isotropic mutation $\mathbf{x} \sim \sigma \mathcal{N}(0, \mathbf{1})$ is defined as

$$Q_{\mathbf{y}}(\mathbf{x}) \coloneqq f\left(\mathbf{y}^{(g)} + \mathbf{x}\right) - f\left(\mathbf{y}^{(g)}\right),\tag{3}$$

and yields in the case of fitness improvement (minimization considered) a negative value $Q_{\mathbf{y}} < 0$. For independent components it is decomposed using $(Q_{\mathbf{y}}(\mathbf{x}))_i = Q_i$ and $(\mathbf{y})_i = y_i$ as

$$Q_{\mathbf{y}}(\mathbf{x}) = \sum_{i=1}^{N} Q_i(x_i) = \sum_{i=1}^{N} f_i \Big(y_i^{(g)} + x_i \Big) - f_i \Big(y_i^{(g)} \Big).$$
(4)

That is, the quality gain corresponds to the difference between fitness values before and after the mutation application. A probabilistic model for the distribution of quality values will be presented below. It will be important for the subsequent progress rate derivations, as selection is based on fitness values.

Analyzing the progress towards the optimizer in search space, the first order progress rate on the Rastrigin function has already been investigated in [11] as a first approach. In this paper, a new approach is presented which significantly improves the prediction quality. The first order progress rate between two generations for the parental component y_i is defined as

$$\varphi_i \coloneqq \mathbf{E}\left[y_i^{(g)} - y_i^{(g+1)} \,\big|\, \mathbf{y}^{(g)}, \sigma^{(g)}\,\right],\tag{5}$$

given parental position $\mathbf{y}^{(g)}$ and mutation strength $\sigma^{(g)}$ at generation g. It is a measure of positional difference in search space and defined to be positive if $y_i^{(g)} > y_i^{(g+1)}$. Assuming (w.l.o.g.) that $y_i^{(g)} > 0$ and $y_i^{(g+1)} > 0$, $\varphi_i > 0$ corresponds to progressing towards the optimizer $\hat{y}_i = 0$. This assumption is only valid as long as the sign of $y_i^{(g+1)}$ does not change, i.e., for small mutations compared to the residual distance. Therefore φ_i has limited applicability when studying the convergence behavior in the vicinity of the optimizer. As has been shown in [4] regarding the performance analysis on the ellipsoid model, a second order progress rate is needed. It is defined as

$$\varphi_i^{\mathrm{II}} \coloneqq \mathrm{E}\left[\left(y_i^{(g)}\right)^2 - \left(y_i^{(g+1)}\right)^2 \, \big| \, \mathbf{y}^{(g)}, \sigma^{(g)} \, \right]. \tag{6}$$

Squaring the positions yields $\varphi_i^{\text{II}} > 0$ independent of the sign, if the distance to $\hat{y}_i = 0$ decreases. Additionally, the derivation will yield expressions containing a progress gain and loss part, which is necessary for a more accurate model of convergence.

Both progress rates will be expressed using integral equations for the expected values and approximations will be necessary to find closed-form solutions. In a second step the progress rates can be applied within difference equations to model the expected dynamics over many generations in order to investigate the global convergence behavior.

Quality Gain Distribution The selection of individuals is based on the attained fitness values. The quality gain measures the fitness change according to (3). When the progress rate of an ES is modeled, the quality gain cumulative distribution function (CDF) $P_Q(q)$ is needed as a function of the location and mutation strength. Obtaining an exact CDF for $Q_{\mathbf{y}}(\mathbf{x})$ is not feasible at this point and also not necessary. Instead, the approach is to model the sum $Q_{\mathbf{y}}(\mathbf{x}) = \sum_{i=1}^{N} Q_i$ over N independently distributed variables as normally distributed in the limit $N \to \infty$ due to the Central Limit Theorem (CLT) as

$$Q_{\mathbf{y}}(\mathbf{x}) = \sum_{i=1}^{N} Q_i(x_i) \overset{N \to \infty}{\sim} \mathcal{N}(\mathrm{E}\left[Q_{\mathbf{y}}(\mathbf{x})\right], \mathrm{Var}\left[Q_{\mathbf{y}}(\mathbf{x})\right]).$$
(7)

This is justified by Lyapunov's condition for the CLT provided that there are no dominating terms within the sum. The following quantities are defined as abbreviations

$$E_Q \coloneqq \operatorname{E}\left[Q_{\mathbf{y}}(\mathbf{x})\right] = \sum_{i=1}^{N} \operatorname{E}\left[Q_i\right]$$
(8)

$$D_Q^2 \coloneqq \operatorname{Var}\left[Q_{\mathbf{y}}(\mathbf{x})\right] = \sum_{i=1}^N \operatorname{Var}\left[Q_i\right].$$
(9)

Now Eq. (7) can be rewritten using a standardized random variate Z

$$Z = \frac{Q_{\mathbf{y}}(\mathbf{x}) - E_Q}{D_Q} \overset{N \to \infty}{\sim} \mathcal{N}(0, 1).$$
(10)

Therefore variate Z converges to a standard normal distribution, such that $P_Q(q)$ and $p_Q(q) = \frac{dP_Q(q)}{dq}$ can be given in terms of the CDF of the normal distribution $\Phi(\cdot)$, and the probability density function (PDF) of the normal distribution according to

$$P_Q(q) = \Phi\left(\frac{q - E_Q}{D_Q}\right) \tag{11}$$

$$p_Q(q) = \frac{1}{\sqrt{2\pi}D_Q} \exp\left[-\frac{1}{2}\left(\frac{q-E_Q}{D_Q}\right)^2\right]$$
(12)

Within the normal approximation (11) 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 p using

$$q = E_Q + D_Q \Phi^{-1}(p). (13)$$

For the derivation of the *i*-th component progress rate the conditional distribution function $P_Q(q|x_i)$ of the quality gain is needed for a given component x_i . In this case expected value and variance are given by

$$E_{Q|x_i} \coloneqq \operatorname{E}\left[Q_{\mathbf{y}}(\mathbf{x})|x_i\right] = Q_i(x_i) + \sum_{j \neq i} \operatorname{E}\left[Q_j\right]$$
(14)

$$D_i^2 \coloneqq \operatorname{Var}\left[Q_{\mathbf{y}}(\mathbf{x})|x_i\right] = \sum_{j \neq i} \operatorname{Var}\left[Q_j\right],\tag{15}$$

where the sum $j \neq i$ is taken for fixed *i* over the remaining N - 1 components. Analogously applying (10), the conditional CDF and PDF read

$$P_Q(q|x_i) = \Phi\left(\frac{q - E_{Q|x_i}}{D_i}\right) \tag{16}$$

$$p_Q(q|x_i) = \frac{1}{\sqrt{2\pi}D_i} \exp\left[-\frac{1}{2}\left(\frac{q - E_Q|x_i}{D_i}\right)^2\right]$$
(17)

Having defined the distribution functions of the quality gain, the quantities $E[Q_i]$ and $Var[Q_i]$ remain to be determined. As the components are independent, it is sufficient to consider a single component and then perform the summation. Starting from definition (4), one can evaluate the quality gain of a single component $Q_i(x_i)$. After applying trigonometric identity $\cos(\alpha(y_i + x_i)) = \cos(\alpha y_i)\cos(\alpha x_i) - \sin(\alpha y_i)\sin(\alpha x_i)$, one gets

$$Q_i(x_i) = f_i(y_i + x_i) - f_i(y_i)$$
(18)

$$= x_i^2 + 2y_i x_i + A\cos\left(\alpha y_i\right) - A\cos\left(\alpha y_i\right)\cos\left(\alpha x_i\right) + A\sin\left(\alpha y_i\right)\sin\left(\alpha x_i\right),\tag{19}$$

of which $E[Q_i]$ is evaluated for $x_i \sim \mathcal{N}(0, \sigma^2)$ in Eq. (A.1) yielding result (A.6) as

$$E_Q = \sum_{i=1}^{N} \sigma^2 + A\cos\left(\alpha y_i\right) \left(1 - e^{-\frac{(\alpha\sigma)^2}{2}}\right).$$
(20)

As $\operatorname{Var}[Q_i] = \operatorname{E}[Q_i^2] - \operatorname{E}[Q_i]^2$, the respective squared quantities also need to be evaluated, see Appendix (A.2). The final result (A.7) reads

$$D_Q^2 = \sum_{i=1}^N 2\sigma^4 + 4y_i^2 \sigma^2 + \frac{A^2}{2} \left[1 - e^{-(\alpha\sigma)^2} \right] \left[1 - \cos(2\alpha y_i) e^{-(\alpha\sigma)^2} \right] + 2A\alpha\sigma^2 e^{-\frac{1}{2}(\alpha\sigma)^2} \left[\alpha\sigma^2 \cos(\alpha y_i) + 2y_i \sin(\alpha y_i) \right].$$
(21)

The quantities $E_{Q|x_i}$ from (14) and from D_i^2 (15) are given analogously by summing over N-1 components. Expressions E_Q and D_Q could be inserted into (11), and $E_{Q|x_i}$ with $Q_i(x_i)$ and D_i into (16). However, it is omitted at this point for better readability.

As an important remark, expression (18) can be linearized w.r.t. mutation x_i to obtain analytically solvable progress rate integrals, see also discussion after Eq. (31). Taylor-expanding f_i around y_i for small x_i gives $f_i(y_i + x_i) = f_i(y_i) + \frac{\partial f_i}{\partial y_i} x_i + O(x_i^2)$, such that after setting $f'_i \coloneqq \frac{\partial f_i}{\partial y_i}$ and evaluating the derivative one has

$$Q_{i}(x_{i}) = f_{i}(y_{i} + x_{i}) - f_{i}(y_{i}) = f'_{i}x_{i} + O(x_{i}^{2})$$

= $(2y_{i} + \alpha A \sin(\alpha y_{i}))x_{i} + O(x_{i}^{2}) = (k_{i} + d_{i})x_{i} + O(x_{i}^{2}),$ (22)

with following definitions applied to (22)

$$f'_i \coloneqq k_i + d_i, \quad \text{with} \quad k_i \coloneqq 2y_i, \quad \text{and} \quad d_i \coloneqq \alpha A \sin(\alpha y_i).$$
 (23)

Component k_i is the derivative of the quadratic term y_i^2 , cf. Eq. (1), which follows the global quadratic structure of the function. Conversely, derivative d_i follows the local oscillation, such that it will be very important for the model of local attraction during the progress rate derivations in Secs. 3 and 4.

3 First Order Progress Rate

While the first order progress rate (5) does not suffice to completely describe the convergence behavior of the ES on Rastrigin, it is a necessary step in the calculation of the second order progress rate in Sec. 4.

Given definition (5) and the parental location $\mathbf{y}^{(g)}$, one has to find the expected value over the *i*-component location $\mathbf{E}\left[y_i^{(g+1)}\right]$. The positional update $\mathbf{y}^{(g)} \to \mathbf{y}^{(g+1)}$ performed by the ES is realized by consecutively applying mutation, selection and recombination, see also Alg. 1, such that one can write

$$\mathbf{y}^{(g+1)} = \frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{y}^{(g)} + \mathbf{x}_{m;\lambda}) = \mathbf{y}^{(g)} + \frac{1}{\mu} \sum_{m=1}^{\mu} \mathbf{x}_{m;\lambda},$$
(24)

where $\mathbf{x}_{m;\lambda}$ denotes the mutation vector of the *m*-th best offspring after selection. Considering the *i*-th component of Eq. (24), abbreviating the mutation component as $x_{m;\lambda} \coloneqq (\mathbf{x}_{m;\lambda})_i$, and taking the expected value thereof yields

$$\mathbf{E}\left[y_{i}^{(g+1)}|\mathbf{y}^{(g)},\sigma^{(g)}\right] = y_{i}^{(g)} + \frac{1}{\mu}\sum_{m=1}^{\mu}\mathbf{E}\left[x_{m;\lambda}|\mathbf{y}^{(g)},\sigma^{(g)}\right].$$
(25)

The progress rate can therefore be evaluated by inserting (25) into (5) giving

$$\varphi_i = -\frac{1}{\mu} \sum_{m=1}^{\mu} \mathbf{E} \left[x_{m;\lambda} | \mathbf{y}^{(g)}, \sigma^{(g)} \right].$$
(26)

From now on the conditional dependency on $\mathbf{y}^{(g)}$ and $\sigma^{(g)}$ will be implicitly assumed as given for better readability of the equations. The expected value of the *i*-th mutation component $x_{m;\lambda}$ after selection can be expressed as an integration over the order statistic density $p_{m;\lambda}(x_i)$ of the *m*-th best individual, such that

$$\varphi_i = -\frac{1}{\mu} \sum_{m=1}^{\mu} \int_{-\infty}^{\infty} x_i \ p_{m;\lambda}(x_i) \,\mathrm{d}x_i \,. \tag{27}$$

The subsequent task will be to derive the density $p_{m;\lambda}$ as a function of mutation and quality gain distributions.

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

$$p_x(x_i) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2} \left(\frac{x_i}{\sigma}\right)^2\right].$$
(28)

Given mutation x_i (and implicitly position \mathbf{y}), a random quality gain value Q is distributed according to a conditional probability density $p_Q(q|x_i)$, see Eq. (17). Given that the *m*-th best individual attains a quality gain within [q, q + dq], there must be 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 one has $\frac{\lambda!}{(m-1)!(\lambda-m)!}$, 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

$$p_{Q;m;\lambda}(q|x_i) = \frac{\lambda!}{(m-1)!(\lambda-m)!} p_Q(q|x_i) P_Q(q)^{m-1} [1 - P_Q(q)]^{\lambda-m}.$$
(29)

By integrating (29) over all attainable quality gain values $q \in [q_l, q_u]$, one arrives at the density

$$p_{m;\lambda}(x_i) = p_x(x_i) \frac{\lambda!}{(m-1)!(\lambda-m)!} \int_{q_i}^{q_u} p_Q(q|x_i) P_Q(q)^{m-1} [1 - P_Q(q)]^{\lambda-m} \,\mathrm{d}q \,. \tag{30}$$

Inserting the order statistic density from (30) into the progress rate (27), one obtains the intermediate result

$$\varphi_i = -\frac{1}{\mu} \sum_{m=1}^{\mu} \frac{\lambda!}{(m-1)!(\lambda-m)!} \int_{-\infty}^{\infty} x_i p_x(x_i) \int_{q_i}^{q_u} p_Q(q|x_i) P_Q(q)^{m-1} [1 - P_Q(q)]^{\lambda-m} \, \mathrm{d}q \, \mathrm{d}x_i \,. \tag{31}$$

A few important remarks can be made regarding Eq. (31). A closed-form analytic solution cannot be obtained without applying further approximations. It can be approached in an analogous way to the φ_i derivation of the Ellipsoid in [8] to obtain a solution in terms of the well-known progress coefficient $c_{\mu/\mu,\lambda}$ [3, p. 216]. However, a closed-form solution with this approach requires a linear relation of Q_i w.r.t. x_i , see relation (22). The effect of a linearized quality gain on the progress rate of the Rastrigin function was already studied in [11] and showed that the progress due to local attraction is not modeled correctly, as the oscillation terms have to be either dropped or linearized for small x_i .

Therefore a different approach is followed here assuming the infinite population limit, an approach which was applied within the analysis of functions with noise-induced multi-modality [6]. Additionally, large dimensionality N will be assumed to simplify the resulting equations. The approach will yield correction terms including the effects of the trigonometric terms from (19), in contrast to only taking linearized terms from (22). Furthermore, an expression for the so-called asymptotic generalized progress coefficients $e_{\vartheta}^{a,b}$ is derived in Appendix B giving

$$e_{\vartheta}^{a,b} = \left[\frac{\mathrm{e}^{-\frac{1}{2}\left[\Phi^{-1}(\vartheta)\right]^{2}}}{\sqrt{2\pi}\vartheta}\right]^{a} \left[-\Phi^{-1}(\vartheta)\right]^{b}.$$
(32)

These are characteristic coefficients describing the progress in the limit $(\mu, \lambda) \to \infty$ with constant truncation ratio $\vartheta = \mu/\lambda$, and are related to the generalized progress coefficients [3, Eq. (5.112)]. These coefficients will reappear during the derivation of φ_i and φ_i^{II} .

The details of the following derivations are in Appendix C.1. Starting from (31) the sum is transformed into an additional integral, which will enable the application of the large population identity. Furthermore, the integration orders are exchanged and the quality gain integral is solved. Then, applying identity (B.1) one arrives at the following intermediate result derived in Appendix C.1, Eq. (C.15), giving

$$\varphi_i \stackrel{\mu \to \infty}{\simeq} -\frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i p_x(x_i) P_Q \left(P_Q^{-1}(\vartheta) \big| x_i \right) \mathrm{d}x_i \,. \tag{33}$$

Now the normal approximation of the quality gain distribution is applied by using the conditional normal distribution function $\Phi\left(\frac{q-E_Q|x_i}{D_i}\right)$, see (16), and its inverse $q = E_Q + D_Q \Phi^{-1}(p)$, see (13), evaluated at probability $p = \vartheta$. Additionally, $E_{Q|x_i}$ from (14) with Q_i -result (19) is reformulated according to

$$E_{Q|x_i} = Q_i(x_i) + \sum_{j \neq i} \mathbb{E}\left[Q_j\right] = k_i x_i + \delta_i(x_i) + E_i, \tag{34}$$

using $k_i = 2y_i$, $\delta_i(x_i) = x_i^2 + A\cos(\alpha y_i) - A\cos(\alpha y_i)\cos(\alpha x_i) + A\sin(\alpha y_i)\sin(\alpha x_i)$, and $E_i = \sum_{j \neq i} \mathbb{E}[Q_i]$ as abbreviations. The expression $\delta_i(x_i)$ can be regarded as a non-linear perturbation term w.r.t. term $k_i x_i$. Now integral (33) yields

$$\varphi_i = -\frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i p_x(x_i) \Phi\left(\frac{E_Q + D_Q \Phi^{-1}(\vartheta) - (k_i x_i + \delta_i(x_i) + E_i)}{D_i}\right) \mathrm{d}x_i \,. \tag{35}$$

A closed form solution of (35) cannot be obtained with $\Phi(\delta_i(x_i))$ containing non-linear terms in x_i . Therefore, an approximate solution is necessary by decomposing the argument as $\Phi(g(x_i) + h(x_i))$ into $g(x_i)$ being a linear function, and $h(x_i)$ a non-linear function of x_i , which are defined in (C.20) and (C.21), respectively. Then, a Taylor expansion of $\Phi(g+h)$ is done up to first order assuming small perturbations h scaling with $1/\sqrt{N}$, such that

$$\Phi(g+h) = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{\mathrm{d}^k \Phi}{\mathrm{d}g^k} h^k = \Phi(g) + \frac{\mathrm{e}^{-\frac{1}{2}g^2}}{\sqrt{2\pi}} h + O\left(\frac{1}{N}\right).$$
(36)

The resulting two integrals are solvable analytically. Within the results, the function $g(x_i)$ being dependent on k_i will give progress contributions of the sphere function, while $h(x_i)$ can be regarded as a perturbation of the sphere containing A and α dependencies. The two integrals after Taylor expansion are given in (C.25) and are successively solved in the Appendix. The resulting expressions can be further simplified assuming large dimensionality N in (C.45) and afterwards the asymptotic progress coefficient is recovered.

