

Analysis of the $(\mu/\mu_I, \lambda)$ - σ -Self-Adaptation Evolution Strategy with Repair by Projection Applied to a Conically Constrained Problem

Patrick Spettel and Hans-Georg Beyer

Abstract—A theoretical performance analysis of the $(\mu/\mu_I, \lambda)$ - σ -Self-Adaptation Evolution Strategy (σ SA-ES) is presented considering a conically constrained problem. Infeasible offspring are repaired using projection onto the boundary of the feasibility region. Closed-form approximations are used for the one-generation progress of the evolution strategy. Approximate deterministic evolution equations are formulated for analyzing the strategy's dynamics. By iterating the evolution equations with the approximate one-generation expressions, the evolution strategy's dynamics can be predicted. The derived theoretical results are compared to experiments for assessing the approximation quality. It is shown that in the steady state the $(\mu/\mu_I, \lambda)$ - σ SA-ES exhibits a performance as if the ES were optimizing a sphere model. Unlike the non-recombinative $(1, \lambda)$ -ES, the parental steady state behavior does not evolve on the cone boundary but stays away from the boundary to a certain extent.

Index Terms—Evolution strategies, repair by projection, conically constrained problem, intermediate recombination.

I. INTRODUCTION

CURRENT research in evolution strategies (ESs) includes the design and analysis of evolution strategies applied to constrained optimization problems. It is of particular interest to gain a deep understanding of evolution strategies on such problems. The insights gained from theory can help to provide guidance in applying evolution strategies to real world constrained problems. First, theory can show what kind of problems are suitably solved by evolution strategies. Moreover, theoretical investigations can guide the design of ES algorithms (for an example, see [1]). Furthermore, it is possible to provide optimal parameter settings based on the theoretically derived suggestions.

The aim of this work to derive (approximate) closed-form expressions for describing the behavior (one-generation

behavior, evolution dynamics, steady state behavior) of the $(\mu/\mu_I, \lambda)$ - σ -Self-Adaptation ES applied to a specific conically constrained problem. The outcomes are useful for algorithm design because the considered problem models the situation of an optimum lying on the constraint boundary. The algorithm needs to adapt the mutation strength to reach the optimum while it has to avoid premature convergence.

Arnold has analyzed a $(1, \lambda)$ -ES with constraint handling by resampling for a single linear constraint [2]. This has been extended in [3] with the analysis of repair by projection for a single linear constraint. This repair approach has been compared with an approach that reflects infeasible points into the feasible region and an approach that truncates infeasible points in [4]. Another idea for constraint handling based on augmented Lagrangian constraint handling has been presented in [5] for a $(1 + 1)$ -ES. There, a single linear inequality constraint with the sphere model is considered. For this, the one-generation behavior was analyzed.

A multi-recombinative variant of this algorithm has been presented in [6] for a single linear constraint and multiple linear constraints in [7]. Markov chains have been used in both cases for a theoretical investigation.

A conically constrained problem is considered in [8]. A very specific case of this problem, in which the objective function gradient coincides with the axis of the conic constraint, is investigated in [9]. For both problem variants, a $(1, \lambda)$ -ES with cumulative step size adaptation (CSA) is considered, in which infeasible offspring are discarded until feasible ones are obtained. In those papers, the one-generation behavior and the steady state behavior are investigated by numerical integration of expected value expressions. The same very specific problem has been investigated in [10] with repair by projection for infeasible offspring. Closed-form approximate expressions for the one-generation behavior, the evolution dynamics, and the steady state behavior have been derived. A remarkable result has been shown. The $(1, \lambda)$ - σ -Self-Adaptation ES with repair by projection applied to the very specific conically constrained problem converges towards the optimizer as if a sphere were to be optimized. This paper extends that work to the multi-recombinative variant, the $(\mu/\mu_I, \lambda)$ - σ -Self-Adaptation ES, which is particularly interesting to analyze for the comparison to the non-recombinative variant. A further interest in the multi-recombinative variant is to understand whether the advantages of the multi-recombinative ES applied to the

Manuscript received Month xx, xxxx; revised Month xx, xxxx and Month xx, xxxx; accepted Month xx, xxxx. Date of publication Month xx, xxxx; date of current version Month xx, xxxx. This work was supported by the Austrian Science Fund FWF under grant P29651-N32. (Corresponding author: Patrick Spettel.)

The authors are with the Research Center Process and Product Engineering at the Vorarlberg University of Applied Sciences, Dornbirn, Austria (e-mail: patrick.spettel@fhv.at; hans-georg.beyer@fhv.at).

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the author. This consists of an appendix containing mathematical derivations and additional figures. This material is x.xx MB in size.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier xx.xxxx/TEVC.xxxx.xxxxxxx

xxxx-xxxx © xxxx IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

unconstrained sphere model [11, Chapter 6] apply to the conically constrained problem as well. In particular, (approximate) closed-form expressions for describing the one-generation and the multi-generation behavior are derived and compared to real ES simulations. The results are discussed with a particular emphasis on the connection to the unconstrained sphere model.

The rest of the paper is organized as follows. The optimization problem is described in Section II. This is followed by a presentation of the algorithm under consideration in Section III. Next, the theoretical results are presented. First, the dynamical systems analysis approach is briefly recapped in Section IV-A. Then, the algorithm's behavior from one generation to the next (microscopic behavior) is investigated in Section IV-B. This is followed by the multi-generation behavior (macroscopic behavior) in Section IV-C. Closed-form approximations under asymptotic assumptions are derived for the microscopic behavior. They are then used in deterministic evolution equations. These evolution equations are iterated in order to predict the mean value dynamics of the ES. The approximations are compared to simulations. Finally, the insights gained are discussed in Section V. In particular, the differences and similarities of the $(1, \lambda)$ -ES and the multi-recombinative variant are discussed.

II. OPTIMIZATION PROBLEM

The optimization problem under consideration is

$$f(\mathbf{x}) = x_1 \rightarrow \min! \quad (1)$$

$$\text{s.t. } x_1^2 - \xi \sum_{k=2}^N x_k^2 \geq 0 \quad (2)$$

$$\text{and } x_1 \geq 0 \quad (3)$$

where $\mathbf{x} = (x_1, \dots, x_N)^T \in \mathbb{R}^N$ and $\xi > 0$. This formulation results in a feasible region that is a so-called second order cone.

The distance x from 0 in x_1 -direction (direction of the cone axis \mathbf{e}_1) and the distance r from the cone's axis \mathbf{e}_1 suffices to describe the state of the ES in the search space. This is denoted as the $(x, r)^T$ -space in the following. Moreover, note that due to the isotropy of the mutations used in the ES, the coordinate system can w.l.o.g. be rotated. Let $\tilde{\mathbf{x}}$ be an individual in the search space with $\tilde{x} = (\tilde{\mathbf{x}})_1$ and $\tilde{r} = \sqrt{\sum_{k=2}^N (\tilde{\mathbf{x}})_k^2}$. Due to the isotropic mutations in the ES, the coordinate system can be rotated around the cone axis \mathbf{e}_1 such that \mathbf{e}_2 points to $(\tilde{\mathbf{x}})_{2..N}$. That is, in this rotated coordinate system, $\tilde{\mathbf{x}} = (\tilde{x}, \tilde{r}, 0, \dots, 0)^T$.

Fig. 1 visualizes the optimization problem. The feasible region comprises the region inside and on the cone. It is horizontally hatched. Its boundary is described by the equation $r = \frac{x_1}{\sqrt{\xi}}$, which follows from Equation (2). One arrives at the equation of the projection line by deriving the equation for the shortest distance to the cone boundary. It is referred to Appendix A for the details. Solving Equation (A.54) for the distance from the cone axis results in $r = -\sqrt{\xi}x_1 + q\left(\sqrt{\xi} + \frac{1}{\sqrt{\xi}}\right)$ as the equation for the projection line. A parent \mathbf{x} and an offspring $\tilde{\mathbf{x}}$ with the mutation $\tilde{\sigma}\mathbf{z}$ are visualized as

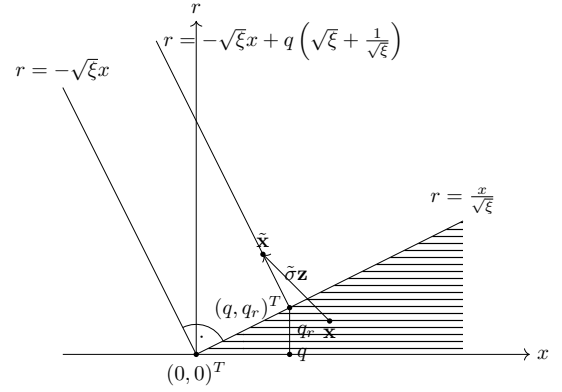


Fig. 1. The conically constrained optimization problem in N dimensions shown in the $(x, r)^T$ -space. The feasible region comprises the region inside and on the cone. It is horizontally hatched.

well. The values q and q_r denote the x_1 and r values after the projection step.

III. ALGORITHM

A $(\mu/\mu_I, \lambda)$ - σ -Self-Adaptation ES is applied to the optimization problem described in Section II. Algorithm 1 shows the pseudo code¹. First, parameters are initialized (lines 1 to 2). After that, the generation loop is entered. λ offspring are generated in lines 6 to 15. A log-normal distribution is used for every offspring to compute its mutation strength $\tilde{\sigma}_l$ by mutating the parental mutation strength $\sigma^{(g)}$ (line 7). Using this determined mutation strength, the offspring's parameter vector is sampled from a multivariate normal distribution with mean $\mathbf{x}^{(g)}$ and standard deviation $\tilde{\sigma}_l$ (line 8). Then, in lines 9 to 11, the offspring is repaired by projection, if necessary ($\text{isFeasible}(\mathbf{x}) = x_1 \geq 0 \wedge x_1^2 - \xi \sum_{k=2}^N x_k^2 \geq 0$). This means that for infeasible offspring, the optimization problem

$$\begin{aligned} \hat{\mathbf{x}} &= \arg \min_{\mathbf{x}'} \|\mathbf{x}' - \mathbf{x}\|^2 \\ \text{s.t. } x_1'^2 - \xi \sum_{k=2}^N x_k'^2 &\geq 0 \\ x_1' &\geq 0 \end{aligned} \quad (4)$$

must be solved where \mathbf{x} is the individual to be projected. For this, a helper function

$$\hat{\mathbf{x}} = \text{projectOntoCone}(\mathbf{x}) \quad (5)$$

is introduced that returns $\hat{\mathbf{x}}$ of (4). A derivation for a closed-form solution to this projection optimization problem (4) is presented in the supplementary material in Appendix A. $x^{(g)}$, $r^{(g)}$, q_l , q'_l , $\langle q \rangle$, and $\langle q_r \rangle$ (lines 4, 5, 13, 14, 19, and 20) are values used in the theoretical analysis. They can be removed in practical implementations of the ES. Table I gives an overview of their definitions. The offspring's fitness is computed in line 12. The next generation's parental individual $\mathbf{x}^{(g+1)}$ (line 17) and the next generation's mutation strength $\sigma^{(g+1)}$ (line 18) are updated next. The parental parameter

¹ $(\mathbf{x})_k$ denotes the k -th element of a vector \mathbf{x} . $\mathbf{x}_{m;\lambda}$ is the order statistic notation. It denotes the m -th best (according to fitness) of λ elements.

symbol	explanation
$x^{(g)}$	first component of the parent (gen. g)
$r^{(g)}$	distance from the cone axis of the parent (gen. g)
q_l	l -th offspring's first component after repair
\mathbf{q}'_l	l -th offspring's parameter vector after repair
$\langle q \rangle, \langle q_r \rangle$	$\frac{1}{\mu} \sum_{m=1}^{\mu} q_m; \lambda, \left\ \frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{q}'_m; \lambda)_{2..N} \right\ $

TABLE I
DEFINITION OF SYMBOLS

vector for the next generation is set to the mean of the parameter vectors of the μ best offspring. Similarly, the parental mutation strength for the next generation is set to the mean of the mutation strengths of the μ best offspring. If the parental parameter vector for the next generation is not feasible, it is projected onto the cone (lines 21 to 23). Note that this repair step is not needed in the real implementation of the algorithm. Since the conical problem is convex, the intermediate recombination of the μ best feasible individuals is feasible as well. However, in the iteration of the evolution equations in Section IV-C, the $(x, r)^T$ can get infeasible due to the approximations used. Finally, the generation counter is updated, which completes one iteration of the generational loop.

Algorithm 1 Pseudo-code of the $(\mu/\mu_I, \lambda)$ - σ -Self-Adaptation ES with repair by projection applied to the conically constrained problem.

```

1: Initialize  $\mathbf{x}^{(0)}, \sigma^{(0)}, \tau, \lambda, \mu$ 
2:  $g \leftarrow 0$ 
3: repeat
4:    $x^{(g)} = (\mathbf{x}^{(g)})_1$ 
5:    $r^{(g)} = \sqrt{\sum_{k=2}^N (\mathbf{x}^{(g)})_k^2}$ 
6:   for  $l \leftarrow 1$  to  $\lambda$  do
7:      $\tilde{\sigma}_l \leftarrow \sigma^{(g)} e^{\tau \mathcal{N}(0,1)}$ 
8:      $\tilde{\mathbf{x}}_l \leftarrow \mathbf{x}^{(g)} + \tilde{\sigma}_l \mathcal{N}(\mathbf{0}, \mathbf{I})$ 
9:     if not isFeasible( $\tilde{\mathbf{x}}_l$ ) then
10:       $\tilde{\mathbf{x}}_l \leftarrow \text{projectOntoCone}(\tilde{\mathbf{x}}_l)$   $\triangleright$  see (4), (5)
11:     end if
12:      $\tilde{f}_l \leftarrow f(\tilde{\mathbf{x}}_l) = (\tilde{\mathbf{x}}_l)_1$ 
13:      $q_l = (\tilde{\mathbf{x}}_l)_1$ 
14:      $\mathbf{q}'_l = \tilde{\mathbf{x}}_l$ 
15:   end for
16:   Sort offspring according to  $\tilde{f}_l$  in ascending order
17:    $\mathbf{x}^{(g+1)} \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\mathbf{x}}_{m;\lambda}$ 
18:    $\sigma^{(g+1)} \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\sigma}_{m;\lambda}$ 
19:    $\langle q \rangle = (\mathbf{x}^{(g+1)})_1$ 
20:    $\langle q_r \rangle = \sqrt{\sum_{k=2}^N (\mathbf{x}^{(g+1)})_k^2}$ 
21:   if not isFeasible( $\mathbf{x}^{(g+1)}$ ) then
22:      $\mathbf{x}^{(g+1)} \leftarrow \text{projectOntoCone}(\mathbf{x}^{(g+1)})$   $\triangleright$  see (4), (5)
23:   end if
24:    $g \leftarrow g + 1$ 
25: until termination criteria are met

```

Fig. 2 shows an example of the x - and r -dynamics that are a result of running Algorithm 1 (solid line). The closed-form approximation iterative system (dotted line) (to be derived in

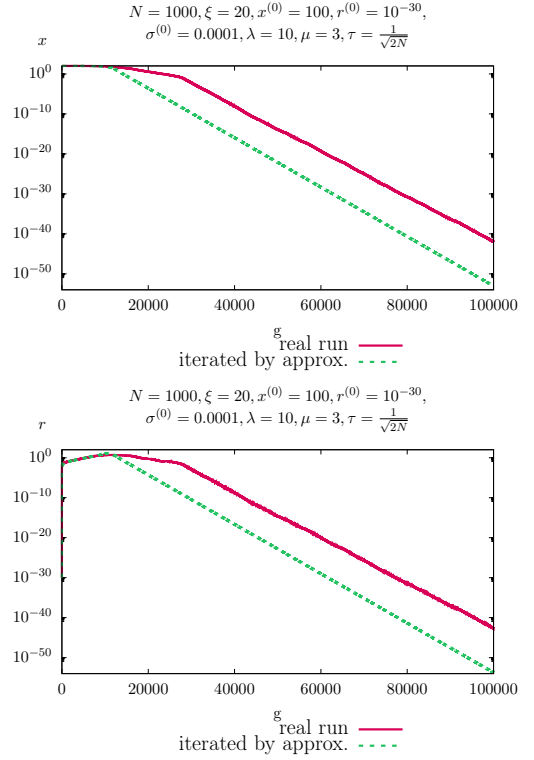


Fig. 2. Comparison of the x - and r -dynamics of a real $(3/3_I, 10)$ -ES run (solid line) with the iteration of the closed-form approximation of the iterative system (dotted line).

the following sections) is shown in comparison. The prediction does not coincide completely with the real run. The ES reaches the steady state later than predicted. The approximations that are derived in this work result in deviations in the transient phase of the ES. However, our main focus is in the steady state. There, the predicted slope of the x and r dynamics is very similar to the one of the real run.

IV. THEORETICAL ANALYSIS

A. The Dynamical Systems Approach

The $(x, r)^T$ -representation described in Section II is used for the analysis of the $(\mu/\mu_I, \lambda)$ -ES (Algorithm 1). The aim is to predict the evolution dynamics of the $(\mu/\mu_I, \lambda)$ -ES. To this end, the dynamical systems method described in [11] is used. For the case here, the three random variables for x , r , and σ describe the state of the system. A Markov process can be used for modeling the transitions between states from one time step to the next. It is often not possible to derive closed-form expressions for the transition densities. Therefore, approximate equations are aimed for. Evolution equations can be stated by making use of the local expected change functions $(\varphi_x, \varphi_r, \psi)$

$$x^{(g+1)} = x^{(g)} - \varphi_x(x^{(g)}, r^{(g)}, \sigma^{(g)}) + \epsilon_x(x^{(g)}, r^{(g)}, \sigma^{(g)}) \quad (6)$$

$$r^{(g+1)} = r^{(g)} - \varphi_r(x^{(g)}, r^{(g)}, \sigma^{(g)}) + \epsilon_r(x^{(g)}, r^{(g)}, \sigma^{(g)}) \quad (7)$$

$$\sigma^{(g+1)} = \sigma^{(g)} + \sigma^{(g)} \psi(x^{(g)}, r^{(g)}, \sigma^{(g)}) + \epsilon_\sigma(x^{(g)}, r^{(g)}, \sigma^{(g)}) \quad (8)$$

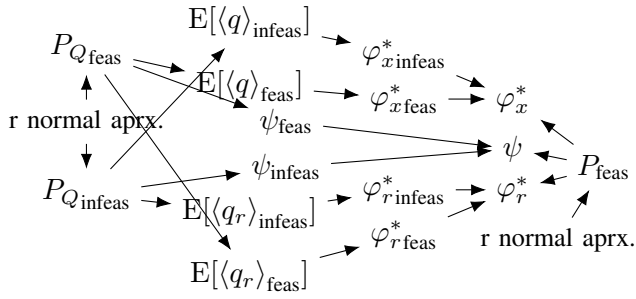


Fig. 3. Graph visualizing the dependencies between the different derivations for the microscopic aspects. An arrow from A to B indicates that A is required for B.

The progress rates are defined as

$$\varphi_x(x^{(g)}, r^{(g)}, \sigma^{(g)}) := E[x^{(g)} - x^{(g+1)} | x^{(g)}, r^{(g)}, \sigma^{(g)}] \quad (9)$$

$$\varphi_r(x^{(g)}, r^{(g)}, \sigma^{(g)}) := E[r^{(g)} - r^{(g+1)} | x^{(g)}, r^{(g)}, \sigma^{(g)}]. \quad (10)$$

They model the expected change in the parameter space from one generation to the next. As a formal decomposition into expected value and fluctuations, the random variables ϵ_x , ϵ_r , and ϵ_σ represent the stochastic part. There are no assumptions about those variables except that $E[\epsilon_x] = 0$, $E[\epsilon_r] = 0$, and $E[\epsilon_\sigma] = 0$ holds. For more details regarding this approach, it is referred to [11, Chapter 7].

In what follows, the normalizations

$$\varphi_x^* := \frac{N\varphi_x}{x^{(g)}}, \quad (11)$$

$$\varphi_r^* := \frac{N\varphi_r}{r^{(g)}}, \quad (12)$$

and

$$\sigma^* := \frac{N\sigma}{r^{(g)}} \quad (13)$$

will be used. For the change of the mutation strength from one generation to the next, a slightly different progress measure will be applied. It is the so-called self-adaptation response (SAR). Its definition [11, Chapter 7, Eq. 7.31] reads

$$\psi(x^{(g)}, r^{(g)}, \sigma^{(g)}) := E \left[\frac{\sigma^{(g+1)} - \sigma^{(g)}}{\sigma^{(g)}} \middle| x^{(g)}, r^{(g)}, \sigma^{(g)} \right]. \quad (14)$$

$\epsilon_x = 0$, $\epsilon_r = 0$, and $\epsilon_\sigma = 0$ is assumed in Equations (6) to (8) for the analysis in this work leading to so-called deterministic evolution equations. That is, the mean-value behavior is simulated with those equations. With the goal of arriving at approximate deterministic evolution equations, approximate expressions for the functions φ_x , φ_r , and ψ need to be derived first.

B. The Microscopic Aspects

The microscopic aspects describe the behavior of the evolution strategy from one generation to the next. They are expressed by the functions introduced in Section IV-A (Equations (9), (10), and (14)).

In the following analyses, two cases are treated separately. It turns out that if the parental individual is in the vicinity

of the cone boundary², the offspring feasibility probability tends to 0 for asymptotic considerations ($N \rightarrow \infty$) (refer to Section V for a discussion about this). This case is denoted as the “infeasible” case using subscripts “infeas” in the derivations for the quantities that are derived separately for both cases. The opposite case of an offspring being feasible with overwhelming probability (intuitively this means “being far away” from the cone boundary) is denoted as the “feasible” case using subscripts “feas” in the derivations for the quantities that are derived separately for both cases. The local progress measures and the SAR are derived separately for the case of being at the cone boundary and for the opposite case. In order to have one approximate closed-form expression, they are combined by weighting with the feasibility probability. It is referred to [10, Sec. 3.1.2.1.2.8, pp. 44-49] for further details. There, the feasibility probability has been derived as

$$P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \simeq \Phi \left[\frac{1}{\sigma^{(g)}} \left(\frac{x^{(g)}}{\sqrt{\xi}} - \bar{r} \right) \right] \quad (15)$$

where \bar{r} is the expected value of the r normal approximation and Φ is the cumulative distribution function of the standard normal variate. The asymptotic feasibility probability correspondence to [9, Eq. (11)] can be shown. In order to simplify the theoretical analysis, the distribution of the r component is approximated by a normal distribution $\mathcal{N}(\bar{r}, \sigma_r^2)$. The supplementary material (Appendix B) presents a detailed derivation of this approximation.

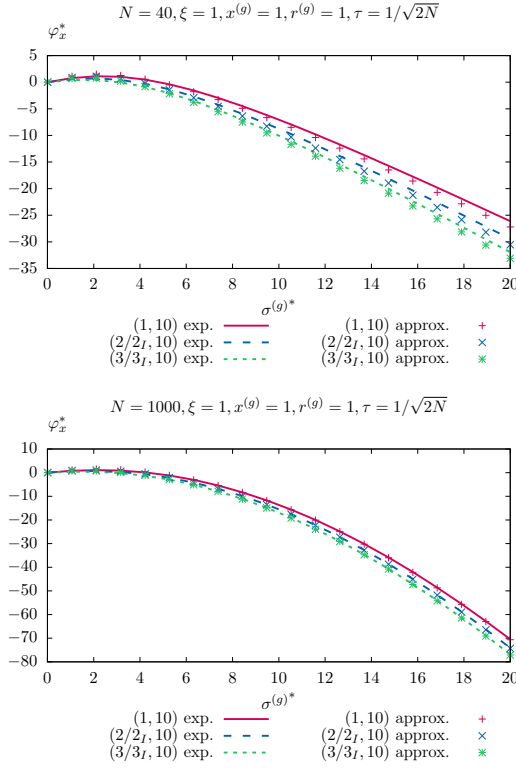
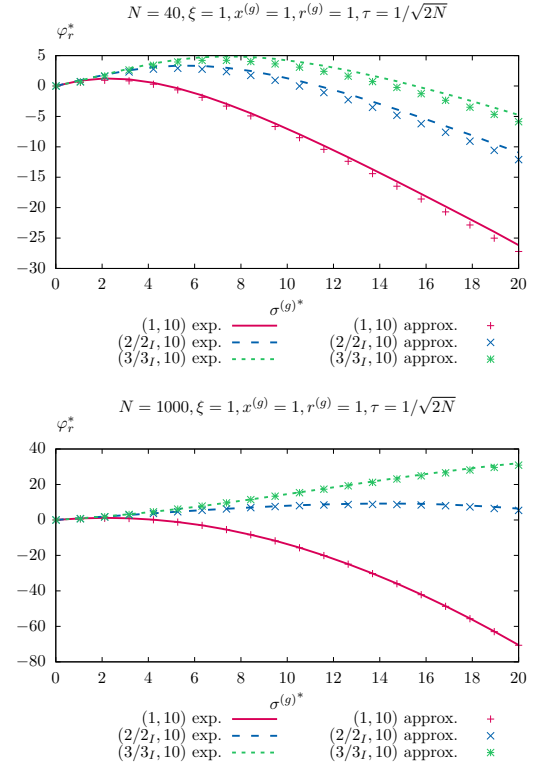
Fig. 3 shows a dependency graph of the different derivations for the microscopic aspects. Some of those derivations are presented in the appendix and cross-referenced where necessary in the following sections. Note especially that the cumulative distribution function of an offspring’s x -value after projection $P_Q(q)$ (with its variants $P_{Q_{\text{feas}}}(q)$ and $P_{Q_{\text{infeas}}}(q)$ for the two cases considered) is required for the derivations of the progress rates and the SAR. Approximate closed-form expressions are derived in the appendix (see Equations (D.90) to (D.119)). As a result, the errors introduced in those approximations have an impact for the derivations that depend on them. Concerning the approximation error, it has been shown in [10, Sec. 3.1.2.1.2, pp. 21-39] that for larger values of ξ , the approximation error gets smaller.

1) Derivation of the x Progress Rate: From the definition of the progress rate (Equation (9)) and the pseudo-code of the ES (Algorithm 1, lines 4 and 19),

$$\varphi_x(x^{(g)}, r^{(g)}, \sigma^{(g)}) = x^{(g)} - E[x^{(g+1)} | x^{(g)}, r^{(g)}, \sigma^{(g)}] \quad (16)$$

$$= x^{(g)} - E[\langle q \rangle | x^{(g)}, r^{(g)}, \sigma^{(g)}] \quad (17)$$

²Refer to the bottom-most subplot of Fig. 8 for a visualization of the algorithm’s distance to the cone boundary in the steady state. It is further discussed in Section V (see Equations (44) and (45) and the respective descriptions).

Fig. 4. Comparison of the x progress rate approximation with simulations.Fig. 5. Comparison of the r progress rate approximation with simulations.

follows. An approximation is derived in the supplementary material (Appendix E) resulting in

$$\begin{aligned} \varphi_x^* \approx & P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \frac{r^{(g)}}{x^{(g)}} \sigma^{(g)*} c_{\mu/\mu, \lambda} \\ & + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] \\ & \left[\frac{N}{1 + \xi} \left(1 - \frac{\sqrt{\xi} r^{(g)}}{x^{(g)}} \sqrt{1 + \frac{\sigma^{(g)*2}}{N}} \right) \right. \\ & \left. + \frac{\sqrt{\xi}}{1 + \xi} \frac{\sqrt{\xi} r^{(g)}}{x^{(g)}} \sigma^{(g)*} c_{\mu/\mu, \lambda} \sqrt{1 + \frac{1}{\xi} \frac{1 + \frac{\sigma^{(g)*2}}{2N}}{1 + \frac{\sigma^{(g)*2}}{N}}} \right] \end{aligned} \quad (18)$$

where $c_{\mu/\mu, \lambda} = e_{\mu, \lambda}^{1,0}$ is the so-called generalized progress coefficient [11, Eq. (5.112), p. 172]. Its definition is stated in the supplementary material (Equation (C.67)). The derivation for the approximation makes use of the r normal approximation and assumes sufficiently small τ (such that $\sigma_l \simeq \sigma^{(g)}$), $N \rightarrow \infty$, and $\sigma^{(g)*} \ll N$.

2) *Derivation of the r Progress Rate:* From the definition of the progress rate (Equation (10)) and the pseudo-code of the ES (Algorithm 1, lines 5 and 20), it follows that

$$\varphi_r(x^{(g)}, r^{(g)}, \sigma^{(g)}) = r^{(g)} - \mathbb{E}[r^{(g+1)} | x^{(g)}, r^{(g)}, \sigma^{(g)}] \quad (19)$$

$$= r^{(g)} - \mathbb{E}[\langle q_r \rangle | x^{(g)}, r^{(g)}, \sigma^{(g)}]. \quad (20)$$

An approximation is derived in the supplementary material (Appendix E) leading to

$$\begin{aligned} \varphi_r^* \approx & P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) N \left(1 - \sqrt{1 + \frac{\sigma^{(g)*2}}{\mu N}} \right) \\ & + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] N \left(1 - \frac{x^{(g)}}{\sqrt{\xi} r^{(g)}} \right. \\ & \left. \times \left(1 - \frac{\varphi_{x_{\text{infeas}}}^*}{N} \right) \sqrt{\frac{1 + \frac{\sigma^{(g)*2}}{\mu N}}{1 + \frac{\sigma^{(g)*2}}{N}}} \right) \end{aligned} \quad (21)$$

where the r normal approximation has been used. Furthermore, the derivation assumes sufficiently small τ (such that $\sigma_l \simeq \sigma^{(g)}$), $N \rightarrow \infty$, and variances of different random variables are assumed to tend to zero (it is referred to Appendix E for those details).

3) *Derivation of the SAR:* The derivation for an approximation of the SAR starting from the expected value in Equation (14) is presented in the supplementary material (Appendix F). It results in

$$\begin{aligned} \psi \approx & \tau^2 \left[\left(\frac{1}{2} + e_{\mu, \lambda}^{1,1} \right) \right] - [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] \\ & \times \frac{\sigma^{(g)*} c_{\mu/\mu, \lambda}}{\sqrt{1 + \xi}}. \end{aligned} \quad (22)$$

In that derivation, it has been assumed that $N \rightarrow \infty$, $\sigma^{(g)*2} \ll N$, and terms of a Taylor expansion (around 0 for functions with N in the denominator) that are higher than linear order in $\sigma^{(g)*}$ have been neglected.

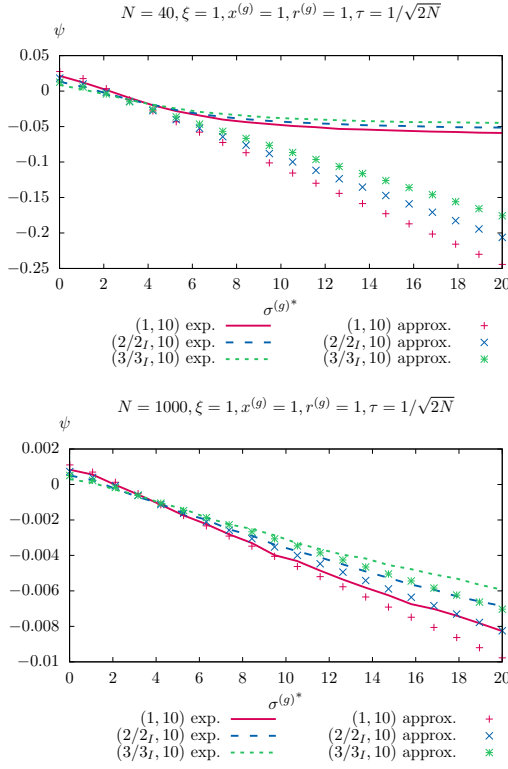


Fig. 6. Comparison of the SAR approximation with simulations. The rather large deviations of the theory from the simulations for larger values of $\sigma^{(g)*}$ are due to the linear approximation of the SAR function. Refer to the second paragraph of Section V for a discussion.

Figs. 4 to 6 show plots comparing the derived closed-form approximation for φ_x^* , φ_r^* , and ψ , respectively, with one-generation experiments. The pluses, crosses, and stars have been calculated by evaluating Equation (18) with Equations (15), (C.66), and (C.73), Equation (21) with Equations (15) and (C.73), and Equation (22) with Equation (15), respectively. The solid, dashed, and dotted lines have been generated by one-generation experiments. For this, the generational loop has been run 10^5 times for a fixed parental individual and constant parameters. The experimentally determined values for φ_x^* , φ_r^* , and ψ from those 10^5 runs have been averaged. As the standard deviation is relatively small for 10^5 runs, it is not shown. Further figures showing comparisons for additional configurations are provided in the supplementary material (Figs. 11 to 19 in Appendix I). The appropriate scaling of the mutation strength allows an arbitrary choice of $x^{(g)}$ for the one-generation experiments due to the symmetry of the problem. It is w.l.o.g. set to 1. The plots in the main text only show the case of the parental individual being on the cone boundary with $\xi = 1$. Hence, $r^{(g)} = x^{(g)}/\sqrt{\xi} = 1$. The plots in the supplementary material additionally show a case of being almost near the cone's axis and a case between the cone axis and the cone boundary.

C. The Evolution Dynamics in the Deterministic Approximation

1) *The Evolution Equations:* With the derived progress rates φ_x^* and φ_r^* and the SAR ψ from Section IV-B, the approx-

imated deterministic evolution equations can be iterated. By doing this, the dynamics can be predicted. The predictions by the approximations are compared to real runs of the ES. Using Equations (6) to (8) together with Equations (11) to (13), the evolution equations read

$$x^{(g+1)} = x^{(g)} - \frac{x^{(g)}\varphi_x^{(g)*}}{N} = x^{(g)} \left(1 - \frac{\varphi_x^{(g)*}}{N} \right) \quad (23)$$

$$r^{(g+1)} = r^{(g)} - \frac{r^{(g)}\varphi_r^{(g)*}}{N} = r^{(g)} \left(1 - \frac{\varphi_r^{(g)*}}{N} \right) \quad (24)$$

$$\sigma^{(g+1)} = \sigma^{(g)} + \sigma^{(g)}\psi^{(g)} = \sigma^{(g)} \left(1 + \psi^{(g)} \right). \quad (25)$$

Fig. 7 shows the mean value dynamics of the $(3/3_I, 10)$ -ES applied to the conically constrained problem for $N = 1000$, $\xi = 1$, and $\tau = \frac{1}{\sqrt{2N}}$. Additional figures for more configurations are provided in the supplementary material (Figs. 20 to 25 in Appendix I). The lines for the real runs have been generated by averaging 100 real runs of the ES. The fluctuations of those runs are visualized by the shaded areas. As the standard deviation around the mean can result in negative values, those show the range of the minimal and maximal values reached in the 100 runs. This allows using a logarithmic scale. Note that for the σ^* plot, the visualization of the fluctuations has been clipped. The lines for the iteration by approximation have been computed by iterating the mean value iterative system with the derived approximations in Section IV-B for $\varphi_x^{(g)*}$, $\varphi_r^{(g)*}$, and $\psi^{(g)}$. Note that due to the approximations used it is possible that the iteration of the mean value iterative system yields infeasible $(x^{(g)}, r^{(g)})^T$ in some generation g . If this situation has occurred, the corresponding $(x^{(g)}, r^{(g)})^T$ have been projected back. Then, in the following iterations, the projected values have been used. This is what is indicated by the repair step in lines 21 to 23 of Algorithm 1 that is only needed for the iteration of the mean value iterative system.

The steady state is reached by the ES (solid line) later than predicted (dotted line). This can be seen in the first three subplots of Fig. 7. The slope of the lines coincides well between the prediction and the ES. But the lines are shifted for some number of generations. The same can be observed in the fourth subplot of Fig. 7 where the σ^* steady state value is attained later by the ES (solid line) than predicted (dotted line).

2) *The ES in the Steady State:* The state of the $(\mu/\mu_I, \lambda)$ -ES on the conically constrained problem is completely described by $(x^{(g)}, r^{(g)}, \sigma^{(g)})^T$ (assuming constant exogenous parameters). For sufficiently large g , the ES transitions into a steady state, in which it can be described by two quantities, the normalized mutation strength σ^* and the ratio r/x . As it turns out, in the steady state the ES moves in the vicinity of the boundary. Therefore, x and r are dependent and the ratio r/x suffices. This results in two conditions for the ES in the steady state.

The first condition is regarding the mutation strength. On average, the normalized mutation strength should be constant in this steady state. That is, in the steady state the expectation

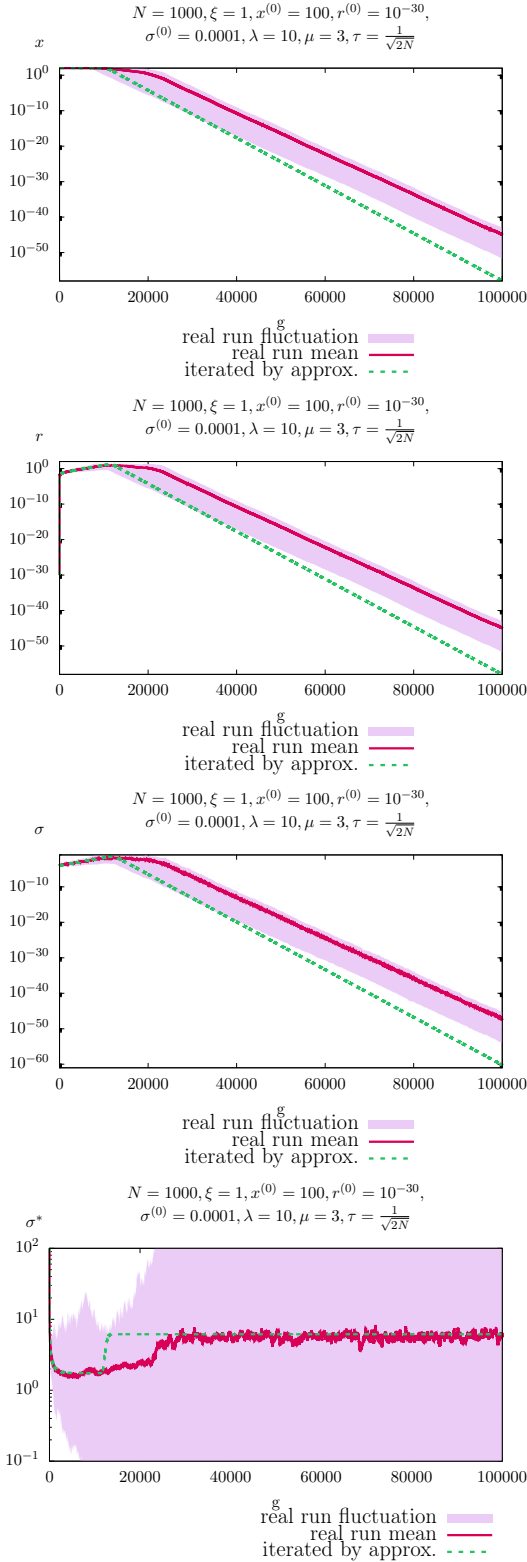


Fig. 7. Mean value dynamics closed-form approximation and real-run comparison of the $(3/3_f, 10)$ -ES with repair by projection applied to the conically constrained problem ($N = 1000$).

of the normalized mutation strength σ_{ss}^* is constant for large time scales

$$\sigma_{ss}^* := \lim_{g \rightarrow \infty} \sigma^{(g)*}. \quad (26)$$

Hence, for sufficiently large g , $\sigma^{(g)*} = \sigma^{(g+1)*} = \sigma_{ss}^*$. Requiring this condition, a steady state condition can be derived. To this end, Equation (25) is normalized yielding

$$\frac{r^{(g+1)} \sigma^{(g+1)*}}{N} = \frac{r^{(g)} \sigma^{(g)*}}{N} (1 + \psi^{(g)}). \quad (27)$$

Setting $\sigma^{(g)*} = \sigma^{(g+1)*}$ results in

$$r^{(g+1)} = r^{(g)} (1 + \psi^{(g)}). \quad (28)$$

Inserting Equation (24) yields the steady state condition

$$\frac{\varphi_r^{(g)*}}{N} = -\psi^{(g)}. \quad (29)$$

The second condition for the ES in the steady state is that it moves in the vicinity of the cone boundary. This can be observed in real ES runs (see the third rows in Figs. 20 to 25 in the supplementary material Appendix I). Hence, for the derivation of analytical approximations for the steady state, the condition $P_{\text{feas}} = 0$ is assumed similar to the $(1, \lambda)$ case. However, in contrast to the $(1, \lambda)$ case, the assumption $\frac{x^{(g)}}{\sqrt{\xi} r^{(g)}} = 1$ is too crude. There exists a distance from the cone boundary for some parameter configurations that cannot be neglected (this can for example be seen in the third row of Fig. 22 in the supplementary material Appendix I). Considering the distance ratio $r^{(g)}/x^{(g)}$, one observes that this ratio approaches a steady state value. This results in the condition

$$\frac{r^{(g)}}{x^{(g)}} = \frac{r^{(g+1)}}{x^{(g+1)}} \quad (30)$$

for sufficiently large values of g . Using the progress rates (Equations (9) to (12)), this can be written as

$$\frac{r_{ss}}{x_{ss}} = \frac{r_{ss} \left(1 - \frac{\varphi_{r_{ss}}^*}{N}\right)}{x_{ss} \left(1 - \frac{\varphi_{x_{ss}}^*}{N}\right)}, \quad (31)$$

which implies

$$\varphi_{r_{ss}}^* = \varphi_{x_{ss}}^*. \quad (32)$$

With the assumption $P_{\text{feas}} = 0$, use of the infeasible case approximations (Equation (C.73) and Equation (E.279)) for treating Equation (32) further, yields

$$N \left(1 - \frac{x_{ss}}{\sqrt{\xi} r_{ss}} \left(1 - \frac{\varphi_{x_{ss} \text{infeas}}^*}{N}\right) \sqrt{\frac{1 + \frac{\sigma_{ss}^{*2}}{\mu N}}{1 + \frac{\sigma_{ss}^{*2}}{N}}}\right) = \varphi_{x_{ss} \text{infeas}}^*. \quad (33)$$

Now, by distributing N over the terms in the bracket on the left-hand side and then solving for the fraction $\frac{x_{ss}}{\sqrt{\xi} r_{ss}}$, one gets³

$$\frac{x_{ss}}{\sqrt{\xi} r_{ss}} = \sqrt{\frac{1 + \frac{\sigma_{ss}^{*2}}{N}}{1 + \frac{\sigma_{ss}^{*2}}{\mu N}}}. \quad (34)$$

³Note that the quantity $\frac{x_{ss}}{\sqrt{\xi} r_{ss}}$ is a more formal notion of the vicinity of the cone boundary.

Considering Equation (32) for Equation (29) results in

$$\frac{\varphi_{x_{ss}}^*}{N} = -\psi_{ss}. \quad (35)$$

Assuming $P_{\text{feas}} = 0$, the approximations for the infeasible case can be used. Insertion of Equation (34) into Equation (C.73)

with the assumption $\frac{1}{\xi} \frac{1 + \frac{\sigma_{ss}^{*2}}{2N}}{1 + \frac{\sigma_{ss}^{*2}}{N}} \simeq \frac{1}{\xi}$ for $N \rightarrow \infty$ yields

$$\varphi_{x_{ss}}^* \approx \frac{N}{1 + \xi} \left(1 - \sqrt{1 + \frac{\sigma_{ss}^{*2}}{\mu N}} \right) + \frac{\sigma_{ss}^* c_{\mu/\mu, \lambda}}{\sqrt{1 + \xi}} \sqrt{\frac{1 + \frac{\sigma_{ss}^{*2}}{\mu N}}{1 + \frac{\sigma_{ss}^{*2}}{N}}} \quad (36)$$

$$\simeq \frac{N}{1 + \xi} \left(1 - \left(1 + \frac{\sigma_{ss}^{*2}}{2\mu N} \right) \right) + \frac{\sigma_{ss}^* c_{\mu/\mu, \lambda}}{\sqrt{1 + \xi}} \quad (37)$$

$$= \frac{N}{1 + \xi} \left(-\frac{\sigma_{ss}^{*2}}{2\mu N} \right) + \frac{\sigma_{ss}^* c_{\mu/\mu, \lambda}}{\sqrt{1 + \xi}}. \quad (38)$$

In the step from Equation (36) to Equation (37), it has been assumed that $\sqrt{\frac{1 + \frac{\sigma_{ss}^{*2}}{\mu N}}{1 + \frac{\sigma_{ss}^{*2}}{N}}} \simeq 1$ for $N \gg \sigma_{ss}^{*2}$ and a Taylor expansion (neglecting the quadratic and higher order terms) has been applied to $\sqrt{1 + \frac{\sigma_{ss}^{*2}}{\mu N}}$. With Equation (38), the steady state equation Equation (29) reads for the asymptotic case of $N \rightarrow \infty$

$$\begin{aligned} \frac{1}{1 + \xi} \left(-\frac{\sigma_{ss}^{*2}}{2\mu N} \right) + \frac{\sigma_{ss}^* c_{\mu/\mu, \lambda}}{N\sqrt{1 + \xi}} \\ = -\tau^2 \left(\frac{1}{2} + e_{\mu, \lambda}^{1,1} - \frac{\sigma_{ss}^* c_{\mu/\mu, \lambda}}{\sqrt{1 + \xi}} \right). \end{aligned} \quad (39)$$

Solving this quadratic equation yields the steady state normalized mutation strength (as the mutation strength is positive, the positive root is taken)

$$\begin{aligned} \sigma_{ss}^* = \sqrt{1 + \xi} \mu c_{\mu/\mu, \lambda} \left[(1 - N\tau^2) + \right. \\ \left. \sqrt{(1 - N\tau^2)^2 + \frac{2\tau^2 N \left(\frac{1}{2} + e_{\mu, \lambda}^{1,1} \right)}{\mu c_{\mu/\mu, \lambda}^2}} \right]. \end{aligned} \quad (40)$$

Fig. 8 shows plots of the steady state computations. The derived closed-form approximation has been compared to real ES runs. The values for visualizing the approximations have been calculated as follows. The steady state σ_{ss}^* value has been computed by evaluating Equation (40). This value was then inserted into the derived approximations for the progress measures to obtain the values for $\varphi_{x_{ss}}^*$ and $\varphi_{r_{ss}}^*$. The values from experiments have been determined by computing the averages of the particular values in real ES runs. For this, the ES was run for $400N$ generations and the particular values have been averaged over these generations. This procedure was repeated 100 times and those 100 times have been averaged to compute the particular steady state values.

V. RESULTS AND CONCLUSIONS

A $(\mu/\mu_I, \lambda)$ -ES applied to a conically constrained problem has been theoretically analyzed.

Approximate expressions for the local progress measures and the SAR have been derived for asymptotic considerations.

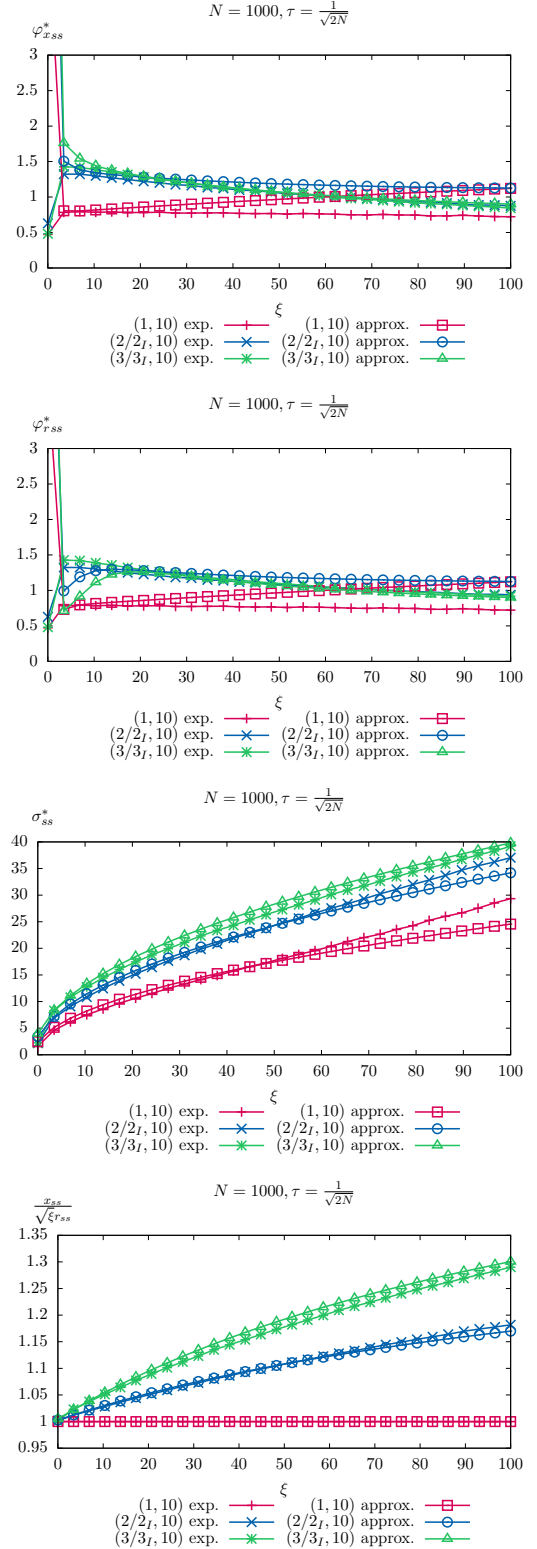


Fig. 8. Steady state closed-form approximation and real-run comparison of the $(\mu/\mu_I, \lambda)$ -ES with repair by projection applied to the conically constrained problem.

Comparison with numerical one-generation experiments shows the approximation quality. As can be seen in Fig. 4 and Fig. 5, the derived approximations for the x and r progress rates coincide well with the experiments, in particular for higher dimensions ($N = 1000$). In the transition from the feasible to the infeasible case, the weighting with the offspring feasibility probability does not fit well in all the cases. This can be seen in the supplementary material (Figs. 11 to 16 in Appendix I). The middle row in these figures always shows a case between the feasible and the infeasible case. However, the deviation in these cases does not affect the steady state considerations because the ES moves in the vicinity of the cone boundary in the steady state. The approximation for the SAR corresponds well to the experiments for small values of $\sigma^{(g)*}$. It deviates for larger values of $\sigma^{(g)*}$. Again, for higher N , the approximation quality fits better as can be seen in Fig. 6. This behavior is inherent in the SAR approximation. First, $N \gg \sigma^{(g)*}$ has been assumed in deriving the SAR approximation. Second, quadratic and higher order terms have been neglected. Therefore, the SAR approximation is linear and hence necessarily deviates for large values of $\sigma^{(g)*}$.

The approximate expressions for the local progress measures and the SAR derived for the $(\mu/\mu_I, \lambda)$ -ES generalize the results obtained for the $(1, \lambda)$ -ES. Considering $\mu = 1$ for Equation (C.66) and Equation (C.73) immediately results in [10, Eq. (3.120), p. 41] and [10, Eq. (3.149), p. 44], respectively, for the x progress rate. For the r progress rate, it is similar. Setting $\mu = 1$ in Equation (21) results in [10, Eq. (3.221), p. 57] for the feasible case and in [10, Eq. (3.209), p. 56] for the infeasible case (after applying the definition of φ^*). The same holds for ψ . It only differs in the progress coefficients. It can easily be verified that Equation (22) with $\mu = 1$ corresponds to [10, Eq. (3.281), p. 69].

For the macroscopic behavior, the local expected progress functions have been iterated to predict the mean value dynamics of the ES and analytic steady state expressions have been derived.

For the evolution dynamics, in Fig. 7 the rate of convergence from the prediction is very similar to the one exhibited by the real ES run. However, the theory predicts an earlier transitioning into the steady state (see the fourth subplot of Fig. 7 showing the σ^* dynamics).

The analytic steady state analysis resulted in a very interesting conclusion. It turns out that the astonishing result from the $(1, \lambda)$ case (see the derivations leading to [10, Eq. (3.308), p. 87]) extends to the multi-recombinative ES. One recognizes the equations for the sphere model in (40) (see [12, Eq. (4.11), p. 35]), i.e., Equation (40) can be written as

$$\sigma_{ss}^* = \sqrt{1 + \xi} \sigma_{ss\text{sphere}}^* \quad (41)$$

where $\sigma_{ss\text{sphere}}^*$ denotes the steady state mutation strength value for the $(\mu/\mu_I, \lambda)$ - σ SA-ES applied to the sphere model. This means that in the steady state the $(\mu/\mu_I, \lambda)$ - σ SA-ES with repair by projection applied to the conically constrained problem with the objective function gradient in direction of the x_1 -axis moves towards the optimal value as if a sphere would be optimized. Additionally, the rate at which the ES approaches this optimal value is the same as if a sphere were

to be optimized, independently of the value of ξ . The $x^{(g)}$ and $r^{(g)}$ dynamics are proportional to $\exp\left(-\frac{\varphi^*}{N}g\right)$ in the steady state. This follows from the definition of the progress rates φ_x^* and φ_r^* . Consequently, linear convergence order with a convergence rate of $\frac{\varphi^*}{N}$ is implied. This means that the conically constrained problem with the objective function gradient in direction of the x_1 -axis has been turned into a sphere model by applying the projection as the repair method. The behavior with other objective function gradient directions still needs to be investigated. However, the behavior is achieved by a $\sqrt{1 + \xi}$ times larger mutation strength (41). That is, the ES performs larger search steps in the search space. These larger mutation steps are automatically realized by the self-adaptation as can be seen in (40).

The figure for the steady state (Fig. 8) allows for a comparison of the $(1, 10)$ -ES, the $(2/2_I, 10)$ -ES, and the $(3/3_I, 10)$ -ES. The mutation strength σ_{ss}^* and $\frac{x_{ss}}{\sqrt{\xi}r_{ss}}$ ratio approximations come close to the ones determined by experiments. For the steady state progress rates, the approximations are not that good. The deviations are particularly high for small values of ξ . The reason for this is that in the derivations for $P_Q(q)$, ξ has been assumed to be sufficiently large. Despite that, the figure gives insights. For the considered ESs, one can observe that the strategies with intermediate recombination exhibit faster progress in the steady state. The difference between the $(2/2_I, 10)$ -ES and the $(3/3_I, 10)$ -ES is marginal, which is also predicted by the theory as derived in the following. Based on Equation (38), one can compute the maximal x progress in the steady state under the assumption $N \gg \sigma_{ss}^*$. Computing the first derivative with respect to σ_{ss}^* of Equation (38) yields

$$\frac{d}{d\sigma_{ss}^*} \varphi_{xss}^* = \frac{c_{\mu/\mu, \lambda}}{\sqrt{1 + \xi}} - \frac{\sigma_{ss}^*}{\mu(\sqrt{1 + \xi})^2} \quad (42)$$

Setting Equation (42) to zero and solving for σ_{ss}^* results in the mutation strength

$$\hat{\sigma}_{ss}^* = \sqrt{1 + \xi} \mu c_{\mu/\mu, \lambda} \quad (43)$$

with which the maximal x progress in the steady state is obtained. Further, by insertion of Equation (43) into Equation (38), the maximal x progress follows as $\hat{\varphi}_{xss}^*(\mu, \lambda) = \frac{\mu c_{\mu/\mu, \lambda}^2}{2}$. Assuming it is achieved by the ES, $\hat{\varphi}_{xss}^*(2, 10) \approx 1.61$ and $\hat{\varphi}_{xss}^*(3, 10) \approx 1.70$ follow, which are relatively close to each other. In contrast to that, $\hat{\varphi}_{xss}^*(1, 10) \approx 1.18$.

Insertion of Equation (43) into Equation (34) yields

$$\frac{\sqrt{\xi}r_{ss}}{x_{ss}} = \sqrt{\frac{1 + \frac{(1+\xi)}{N} \mu c_{\mu/\mu, \lambda}^2}{1 + \frac{(1+\xi)}{N} \mu c_{\mu/\mu, \lambda}^2 \mu}}. \quad (44)$$

This gives insights about the deviation of the parental centroid from the cone boundary. Assuming $\frac{(1+\xi)}{N} \mu c_{\mu/\mu, \lambda}^2$ in Equation (44) to be sufficiently small, a Taylor expansion around 0 for $\frac{(1+\xi)}{N} \mu c_{\mu/\mu, \lambda}^2$ can be applied. Together with subsequent cutoff after the linear term, it results in

$$\frac{\sqrt{\xi}r_{ss}}{x_{ss}} \simeq 1 + \frac{(1 + \xi)}{N} \frac{\mu c_{\mu/\mu, \lambda}^2}{2} (1 - \mu). \quad (45)$$

As can be seen, in the case without recombination ($\mu = 1$), the parental centroid moves on the cone boundary. However,

with recombination ($\mu > 1$), the centroid leaves this boundary and moves into the interior of the cone. The deviation from the cone boundary increases quadratically with μ . With these observations, it can now be formally shown that the asymptotical offspring feasibility probability $P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \simeq \Phi\left[\frac{1}{\sigma^{(g)}}\left(\frac{x^{(g)}}{\sqrt{\xi}} - \bar{r}\right)\right] = \Phi\left[\frac{N}{\sigma^{(g)*}}\left(\frac{x^{(g)}}{r^{(g)}\sqrt{\xi}} - \frac{\bar{r}}{r^{(g)}}\right)\right]$ tends to zero in the steady state movement near the cone boundary. For $N \rightarrow \infty$ with $\xi/N \ll 1$ and $\mu/N \ll 1$, $(\sqrt{\xi}r_{ss})/x_{ss} \simeq 1$. Further, $\bar{r} > r^{(g)}$ follows from Equation (B.62). Hence, the feasibility probability tends to 0 for the steady state movement of the ES near the cone boundary assuming $N \rightarrow \infty$ with $N \gg \sigma^{(g)*}$.