Finally, the result is given in terms of derivative components k_i and d_i , see (23), quality gain variance D_Q^2 from (20) and $c_{\vartheta} = e_{\vartheta}^{1,0}$ from (32) by applying the limits $N \to \infty$ and $(\mu, \lambda) \to \infty$ (constant $\vartheta = \mu/\lambda$) as

$$\varphi_{i} = c_{\vartheta} \frac{\sigma^{2}}{D_{Q}} \left(k_{i} + e^{-\frac{1}{2}(\alpha\sigma)^{2}} d_{i} \right)$$

$$= c_{\vartheta} \frac{\sigma^{2}}{D_{Q}} \left(2y_{i} + e^{-\frac{1}{2}(\alpha\sigma)^{2}} \alpha A \sin\left(\alpha y_{i}\right) \right).$$
(37)

The expressions for c_{ϑ} and D_Q were not inserted in order to improve readability. Result (37) shows very interesting properties compared to [11, Eq. (26)], where a linearized quality gain approximation resulted in

$$\varphi_{i,\text{lin}} = c_{\mu/\mu,\lambda} \frac{\sigma^2}{\sqrt{(f'_i\sigma)^2 + D_i^2}} f'_i.$$
(38)

First note that the progress coefficient was replaced by its asymptotic form $c_{\mu/\mu,\lambda} \to c_{\vartheta}$. The difference for the variance terms in the denominators of (37) and (38) is negligible for large N with $D_Q^2 \approx D_i^2 + (f'_i \sigma)^2$, see also (C.45). However, the most notable difference lies between the derivative term $f'_i = k_i + d_i$, see definition (23), and the newly obtained term $k_i + e^{-\frac{1}{2}(\alpha\sigma)^2}d_i$. It contains an unchanged sphere-dependent term k_i and an exponentially decaying Rastrigin-specific term d_i . This characteristic form will be discussed in the subsequent part.

One-generation experiments At this point one-generation experiments can be performed and compared to result (37). To this end, a random position vector \mathbf{y} is initialized isotropically with $\|\mathbf{y}\| = R$ given some residual distance R. Then, repeated simulations are performed and quantity (5) is averaged over 10^6 trials. The issue with the choice of R is that the "interesting" region with high density of local minima scales with N, such that a relation R(N) is needed. The following argumentation can be given. Assuming w.l.o.g. $\mathbf{y} > \mathbf{0}$ and that all components of the parental position are at some given local minimum denoted by $\hat{y}^{(j)}$. Index j identifies the local attractor along the half-axis, e.g. $j \in \{1, 2, 3\}$ in Fig. 1 on the right side. For N = 1 one has $\mathbf{y} = [\hat{y}^{(j)}]$ and therefore $R^2 = (\hat{y}^{(j)})^2$. Having N components at the same j-th local minimum yields $\mathbf{y} = [\hat{y}^{(j)}, \hat{y}^{(j)}, ..., \hat{y}^{(j)}]$, such that $R^2 = N(\hat{y}^{(j)})^2$. A scaling $R = O(\sqrt{N})$ is therefore needed to stay within a certain region of local attractors when N is increased.

The progress rates of two exemplary components for a single experiment are shown in Fig. 2. For both plots $\sigma \in [0, 1]$ was chosen in order to investigate the effects of the oscillation as $\alpha = 2\pi$. On the left, one observes enhanced progress for moderate σ -values due to local attraction, as both local and global attractor are aligned along the same direction. On the right, there is negative progress for moderate σ , as the local attractor is driving the ES away from the global attractor. For larger σ , the overall spherical shape is dominating and both exhibit positive progress. A decomposition of the progress rate in terms of $\varphi_i = \varphi_i(d_i, k_i)|_{k_i=0} + \varphi_i(d_i, k_i)|_{d_i=0}$ is displayed in Fig. 2. It shows the large-scale behavior of the k_i -term, dashed cyan, and limited range of the d_i -term, dotted green. As $k_i = \partial(y_i^2)/\partial y_i$, its progress term models the global quadratic structure of Rastrigin, see derivative definitions (23). The second term $e^{-\frac{1}{2}(\alpha\sigma)^2}d_i$ models the Rastrigin-specific local oscillation having limited range depending on the mutation strength σ (or α). By defining scale-invariant mutations using (2) with $\sigma = \sigma^* R/N$, the oscillations vanish via $e^{-\frac{1}{2}(\alpha\sigma^* R/N)^2}$ for large residual distance R, where the sphere function is recovered. This model significantly improves the progress rate formula (38) from [11].

As a note, changing one of the fitness parameters A or α directly affects Fig. 2. The change of amplitude A rescales both the (local) peak and dip heights accordingly, increasing the effects of local attraction for larger

A. Increasing frequency α has mostly short-range effects as the overall range is reduced due to suppression via $e^{-\frac{1}{2}(\alpha\sigma)^2}$ of (37). In the subsequent parts, the progress rate is investigated for A = 1 and $\alpha = 2\pi$ as an example.

In Figs. 3 and 4 the progress rate is evaluated over scale-invariant σ^* for two different N-values and population sizes. One can see that the approximation quality improves for larger N and μ , as expected from the applied approximations. The overall agreement between simulation and approximation is good for larger and smaller residual distances R, see left and right plots, respectively. The σ^* -range was chosen large enough, such that the progress rate of the corresponding sphere function [3, Eq. (6.54)] reaches negative values due to mutations being too large. This boundary directly translates to Rastrigin, as the global structure is the same. However, due to φ_i being first order, no negative progress occurs even for large σ^* . Therefore the second order progress rate φ_i^{II} needs to be derived in Sec. 4, where loss terms will provide additional correction terms.



Figure 2: One-generation experiments with (10/10, 40)-ES for N = 20, A = 10, $\alpha = 2\pi$ at randomly chosen $\|\mathbf{y}\| = R = \sqrt{N}$. The results for φ_i of Eq. (37) are shown for the exemplary components i = 2 with $y_i = 1.16$ (left) and i = 12 with $y_i = 0.78$ (right) to illustrate the effect of local attraction on the progress rate. The plots show additionally Eq. (37) with $\varphi_i(k_i) = \varphi_i(d_i, k_i)|_{d_i=0}$ [cyan, dashed] and $\varphi_i(d_i) = \varphi_i(d_i, k_i)|_{k_i=0}$ [green, dotted], respectively.



Figure 3: Progress rate φ_i as a function of the normalized mutation σ^* for (10/10, 40)-ES with N = 20, A = 1, $\alpha = 2\pi$, at two residual distances $R = 10\sqrt{N}$ with $y_i = 11.6$ (left) and $R = 0.1\sqrt{N}$ with $y_i = 0.116$ (right). As in Fig. 2, black dots depict the simulation, while the red line shows result (37). The error bars are very small and therefore not visible.



Figure 4: Progress rate φ_i as a function of the normalized mutation σ^* for (100/100, 200)-ES with N = 100, A = 1, $\alpha = 2\pi$, at two residual distances $R = 10\sqrt{N}$ with $y_i = 11.9$ (left) and $R = 0.1\sqrt{N}$ with $y_i = 0.119$ (right). The approximation quality improves compared to Fig. 3 and shows very good agreement.

4 Second Order Progress Rate

The second order progress rate (6) requires the evaluation of $\mathbf{E}\left[(y_i^{(g+1)})^2\right]$. Starting again with (24), referring to the *i*-th component and squaring yields

$$\left(y_i^{(g+1)}\right)^2 = \left(y_i^{(g)} + \frac{1}{\mu} \sum_{m=1}^{\mu} x_{m;\lambda}\right)^2$$

$$= \left(y_i^{(g)}\right)^2 + 2y_i^{(g)} \frac{1}{\mu} \sum_{m=1}^{\mu} x_{m;\lambda} + \frac{1}{\mu^2} \left(\sum_{m=1}^{\mu} x_{m;\lambda}\right)^2.$$

$$(39)$$

Squaring the last term can be evaluated by separating the sum into equal and unequal indices

$$\left(\sum_{m=1}^{\mu} x_{m;\lambda}\right)^{2} = \left(\sum_{k=1}^{\mu} x_{k;\lambda}\right) \left(\sum_{l=1}^{\mu} x_{l;\lambda}\right) = \sum_{m=1}^{\mu} (x_{m;\lambda})^{2} + \sum_{k\neq l} x_{k;\lambda} x_{l;\lambda}$$

$$= \sum_{m=1}^{\mu} (x_{m;\lambda})^{2} + 2 \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} x_{k;\lambda} x_{l;\lambda}.$$
(40)

Inserting (40) into (39) and taking the expected value (conditional variables $\mathbf{y}^{(g)}$ and $\sigma^{(g)}$ are implicitly assumed to be given) yields

$$\mathbf{E}\left[\left(y_{i}^{(g+1)}\right)^{2}\right] = \left(y_{i}^{(g)}\right)^{2} + 2y_{i}^{(g)}\frac{1}{\mu}\sum_{m=1}^{\mu}\mathbf{E}\left[x_{m;\lambda}\right] + \frac{1}{\mu^{2}}\sum_{m=1}^{\mu}\mathbf{E}\left[\left(x_{m;\lambda}\right)^{2}\right] + \frac{2}{\mu^{2}}\sum_{l=2}^{\mu}\sum_{k=1}^{l-1}\mathbf{E}\left[x_{k;\lambda}x_{l;\lambda}\right].$$
 (41)

Noting that $\varphi_i = -\frac{1}{\mu} \sum_{m=1}^{\mu} \mathbb{E}[x_{m;\lambda}]$, see Eq. (26), and using (41) in φ_i^{II} -definition (6) yields the second order *i*-th component progress rate

$$\varphi_i^{\text{II}} = 2y_i^{(g)}\varphi_i - \frac{1}{\mu^2}E^{(2)} - \frac{2}{\mu^2}E^{(1,1)}, \qquad (42)$$

for which the two following expected values need to be determined

$$\frac{1}{\mu^2} E^{(2)} \coloneqq \frac{1}{\mu^2} \sum_{m=1}^{\mu} E\left[(x_{m;\lambda})^2 \right]$$
(43)

$$\frac{1}{\mu^2} E^{(1,1)} \coloneqq \frac{1}{\mu^2} \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} \mathbb{E} \left[x_{k;\lambda} x_{l;\lambda} \right].$$
(44)

The solution of (43) is presented in Appendix C.2.1. It uses order statistic density (30) for the *m*-th individual, large population identity (B.1), and the expansion of the normal CDF (36) up to first order. The resulting two integrations can then be solved analytically and within the limit $N \to \infty$ the results simplify significantly. Finally, the asymptotic generalized progress coefficient definition (32) is applied and the result yields

$$\frac{1}{\mu^2} E^{(2)} = \frac{\sigma^2}{\mu} \Biggl\{ 1 + e_{\vartheta}^{1,1} \frac{(2y_i)^2 \sigma^2}{D_Q^2} - \frac{c_{\vartheta}}{D_Q} \Biggl[3\sigma^2 + A\cos\left(\alpha y_i\right) \left(1 - e^{-\frac{1}{2}(\alpha\sigma)^2} + \alpha^2 \sigma^2 e^{-\frac{1}{2}(\alpha\sigma)^2} \right) \Biggr] \Biggr\}.$$
(45)

The solution of the second expected value (44) is shown in Appendix C.2.2. For this purpose, the joint probability density $p_{k,l;\lambda}$ for two individuals k and l is derived assuming $1 \leq k < l \leq \lambda$, i.e., k yielding a smaller (better) quality value than l out of λ individuals. Then, the resulting five-fold integration is restructured and successively solved. Interestingly, the final integral can be related to (squared) first order progress integral (33), which is a remarkably simple result for $E^{(1,1)}$ in terms of φ_i . Thereafter, result (37) for φ_i can be inserted and $c_{\vartheta}^2 = e_{\vartheta}^{2,0}$, see (32), is identified. One gets

$$\frac{1}{\mu^2} E^{(1,1)} \simeq \frac{1}{2} \frac{\mu - 1}{\mu} \varphi_i^2
= \frac{1}{2} \frac{\sigma^2}{\mu} (\mu - 1) e_{\vartheta}^{2,0} \frac{\sigma^2}{D_Q^2} \left(2y_i + e^{-\frac{1}{2}(\alpha\sigma)^2} \alpha A \sin(\alpha y_i) \right)^2.$$
(46)

Using expressions (45) and (46), the result for the second order progress rate (42) yields

$$\varphi_{i}^{\mathrm{II}} = c_{\vartheta} \frac{\sigma^{2}}{D_{Q}} \left(4y_{i}^{2} + \mathrm{e}^{-\frac{1}{2}(\alpha\sigma)^{2}} 2\alpha Ay_{i} \sin\left(\alpha y_{i}\right) \right)
- \frac{\sigma^{2}}{\mu} \left\{ 1 + e_{\vartheta}^{1,1} \frac{(2y_{i})^{2}\sigma^{2}}{D_{Q}^{2}} - \frac{c_{\vartheta}}{D_{Q}} \left[3\sigma^{2} + A\cos\left(\alpha y_{i}\right) \left(1 - \mathrm{e}^{-\frac{1}{2}(\alpha\sigma)^{2}} + \alpha^{2}\sigma^{2}\mathrm{e}^{-\frac{1}{2}(\alpha\sigma)^{2}} \right) \right] + (\mu - 1)e_{\vartheta}^{2,0} \frac{\sigma^{2}}{D_{Q}^{2}} \left(2y_{i} + \mathrm{e}^{-\frac{1}{2}(\alpha\sigma)^{2}} \alpha A\sin\left(\alpha y_{i}\right) \right)^{2} \right\}.$$
(47)

For future investigations of the convergence and step-size adaption properties of the $(\mu/\mu_I, \lambda)$ -ES, a simpler expression than (47) will be needed. As the limit $N \to \infty$ was assumed throughout the derivation, it can be applied once more to simplify the loss term within $\{\cdot\}$ of (47) significantly. The motivation stems also from φ_i^{II} -investigations done on the ellipsoid function [4], where the loss term could be successfully simplified without large degradation of approximation quality.

The analysis of the loss-term N-scaling behavior is shown in Appendix C.2.3. It yields that in $\{\cdot\}$ of (47) all terms except "1" are vanishing for large N as long as $\mu(N)$ scales sub-linearly. Theoretical results concerning population sizing, i.e., choosing the necessary $\mu(N)$ to achieve high global convergence probability P_S (success probability), are not available at this point. It is one of the main future goals of the current research project. Note that treating μ as a constant is also not satisfactory, since for large N an increase of μ is necessary to maintain a high success rate on a highly multimodal problem. However, experimental investigations on the Rastrigin function including step-size adaption have shown a sub-linear relation, which validates the approximation.

Finally, expression (47) is simplified applying the argumentation above to intermediate result (C.124), such that one obtains the second order progress rate for the limits $N \to \infty$ and $(\mu, \lambda) \to \infty$ with constant $\vartheta = \mu/\lambda$ and $\mu = o(N)$ as

$$\varphi_i^{\rm II} = c_\vartheta \frac{\sigma^2}{D_Q} \left(4y_i^2 + e^{-\frac{1}{2}(\alpha\sigma)^2} 2\alpha A y_i \sin\left(\alpha y_i\right) \right) - \frac{\sigma^2}{\mu}.$$
(48)

The expressions for $c_{\vartheta} = e_{\vartheta}^{1,0}$ in (32) and D_Q in (21) were not inserted to improve readability. The result of (48) can be mapped to the Evolutionary Progress Principle [3] as it contains a progress gain and loss term, respectively. Here, the gain part scales with c_{ϑ} and it is a y_i -dependent expression. Hence, depending on the sign of $y_i \sin(\alpha y_i)$ it may also yield negative contributions due to local attraction moving the ES away from the global optimizer, cf. Fig. 2. The loss term $-\sigma^2/\mu$ is characteristic for intermediate recombination. It introduces significant loss for large σ , but can be decreased using a larger μ due to recombination effects.

One-generation experiments Results of one-generation experiments are presented in Figs. 5 and 6 by evaluating (6) over 10^6 trials (black dots with vanishing error bars) and comparing with the obtained approximations. The red dash-dotted line is showing simplified result (48), while the blue dashed line is showing (47). The positions **y** were initialized randomly (given *R*) and kept constant over all repetitions.

First thing to note is that the loss terms now predict negative progress for large σ^* , which was not the case for φ_i . The approximation quality is relatively consistent varying R (left and right, respectively) and improves significantly for larger N and μ in Fig. 6, which was expected. Simplified expression φ_i^{II} from (48) [red, dash-dotted] yields good results compared to (47) [blue, dashed], with (47) giving slightly better results for smaller σ^* and (48) better results at larger σ^* . This indicates that additional terms of the Taylor expansion (36) would be needed to further improve the results of (47). However, this would make the expression more involved. Furthermore, the results of Fig. 5 are relatively good considering that a rather small population (10/10, 40)-ES was used at low dimensionality N = 20.