The theory developed allows further to compute the optimal value $\hat{\tau}$ for the learning parameter τ . Requiring that the steady state mutation strength σ_{ss}^* from (40) attains $\hat{\sigma}_{ss}^*$ from (43), one gets

$$\sqrt{1 + \xi\mu c_{\mu/\mu,\lambda}} \stackrel{!}{=} \sqrt{1 + \xi\mu c_{\mu/\mu,\lambda}} \left[(1 - N\tau^2) + \sqrt{(1 - N\tau^2)^2 + \frac{2\tau^2 N \left(\frac{1}{2} + e_{\mu,\lambda}^{1,1}\right)}{\mu c_{\mu/\mu,\lambda}^2}} \right]. \quad (46)$$

Solving Equation (46) for τ yields the optimal learning parameter

$$\hat{\tau}(\mu, \lambda) = \frac{1}{\sqrt{2N}} \sqrt{\frac{\mu c_{\mu/\mu,\lambda}^2}{\mu c_{\mu/\mu,\lambda}^2 - \frac{1}{2} - e_{\mu,\lambda}^{1,1}}}, \quad (47)$$

which equals the optimal learning parameter of the sphere model (cf. [12, Eq. (4.116), p. 71]). For $\lambda \rightarrow \infty$ ($\lambda/N \ll 1$) with $\mu/\lambda = \text{const.}$, one has $\hat{\tau} \simeq \frac{1}{\sqrt{2N}}$, which has been used in the simulations as an approximation to the optimal value⁴. For $N = 1000$, one has $\frac{1}{\sqrt{2N}} \approx 0.022$ and $\hat{\tau}(1, 10) \approx 0.017$, $\hat{\tau}(2, 10) \approx 0.021$, and $\hat{\tau}(3, 10) \approx 0.022$. Hence, already for $\lambda = 10$, $\frac{1}{\sqrt{2N}}$ approximates the optimal learning parameter relatively well.

Let us finally outline research goals to be addressed in the future. Investigations concerning the more general conically constrained problem ($\alpha \neq 0$) are interesting to pursue further. In addition, the analysis and comparison of different σ control methods is a topic for future work. It is of interest to analyze the ES with CSA and Meta-ESs instead of self-adaptation for adapting σ . A comparison between the different approaches may give insights for practical recommendations of their application. As another research direction, it is of interest to investigate and compare different repair methods. Truncation and reflection are two examples. In the truncation approach, an infeasible offspring is repaired by the intersection of its mutation vector and the cone. For the reflection, infeasible offspring are mirrored into the feasible region about the cone boundary.

REFERENCES

- [1] P. Spettel, H.-G. Beyer, and M. Hellwig, "A covariance matrix self-adaptation evolution strategy for optimization under linear constraints," *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 3, pp. 514–524, June 2019.
- [2] D. V. Arnold, "On the behaviour of the $(1, \lambda)$ -ES for a simple constrained problem," in *Proceedings of the 2011 Conference on Foundations of Genetic Algorithms*. ACM, 2011, pp. 15–24.
- [3] —, "Analysis of a repair mechanism for the $(1, \lambda)$ -ES applied to a simple constrained problem," in *Proceedings of the 2011 Conference on Genetic and Evolutionary Computation*. ACM, 2011, pp. 853–860.
- [4] M. Hellwig and D. V. Arnold, "Comparison of constraint-handling mechanisms for the $(1, \lambda)$ -ES on a simple constrained problem," *Evolutionary Computation*, vol. 24, no. 1, pp. 1–23, 2016.
- [5] D. V. Arnold and J. Porter, "Towards an augmented lagrangian constraint handling approach for the $(1+1)$ -ES," in *Proceedings of the 2015 Conference on Genetic and Evolutionary Computation*. ACM, 2015, pp. 249–256.
- [6] A. Atamna, A. Auger, and N. Hansen, "Augmented lagrangian constraint handling for CMA-ES - case of a single linear constraint," in *International Conference on Parallel Problem Solving from Nature*. Springer, 2016, pp. 181–191.
- [7] —, "Linearly convergent evolution strategies via augmented lagrangian constraint handling," in *Proceedings of the 2017 Conference on Foundations of Genetic Algorithms*. ACM, 2017, pp. 149–161.
- [8] D. V. Arnold, "On the behaviour of the $(1, \lambda)$ -ES for conically constrained linear problems," *Evolutionary Computation*, vol. 22, no. 3, pp. 503–523, 2014.
- [9] —, "On the behaviour of the $(1, \lambda)$ -ES for a conically constrained problem," in *Proceedings of the 2013 Conference on Genetic and Evolutionary Computation*. ACM, 2013, pp. 423–430.
- [10] P. Spettel and H.-G. Beyer, "Technical report: Analysis of the $(1, \lambda)$ - σ -self-adaptation evolution strategy with repair by projection applied to a conically constrained problem," Vorarlberg University of Applied Sciences, Tech. Rep. TR-SAESCONE-18, 2018, available online: https://www.fhv.at/fileadmin/user_upload/fhv/files/forschung/ppe/working-papers/Analysis_onecommalambdasigmaSA-ES_Projection_Cone_TR.pdf (accessed Oct 01, 2018).
- [11] H.-G. Beyer, *The Theory of Evolution Strategies*, ser. Natural Computing Series. Springer, 2001.
- [12] S. Meyer-Nieberg, "Self-adaptation in evolution strategies," Ph.D. dissertation, Dortmund University of Technology, 12 2007.

⁴It is also referred to [11, Sec. 6.3], where investigations for the unconstrained sphere model regarding $\lambda \rightarrow \infty$ ($\lambda/N \ll 1$) with $\mu/\lambda = \text{const.}$ are described.

Analysis of the $(\mu/\mu_I, \lambda)$ - σ -Self-Adaptation Evolution Strategy with Repair by Projection Applied to a Conically Constrained Problem

Patrick Spettel and Hans-Georg Beyer

SUPPLEMENTARY MATERIAL

APPENDIX

APPENDIX A

DERIVATION OF CLOSED-FORM EXPRESSIONS FOR THE PROJECTION

In this section, a geometrical approach for the projection of points that are outside of the feasible region onto the cone boundary is described¹. Fig. 9 shows a visualization of the 2D-plane spanned by an offspring $\tilde{\mathbf{x}}$ and the x_1 coordinate axis. The equation for the cone boundary is a direct consequence of the problem definition (Equation (2)). The projected vector is indicated by \mathbf{x}_Π . The unit vectors \mathbf{e}_1 , $\mathbf{e}_{\tilde{\mathbf{x}}}$, $\mathbf{e}_{\tilde{\mathbf{r}}}$, and $\mathbf{e}_{\tilde{\mathbf{c}}}$ are introduced. They are unit vectors in the direction of the x_1 axis, $\tilde{\mathbf{x}}$ vector, $\tilde{\mathbf{r}} = (0, \tilde{x}_2, \dots, \tilde{x}_N)^T$ vector, and the cone boundary, respectively. The projection line is indicated by the dashed line. It indicates the shortest Euclidean distance from $\tilde{\mathbf{x}}$ to the cone boundary. The goal is to compute the parameter vector after projection \mathbf{x}_Π . By inspecting Fig. 9 one can see that the projected vector \mathbf{x}_Π can be expressed as

$$\mathbf{x}_\Pi = \begin{cases} (\mathbf{e}_{\tilde{\mathbf{c}}}^T \tilde{\mathbf{x}}) \mathbf{e}_{\tilde{\mathbf{c}}} & \text{if } \mathbf{e}_{\tilde{\mathbf{c}}}^T \tilde{\mathbf{x}} > 0 \\ \mathbf{0} & \text{otherwise.} \end{cases} \quad (\text{A.48})$$

From Equation (2), the equation of the cone boundary $\frac{x}{\sqrt{\xi}}$ follows. Using this, the vector $\mathbf{e}_{\tilde{\mathbf{c}}}$ can be expressed as a linear combination of the vectors \mathbf{e}_1 and $\mathbf{e}_{\tilde{\mathbf{r}}}$. After normalization (such that $\|\mathbf{e}_{\tilde{\mathbf{c}}}\| = 1$) this writes

$$\mathbf{e}_{\tilde{\mathbf{c}}} = \frac{\mathbf{e}_1 + \frac{1}{\sqrt{\xi}} \mathbf{e}_{\tilde{\mathbf{r}}}}{\sqrt{1 + \frac{1}{\xi}}}. \quad (\text{A.49})$$

The unit vectors $\mathbf{e}_1 = (1, 0, \dots, 0)^T$ and $\mathbf{e}_{\tilde{\mathbf{x}}} = \frac{\tilde{\mathbf{x}}}{\|\tilde{\mathbf{x}}\|}$ are known. The vector $\mathbf{e}_{\tilde{\mathbf{r}}}$ is a unit vector in direction of $\tilde{\mathbf{x}}$ in the dimensions 2 to N , i.e.,

$$\mathbf{e}_{\tilde{\mathbf{r}}} = (0, (\tilde{\mathbf{x}})_2, \dots, (\tilde{\mathbf{x}})_N)^T / \|\tilde{\mathbf{r}}\|. \quad (\text{A.50})$$

Inserting Equation (A.50) into Equation (A.49) results in

$$\mathbf{e}_{\tilde{\mathbf{c}}} = \left(1, \frac{(\tilde{\mathbf{x}})_2}{\sqrt{\xi} \|\tilde{\mathbf{r}}\|}, \dots, \frac{(\tilde{\mathbf{x}})_N}{\sqrt{\xi} \|\tilde{\mathbf{r}}\|} \right)^T / \sqrt{1 + \frac{1}{\xi}}. \quad (\text{A.51})$$

Using Equation (A.51) it follows that

$$\begin{aligned} \mathbf{e}_{\tilde{\mathbf{c}}}^T \tilde{\mathbf{x}} &= \left((\tilde{\mathbf{x}})_1 + \frac{(\tilde{\mathbf{x}})_2^2 + \dots + (\tilde{\mathbf{x}})_N^2}{\sqrt{\xi} \|\tilde{\mathbf{r}}\|} \right) / \sqrt{1 + \frac{1}{\xi}} \\ &= \left((\tilde{\mathbf{x}})_1 + \frac{\|\tilde{\mathbf{r}}\|^2}{\sqrt{\xi} \|\tilde{\mathbf{r}}\|} \right) / \sqrt{1 + \frac{1}{\xi}} \\ &= \left((\tilde{\mathbf{x}})_1 + \frac{\|\tilde{\mathbf{r}}\|}{\sqrt{\xi}} \right) / \sqrt{1 + \frac{1}{\xi}} = \left((\tilde{\mathbf{x}})_1 \sqrt{\xi} + \|\tilde{\mathbf{r}}\| \right) / \sqrt{\xi + 1}. \end{aligned} \quad (\text{A.52})$$

¹Note that the same closed-form projection equations can also be derived using the KKT conditions for minimizing the Lagrangian

$$L(q, \mathbf{r}_\Pi, \lambda) = \frac{1}{2} \|\mathbf{r}_\Pi - \tilde{\mathbf{r}}\|^2 + \frac{1}{2} (q - x_1)^2 - \lambda (\xi \|\mathbf{r}_\Pi\|^2 - q^2).$$

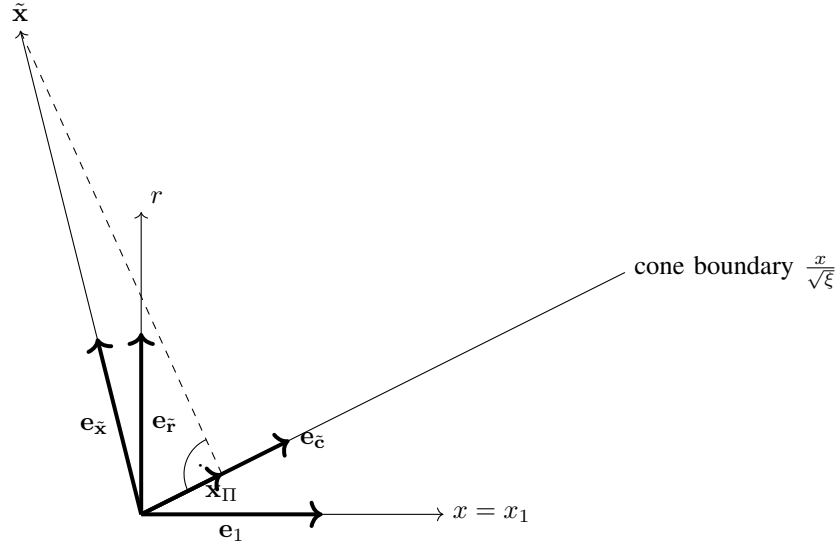


Fig. 9. 2D-plane spanned by an offspring $\tilde{\mathbf{x}}$ and the x_1 coordinate axis. Vectors introduced for expressing the projection onto the feasible region of Fig. 1 are visualized. The vectors \mathbf{e}_1 , $\mathbf{e}_{\tilde{\mathbf{x}}}$, $\mathbf{e}_{\tilde{\mathbf{r}}}$, and $\mathbf{e}_{\tilde{\mathbf{c}}}$ are unit vectors of the corresponding vectors. The dashed line indicates the orthogonal projection of $\tilde{\mathbf{x}}$ onto the cone boundary.

Use of Equations (A.48), (A.51), and (A.52) yields for the case $\mathbf{e}_{\tilde{\mathbf{c}}}^T \tilde{\mathbf{x}} > 0$

$$\begin{aligned} \mathbf{x}_{\Pi} &= (\mathbf{e}_{\tilde{\mathbf{c}}}^T \tilde{\mathbf{x}}) \mathbf{e}_{\tilde{\mathbf{c}}} \\ &= \frac{1}{1 + \frac{1}{\xi}} \left((\tilde{\mathbf{x}})_1 + \frac{(\tilde{\mathbf{x}})_2^2 + \dots + (\tilde{\mathbf{x}})_N^2}{\sqrt{\xi} \|\tilde{\mathbf{r}}\|} \right) \left(1, \frac{(\tilde{\mathbf{x}})_2}{\sqrt{\xi} \|\tilde{\mathbf{r}}\|}, \dots, \frac{(\tilde{\mathbf{x}})_N}{\sqrt{\xi} \|\tilde{\mathbf{r}}\|} \right)^T. \end{aligned} \quad (\text{A.53})$$

The first vector component of \mathbf{x}_{Π} writes

$$q = (\mathbf{x}_{\Pi})_1 = \frac{1}{1 + \frac{1}{\xi}} \left((\tilde{\mathbf{x}})_1 + \frac{(\tilde{\mathbf{x}})_2^2 + \dots + (\tilde{\mathbf{x}})_N^2}{\sqrt{\xi} \|\tilde{\mathbf{r}}\|} \right) = \frac{\xi}{\xi + 1} \left((\tilde{\mathbf{x}})_1 + \frac{\|\tilde{\mathbf{r}}\|}{\sqrt{\xi}} \right). \quad (\text{A.54})$$

The k -th vector component of \mathbf{x}_{Π} for $k \in \{2, \dots, N\}$ writes

$$(\mathbf{x}_{\Pi})_k = \frac{\xi}{\xi + 1} \left((\tilde{\mathbf{x}})_1 + \frac{\|\tilde{\mathbf{r}}\|}{\sqrt{\xi}} \right) \frac{(\tilde{\mathbf{x}})_k}{\sqrt{\xi} \|\tilde{\mathbf{r}}\|} = \frac{\xi}{\xi + 1} \left(\frac{(\tilde{\mathbf{x}})_1}{\sqrt{\xi} \|\tilde{\mathbf{r}}\|} + \frac{1}{\xi} \right) (\tilde{\mathbf{x}})_k. \quad (\text{A.55})$$

And the projected distance from the cone axis $\|\mathbf{r}_{\Pi}\|$ writes

$$\begin{aligned} q_r = \|\mathbf{r}_{\Pi}\| &= \sqrt{\sum_{k=2}^N (\mathbf{x}_{\Pi})_k^2} = \sqrt{\sum_{k=2}^N \left(\frac{\xi}{\xi + 1} \left(\frac{(\tilde{\mathbf{x}})_1}{\sqrt{\xi} \|\tilde{\mathbf{r}}\|} + \frac{1}{\xi} \right) \right)^2 (\tilde{\mathbf{x}})_k^2} \\ &= \left(\frac{\xi}{\xi + 1} \left(\frac{(\tilde{\mathbf{x}})_1}{\sqrt{\xi} \|\tilde{\mathbf{r}}\|} + \frac{1}{\xi} \right) \right) \|\tilde{\mathbf{r}}\|. \end{aligned} \quad (\text{A.56})$$

Note that in terms of the $(x, r)^T$ representation, the offspring $\tilde{\mathbf{x}}$ can be expressed as $(\mathbf{e}_1^T \tilde{\mathbf{x}}, \mathbf{e}_{\tilde{\mathbf{r}}}^T \tilde{\mathbf{x}})^T = ((\tilde{\mathbf{x}})_1, \|\tilde{\mathbf{r}}\|)^T$. This is exactly what is expected. The first component is the value in direction of the cone axis. The second component is the distance from the cone axis.

APPENDIX B

DERIVATION OF THE NORMAL APPROXIMATION FOR THE OFFSPRING DENSITY IN r DIRECTION

The density of an offspring in r direction before projection can be derived from Algorithm 1 (Lines 7 and 8). Remember that an individual can be uniquely described by its distance x from 0 in direction of the cone axis and the distance r from the cone axis. In addition, due to the symmetry of the problem, the coordinate system can w.l.o.g. be rotated such that an individual in the $(x, r)^T$ -space, $\left(\tilde{x}_1, \tilde{r} = \sqrt{\sum_{k=2}^N \tilde{x}_k^2}\right)^T$ is $(\tilde{x}_1, \tilde{r}, 0, \dots, 0)^T$ in the rotated coordinate system. The offspring's distance from the cone's axis, $\tilde{r} = \sqrt{\sum_{k=2}^N \tilde{x}_k^2}$, is distributed according to a non-central χ distribution with $N - 1$ degrees of freedom. Under the assumption $\tilde{\sigma}_l \approx \sigma^{(g)}$, $\tilde{x}_2 = r^{(g)} + \sigma^{(g)} z_2 \sim \mathcal{N}(r^{(g)}, \sigma^{(g)2})$ and $\tilde{x}_k = \sigma^{(g)} z_k \sim \mathcal{N}(0, \sigma^{(g)2})$ for $k \in \{3, \dots, N\}$, it is the (positive) square root of a sum of squared normally distributed random variables.

It is the goal of this section to derive a normal approximation for \tilde{r} . From the considerations explained in the previous paragraph, it follows that the distance from the cone's axis of the offspring is

$$\tilde{r} = \sqrt{r^{(g)2} + 2\sigma^{(g)}r^{(g)}z_2 + \sigma^{(g)2}z_2^2 + \sigma^{(g)2}\sum_{k=3}^N z_k^2}. \quad (\text{B.57})$$

Now, $2\sigma^{(g)}r^{(g)}z_2 + \sigma^{(g)2}z_2^2$ and $\sigma^{(g)2}\sum_{k=3}^N z_k^2$ are replaced with normally distributed expressions with mean values and standard deviations of the corresponding expressions. With the moments of the i.i.d. variables according to a standard normal distribution z_k , $E[z_k] = 0$, and $E[z_k^2] = 1$, the expected values $E[2\sigma^{(g)}r^{(g)}z_2 + \sigma^{(g)2}z_2^2] = \sigma^{(g)2}$ and $E[\sigma^{(g)2}\sum_{k=3}^N z_k^2] = \sigma^{(g)2}(N - 2)$ follow. The variances are computed as

$$\begin{aligned} & \text{Var}[2\sigma^{(g)}r^{(g)}z_2 + \sigma^{(g)2}z_2^2] \\ &= E[(2\sigma^{(g)}r^{(g)}z_2 + \sigma^{(g)2}z_2^2)^2] - E[2\sigma^{(g)}r^{(g)}z_2 + \sigma^{(g)2}z_2^2]^2 \\ &= E[4\sigma^{(g)2}r^{(g)2}z_2^2 + 4\sigma^{(g)3}r^{(g)}z_2^3 + \sigma^{(g)4}z_2^4] - \sigma^{(g)4} \\ &= 4\sigma^{(g)2}r^{(g)2} + 3\sigma^{(g)4} - \sigma^{(g)4} = 4\sigma^{(g)2}r^{(g)2} + 2\sigma^{(g)4} \end{aligned} \quad (\text{B.58})$$

and

$$\begin{aligned} \text{Var}\left[\sigma^{(g)2}\sum_{i=3}^N z_i^2\right] &= \sigma^{(g)4}\sum_{i=3}^N \text{Var}[z_i^2] = \sigma^{(g)4}\sum_{i=3}^N (E[z_i^4] - E[z_i^2]^2) \\ &= \sigma^{(g)4}\sum_{i=3}^N 2 = 2\sigma^{(g)4}(N - 2) \end{aligned} \quad (\text{B.59})$$

where the moments of the i.i.d. variables according to a standard normal distribution z_i $E[z_i^2] = 1$, $E[z_i^3] = 0$, and $E[z_i^4] = 3$ were used.

With these results, the normal approximation follows as

$$\begin{aligned} \tilde{r} &\approx \sqrt{r^{(g)2} + \mathcal{N}(\sigma^{(g)2}, 4\sigma^{(g)2}r^{(g)2} + 2\sigma^{(g)4}) + \mathcal{N}(\sigma^{(g)2}(N - 2), 2\sigma^{(g)4}(N - 2))} \\ &= \sqrt{r^{(g)2} + \sigma^{(g)2}(N - 1) + \sqrt{4\sigma^{(g)2}r^{(g)2} + 2\sigma^{(g)4}(N - 1)}\mathcal{N}(0, 1)}. \end{aligned} \quad (\text{B.60})$$

Substitution of the normalized quantity $\sigma^{(g)*} = \frac{\sigma^{(g)}N}{r^{(g)}}$ yields after simplification

$$\tilde{r} \approx r^{(g)}\sqrt{1 + \frac{\sigma^{(g)*2}}{N}\left(1 - \frac{1}{N}\right)}\sqrt{1 + \frac{2\sigma^{(g)*}}{N}\frac{\sqrt{1 + \frac{\sigma^{(g)*2}}{2N}\left(1 - \frac{1}{N}\right)}}{1 + \frac{\sigma^{(g)*2}}{N}\left(1 - \frac{1}{N}\right)}}\mathcal{N}(0, 1). \quad (\text{B.61})$$

As the expression $\frac{2\sigma^{(g)*}}{N}\frac{\sqrt{1 + \frac{\sigma^{(g)*2}}{2N}\left(1 - \frac{1}{N}\right)}}{1 + \frac{\sigma^{(g)*2}}{N}\left(1 - \frac{1}{N}\right)}\mathcal{N}(0, 1) \rightarrow 0$ with $N \rightarrow \infty$ ($\sigma^{(g)*} < \infty$), a further asymptotically simplified expression can be obtained by Taylor expansion of the square root at 0 and cutoff after the linear term

$$\tilde{r} \approx \underbrace{r^{(g)}\sqrt{1 + \frac{\sigma^{(g)*2}}{N}\left(1 - \frac{1}{N}\right)}}_{\tilde{r}} + \underbrace{r^{(g)}\frac{\sigma^{(g)*}}{N}\sqrt{\frac{1 + \frac{\sigma^{(g)*2}}{2N}\left(1 - \frac{1}{N}\right)}{1 + \frac{\sigma^{(g)*2}}{N}\left(1 - \frac{1}{N}\right)}}}_{\sigma_r}\mathcal{N}(0, 1). \quad (\text{B.62})$$

Consequently, the mean of the asymptotic normal approximation of \tilde{r} is \bar{r} and its standard deviation is σ_r

$$p_r(r) \approx \frac{1}{\sigma_r} \phi\left(\frac{r - \bar{r}}{\sigma_r}\right) = \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left[-\frac{1}{2}\left(\frac{r - \bar{r}}{\sigma_r}\right)^2\right]. \quad (\text{B.63})$$

It is referred to [13, Sec. 3.1.2.1.2.2, pp. 22-29] for approximation quality investigations. In particular, Fig. 3.3 in that work shows plots comparing the density of the normal approximation to the exact density.

APPENDIX C DERIVATION OF THE x PROGRESS RATE

From the definition of the progress rate (Equation (9)) and the pseudo-code of the ES (Algorithm 1, lines 4 and 19),

$$\varphi_x(x^{(g)}, r^{(g)}, \sigma^{(g)}) = x^{(g)} - \mathbb{E}[x^{(g+1)} | x^{(g)}, r^{(g)}, \sigma^{(g)}] \quad (\text{C.64})$$

$$= x^{(g)} - \mathbb{E}[\langle q \rangle | x^{(g)}, r^{(g)}, \sigma^{(g)}] \quad (\text{C.65})$$

follows. Therefore, $\mathbb{E}[\langle q \rangle] := \mathbb{E}[\langle q \rangle | x^{(g)}, r^{(g)}, \sigma^{(g)}]$ needs to be derived to proceed further. The derivation of an approximation for it is presented in detail in Appendix D. It makes use of the r normal approximation and assumes sufficiently small τ (such that $\sigma_l \simeq \sigma^{(g)}$), $N \rightarrow \infty$, and $\sigma^{(g)*} \ll N$. In addition, the needed and derived approximation of P_Q gets better for larger values of ξ .

A. The Approximate x Progress Rate in the case that the probability of feasible offspring tends to 1

Using the result from Equation (D.134), the normalized x progress rate for the feasible case $\varphi_{x_{\text{feas}}}^*$ can be formulated as

$$\varphi_{x_{\text{feas}}}^* = \frac{N(x^{(g)} - \mathbb{E}[\langle q \rangle_{\text{feas}}])}{x^{(g)}} \approx \frac{N(x^{(g)} - x^{(g)} + \sigma^{(g)} c_{\mu/\mu, \lambda})}{x^{(g)}} = \frac{r^{(g)}}{x^{(g)}} \sigma^{(g)*} c_{\mu/\mu, \lambda}. \quad (\text{C.66})$$

where $c_{\mu/\mu, \lambda} = e_{\mu, \lambda}^{1,0}$ is the so-called generalized progress coefficient [14, Eq. (5.112), p. 172]. It is defined as,

$$e_{\mu, \lambda}^{\alpha, \beta} := \frac{\lambda - \mu}{(\sqrt{2\pi})^{\alpha+1}} \binom{\lambda}{\mu} \int_{t=-\infty}^{t=\infty} t^{\beta} e^{-\frac{\alpha+1}{2} t^2} [\Phi(t)]^{\lambda-\mu-1} [1 - \Phi(t)]^{\mu-\alpha} dt. \quad (\text{C.67})$$

B. The Approximate x Progress Rate in the case that the probability of feasible offspring tends to 0

Similarly, for the infeasible case, the progress rate $\varphi_{x_{\text{infeas}}}^*$ can be calculated from Equation (D.168):

$$\varphi_{x_{\text{infeas}}}^* = \frac{N(x^{(g)} - \mathbb{E}[\langle q \rangle_{\text{infeas}}])}{x^{(g)}} \quad (\text{C.68})$$

$$= \frac{N}{x^{(g)}} \left(\frac{1 + \xi}{1 + \xi} x^{(g)} - \mathbb{E}[\langle q \rangle_{\text{infeas}}] \right) \quad (\text{C.69})$$

$$\approx \frac{N}{x^{(g)}} \left[\frac{x^{(g)} + \xi x^{(g)} - \xi x^{(g)}}{1 + \xi} - \frac{\sqrt{\xi} \bar{r}}{1 + \xi} + \frac{\xi}{1 + \xi} \left(\sqrt{\sigma^{(g)2} + \sigma_r^2 / \xi} \right) c_{\mu/\mu, \lambda} \right]. \quad (\text{C.70})$$

This can further be simplified using σ_r of the normal approximation (Equation (B.62)) together with σ -normalization yielding

$$\varphi_{x_{\text{infeas}}}^* = \frac{N}{1 + \xi} \left[1 - \sqrt{\xi} \frac{r^{(g)}}{x^{(g)}} \sqrt{1 + \frac{\sigma^{(g)*2}}{N} \left(1 - \frac{1}{N} \right)} + \xi \frac{r^{(g)}}{x^{(g)}} \sqrt{\frac{\sigma^{(g)*2}}{N^2} + \frac{1}{\xi} \frac{\sigma^{(g)*2}}{N^2} \frac{1 + \frac{\sigma^{(g)*2}}{2N} (1 - \frac{1}{N})}{1 + \frac{\sigma^{(g)*2}}{N} (1 - \frac{1}{N})}} c_{\mu/\mu, \lambda} \right] \quad (\text{C.71})$$

$$\simeq \frac{N}{1 + \xi} \left[1 - \sqrt{\xi} \frac{r^{(g)}}{x^{(g)}} \sqrt{1 + \frac{\sigma^{(g)*2}}{N}} + \xi \frac{r^{(g)}}{x^{(g)}} \frac{\sigma^{(g)*}}{N} \sqrt{1 + \frac{1}{\xi} \frac{1 + \frac{\sigma^{(g)*2}}{2N}}{1 + \frac{\sigma^{(g)*2}}{N}}} c_{\mu/\mu, \lambda} \right]. \quad (\text{C.72})$$

From Equation (C.71) to Equation (C.72), $\frac{1}{N}$ has been neglected compared to 1 as $N \rightarrow \infty$. It can be rewritten further yielding

$$\varphi_{x_{\text{infeas}}}^* = \frac{N}{1 + \xi} \left(1 - \frac{\sqrt{\xi} r^{(g)}}{x^{(g)}} \sqrt{1 + \frac{\sigma^{(g)*2}}{N}} \right) + \frac{\sqrt{\xi}}{1 + \xi} \frac{\sqrt{\xi} r^{(g)}}{x^{(g)}} \sigma^{(g)*} c_{\mu/\mu, \lambda} \sqrt{1 + \frac{1}{\xi} \frac{1 + \frac{\sigma^{(g)*2}}{2N}}{1 + \frac{\sigma^{(g)*2}}{N}}}. \quad (\text{C.73})$$

C. The Approximate x Progress Rate - Combination Using the Single Offspring Feasibility Probability

Similarly to [13, Sec. 3.1.2.1.2.8, pp. 44-49], the feasible and infeasible cases can be combined using the single offspring feasibility probability

$$\varphi_x^* \approx P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \varphi_{x_{\text{feas}}}^* + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] \varphi_{x_{\text{infeas}}}^* \quad (\text{C.74})$$

$$\begin{aligned} &= P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \frac{r^{(g)}}{x^{(g)}} \sigma^{(g)*} c_{\mu/\mu, \lambda} + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] \left[\frac{N}{1 + \xi} \left(1 - \frac{\sqrt{\xi} r^{(g)}}{x^{(g)}} \sqrt{1 + \frac{\sigma^{(g)*2}}{N}} \right) \right. \\ &\quad \left. + \frac{\sqrt{\xi}}{1 + \xi} \frac{\sqrt{\xi} r^{(g)}}{x^{(g)}} \sigma^{(g)*} c_{\mu/\mu, \lambda} \sqrt{1 + \frac{1}{\xi} \frac{1 + \frac{\sigma^{(g)*2}}{2N}}{1 + \frac{\sigma^{(g)*2}}{N}}} \right]. \end{aligned} \quad (\text{C.75})$$

APPENDIX D
DERIVATION OF $E[\langle q \rangle]$

With Lines 13, 17, and 19 of Algorithm 1,

$$E[\langle q \rangle] := E[\langle q \rangle \mid x^{(g)}, r^{(g)}, \sigma^{(g)}] \quad (D.76)$$

$$= E \left[\frac{1}{\mu} \sum_{m=1}^{\mu} q_{m;\lambda} \mid x^{(g)}, r^{(g)}, \sigma^{(g)} \right] \quad (D.77)$$

$$= \frac{1}{\mu} \sum_{m=1}^{\mu} E \left[q_{m;\lambda} \mid x^{(g)}, r^{(g)}, \sigma^{(g)} \right] \quad (D.78)$$

follows, where $q_{m;\lambda} := (\tilde{x}_{m;\lambda})_1$ denotes the x -value after (possible) projection of the m -th best offspring. It is the next step to derive the expected value of this quantity:

$$E[q_{m;\lambda}] := E[q_{m;\lambda} \mid x^{(g)}, r^{(g)}, \sigma^{(g)}] \quad (D.79)$$

$$= \int_{q=0}^{q=\infty} q p_{q_{m;\lambda}}(q) dq \quad (D.80)$$

$$= \int_{q=0}^{q=\infty} q \frac{\lambda!}{(\lambda-m)!(m-1)!} p_Q(q) [1 - P_Q(q)]^{\lambda-m} [P_Q(q)]^{m-1} dq. \quad (D.81)$$

Equation (D.80) follows directly from the definition of expectation where

$$p_{q_{m;\lambda}}(q) := p_{q_{m;\lambda}}(q \mid x^{(g)}, r^{(g)}, \sigma^{(g)})$$

indicates the probability density function of the m -th best (offspring with m -th smallest q value) offspring's q value. The random variable Q denotes the random x values after projection. The step to Equation (D.81) follows from the calculation of $p_{q_{m;\lambda}}(q)$. Because the objective function (Equation (1)) is defined to return an individual's x value, $p_{q_{m;\lambda}}(q)$ is the probability density function of the m -th smallest q value among λ values. This calculation is well-known in order statistics (see, e.g., [15]). A short derivation is presented here for the case under consideration. A single offspring's q value has a probability density of $p_Q(q)$. In order for this particular q value to be the m -th smallest, there must be $(\lambda - m)$ offspring greater than that, and $(m - 1)$ offspring smaller. This results in the density $p_Q(q) [1 - P_Q(q)]^{\lambda-m} [P_Q(q)]^{m-1}$ where $P_Q(q) = \Pr[Q \leq q]$. Because the number of such combinations is $\lambda \binom{\lambda-1}{m-1} = \frac{\lambda!}{(\lambda-m)!(m-1)!}$ one obtains

$$p_{q_{m;\lambda}}(q) = \frac{\lambda!}{(\lambda-m)!(m-1)!} p_Q(q) [1 - P_Q(q)]^{\lambda-m} [P_Q(q)]^{m-1}. \quad (D.82)$$

Insertion of Equation (D.81) into Equation (D.78) results in

$$E[\langle q \rangle] = \frac{1}{\mu} \sum_{m=1}^{\mu} \int_{q=0}^{q=\infty} q \frac{\lambda!}{(\lambda-m)!(m-1)!} p_Q(q) [1 - P_Q(q)]^{\lambda-m} [P_Q(q)]^{m-1} \quad (D.83)$$

$$= \frac{\lambda!}{\mu} \int_{q=0}^{q=\infty} q p_Q(q) \sum_{m=1}^{\mu} \frac{[1 - P_Q(q)]^{\lambda-m} [P_Q(q)]^{m-1}}{(\lambda-m)!(m-1)!}. \quad (D.84)$$

The identity

$$\sum_{m=1}^{\mu} \frac{P^{m-1} [1 - P]^{\lambda-m}}{(m-1)!(\lambda-m)!} = \frac{1}{(\lambda-\mu-1)!(\mu-1)!} \int_0^{1-P} z^{\lambda-\mu-1} (1-z)^{\mu-1} dz \quad (D.85)$$

stated in [14, Eq. (5.14), p. 147] allows expressing the sum in Equation (D.84) as an integral. Application of Equation (D.85) to Equation (D.84) and expressing the fraction using the binomial coefficient yields

$$E[\langle q \rangle] = (\lambda - \mu) \binom{\lambda}{\mu} \int_{q=0}^{q=\infty} q p_Q(q) \int_{z=0}^{z=1-P_Q(q)} z^{\lambda-\mu-1} (1-z)^{\mu-1} dz dq. \quad (D.86)$$

To proceed further, $z = 1 - P_Q(y)$ is substituted in the inner integral. This implies $y = P_Q^{-1}(1 - z)$ and $dz = -p_Q(y)$. The upper and lower bounds for the substituted integral follow as $y_u = P_Q^{-1}(1 - 1 + P_Q(q)) = q$ and $y_l = P_Q^{-1}(1 - 0) = \infty$, respectively. Therefore, one obtains

$$E[\langle q \rangle] = (\lambda - \mu) \binom{\lambda}{\mu} \int_{q=0}^{q=\infty} q p_Q(q) \int_{y=\infty}^{y=q} [1 - P_Q(y)]^{\lambda-\mu-1} [P_Q(y)]^{\mu-1} (-p_Q(y)) dy dq \quad (D.87)$$

$$= (\lambda - \mu) \binom{\lambda}{\mu} \int_{q=0}^{q=\infty} q p_Q(q) \int_{y=q}^{y=\infty} p_Q(y) [1 - P_Q(y)]^{\lambda-\mu-1} [P_Q(y)]^{\mu-1} dy dq. \quad (D.88)$$

Changing the order of integration, one finally gets

$$\mathbb{E}[\langle q \rangle] = (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=0}^{y=\infty} p_Q(y) [1 - P_Q(y)]^{\lambda-\mu-1} [P_Q(y)]^{\mu-1} \int_{q=0}^{q=y} q p_Q(q) dq dy. \quad (\text{D.89})$$

$P_Q(q)$ and $p_Q(q)$ need to be derived next. To this end, $\Pr[Q \leq q]$ is calculated by integration. This results in $P_Q(q) = \Pr[Q \leq q]$. Using this, $p_Q(q)$ follows by computing its derivative. The integral can be written as

$$\begin{aligned} P_Q(q) &= \int_{x=-\infty}^{x=q} \int_{r=0}^{r=\infty} p_{1;1}(x, r) dr dx \\ &\quad - \int_{x=-\infty}^{x=q} \int_{r=-\sqrt{\xi}x + (\sqrt{\xi}+1/\sqrt{\xi})q}^{r=\infty} p_{1;1}(x, r) dr dx. \end{aligned} \quad (\text{D.90})$$

Taking a look at Fig. 1, the integral is calculated as follows. First, the whole area above the x -axis up to $x = q$, is calculated. Then, the area enclosed by the projection line and the vertical line at $x = q$ is subtracted from it. The function $p_{1;1}(x, r) := p_{1;1}(x, r | x^{(g)}, r^{(g)}, \sigma^{(g)})$ denotes the joint probability density function of a single descendant before projection. Note that this density includes the log-normal mutation of $\sigma^{(g)}$. Due to difficulties analytically handling the integral further in this exact form, assuming τ sufficiently small simplifies the analysis considerably. More formally, with $\tilde{\sigma}_l \approx \sigma^{(g)}$ the mutation of $\sigma^{(g)}$ can be neglected in the analysis. In particular, it allows writing $p_{1;1}(x, r) \approx p_x(x)p_r(r)$, where $p_x(x)$ is the x -density before projection and $p_r(r)$ is the normal approximation of the r -density before projection (Equation (B.63)). Note that they are independent because the mutations in the different dimensions are independent of each other. $p_x(x)$ follows from the offspring generation in Algorithm 1 (Line 8). Every component of the offspring's parameter vector $\tilde{\mathbf{x}}$ is independently and identically distributed according to a normal random variable. With $\tilde{\sigma}_l \approx \sigma^{(g)}$,

$$p_x(x) \approx \frac{1}{\sigma^{(g)}} \phi\left(\frac{x - x^{(g)}}{\sigma^{(g)}}\right) = \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{x - x^{(g)}}{\sigma^{(g)}}\right)^2\right] \quad (\text{D.91})$$

follows. Now,

$$\begin{aligned} P_Q(q) &\approx \int_{x=-\infty}^{x=q} \int_{r=0}^{r=\infty} p_x(x)p_r(r) dr dx \\ &\quad - \int_{x=-\infty}^{x=q} \int_{r=-\sqrt{\xi}x + (\sqrt{\xi}+1/\sqrt{\xi})q}^{r=\infty} p_x(x)p_r(r) dr dx \end{aligned} \quad (\text{D.92})$$

$$\begin{aligned} &= \int_{x=-\infty}^{x=q} p_x(x) \underbrace{\int_{r=0}^{r=\infty} p_r(r) dr}_{=1} dx \\ &\quad - \int_{x=-\infty}^{x=q} p_x(x) \left[\int_{r=-\sqrt{\xi}x + (\sqrt{\xi}+1/\sqrt{\xi})q}^{r=\infty} p_r(r) dr \right] dx \end{aligned} \quad (\text{D.93})$$

$$\begin{aligned} &\approx \int_{x=-\infty}^{x=q} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{x - x^{(g)}}{\sigma^{(g)}}\right)^2\right] dx \\ &\quad - \int_{x=-\infty}^{x=q} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{x - x^{(g)}}{\sigma^{(g)}}\right)^2\right] \\ &\quad \times \left\{ \int_{r=-\sqrt{\xi}x + (\sqrt{\xi}+1/\sqrt{\xi})q}^{r=\infty} \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left[-\frac{1}{2}\left(\frac{r - \bar{r}}{\sigma_r}\right)^2\right] dr \right\} dx. \end{aligned} \quad (\text{D.94})$$

follows. Expressing the integrals where possible using the cumulative distribution function of the standard normal distribution

$\Phi(\cdot)$, Equation (D.94) can be turned into

$$P_Q(q) \approx \Phi\left(\frac{q - x^{(g)}}{\sigma^{(g)}}\right) - \int_{x=-\infty}^{x=q} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{x - x^{(g)}}{\sigma^{(g)}}\right)^2\right] \times \left[1 - \Phi\left(\frac{-\sqrt{\xi}x + (\sqrt{\xi} + 1/\sqrt{\xi})q - \bar{r}}{\sigma_r}\right)\right] dx \quad (D.95)$$

$$= \Phi\left(\frac{q - x^{(g)}}{\sigma^{(g)}}\right) - \Phi\left(\frac{q - x^{(g)}}{\sigma^{(g)}}\right) + \int_{x=-\infty}^{x=q} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{x - x^{(g)}}{\sigma^{(g)}}\right)^2\right] \times \Phi\left(\frac{-\sqrt{\xi}x + (\sqrt{\xi} + 1/\sqrt{\xi})q - \bar{r}}{\sigma_r}\right) dx. \quad (D.96)$$

Equation (D.96) can further be simplified and then upper bounded yielding

$$P_Q(q) \approx \int_{x=-\infty}^{x=q} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{x - x^{(g)}}{\sigma^{(g)}}\right)^2\right] \times \Phi\left(\frac{-\sqrt{\xi}x + (\sqrt{\xi} + 1/\sqrt{\xi})q - \bar{r}}{\sigma_r}\right) dx \quad (D.97)$$

$$\leq \int_{x=-\infty}^{x=\infty} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{x - x^{(g)}}{\sigma^{(g)}}\right)^2\right] \times \Phi\left(\frac{-\sqrt{\xi}x + (\sqrt{\xi} + 1/\sqrt{\xi})q - \bar{r}}{\sigma_r}\right) dx \quad (D.98)$$

$$= \Phi\left(\frac{(1 + 1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}}\right). \quad (D.99)$$

The step from Equation (D.97) to Equation (D.98) is justified because the expression inside the integral is non-negative. Therefore, an increase in the upper bound of the integral results in an increase of the result of evaluating the integral. While the integral in Equation (D.97) has an upper bound of q and a closed form solution is not apparent, the integral in Equation (D.98) has an upper bound of ∞ and can be solved analytically. To this end, the identity

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} \Phi(at + b) = \sqrt{2\pi} \Phi\left(\frac{b}{\sqrt{1 + a^2}}\right), \quad (D.100)$$

which has been derived and proven in [14, Equation A.10], can be used with the substitution $t := \frac{x - x^{(g)}}{\sigma^{(g)}}$. After simplification, one arrives at Equation (D.99).

It turns out that this upper bound is a satisfactory approximation for $P_Q(q)$ (see [13, Sec. 3.1.2.1.2, pp. 21-39] for more details). By taking a close look at Equation (D.97), one can see that the expression can be further simplified under certain conditions. Normalization of σ_r and multiplication of the argument to $\Phi(\cdot)$ by $\frac{1/\sqrt{\xi}}{1/\sqrt{\xi}}$ in Equation (D.97) results in

$$P_Q(q) \approx \int_{x=-\infty}^{x=q} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{x - x^{(g)}}{\sigma^{(g)}}\right)^2\right] \times \Phi\left(\frac{-x + (1 + 1/\xi)q - \bar{r}/\sqrt{\xi}}{\sigma_r^* r^{(g)} / (N\sqrt{\xi})}\right) dx \quad (D.101)$$

$$= \int_{x=-\infty}^{x=q} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{x - x^{(g)}}{\sigma^{(g)}}\right)^2\right] \times \Phi\left(\frac{N\sqrt{\xi}(-x/r^{(g)} + (1 + 1/\xi)q/r^{(g)} - \bar{r}/(r^{(g)}\sqrt{\xi}))}{\sigma_r^*}\right) dx. \quad (D.102)$$

Assuming $N\sqrt{\xi} \rightarrow \infty$, one observes that

$$\Phi \left(N\sqrt{\xi} \frac{-\frac{x}{r^{(g)}} + (1 + \frac{1}{\xi})\frac{q}{r^{(g)}} - \frac{\bar{r}}{r^{(g)}\sqrt{\xi}}}{\sigma_r^*} \right) \approx 1 \iff \left(1 + \frac{1}{\xi} \right) \frac{q}{r^{(g)}} - \frac{x}{r^{(g)}} - \frac{\bar{r}}{r^{(g)}\sqrt{\xi}} > 0 \quad (\text{D.103})$$

$$\iff x < \left(1 + \frac{1}{\xi} \right) q - \frac{\bar{r}}{\sqrt{\xi}} \quad (\text{D.104})$$

$$\iff q < \left(1 + \frac{1}{\xi} \right) q - \frac{\bar{r}}{\sqrt{\xi}} \quad (\text{D.105})$$

$$\iff 0 < \frac{q}{\xi} - \frac{\bar{r}}{\sqrt{\xi}} \quad (\text{D.106})$$

$$\iff \frac{\bar{r}}{\sqrt{\xi}} < \frac{q}{\xi} \quad (\text{D.107})$$

$$\iff \frac{\bar{r}\sqrt{\xi}}{q} < 1. \quad (\text{D.108})$$

From Equation (D.104) to Equation (D.105), $x \leq q$ that follows from the bounds of the integration, has been used. The resulting condition (Equation (D.108)) means that the simplification can be used if the probability of generating feasible offspring is high. This can be seen in the following way. The probability for a particular value q to be feasible is given by

$$\Pr \left[\sqrt{\xi}\tilde{r} \leq |q| \right] = \int_{\tilde{r}=0}^{\tilde{r}=q/\sqrt{\xi}} p_r(\tilde{r}) d\tilde{r} \approx \Phi \left(\frac{q/\sqrt{\xi} - \bar{r}}{\sigma_r} \right) = \Phi \left(N \frac{q/\sqrt{\xi} - \bar{r}}{\sigma_r^* r^{(g)}} \right). \quad (\text{D.109})$$

It is the integration from the cone axis up to the cone boundary at the given value q . For $N \rightarrow \infty$, it follows that

$$\Phi \left(N \frac{q/\sqrt{\xi} - \bar{r}}{\sigma_r^* r^{(g)}} \right) \approx 1 \iff \frac{q/\sqrt{\xi} - \bar{r}}{\sigma_r^* r^{(g)}} > 0 \quad (\text{D.110})$$

$$\iff q/\sqrt{\xi} - \bar{r} > 0 \quad (\text{D.111})$$

$$\iff q/\sqrt{\xi} > \bar{r} \quad (\text{D.112})$$

$$\iff q > \bar{r}\sqrt{\xi} \quad (\text{D.113})$$

$$\iff \frac{\bar{r}\sqrt{\xi}}{q} < 1. \quad (\text{D.114})$$

Under this condition, Equation (D.101) can be simplified in the asymptotic case $N \rightarrow \infty$ to

$$P_Q(q) \approx \Phi \left(\frac{q - x^{(g)}}{\sigma^{(g)}} \right) \text{ for } q > \bar{r}\sqrt{\xi}. \quad (\text{D.115})$$

These derivations lead to the $P_Q(q)$ and $p_Q(q)$ approximations used in the further derivations:

$$P_Q(q) \approx \begin{cases} P_{Q_{\text{feas}}}(q) := \Phi \left(\frac{q - x^{(g)}}{\sigma^{(g)}} \right), & \text{for } q > \bar{r}\sqrt{\xi} \end{cases} \quad (\text{D.116})$$

$$P_{Q_{\text{infeas}}}(q) := \Phi \left(\frac{(1 + 1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}} \right), \text{ otherwise.} \quad (\text{D.117})$$

Taking the derivative with respect to q yields

$$p_{Q_{\text{feas}}}(q) = \frac{1}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2} \left(\frac{q - x^{(g)}}{\sigma^{(g)}} \right)^2}, \text{ for } q > \bar{r}\sqrt{\xi} \quad (\text{D.118})$$

$$p_Q(q) \approx \begin{cases} p_{Q_{\text{infeas}}}(q) = \left(\frac{(1 + 1/\xi)}{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}} \right) \\ \times \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{(1 + 1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}} \right)^2 \right], \\ \text{otherwise.} \end{cases} \quad (\text{D.119})$$

It is referred to [13, Sec. 3.1.2.1.2, pp. 21-39] for more details in the presented analysis and for investigations regarding the quality of the approximations.

A. Derivation of $E[\langle q \rangle]_{feas}$

For treating the feasible case, insertion of Equations (D.116) and (D.118) into Equation (D.89) yields

$$E[\langle q \rangle_{feas}] \approx (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=0}^{y=\infty} p_{Q_{feas}}(y) [1 - P_{Q_{feas}}(y)]^{\lambda-\mu-1} [P_{Q_{feas}}(y)]^{\mu-1} \times \int_{q=0}^{q=y} q p_{Q_{feas}}(q) dq dy \quad (D.120)$$

$$= (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=\bar{r}\sqrt{\xi}}^{y=\infty} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2}\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)^2} \left[1 - \Phi\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)\right]^{\lambda-\mu-1} \times \left[\Phi\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)\right]^{\mu-1} \int_{q=\bar{r}\sqrt{\xi}}^{q=y} q \frac{1}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)^2} dq dy. \quad (D.121)$$

For solving the inner integral, $t := \frac{q-x^{(g)}}{\sigma^{(g)}}$ is substituted. It implies $q = \sigma^{(g)}t + x^{(g)}$ and $dq = \sigma^{(g)} dt$. The lower bound follows with σ -normalization, assuming $N \rightarrow \infty$, $N \gg \sigma^{*(g)}$, and knowing that $x^{(g)} \geq \bar{r}\sqrt{\xi}$ holds for the feasible case. It reads

$$t_l = \frac{\bar{r}\sqrt{\xi} - x^{(g)}}{\sigma^{(g)}} = \frac{N}{\sigma^{*(g)}} \frac{\bar{r}\sqrt{\xi} - x^{(g)}}{r^{(g)}} \simeq -\infty.$$

The application of the substitution results in

$$\int_{q=\bar{r}\sqrt{\xi}}^{q=y} q \frac{1}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)^2} dq \simeq \int_{t=-\infty}^{t=\frac{y-x^{(g)}}{\sigma^{(g)}}} (\sigma^{(g)}t + x^{(g)}) \frac{1}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2}t^2} \sigma^{(g)} dt \quad (D.122)$$

$$= \sigma^{(g)} \frac{1}{\sqrt{2\pi}} \int_{t=-\infty}^{t=\frac{y-x^{(g)}}{\sigma^{(g)}}} t e^{-\frac{1}{2}t^2} dt + x^{(g)} \int_{t=-\infty}^{t=\frac{y-x^{(g)}}{\sigma^{(g)}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt. \quad (D.123)$$

The integral in the second summand in Equation (D.123) can be expressed using the cumulative distribution function of the normal distribution $\Phi(\cdot)$. The integral in the first summand in Equation (D.123) can be solved using the identity

$$\int_{-\infty}^x t e^{-\frac{1}{2}t^2} dt = -e^{-\frac{1}{2}x^2} \quad (D.124)$$

derived in [14, Eq. (A.16), p. 331]. Hence, one obtains

$$\int_{q=\bar{r}\sqrt{\xi}}^{q=y} q \frac{1}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)^2} dq \simeq x^{(g)} \Phi\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right) - \sigma^{(g)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)^2}. \quad (D.125)$$

Equation (D.125) can now be inserted into Equation (D.121) yielding

$$E[\langle q \rangle_{feas}] \approx (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=\bar{r}\sqrt{\xi}}^{y=\infty} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2}\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)^2} \left[1 - \Phi\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)\right]^{\lambda-\mu-1} \times \left[\Phi\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)\right]^{\mu-1} x^{(g)} \Phi\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right) dy - (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=\bar{r}\sqrt{\xi}}^{y=\infty} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2}\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)^2} \left[1 - \Phi\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)\right]^{\lambda-\mu-1} \times \left[\Phi\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)\right]^{\mu-1} \sigma^{(g)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)^2} dy. \quad (D.126)$$

The first term in Equation (D.126) can be simplified by using $P_{Q_{feas}}(y) = \Phi\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)$. It yields

$$(\lambda - \mu) \binom{\lambda}{\mu} \int_{y=\bar{r}\sqrt{\xi}}^{y=\infty} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2}\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)^2} \left[1 - \Phi\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)\right]^{\lambda-\mu-1} \times \left[\Phi\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)\right]^{\mu-1} x^{(g)} \Phi\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right) dy = x^{(g)} (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=\bar{r}\sqrt{\xi}}^{y=\infty} p_{Q_{feas}}(y) [1 - P_{Q_{feas}}(y)]^{\lambda-\mu-1} [P_{Q_{feas}}(y)]^{\mu} dy. \quad (D.127)$$

This can be rewritten resulting in

$$\begin{aligned} & x^{(g)}(\lambda - \mu) \binom{\lambda}{\mu} \int_{y=\bar{r}\sqrt{\xi}}^{y=\infty} p_{Q_{\text{feas}}}(y) [1 - P_{Q_{\text{feas}}}(y)]^{\lambda-\mu-1} [P_{Q_{\text{feas}}}(y)]^{\mu} dy \\ &= x^{(g)}(\lambda - \mu) \frac{\lambda!}{(\lambda - \mu)! \mu!} \int_{y=\bar{r}\sqrt{\xi}}^{y=\infty} p_{Q_{\text{feas}}}(y) [1 - P_{Q_{\text{feas}}}(y)]^{\lambda-\mu-1} [P_{Q_{\text{feas}}}(y)]^{\mu} dy \end{aligned} \quad (\text{D.128})$$

$$\begin{aligned} &= x^{(g)} \frac{\lambda!}{(\lambda - (\mu + 1))! ((\mu + 1) - 1)!} \int_{y=\bar{r}\sqrt{\xi}}^{y=\infty} p_{Q_{\text{feas}}}(y) [1 - P_{Q_{\text{feas}}}(y)]^{\lambda-(\mu+1)} \\ &\quad \times [P_{Q_{\text{feas}}}(y)]^{(\mu+1)-1} dy. \end{aligned} \quad (\text{D.129})$$

Now, by comparing the expression inside the integral with Equation (D.82), one observes that this is exactly the probability density function of the $(\mu + 1)$ -th best offspring. Because in this section the case under consideration is the feasible case, in the limit the whole mass is in $p_{q_{\text{feas}}}$ and no mass in $p_{q_{\text{infeas}}}$. Hence, integration over all values of q equals 1. This reads

$$\begin{aligned} & x^{(g)} \frac{\lambda!}{(\lambda - (\mu + 1))! ((\mu + 1) - 1)!} \int_{y=0}^{y=\infty} p_{Q_{\text{feas}}}(y) [1 - P_{Q_{\text{feas}}}(y)]^{\lambda-(\mu+1)} \\ &\quad \times [P_{Q_{\text{feas}}}(y)]^{(\mu+1)-1} dy = x^{(g)} \underbrace{\int_{y=0}^{y=\infty} p_{q_{\text{feas}}(\mu+1);\lambda}(y) dy}_{=1} = x^{(g)}. \end{aligned} \quad (\text{D.130})$$

The second term in Equation (D.126) can be simplified by solving the integral using substitution. The substitution $-t := \frac{y-x^{(g)}}{\sigma^{(g)}}$ is performed. It implies $y = -\sigma^{(g)}t + x^{(g)}$ and $dy = -\sigma^{(g)} dt$. The lower bound follows with σ -normalization, assuming $N \rightarrow \infty$, $N \gg \sigma^{*(g)}$, and knowing that $x^{(g)} \geq \bar{r}\sqrt{\xi}$ holds for the feasible case. It reads $t_l = -\frac{\bar{r}\sqrt{\xi}-x^{(g)}}{\sigma^{(g)}} = -\frac{N}{\sigma^{*(g)}} \frac{\bar{r}\sqrt{\xi}-x^{(g)}}{r^{(g)}} \simeq \infty$. Similarly, the upper bound follows with σ -normalization and assuming $N \rightarrow \infty$ and $N \gg \sigma^{*(g)}$ as $t_u = \lim_{y \rightarrow \infty} \frac{-y+x^{(g)}}{\sigma^{(g)}} = \lim_{y \rightarrow \infty} \frac{N}{\sigma^{*(g)}} \frac{-y+x^{(g)}}{r^{(g)}} = \frac{N}{\sigma^{*(g)} r^{(g)}} \lim_{y \rightarrow \infty} (-y + x^{(g)}) \simeq -\infty$. The integral after substitution reads

$$\begin{aligned} & -(\lambda - \mu) \binom{\lambda}{\mu} \int_{y=\bar{r}\sqrt{\xi}}^{y=\infty} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2}\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)^2} \left[1 - \Phi\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)\right]^{\lambda-\mu-1} \\ &\quad \times \left[\Phi\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)\right]^{\mu-1} \sigma^{(g)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-x^{(g)}}{\sigma^{(g)}}\right)^2} dy \\ &= -\sigma^{(g)} \frac{\lambda - \mu}{(\sqrt{2\pi})^{1+1}} \binom{\lambda}{\mu} (-1) \int_{t=-\infty}^{t=\infty} e^{-\frac{1+1}{2}t^2} [1 - \Phi(-t)]^{\lambda-\mu-1} [\Phi(-t)]^{\mu-1} dt \end{aligned} \quad (\text{D.131})$$

$$= -\sigma^{(g)} \frac{\lambda - \mu}{(\sqrt{2\pi})^{1+1}} \binom{\lambda}{\mu} \int_{t=-\infty}^{t=\infty} e^{-\frac{1+1}{2}t^2} [\Phi(t)]^{\lambda-\mu-1} [1 - \Phi(t)]^{\mu-1} dt = -\sigma^{(g)} c_{\mu/\mu, \lambda}. \quad (\text{D.132})$$

From Equation (D.131) to Equation (D.132) the fact that taking the negative of an integral can be expressed by exchanging the lower and upper bounds and the identity $1 - \Phi(-t) = \Phi(t)$ have been used. In Equation (D.132), the progress coefficient $c_{\mu/\mu, \lambda}$ defined in [14, Eq. (6.102), p. 247] appears. It is defined as

$$c_{\mu/\mu, \lambda} := \frac{\lambda - \mu}{2\pi} \binom{\lambda}{\mu} \int_{t=-\infty}^{t=\infty} e^{-t^2} [\Phi(t)]^{\lambda-\mu-1} [1 - \Phi(t)]^{\mu-1} dt. \quad (\text{D.133})$$

Insertion of Equation (D.130) and Equation (D.132) into Equation (D.126) results in

$$\mathbb{E}[\langle q \rangle_{\text{feas}}] \approx x^{(g)} - \sigma^{(g)} c_{\mu/\mu, \lambda}. \quad (\text{D.134})$$