One can conclude that (48) yields very good results considering its "simplicity". It will therefore be used in Sec. 5 to investigate the dynamical behavior of the ES. It should be noted that D_Q from (21) is still a relatively complex expression and does not allow for an easy interpretation of the **y**-dependence on the overall variance. However, for $\sigma \to \infty$ (equivalent to $R \to \infty$ setting $\sigma = \sigma^* R/N$ with constant σ^* and N), the exponential factors are vanishing and the term $NA^2/2$ is negligible, see Eqs. (A.8) and (A.9). In this limit the sphere variance $D_Q^2(R) = 4R^2\sigma^2 + 2N\sigma^4$ is recovered, where Rastrigin-specific oscillation effects are vanishing. Investigations to simplify $D_Q^2(\mathbf{y})$ or to provide an *R*-dependent (average) formulation $D_Q^2(R)$ are part of future research. Furthermore, at this point there is no aggregated progress measure over all *N* components, such as the *R*-dependent sphere progress rate. Given some $\mathbf{y}^{(g)}$ one can evaluate all i = 1, ..., Nvalues for φ_i^{II} and obtain a progress vector, but the overall effect on $R^{(g)} \to R^{(g+1)}$ is not known. This will also be part of future research. However, the cumulative effect of all *N* progress rates can be evaluated within a dynamical systems model to be shown in the next chapter.



Figure 5: Second order progress rate φ_i^{II} as a function of σ^* for (10/10, 40)-ES with N = 20, A = 1, $\alpha = 2\pi$, at two residual distances $R = 10\sqrt{N}$ with $y_i = 11.6$ (left) and $R = 0.1\sqrt{N}$ with $y_i = 0.116$ (right).



Figure 6: Second order progress rate φ_i^{II} as a function of σ^* for (100/100, 200)-ES with N = 100, A = 1, $\alpha = 2\pi$, at two residual distances $R = 10\sqrt{N}$ with $y_i = 11.9$ (left) and $R = 0.1\sqrt{N}$ with $y_i = 0.119$ (right).

5 Evolution Equations

In the previous sections one-generation experiments were conducted and compared against progress rate results (37) and (48). In order to have an aggregated measure over all components and many generations, φ_i and φ_i^{II} will be used within the so-called evolution equations and compared to real optimization runs of Alg. 1. Using this method the (mean) global convergence behavior can be investigated.

Given definitions for first and second order progress (5) and (6), the expressions can be reformulated as stochastic iterative mappings between two generations $g \rightarrow g + 1$ according to

$$y_i^{(g+1)} = y_i^{(g)} - \varphi_i(\sigma^{(g)}, \mathbf{y}^{(g)}) + \epsilon^{(1)}(\sigma^{(g)}, \mathbf{y}^{(g)})$$
(49)

$$\left(y_{i}^{(g+1)}\right)^{2} = \left(y_{i}^{(g)}\right)^{2} - \varphi_{i}^{\mathrm{II}}(\sigma^{(g)}, \mathbf{y}^{(g)}) + \epsilon^{(2)}(\sigma^{(g)}, \mathbf{y}^{(g)}).$$
(50)

The two terms $\epsilon^{(1)}$ and $\epsilon^{(2)}$ can be interpreted as fluctuations w.r.t. the expected values (provided by φ_i and φ_i^{II}), such that condition $\text{E}\left[\epsilon^{(1)}\right] = 0 = \text{E}\left[\epsilon^{(2)}\right]$ is necessary. However, the exact transition densities for $g \to g + 1$ are not known at this point. In principle, they could be approximated using a finite number of higher order moments (or cumulants) to model the fluctuations [3, Ch. 7]. At this point, for a first study of the progress rate results on the dynamics, the fluctuations are neglected setting $\epsilon^{(1)} = 0 = \epsilon^{(2)}$. One arrives at the (deterministic) equations describing the mean-value dynamics of the position coordinates

$$y_i^{(g+1)} = y_i^{(g)} - \varphi_i(\sigma^{(g)}, \mathbf{y}^{(g)})$$
(51)

$$\left(y_{i}^{(g+1)}\right)^{2} = \left(y_{i}^{(g)}\right)^{2} - \varphi_{i}^{\mathrm{II}}(\sigma^{(g)}, \mathbf{y}^{(g)}),$$
(52)

with constant normalized mutation strength σ^* from Eq. (2) giving

$$\sigma^{(g)} = \sigma^* \left\| \mathbf{y}^{(g)} \right\| / N.$$
(53)

Two important issues need to be discussed. Firstly, the positional iterations are defined for 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 $R = \|\mathbf{y}^{(g)}\|$. Secondly, for the evaluation of φ_i^{II} being a function of $\mathbf{y}^{(g)}$, the square root of the components $(y_i^{(g)})^2$ has to be taken after iteration giving two solutions $\pm y_i^{(g)}$. As the corresponding terms of φ_i^{II} and $D_Q^2(\mathbf{y})$ are even in $y_i^{(g)}$, both solutions are equivalent.

Dynamic experiments In the following, the deterministic iterations (51) and (52) using mutation rescaling (53) are compared to real optimization runs. For the initialization, $\mathbf{y}^{(0)}$ is chosen randomly such that $\|\mathbf{y}^{(0)}\| = R^{(0)}$ for a given $R^{(0)}$. The starting position is kept constant for consecutive runs of the same experiment. For the magnitude of $R^{(0)}$ it is ensured that the strategy starts far enough away from the local minima landscape. Given Fig. 1 with A = 1, the farthermost local minimizer is at $y_i \approx 3$ with resulting $R = 3\sqrt{N}$ for N-components, such that $R^{(0)} = 20\sqrt{N} > 3\sqrt{N}$ is chosen.

Considering the choice of σ^* one observes in experiments that larger mutation strengths increase the success probability P_S of individual trials to converge to the global optimizer. This is due to the fact that large steps tend to overcome local attraction more easily. However, this comes at the expense of efficiency, since large steps are often overshooting the global optimizer. Therefore in Fig. 7, σ^* is chosen larger than the sphere-optimal value $\hat{\sigma}^*_{sph}$, which can be obtained by numerically solving [3, Eq. (6.54)], but small enough to prevent negative progress. The aim was to obtain $P_S \approx 1$.

In order to aggregate the $R^{(g)}$ -data of multiple dynamic experiments, the median has shown to be a suitable measure of central tendency. The main issue is that due to fluctuations the $R^{(g)}$ -values of distinct runs may differ by orders of magnitude, such that the mean yields biased results due to a skewed distribution. The median is more suitable in this case and a more stable measure.

In Fig. 7 one can observe three phases within the dynamics. First, linear convergence is observed for large $R^{(g)}$ -values, where the sphere function dominates. Then, a slow down is observed due to increasing effects of local attraction. For small $R^{(g)}$ -values, the ES descends into the global attractor basin and linear convergence can be observed again. One can see that the φ_i -iteration (blue) shows by far too much progress compared to φ_i^{II} -iteration. This is due to the first order model, which does not include loss terms and overestimates the progress significantly, see also discussion of result (37). Iteration via φ_i^{II} (red) shows good results compared to the median curve, especially for larger μ and N (right plot). Better agreement for large populations is also due to reduced fluctuation effects, which were neglected at the beginning of Sec. 5.

In Fig. 8 the effect of reduced σ^* is investigated, which increases the probability of local convergence. The left plot shows $\sigma^* = 5$ with no globally converging runs, as the mutation strength is too low. Technically, for constant σ^* there is no local convergence as the algorithm never stops if R is not decreasing. Still, the experiments are stopped after some g-threshold is reached. The stagnating behavior of the ES around some $R^{(g)}$ can be illustrated using Fig. 2. For $\sigma = 0.2$ one has $\sigma^* = \sigma N/R \approx 0.9$, which is small compared to $\hat{\sigma}^*_{\text{sph}} \approx 5.7$. Both left and right progress components of Fig. 2 are significantly influenced by the local attraction region at $\sigma = 0.2$. While some components may be improved (positive value left), others are worsened (negative value right) resulting in a cumulative effect of $R^{(g)}$ -stagnation. One way out can be increasing σ (or equivalently σ^*). However, the local minima landscape changes with changing R and arbitrary σ^* -increase is not possible. Stagnation may appear at higher σ^* for different $R^{(g)}$ -values, if the Rastrigin function is hard to optimize for given ES parameters. For an active step-size adaption, changing σ appropriately – without converging locally – also poses a major challenge.

In the central plot of Fig. 8 roughly half of the runs are globally converging at increased $\sigma^* = \hat{\sigma}^*_{\rm sph}$. In this case the deterministic iteration follows a single converging path, as no fluctuations are modeled. The residual distance of the locally converging runs is reduced compared to $\sigma^* = 5$. Note that the convergence speed is faster (steeper negative slope) for the globally converging runs compared to $\sigma^* = 30$ of Fig. 7 due to sphere-optimal $\hat{\sigma}^*_{\rm sph}$. However, this comes with the disadvantage of a lower P_S , as more trials are converging runs. Again, the ES convergence speed is faster closer to $\hat{\sigma}^*_{\rm sph}$, but shows a slightly reduced P_S -value. The overall prediction quality of the iteration is good and the results affirm the expectation, that large mutations are favorable to maximize P_S on the Rastrigin function.

To confirm the expectation that the approximation quality increases further for larger μ and N, experiments are shown in Fig. 9. First thing to notice is that positional fluctuations of the ES trials decrease further, such that nearly all runs follow a similar path in R-space. This is related to the intermediate recombination, see Eq. (24), as position $\mathbf{y}^{(g+1)}$ is obtained by averaging over a large number of individuals. One can see good agreement, but for the left plot there is still some room for improvement. This is related to truncation ratio $\vartheta = 1/4$, such that the Taylor expansion point in Eq. (36) via function $g(x_i)$ is shifted by $\Phi^{-1}(\vartheta)$. For $\vartheta = 1/2$ and even larger N and μ (right plot), very good agreement is observed.



Figure 7: Comparison of real optimization runs with mean value dynamics using progress rates φ_i via (49) [dashed blue] and φ_i^{II} via (50) [dash-dotted red]. Gray lines show all 100 successful runs of Alg. 1 and the black line shows the median thereof. The left plot shows (10/10, 40)-ES for N = 20 with $\sigma^* = 7$ ($\hat{\sigma}^*_{\text{sph}} = 5.7$) and the right one (100/100, 200)-ES for N = 100 with $\sigma^* = 30$ ($\hat{\sigma}^*_{\text{sph}} = 18.3$). For both experiments A = 1, and $\alpha = 2\pi$ are chosen. The resulting success probability $P_S = 1$.



Figure 8: Variation of σ^* for (100/100, 200)-ES for N = 100, A = 1, and $\alpha = 2\pi$. From left to right $\sigma^* = \{5, 18.3, 25\}$, with $\hat{\sigma}^*_{sph} = 18.3$, and success rate $P_S = \{0, 0.45, 0.97\}$. The experiment with $\sigma^* = 30$ ($P_S = 1$) was already shown in Fig. 7. Globally converging trials are shown in gray, and non-converging runs in light-orange. The median is taken over the globally converging runs, except for the left plot where none exist, in which the median over all unsuccessful runs is taken.



Figure 9: The left plot shows (1000/1000, 4000)-ES with $\sigma^* = 110$ for N = 1000, A = 1, and $\alpha = 2\pi$. The right plot shows (10000/10000, 20000)-ES with $\sigma^* = 400$ for N = 10000 (same α and A), evaluated for 50 trials due to CPU resource restrictions.

6 Conclusion and Outlook

In this paper the full first and second order progress rate analysis of the $(\mu/\mu_I, \lambda)$ -ES has been presented. In order to obtain closed-form expressions for φ_i and φ_i^{II} it was necessary to consider the asymptotic $N \to \infty$ and the large population assumption. While the latter does not present a serious issue because large populations are needed to ensure global convergence, it was the key prerequisite to simplify the expected value integrals. As the experiments have shown, the approximation quality of the progress rate expressions is rather good even for N as small as 20 and comparably small populations of $\mu = 10$. For larger N and μ the approximation quality improves further, as expected. The first order progress rate result is able to model the local attraction effects on the Rastrigin function. This is a very important step, as all subsequent investigations in this paper are based on φ_i -results. The second order progress rate derivation was needed to obtain additional loss terms improving the progress model further, especially for larger mutation strengths and close to the global optimizer.

Using the progress rate expressions, the dynamics of the evolution process have been investigated. There is a good agreement between the iterations and real ES-runs using median aggregation of the residual distance R to the global optimizer. As has been shown, depending on the choice of the normalized mutation strength, one can model global as well as local convergence behavior. Additionally, one observes a trade-off between efficiency and success rate, as relatively large mutations have to be chosen to maximize the success probability. The conducted experiments assume scale-invariance, i.e., the mutation strength is controlled by the residual distance R. This is in contrast to the full self-adaptive ES where σ evolves during the ES run either by mutative self-adaptation, cumulative step-size adaptation or Meta-ES.

The incorporation of the self-adaptation process will be the next step completing the analysis of the $(\mu/\mu_I, \lambda)$ -ES on Rastrigin. To this end, the self-adaptation response (SAR) function must be derived. Combining N progress rates with the SAR function yields N + 1 evolution equations. In order to get manageable expressions that allow for analytic population sizing and expected runtime investigations, additional aggregation is needed. One possible approach would be the aggregation of individual parental y_i components into the parental distance R modeling the expected progress as a function of the residual distance. This would reduce the number of evolution equations to two and making further analytic treatment more accessible.

Finally, the presented approach to model the ES-dynamics is based on mean value considerations. That is, fluctuations are not considered so far. Whether the approach presented can be extended to allow for the calculation of the global attractor convergence probability as a function of strategy and fitness parameters remains an open question.

Acknowledgments

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

Appendix A Expected Value and Variance of Quality Gain

The derivation of D_i^2 was already published in the supplementary material of [11]. For the sake of completeness, its derivation is sketched again and the results are applied to obtain E_Q and D_Q^2 (and analogously $E_{Q|x_i}$ and D_i^2) needed in Sec. 2. The results are also needed during the progress rate derivation. As a remark, terms containing moments of $x \sim \mathcal{N}(0, \sigma^2)$, i.e. $\mathbb{E}[x^k]$ with $k \geq 1$, are silently evaluated as they are assumed to be widely known.

Starting with $E[Q_i]$ applied to (19) yields the intermediate result

$$\mathbf{E}\left[Q_{i}\right] = \sigma^{2} - A\cos\left(\alpha y_{i}\right) \mathbf{E}\left[\cos\left(\alpha x_{i}\right)\right],\tag{A.1}$$

where odd powers of $E[x^k] = 0$, which also yields $E[\sin(\alpha x_i)] = 0$. Analogously, evaluating $Var[Q_i]$ yields after collecting terms

$$\operatorname{Var} \left[Q_{i}\right] = \operatorname{E} \left[Q_{i}^{2}\right] - \operatorname{E} \left[Q_{i}\right]^{2}$$

$$= 2\sigma^{4} + 4y_{i}^{2}\sigma^{2} + A^{2}\sin^{2}\left(\alpha y_{i}\right)\operatorname{Var}\left[\sin\left(\alpha x_{i}\right)\right]$$

$$+ A^{2}\cos^{2}\left(\alpha y_{i}\right)\operatorname{Var}\left[\cos\left(\alpha x_{i}\right)\right] - 2A\cos\left(\alpha y_{i}\right)\operatorname{E}\left[x^{2}\cos\left(\alpha x_{i}\right)\right]$$

$$+ 2A\sigma^{2}\cos\left(\alpha y_{i}\right)\operatorname{E}\left[\cos\left(\alpha x_{i}\right)\right] + 4Ay_{i}\sin\left(\alpha y_{i}\right)\operatorname{E}\left[x\sin\left(\alpha x_{i}\right)\right].$$
(A.2)

In the general case, expectations of the form $E[x^k \cos \alpha x]$ and $E[x^k \sin \alpha x]$ for $k \ge 0$ can be obtained by using the definition of the characteristic function χ of $x \sim \mathcal{N}(\mu, \sigma^2)$ and its known result [1]

$$\chi_x(\alpha) = \mathbf{E}\left[e^{i\alpha x}\right] = e^{i\alpha\mu - \frac{1}{2}\alpha^2\sigma^2} = e^{-\frac{1}{2}\alpha^2\sigma^2} [\cos\left(\alpha\mu\right) + i\sin\left(\alpha\mu\right)].$$
(A.3)

Now the k-th derivatives with respect to α can be applied to both sides

$$\frac{\mathrm{d}^{k}}{\mathrm{d}\alpha^{k}} \operatorname{E}\left[\mathrm{e}^{i\alpha x}\right] = \operatorname{E}\left[\frac{\mathrm{d}^{k}}{\mathrm{d}\alpha^{k}}\mathrm{e}^{i\alpha x}\right] = \operatorname{E}\left[\frac{\mathrm{d}^{k}}{\mathrm{d}\alpha^{k}}\cos(\alpha x)\right] + i\operatorname{E}\left[\frac{\mathrm{d}^{k}}{\mathrm{d}\alpha^{k}}\sin(\alpha x)\right]$$
$$\stackrel{!}{=}\frac{\mathrm{d}^{k}}{\mathrm{d}\alpha^{k}}\left[\mathrm{e}^{-\frac{(\alpha\sigma)^{2}}{2}}\left[\cos\left(\alpha\mu\right) + i\sin\left(\alpha\mu\right)\right]\right],$$
(A.4)

such that corresponding real and imaginary parts can be identified. Given $\mu = 0$ for $k = \{0, 1, 2\}$ the required expectations of trigonometric terms can be derived. Additionally, trigonometric identities $\cos^2(x) = 1/2 + \cos(2x)/2$ and $\sin^2(x) = 1/2 - \cos(2x)/2$ are used. The results are

$$E \left[\cos \left(\alpha x \right) \right] = e^{-\frac{(\alpha \sigma)^2}{2}}, \quad E \left[\cos^2 \left(\alpha x \right) \right] = \frac{1}{2} + \frac{1}{2} e^{-\frac{(2\alpha \sigma)^2}{2}}$$
$$E \left[\sin^2 \left(\alpha x \right) \right] = \frac{1}{2} - \frac{1}{2} e^{-\frac{(2\alpha \sigma)^2}{2}}, \quad E \left[x \sin \left(\alpha x \right) \right] = \alpha \sigma^2 e^{-\frac{(\alpha \sigma)^2}{2}}$$
$$(A.5)$$
$$E \left[x^2 \cos \left(\alpha x \right) \right] = \left(\sigma^2 - \alpha^2 \sigma^4 \right) e^{-\frac{(\alpha \sigma)^2}{2}}, \quad Var \left[(\cdot) \right] = E \left[(\cdot)^2 \right] - E \left[(\cdot) \right]^2.$$

Inserting relations (A.5) into (A.1) and (A.2), summing over all N components and collecting the resulting terms one obtains the expected value and variance of the Rastrigin quality gain as

$$E_Q = \sum_{i=1}^{N} \sigma^2 + A\cos\left(\alpha y_i\right) \left(1 - e^{-\frac{(\alpha\sigma)^2}{2}}\right)$$
(A.6)

$$D_Q^2 = \sum_{i=1}^N 2\sigma^4 + 4y_i^2 \sigma^2 + \frac{A^2}{2} \left[1 - e^{-(\alpha\sigma)^2} \right] \left[1 - \cos(2\alpha y_i) e^{-(\alpha\sigma)^2} \right] + 2A\alpha\sigma^2 e^{-\frac{1}{2}(\alpha\sigma)^2} \left[\alpha\sigma^2 \cos(\alpha y_i) + 2y_i \sin(\alpha y_i) \right].$$
(A.7)

Variance expression (A.7) can be simplified to some extent by first performing the summation over the first two terms applying $\sum_i y_i^2 = R^2$. Additionally, for large σ (or equivalently large R with $\sigma = \sigma^* R/N$ using normalization (2)) the exponential terms are vanishing and the term $NA^2/2$ is negligible. One gets

$$D_Q^2 = 2N\sigma^4 + 4R^2\sigma^2 + \sum_{i=1}^N \frac{A^2}{2} \left[1 - e^{-(\alpha\sigma)^2} \right] \left[1 - \cos(2\alpha y_i) e^{-(\alpha\sigma)^2} \right] + 2A\alpha\sigma^2 e^{-\frac{1}{2}(\alpha\sigma)^2} \left[\alpha\sigma^2 \cos(\alpha y_i) + 2y_i \sin(\alpha y_i) \right]$$
(A.8)

$$\simeq 2N\sigma^4 + 4R^2\sigma^2. \tag{A.9}$$

Expression (A.9) recovers the well-known variance of the sphere function [3, Ch. 4].