B. Derivation of $E[\langle q \rangle]_{\text{infeas}}$

For treating the infeasible case, insertion of Equations (D.117) and (D.119) into Equation (D.89) yields

$$E[\langle q \rangle_{\text{infeas}}] \approx (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=0}^{y=\bar{r}\sqrt{\xi}} p_{Q_{\text{infeas}}}(y) [1 - P_{Q_{\text{infeas}}}(y)]^{\lambda-\mu-1} [P_{Q_{\text{infeas}}}(y)]^{\mu-1} \times \int_{q=0}^{q=y} q p_{Q_{\text{infeas}}}(q) dq dy \quad (\text{D.135})$$

$$= (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=0}^{y=\bar{r}\sqrt{\xi}} p_{Q_{\text{infeas}}}(y) [1 - P_{Q_{\text{infeas}}}(y)]^{\lambda-\mu-1} [P_{Q_{\text{infeas}}}(y)]^{\mu-1} \times \int_{q=0}^{q=y} q \left(\frac{(1 + 1/\xi)}{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}} \right) \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{(1 + 1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}} \right)^2 \right] dq dy. \quad (\text{D.136})$$

For solving the inner integral in Equation (D.136), the substitution

$$\frac{(1 + 1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}} := t \quad (\text{D.137})$$

is used. It follows that

$$q = \frac{1}{(1 + 1/\xi)} \left(\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi} t + x^{(g)} + \bar{r}/\sqrt{\xi} \right) \quad (\text{D.138})$$

and

$$dq = \frac{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}}{(1 + 1/\xi)} dt. \quad (\text{D.139})$$

Using the normalized $\sigma^{(g)*}$, $\sigma_r \simeq \sigma^{(g)}$ for $N \rightarrow \infty$, and $\sigma^{(g)*} \ll N$ (derived from Equation (B.62)), t can be expressed as

$$t = \frac{(1 + 1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\frac{\sigma^{(g)*2} r^{(g)2}}{N^2} + \frac{\sigma^{(g)*2} r^{(g)2}}{N^2 \xi}}} \quad (\text{D.140})$$

$$= \frac{(1 + 1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\frac{1}{N\sqrt{\xi}} \sqrt{\xi \sigma^{(g)*2} r^{(g)2} + \sigma^{(g)*2} r^{(g)2}}} \quad (\text{D.141})$$

$$= N\sqrt{\xi} \left[\frac{(1 + 1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\xi \sigma^{(g)*2} r^{(g)2} + \sigma^{(g)*2} r^{(g)2}}} \right] \quad (\text{D.142})$$

$$= N\sqrt{\xi} \left[\frac{(1 + 1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sigma^{(g)*} r^{(g)} \sqrt{\xi + 1}} \right]. \quad (\text{D.143})$$

The lower bound in the transformed integral therefore follows assuming $\xi \gg 1$, $N \rightarrow \infty$, using $\infty > \bar{r} \simeq r^{(g)} \geq 0$, and knowing that $0 \leq x^{(g)} < \infty$ as

$$t_l = N\sqrt{\xi} \left[\frac{(1 + 1/\xi)0 - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sigma^{(g)*} r^{(g)} \sqrt{\xi + 1}} \right] \quad (\text{D.144})$$

$$= N\sqrt{\xi} \left[\frac{-x^{(g)} - \bar{r}/\sqrt{\xi}}{\sigma^{(g)*} r^{(g)} \sqrt{\xi + 1}} \right] \quad (\text{D.145})$$

$$\simeq N \underbrace{\left[\frac{-x^{(g)} - \bar{r}/\sqrt{\xi}}{\sigma^{(g)*} r^{(g)}} \right]}_{\leq 0} \quad (\text{D.146})$$

$$\simeq -\infty. \quad (\text{D.147})$$

The transformed integral writes

$$\begin{aligned} & \int_{q=0}^{q=y} q \left(\frac{(1+1/\xi)}{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}} \right) \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{(1+1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}} \right)^2 \right] dq \\ &= \int_{t=-\infty}^{t=\frac{(1+1/\xi)y - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}}} \frac{1}{(1+1/\xi)} \left(\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi} t + x^{(g)} + \bar{r}/\sqrt{\xi} \right) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt. \end{aligned} \quad (\text{D.148})$$

Use of Equation (D.124) for the first summand and the cumulative distribution function of the standard normal distribution $\Phi(\cdot)$ for the second and third summands, this can further be rewritten resulting in

$$\begin{aligned} & \int_{t=-\infty}^{t=\frac{(1+1/\xi)y - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}}} \frac{1}{(1+1/\xi)} \left(\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi} t + x^{(g)} + \bar{r}/\sqrt{\xi} \right) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt \\ &= -\frac{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}}{(1+1/\xi)} \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{(1+1/\xi)y - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}} \right)^2 \right] \\ & \quad + \underbrace{\frac{1}{(1+1/\xi)} x^{(g)} \Phi \left(\frac{(1+1/\xi)y - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}} \right)}_{=P_{Q_{\text{infeas}}}} \\ & \quad + \underbrace{\frac{1}{(1+1/\xi)} \bar{r}/\sqrt{\xi} \Phi \left(\frac{(1+1/\xi)y - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}} \right)}_{=P_{Q_{\text{infeas}}}} dt. \end{aligned} \quad (\text{D.149})$$

Insertion of Equation (D.149) back into Equation (D.136) leads to

$$\begin{aligned} \mathbb{E}[\langle q \rangle_{\text{infeas}}] &\approx (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=0}^{y=\bar{r}\sqrt{\xi}} p_{Q_{\text{infeas}}}(y) [1 - P_{Q_{\text{infeas}}}(y)]^{\lambda-\mu-1} [P_{Q_{\text{infeas}}}(y)]^{\mu-1} \\ &\quad \times (-1) \frac{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}}{(1+1/\xi)} \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{(1+1/\xi)y - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}} \right)^2 \right] dy \\ &\quad + \underbrace{\frac{1}{(1+1/\xi)} x^{(g)} \int_{y=0}^{\infty} p_{q_{\text{infeas}}(\mu+1);\lambda} dy}_{=1} \\ &\quad + \underbrace{\frac{1}{(1+1/\xi)} \bar{r}/\sqrt{\xi} \int_{y=0}^{\infty} p_{q_{\text{infeas}}(\mu+1);\lambda} dy}_{=1}. \end{aligned} \quad (\text{D.150})$$

The second and third summands are the result of the same argument as the one leading to Equation (D.130). For solving the integral in the first summand,

$$\frac{(1+1/\xi)y - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}} := -t \quad (\text{D.151})$$

is substituted. It follows that

$$y = -\frac{1}{(1+1/\xi)} \left(\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi} t + x^{(g)} + \bar{r}/\sqrt{\xi} \right) \quad (\text{D.152})$$

and

$$dy = -\frac{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}}{(1+1/\xi)} dt. \quad (\text{D.153})$$

Using the normalized $\sigma^{(g)*}$, $\sigma_r \simeq \sigma^{(g)}$ for $N \rightarrow \infty$, and $\sigma^{(g)*} \ll N$ (derived from Equation (B.62)), t can be expressed as

$$t = -\frac{(1 + 1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\frac{\sigma^{(g)*2} r^{(g)2}}{N^2} + \frac{\sigma^{(g)*2} r^{(g)2}}{N^2 \xi}}} \quad (\text{D.154})$$

$$= -\frac{(1 + 1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\frac{1}{N\sqrt{\xi}} \sqrt{\xi \sigma^{(g)*2} r^{(g)2} + \sigma^{(g)*2} r^{(g)2}}} \quad (\text{D.155})$$

$$= -N\sqrt{\xi} \left[\frac{(1 + 1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\xi \sigma^{(g)*2} r^{(g)2} + \sigma^{(g)*2} r^{(g)2}}} \right] \quad (\text{D.156})$$

$$= -N\sqrt{\xi} \left[\frac{(1 + 1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sigma^{(g)*} r^{(g)} \sqrt{\xi + 1}} \right]. \quad (\text{D.157})$$

The lower bound in the transformed integral therefore follows assuming $\xi \gg 1$, $N \rightarrow \infty$, using $\infty > \bar{r} \simeq r^{(g)} \geq 0$, and knowing that $0 \leq x^{(g)} < \infty$ as

$$t_l = -N\sqrt{\xi} \left[\frac{(1 + 1/\xi)0 - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sigma^{(g)*} r^{(g)} \sqrt{\xi + 1}} \right] \quad (\text{D.158})$$

$$= -N\sqrt{\xi} \left[\frac{-x^{(g)} - \bar{r}/\sqrt{\xi}}{\sigma^{(g)*} r^{(g)} \sqrt{\xi + 1}} \right] \quad (\text{D.159})$$

$$\simeq -N \underbrace{\left[\frac{-x^{(g)} - \bar{r}/\sqrt{\xi}}{\sigma^{(g)*} r^{(g)}} \right]}_{\leq 0} \quad (\text{D.160})$$

$$\simeq \infty. \quad (\text{D.161})$$

Similarly, the upper bound follows with the same assumptions and using the fact that the case under consideration is the infeasible case, i.e., $x^{(g)} \leq \bar{r}\sqrt{\xi}$

$$t_u = -N\sqrt{\xi} \left[\frac{(1 + 1/\xi)\bar{r}\sqrt{\xi} - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sigma^{(g)*} r^{(g)} \sqrt{\xi + 1}} \right] \quad (\text{D.162})$$

$$= -N \underbrace{\left[\frac{\bar{r}\sqrt{\xi} - x^{(g)}}{\sigma^{(g)*} r^{(g)}} \right]}_{\geq 0} \quad (\text{D.163})$$

$$\simeq -\infty. \quad (\text{D.164})$$

The transformed integral writes

$$\begin{aligned} & (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=0}^{y=\bar{r}\sqrt{\xi}} p_{Q_{\text{infeas}}}(y) [1 - P_{Q_{\text{infeas}}}(y)]^{\lambda-\mu-1} [P_{Q_{\text{infeas}}}(y)]^{\mu-1} \\ & \times (-1) \frac{\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi}}{(1 + 1/\xi)} \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{(1 + 1/\xi)y - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi}} \right)^2 \right] dy \\ & = - \int_{t=-\infty}^{t=0} (\lambda - \mu) \binom{\lambda}{\mu} (-1) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} [1 - \Phi(-t)]^{\lambda-\mu-1} [\Phi(-t)]^{\mu-1} \frac{\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi}}{(1 + 1/\xi)} dt \end{aligned} \quad (\text{D.165})$$

$$= - \underbrace{\frac{\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi}}{(1 + 1/\xi)} \frac{\lambda - \mu}{(\sqrt{2\pi})^{1+1}} \binom{\lambda}{\mu} \int_{t=-\infty}^{t=0} e^{-\frac{1}{2}t^2} [\Phi(t)]^{\lambda-\mu-1} [1 - \Phi(t)]^{\mu-1} dt}_{=c_{\mu/\mu,\lambda}}. \quad (\text{D.166})$$

Insertion of Equation (D.166) into Equation (D.150) finally yields

$$E[\langle q \rangle_{\text{infeas}}] \approx \frac{x^{(g)}}{(1 + 1/\xi)} + \frac{\bar{r}}{\sqrt{\xi}(1 + 1/\xi)} - \frac{\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi}}{(1 + 1/\xi)} c_{\mu/\mu,\lambda} \quad (\text{D.167})$$

$$= \frac{\xi}{1 + \xi} \left(x^{(g)} + \bar{r}/\sqrt{\xi} \right) - \frac{\xi}{1 + \xi} \sqrt{\sigma^{(g)2} + \sigma_r^2/\xi} c_{\mu/\mu,\lambda}. \quad (\text{D.168})$$

APPENDIX E

DERIVATION OF THE r PROGRESS RATE

From the definition of the progress rate (Equation (10)) and the pseudo-code of the ES (Algorithm 1, Lines 5 and 20) it follows that

$$\varphi_r(x^{(g)}, r^{(g)}, \sigma^{(g)}) = r^{(g)} - \mathbb{E}[r^{(g+1)} \mid x^{(g)}, r^{(g)}, \sigma^{(g)}] \quad (\text{E.169})$$

$$= r^{(g)} - \mathbb{E}[\langle q_r \rangle \mid x^{(g)}, r^{(g)}, \sigma^{(g)}]. \quad (\text{E.170})$$

The computation $\mathbb{E}[\langle q_r \rangle \mid x^{(g)}, r^{(g)}, \sigma^{(g)}]$ requires the evaluation of the square root of a sum of squares. This is because $\langle q_r \rangle$ is the distance from the cone axis *after* centroid computation. Therefore, the computation of $\mathbb{E}[\langle q_r \rangle \mid x^{(g)}, r^{(g)}, \sigma^{(g)}]$ is not analogous to the computation of the x progress rate. For the computation of the r progress rate, two cases are distinguished. First, the case of “being far” from the cone boundary is considered (feasible offspring with overwhelming probability). Second, the case of “being on” the cone boundary is considered (infeasible offspring with overwhelming probability). Both cases are then combined with the single offspring feasibility probability. The feasible case is similar to the sphere model because the projection can be ignored. For the infeasible case, a geometric approach is pursued.

A. The Approximate r Progress Rate in the case that the probability of feasible offspring tends to 1

For the case that the probability for infeasible offspring is negligible, generated offspring do not need to be repaired. The mutation vectors are therefore not altered. Hence, from Lines 7 to 8 of Algorithm 1,

$$\tilde{\mathbf{x}}_{m;\lambda} = \mathbf{x}^{(g)} + \tilde{\sigma}_{m;\lambda} \mathbf{z}_{m;\lambda} \quad (\text{E.171})$$

follows directly. Now, τ is assumed to be small such that $\tilde{\sigma}_l \simeq \sigma^{(g)}$ for $l \in \{1, \dots, \lambda\}$. Hence,

$$\tilde{\mathbf{x}}_{m;\lambda} \simeq \mathbf{x}^{(g)} + \sigma^{(g)} \mathbf{z}_{m;\lambda} \quad (\text{E.172})$$

and subsequently

$$\langle \tilde{\mathbf{x}} \rangle \simeq \mathbf{x}^{(g)} + \sigma^{(g)} \frac{1}{\mu} \sum_{m=1}^{\mu} \mathbf{z}_{m;\lambda} \quad (\text{E.173})$$

can be written. With this,

$$\begin{aligned} \langle q_r \rangle_{\text{feas}} = \langle \tilde{r} \rangle &\simeq \sqrt{\sum_{k=2}^N \left((\mathbf{x}^{(g)})_k + \sigma^{(g)} \frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^2} \\ &= \sqrt{\underbrace{\sum_{k=2}^N (\mathbf{x}^{(g)})_k^2 + \sum_{k=2}^N \left(2\sigma^{(g)} (\mathbf{x}^{(g)})_k \frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right) + \sum_{k=2}^N \left(\sigma^{(g)} \frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^2}_{=\langle \tilde{r} \rangle^2}} \end{aligned} \quad (\text{E.174})$$

follows for the distance from the cone axis $\langle \tilde{r} \rangle$ of $\langle \tilde{\mathbf{x}} \rangle$. To proceed with Equation (E.174), it is assumed that the deviation of the expression under the square root from its mean is small. This allows the use of a Taylor expansion

$$\sqrt{\langle \tilde{r} \rangle^2} = \sqrt{\mathbb{E}[\langle \tilde{r} \rangle^2]} + \frac{1}{2} \frac{\langle \tilde{r} \rangle^2 - \mathbb{E}[\langle \tilde{r} \rangle^2]}{\sqrt{\mathbb{E}[\langle \tilde{r} \rangle^2]}} - \frac{1}{8} \frac{(\langle \tilde{r} \rangle^2 - \mathbb{E}[\langle \tilde{r} \rangle^2])^2}{(\mathbb{E}[\langle \tilde{r} \rangle^2])^{3/2}} + \dots \quad (\text{E.175})$$

Taking the expected value, the second summand vanishes and Equation (E.175) writes

$$\mathbb{E}[\sqrt{\langle \tilde{r} \rangle^2}] = \sqrt{\mathbb{E}[\langle \tilde{r} \rangle^2]} - \frac{1}{8} \frac{\text{Var}[\langle \tilde{r} \rangle^2]}{(\mathbb{E}[\langle \tilde{r} \rangle^2])^{3/2}} + \dots \quad (\text{E.176})$$

Assuming that the quadratic term $\text{Var}[\langle \tilde{r} \rangle^2]$ and the higher order terms are negligible, the approximation writes

$$\mathbb{E}[\sqrt{\langle \tilde{r} \rangle^2}] \approx \sqrt{\mathbb{E}[\langle \tilde{r} \rangle^2]}. \quad (\text{E.177})$$

Reinsertion of Equation (E.174) into the approximation (Equation (E.177)) results in

$$\begin{aligned} \mathbb{E}[\langle \tilde{r} \rangle] &\approx \sqrt{\underbrace{\sum_{k=2}^N (\mathbf{x}^{(g)})_k^2}_{=r^{(g)2}} + \sum_{k=2}^N \left(2\sigma^{(g)} (\mathbf{x}^{(g)})_k \frac{1}{\mu} \sum_{m=1}^{\mu} \mathbb{E}[(\mathbf{z}_{m;\lambda})_k] \right) + \sum_{k=2}^N \frac{\sigma^{(g)2}}{\mu^2} \mathbb{E} \left[\left(\sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^2 \right]} \end{aligned} \quad (\text{E.178})$$

As selection is done in direction of the cone axis, the $(\mathbf{z}_{m;\lambda})_k$ for $k \in \{2, \dots, N\}$ are selectively neutral (because they are perpendicular to the cone axis). Hence, $(\mathbf{z}_{m;\lambda})_k \sim \mathcal{N}(0, 1)$ and consequently

$$\mathbb{E}[(\mathbf{z}_{m;\lambda})_k] = 0 \quad (\text{E.179})$$

holds. The expression $\sum_{k=2}^N \frac{\sigma^{(g)^2}}{\mu^2} \mathbb{E} \left[\left(\sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^2 \right]$ can be rewritten by expanding the squared sum term. It writes

$$\sum_{k=2}^N \frac{\sigma^{(g)^2}}{\mu^2} \mathbb{E} \left[\left(\sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^2 \right] = \frac{\sigma^{(g)^2}}{\mu^2} \sum_{k=2}^N \sum_{m=1}^{\mu} \sum_{p=1}^{\mu} \mathbb{E} [(\mathbf{z}_{m;\lambda})_k (\mathbf{z}_{p;\lambda})_k] \quad (\text{E.180})$$

$$= \frac{\sigma^{(g)^2}}{\mu^2} \sum_{k=2}^N \sum_{m=1}^{\mu} \mathbb{E} [(\mathbf{z}_{m;\lambda})_k^2] + \frac{\sigma^{(g)^2}}{\mu^2} \sum_{k=2}^N \sum_{m=1}^{\mu} \sum_{\substack{p=1 \\ p \neq m}}^{\mu} \mathbb{E} [(\mathbf{z}_{m;\lambda})_k (\mathbf{z}_{p;\lambda})_k] \quad (\text{E.181})$$

It is important to note that the second summand in Equation (E.181) vanishes. The random variables $(\mathbf{z}_{m;\lambda})_k$ for $k \in \{2..N\}$ are isotropically distributed in the $2..N$ subspace due to the selective neutrality of the i.i.d. normally distributed mutations in this subspace. Hence, the cross terms written explicitly as the second summand of Equation (E.181), are immediately zero. Consequently, for fixed m and p with $m \neq p$,

$$\begin{aligned} \sum_{k=2}^N \mathbb{E} [(\mathbf{z}_{m;\lambda})_k (\mathbf{z}_{p;\lambda})_k] &= \mathbb{E} [(\mathbf{z}_{m;\lambda})_{2..N}^T (\mathbf{z}_{p;\lambda})_{2..N}] \\ &= \mathbb{E} [(\mathbf{z}_{m;\lambda})_{2..N}^T] \mathbb{E} [(\mathbf{z}_{p;\lambda})_{2..N}] = 0 \end{aligned} \quad (\text{E.182})$$

holds. With Equation (E.179) and Equation (E.182), Equation (E.178) simplifies to

$$\mathbb{E}[\langle \tilde{r} \rangle] \approx \sqrt{r^{(g)^2} + \frac{\sigma^{(g)^2}}{\mu^2} \sum_{m=1}^{\mu} \mathbb{E} \left[\sum_{k=2}^N (\mathbf{z}_{m;\lambda})_k^2 \right]} \quad (\text{E.183})$$

Again, because selection is done in the x_1 direction, the expression $\mathbb{E} \left[\sum_{k=2}^N (\mathbf{z}_{m;\lambda})_k^2 \right]$ is the expected value of the sum of $N - 1$ standard normally distributed random variables. It is therefore the expected value of a χ -distributed random variable with $N - 1$ degrees of freedom. Hence,

$$\mathbb{E} \left[\sum_{k=2}^N (\mathbf{z}_{m;\lambda})_k^2 \right] = N - 1 \quad (\text{E.184})$$

holds which implies

$$\mathbb{E}[\langle q_r \rangle_{\text{feas}}] = \mathbb{E}[\langle \tilde{r} \rangle] \approx \sqrt{r^{(g)^2} + \frac{\sigma^{(g)^2}}{\mu} (N - 1)} \quad (\text{E.185})$$

$$= r^{(g)} \sqrt{1 + \frac{\sigma^{(g)*2}}{N^2 \mu} (N - 1)} \quad (\text{E.186})$$

$$\simeq r^{(g)} \sqrt{1 + \frac{\sigma^{(g)*2}}{\mu N}} \quad (\text{E.187})$$

With this result, the normalized r progress rate for the feasible case can be formulated as

$$\varphi_{r \text{ feas}}^* = \frac{N(r^{(g)} - \mathbb{E}[\langle q_r \rangle_{\text{feas}}])}{r^{(g)}} \quad (\text{E.188})$$

$$\approx N \left(1 - \sqrt{1 + \frac{\sigma^{(g)*2}}{\mu N}} \right) \quad (\text{E.189})$$

In order to justify the approximation in (E.177), it is shown that the fluctuations of $\langle \tilde{r} \rangle^2$ around its mean value are small. This is done by showing that $\frac{\sqrt{\text{Var}[\langle \tilde{r} \rangle^2]}}{\mathbb{E}[\langle \tilde{r} \rangle^2]}$ tends to zero for $N \rightarrow \infty$ and assuming that the quadratic term dominates the sum of all the terms of higher than quadratic order in Equation (E.176). The starting point of the calculation of $\text{Var}[\langle \tilde{r} \rangle^2]$ is the

term under the square root of Equation (E.178). Use of stochastic independence of the different components for $k \in \{2..N\}$ (because selection is done in x_1 -direction) results in

$$\text{Var} [\langle \hat{r} \rangle^2] \approx \sum_{k=2}^N \text{Var} \left[\left(2\sigma^{(g)}(\mathbf{x}^{(g)})_k \frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right) + \left(\sigma^{(g)} \frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^2 \right] \quad (\text{E.190})$$

$$\begin{aligned} &= \sum_{k=2}^N \text{Var} \left[\left(2\sigma^{(g)}(\mathbf{x}^{(g)})_k \frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right) \right] + \text{Var} \left[\left(\sigma^{(g)} \frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^2 \right] \\ &\quad + 2 \text{Cov} \left[\underbrace{\left(2\sigma^{(g)}(\mathbf{x}^{(g)})_k \frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)}_{=:A}, \underbrace{\left(\sigma^{(g)} \frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^2}_{=:B} \right]. \end{aligned} \quad (\text{E.191})$$

By the properties of the variance calculation and the independence due to the selection in x_1 -direction, Equation (E.191) can be rewritten to

$$\text{Var} [\langle \hat{r} \rangle^2] \approx \sum_{k=2}^N 4\sigma^{(g)2}(\mathbf{x}^{(g)})_k^2 \frac{1}{\mu^2} \sum_{m=1}^{\mu} \text{Var} [(\mathbf{z}_{m;\lambda})_k] + \sigma^{(g)4} \frac{1}{\mu^4} \text{Var} \left[\left(\sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^2 \right] + 2 \text{Cov} [A, B]. \quad (\text{E.192})$$

Again, because selection only depends on the x_1 -direction, the random variables $(\mathbf{z}_{m;\lambda})_k$ for $k \in \{2..N\}$ are isotropically distributed in the $2..N$ subspace. Because the projection does not have an influence for the “feasible” case under consideration, the $(\mathbf{z}_{m;\lambda})_k$ are i.i.d. standard normal random variables. Consequently, $\text{Var} [(\mathbf{z}_{m;\lambda})_k] = 1$ holds and

$$4\sigma^{(g)2}(\mathbf{x}^{(g)})_k^2 \frac{1}{\mu^2} \sum_{m=1}^{\mu} \text{Var} [(\mathbf{z}_{m;\lambda})_k] = \frac{4\sigma^{(g)2}}{\mu} (\mathbf{x}^{(g)})_k^2. \quad (\text{E.193})$$

$\sigma^{(g)4} \frac{1}{\mu^4} \text{Var} \left[\left(\sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^2 \right]$ can be derived by writing the cross terms explicitly

$$\sigma^{(g)4} \frac{1}{\mu^4} \text{Var} \left[\left(\sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^2 \right] = \frac{\sigma^{(g)4}}{\mu^4} \text{Var} \left[\sum_{m=1}^{\mu} \sum_{p=1}^{\mu} (\mathbf{z}_{m;\lambda})_k (\mathbf{z}_{p;\lambda})_k \right] \quad (\text{E.194})$$

$$\begin{aligned} &= \frac{\sigma^{(g)4}}{\mu^4} \sum_{m=1}^{\mu} \text{Var} [(\mathbf{z}_{m;\lambda})_k^2] \\ &\quad + \frac{\sigma^{(g)4}}{\mu^4} \sum_{m=1}^{\mu} \sum_{\substack{p=1 \\ p \neq m}}^{\mu} \text{Var} [(\mathbf{z}_{m;\lambda})_k (\mathbf{z}_{p;\lambda})_k]. \end{aligned} \quad (\text{E.195})$$

$\text{Var} [(\mathbf{z}_{m;\lambda})_k^2] = 2$ because $(\mathbf{z}_{m;\lambda})_k^2$ is χ^2 distributed with one degree of freedom. $\text{Var} [(\mathbf{z}_{m;\lambda})_k (\mathbf{z}_{p;\lambda})_k] = \text{E} [(\mathbf{z}_{m;\lambda})_k^2] \text{E} [(\mathbf{z}_{p;\lambda})_k^2] - \text{E} [(\mathbf{z}_{m;\lambda})_k]^2 \text{E} [(\mathbf{z}_{p;\lambda})_k]^2 = 1$. Those observations follow with the same arguments that the $(\mathbf{z}_{m;\lambda})_k$ are i.i.d. standard normal random variables and therefore $\text{E} [(\mathbf{z}_{m;\lambda})_k^2] = 1$ and $\text{E} [(\mathbf{z}_{m;\lambda})_k] = 0$. Hence,

$$\sigma^{(g)4} \frac{1}{\mu^4} \text{Var} \left[\left(\sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^2 \right] = \frac{2\sigma^{(g)4}\mu}{\mu^4} + \frac{\sigma^{(g)4}\mu(\mu-1)}{\mu^4} = \frac{\sigma^{(g)4}}{\mu^3} (1 + \mu). \quad (\text{E.196})$$

Using the abbreviations A and B introduced in Equation (E.191), the covariance term can be written as

$$\text{Cov} [A, B] = \text{E} [AB] - \text{E} [A] \text{E} [B] \quad (\text{E.197})$$

$$\begin{aligned} &= \text{E} \left[\left(2\sigma^{(g)}(\mathbf{x}^{(g)})_k \frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right) \left(\sigma^{(g)} \frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^2 \right] \\ &\quad - \text{E} \left[\left(2\sigma^{(g)}(\mathbf{x}^{(g)})_k \frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right) \right] \text{E} \left[\left(\sigma^{(g)} \frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^2 \right]. \end{aligned} \quad (\text{E.198})$$

Equation (E.198) can be rewritten using linearity of expectation yielding

$$\text{Cov} [A, B] = 2\sigma^{(g)3}(\mathbf{x}^{(g)})_k \frac{1}{\mu^3} \text{E} \left[\left(\sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^3 \right] - 2\sigma^{(g)3}(\mathbf{x}^{(g)})_k \frac{1}{\mu^3} \sum_{m=1}^{\mu} \text{E} [(\mathbf{z}_{m;\lambda})_k] \text{E} \left[\left(\sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^2 \right]. \quad (\text{E.199})$$

$E[(\mathbf{z}_{m;\lambda})_k] = 0$ is used again and hence the second term on the right-hand side in Equation (E.199) is zero, i.e.,

$$2\sigma^{(g)3}(\mathbf{x}^{(g)})_k \frac{1}{\mu^3} E[(\mathbf{z}_{m;\lambda})_k] E \left[\left(\sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^2 \right] = 0. \quad (\text{E.200})$$

For the first term on the right-hand side of Equation (E.199), the sum is rewritten to

$$\begin{aligned} \left(\sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^3 &= \sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k^3 + \sum_{m=1}^{\mu} \sum_{\substack{p=1 \\ p \neq m}}^{\mu} 3(\mathbf{z}_{m;\lambda})_k^2 (\mathbf{z}_{p;\lambda})_k \\ &\quad + \sum_{m=1}^{\mu} \sum_{\substack{p=1 \\ p \neq m}}^{\mu} \sum_{\substack{s=1 \\ s \neq p \\ s \neq m}}^{\mu} (\mathbf{z}_{m;\lambda})_k (\mathbf{z}_{p;\lambda})_k (\mathbf{z}_{s;\lambda})_k. \end{aligned} \quad (\text{E.201})$$

Because the projection does not have an influence for the “feasible” case under consideration and selection is done along the x_1 -direction, the $(\mathbf{z}_{m;\lambda})_k$ are i.i.d. standard normal random variables. Additionally, using linearity of expectation, calculating the expected value of Equation (E.201) results in

$$\begin{aligned} E \left[\left(\sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^3 \right] &= \sum_{m=1}^{\mu} E[(\mathbf{z}_{m;\lambda})_k^3] + 3 \sum_{m=1}^{\mu} \sum_{\substack{p=1 \\ p \neq m}}^{\mu} E[(\mathbf{z}_{m;\lambda})_k^2] E[(\mathbf{z}_{p;\lambda})_k] \\ &\quad + \sum_{m=1}^{\mu} \sum_{\substack{p=1 \\ p \neq m}}^{\mu} \sum_{\substack{s=1 \\ s \neq p \\ s \neq m}}^{\mu} E[(\mathbf{z}_{m;\lambda})_k] E[(\mathbf{z}_{p;\lambda})_k] E[(\mathbf{z}_{s;\lambda})_k]. \end{aligned} \quad (\text{E.202})$$

All three terms in Equation (E.202) are zero due to the moments of the standard normal distribution. Consequently,

$$E \left[\left(\sum_{m=1}^{\mu} (\mathbf{z}_{m;\lambda})_k \right)^3 \right] = 0 \quad (\text{E.203})$$

follows and therefore with Equation (E.200) and Equation (E.203) one obtains

$$\text{Cov}[A, B] = 0. \quad (\text{E.204})$$

With these results, insertion of Equation (E.193), Equation (E.196), and Equation (E.204) into Equation (E.192) results in

$$\text{Var}[\langle \tilde{r} \rangle^2] \approx \left(\sum_{k=2}^N \frac{4\sigma^{(g)2}}{\mu} (\mathbf{x}^{(g)})_k^2 \right) + \frac{\sigma^{(g)4}(N-1)}{\mu^3} (1+\mu) \quad (\text{E.205})$$

$$= \frac{4\sigma^{(g)2} r^{(g)2}}{\mu} + \frac{\sigma^{(g)4}(N-1)}{\mu^3} (1+\mu) \quad (\text{E.206})$$

$$= \frac{4\sigma^{(g)*2} r^{(g)4}}{N^2 \mu} + \frac{\sigma^{(g)*4} r^{(g)4} (N-1)(1+\mu)}{N^4 \mu^3}. \quad (\text{E.207})$$

For the expected value,

$$E[\langle \tilde{r} \rangle^2] \simeq r^{(g)2} \quad (\text{E.208})$$

follows directly from the derivations leading to Equation (E.187) for $N \rightarrow \infty$ with $N \gg \sigma^{(g)*}$. Therefore, with Equation (E.207) and Equation (E.208) one obtains

$$\begin{aligned} \frac{\sqrt{\text{Var}[\langle \tilde{r} \rangle^2]}}{E[\langle \tilde{r} \rangle^2]} &\approx \sqrt{\frac{4\sigma^{(g)*2} r^{(g)4}}{N^2 \mu r^{(g)4}} + \frac{\sigma^{(g)*4} r^{(g)4} (N-1)(1+\mu)}{N^4 \mu^3 r^{(g)4}}} = \sqrt{\frac{4\sigma^{(g)*2}}{N^2 \mu} + \frac{\sigma^{(g)*4} (N-1)(1+\mu)}{N^4 \mu^3}} \\ &\simeq \frac{\sigma^{(g)*}}{N\sqrt{\mu}} \sqrt{4 + \frac{\sigma^{(g)*2} (1+\mu)}{N\mu^2}}, \end{aligned} \quad (\text{E.209})$$

where in the last step $N \gg 1$ has been assumed. Hence, (E.209) tends to zero for $N \rightarrow \infty$ with $N \gg \sigma^{(g)*2}$.

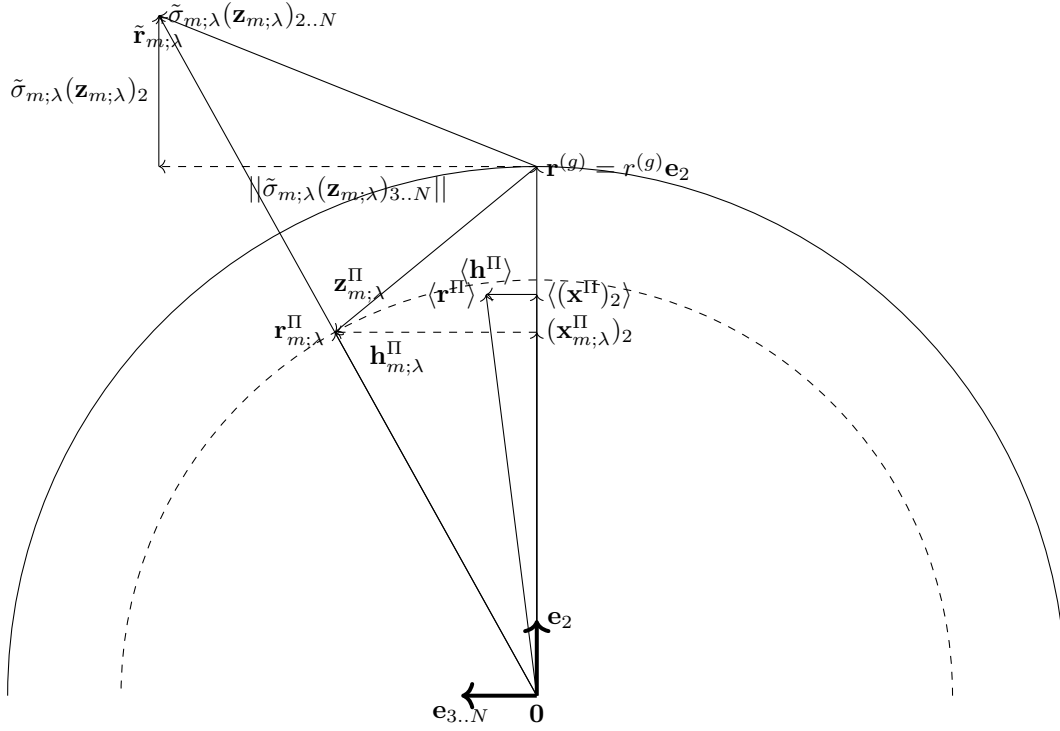


Fig. 10. Visualization of the geometrical approach for deriving $\varphi_{r_{\text{infeas}}}$ for the multi-recombinative ES applied to the conically constrained problem.

B. The Approximate r Progress Rate in the case that the probability of feasible offspring tends to 0

For the case that the probability of feasible offspring tends to 0, the aim is to relate the r progress rate to the x progress rate. It has also been done in this way for the $(1, \lambda)$ case in Section IV-B2. Here, a geometrical approach is pursued that is visualized in Fig. 10. The visualization shows the $2..N$ space of the conically constrained problem. That is, the view in direction of the cone axis (x_1 direction) is shown. The x_1 direction can be imagined “coming out” of the paper perpendicular to \mathbf{e}_2 and $\mathbf{e}_{3..N}$. For this constellation, the optimal value can be imagined being directly “on the paper”. The origin is indicated as $\mathbf{0}$ and the coordinate axes in the $2..N$ space are indicated as \mathbf{e}_2 and $\mathbf{e}_{3..N}$. The coordinate system is rotated such that the distance from the cone axis of the parental individual corresponds to the parental parameter vector’s value in the second dimension, i.e., $\mathbf{x}^{(g)} = (x^{(g)}, r^{(g)}, 0, \dots, 0)^T$. The parental individual is therefore located at the position indicated by $\mathbf{r}^{(g)}$. The solid half circle indicates the cone boundary at $x_1 = x^{(g)}$. As the case under consideration is the one in which the probability of feasible offspring tends to 0, the parental individual is assumed to be on the cone boundary². The mutation vector $\tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{2..N}$ indicates the mutation vector in the $2..N$ space leading to the m -th best offspring before projection $\tilde{\mathbf{r}}_{m;\lambda}$. This mutation vector can be decomposed into vectors in direction of \mathbf{e}_2 and $\mathbf{e}_{3..N}$. The lengths of those vectors are indicated as $\tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_2$ and $\|\tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{3..N}\|$, respectively. The value in direction of the cone boundary of the m -th best offspring after projection $\mathbf{x}_{m;\lambda}^{\Pi}$ is assumed such that the dashed half circle indicates the cone boundary corresponding to $(\mathbf{x}_{m;\lambda}^{\Pi})_1$. As the dashed half circle is drawn smaller than the solid half circle, it is assumed that the m -th best offspring after projection $\mathbf{r}_{m;\lambda}^{\Pi}$ has a smaller value in direction of x_1 than the parental individual. The m -th best offspring under consideration is considered infeasible. Hence, it is projected onto the cone boundary. Its decomposition into vectors along the \mathbf{e}_2 and $\mathbf{e}_{3..N}$ axes are indicated as $(\mathbf{x}_{m;\lambda}^{\Pi})_2$ and $\mathbf{h}_{m;\lambda}^{\Pi}$, respectively. The centroid of the best μ offspring after projection has to be computed for the calculation of the r progress rate. This centroid is indicated as $\langle \mathbf{r}^{\Pi} \rangle$ with its decompositions in \mathbf{e}_2 and $\mathbf{e}_{3..N}$ directions $\langle (\mathbf{x}^{\Pi})_2 \rangle$ and $\langle \mathbf{h}^{\Pi} \rangle$, respectively.

$\varphi_{r_{\text{infeas}}}$ can be computed by geometric considerations with the help of Fig. 10. Note that $\langle q_r \rangle_{\text{infeas}} = \|\langle \mathbf{r}^{\Pi} \rangle\|$ holds. Hence,

$$\varphi_{r_{\text{infeas}}} = r^{(g)} - \mathbb{E}[\|\langle \mathbf{r}^{\Pi} \rangle\|] \quad (\text{E.210})$$

²An asymptotic expression for the feasibility probability has been derived as $P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \simeq \Phi \left[\frac{1}{\sigma^{(g)}} \left(\frac{x^{(g)}}{\sqrt{\xi}} - \bar{r} \right) \right] = \Phi \left[\frac{N}{\sigma^{(g)*}} \left(\frac{x^{(g)}}{r^{(g)}\sqrt{\xi}} - \frac{\bar{r}}{r^{(g)}} \right) \right]$ (see [13, Sec. 3.1.2.1.2.8, pp. 44-49]). Note that for the case that the parental individual is on the cone boundary, $r^{(g)} = \frac{x^{(g)}}{\sqrt{\xi}}$ holds. Further, $\bar{r} > r^{(g)}$ follows from Equation (B.62). Hence, the feasibility probability tends to 0 for the case of the parental individual being on the cone boundary for $N \rightarrow \infty$ with $N \gg \sigma^{(g)*}$.

follows and $E[||\langle \mathbf{r}^\Pi \rangle ||]$ needs to be derived. By Pythagoras' theorem,

$$E[||\langle \mathbf{r}^\Pi \rangle ||] = E \left[\sqrt{\langle (\mathbf{x}^\Pi)_2 \rangle^2 + ||\langle \mathbf{h}^\Pi \rangle ||^2} \right] \quad (\text{E.211})$$

follows. With the assumption that the expression under the square root deviates only slightly from its mean, a Taylor expansion can be used. The idea is the same as for the derivation leading to Equation (E.177) and results in

$$E[||\langle \mathbf{r}^\Pi \rangle ||] = E \left[\sqrt{\langle (\mathbf{x}^\Pi)_2 \rangle^2 + ||\langle \mathbf{h}^\Pi \rangle ||^2} \right] = \sqrt{E \left[\langle (\mathbf{x}^\Pi)_2 \rangle^2 + ||\langle \mathbf{h}^\Pi \rangle ||^2 \right]} + \dots \quad (\text{E.212})$$

$$\approx \sqrt{E \left[\langle (\mathbf{x}^\Pi)_2 \rangle^2 + ||\langle \mathbf{h}^\Pi \rangle ||^2 \right]} \quad (\text{E.213})$$

$$= \sqrt{E \left[\langle (\mathbf{x}^\Pi)_2 \rangle^2 \right] + E \left[||\langle \mathbf{h}^\Pi \rangle ||^2 \right]} \quad (\text{E.214})$$

where the last step follows from linearity of expectation. To proceed further, $E \left[\langle (\mathbf{x}^\Pi)_2 \rangle^2 \right]$ and $E \left[||\langle \mathbf{h}^\Pi \rangle ||^2 \right]$ need to be derived.

For $E \left[\langle (\mathbf{x}^\Pi)_2 \rangle^2 \right]$,

$$E \left[\langle (\mathbf{x}^\Pi)_2 \rangle^2 \right] = E \left[\left(\frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{x}_{m;\lambda}^\Pi)_2 \right)^2 \right] \quad (\text{E.215})$$

can be written by expanding the notation for the centroid computation. By the intercept theorem,

$$\frac{||\mathbf{r}_{m;\lambda}^\Pi||}{(\mathbf{x}_{m;\lambda}^\Pi)_2} = \frac{||\tilde{\mathbf{r}}_{m;\lambda}||}{r^{(g)} + \tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_2} \quad (\text{E.216})$$

follows (see Fig. 10). This implies

$$(\mathbf{x}_{m;\lambda}^\Pi)_2 = \frac{||\mathbf{r}_{m;\lambda}^\Pi|| (r^{(g)} + \tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_2)}{||\tilde{\mathbf{r}}_{m;\lambda}||}. \quad (\text{E.217})$$

$||\mathbf{r}_{m;\lambda}^\Pi||$ can be expressed in a different way by using the projection equation (see Equation (A.56)) from Appendix A yielding

$$||\mathbf{r}_{m;\lambda}^\Pi|| = \frac{(\mathbf{x}_{m;\lambda}^\Pi)_1}{\sqrt{\xi}}. \quad (\text{E.218})$$

$||\tilde{\mathbf{r}}_{m;\lambda}||$ can be written in terms of the parental individual and the mutation as

$$||\tilde{\mathbf{r}}_{m;\lambda}|| = ||r^{(g)} \mathbf{e}_2 + \tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{2..N}||. \quad (\text{E.219})$$

Insertion of Equation (E.218) and Equation (E.219) into Equation (E.217) results in

$$(\mathbf{x}_{m;\lambda}^\Pi)_2 = \frac{(\mathbf{x}_{m;\lambda}^\Pi)_1 (r^{(g)} + \tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_2)}{\sqrt{\xi} ||r^{(g)} \mathbf{e}_2 + \tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{2..N}||}. \quad (\text{E.220})$$

Inserting Equation (E.220) into Equation (E.215) yields

$$E \left[\langle (\mathbf{x}^\Pi)_2 \rangle^2 \right] = E \left[\left(\frac{1}{\mu} \sum_{m=1}^{\mu} \frac{(\mathbf{x}_{m;\lambda}^\Pi)_1 (r^{(g)} + \tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_2)}{\sqrt{\xi} ||r^{(g)} \mathbf{e}_2 + \tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{2..N}||} \right)^2 \right]. \quad (\text{E.221})$$

In order to proceed further, sufficiently small τ is assumed such that $\tilde{\sigma}_{m;\lambda} \simeq \sigma^{(g)}$ holds. Additionally, $\sigma^{(g)}$ is assumed to be sufficiently small as well. These assumptions allow writing

$$r^{(g)} + \tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_2 \simeq r^{(g)} + \sigma^{(g)}(\mathbf{z}_{m;\lambda})_2 \simeq r^{(g)}. \quad (\text{E.222})$$

For simplifying the expression $||r^{(g)} \mathbf{e}_2 + \tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{2..N}||$, note that $(\mathbf{z}_{m;\lambda})_{2..N}$ corresponds to $(\mathbf{z})_{2..N}$ because selection is done in the x_1 direction. This allows writing

$$||r^{(g)} \mathbf{e}_2 + \tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{2..N}|| \simeq ||r^{(g)} \mathbf{e}_2 + \sigma^{(g)}(\mathbf{z})_{2..N}|| = \tilde{r} \simeq r^{(g)} \sqrt{1 + \frac{\sigma^{(g)*2}}{N}} \quad (\text{E.223})$$

where the last step follows from the normal approximation of the distribution of a single offspring's distance from the cone axis \tilde{r} (see Equation (B.62)).

Consideration of Equation (E.222) and Equation (E.223) for Equation (E.221) results in

$$\mathbb{E} \left[\langle (\mathbf{x}^\Pi)_2 \rangle^2 \right] \simeq \mathbb{E} \left[\left(\frac{1}{\mu} \sum_{m=1}^{\mu} \frac{(\mathbf{x}_{m;\lambda}^\Pi)_1 r^{(g)}}{\sqrt{\xi} r^{(g)} \sqrt{1 + \frac{\sigma^{(g)*2}}{N}}} \right)^2 \right] \quad (\text{E.224})$$

$$= \mathbb{E} \left[\left(\frac{1}{\sqrt{\xi} \sqrt{1 + \frac{\sigma^{(g)*2}}{N}}} \frac{1}{\mu} \sum_{m=1}^{\mu} (\mathbf{x}_{m;\lambda}^\Pi)_1 \right)^2 \right] \quad (\text{E.225})$$

$$= \mathbb{E} \left[\left(\frac{\langle (\mathbf{x}^\Pi)_1 \rangle}{\sqrt{\xi} \sqrt{1 + \frac{\sigma^{(g)*2}}{N}}} \right)^2 \right]. \quad (\text{E.226})$$

For $\mathbb{E} \left[\|\langle \mathbf{h}^\Pi \rangle\|^2 \right]$,

$$\mathbb{E} \left[\|\langle \mathbf{h}^\Pi \rangle\|^2 \right] = \mathbb{E} \left[\left(\frac{1}{\mu} \sum_{m=1}^{\mu} \mathbf{h}_{m;\lambda}^\Pi \right)^2 \right] \quad (\text{E.227})$$

$$= \frac{1}{\mu^2} \sum_{m=1}^{\mu} \sum_{p=1}^{\mu} \mathbb{E} \left[\mathbf{h}_{m;\lambda}^\Pi{}^T \mathbf{h}_{p;\lambda}^\Pi \right] \quad (\text{E.228})$$

$$= \frac{1}{\mu^2} \sum_{m=1}^{\mu} \mathbb{E} \left[\mathbf{h}_{m;\lambda}^\Pi{}^T \mathbf{h}_{m;\lambda}^\Pi \right] + \frac{1}{\mu^2} \sum_{m=1}^{\mu} \sum_{\substack{p=1 \\ p \neq m}}^{\mu} \mathbb{E} \left[\mathbf{h}_{m;\lambda}^\Pi{}^T \mathbf{h}_{p;\lambda}^\Pi \right] \quad (\text{E.229})$$

can be written by expanding the notation for the centroid computation. The second summand in Equation (E.229) vanishes. Note that the $\mathbf{h}_{m;\lambda}$ and $\mathbf{h}_{p;\lambda}$ vectors for particular m and p are in the $3..N$ space. Because selection is done in x_1 direction, they are statistically independent. Additionally, every direction for those vectors is equally probable because the $3..N$ is spherical (assuming the distortion through the projection is negligible). As a result, for fixed m and p ,

$$\mathbb{E} \left[\mathbf{h}_{m;\lambda}^\Pi{}^T \mathbf{h}_{p;\lambda}^\Pi \right] = \mathbb{E} \left[\mathbf{h}_{m;\lambda}^\Pi{}^T \right] \mathbb{E} \left[\mathbf{h}_{p;\lambda}^\Pi \right] = 0 \quad (\text{E.230})$$

holds. Consequently, these considerations result in

$$\mathbb{E} \left[\|\langle \mathbf{h}^\Pi \rangle\|^2 \right] = \frac{1}{\mu^2} \sum_{m=1}^{\mu} \mathbb{E} \left[\mathbf{h}_{m;\lambda}^\Pi{}^T \mathbf{h}_{m;\lambda}^\Pi \right] = \frac{1}{\mu^2} \sum_{m=1}^{\mu} \mathbb{E} \left[\|\mathbf{h}_{m;\lambda}^\Pi\|^2 \right]. \quad (\text{E.231})$$

Thus, $\mathbb{E} \left[\|\mathbf{h}_{m;\lambda}^\Pi\|^2 \right]$ needs to be derived to proceed. By the intercept theorem,

$$\frac{\|\mathbf{h}_{m;\lambda}^\Pi\|}{\|\mathbf{r}_{m;\lambda}^\Pi\|} = \frac{\|\tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{3..N}\|}{\|\tilde{\mathbf{r}}_{m;\lambda}\|} \quad (\text{E.232})$$

follows (see Fig. 10). With this and consideration of Equation (E.218),

$$\mathbb{E} \left[\|\mathbf{h}_{m;\lambda}^\Pi\|^2 \right] = \mathbb{E} \left[\left(\frac{(\mathbf{x}_{m;\lambda}^\Pi)_1 \|\tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{3..N}\|}{\sqrt{\xi} \|\tilde{\mathbf{r}}_{m;\lambda}\|} \right)^2 \right] \quad (\text{E.233})$$

can be written. In order to simplify Equation (E.233), it is assumed that the deviation of $\|\tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{3..N}\|^2$ and $\|\tilde{\mathbf{r}}_{m;\lambda}\|^2$ from their mean values is negligible. This allows replacing those random variable expressions with their mean values yielding

$$\mathbb{E} \left[\|\mathbf{h}_{m;\lambda}^\Pi\|^2 \right] \approx \mathbb{E} \left[\frac{(\mathbf{x}_{m;\lambda}^\Pi)_1^2}{\xi} \frac{\mathbb{E} \left[\|\tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{3..N}\|^2 \right]}{\mathbb{E} \left[\|\tilde{\mathbf{r}}_{m;\lambda}\|^2 \right]} \right] \quad (\text{E.234})$$

$$= \mathbb{E} \left[\frac{(\mathbf{x}_{m;\lambda}^\Pi)_1^2}{\xi} \frac{\mathbb{E} \left[\|\tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{3..N}\|^2 \right]}{\mathbb{E} \left[\|\tilde{r}^{(g)} \mathbf{e}_2 + \tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{2..N}\|^2 \right]} \right]. \quad (\text{E.235})$$

It can be seen as follows that the fluctuations tend to zero in relation to the mean value for asymptotic considerations.

$$\text{Var} \left[\|\tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{3..N}\|^2 \right] = \text{Var} \left[\sum_{k=3}^N \tilde{\sigma}_{m;\lambda}^2(\mathbf{z}_{m;\lambda})_k^2 \right] \quad (\text{E.236})$$

can be derived. Assuming again that the σ mutation can be neglected, we have $\tilde{\sigma}_{m;\lambda} \approx \sigma^{(g)}$. The random variables $(\mathbf{z}_{m;\lambda})_k$ for $k \in \{2..N\}$ are isotropically distributed in the $2..N$ subspace due to the selective neutrality of the i.i.d. normally distributed mutations in this subspace. The sum of the mutations is assumed to be χ^2 distributed and it follows that

$$\text{Var} [||\tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{3..N}||^2] \approx 2\sigma^{(g)^4}(N-2) \quad (\text{E.237})$$

holds. With the same arguments,

$$\text{E} [||\tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{3..N}||^2] \approx \sigma^{(g)^2}(N-2) \quad (\text{E.238})$$

follows. With Equation (E.237) and Equation (E.238), it can now be shown that the standard deviation in relation to the expected value tends to zero for $N \rightarrow \infty$:

$$\frac{\sqrt{\text{Var} [||\tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{3..N}||^2]}}{\text{E} [||\tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{3..N}||^2]} \approx \frac{\sqrt{2\sigma^{(g)^4}(N-2)}}{\sigma^{(g)^2}(N-2)} = \frac{\sqrt{2}}{\sqrt{N-2}}. \quad (\text{E.239})$$

In order to show the same for $\frac{\sqrt{\text{Var} [||r^{(g)}\mathbf{e}_2 + \tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{2..N}||^2]}}{\text{E} [||r^{(g)}\mathbf{e}_2 + \tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{2..N}||^2]}$, it is again assumed that $\tilde{\sigma}_{m;\lambda} \approx \sigma^{(g)}$ and that the selection has no influence on the $2..N$ components. Hence, use of the calculation leading to Equation (B.58) and Equation (B.59) results in

$$\text{Var} [||r^{(g)}\mathbf{e}_2 + \tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{2..N}||^2] \approx 4\sigma^{(g)^2}r^{(g)^2} + 2\sigma^{(g)^4}(N-1) = \frac{4\sigma^{(g)*2}r^{(g)^4}}{N^2} + \frac{2\sigma^{(g)*4}r^{(g)^4}(N-1)}{N^4}. \quad (\text{E.240})$$

Note that with the same assumptions used for deriving Equation (E.240),

$$\text{E} [||r^{(g)}\mathbf{e}_2 + \tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{2..N}||^2] \approx \text{E} [\tilde{r}^2] \approx r^{(g)^2} \left(1 + \frac{\sigma^{(g)*2}}{N} \right) \simeq r^{(g)^2} \quad (\text{E.241})$$

follows for $N \rightarrow \infty$ with $N \gg \sigma^{(g)*2}$ using the normal approximation of the distribution of a single offspring's distance from the cone axis \tilde{r} (see Equation (B.62)). With Equation (E.240) and Equation (E.241), we can now write

$$\begin{aligned} \frac{\sqrt{\text{Var} [||r^{(g)}\mathbf{e}_2 + \tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{2..N}||^2]}}{\text{E} [||r^{(g)}\mathbf{e}_2 + \tilde{\sigma}_{m;\lambda}(\mathbf{z}_{m;\lambda})_{2..N}||^2]} &\approx \sqrt{\frac{4N^2\sigma^{(g)*2}r^{(g)^4} + 2\sigma^{(g)*4}r^{(g)^4}(N-1)}{N^4r^{(g)^4}}} \simeq \sqrt{\frac{4\sigma^{(g)*2}}{N^2} + \frac{2\sigma^{(g)*4}}{N^3}} \\ &= \frac{\sigma^{(g)*}}{N} \sqrt{4 + \frac{2\sigma^{(g)*2}}{N}}, \end{aligned} \quad (\text{E.242})$$

where in the second to last step $N \gg 1$ has been assumed. This result shows that for $N \rightarrow \infty$ with $N \gg \sigma^{(g)*2}$, the fluctuations around the mean are negligible.

Now τ is assumed to be sufficiently small such that $\tilde{\sigma}_{m;\lambda} \simeq \sigma^{(g)}$. Moreover, because selection is done in the x_1 direction, the $\mathbf{z}_{m;\lambda}$ components from 2 to N are all standard normally distributed. With these observations,

$$\text{E} [||\mathbf{h}_{m;\lambda}^\Pi||^2] \approx \text{E} \left[\frac{(\mathbf{x}_{m;\lambda}^\Pi)_1^2}{\xi} \frac{\sigma^{(g)^2}(N-2)}{\text{E} [||r^{(g)}\mathbf{e}_2 + \sigma^{(g)}(\mathbf{z})_{2..N}||^2]} \right] \quad (\text{E.243})$$

$$= \text{E} \left[\frac{(\mathbf{x}_{m;\lambda}^\Pi)_1^2}{\xi} \frac{\sigma^{(g)^2}(N-2)}{\text{E} [\tilde{r}^2]} \right] \quad (\text{E.244})$$

$$\simeq \text{E} \left[\frac{(\mathbf{x}_{m;\lambda}^\Pi)_1^2}{\xi} \frac{\sigma^{(g)^2}(N-2)}{r^{(g)^2} \left(1 + \frac{\sigma^{(g)*2}}{N} \right)} \right] \quad (\text{E.245})$$

can be derived where the last step follows from the normal approximation of the distribution of a single offspring's distance from the cone axis \tilde{r} (see Equation (B.62)). Insertion of Equation (E.245) into Equation (E.231) yields

$$\text{E} [||\langle \mathbf{h}^\Pi \rangle||^2] \approx \frac{1}{\mu^2} \sum_{m=1}^{\mu} \text{E} \left[\frac{(\mathbf{x}_{m;\lambda}^\Pi)_1^2}{\xi} \frac{\sigma^{(g)^2}(N-2)}{r^{(g)^2} \left(1 + \frac{\sigma^{(g)*2}}{N} \right)} \right] \quad (\text{E.246})$$

$$= \left(\frac{1}{\mu} \frac{1}{\xi} \frac{\sigma^{(g)^2}(N-2)}{r^{(g)^2} \left(1 + \frac{\sigma^{(g)*2}}{N} \right)} \right) \left(\frac{1}{\mu} \sum_{m=1}^{\mu} \text{E} [(\mathbf{x}_{m;\lambda}^\Pi)_1^2] \right) \quad (\text{E.247})$$

$$= \left(\frac{1}{\mu} \frac{1}{\xi} \frac{\sigma^{(g)^2}(N-2)}{r^{(g)^2} \left(1 + \frac{\sigma^{(g)*2}}{N} \right)} \right) \text{E} [(\langle \mathbf{x}^\Pi \rangle_1^2)]. \quad (\text{E.248})$$

Insertion of Equation (E.226) and Equation (E.248) into Equation (E.214) results in

$$\mathbb{E}[\|\langle \mathbf{x}^\Pi \rangle\|] \approx \sqrt{\frac{\mathbb{E}[\langle (\mathbf{x}^\Pi)_1 \rangle^2]}{\xi \left(1 + \frac{\sigma^{(g)*2}}{N}\right)} + \left(\frac{1}{\mu} \frac{1}{\xi} \frac{\sigma^{(g)2}(N-2)}{r^{(g)2} \left(1 + \frac{\sigma^{(g)*2}}{N}\right)}\right) \mathbb{E}[\langle (\mathbf{x}^\Pi)_1^2 \rangle]} \quad (\text{E.249})$$

$$\approx \sqrt{\frac{\mathbb{E}[\langle (\mathbf{x}^\Pi)_1 \rangle^2]}{\xi \left(1 + \frac{\sigma^{(g)*2}}{N}\right)} + \left(\frac{1}{\mu} \frac{1}{\xi} \frac{\sigma^{(g)2}(N-2)}{r^{(g)2} \left(1 + \frac{\sigma^{(g)*2}}{N}\right)}\right) \mathbb{E}[\langle (\mathbf{x}^\Pi)_1^2 \rangle]} \quad (\text{E.250})$$

where in the last step it has been assumed that the deviation of $\langle (\mathbf{x}^\Pi)_1 \rangle$ from its mean is negligible. More formally, $\langle (\mathbf{x}^\Pi)_1 \rangle$ can be written as the sum of its expected value and a stochastic term

$$\langle (\mathbf{x}^\Pi)_1 \rangle = \mathbb{E}[\langle (\mathbf{x}^\Pi)_1 \rangle] + \epsilon_{\langle (\mathbf{x}^\Pi)_1 \rangle}. \quad (\text{E.251})$$

This stochastic term $\epsilon_{\langle (\mathbf{x}^\Pi)_1 \rangle}$ is assumed to be negligible which leads to

$$\mathbb{E}[\langle (\mathbf{x}^\Pi)_1 \rangle^2] \approx \mathbb{E}[\langle (\mathbf{x}^\Pi)_1 \rangle]^2. \quad (\text{E.252})$$

Note that the x_1 value after projection corresponds to q .