Appendix B Large Population Identity

Solving the expected value integrals of progress rates φ_i and φ_i^{II} for large populations requires the evaluation of subsequent integral (B.1). The identity will be proven by Taylor expanding the integrand around the (sharp) maximum of the population-dependent term. It is shown that in the limit of infinitely large populations and constant truncation ratio only the 0-th order term yields relevant contributions. In the end, the obtained result can also be applied to derive the asymptotic generalized progress coefficients in (B.30).

Theorem. Let $\lambda > \mu + 1$ and $\mu > a$ with $a \ge 1$ and $\vartheta = \mu/\lambda$ with $0 < \vartheta < 1$, 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 a function defined and differentiable on (0,1), and let $B(\cdot, \cdot)$ be the beta function. For infinitely large $(\mu, \lambda) \to \infty$ and constant $\vartheta = \mu/\lambda$ the asymptotic equality holds

$$I^{a}_{\mu,\lambda}[f] = \frac{1}{B(\lambda - \mu, \mu)} \int_{0}^{1} f(t) t^{\lambda - \mu - 1} (1 - t)^{\mu - a} dt \simeq \frac{f(1 - \vartheta)}{\vartheta^{a - 1}},$$
(B.1)

with higher order terms vanishing with $O(1/\mu)$ and $O(1/\lambda)$.

Proof. 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,



Figure B.1: Integrand $I(t) = \frac{1}{B(\lambda-\mu,\mu)} t^{\lambda-\mu-1} (1-t)^{\mu-1}$ 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's sharpness increases with growing λ and the factor $1/B(\lambda-\mu,\mu)$ rescales the peak heights.

see also Fig. B.1. This observation suggests performing a Taylor series expansion of the function f(t) around the sharp peak located at \hat{t} defined by

$$\hat{t} = \underset{t \in [0,1]}{\operatorname{argmax}} \left[t^{\lambda - \mu - 1} (1 - t)^{\mu - a} \right].$$
(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

$$\frac{\mathrm{d}}{\mathrm{d}t} [t^{\lambda-\mu-1}(1-t)^{\mu-a}] \stackrel{!}{=} 0 \tag{B.3}$$

$$\Rightarrow \hat{t} = \frac{\lambda - \mu - 1}{\lambda - a - 1} = 1 - \frac{\mu - a}{\lambda - a - 1} = 1 - \frac{\mu(1 - a/\mu)}{\lambda(1 - a/\lambda - 1/\lambda)}.$$
(B.4)

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

$$\lim_{\substack{(\mu,\lambda)\to\infty\\\vartheta=\text{const.}}} \hat{t} = 1 - \vartheta. \tag{B.5}$$

Taylor-expanding f(t) around \hat{t} yields

$$f(t) = \sum_{k=0}^{\infty} \frac{1}{k!} \left. \frac{\partial^k f}{\partial t^k} \right|_{t=\hat{t}} (t-\hat{t})^k, \tag{B.6}$$

such that integral (B.1) is expressed as

$$\begin{split} I^{a}_{\mu,\lambda}[f] &= \frac{1}{\mathcal{B}(\lambda-\mu,\mu)} \int_{0}^{1} \sum_{k=0}^{\infty} \frac{1}{k!} \left. \frac{\partial^{k} f}{\partial t^{k}} \right|_{t=\hat{t}} (t-\hat{t})^{k} t^{\lambda-\mu-1} (1-t)^{\mu-a} \, \mathrm{d}t \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} \left. \frac{\partial^{k} f}{\partial t^{k}} \right|_{t=\hat{t}} \frac{1}{\mathcal{B}(\lambda-\mu,\mu)} \int_{0}^{1} t^{\lambda-\mu-1} (1-t)^{\mu-a} (t-\hat{t})^{k} \, \mathrm{d}t \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} \left. \frac{\partial^{k} f}{\partial t^{k}} \right|_{t=\hat{t}} C^{(k)}. \end{split}$$
(B.7)

The introduced coefficients $C^{(k)}$ are defined as

$$C^{(k)} \coloneqq \frac{1}{\mathrm{B}(\lambda - \mu, \mu)} \int_0^1 t^{\lambda - \mu - 1} (1 - t)^{\mu - a} (t - \hat{t})^k \,\mathrm{d}t \,. \tag{B.8}$$

It will be shown that only the 0-th order coefficient $C^{(0)}$ will yield significant contributions and all higher orders $k \ge 1$ will vanish with $O(1/\lambda)$ for large populations.

Starting with k = 0 the coefficient can be evaluated using relations between beta function B, gamma function Γ and factorial, namely $B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt$, $B(x, y) = \frac{\Gamma(x) \Gamma(y)}{\Gamma(x+y)}$ and $\Gamma(n) = (n-1)!$, such that

$$\begin{split} C^{(0)} &= \frac{1}{\mathcal{B}(\lambda - \mu, \mu)} \int_{0}^{1} t^{\lambda - \mu - 1} (1 - t)^{\mu - a} \, \mathrm{d}t \\ &= \frac{\mathcal{B}(\lambda - \mu, \mu - a + 1)}{\mathcal{B}(\lambda - \mu, \mu)} = \frac{\Gamma(\lambda - \mu) \, \Gamma(\mu - a + 1)}{\Gamma(\lambda - a + 1)} \frac{\Gamma(\lambda)}{\Gamma(\lambda - \mu) \, \Gamma(\mu)} \\ &= \frac{(\lambda - \mu - 1)!(\mu - a)!}{(\lambda - a)!} \frac{(\lambda - 1)!}{(\lambda - \mu - 1)!(\mu - 1)!} \\ &= \frac{(\lambda - 1)!(\mu - a)!}{(\lambda - a)!(\mu - 1)!} = \frac{\prod_{n=1}^{a-1} \lambda - n}{\prod_{n=1}^{a-1} \mu - n} = \prod_{n=1}^{a-1} \frac{\lambda}{\mu} \frac{1 - n/\lambda}{1 - n/\mu} \\ &= \begin{cases} 1 & \text{for } a = 1, \\ \frac{1}{\vartheta^{a-1}} \prod_{n=1}^{a-1} \frac{1 - n/\lambda}{1 - n/\mu}. & \text{for } a > 1. \end{cases} \end{split}$$

It was used that for a > 1 one has

$$\frac{(\lambda-1)!}{(\lambda-a)!} = \prod_{n=1}^{a-1} (\lambda-n) \quad \text{and} \quad \frac{(\mu-a)!}{(\mu-1)!} = \frac{1}{\prod_{n=1}^{a-1} (\mu-n)}.$$
 (B.10)

Therefore the limit yields for any $a \ge 1$ and k = 0

$$\lim_{\substack{(\mu,\lambda)\to\infty\\\vartheta=\text{const.}}} C^{(0)} = \frac{1}{\vartheta^{a-1}},\tag{B.11}$$

with $O(1/\mu)$ and $O(1/\lambda)$. The analysis of $C^{(k)}$ with $k \ge 1$ is slightly more involved. Noting that $(t - \hat{t})^k = (-\hat{t})^k (1 - t/\hat{t})^k$ one has

$$C^{(k)} = \frac{(-\hat{t})^k}{B(\lambda - \mu, \mu)} \int_0^1 t^{\lambda - \mu - 1} (1 - t)^{\mu - a} \left(1 - \frac{t}{\hat{t}}\right)^k \mathrm{d}t \,. \tag{B.12}$$

By applying the binomial theorem the expression can be reformulated

$$C^{(k)} = \frac{(-\hat{t})^k}{B(\lambda - \mu, \mu)} \int_0^1 t^{\lambda - \mu - 1} (1 - t)^{\mu - a} \sum_{i=0}^k \binom{k}{i} 1^{k - i} \left(-\frac{t}{\hat{t}} \right)^i dt$$

$$= (-\hat{t})^k \sum_{i=0}^k \binom{k}{i} (-1)^i \left(\frac{1}{\hat{t}} \right)^i \frac{1}{B(\lambda - \mu, \mu)} \int_0^1 t^i t^{\lambda - \mu - 1} (1 - t)^{\mu - a} dt \qquad (B.13)$$

$$= (-\hat{t})^k \sum_{i=0}^k \binom{k}{i} (-1)^i F_1^{i,a} F_2^{i,a},$$

with additional treatment needed for the factors $F_1^{i,a}$ and $F_2^{i,a}$

$$F_1^{i,a} \coloneqq \left(\frac{1}{\hat{t}}\right)^i \tag{B.14}$$

$$F_2^{i,a} \coloneqq \frac{1}{\mathcal{B}(\lambda - \mu, \mu)} \int_0^1 t^{\lambda - \mu - 1 + i} (1 - t)^{\mu - a} \, \mathrm{d}t \,. \tag{B.15}$$

Factor $F_1^{i,a}$ is easily evaluated using (B.4) and yields

$$F_1^{i,a} = \left(\frac{1}{\hat{t}}\right)^i = \left(\frac{\lambda - a - 1}{\lambda - \mu - 1}\right)^i = \left(\frac{1 - a/\lambda - 1/\lambda}{1 - \vartheta - 1/\lambda}\right)^i.$$
 (B.16)

Factor $F_2^{i,a}$ yields after applying the beta function definition

$$F_{2}^{i,a} = \frac{B(\lambda - \mu + i, \mu - a + 1)}{B(\lambda - \mu, \mu)}$$

= $\frac{(\lambda - \mu - 1 + i)!(\mu - a)!}{(\lambda - a + i)!} \frac{(\lambda - 1)!}{(\lambda - \mu - 1)!(\mu - 1)!}$
= $\frac{(\lambda - \mu - 1 + i)!}{(\lambda - \mu - 1)!} \frac{(\lambda - 1)!(\mu - a)!}{(\lambda - a + i)!(\mu - 1)!}.$ (B.17)

The first ratio of (B.17) yields

$$\frac{(\lambda - \mu - 1 + i)!}{(\lambda - \mu - 1)!} = \prod_{j=1}^{i} (\lambda - \mu - 1 + j) = \lambda^{i} \prod_{j=1}^{i} (1 - \vartheta - 1/\lambda + j/\lambda).$$
(B.18)

For the second ratio of (B.17) one can use (B.10) and write $(\lambda - 1)! = (\lambda - a)! \prod_{n=1}^{a-1} (\lambda - n)$, such that

$$\frac{(\lambda-1)!(\mu-a)!}{(\lambda-a+i)!(\mu-1)!} \stackrel{a\geq 1}{=} \frac{(\lambda-a)!}{(\lambda-a+i)!} \frac{\prod_{n=1}^{a-1}\lambda-n}{\prod_{n=1}^{a-1}\mu-n} = \frac{1}{\prod_{j=1}^{i}(\lambda-a+j)} \frac{\prod_{n=1}^{a-1}\lambda-n}{\prod_{n=1}^{i}\mu-n} = \frac{1}{\lambda^{i}\prod_{j=1}^{i}(1-a/\lambda+j/\lambda)} \frac{\lambda^{a-1}\prod_{n=1}^{a-1}1-n/\lambda}{\mu^{a-1}\prod_{n=1}^{a-1}1-n/\mu}, \qquad (B.19)$$

$$\frac{(\lambda-1)!(\mu-a)!}{(\lambda-a+i)!(\mu-1)!} \stackrel{a=1}{=} \frac{1}{\lambda^{i}\prod_{j=1}^{i}(1-1/\lambda+j/\lambda)}.$$

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 $a \ge 1$

$$F_{2}^{i,a} = \frac{\lambda^{i} \prod_{j=1}^{i} (1 - \vartheta - 1/\lambda + j/\lambda)}{\lambda^{i} \prod_{j=1}^{i} (1 - a/\lambda + j/\lambda)} \frac{\lambda^{a-1} \prod_{n=1}^{a-1} (1 - n/\lambda)}{\mu^{a-1} \prod_{n=1}^{a-1} (1 - n/\mu)}$$

$$= \frac{1}{\vartheta^{a-1}} \prod_{j=1}^{i} \frac{(1 - \vartheta - 1/\lambda + j/\lambda)}{(1 - a/\lambda + j/\lambda)} \prod_{n=1}^{a-1} \frac{(1 - n/\lambda)}{(1 - n/\mu)}.$$
(B.20)

Finally the result for $C^{(k)}$ from (B.13) can be evaluated using (B.16) and (B.20)

$$C^{(k)} = (-\hat{t})^{k} \sum_{i=0}^{k} \binom{k}{i} (-1)^{i} F_{1}^{i,a} F_{2}^{i,a}$$

$$= (-\hat{t})^{k} \sum_{i=0}^{k} \binom{k}{i} (-1)^{i} \frac{(1-a/\lambda-1/\lambda)^{i}}{(1-\vartheta-1/\lambda)^{i}} \frac{1}{\vartheta^{a-1}} \prod_{j=1}^{i} \frac{(1-\vartheta-1/\lambda+j/\lambda)}{(1-a/\lambda+j/\lambda)} \prod_{n=1}^{a-1} \frac{(1-n/\lambda)}{(1-n/\mu)}$$

$$= \frac{(-\hat{t})^{k}}{\vartheta^{a-1}} \prod_{n=1}^{a-1} \frac{(1-n/\lambda)}{(1-n/\mu)} \sum_{i=0}^{k} \binom{k}{i} (-1)^{i} \prod_{j=1}^{i} \frac{(1-\vartheta-1/\lambda+j/\lambda)(1-a/\lambda-1/\lambda)}{(1-\vartheta-1/\lambda)(1-a/\lambda+j/\lambda)}.$$
(B.21)

In the second line of (B.21) factors $F_1^{i,a}$ and $F_2^{i,a}$ were inserted and independent terms of *i* were moved out of the sum. Additionally the factors $(1 - a/\lambda - 1/\lambda)^i$ and $1/(1 - \vartheta - 1/\lambda)^i$ were moved into the product over j = 1, ..., i which is important for the following limit consideration.

Applying the limit $(\mu, \lambda) \to \infty$ 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 $k \ge 1$, one obtains the limit

$$\lim_{\substack{(\mu,\lambda)\to\infty\\\vartheta=\text{const.}}} C^{(k)} = \frac{(-\hat{t})^k}{\vartheta^{a-1}} \sum_{i=0}^k \binom{k}{i} (-1)^i = 0, \quad \text{for} \quad k \ge 1.$$
(B.22)

Now the results can be collected. 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

$$\lim_{\substack{(\mu,\lambda)\to\infty\\\vartheta=\text{const.}}} I^a_{\mu,\lambda}[f] = \lim_{\substack{(\mu,\lambda)\to\infty\\\vartheta=\text{const.}}} \left[\sum_{k=0}^\infty \frac{1}{k!} \left. \frac{\partial^k f}{\partial t^k} \right|_{t=\hat{t}} C^{(k)} \right]$$
$$= \frac{1}{\vartheta^{a-1}} f(1-\vartheta), \tag{B.23}$$

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} . **Generalized Progress Coefficient** An important application emerges investigating the generalized progress coefficients introduced in [3, Eq. (5.112)]

$$e_{\mu,\lambda}^{a,b} = \frac{\lambda - \mu}{(2\pi)^{\frac{a+1}{2}}} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} x^{b} e^{-\frac{a+1}{2}x^{2}} \left[\Phi(x)\right]^{\lambda - \mu - 1} \left[1 - \Phi(x)\right]^{\mu - a} dx, \qquad (B.24)$$

for which asymptotic properties can be derived assuming large populations. The population depend prefactors are rewritten as

$$(\lambda - \mu) \binom{\lambda}{\mu} = \frac{\lambda}{\mu} \frac{(\lambda - 1)!}{(\lambda - \mu - 1)!(\mu - 1)!} = \frac{1}{\vartheta} \frac{1}{B(\lambda - \mu, \mu)}.$$
(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 $0 \le t \le 1$, the progress coefficients yields

$$e^{a,b}_{\mu,\lambda} = \frac{1}{\vartheta} \frac{1}{B(\lambda-\mu,\mu)} \frac{1}{(2\pi)^{a/2}} \int_0^1 \left[\Phi^{-1}(t) \right]^b e^{-\frac{a}{2} \left[\Phi^{-1}(t) \right]^2} t^{\lambda-\mu-1} (1-t)^{\mu-a} dt \,. \tag{B.26}$$

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

$$f^{a,b}(t)\big|_{t=1-\vartheta} = \left[\Phi^{-1}(t)\right]^{b} e^{-\frac{a}{2}\left[\Phi^{-1}(t)\right]^{2}}\Big|_{t=1-\vartheta}$$

$$= \left[\Phi^{-1}(1-\vartheta)\right]^{b} e^{-\frac{a}{2}\left[\Phi^{-1}(1-\vartheta)\right]^{2}}.$$
(B.27)