A condition for the justification of the approximation in Equation (E.252) is derived in the following. The idea is to show that $\frac{\sqrt{\text{Var}[\langle (\mathbf{x}^\Pi)_1 \rangle]}}{\mathbb{E}[\langle (\mathbf{x}^\Pi)_1 \rangle]}$ tends to zero for $N \rightarrow \infty$, which means that the fluctuations relative to the expected value are negligible. The variance $\text{Var}[\langle (\mathbf{x}^\Pi)_1 \rangle] = \text{Var}[q]_{\text{infeas}}$ can be calculated as

$$\text{Var}[q]_{\text{infeas}} = \mathbb{E}[\langle q \rangle_{\text{infeas}}^2] - \mathbb{E}[\langle q \rangle_{\text{infeas}}]^2 \quad (\text{E.253})$$

according to the König-Huygens' theorem. The first addend can be written as

$$\mathbb{E}[\langle q \rangle_{\text{infeas}}^2] = \mathbb{E}\left[\left(\frac{1}{\mu} \sum_{m=1}^{\mu} q_{m;\lambda_{\text{infeas}}}\right)^2\right] = \frac{1}{\mu^2} \sum_{m=1}^{\mu} \sum_{p=1}^{\mu} \mathbb{E}[q_{m;\lambda_{\text{infeas}}} q_{p;\lambda_{\text{infeas}}}], \quad (\text{E.254})$$

where the square of the sum has been rewritten and linearity of expectation has been applied in the last step. Since the q values are positive due to the formulation of the problem and smaller q values are considered better w.r.t. fitness, Equation (E.254) can be bounded above using the worst q values

$$\mathbb{E}[\langle q \rangle_{\text{infeas}}^2] \leq \mathbb{E}[q_{\mu;\lambda_{\text{infeas}}}^2]. \quad (\text{E.255})$$

An approximation for $\mathbb{E}[q_{\mu;\lambda_{\text{infeas}}}^2]$ has already been derived in Equation (G.384) as

$$\mathbb{E}[q_{\mu;\lambda_{\text{infeas}}}^2] \approx \frac{(\sigma^{(g)2} + \sigma_r^2/\xi)}{(1 + 1/\xi)^2} e_{(\mu-1),\lambda}^{0,2} - \frac{\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi}}{(1 + 1/\xi)^2} 2(x^{(g)} + \bar{r}/\sqrt{\xi}) e_{(\mu-1),\lambda}^{0,1} + \frac{(x^{(g)} + \bar{r}/\sqrt{\xi})^2}{(1 + 1/\xi)^2}. \quad (\text{E.256})$$

For $\mathbb{E}[\langle q \rangle_{\text{infeas}}]^2$, insertion of Equation (D.168) into Equation (E.252) with subsequent calculation of the square and rewriting results in

$$\begin{aligned} \mathbb{E}[\langle (\mathbf{x}^\Pi)_1 \rangle]^2 &= \mathbb{E}[\langle q \rangle_{\text{infeas}}]^2 \\ &\approx \left[\frac{\xi}{1 + \xi} \left(x^{(g)} + \bar{r}/\sqrt{\xi} \right) - \frac{\xi}{1 + \xi} \sqrt{\sigma^{(g)2} + \sigma_r^2/\xi} c_{\mu/\mu,\lambda} \right]^2 \end{aligned} \quad (\text{E.257})$$

$$= \left(\frac{(\sigma^{(g)2} + \sigma_r^2/\xi)}{(1 + 1/\xi)^2} c_{\mu/\mu,\lambda}^2 - \frac{\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi}}{(1 + 1/\xi)^2} 2(x^{(g)} + \bar{r}/\sqrt{\xi}) c_{\mu/\mu,\lambda} + \frac{(x^{(g)} + \bar{r}/\sqrt{\xi})^2}{(1 + 1/\xi)^2} \right). \quad (\text{E.258})$$

With Equation (E.253) and Equation (E.255), one obtains

$$\text{Var}[\langle q \rangle_{\text{infeas}}] \leq \mathbb{E}[q_{\mu;\lambda_{\text{infeas}}}^2] - \mathbb{E}[\langle q \rangle_{\text{infeas}}]^2 \quad (\text{E.259})$$

$$\approx \frac{(\sigma^{(g)2} + \sigma_r^2/\xi)}{(1 + 1/\xi)^2} \left(e_{(\mu-1),\lambda}^{0,2} - c_{\mu/\mu,\lambda}^2 \right) + \frac{\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi}}{(1 + 1/\xi)^2} 2(x^{(g)} + \bar{r}/\sqrt{\xi}) \left(c_{\mu/\mu,\lambda} - e_{(\mu-1),\lambda}^{0,1} \right) \quad (\text{E.260})$$

$$\simeq \frac{\sigma^{(g)*2} r^{(g)2}}{N^2(1 + 1/\xi)} \left(e_{(\mu-1),\lambda}^{0,2} - c_{\mu/\mu,\lambda}^2 \right) + \frac{\sigma^{(g)*} r^{(g)} \sqrt{1 + 1/\xi}}{N(1 + 1/\xi)^2} 2(x^{(g)} + \bar{r}/\sqrt{\xi}) \left(c_{\mu/\mu,\lambda} - e_{(\mu-1),\lambda}^{0,1} \right). \quad (\text{E.261})$$

From Equation (E.260) to Equation (E.261), the observation $\sigma_r \simeq \sigma^{(g)}$ for $N \rightarrow \infty$ from Equation (B.62) has been used and the normalization $\sigma^{(g)*} = \frac{\sigma^{(g)} N}{r^{(g)}}$ has been applied. Considering the expected value of $\langle q \rangle_{\text{infeas}}$, Equation (9) and Equation (11) yield

$$\mathbb{E}[\langle (\mathbf{x}^\Pi)_1 \rangle] = \mathbb{E}[\langle q \rangle_{\text{infeas}}] = x^{(g)} \left(1 - \frac{\varphi_{x\text{infeas}}^*}{N} \right). \quad (\text{E.262})$$

With the assumption of constant normalized progress in the steady state and the assumption of $N \gg \varphi_{x\text{infeas}}^*$ for $N \rightarrow \infty$, $\mathbb{E}[\langle q \rangle_{\text{infeas}}] \simeq x^{(g)}$. Hence, for showing that the fluctuation around $\mathbb{E}[\langle (\mathbf{x}^\Pi)_1 \rangle]$ is relatively small for the asymptotic case, the quotient $\frac{\sqrt{\text{Var}[\langle (\mathbf{x}^\Pi)_1 \rangle]}}{\mathbb{E}[\langle (\mathbf{x}^\Pi)_1 \rangle]}$ is considered for $N \rightarrow \infty$. Using Equation (E.261) and Equation (E.262), one can write

$$\frac{\sqrt{\text{Var}[\langle (\mathbf{x}^\Pi)_1 \rangle]}}{\mathbb{E}[\langle (\mathbf{x}^\Pi)_1 \rangle]} \leq \sqrt{\frac{\sigma^{(g)*2} r^{(g)2}}{N^2(1+1/\xi)x^{(g)2}} \left(e_{(\mu-1),\lambda}^{0,2} - c_{\mu/\mu,\lambda}^2 \right) + \frac{\sigma^{(g)*} r^{(g)} \sqrt{1+1/\xi}}{N(1+1/\xi)^2 x^{(g)2}} 2(x^{(g)} + \bar{r}/\sqrt{\xi}) \left(c_{\mu/\mu,\lambda} - e_{(\mu-1),\lambda}^{0,1} \right)}. \quad (\text{E.263})$$

By Equation (B.62), $\bar{r} \simeq r^{(g)}$ for $N \rightarrow \infty$. In addition, it is assumed that the ES has transitioned into the steady state and therefore $\frac{\sqrt{\xi} r_{ss}}{x_{ss}} \simeq 1$ holds by the arguments of Section V. Using these observations, Equation (E.263) results in

$$\frac{\sqrt{\text{Var}[\langle (\mathbf{x}^\Pi)_1 \rangle]}}{\mathbb{E}[\langle (\mathbf{x}^\Pi)_1 \rangle]} \leq \sqrt{\frac{\sigma_{ss}^{*2}}{N^2(1+1/\xi)\xi} \left(e_{(\mu-1),\lambda}^{0,2} - c_{\mu/\mu,\lambda}^2 \right) + \left(\frac{2\sigma_{ss}^* \sqrt{1+1/\xi}}{N(1+1/\xi)^2 \sqrt{\xi}} + \frac{\sigma_{ss}^* \sqrt{1+1/\xi}}{N(1+1/\xi)^2 \sqrt{\xi}} \right) \left(c_{\mu/\mu,\lambda} - e_{(\mu-1),\lambda}^{0,1} \right)}. \quad (\text{E.264})$$

For constant μ and λ , the progress coefficients are constant. Hence, Equation (E.264) tends to zero in the steady state for $N \rightarrow \infty$ with $N\xi \gg \sigma^{(g)*}$.

Due to Equation (9) and Equation (11),

$$\mathbb{E}[\langle (\mathbf{x}^\Pi)_1 \rangle] = \mathbb{E}[\langle q \rangle_{\text{infeas}}] = x^{(g)} \left(1 - \frac{\varphi_{x\text{infeas}}^*}{N} \right) \quad (\text{E.265})$$

holds and an approximation for $\varphi_{x\text{infeas}}^*$ has already been derived as Equation (C.73). Because

$$\mathbb{E}[\langle (\mathbf{x}^\Pi)_1^2 \rangle] = \mathbb{E}[\langle q^2 \rangle_{\text{infeas}}], \quad (\text{E.266})$$

the approximation derived in Appendix H (Equation (H.398)³) can be used. In order to simplify Equation (E.250) further, an additional assumption is made. It is assumed that

$$\mathbb{E}[\langle (\mathbf{x}^\Pi)_1^2 \rangle] \approx \mathbb{E}[\langle (\mathbf{x}^\Pi)_1 \rangle^2]. \quad (\text{E.267})$$

In order to derive a condition for which the assumption in Equation (E.267) is justified, Equations (E.252), (E.266), and (E.267) are investigated further. By Equations (E.266) and (H.398), one has

$$\begin{aligned} \mathbb{E}[\langle (\mathbf{x}^\Pi)_1^2 \rangle] &= \mathbb{E}[\langle q^2 \rangle_{\text{infeas}}] = \left(\frac{(\sigma^{(g)})^2 + \sigma_r^2/\xi}{(1+1/\xi)^2} \left[1 + e_{\mu,\lambda}^{1,1} \right] \right. \\ &\quad \left. - \frac{\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi}}{(1+1/\xi)^2} 2(x^{(g)} + \bar{r}/\sqrt{\xi}) c_{\mu/\mu,\lambda} + \frac{(x^{(g)} + \bar{r}/\sqrt{\xi})^2}{(1+1/\xi)^2} \right). \end{aligned} \quad (\text{E.268})$$

Insertion of Equation (D.168) into Equation (E.252) with subsequent calculation of the square and rewriting results in

$$\begin{aligned} \mathbb{E}[\langle (\mathbf{x}^\Pi)_1 \rangle^2] &= \mathbb{E}[\langle q \rangle_{\text{infeas}}]^2 \\ &\approx \left[\frac{\xi}{1+\xi} \left(x^{(g)} + \bar{r}/\sqrt{\xi} \right) - \frac{\xi}{1+\xi} \sqrt{\sigma^{(g)2} + \sigma_r^2/\xi} c_{\mu/\mu,\lambda} \right]^2 \end{aligned} \quad (\text{E.269})$$

$$\begin{aligned} &= \left(\frac{(\sigma^{(g)})^2 + \sigma_r^2/\xi}{(1+1/\xi)^2} c_{\mu/\mu,\lambda}^2 \right. \\ &\quad \left. - \frac{\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi}}{(1+1/\xi)^2} 2(x^{(g)} + \bar{r}/\sqrt{\xi}) c_{\mu/\mu,\lambda} + \frac{(x^{(g)} + \bar{r}/\sqrt{\xi})^2}{(1+1/\xi)^2} \right). \end{aligned} \quad (\text{E.270})$$

Comparison of Equation (E.268) with Equation (E.270) reveals that only the first term differs. For $N \rightarrow \infty$, one has $\sigma_r \simeq r^{(g)} \frac{\sigma^{(g)*}}{N}$ by Equation (B.62). Together with $\sigma^{(g)} = r^{(g)} \frac{\sigma^{(g)*}}{N}$, which follows from the definition of $\sigma^{(g)*} = \frac{\sigma^{(g)} N}{r^{(g)}}$, this yields

$$\frac{(\sigma^{(g)})^2 + \sigma_r^2/\xi}{(1+1/\xi)^2} \simeq \frac{\sigma^{(g)2} + \left(r^{(g)} \frac{\sigma^{(g)*}}{N} \right)^2 / \xi}{(1+1/\xi)^2} = \frac{\left(r^{(g)} \frac{\sigma^{(g)*}}{N} \right)^2}{(1+1/\xi)} = r^{(g)2} \frac{\sigma^{(g)*2}}{N^2(1+1/\xi)}. \quad (\text{E.271})$$

³Note that for the infeasible case only the second summand of Equation (H.398) is relevant.

Hence, for $N^2(1 + 1/\xi) \gg \sigma^{(g)*2}$, Equation (E.268) and Equation (E.270) are approximately equal, which justifies Equation (E.267).

As it has already been assumed that $E[\langle(\mathbf{x}^\Pi)_1\rangle^2] \approx E[\langle(\mathbf{x}^\Pi)_1\rangle]^2$,

$$E[|\langle\mathbf{r}^\Pi\rangle|] \approx \sqrt{\frac{E[\langle(\mathbf{x}^\Pi)_1\rangle]^2}{\xi \left(1 + \frac{\sigma^{(g)*2}}{N}\right)} + \left(\frac{1}{\mu} \frac{1}{\xi} \frac{\sigma^{(g)2}(N-2)}{r^{(g)2} \left(1 + \frac{\sigma^{(g)*2}}{N}\right)}\right) E[\langle(\mathbf{x}^\Pi)_1\rangle]^2} \quad (\text{E.272})$$

$$= \sqrt{\frac{E[\langle(\mathbf{x}^\Pi)_1\rangle]^2}{\xi \left(1 + \frac{\sigma^{(g)*2}}{N}\right)} \left(1 + \frac{1}{\mu} \frac{\sigma^{(g)2}(N-2)}{r^{(g)2}}\right)} \quad (\text{E.273})$$

$$= \frac{E[\langle(\mathbf{x}^\Pi)_1\rangle]}{\sqrt{\xi} \sqrt{1 + \frac{\sigma^{(g)*2}}{N}}} \sqrt{1 + \frac{1}{\mu} \frac{\sigma^{(g)*2}(N-2)}{N^2}} \quad (\text{E.274})$$

$\simeq \frac{\sigma^{(g)*2}}{N}$ for $N \rightarrow \infty$

$$\simeq \frac{E[\langle(\mathbf{x}^\Pi)_1\rangle]}{\sqrt{\xi} \sqrt{1 + \frac{\sigma^{(g)*2}}{N}}} \sqrt{1 + \frac{\sigma^{(g)*2}}{\mu N}} \quad (\text{E.275})$$

$$= \frac{E[\langle q \rangle_{\text{infeas}}]}{\sqrt{\xi}} \sqrt{\frac{1 + \frac{\sigma^{(g)*2}}{\mu N}}{1 + \frac{\sigma^{(g)*2}}{N}}} \quad (\text{E.276})$$

can now be written under the assumptions stated before. Insertion of Equation (E.276) into Equation (E.210) yields

$$\varphi_{r\text{infeas}} \approx r^{(g)} - \frac{E[\langle q \rangle_{\text{infeas}}]}{\sqrt{\xi}} \sqrt{\frac{1 + \frac{\sigma^{(g)*2}}{\mu N}}{1 + \frac{\sigma^{(g)*2}}{N}}} \quad (\text{E.277})$$

$$= r^{(g)} - \frac{x^{(g)} - \varphi_{x\text{infeas}}}{\sqrt{\xi}} \sqrt{\frac{1 + \frac{\sigma^{(g)*2}}{\mu N}}{1 + \frac{\sigma^{(g)*2}}{N}}}. \quad (\text{E.278})$$

Normalization of φ_r and φ_x yields

$$\varphi_{r\text{infeas}}^* \approx N \left(1 - \frac{x^{(g)}}{\sqrt{\xi} r^{(g)}} \left(1 - \frac{\varphi_{x\text{infeas}}^*}{N} \right) \sqrt{\frac{1 + \frac{\sigma^{(g)*2}}{\mu N}}{1 + \frac{\sigma^{(g)*2}}{N}}} \right) \quad (\text{E.279})$$

where an approximation for $\varphi_{x\text{infeas}}^*$ has already been derived as Equation (C.73).

C. The Approximate r Progress Rate - Combination Using the Single Offspring Feasibility Probability

Similarly to the x progress rate, the feasible and infeasible cases can be combined using the single offspring feasibility probability yielding

$$\varphi_r^* \approx P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \varphi_{r\text{feas}}^* + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] \varphi_{r\text{infeas}}^* \quad (\text{E.280})$$

$$\begin{aligned} &\approx P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) N \left(1 - \sqrt{1 + \frac{\sigma^{(g)*2}}{\mu N}} \right) \\ &+ [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] N \left(1 - \frac{x^{(g)}}{\sqrt{\xi} r^{(g)}} \left(1 - \frac{\varphi_{x\text{infeas}}^*}{N} \right) \sqrt{\frac{1 + \frac{\sigma^{(g)*2}}{\mu N}}{1 + \frac{\sigma^{(g)*2}}{N}}} \right) \end{aligned} \quad (\text{E.281})$$

APPENDIX F

DERIVATION OF THE SAR

From the definition of the SAR (Equation (14)) and the pseudo-code of the ES (Algorithm 1, Line 18) it follows that

$$\psi(x^{(g)}, r^{(g)}, \sigma^{(g)}) = \mathbb{E} \left[\frac{\sigma^{(g+1)} - \sigma^{(g)}}{\sigma^{(g)}} \mid x^{(g)}, r^{(g)}, \sigma^{(g)} \right] \quad (\text{F.282})$$

$$= \mathbb{E} \left[\frac{\left(\frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\sigma}_{m;\lambda} \right) - \sigma^{(g)}}{\sigma^{(g)}} \mid x^{(g)}, r^{(g)}, \sigma^{(g)} \right] \quad (\text{F.283})$$

$$= \mathbb{E} \left[\frac{1}{\mu} \sum_{m=1}^{\mu} \frac{\tilde{\sigma}_{m;\lambda} - \sigma^{(g)}}{\sigma^{(g)}} \mid x^{(g)}, r^{(g)}, \sigma^{(g)} \right] \quad (\text{F.284})$$

$$= \frac{1}{\mu} \sum_{m=1}^{\mu} \mathbb{E} \left[\frac{\tilde{\sigma}_{m;\lambda} - \sigma^{(g)}}{\sigma^{(g)}} \mid x^{(g)}, r^{(g)}, \sigma^{(g)} \right] \quad (\text{F.285})$$

$$= \frac{1}{\mu} \sum_{m=1}^{\mu} \int_{\sigma=0}^{\sigma=\infty} \left(\frac{\sigma - \sigma^{(g)}}{\sigma^{(g)}} \right) p_{\tilde{\sigma}_{m;\lambda}}(\sigma) d\sigma \quad (\text{F.286})$$

where $p_{\tilde{\sigma}_{m;\lambda}}(\sigma) := p_{\tilde{\sigma}_{m;\lambda}}(\sigma \mid x^{(g)}, r^{(g)}, \sigma^{(g)})$ denotes the probability density function of the m -th best offspring's mutation strength. Note that $\tilde{\sigma}_{m;\lambda}$ is not obtained by direct selection. It is the σ value of the individual with the m -th best objective function value among the λ offspring. Its probability density function can be derived as follows. A random sample σ from the conditional distribution density

$$p_{\sigma}(\sigma \mid \sigma^{(g)}) = \frac{1}{\sqrt{2\pi\tau}} \frac{1}{\sigma} \exp \left[-\frac{1}{2} \left(\frac{\ln(\sigma/\sigma^{(g)})}{\tau} \right)^2 \right] \quad (\text{F.287})$$

is considered. In order for this σ to be selected, the corresponding offspring's q value has to be the m -th best value among all the λ offspring. This is the case if $(\lambda - m)$ offspring have a greater projected value, and $(m - 1)$ offspring have a smaller projected value. The projected q density for a given mutation strength is given by $p_Q(q \mid x^{(g)}, r^{(g)}, \sigma)$. As there are $\lambda \binom{\lambda-1}{m-1} = \frac{\lambda!}{(\lambda-m)!(m-1)!}$ different possibilities for one offspring being the m -th best among λ descendants, it must be multiplied by this factor. Integrating over all possible q values yields

$$p_{\tilde{\sigma}_{m;\lambda}}(\sigma) = p_{\sigma}(\sigma \mid \sigma^{(g)}) \frac{\lambda!}{(\lambda-m)!(m-1)!} \int_{q=0}^{q=\infty} p_Q(q \mid x^{(g)}, r^{(g)}, \sigma) \times [1 - P_Q(q)]^{\lambda-m} [P_Q(q)]^{m-1} dq. \quad (\text{F.288})$$

This result can now be inserted into Equation (F.286) resulting in

$$\psi(x^{(g)}, r^{(g)}, \sigma^{(g)}) = \frac{1}{\mu} \sum_{m=1}^{\mu} \int_{\sigma=0}^{\sigma=\infty} \left(\frac{\sigma - \sigma^{(g)}}{\sigma^{(g)}} \right) p_{\sigma}(\sigma \mid \sigma^{(g)}) \quad (\text{F.289})$$

$$\begin{aligned} & \times \frac{\lambda!}{(\lambda-m)!(m-1)!} \int_{q=0}^{q=\infty} p_Q(q \mid x^{(g)}, r^{(g)}, \sigma) [1 - P_Q(q)]^{\lambda-m} [P_Q(q)]^{m-1} dq d\sigma \\ & = \int_{\sigma=0}^{\sigma=\infty} \left(\frac{\sigma - \sigma^{(g)}}{\sigma^{(g)}} \right) p_{\sigma}(\sigma \mid \sigma^{(g)}) \\ & \times \frac{\lambda!}{\mu} \int_{q=0}^{q=\infty} p_Q(q \mid x^{(g)}, r^{(g)}, \sigma) \sum_{m=1}^{\mu} \frac{[1 - P_Q(q)]^{\lambda-m} [P_Q(q)]^{m-1}}{(\lambda-m)!(m-1)!} dq d\sigma. \end{aligned} \quad (\text{F.290})$$

Use of Equation (D.85) with subsequent expression of the fraction using the binomial coefficient, one can rewrite Equation (F.290) to get

$$\begin{aligned} \psi(x^{(g)}, r^{(g)}, \sigma^{(g)}) & = \int_{\sigma=0}^{\sigma=\infty} \left(\frac{\sigma - \sigma^{(g)}}{\sigma^{(g)}} \right) p_{\sigma}(\sigma \mid \sigma^{(g)}) \\ & \times (\lambda - \mu) \binom{\lambda}{\mu} \int_{q=0}^{q=\infty} p_Q(q \mid x^{(g)}, r^{(g)}, \sigma) \int_{z=0}^{z=1-P_Q(q)} z^{\lambda-\mu-1} (1-z)^{\mu-1} dz dq d\sigma. \end{aligned} \quad (\text{F.291})$$

To proceed further, $z = 1 - P_Q(y)$ is substituted in the inner integral. This implies $y = P_Q^{-1}(1 - z)$ and $dz = -p_Q(y)$. The upper and lower bounds for the substituted integral follow as $y_u = P_Q^{-1}(1 - 1 + P_Q(q)) = q$ and $y_l = P_Q^{-1}(1 - 0) = \infty$, respectively. Therefore, one obtains

$$\begin{aligned} \psi(x^{(g)}, r^{(g)}, \sigma^{(g)}) &= \int_{\sigma=0}^{\sigma=\infty} \left(\frac{\sigma - \sigma^{(g)}}{\sigma^{(g)}} \right) p_{\sigma}(\sigma | \sigma^{(g)}) (\lambda - \mu) \binom{\lambda}{\mu} \\ &\quad \times \int_{q=0}^{q=\infty} p_Q(q | x^{(g)}, r^{(g)}, \sigma) \end{aligned} \quad (\text{F.292})$$

$$\begin{aligned} &\quad \times \int_{y=\infty}^{y=q} [1 - P_Q(y)]^{\lambda-\mu-1} [P_Q(y)]^{\mu-1} (-p_Q(y)) dy dq d\sigma \\ \psi(x^{(g)}, r^{(g)}, \sigma^{(g)}) &= \int_{\sigma=0}^{\sigma=\infty} \left(\frac{\sigma - \sigma^{(g)}}{\sigma^{(g)}} \right) p_{\sigma}(\sigma | \sigma^{(g)}) (\lambda - \mu) \binom{\lambda}{\mu} \\ &\quad \times \int_{q=0}^{q=\infty} p_Q(q | x^{(g)}, r^{(g)}, \sigma) \end{aligned} \quad (\text{F.293})$$

$$\times \int_{y=q}^{y=\infty} p_Q(y) [1 - P_Q(y)]^{\lambda-\mu-1} [P_Q(y)]^{\mu-1} dy dq d\sigma.$$

Changing the order of the two inner-most integrals one finally gets

$$\begin{aligned} \psi(x^{(g)}, r^{(g)}, \sigma^{(g)}) &= \int_{\sigma=0}^{\sigma=\infty} \left(\frac{\sigma - \sigma^{(g)}}{\sigma^{(g)}} \right) p_{\sigma}(\sigma | \sigma^{(g)}) (\lambda - \mu) \binom{\lambda}{\mu} \\ &\quad \times \int_{y=0}^{y=\infty} p_Q(y) [1 - P_Q(y)]^{\lambda-\mu-1} [P_Q(y)]^{\mu-1} \end{aligned} \quad (\text{F.294})$$

$$\times \int_{q=0}^{q=y} p_Q(q | x^{(g)}, r^{(g)}, \sigma) dq dy d\sigma$$

$$\begin{aligned} \psi(x^{(g)}, r^{(g)}, \sigma^{(g)}) &= \int_{\sigma=0}^{\sigma=\infty} \left(\frac{\sigma - \sigma^{(g)}}{\sigma^{(g)}} \right) p_{\sigma}(\sigma | \sigma^{(g)}) \\ &\quad \times (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=0}^{y=\infty} p_Q(y) [1 - P_Q(y)]^{\lambda-\mu-1} [P_Q(y)]^{\mu-1} \end{aligned} \quad (\text{F.295})$$

$$\times P_Q(y | x^{(g)}, r^{(g)}, \sigma) dy d\sigma.$$

Due to difficulties in analytically solving Equation (F.295), an approximate solution is derived. To this end, the approach used in [14, Sec. 7.3.2.4] is followed. For this, Equation (F.295) is written as

$$\psi(x^{(g)}, r^{(g)}, \sigma^{(g)}) = \int_{\sigma=0}^{\sigma=\infty} f(\sigma) p_{\sigma}(\sigma | \sigma^{(g)}) d\sigma \quad (\text{F.296})$$

$$= \int_{\sigma=0}^{\sigma=\infty} f(\sigma) \frac{1}{\sqrt{2\pi\tau}} \frac{1}{\sigma} \exp \left[-\frac{1}{2} \left(\frac{\ln(\sigma/\sigma^{(g)})}{\tau} \right)^2 \right] d\sigma \quad (\text{F.297})$$

with

$$\begin{aligned} f(\sigma) &= \left(\frac{\sigma - \sigma^{(g)}}{\sigma^{(g)}} \right) (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=0}^{y=\infty} p_Q(y) [1 - P_Q(y)]^{\lambda-\mu-1} [P_Q(y)]^{\mu-1} \\ &\quad P_Q(y | x^{(g)}, r^{(g)}, \sigma) dy. \end{aligned} \quad (\text{F.298})$$