Therefore the coefficients can be expressed as

$$e_{\mu,\lambda}^{a,b} = \frac{1}{\vartheta} \frac{1}{(2\pi)^{a/2}} I_{\mu,\lambda}^{a} \left[f^{a,b} \right] \simeq \frac{1}{\vartheta} \frac{1}{(2\pi)^{a/2}} \frac{f^{a,b}(1-\vartheta)}{\vartheta^{a-1}} = \frac{1}{(2\pi)^{a/2}} \frac{1}{\vartheta^{a}} \left[\Phi^{-1}(1-\vartheta) \right]^{b} e^{-\frac{a}{2} \left[\Phi^{-1}(1-\vartheta) \right]^{2}} = \left[\frac{e^{-\frac{1}{2} \left[\Phi^{-1}(\vartheta) \right]^{2}}}{\sqrt{2\pi} \vartheta} \right]^{a} \left[-\Phi^{-1}(\vartheta) \right]^{b}.$$
(B.28)

In the first line the asymptotic equality is used and in the second line expression (B.27). 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

$$e_{\vartheta}^{a,b} \coloneqq \lim_{\substack{(\mu,\lambda) \to \infty \\ \vartheta = \text{const.}}} e_{\mu,\lambda}^{a,b}, \tag{B.29}$$

the final result yields

$$e_{\vartheta}^{a,b} = \left[\frac{\mathrm{e}^{-\frac{1}{2}\left[\Phi^{-1}(\vartheta)\right]^{2}}}{\sqrt{2\pi}\vartheta}\right]^{a} \left[-\Phi^{-1}(\vartheta)\right]^{b}.$$
 (B.30)

Appendix C Derivation of the Progress Rates

C.1 First Order Progress Rate

The first order progress rate is derived starting from Eq. (31). Moving the sum and the *m*-dependent prefactors into the innermost integration yields

$$\varphi_i = -\frac{\lambda!}{\mu} \int_{-\infty}^{\infty} x_i p_x(x_i) \int_{q_l}^{q_u} p_Q(q|x_i) \sum_{m=1}^{\mu} \frac{P_Q(q)^{m-1} [1 - P_Q(q)]^{\lambda - m}}{(m-1)! (\lambda - m)!} \, \mathrm{d}q \, \mathrm{d}x_i \,. \tag{C.1}$$

Now a transformation can be applied for the sum $\sum_{m}(\cdot)$ yielding an expression as a function of the regularized incomplete beta function [3, p. 147]. This will later enable the application of the large population identity shown in Appendix B. One has

$$\sum_{m=1}^{\mu} \frac{P(q)^{m-1} [1 - P(q)]^{\lambda - m}}{(m-1)! (\lambda - m)!}$$

$$= \frac{1}{(\lambda - \mu - 1)! (\mu - 1)!} \int_{0}^{1 - P(q)} t^{\lambda - \mu - 1} (1 - t)^{\mu - 1} dt.$$
(C.2)

Furthermore, one can rewrite the population dependent factor as follows

$$\frac{\lambda!}{\mu} \frac{1}{(\lambda - \mu - 1)!(\mu - 1)!} = \frac{\lambda}{\mu} \frac{(\lambda - 1)!}{(\lambda - \mu - 1)!(\mu - 1)!} = \frac{\lambda}{\mu} \frac{\Gamma(\lambda)}{\Gamma(\lambda - \mu)\Gamma(\mu)} = \frac{\lambda}{\mu} \frac{1}{B(\lambda - \mu, \mu)}, \quad (C.3)$$

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. After replacing the sum and refactoring we arrive at the following progress rate integral

$$\varphi_i = -\frac{\lambda}{\mu} \frac{1}{\mathcal{B}(\lambda - \mu, \mu)} \int_{x_i = -\infty}^{x_i = \infty} x_i p_x(x_i) \int_{q = q_l}^{q = q_u} p_Q(q|x_i) \int_{t=0}^{t=1-P_Q(q)} t^{\lambda - \mu - 1} (1 - t)^{\mu - 1} \, \mathrm{d}t \, \mathrm{d}q \, \mathrm{d}x_i \,.$$
(C.4)

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

$$q_l \le q \le q_u, \qquad 0 \le t \le 1 - P_Q(q). \tag{C.5}$$

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

$$0 \le t \le 1, \qquad q_l \le q \le P_Q^{-1}(1-t).$$
 (C.6)

The progress rate changes to

$$\varphi_i = -\frac{\lambda}{\mu} \frac{1}{B(\lambda - \mu, \mu)} \int_{x_i = -\infty}^{x_i = \infty} x_i p_x(x_i) \int_{t=0}^{t=1} t^{\lambda - \mu - 1} (1 - t)^{\mu - 1} \int_{q=q_l}^{q=P_Q^{-1}(1 - t)} p_Q(q|x_i) \, \mathrm{d}q \, \mathrm{d}t \, \mathrm{d}x_i \,.$$
(C.7)

Now the innermost integral can be solved

$$\int_{q_l}^{P_Q^{-1}(1-t)} p_Q(q|x_i) \,\mathrm{d}q = \left[P_Q(q|x_i) \right]_{q_l}^{P_Q^{-1}(1-t)} \tag{C.8}$$

$$= P_Q(P_Q^{-1}(1-t)|x_i) - P_Q(q_l|x_i)$$
(C.9)

$$= P_Q(P_Q^{-1}(1-t)|x_i)$$
(C.10)

$$=: f(t, x_i). \tag{C.11}$$

where the probability $P_Q(q_l|x_i) = \Pr(Q \le q_l|x_i) = 0$ for any lower bound value q_l . For better readability the function $f(t, x_i)$ was introduced. Thus, we arrive at the following progress rate integral

$$\varphi_i = -\frac{\lambda}{\mu} \int_{x_i = -\infty}^{x_i = \infty} x_i p_x(x_i) \frac{1}{B(\lambda - \mu, \mu)} \int_{t=0}^{t=1} t^{\lambda - \mu - 1} (1 - t)^{\mu - 1} f(t, x_i) \, \mathrm{d}t \, \mathrm{d}x_i \,. \tag{C.12}$$

Large Population Approximation Unfortunately a closed form solution of (C.12) is not possible due to the factor $f(t, x_i) = P_Q(P_Q^{-1}(1-t)|x_i)$. But within the large-population limit with $(\mu, \lambda) \to \infty$ and constant truncation ratio $\vartheta = \mu/\lambda$, a solution for the *t*-integration can be given using the results of Appendix B. Comparing (C.12) with identity (B.1), one can identify integral $I^a_{\mu,\lambda}[f]$ with parameters a = 1 and b = 0 giving

$$\varphi_i = -\frac{\lambda}{\mu} \int_{x_i = -\infty}^{x_i = \infty} x_i p_x(x_i) I^1_{\mu,\lambda}[f] \,\mathrm{d}x_i \,. \tag{C.13}$$

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

$$f(t,x_i)|_{t=1-\vartheta} = P_Q(P_Q^{-1}(1-t)|x_i)|_{t=1-\vartheta} = P_Q(P_Q^{-1}(\vartheta)|x_i).$$
(C.14)

Applying $I^1_{\mu,\lambda}[f] \simeq P_Q(P_Q^{-1}(\vartheta)|x_i)$ to integral (C.13) yields

$$\varphi_i \simeq -\frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i p_x(x_i) P_Q(P_Q^{-1}(\vartheta) | x_i) \, \mathrm{d}x_i \,, \tag{C.15}$$

which now consists only of a single integration over the *i*-th mutation component x_i .

At this point expressions for the quality gain CDF and its inverse are needed in (C.15). The normal approximation for the quality gain distribution from Sec. 2 is used with

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

$$q = P_Q^{-1}(\vartheta) = E_Q + D_Q \Phi^{-1}(\vartheta) \qquad (C.16)$$

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

Now $E_{Q|x_i}$ from (14) with Q_i -result (19) is reformulated according to

$$E_{Q|x_i} = Q_i(x_i) + \sum_{j \neq i} E[Q_j] = k_i x_i + \delta_i(x_i) + E_i,$$
(C.17)

with following definitions

$$k_{i} \coloneqq 2y_{i}$$

$$\delta_{i}(x_{i}) \coloneqq x_{i}^{2} + A\cos\left(\alpha y_{i}\right)\left(1 - \cos\left(\alpha x_{i}\right)\right) + A\sin\left(\alpha y_{i}\right)\sin\left(\alpha x_{i}\right)$$

$$E_{i} \coloneqq \sum_{j \neq i} \mathbb{E}\left[Q_{i}\right].$$
(C.18)

These definitions are used as abbreviations to distinguish the (non-linear) perturbation term $\delta(x_i)$ from the linear term $k_i x_i$. Inserting relation (C.17) into (C.16) and the result into (C.15) yields

$$\varphi_i = -\frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i p_x(x_i) \Phi\left(\frac{E_Q + D_Q \Phi^{-1}(\vartheta) - (k_i x_i + \delta_i(x_i) + E_i)}{D_i}\right) \mathrm{d}x_i \,. \tag{C.19}$$

To obtain a function of the form $\Phi(g(x_i) + h(x_i))$ and apply the Taylor expansion (36), the functions g and h are defined as

$$g(x_i) \coloneqq -\frac{k_i}{D_i} x_i + \frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_i}$$
(C.20)

$$h(x_i) \coloneqq -\frac{\delta(x_i)}{D_i},\tag{C.21}$$

with g being a linear function of x_i and h a non-linear function thereof. Additionally, the abbreviation $E_{Q_i} = E_Q - E_i = E[Q_i]$, cf. Eq. (8), is used to denote the expected value of the *i*-th summand of the quality gain (4). For the expansion of $\Phi(g+h)$ in Eq. (C.19), the definition of the probabilist's Hermite polynomials will be useful. They are defined in terms of the standard normal density $d\Phi(x)/dx = \phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ as

$$\frac{\mathrm{d}^n \phi(x)}{\mathrm{d}x^n} = (-1)^n \operatorname{He}_n(x) \phi(x).$$
(C.22)

Using (C.22), Taylor expansion (36) can therefore be written as

$$\Phi(g+h) = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\mathrm{d}^{n} \Phi}{\mathrm{d} g^{n}} h^{n} = \Phi(g) + \phi(g)h + \sum_{n=1}^{\infty} \frac{1}{(n+1)!} \frac{\mathrm{d}^{n} \phi}{\mathrm{d} g^{n}} h^{n+1}$$
$$= \Phi(g) + \phi(g)h + \sum_{n=1}^{\infty} \frac{(-1)^{n}}{(n+1)!} \operatorname{He}_{n}(g) \phi(g) h^{n+1}.$$
(C.23)

Plugging (C.23) into integral (C.19) yields

$$\varphi_i = -\frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i p_x(x_i) \left[\Phi(g(x_i)) + \phi(g(x_i))h(x_i) + \phi(g(x_i)) \sum_{n=1}^{\infty} \frac{(-1)^n}{(n+1)!} \operatorname{He}_n(g(x_i))h(x_i)^{n+1} \right] \mathrm{d}x_i \,. \quad (C.24)$$

The zeroth order term in $[\cdot]$ of (C.24) will yield a closed form solution due to $g(x_i)$ being linear w.r.t. x_i . The first order term can be solved by applying quadratic completion to the Gaussian product $p_x(x_i)\phi(g(x_i))$ yielding an expected value over a normal density. The sum in (C.24) will be neglected in the limit $N \to \infty$ compared to the first order term. First, note that function h of (C.21) models the perturbation of a single component divided by D_i , which scales with $\sqrt{N-1} \approx \sqrt{N}$ for $N \to \infty$, see variance (15). The term $\operatorname{He}_n(g(x_i))$ yields an n-th order polynomial in $g(x_i)$, see (C.20). The first term of $g(x_i)$ is vanishing for $N \to \infty$ and therefore not critical. For the second term of $g(x_i)$ ratio $E_{Q_i}/D_i = O(1/\sqrt{N})$ is vanishing and the asymptotic equality $D_Q \simeq D_i$ holds, see more detailed discussion of Eq. (C.45). As $\text{He}_n(g)$ is asymptotically constant in the limit $N \to \infty$, one can conclude the scaling of the sum as $O(h^2) = O(1/N)$. Hence, Eq. (C.24) can be expressed in terms of two integrals I_i^0 and I_i^1 and higher order negligible terms as

$$\varphi_i = I_i^0 + I_i^1 + O\left(\frac{1}{N}\right), \quad \text{with} \tag{C.25}$$

$$I_i^0 \coloneqq -\frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i p_x(x_i) \Phi(g(x_i)) \, \mathrm{d}x_i \,, \quad \text{and}$$
(C.26)

$$I_i^1 \coloneqq -\frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i h(x_i) p_x(x_i) \phi(g(x_i)) \,\mathrm{d}x_i \,. \tag{C.27}$$

Solving integral I_i^0 Starting with (C.26) and using definition (C.20), the equation can be rewritten as

$$I_i^0 = -\frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i p_x(x_i) \Phi\left(-\frac{k_i}{D_i} x_i + \frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_i}\right) \mathrm{d}x_i \,. \tag{C.28}$$

Inserting the mutation density $p_x(x_i)$ from Eq. (28) and applying the substitution $z = x_i/\sigma$ one gets

$$I_i^0 = -\frac{\sigma}{\sqrt{2\pi\vartheta}} \int_{-\infty}^{\infty} z \mathrm{e}^{-\frac{1}{2}z^2} \Phi\left(-\frac{k_i\sigma}{D_i}z + \frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_i}\right) \mathrm{d}z\,. \tag{C.29}$$

At this point the following integral identity [3, Eq. (A.12)] can be applied

$$\int_{-\infty}^{\infty} t e^{-\frac{1}{2}t^2} \Phi(at+b) dt = \frac{a}{\sqrt{1+a^2}} \exp\left[-\frac{1}{2}\frac{b^2}{1+a^2}\right].$$
 (C.30)

The corresponding coefficients can be identified as

$$a = -\frac{k_i \sigma}{D_i} \tag{C.31}$$

$$b = \frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_i}.$$
(C.32)

Evaluating the factor $a/\sqrt{1+a^2}$ from (C.30) gives

$$\frac{a}{\sqrt{1+a^2}} = -\frac{\frac{k_i\sigma}{D_i}}{\sqrt{\frac{D_i^2}{D_i^2} + \frac{(k_i\sigma)^2}{D_i^2}}} = -\frac{k_i\sigma}{\sqrt{D_i^2 + (k_i\sigma)^2}} = -\frac{k_i\sigma}{D_+},\tag{C.33}$$

where following definition was introduced

$$D_+^2 \coloneqq D_i^2 + (k_i \sigma)^2. \tag{C.34}$$

The factor $\exp\left[-b^2/2(1+a^2)\right]$ in (C.30) yields

$$\exp\left[-\frac{1}{2}\frac{b^{2}}{1+a^{2}}\right] = \exp\left[-\frac{1}{2}\left(\frac{E_{Q_{i}}+D_{Q}\Phi^{-1}(\vartheta)}{D_{i}}\right)^{2}\frac{1}{\frac{D_{i}^{2}}{D_{i}^{2}}+\frac{(k_{i}\sigma)^{2}}{D_{i}^{2}}}\right]$$
$$= \exp\left[-\frac{1}{2}\left(\frac{E_{Q_{i}}+D_{Q}\Phi^{-1}(\vartheta)}{D_{+}}\right)^{2}\right].$$
(C.35)

Inserting results (C.33) and (C.35) into identity relation (C.30), the integral (C.29) yields

$$I_{i}^{0} = \frac{1}{\sqrt{2\pi}} \frac{1}{\vartheta} \exp\left[-\frac{1}{2} \left(\frac{E_{Q_{i}} + D_{Q} \Phi^{-1}(\vartheta)}{D_{+}}\right)^{2}\right] \frac{k_{i} \sigma^{2}}{D_{+}}.$$
 (C.36)

Solving integral I_i^1 The second progress rate integral (C.27) with density $p_x(x)$ from (28) and $\phi(g) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}g(x_i)^2}$ yields

$$I_{i}^{1} = -\frac{1}{\sqrt{2\pi\vartheta}} \int_{-\infty}^{\infty} x_{i}h(x_{i})p_{x}(x_{i})e^{-\frac{1}{2}g(x_{i})^{2}} dx_{i}$$
(C.37)

$$= -\frac{1}{\sqrt{2\pi\vartheta}} \int_{-\infty}^{\infty} x_i h(x_i) \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2} \left(\frac{x_i}{\sigma}\right)^2} e^{-\frac{1}{2}g(x_i)^2} \,\mathrm{d}x_i \,. \tag{C.38}$$

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

$$I_i^1 \stackrel{!}{=} -\frac{C}{2\pi\vartheta\sigma} \int_{-\infty}^{\infty} x_i h(x_i) \mathrm{e}^{-\frac{1}{2}\left(\frac{x_i-m}{s}\right)^2} \,\mathrm{d}x_i\,,\tag{C.39}$$

which will be determined using quadratic completion. Recalling $h(x_i)$ -definition (C.21) and perturbation term $\delta(x_i)$ from (C.18) we have

$$h(x_i) = -\frac{\delta(x_i)}{D_i} = -\frac{x_i^2 + A\cos(\alpha y_i)(1 - \cos(\alpha x_i)) + A\sin(\alpha y_i)\sin(\alpha x_i)}{D_i}.$$
 (C.40)

Using this relation, the integral (C.38) will be reformulated later as an expected value of the function $x_i h(x_i)$ over the normal density $\mathcal{N}(m, s^2)$. Returning to quadratic completion, one demands

$$e^{-\frac{1}{2}\frac{x_i^2}{\sigma^2}}e^{-\frac{1}{2}g(x_i)^2} \stackrel{!}{=} Ce^{-\frac{1}{2}\frac{(x_i-m)^2}{s^2}}.$$
 (C.41)

The calculation thereof is straightforward and the details are omitted here. Using definition (C.34) with $D_{+}^{2} := D_{i}^{2} + (k_{i}\sigma)^{2}$, the mean value *m* and standard deviation *s*, respectively, yield

$$m = \frac{[E_{Q_i} + D_Q \Phi^{-1}(\vartheta)]k_i \sigma^2}{D_+^2}$$
(C.42)

$$s = \frac{D_i \sigma}{D_+}.\tag{C.43}$$

The prefactor C from (C.41) is evaluated as

$$C = \exp\left[-\frac{1}{2}\left(\frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_+}\right)^2\right].$$
(C.44)

The result is the same exponential factor as in I_i^0 from Eq. (C.36), which will later be related to the asymptotic progress coefficient $e_{\vartheta}^{1,0}$.

Large dimensionality approximation Now the large dimensionality approximation can be applied to further simplify m, s, and C. This will also simplify the expected value calculation of the involved trigonometric terms in (C.39) considerably.

The variance quantities D_Q^2 , D_i^2 , and D_+^2 differ only by a single component, the *i*-th component. Assuming that the contribution of the *i*-th component is not dominating the overall variance of the remaining N-1 components, its contribution is negligible in the limit $N \to \infty$. Using Eqs. (9), (15), and (C.34) for D_Q^2 , D_i^2 , and D_+^2 , respectively, therefore yields asymptotically

$$D_Q^2 = \sum_{i=1}^N \operatorname{Var}\left[Q_i\right] \simeq \sum_{j \neq i} \operatorname{Var}\left[Q_j\right] \simeq \sum_{j \neq i} \operatorname{Var}\left[Q_j\right] + (k_i \sigma)^2.$$
(C.45)

Hence, the following asymptotic equalities hold

$$D_i^2 \simeq D_Q^2$$
, and $D_+^2 \simeq D_Q^2$. (C.46)

Using relation (C.46), m from Eq. (C.42) and s from Eq. (C.43) are evaluated as

$$m \simeq \frac{\left(E_{Q_i} + D_Q \Phi^{-1}(\vartheta)\right) k_i \sigma^2}{D_Q^2} = \left(\frac{E_{Q_i}}{D_Q^2} + \frac{\Phi^{-1}(\vartheta)}{D_Q}\right) k_i \sigma^2 \tag{C.47}$$

$$s \simeq \frac{D_Q}{D_Q}\sigma = \sigma.$$
 (C.48)

Since Eq. (C.47) contains D_Q and D_Q^2 in its denominator (where a sum over N terms is taken), it can further be simplified for large N. From the variance result in Appendix (A.7) we observe that D_Q^2 scales with the number of components N, such that D_Q scales with \sqrt{N} . Within (·) of Eq. (C.47), the first term $E_{Q_i} = E_Q - E_i = E[Q_i]$, cf. Eq. (8), 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 ratios. Both terms are suppressed by N and \sqrt{N} , respectively, and the infinite dimension limit can be evaluated as

$$\lim_{N \to \infty} m(N) = \lim_{N \to \infty} \left(\frac{E_{Q_i}}{D_Q^2} + \frac{\Phi^{-1}(\vartheta)}{D_Q} \right) k_i \sigma^2 = 0,$$
(C.49)

which is valid for any finite σ . The approximations (C.46) and (C.49) also change the exponential factor (C.44) as follows

$$C = \exp\left[-\frac{1}{2}\left(\frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_+}\right)^2\right] \simeq \exp\left[-\frac{1}{2}\Phi^{-1}(\vartheta)^2\right].$$
(C.50)

The obtained results for $N \to \infty$ are summarized as

$$s \simeq \sigma$$
, and $m \simeq 0$, (C.51)

which changes the density of $x_i \sim \mathcal{N}(m, s^2)$ to the density $x_i \sim \mathcal{N}(0, \sigma^2)$ in Eq. (C.39). The expected value of terms in $x_i h(x_i)$ containing odd powers of x_i is therefore zero, such that only one term being a function of $x_i \sin(\alpha x_i)$ needs to be evaluated. Collecting all approximation results and inserting for $x_i h(x_i)$ the term $-A \sin(\alpha y_i) x_i \sin(\alpha x_i) / D_Q$ (with $D_i \simeq D_Q$), integral (C.39) is therefore evaluated as

$$I_i^1 = \frac{\exp\left[-\frac{1}{2}\Phi^{-1}(\vartheta)^2\right]A\sin\left(\alpha y_i\right)}{2\pi\vartheta\sigma D_Q}\int_{-\infty}^{\infty}x_i\sin\left(\alpha x_i\right)\mathrm{e}^{-\frac{1}{2}\left(\frac{x_i}{\sigma}\right)^2}\,\mathrm{d}x_i\tag{C.52}$$

$$= \frac{\exp\left[-\frac{1}{2}\Phi^{-1}(\vartheta)^2\right]}{\sqrt{2\pi}\vartheta} \frac{A\sin\left(\alpha y_i\right)}{D_Q} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x_i \sin\left(\alpha x_i\right) \mathrm{e}^{-\frac{1}{2}\left(\frac{x_i}{\sigma}\right)^2} \mathrm{d}x_i \tag{C.53}$$

$$= c_{\vartheta} \frac{A\sin\left(\alpha y_{i}\right)}{D_{Q}} \operatorname{E}\left[x_{i}\sin\left(\alpha x_{i}\right)\right]$$
(C.54)

$$= c_{\vartheta} \frac{A \sin\left(\alpha y_{i}\right)}{D_{Q}} \alpha \sigma^{2} \mathrm{e}^{-\frac{1}{2}(\alpha \sigma)^{2}} \tag{C.55}$$

$$= c_{\vartheta} \frac{d_i \sigma^2}{D_O} \mathrm{e}^{-\frac{1}{2}(\alpha \sigma)^2}. \tag{C.56}$$

using (C.50) and (C.51) in the first line, rearranging the terms in the second line, applying definition (32) for $e_{\vartheta}^{1,0} = c_{\vartheta}$ and the definition of the expected value of the term $x_i \sin(\alpha x_i)$ in the third line, evaluating the expectation using (A.5) in the fourth line, and finally recognizing the derivative $d_i = \alpha A \sin(\alpha y_i)$ from (23) in the last line.

Analogously, the large dimensionality approximation can be applied to I_i^0 from (C.36) with $e_{\vartheta}^{1,0} = c_{\vartheta}$ giving

$$I_i^0 = \frac{1}{\sqrt{2\pi}} \frac{1}{\vartheta} \exp\left[-\frac{1}{2} \left(\frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_+}\right)^2\right] \frac{k_i \sigma^2}{D_+} \simeq c_\vartheta \frac{k_i \sigma^2}{D_Q}.$$
 (C.57)

Collecting results (C.57) and (C.56), and inserting them back into progress rate (C.25), the final result for the first order progress rate is obtained with derivative components $k_i = 2y_i$ and $d_i = \alpha A \sin(\alpha y_i)$ in the limits $N \to \infty$ and $(\mu, \lambda) \to \infty$ with constant $\vartheta = \mu/\lambda$ as

$$\varphi_i = I_i^0 + I_i^1$$

= $c_\vartheta \frac{\sigma^2}{D_Q} \left(k_i + e^{-\frac{1}{2}(\alpha\sigma)^2} d_i \right) = c_\vartheta \frac{\sigma^2}{D_Q} \left(2y_i + e^{-\frac{1}{2}(\alpha\sigma)^2} \alpha A \sin\left(\alpha y_i\right) \right).$ (C.58)

C.2 Second Order Progress Rate

C.2.1 Expectation of $E^{(2)}$

Starting from (43) and applying the order statistic density from (27) one has

$$\frac{1}{\mu^2} E^{(2)} = \frac{1}{\mu^2} \sum_{m=1}^{\mu} E\left[x_{m;\lambda}^2\right] = \frac{1}{\mu^2} \sum_{m=1}^{\mu} \int_{-\infty}^{\infty} x_i^2 \ p_{m;\lambda}(x_i) \,\mathrm{d}x_i \,. \tag{C.59}$$

Both (27) and (C.59) have the same structure after inserting the order statistic density $p_{m;\lambda}(x_i)$ from (30) and the integration over the squared mutation component is performed as the last step. The results of

Appendix C.1 can therefore be applied to Eq. (C.59). Starting from Eq. (C.1) one arrives at intermediate result (C.12) with squared mutation component, such that (C.59) yields

$$\frac{1}{\mu^2} E^{(2)} = \frac{1}{\mu} \left[\frac{1}{\mu} \sum_{m=1}^{\mu} \int_{-\infty}^{\infty} x_i^2 p_{m;\lambda}(x_i | \mathbf{y}) \, \mathrm{d}x_i \right]
= \frac{1}{\mu} \left[\frac{\lambda}{\mu} \int_{x_i = -\infty}^{x_i = \infty} x_i^2 p_x(x_i) \frac{1}{\mathrm{B}(\lambda - \mu, \mu)} \int_{t=0}^{t=1} t^{\lambda - \mu - 1} (1 - t)^{\mu - 1} P_Q(P_Q^{-1}(1 - t) | x_i) \, \mathrm{d}t \, \mathrm{d}x_i \right].$$
(C.60)

Solving the t-integration in Eq. (C.60), the large population identity (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

$$\frac{1}{\mu^2} E^{(2)} \simeq \frac{1}{\mu} \frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i^2 p_x(x_i) P_Q(P_Q^{-1}(\vartheta) | x_i) \,\mathrm{d}x_i \,, \tag{C.61}$$

which is analogous to (C.15). Inserting the normal approximation of the quality gain (C.16) into (C.61)leads again to an analytically not solvable integration due to non-linear terms in x_i within $\Phi(\cdot)$. Applying the expansion (C.23) of the normal CDF and using 0-th and first order terms yields

$$\frac{1}{\mu^2} E^{(2)} = \frac{1}{\mu} \left[\frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i^2 p_x(x_i) \Phi(g(x_i)) \, \mathrm{d}x_i + \frac{1}{\sqrt{2\pi\vartheta}} \int_{-\infty}^{\infty} x_i^2 p_x(x_i) h(x_i) \mathrm{e}^{-\frac{1}{2}g(x_i)^2} \, \mathrm{d}x_i \right]$$

=: $I_i^0 + I_i^1$, (C.62)

with the two integrals abbreviated as I_i^0 and I_i^1 , which are evaluated now. Starting with the first integration I_i^0 , it is rewritten analogously to (C.29) using $g(x_i)$ from (C.20) and the substitution $z = x_i/\sigma$ giving

$$I_i^0 = \frac{1}{\mu\vartheta} \int_{-\infty}^{\infty} x_i^2 p_x(x_i) \Phi(g(x_i)) \,\mathrm{d}x_i = \frac{\sigma^2}{\sqrt{2\pi}\mu\vartheta} \int_{-\infty}^{\infty} z^2 \mathrm{e}^{-\frac{1}{2}z^2} \Phi\left(-\frac{k_i\sigma}{D_i}z + \frac{E_{Q_i} + D_Q\Phi^{-1}(\vartheta)}{D_i}\right) \,\mathrm{d}z\,. \quad (C.63)$$

At this point the result of integral identity (D.1) is needed to solve (C.63). Defining the coefficients

$$a = -\frac{k_i \sigma}{D_i}, \quad b = \frac{E_{Q_i} + D_Q \Phi^{-1}(\vartheta)}{D_i}, \tag{C.64}$$

expressions needed for (D.1) are evaluated as

$$(1+a^{2})^{1/2} = \sqrt{\frac{D_{i}^{2}}{D_{i}^{2}} + \left(\frac{k_{i}\sigma}{D_{i}}\right)^{2}} = \sqrt{\frac{D_{+}^{2}}{D_{+}^{2}}} = \frac{D_{+}}{D_{i}}$$

$$\frac{a^{2}b}{(1+a^{2})^{3/2}} = \frac{(k_{i}\sigma)^{2}}{D_{+}^{2}} \frac{E_{Q_{i}} + D_{Q}\Phi^{-1}(\vartheta)}{D_{+}}$$

$$e^{-\frac{1}{2}\frac{b^{2}}{1+a^{2}}} = \exp\left[-\frac{1}{2}\left(\frac{E_{Q_{i}} + D_{Q}\Phi^{-1}(\vartheta)}{D_{+}}\right)^{2}\right],$$
(C.65)

using $D_{+}^{2} = D_{i}^{2} + (k_{i}\sigma)^{2}$ from (C.34). Applying the large dimensionality approximation from (C.46) and limit (C.49), the expression (C.65) simplifies as

$$(1+a^2)^{1/2} \simeq 1 \frac{a^2 b}{(1+a^2)^{3/2}} \simeq \frac{(k_i \sigma)^2}{D_Q^2} \Phi^{-1}(\vartheta) e^{-\frac{1}{2}\frac{b^2}{1+a^2}} \simeq \exp\left[-\frac{1}{2} \left[\Phi^{-1}(\vartheta)\right]^2\right].$$
 (C.66)

Now integral identity (D.1) can be evaluated using (C.66) and yields the result for integral (C.63)

$$I_{i}^{0} \simeq \frac{\sigma^{2}}{\mu\vartheta} \left[\Phi\left(\Phi^{-1}(\vartheta)\right) - \frac{1}{\sqrt{2\pi}} \frac{(k_{i}\sigma)^{2}}{D_{Q}^{2}} \Phi^{-1}(\vartheta) \exp\left[-\frac{1}{2} \left[\Phi^{-1}(\vartheta)\right]^{2}\right] \right]$$
$$= \frac{\sigma^{2}}{\mu} \left[1 - \Phi^{-1}(\vartheta) \left[\frac{e^{-\frac{1}{2}\left[\Phi^{-1}(\vartheta)\right]^{2}}}{\sqrt{2\pi}\vartheta}\right] \frac{(k_{i}\sigma)^{2}}{D_{Q}^{2}} \right].$$
(C.67)

Given (C.67), the asymptotic generalized progress coefficient definition $e_{\vartheta}^{1,1}$ from (B.30) can be applied with parameters a = 1 and b = 1

$$e_{\vartheta}^{1,1} = \left[-\Phi^{-1}(\vartheta)\right] \left[\frac{\mathrm{e}^{-\frac{1}{2}\left[\Phi^{-1}(\vartheta)\right]^{2}}}{\sqrt{2\pi}\vartheta}\right].$$
 (C.68)

This leads to following result for the first integral I_i^0

$$I_{i}^{0} = \frac{\sigma^{2}}{\mu} \left[1 + e_{\vartheta}^{1,1} \frac{(k_{i}\sigma)^{2}}{D_{Q}^{2}} \right].$$
(C.69)

Second integration I_i^1 from (C.62) is defined as

$$I_{i}^{1} = \frac{1}{\sqrt{2\pi\mu\vartheta}} \int_{-\infty}^{\infty} x_{i}^{2} p_{x}(x_{i}) h(x_{i}) \mathrm{e}^{-\frac{1}{2}g(x_{i})^{2}} \,\mathrm{d}x_{i}$$
(C.70)

with $g(x_i)$ and $h(x_i)$ defined in (C.20) and (C.21), respectively. Quadratic completion for the Gaussians of (C.70) was already evaluated in Eq. (C.39) with parameters m, s, and C given in (C.42), (C.43), and (C.44), respectively. Again, the large dimensionality approximation is applied to simplify the lengthy expressions and the results of (C.50) and (C.51) are applicable with

$$m \simeq 0, \quad s \simeq \sigma, \quad C \simeq \exp\left[-\frac{1}{2}\left[\Phi^{-1}(\vartheta)\right]^2\right].$$
 (C.71)

Therefore integral (C.70) with quadratic completion (C.39) assuming large N yields

$$I_{i}^{1} = \frac{1}{\mu} \frac{C}{2\pi \vartheta \sigma} \int_{-\infty}^{\infty} x_{i}^{2} h(x_{i}) e^{-\frac{1}{2} \left(\frac{x_{i} - m}{s}\right)^{2}} dx_{i}$$

$$\simeq \frac{1}{\mu} \frac{e^{-\frac{1}{2} \left[\Phi^{-1}(\vartheta)\right]^{2}}}{\sqrt{2\pi} \vartheta} \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} x_{i}^{2} h(x_{i}) e^{-\frac{1}{2} \frac{x_{i}^{2}}{\sigma^{2}}} dx_{i}.$$
 (C.72)

Given (C.72), one can compare coefficients with the asymptotic generalized progress coefficient from (B.30) and identify following expression using a = 1 and b = 0

$$\frac{e^{-\frac{1}{2}\left[\Phi^{-1}(\vartheta)\right]^2}}{\sqrt{2\pi}\vartheta} = e_{\vartheta}^{1,0} = c_{\vartheta},\tag{C.73}$$

Additionally in (C.72), 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 (C.40) with $D_i \simeq D_Q$, expression (C.72) is reformulated

$$I_i^1 = -\frac{c_{\vartheta}}{\mu D_Q} \left(\mathbf{E} \left[x_i^4 \right] + A\sin\left(\alpha y_i\right) \mathbf{E} \left[x_i^2 \sin\left(\alpha x_i\right) \right] + A\cos\left(\alpha y_i\right) \mathbf{E} \left[x_i^2 \right] - A\cos\left(\alpha y_i\right) \mathbf{E} \left[x_i^2 \cos\left(\alpha x_i\right) \right] \right).$$
(C.74)

One has $E[x_i^4] = 3\sigma^4$ and $E[x_i^2] = \sigma^2$. Using results from Appendix A the remaining expected values read

$$E\left[x_i^2 \sin\left(\alpha x_i\right)\right] = 0$$

$$E\left[x_i^2 \cos\left(\alpha x_i\right)\right] = (\sigma^2 - \alpha^2 \sigma^4) e^{-\frac{1}{2}(\alpha \sigma)^2}.$$
(C.75)

Therefore one gets

$$I_{i}^{1} = -\frac{c_{\vartheta}\sigma^{2}}{\mu D_{Q}} \left[3\sigma^{2} + A\cos\left(\alpha y_{i}\right) \left(1 - e^{-\frac{1}{2}(\alpha\sigma)^{2}} + \alpha^{2}\sigma^{2}e^{-\frac{1}{2}(\alpha\sigma)^{2}} \right) \right].$$
(C.76)

Collecting the results (C.69) and (C.76) with $k_i = 2y_i$ and inserting them back into (C.62) the expectation value reads

$$\frac{1}{\mu^{2}}E^{(2)} = \frac{\sigma^{2}}{\mu} \Biggl\{ 1 + e_{\vartheta}^{1,1} \frac{(2y_{i})^{2}\sigma^{2}}{D_{Q}^{2}} - \frac{c_{\vartheta}}{D_{Q}} \Biggl[3\sigma^{2} + A\cos\left(\alpha y_{i}\right) \left(1 - e^{-\frac{1}{2}(\alpha\sigma)^{2}} + \alpha^{2}\sigma^{2}e^{-\frac{1}{2}(\alpha\sigma)^{2}} \right) \Biggr] \Biggr\}.$$
(C.77)

C.2.2 Expectation of $E^{(1,1)}$

The expectation value to be evaluated is given in Eq. (44) by

$$\frac{1}{\mu^2} E^{(1,1)} = \frac{1}{\mu^2} \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} \mathbb{E} \left[x_{k;\lambda} x_{l;\lambda} \right].$$
(C.78)

The derivation includes the following steps. First, a joint order statistic density has to be derived for the expected value. Second, the double sum of (C.78) is converted into a single integration using a known identity. Then, the resulting five-fold integration is restructured by exchanging bounds and successively solved. A remarkably simple intermediate result will be obtained in Eq. (C.105) within the limit of $\mu \to \infty$. Finally, the previously derived φ_i -result can be applied.

The double sum in (C.78) 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 $1 \le k < l \le \lambda$ is assumed with k yielding a smaller (better) quality value $q_1 < q_2$. Calculating (C.78) the joint probability density $p_{k,l;\lambda}(x_1, x_2)$ is needed, such that the expected value can be formulated as

$$\frac{1}{\mu^2} E^{(1,1)} = \frac{1}{\mu^2} \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p_{k,l;\lambda}(x_1, x_2) \,\mathrm{d}x_2 \,\mathrm{d}x_1 \,. \tag{C.79}$$

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 smaller than q_1 , l-k-1 values between q_1 and q_2 and $\lambda - l$ values larger than q_2 with probabilities

$$\Pr\{Q \le q_1\}^{k-1} = P_Q(q_1)^{k-1}$$

$$\Pr\{q_1 \le Q \le q_2\}^{l-k-1} = [P_Q(q_2) - P_Q(q_1)]^{l-k-1}$$

$$\Pr\{Q > q_2\}^{\lambda-l} = [1 - P_Q(q_2)]^{\lambda-l},$$
(C.80)

and $P_Q(q)$ denoting the quality gain CDF. The joint probability density can therefore be written as

$$p_{k,l;\lambda}(x_1, x_2) = p_x(x_1)p_x(x_2) \int_{q_{\min}}^{\infty} p_Q(q_1|x_1) \int_{q_1}^{\infty} p_Q(q_2|x_2) \times \lambda! \frac{P_Q(q_1)^{k-1}[P_Q(q_2) - P_Q(q_1)]^{l-k-1}[1 - P_Q(q_2)]^{\lambda-l}}{(k-1)!(l-k-1)!(\lambda-l)!} \,\mathrm{d}q_2 \,\mathrm{d}q_1 \,,$$
(C.81)

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 (C.80). Plugging (C.81) into (C.79) and moving the sum into the integration one gets

$$\frac{1}{\mu^2} E^{(1,1)} = \frac{\lambda!}{\mu^2} \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) \int_{q_{\min}}^{\infty} p_Q(q_1|x_1) \int_{q_1}^{\infty} p_Q(q_2|x_2) \\ \times \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} \frac{P_Q(q_1)^{k-1} [P_Q(q_2) - P_Q(q_1)]^{l-k-1} [1 - P_Q(q_2)]^{\lambda-l}}{(k-1)!(l-k-1)!(\lambda-l)!} \, \mathrm{d}q_2 \, \mathrm{d}q_1 \, \mathrm{d}x_2 \, \mathrm{d}x_1 \,.$$
(C.82)

The double sum of (C.82) 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

$$\sum_{l=2}^{\mu} \sum_{k=1}^{l-1} \frac{Q_1^{\lambda-l} [Q_2 - Q_1]^{l-k-1} [1 - Q_2]^{k-1}}{(\lambda - l)! (l-k-1)! (k-1)!} = \frac{1}{(\lambda - \mu - 1)! (\mu - 2)!} \int_0^{Q_1} t^{\lambda - \mu - 1} (1 - t)^{\mu - 2} \,\mathrm{d}t \,, \tag{C.83}$$

for real values Q_1 and Q_2 , with 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 (C.82) can be recognized by comparing with (C.83). Applying the identity therefore yields

$$\sum_{l=2}^{\mu} \sum_{k=1}^{l-1} \frac{[1 - P_Q(q_2)]^{\lambda - l} [P_Q(q_2) - P_Q(q_1)]^{l-k-1} [P_Q(q_1)]^{k-1}}{(\lambda - l)! (l-k-1)! (k-1)!}$$

$$= \frac{1}{(\lambda - \mu - 1)! (\mu - 2)!} \int_0^{1 - P_Q(q_2)} t^{\lambda - \mu - 1} (1 - t)^{\mu - 2} dt.$$
(C.84)

Hence, Eq. (C.82) is evaluated as

$$\frac{1}{\mu^2} E^{(1,1)} = \frac{\lambda!}{\mu^2} \frac{1}{(\lambda - \mu - 1)!(\mu - 2)!} \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) \times \int_{q_{\min}}^{\infty} p_Q(q_1|x_1) \int_{q_1}^{\infty} p_Q(q_2|x_2) \int_{0}^{1 - P_Q(q_2)} t^{\lambda - \mu - 1} (1 - t)^{\mu - 2} dt dq_2 dq_1 dx_2 dx_1.$$
(C.85)

The prefactor of Eq. (C.85) can be evaluated as

$$\frac{\lambda!}{\mu^2} \frac{1}{(\lambda - \mu - 1)!(\mu - 2)!} = \frac{\lambda(\lambda - 1)!(\mu - 1)}{\mu^2(\lambda - \mu - 1)!(\mu - 1)!} = \frac{\lambda}{\mu} \frac{\mu - 1}{\mu} \frac{(\lambda - 1)!}{(\lambda - \mu - 1)!(\mu - 1)!} = \frac{1}{\vartheta} \frac{\mu - 1}{\mu} \frac{1}{B(\lambda - \mu, \mu)},$$
(C.86)

which will be useful during subsequent calculations.

Now the integration order will be exchanged twice in (C.85). 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

$$q_1 \le q_2 < \infty, \quad 0 \le t \le 1 - P_Q(q_2),$$
 (C.87)

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

$$0 \le t \le 1 - P_Q(q_1), \quad q_1 \le q_2 \le P_Q^{-1}(1-t).$$
 (C.88)

Using factor (C.86) and exchanged bounds (C.88), the expression (C.85) is reformulated as

$$\frac{1}{\mu^2} E^{(1,1)} = \frac{1}{\vartheta} \frac{\mu - 1}{\mu} \frac{1}{B(\lambda - \mu, \mu)} \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) \times \int_{q_{\min}}^{\infty} p_Q(q_1 | x_1) \int_{0}^{1 - P_Q(q_1)} t^{\lambda - \mu - 1} (1 - t)^{\mu - 2} \int_{q_1}^{P_Q^{-1}(1 - t)} p_Q(q_2 | x_2) \, \mathrm{d}q_2 \, \mathrm{d}t \, \mathrm{d}q_1 \, \mathrm{d}x_2 \, \mathrm{d}x_1.$$
(C.89)

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

$$q_{\min} \le q_1 < \infty, \quad 0 \le t \le 1 - P_Q(q_1),$$
 (C.90)

yielding exchanged bounds

$$0 \le t \le 1, \quad q_{\min} \le q_1 \le P_Q^{-1}(1-t).$$
 (C.91)

Therefore, one arrives at the following integral to be solved (beta function has been moved inside as it will be evaluated during the t-integration)

$$\frac{1}{\mu^{2}}E^{(1,1)} = \frac{1}{\vartheta}\frac{\mu-1}{\mu}\int_{-\infty}^{\infty}x_{1}p_{x}(x_{1})\int_{-\infty}^{\infty}x_{2}p_{x}(x_{2}) \\
\times \left(\frac{1}{B(\lambda-\mu,\mu)}\int_{0}^{1}t^{\lambda-\mu-1}(1-t)^{\mu-2} \\
\times \left[\int_{q_{\min}}^{P_{Q}^{-1}(1-t)}p_{Q}(q_{1}|x_{1})\left\{\int_{q_{1}}^{P_{Q}^{-1}(1-t)}p_{Q}(q_{2}|x_{2})\,\mathrm{d}q_{2}\right\}\mathrm{d}q_{1}\right]\mathrm{d}t\right)\mathrm{d}x_{2}\,\mathrm{d}x_{1}\,.$$
(C.92)

Now the integrals in (C.92) will be successively solved. Starting with integral $\{\cdot\}$ over q_2 one has

$$\int_{q_1}^{P_Q^{-1}(1-t)} p_Q(q_2|x_2) \,\mathrm{d}q_2 = \left[P_Q(q_2|x_2) \right]_{q_1}^{P_Q^{-1}(1-t)} = P_Q(P_Q^{-1}(1-t)|x_2) - P_Q(q_1|x_2). \tag{C.93}$$

The q_1 -integration within [·] using (C.93) yields

$$\int_{q_{\min}}^{P_Q^{-1}(1-t)} p_Q(q_1|x_1) \left(P_Q(P_Q^{-1}(1-t)|x_2) - P_Q(q_1|x_2) \right) dq_1$$
(C.94)

$$= P_Q(P_Q^{-1}(1-t)|x_2) \int_{q_{\min}}^{P_Q^{-1}(1-t)} p_Q(q_1|x_1) \,\mathrm{d}q_1 \tag{C.95}$$

$$-\int_{q_{\min}}^{P_Q^{-1}(1-t)} p_Q(q_1|x_1) P_Q(q_1|x_2) \,\mathrm{d}q_1 \tag{C.96}$$

First integral (C.95) is easily evaluated, as the conditional density is integrated over its support giving

$$P_Q(P_Q^{-1}(1-t)|x_2) \int_{q_{\min}}^{P_Q^{-1}(1-t)} p_Q(q_1|x_1) \, \mathrm{d}q_1 = P_Q(P_Q^{-1}(1-t)|x_2) \Big[P_Q(q_1|x_1) \Big]_{q_{\min}}^{P_Q^{-1}(1-t)}$$

$$= P_Q(P_Q^{-1}(1-t)|x_2) P_Q(P_Q^{-1}(1-t)|x_1)$$
(C.97)

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

The second integral (C.96) 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

$$\int_{q_{\min}}^{P_Q^{-1}(1-t)} p_Q(q_1|x_1) P_Q(q_1|x_2) \, \mathrm{d}q_1$$

$$= P_Q(P_Q^{-1}(1-t)|x_1) P_Q(P_Q^{-1}(1-t)|x_2) - \int_{q_{\min}}^{P_Q^{-1}(1-t)} P_Q(q_1|x_1) p_Q(q_1|x_2) \, \mathrm{d}q_1 \,.$$
(C.98)

Equation (C.98) inserted into (C.92) has to be integrated over x_1 and x_2 , of which the order can be exchanged. For the following step the *t*-integration and the prefactors of (C.92) have no influence, such that they are dropped for better readability. Integrating both sides of (C.98) yields

$$\int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) \int_{q_{\min}}^{P_Q^{-1}(1-t)} p_Q(q_1|x_1) P_Q(q_1|x_2) \, dq_1 \, dx_2 \, dx_1$$

$$= \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) P_Q(P_Q^{-1}(1-t)|x_1) P_Q(P_Q^{-1}(1-t)|x_2) \, dx_2 \, dx_1 \qquad (C.99)$$

$$- \int_{-\infty}^{\infty} x_2 p_x(x_2) \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{q_{\min}}^{P_Q^{-1}(1-t)} P_Q(q_1|x_2) p_Q(q_1|x_1) \, dq_1 \, dx_1 \, dx_2 \, ,$$

where in the last line the integration order of x_1 and x_2 was exchanged, such that an expression equivalent to the left-hand side of (C.99) is obtained with given arguments for p_Q and P_Q . Collecting the terms, Eq. (C.99) can be formulated as

$$\int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) \int_{q_{\min}}^{P_Q^{-1}(1-t)} p_Q(q_1|x_1) P_Q(q_1|x_2) \, \mathrm{d}q_1 \, \mathrm{d}x_2 \, \mathrm{d}x_1$$

$$= \frac{1}{2} \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) P_Q(P_Q^{-1}(1-t)|x_1) P_Q(P_Q^{-1}(1-t)|x_2) \, \mathrm{d}x_2 \, \mathrm{d}x_1 \, .$$
(C.100)

Noting that the right-hand side of result (C.100) is one half of the first integration result (C.97) after x-integration and noting the minus sign in (C.96), one gets for (C.94) the expression

$$\int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) \int_{q_{\min}}^{P_Q^{-1}(1-t)} p_Q(q_1|x_1) \left(P_Q(P_Q^{-1}(1-t)|x_2) - P_Q(q_1|x_2) \right) dq_1 dx_2 dx_1$$

$$= \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) \left(1 - \frac{1}{2} \right) P_Q(P_Q^{-1}(1-t)|x_1) P_Q(P_Q^{-1}(1-t)|x_2) dx_2 dx_1.$$
(C.101)

Including prefactors and integration over t again, the result within $[\cdot]$ of (C.92) simplifies significantly giving

$$\frac{1}{\mu^2} E^{(1,1)} = \frac{1}{2} \frac{1}{\vartheta} \frac{\mu - 1}{\mu} \int_{-\infty}^{\infty} x_1 p_x(x_1) \int_{-\infty}^{\infty} x_2 p_x(x_2) \\
\times \left(\frac{1}{\mathrm{B}(\lambda - \mu, \mu)} \int_0^1 t^{\lambda - \mu - 1} (1 - t)^{\mu - 2} P_Q(P_Q^{-1}(1 - t)|x_1) P_Q(P_Q^{-1}(1 - t)|x_2) \,\mathrm{d}t \right) \mathrm{d}x_2 \,\mathrm{d}x_1 \,.$$
(C.102)

For the integral in (·) of Eq. (C.102) the large population identity of (B.1) can be applied for $(\mu, \lambda) \to \infty$ with constant ϑ . 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

$$\frac{1}{\mathrm{B}(\lambda-\mu,\mu)} \int_0^1 t^{\lambda-\mu-1} (1-t)^{\mu-2} P_Q(P_Q^{-1}(1-t)|x_1) P_Q(P_Q^{-1}(1-t)|x_2) \,\mathrm{d}t$$

$$\simeq \frac{1}{\vartheta} P_Q(P_Q^{-1}(\vartheta)|x_1) P_Q(P_Q^{-1}(\vartheta)|x_2).$$
(C.103)

Using asymptotic equality (C.103) and noting that the terms containing x_1 and x_2 can be separated accordingly, Eq. (C.102) becomes

$$\frac{1}{\mu^2} E^{(1,1)} \simeq \frac{1}{2} \frac{1}{\vartheta^2} \frac{\mu - 1}{\mu} \int_{-\infty}^{\infty} x_1 p_x(x_1) P_Q(P_Q^{-1}(\vartheta) | x_1) \, \mathrm{d}x_1 \int_{-\infty}^{\infty} x_2 p_x(x_2) P_Q(P_Q^{-1}(\vartheta) | x_2) \, \mathrm{d}x_2$$

$$= \frac{1}{2} \frac{\mu - 1}{\mu} \left[\frac{1}{\vartheta} \int_{-\infty}^{\infty} x_i p_x(x_i) P_Q(P_Q^{-1}(\vartheta) | x_i) \, \mathrm{d}x_i \right]^2$$
(C.104)

where the integration variable is now denoted as x_i referring to the *i*-th mutation component. Additionally, the factor $1/\vartheta$ was moved into [·].

Remarkably, 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. (C.15). The result of (C.104) can therefore be expressed as

$$\frac{1}{\mu^2} E^{(1,1)} \simeq \frac{1}{2} \frac{\mu - 1}{\mu} \varphi_i^2. \tag{C.105}$$

Now the result for φ_i from (C.58) can be inserted into (C.105) giving

$$\frac{1}{\mu^{2}}E^{(1,1)} \simeq \frac{1}{2}\frac{\mu-1}{\mu}\varphi_{i}^{2}
= \frac{1}{2}\frac{\mu-1}{\mu}\left(c_{\vartheta}\frac{\sigma^{2}}{D_{Q}}\left(k_{i}+e^{-\frac{1}{2}(\alpha\sigma)^{2}}d_{i}\right)\right)^{2}
= \frac{1}{2}\frac{\mu-1}{\mu}e^{2,0}\frac{\sigma^{4}}{D_{Q}^{2}}\left(2y_{i}+e^{-\frac{1}{2}(\alpha\sigma)^{2}}\alpha A\sin\left(\alpha y_{i}\right)\right)^{2},$$
(C.106)

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 (B.30) according to

$$c_{\vartheta}^{2} = \left(e_{\vartheta}^{1,0}\right)^{2} = \left[\frac{\mathrm{e}^{-\frac{1}{2}\left[\Phi^{-1}(\vartheta)\right]^{2}}}{\sqrt{2\pi}\vartheta}\right]^{2} = e_{\vartheta}^{2,0}.$$
(C.107)

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

$$\frac{1}{\mu^2} E^{(1,1)} = \frac{1}{2} \frac{\sigma^2}{\mu} (\mu - 1) e_{\vartheta}^{2,0} \frac{\sigma^2}{D_Q^2} \left(2y_i + e^{-\frac{1}{2}(\alpha\sigma)^2} \alpha A \sin(\alpha y_i) \right)^2.$$
(C.108)

C.2.3 Investigation of Loss Terms

The main goal is to further simplify φ_i^{II} -result from Eq. (47) by investigating the loss term within $\{\cdot\}$ in the limit $N \to \infty$. The terms are abbreviated according to their respective factors as $e_{\vartheta}^{1,1}$, c_{ϑ}/D_Q and $e_{\vartheta}^{2,0}$. As the φ_i^{II} -approximation shall be valid for constant normalized mutations σ^* given some residual distance R, the transformed mutation is given by

$$\sigma = \frac{\sigma^* R}{N},\tag{C.109}$$

and will be expanded within the exponential functions of (47) for $N \to \infty$. Within this limit attention must be paid considering the relation R(N), as the (interesting) *R*-range with high density of local minima increases as well. This can also be observed experimentally in Fig. 9 as the transition region shifts to larger *R*. Therefore the expansion orders will be displayed as functions of R/N for the following derivations. To the end, the two cases R = const. and $R = O(\sqrt{N})$ will be investigated. The scaling $R = O(\sqrt{N})$ was already motivated in Sec. 3 presenting the experimental setup.

First the variance (21) is analyzed. The exponentials are expanded using

$$e^{-c\left(\alpha\sigma^*\frac{R}{N}\right)^2} = 1 - c\left(\alpha\sigma^*\frac{R}{N}\right)^2 + O\left(\frac{R^4}{N^4}\right),\tag{C.110}$$

with $c \in \{1/2, 1\}$. Using (C.109) in (21) the variance as a function of σ^* yields

$$D_Q^2 = \sum_{i=1}^N 2\left(\sigma^* \frac{R}{N}\right)^4 + 4y_i^2 \left(\sigma^* \frac{R}{N}\right)^2 + \frac{A^2}{2} \left(1 - e^{-\left(\alpha\sigma^* \frac{R}{N}\right)^2}\right) \left(1 - \cos\left(2\alpha y_i\right) e^{-\left(\alpha\sigma^* \frac{R}{N}\right)^2}\right) + 2\alpha A \left(\sigma^* \frac{R}{N}\right)^2 e^{-\frac{1}{2}\left(\alpha\sigma^* \frac{R}{N}\right)^2} \left(\alpha \left(\sigma^* \frac{R}{N}\right)^2 \cos\left(\alpha y_i\right) + 2y_i \sin\left(\alpha y_i\right)\right).$$
(C.111)

Applying expansion (C.110) and collecting higher order terms one gets

$$D_Q^2 = \sum_{i=1}^N 4y_i^2 \left(\sigma^* \frac{R}{N}\right)^2 + O\left(\frac{R^4}{N^4}\right) \\ + \frac{A^2}{2} \left(1 - \left[1 - \left(\alpha\sigma^* \frac{R}{N}\right)^2 + O\left(\frac{R^4}{N^4}\right)\right]\right) \left(1 - \cos\left(2\alpha y_i\right) \left[1 - \left(\alpha\sigma^* \frac{R}{N}\right)^2 + O\left(\frac{R^4}{N^4}\right)\right]\right) \quad (C.112) \\ + 2\alpha A \left(\sigma^* \frac{R}{N}\right)^2 \left[1 - \frac{1}{2} \left(\alpha\sigma^* \frac{R}{N}\right)^2 + O\left(\frac{R^4}{N^4}\right)\right] \left(\alpha \left(\sigma^* \frac{R}{N}\right)^2 \cos\left(\alpha y_i\right) + 2y_i \sin\left(\alpha y_i\right)\right).$$

The summand with prefactor $A^2/2$ of (C.112) yields

3.7

$$\frac{A^2}{2} \left(1 - \left[1 - \left(\alpha \sigma^* \frac{R}{N} \right)^2 + O\left(\frac{R^4}{N^4} \right) \right] \right) \left(1 - \cos\left(2\alpha y_i \right) \left[1 - \left(\alpha \sigma^* \frac{R}{N} \right)^2 + O\left(\frac{R^4}{N^4} \right) \right] \right) \\
= \frac{A^2}{2} \left(\left(\alpha \sigma^* \frac{R}{N} \right)^2 + O\left(\frac{R^4}{N^4} \right) \right) \left(1 - \cos\left(2\alpha y_i \right) + \cos\left(2\alpha y_i \right) \left(\alpha \sigma^* \frac{R}{N} \right)^2 + O\left(\frac{R^4}{N^4} \right) \right) \\
= \frac{A^2}{2} \left(\alpha \sigma^* \frac{R}{N} \right)^2 \left(1 - \cos\left(2\alpha y_i \right) \right) + O\left(\frac{R^4}{N^4} \right) \\
= \left(A\alpha \sigma^* \frac{R}{N} \right)^2 \sin^2(\alpha y_i) + O\left(\frac{R^4}{N^4} \right).$$
(C.113)

For the last line of (C.113) it was used that $1 - \cos 2x = 2 \sin^2 x$. The last summand of (C.112) yields

$$2\alpha A \left(\sigma^* \frac{R}{N}\right)^2 \left[1 - \frac{1}{2} \left(\alpha \sigma^* \frac{R}{N}\right)^2 + O\left(\frac{R^4}{N^4}\right)\right] \left(\alpha \left(\sigma^* \frac{R}{N}\right)^2 \cos\left(\alpha y_i\right) + 2y_i \sin\left(\alpha y_i\right)\right)$$

= $2\alpha A \left(\sigma^* \frac{R}{N}\right)^2 2y_i \sin\left(\alpha y_i\right) + O\left(\frac{R^4}{N^4}\right).$ (C.114)

Collecting results (C.113) and (C.114) the variance simplifies

$$D_Q^2 = \sum_{i=1}^{N} 4y_i^2 \left(\sigma^* \frac{R}{N}\right)^2 + 2\alpha A \left(\sigma^* \frac{R}{N}\right)^2 2y_i \sin(\alpha y_i) + \left(\sigma^* \frac{R}{N}\right)^2 (\alpha A \sin(\alpha y_i))^2 + O\left(\frac{R^4}{N^4}\right) = \left(\sigma^* \frac{R}{N}\right)^2 \sum_{i=1}^{N} (2y_i + \alpha A \sin(\alpha y_i))^2 + \sum_{i=1}^{N} O\left(\frac{R^4}{N^4}\right) = \left(\sigma^* \frac{R}{N}\right)^2 \sum_{i=1}^{N} (f_i')^2 + O\left(\frac{R^4}{N^3}\right).$$
(C.115)

using definition (23) for the derivative f'_i . Given (C.115), the scaling of $\sum_{i=1}^N (f'_i)^2$ w.r.t. N and R can be deduced applying the triangle inequality. Considering the positional vector \mathbf{y} and definition $\sin(\alpha \mathbf{y}) \coloneqq \sin(\alpha y_1)\mathbf{e}_1 + \sin(\alpha y_2)\mathbf{e}_2 + \ldots + \sin(\alpha y_N)\mathbf{e}_N$ with \mathbf{e}_i being the *i*-th unit vector, one has

$$\sum_{i=1}^{N} (f_i')^2 = \|2\mathbf{y} + \alpha A \sin(\alpha \mathbf{y})\|^2.$$
(C.116)

Using inequality $\|\mathbf{a} + \mathbf{b}\| \le \|\mathbf{a}\| + \|\mathbf{b}\|$ and therefore $\|\mathbf{a} + \mathbf{b}\|^2 \le (\|\mathbf{a}\| + \|\mathbf{b}\|)^2$, and using $\|\mathbf{y}\|^2 = R^2$, an upper bound for expression (C.116) can be given as

$$\begin{aligned} \|2\mathbf{y} + \alpha A\sin\left(\alpha\mathbf{y}\right)\|^{2} &\leq 4\|\mathbf{y}\|^{2} + 4\alpha A\|\mathbf{y}\|\|\sin\left(\alpha\mathbf{y}\right)\| + (\alpha A)^{2}\|\sin\left(\alpha\mathbf{y}\right)\|^{2} \\ &= 4R^{2} + 4\alpha AR\sqrt{\sum_{i=1}^{N}\sin^{2}\left(\alpha y_{i}\right)} + (\alpha A)^{2}\sum_{i=1}^{N}\sin^{2}\left(\alpha y_{i}\right) \\ &\leq 4R^{2} + 4\alpha AR\sqrt{N} + (\alpha A)^{2}N = (2R + \alpha A\sqrt{N})^{2}. \end{aligned}$$
(C.117)

From (C.116) and (C.117) one can deduce the (upper bound) scaling $\sum_{i=1}^{N} (f'_i)^2 = O(N)$, which is valid for both constant R and $R = O(\sqrt{N})$. Applying the scaling to the result of (C.115), one obtains the scaling

relation of D_Q^2 for large N by neglecting higher orders as

$$D_Q^2 \simeq \frac{(\sigma^* R)^2}{N}.$$
(C.118)

Having obtained the scaling of D_Q^2 , the terms within $\{\cdot\}$ of (47) are investigated. The term with $e_{\vartheta}^{1,1}$ is easily evaluated. Inserting relation (C.118) for D_Q^2 and $\sigma = \sigma^* R/N$ one gets

$$e_{\vartheta}^{1,1} \frac{\sigma^2}{D_Q^2} (2y_i)^2 = e_{\vartheta}^{1,1} \frac{\left(\sigma^* \frac{R}{N}\right)^2}{\frac{(\sigma^* R)^2}{N}} (2y_i)^2 = e_{\vartheta}^{1,1} \frac{(2y_i)^2}{N} = O\left(\frac{1}{N}\right).$$
(C.119)

The second term with c_{ϑ}/D_Q is evaluated using normalization (C.109) and expansion (C.110) as

$$\frac{c_{\vartheta}}{D_Q} \left\{ 3 \left(\sigma^* \frac{R}{N} \right)^2 + A \cos\left(\alpha y_i\right) \left(1 - \left[1 - \frac{1}{2} \left(\alpha \sigma^* \frac{R}{N} \right)^2 + O\left(\frac{R^4}{N^4} \right) \right] \right) \\
+ \left(\alpha \sigma^* \frac{R}{N} \right)^2 \left[1 - \frac{1}{2} \left(\alpha \sigma^* \frac{R}{N} \right)^2 + O\left(\frac{R^4}{N^4} \right) \right] \right) \right\}$$

$$= \frac{c_{\vartheta}}{D_Q} \left\{ 3 \left(\sigma^* \frac{R}{N} \right)^2 + A \cos\left(\alpha y_i\right) \left(\frac{1}{2} \left(\alpha \sigma^* \frac{R}{N} \right)^2 + \left(\alpha \sigma^* \frac{R}{N} \right)^2 + O\left(\frac{R^4}{N^4} \right) \right) \right\}$$

$$= c_{\vartheta} \frac{\left(\sigma^* \frac{R}{N} \right)^2}{D_Q} \left\{ 3 + \frac{3}{2} \alpha^2 A \cos\left(\alpha y_i\right) + O\left(\frac{R^2}{N^2} \right) \right\}.$$
(C.120)

Inserting $D_Q \simeq \sigma^* R / \sqrt{N}$ from (C.118) into (C.120) yields

$$c_{\vartheta} \frac{\left(\sigma^* \frac{R}{N}\right)^2}{\frac{\sigma^* R}{\sqrt{N}}} \left\{ 3 + \frac{3}{2} \alpha^2 A \cos\left(\alpha y_i\right) + O\left(\frac{R^2}{N^2}\right) \right\}$$

$$= O\left(\frac{R}{N^{3/2}}\right) = \begin{cases} O\left(\frac{1}{N^{3/2}}\right) & \text{if } R = \text{const.} \\ O\left(\frac{1}{N}\right) & \text{if } R = O(\sqrt{N}). \end{cases}$$
(C.121)

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

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

$$= (\mu - 1)e_{\vartheta}^{2,0} \frac{\left(\sigma^* \frac{R}{N}\right)^2}{D_Q^2} \left(2y_i + \alpha A\sin\left(\alpha y_i\right) + O\left(\frac{R^2}{N^2}\right)\right)^2.$$
(C.122)

Using scaling (C.118) for D_Q^2 and writing $\mu(N)$ to denote the (unknown) population dependency on N one gets

$$(\mu - 1)e_{\vartheta}^{2,0} \frac{\left(\sigma^* \frac{R}{N}\right)^2}{\left(\frac{\sigma^* R}{N}\right)^2} \left(2y_i + \alpha A \sin\left(\alpha y_i\right) + O\left(\frac{R^2}{N^2}\right)\right)^2$$

= $\mu(N)O\left(\frac{1}{N}\right) = \begin{cases} O\left(\frac{1}{N}\right) & \text{if } \mu(N) = \text{const.} \\ O\left(\frac{\mu(N)}{N}\right) & \text{else.} \end{cases}$ (C.123)

Finally, inserting the scaling results for the three terms (C.119), (C.121), and (C.123) back into $\{\cdot\}$ of the quadratic progress rate (47), one gets for large dimensionality $N \to \infty$ and residual distance scaling $R = O(\sqrt{N})$ the relation

$$\varphi_i^{\mathrm{II}} = c_\vartheta \frac{\sigma^2}{D_Q} \left(4y_i^2 + \mathrm{e}^{-\frac{1}{2}(\alpha\sigma)^2} 2\alpha A y_i \sin\left(\alpha y_i\right) \right) - \frac{\sigma^2}{\mu} \left\{ 1 + O\left(\frac{1}{N}\right) + O\left(\frac{\mu(N)}{N}\right) \right\}.$$
 (C.124)

Provided that the population size $\mu = o(N)$, i.e. increasing sub-linearily with N, all terms except "1" can be neglected for $N \to \infty$. Recalling that $(\mu, \lambda) \to \infty$ with constant $\vartheta = \mu/\lambda$ and the aforementioned conditions, the final result yields

$$\varphi_i^{\text{II}} = c_\vartheta \frac{\sigma^2}{D_Q} \left(4y_i^2 + e^{-\frac{1}{2}(\alpha\sigma)^2} 2\alpha A y_i \sin\left(\alpha y_i\right) \right) - \frac{\sigma^2}{\mu}.$$
(C.125)

Appendix D Identities

Identity. For real parameters a and b it holds

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 \mathrm{e}^{-\frac{1}{2}t^2} \Phi(at+b) \,\mathrm{d}t = \Phi\left(\frac{b}{(1+a^2)^{1/2}}\right) - \frac{1}{\sqrt{2\pi}} \frac{a^2b}{(1+a^2)^{3/2}} \mathrm{e}^{-\frac{1}{2}\frac{b^2}{1+a^2}}.$$
 (D.1)

Proof. It can be proven starting from the known identity [3, Eq. (A.9)]

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 \mathrm{e}^{-\frac{1}{2}t^2} \mathrm{e}^{-\frac{1}{2}(at+b)^2} \,\mathrm{d}t = \frac{1+a^2+a^2b^2}{(1+a^2)^{5/2}} \mathrm{e}^{-\frac{1}{2}\frac{b^2}{1+a^2}}.$$
 (D.2)

Both sides can be integrated with respect to b, such that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 \mathrm{e}^{-\frac{1}{2}t^2} \int_{-\infty}^{b'} \mathrm{e}^{-\frac{1}{2}(at+b)^2} \,\mathrm{d}b \,\mathrm{d}t = \int_{-\infty}^{b'} \left[\frac{1+a^2}{(1+a^2)^{5/2}} + \frac{a^2b^2}{(1+a^2)^{5/2}} \right] \mathrm{e}^{-\frac{1}{2}\frac{b^2}{1+a^2}} \,\mathrm{d}b \tag{D.3}$$

Integration of left-hand side yields simply

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 \mathrm{e}^{-\frac{1}{2}t^2} \int_{-\infty}^{b'} \mathrm{e}^{-\frac{1}{2}(at+b)^2} \,\mathrm{d}b \,\mathrm{d}t = \int_{-\infty}^{\infty} t^2 \mathrm{e}^{-\frac{1}{2}t^2} \Phi(at+b') \,\mathrm{d}t \,, \tag{D.4}$$

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

$$\frac{1}{(1+a^2)^{3/2}} \int_{-\infty}^{b'} e^{-\frac{1}{2}\frac{b^2}{1+a^2}} db = \frac{\sqrt{2\pi}}{1+a^2} \Phi\left(\frac{b'}{(1+a^2)^{1/2}}\right).$$
 (D.5)

For the second term on the right-hand side following integral is used

$$\int x e^{-\frac{1}{2}\frac{x^2}{s^2}} dx = \int x \frac{s^2}{x} e^{-y} dy = -s^2 e^{-y} = -s^2 e^{-\frac{1}{2}\frac{x^2}{s^2}},$$
(D.6)

using the substitution $y = \frac{x^2}{2s^2}$ with $dx = s^2 dy / x$. The second term of right-hand side of (D.3) is partially integrated using (D.6), such that

$$\frac{a^2}{(1+a^2)^{5/2}} \int_{-\infty}^{b'} b \left[b e^{-\frac{1}{2} \frac{b^2}{1+a^2}} \right] db$$

$$= \frac{a^2}{(1+a^2)^{5/2}} \left\{ \left[-b(1+a^2) e^{-\frac{1}{2} \frac{b^2}{1+a^2}} \right]_{-\infty}^{b'} + \int_{-\infty}^{b'} (1+a^2) e^{-\frac{1}{2} \frac{b^2}{1+a^2}} db \right\}$$

$$= -\frac{a^2 b'}{(1+a^2)^{3/2}} e^{-\frac{1}{2} \frac{b'^2}{1+a^2}} + a^2 \frac{\sqrt{2\pi}}{1+a^2} \Phi \left(\frac{b'}{(1+a^2)^{1/2}} \right).$$
(D.7)

Adding results (D.5) and (D.7) for the right-hand side, 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.

References

- 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] H.-G. Beyer. The Theory of Evolution Strategies. Natural Computing Series. Springer, Heidelberg, 2001. DOI: 10.1007/978-3-662-04378-3.
- [4] 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.
- [5] H.-G. Beyer and H.-P. Schwefel. Evolution Strategies: A Comprehensive Introduction. Natural Computing, 1(1):3–52, 2002.
- [6] H.-G. Beyer and B. Sendhoff. Toward a Steady-State Analysis of an Evolution Strategy on a Robust Optimization Problem with Noise-Induced Multi-Modality. *IEEE Transactions on Evolutionary Computation*, 21(4):629–643, 2017. DOI: 10.1109/TEVC.2017.2668068.
- [7] N. Hansen and S. Kern. Evaluating the CMA Evolution Strategy on Multimodal Test Functions. In X. Yao et al., editor, *Parallel Problem Solving from Nature 8*, pages 282–291, Berlin, 2004. Springer.
- [8] A. Melkozerov and H.-G. Beyer. On the Analysis of Self-Adaptativ Evolution Strategies on Elliptic Model: First Results. In J. Branke et al., editor, *GECCO'10: Proceedings of the Genetic and Evolu*tionary Computation Conference, pages 369–376, New York, 2010. ACM.
- [9] S. Meyer-Nieberg. Self-Adaptation in Evolution Strategies. PhD thesis, University of Dortmund, CS Department, Dortmund, Germany, 2007.
- [10] N. Müller and T. Glasmachers. Non-local optimization: imposing structure on optimization problems by relaxation. In *Foundations of Genetic Algorithms*, 16, pages 1–10. ACM, 2021.
- [11] A. Omeradzic and H.-G. Beyer. Progress Rate Analysis of Evolution Strategies on the Rastrigin Function: First Results. In G. Rudolph, A. V. Kononova, H. Aguirre, P. Kerschke, G. Ochoa, and T. Tušar, editors, *Parallel Problem Solving from Nature – PPSN XVII*, pages 499–511. Springer International Publishing, 2022.