Now, τ is assumed to be small. Therefore, Equation (F.297) can be expanded into a Taylor series at $\tau = 0$. Further, as τ is small, the probability mass of the log-normally distributed values is concentrated around $\sigma^{(g)}$. Therefore, $f(\sigma)$ can be expanded into a Taylor series at $\sigma = \sigma^{(g)}$. After further calculation (it is referred to [14, Sec. 7.3.2.4] for all the details) one obtains

$$\psi(x^{(g)}, r^{(g)}, \sigma^{(g)}) = f(\sigma^{(g)}) + \frac{\tau^2}{2} \sigma^{(g)} \frac{\partial f}{\partial \sigma} \bigg|_{\sigma=\sigma^{(g)}} + \frac{\tau^2}{2} \sigma^{(g)2} \frac{\partial^2 f}{\partial \sigma^2} \bigg|_{\sigma=\sigma^{(g)}} + O(\tau^4). \quad (\text{F.299})$$

To proceed further, the expressions $f(\sigma^{(g)})$, $\frac{\partial f}{\partial \sigma} \big|_{\sigma=\sigma^{(g)}}$, and $\frac{\partial^2 f}{\partial \sigma^2} \big|_{\sigma=\sigma^{(g)}}$ need to be evaluated. Because $f(\sigma^{(g)}) = \frac{\sigma^{(g)} - \sigma^{(g)}}{\sigma^{(g)}} \times \dots = 0$,

$$f(\sigma^{(g)}) = 0 \quad (\text{F.300})$$

immediately follows. For $\frac{\partial f}{\partial \sigma} \Big|_{\sigma=\sigma^{(g)}}$, one derives with the product rule

$$\begin{aligned} \frac{\partial f}{\partial \sigma} &= \frac{1}{\sigma^{(g)}} (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=0}^{y=\infty} p_Q(y) [1 - P_Q(y)]^{\lambda-\mu-1} [P_Q(y)]^{\mu-1} P_Q(y | x^{(g)}, r^{(g)}, \sigma) dy \\ &\quad + \left(\frac{\sigma - \sigma^{(g)}}{\sigma^{(g)}} \right) (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=0}^{y=\infty} p_Q(y) [1 - P_Q(y)]^{\lambda-\mu-1} [P_Q(y)]^{\mu-1} \\ &\quad \times \frac{\partial}{\partial \sigma} P_Q(y | x^{(g)}, r^{(g)}, \sigma) dy. \end{aligned} \quad (\text{F.301})$$

Therefore, it follows that

$$\begin{aligned} \frac{\partial f}{\partial \sigma} \Big|_{\sigma=\sigma^{(g)}} &= \frac{1}{\sigma^{(g)}} (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=0}^{y=\infty} p_Q(y) [1 - P_Q(y)]^{\lambda-\mu-1} [P_Q(y)]^{\mu-1} \\ &\quad \times \underbrace{P_Q(y | x^{(g)}, r^{(g)}, \sigma^{(g)})}_{=P_Q(y)} dy \\ &\quad + \underbrace{\left(\frac{\sigma^{(g)} - \sigma^{(g)}}{\sigma^{(g)}} \right)}_{=0} (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=0}^{y=\infty} p_Q(y) [1 - P_Q(y)]^{\lambda-\mu-1} [P_Q(y)]^{\mu-1} \end{aligned} \quad (\text{F.302})$$

$$\begin{aligned} &\quad \times \frac{\partial}{\partial \sigma} P_Q(y | x^{(g)}, r^{(g)}, \sigma) dy \Big|_{\sigma=\sigma^{(g)}} \\ &= \frac{1}{\sigma^{(g)}} \frac{\lambda!}{(\lambda - (\mu + 1))! ((\mu + 1) - 1)!} \\ &\quad \times \int_{y=0}^{y=\infty} p_Q(y) [1 - P_Q(y)]^{\lambda-(\mu+1)} [P_Q(y)]^{(\mu+1)-1} dy \end{aligned} \quad (\text{F.303})$$

$$\begin{aligned} &= \frac{1}{\sigma^{(g)}} \underbrace{\int_{y=0}^{y=\infty} p_{q_{(\mu+1);\lambda}}(y) dy}_{=1} \\ &= \frac{1}{\sigma^{(g)}}. \end{aligned} \quad (\text{F.304})$$

$$= \frac{1}{\sigma^{(g)}}. \quad (\text{F.305})$$

Computing the derivative with respect to σ of Equation (F.301) using the product rule for the second summand results in

$$\begin{aligned} \frac{\partial^2 f}{\partial \sigma^2} &= \frac{1}{\sigma^{(g)}} (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=0}^{y=\infty} p_Q(y) [1 - P_Q(y)]^{\lambda-\mu-1} [P_Q(y)]^{\mu-1} \\ &\quad \times \frac{\partial}{\partial \sigma} P_Q(y | x^{(g)}, r^{(g)}, \sigma) dy \\ &\quad + \frac{1}{\sigma^{(g)}} (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=0}^{y=\infty} p_Q(y) [1 - P_Q(y)]^{\lambda-\mu-1} [P_Q(y)]^{\mu-1} \\ &\quad \times \frac{\partial}{\partial \sigma} P_Q(y | x^{(g)}, r^{(g)}, \sigma) dy \\ &\quad + \left(\frac{\sigma - \sigma^{(g)}}{\sigma^{(g)}} \right) (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=0}^{y=\infty} p_Q(y) [1 - P_Q(y)]^{\lambda-\mu-1} [P_Q(y)]^{\mu-1} \\ &\quad \times \frac{\partial^2}{\partial \sigma^2} P_Q(y | x^{(g)}, r^{(g)}, \sigma) dy. \end{aligned} \quad (\text{F.306})$$

This implies

$$\begin{aligned} \frac{\partial^2 f}{\partial \sigma^2} \Big|_{\sigma=\sigma^{(g)}} &= \frac{2}{\sigma^{(g)}} (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=0}^{y=\infty} p_Q(y) [1 - P_Q(y)]^{\lambda-\mu-1} [P_Q(y)]^{\mu-1} \\ &\quad \times \frac{\partial}{\partial \sigma} P_Q(y | x^{(g)}, r^{(g)}, \sigma) dy \Big|_{\sigma=\sigma^{(g)}}. \end{aligned} \quad (\text{F.307})$$

The $P_Q(q)$ and $p_Q(q)$ approximations Equations (D.116) to (D.119) are used to proceed further. Two cases (being feasible with overwhelming probability and being infeasible with overwhelming probability) have been distinguished in the derivation of the P_Q and p_Q approximations. Therefore, those two cases are treated separately for Equation (F.307).

A. The Approximate SAR in the case that the probability of feasible offspring tends to 1

In order to treat Equation (F.307) further for the feasible case, $\frac{\partial}{\partial \sigma} P_{Q_{\text{feas}}}(y | x^{(g)}, r^{(g)}, \sigma)$ needs to be derived. Insertion of Equation (D.116) into Equation (F.307) and subsequent application of the chain rule results in

$$\frac{\partial}{\partial \sigma} P_{Q_{\text{feas}}}(y | x^{(g)}, r^{(g)}, \sigma) = \frac{\partial}{\partial \sigma} \Phi \left(\frac{y - x^{(g)}}{\sigma} \right) \quad (\text{F.308})$$

$$= \phi \left(\frac{y - x^{(g)}}{\sigma} \right) (-1) \left(\frac{y - x^{(g)}}{\sigma^2} \right) \quad (\text{F.309})$$

$$= -\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{y - x^{(g)}}{\sigma} \right)^2} \left(\frac{y - x^{(g)}}{\sigma^2} \right). \quad (\text{F.310})$$

Insertion of Equations (D.116), (D.118), and (F.310) into Equation (F.307) yields

$$\begin{aligned} \frac{\partial^2 f_{\text{feas}}}{\partial \sigma^2} \Big|_{\sigma=\sigma^{(g)}} &= \frac{2}{\sigma^{(g)}} (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=\bar{r}\sqrt{\xi}}^{y=\infty} \frac{1}{\sigma^{(g)}} \left(-\frac{y - x^{(g)}}{\sigma^{(g)}} \right) \frac{1}{(\sqrt{2\pi})^2 \sigma^{(g)}} e^{-\frac{1}{2} \left(\frac{y - x^{(g)}}{\sigma^{(g)}} \right)^2} \\ &\quad \times \left[1 - \Phi \left(\frac{y - x^{(g)}}{\sigma^{(g)}} \right) \right]^{\lambda - \mu - 1} \left[\Phi \left(\frac{y - x^{(g)}}{\sigma^{(g)}} \right) \right]^{\mu - 1} dy. \end{aligned} \quad (\text{F.311})$$

For solving this integral, $-t := \frac{y - x^{(g)}}{\sigma^{(g)}}$ is substituted. It implies $y = -\sigma^{(g)}t + x^{(g)}$ and $dy = -\sigma^{(g)} dt$. The lower bound follows with σ -normalization, assuming $N \rightarrow \infty$, $N \gg \sigma^{*(g)}$, and knowing that $x^{(g)} \geq \bar{r}\sqrt{\xi}$ holds for the feasible case. It reads

$$t_l = -\frac{\bar{r}\sqrt{\xi} - x^{(g)}}{\sigma^{(g)}} = -\frac{N}{\sigma^{*(g)}} \frac{\bar{r}\sqrt{\xi} - x^{(g)}}{r^{(g)}} \simeq \infty.$$

Similarly, the upper bound

$$t_u = \lim_{y \rightarrow \infty} -\frac{N}{\sigma^{*(g)} r^{(g)}} (y - x^{(g)}) = -\infty$$

follows. Hence, the transformed integral reads

$$\begin{aligned} \frac{\partial^2 f_{\text{feas}}}{\partial \sigma^2} \Big|_{\sigma=\sigma^{(g)}} &= -\frac{2}{\sigma^{(g)}} (\lambda - \mu) \binom{\lambda}{\mu} \int_{t=-\infty}^{t=\infty} t \frac{1}{(\sqrt{2\pi})^2 \sigma^{(g)}} e^{-\frac{1}{2} t^2} \\ &\quad \times [1 - \Phi(-t)]^{\lambda - \mu - 1} [\Phi(-t)]^{\mu - 1} dt. \end{aligned} \quad (\text{F.312})$$

This integral can further be rewritten by making use of the fact that taking the negative of an integral is equivalent to exchanging the lower and upper bounds. Further, the identity $\Phi(t) = 1 - \Phi(-t)$ is used. The resulting expression reads

$$\begin{aligned} \frac{\partial^2 f_{\text{feas}}}{\partial \sigma^2} \Big|_{\sigma=\sigma^{(g)}} &= \frac{2}{\sigma^{(g)}} (\lambda - \mu) \binom{\lambda}{\mu} \int_{t=-\infty}^{t=\infty} t \frac{1}{(\sqrt{2\pi})^2 \sigma^{(g)}} e^{-\frac{1}{2} t^2} \\ &\quad \times [\Phi(t)]^{\lambda - \mu - 1} [1 - \Phi(t)]^{\mu - 1} dt \end{aligned} \quad (\text{F.313})$$

$$= \frac{2}{\sigma^{(g)^2}} \underbrace{\frac{\lambda - \mu}{(\sqrt{2\pi})^{1+1}} \binom{\lambda}{\mu} \int_{t=-\infty}^{t=\infty} t e^{-\frac{1}{2} t^2} [\Phi(t)]^{\lambda - \mu - 1} [1 - \Phi(t)]^{\mu - 1} dt}_{= e_{\mu, \lambda}^{1,1}}. \quad (\text{F.314})$$

In Equation (F.314), one of the so-called generalized progress coefficients (see Equation (G.366)) appears. Therefore,

$$\frac{\partial^2 f_{\text{feas}}}{\partial \sigma^2} \Big|_{\sigma=\sigma^{(g)}} = \frac{2}{\sigma^{(g)^2}} e_{\mu, \lambda}^{1,1} \quad (\text{F.315})$$

follows. Insertion of Equation (F.315) into Equation (F.299) yields the SAR for the feasible case

$$\psi_{\text{feas}} \approx 0 + \frac{\tau^2}{2} + \frac{\tau^2}{2} \sigma^{(g)^2} \frac{2}{\sigma^{(g)^2}} e_{\mu, \lambda}^{1,1} = \frac{\tau^2}{2} \left(1 + 2e_{\mu, \lambda}^{1,1} \right) = \tau^2 \left(\frac{1}{2} + e_{\mu, \lambda}^{1,1} \right). \quad (\text{F.316})$$

B. The Approximate SAR in the case that the probability of feasible offspring tends to 0

In order to treat Equation (F.307) further for the infeasible case, $\frac{\partial}{\partial \sigma} P_{Q_{\text{infeas}}}(y | x^{(g)}, r^{(g)}, \sigma)$ needs to be derived. To this end, Equation (D.117) is simplified further using $\sqrt{1+x} \simeq 1 + \frac{x}{2} + O(x^2)$

$$\bar{r} \simeq \sqrt{1 + \frac{\sigma^{(g)^2} N}{r^{(g)^2}}} \simeq 1 + \frac{\sigma^{(g)^2} N}{2r^{(g)^2}} = 1 + \frac{\sigma^{(g)*2}}{2N} \quad (\text{F.317})$$

(which is a justified approximation if $\sigma^{(g)*2} \ll N$) and for $N \rightarrow \infty$

$$\sigma^{(g)} \simeq \sigma_r \quad (\text{F.318})$$

(see Equation (B.62)) yielding

$$P_{Q_{\text{infeas}}}(q | x^{(g)}, r^{(g)}, \sigma) \approx \Phi \left(\frac{\left(1 + \frac{1}{\xi}\right) q - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}} \left(1 + \frac{\sigma^2 N}{2r^{(g)2}}\right)}{\sigma \sqrt{1 + \frac{1}{\xi}}} \right) \quad (\text{F.319})$$

and

$$p_{Q_{\text{infeas}}}(q | x^{(g)}, r^{(g)}, \sigma) \quad (\text{F.320})$$

$$= \frac{d}{dq} P_{Q_{\text{infeas}}}(q) \quad (\text{F.321})$$

$$\approx \frac{1}{\sqrt{2\pi}} \frac{\sqrt{1 + \frac{1}{\xi}}}{\sigma} \exp \left[-\frac{1}{2} \left(\frac{\left(1 + \frac{1}{\xi}\right) q - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}} \left(1 + \frac{\sigma^2 N}{2r^{(g)2}}\right)}{\sigma \sqrt{1 + \frac{1}{\xi}}} \right)^2 \right]. \quad (\text{F.322})$$

Insertion of Equation (F.319) into Equation (F.307) and subsequent application of the chain rule results in

$$\frac{\partial}{\partial \sigma} P_{Q_{\text{infeas}}}(y | x^{(g)}, r^{(g)}, \sigma) \approx \frac{\partial}{\partial \sigma} \Phi \left(\frac{\left(1 + \frac{1}{\xi}\right) y - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}} \left(1 + \frac{\sigma^2 N}{2r^{(g)2}}\right)}{\sigma \sqrt{1 + \frac{1}{\xi}}} \right) \quad (\text{F.323})$$

$$= \frac{\partial}{\partial \sigma} \Phi \left(\frac{\left(1 + \frac{1}{\xi}\right) y - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}}}{\sigma \sqrt{1 + \frac{1}{\xi}}} - \frac{\sigma N}{2r^{(g)} \sqrt{\xi} \sqrt{1 + \frac{1}{\xi}}} \right) \quad (\text{F.324})$$

$$= \phi \left(\frac{\left(1 + \frac{1}{\xi}\right) y - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}}}{\sigma \sqrt{1 + \frac{1}{\xi}}} - \frac{\sigma N}{2r^{(g)} \sqrt{\xi} \sqrt{1 + \frac{1}{\xi}}} \right) \quad (\text{F.325})$$

$$\times \left(-\frac{\left(1 + \frac{1}{\xi}\right) y - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}}}{\sigma^2 \sqrt{1 + \frac{1}{\xi}}} - \frac{N}{2r^{(g)} \sqrt{\xi} \sqrt{1 + \frac{1}{\xi}}} \right) \\ = \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\left(1 + \frac{1}{\xi}\right) y - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}}}{\sigma \sqrt{1 + \frac{1}{\xi}}} - \frac{\sigma N}{2r^{(g)} \sqrt{\xi} \sqrt{1 + \frac{1}{\xi}}} \right)^2 \right] \\ \times \left(-\frac{\left(1 + \frac{1}{\xi}\right) y - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}}}{\sigma^2 \sqrt{1 + \frac{1}{\xi}}} - \frac{N}{2r^{(g)} \sqrt{\xi} \sqrt{1 + \frac{1}{\xi}}} \right). \quad (\text{F.326})$$

Insertion of Equations (F.319), (F.322), and (F.326) into Equation (F.307) yields

$$\begin{aligned}
\frac{\partial^2 f}{\partial \sigma^2} \Big|_{\sigma=\sigma^{(g)}} &= \frac{2}{\sigma^{(g)}} (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=0}^{y=\infty} \frac{1}{\sqrt{2\pi}} \frac{\sqrt{1 + \frac{1}{\xi}}}{\sigma^{(g)}} \\
&\times \exp \left[-\frac{1}{2} \left(\frac{\left(1 + \frac{1}{\xi}\right) y - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}} \left(1 + \frac{\sigma^{(g)2} N}{2r^{(g)2}}\right)}{\sigma^{(g)} \sqrt{1 + \frac{1}{\xi}}} \right)^2 \right] \\
&\times \left[1 - \Phi \left(\frac{\left(1 + \frac{1}{\xi}\right) y - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}} \left(1 + \frac{\sigma^{(g)2} N}{2r^{(g)2}}\right)}{\sigma^{(g)} \sqrt{1 + \frac{1}{\xi}}} \right) \right]^{\lambda - \mu - 1} \\
&\times \left[\Phi \left(\frac{\left(1 + \frac{1}{\xi}\right) y - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}} \left(1 + \frac{\sigma^{(g)2} N}{2r^{(g)2}}\right)}{\sigma^{(g)} \sqrt{1 + \frac{1}{\xi}}} \right) \right]^{\mu - 1} \\
&\times \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\left(1 + \frac{1}{\xi}\right) y - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}} - \frac{\sigma^{(g)} N}{2r^{(g)} \sqrt{\xi} \sqrt{1 + \frac{1}{\xi}}} \right)^2 \right] \\
&\quad = \exp \left[-\frac{1}{2} \left(\frac{\left(1 + \frac{1}{\xi}\right) y - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}} \left(1 + \frac{\sigma^{(g)2} N}{2r^{(g)2}}\right)}{\sigma^{(g)} \sqrt{1 + \frac{1}{\xi}}} \right)^2 \right] \\
&\quad \times \left(-\frac{\left(1 + \frac{1}{\xi}\right) y - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}}}{\sigma^{(g)2} \sqrt{1 + \frac{1}{\xi}}} - \frac{N}{2r^{(g)} \sqrt{\xi} \sqrt{1 + \frac{1}{\xi}}} \right) dy.
\end{aligned} \tag{F.327}$$

For solving this integral,

$$\frac{\left(1 + \frac{1}{\xi}\right) y - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}} \left(1 + \frac{\sigma^{(g)2} N}{2r^{(g)2}}\right)}{\sigma^{(g)} \sqrt{1 + \frac{1}{\xi}}} := -t \tag{F.328}$$

is substituted. It further implies

$$\frac{dt}{dy} = -\frac{1 + 1/\xi}{\sigma^{(g)} \sqrt{1 + 1/\xi}} \tag{F.329}$$

$$dy = -\frac{\sigma^{(g)}}{\sqrt{1 + 1/\xi}} dt. \tag{F.330}$$

Expressing t with normalized $\sigma^{(g)} = \frac{r^{(g)} \sigma^{(g)*}}{N}$ one obtains

$$t = -N \frac{\left(1 + \frac{1}{\xi}\right) y - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}} \left(1 + \frac{\sigma^{(g)*2} r^{(g)2} N}{2r^{(g)2} N^2}\right)}{\sigma^{(g)*} r^{(g)} \sqrt{1 + \frac{1}{\xi}}} \tag{F.331}$$

$$= -N \frac{\left(1 + \frac{1}{\xi}\right) y - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}} \left(1 + \frac{\sigma^{(g)*2}}{2N}\right)}{\sigma^{(g)*} r^{(g)} \sqrt{1 + \frac{1}{\xi}}}. \tag{F.332}$$

For $N \rightarrow \infty$, the integration bounds after substitution follow as $t_l = \infty$ and $t_u = -\infty$. The transformed integral reads

$$\left. \frac{\partial^2 f_{\text{infeas}}}{\partial \sigma^2} \right|_{\sigma=\sigma^{(g)}} = -\frac{2}{\sigma^{(g)}}(\lambda - \mu) \binom{\lambda}{\mu} \int_{t=-\infty}^{t=\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} [1 - \Phi(-t)]^{\lambda-\mu-1} [\Phi(-t)]^{\mu-1} \times \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \left(\frac{t}{\sigma^{(g)}} - \frac{2N}{2r^{(g)}\sqrt{\xi}\sqrt{1+\frac{1}{\xi}}} \right) dt \quad (\text{F.333})$$

$$= \frac{2}{\sigma^{(g)^2}} \underbrace{\frac{\lambda - \mu}{(\sqrt{2\pi})^{1+1}} \binom{\lambda}{\mu} \int_{t=-\infty}^{t=\infty} t e^{-\frac{1+1}{2}t^2} [\Phi(t)]^{\lambda-\mu-1} [1 - \Phi(t)]^{\mu-1} dt}_{=e_{\mu,\lambda}^{1,1}} - \frac{2}{\sigma^{(g)}} \frac{N}{r^{(g)}\sqrt{\xi}\sqrt{1+\frac{1}{\xi}}} \quad (\text{F.334})$$

$$\times \underbrace{\frac{\lambda - \mu}{(\sqrt{2\pi})^{1+1}} \binom{\lambda}{\mu} \int_{t=-\infty}^{t=\infty} e^{-\frac{1+1}{2}t^2} [\Phi(t)]^{\lambda-\mu-1} [1 - \Phi(t)]^{\mu-1} dt}_{=c_{\mu/\mu,\lambda}} = \frac{2}{\sigma^{(g)^2}} e_{\mu,\lambda}^{1,1} - \frac{2}{\sigma^{(g)}} \frac{N}{r^{(g)}\sqrt{\xi}\sqrt{1+\frac{1}{\xi}}} c_{\mu/\mu,\lambda}. \quad (\text{F.335})$$

From Equation (F.333) to Equation (F.334) the fact that taking the negative of an integral can be expressed by exchanging the lower and upper bounds, and the identity $\Phi(t) = 1 - \Phi(-t)$ have been used. Two of the generalized progress coefficients (see Equation (G.366)) appear and have been inserted into Equation (F.335). Insertion of Equation (F.335) into Equation (F.299) yields the SAR for the infeasible case

$$\psi_{\text{infeas}} \approx 0 + \frac{\tau^2}{2} + \frac{\tau^2}{2} \sigma^{(g)^2} \left(\frac{2}{\sigma^{(g)^2}} e_{\mu,\lambda}^{1,1} - \frac{2}{\sigma^{(g)}} \frac{N}{r^{(g)}\sqrt{\xi}\sqrt{1+\frac{1}{\xi}}} c_{\mu/\mu,\lambda} \right) \quad (\text{F.336})$$

$$= \frac{\tau^2}{2} \left(1 + 2e_{\mu,\lambda}^{1,1} - \frac{2N\sigma^{(g)}c_{\mu/\mu,\lambda}}{r^{(g)}\sqrt{1+\xi}} \right) \quad (\text{F.337})$$

$$= \tau^2 \left(\frac{1}{2} + e_{\mu,\lambda}^{1,1} - \frac{\sigma^{(g)*}c_{\mu/\mu,\lambda}}{\sqrt{1+\xi}} \right) \quad (\text{F.338})$$

C. The Approximate SAR Progress Rate - Combination Using the Single Offspring Feasibility Probability

Both cases are combined into

$$\psi \approx P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \psi_{\text{feas}} + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] \psi_{\text{infeas}} \quad (\text{F.339})$$

$$= P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \left[\tau^2 \left(\frac{1}{2} + e_{\mu,\lambda}^{1,1} \right) \right] + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] \left[\tau^2 \left(\frac{1}{2} + e_{\mu,\lambda}^{1,1} - \frac{\sigma^{(g)*}c_{\mu/\mu,\lambda}}{\sqrt{1+\xi}} \right) \right] \quad (\text{F.340})$$

$$= \tau^2 \left[\left(\frac{1}{2} + e_{\mu,\lambda}^{1,1} \right) \right] - [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] \frac{\sigma^{(g)*}c_{\mu/\mu,\lambda}}{\sqrt{1+\xi}}. \quad (\text{F.341})$$

APPENDIX G

DERIVATION OF $E[q_{r m; \lambda}^2]$ AND $E[q_{m; \lambda}^2]$

$E[q_{r m; \lambda}^2]$ can be derived similarly to the $(1, \lambda)$ case as done for deriving $E[q_{r 1; \lambda}]$ (see [13, Sec. 3.1.2.2, pp. 49-56]):

$$E[q_{r m; \lambda}^2] := E[q_{r m; \lambda}^2 | x^{(g)}, r^{(g)}, \sigma^{(g)}] = \int_{q_r=0}^{q_r=\infty} q_r^2 p_{q_{r m; \lambda}}(q_r) dq_r. \quad (\text{G.342})$$

The density

$$p_{q_{r m; \lambda}}(q_r) := p_{q_{r m; \lambda}}(q_r | x^{(g)}, r^{(g)}, \sigma^{(g)}) \quad (\text{G.343})$$

indicates the probability density function of the m -th best (offspring with m -th smallest q value) offspring's q_r value. The derivation is analogous to the $(1, \lambda)$ case with the additional consideration for the order statistics (m -th best instead of best as done for the multi-recombinative x progress rate (Equation (D.82))). It yields

$$p_{q_{r m; \lambda}}(q_r) = \frac{\lambda!}{(\lambda - m)!(m - 1)!} \int_{q=0}^{q=\infty} p_{Q, Q_r}(q, q_r) [1 - P_Q(q)]^{\lambda - m} [P_Q(q)]^{m - 1} dq. \quad (\text{G.344})$$

Insertion of Equation (G.344) into Equation (G.342) results in

$$E[q_{r m; \lambda}^2] = \int_{q_r=0}^{q_r=\infty} q_r^2 \frac{\lambda!}{(\lambda - m)!(m - 1)!} \int_{q=0}^{q=\infty} p_{Q, Q_r}(q, q_r) \times [1 - P_Q(q)]^{\lambda - m} [P_Q(q)]^{m - 1} dq dq_r. \quad (\text{G.345})$$

By changing the order of integration, Equation (G.345) can be rewritten to

$$E[q_{r m; \lambda}^2] = \frac{\lambda!}{(\lambda - m)!(m - 1)!} \int_{q=0}^{q=\infty} \underbrace{\left[\int_{q_r=0}^{q_r=\infty} q_r^2 p_{Q, Q_r}(q, q_r) dq_r \right]}_{=: I_2(q)} \times [1 - P_Q(q)]^{\lambda - m} [P_Q(q)]^{m - 1} dq. \quad (\text{G.346})$$

Note that $I_2(q)$ is very similar to $I(q)$ from [13, Sec. 3.1.2.2, pp. 49-56]. It only differs in that $I_2(q)$ contains q_r^2 and $I(q)$ contains q_r . As already done for $I(q)$, $I_2(q)$ is expressed in terms of the values before projection. Because q_r^2 represents the squared distance from the cone's axis after projection, it only takes values in the interval $[0, (x/\sqrt{\xi})^2]$ for $x \in \mathbb{R}, x \geq 0$. This is because all individuals that happen to be generated outside the cone are projected onto the cone boundary. The integral in $I_2(q)$ is therefore split into two summands. The first summand represents the case of immediately feasible offspring individuals. The second summand represents the case of projected offspring individuals. All points that lie on the projection line for a given value of q are projected to $q/\sqrt{\xi}$. This writes

$$I_2(q) dq = \int_{\tilde{r}=0}^{\tilde{r}=q/\sqrt{\xi}} \tilde{r}^2 p_{1;1}(q, \tilde{r}) d\tilde{r} dq + \left(\frac{q}{\sqrt{\xi}} \right)^2 \underbrace{\left(p_Q(q) dq - \int_{\tilde{r}=0}^{\tilde{r}=q/\sqrt{\xi}} p_{1;1}(q, \tilde{r}) d\tilde{r} dq \right)}_{= dP_{\text{line}}} \quad (\text{G.347})$$

With the same arguments and assumptions leading to [13, Eq. (3.206), page 56], the approximation

$$E[q_{r m; \lambda_{\text{infeas}}}^2] \approx \frac{1}{\xi} E[q_{m; \lambda_{\text{infeas}}}^2] \quad (\text{G.348})$$

can be derived for the case that the m -th best offspring is infeasible with high probability. For the case that the m -th best offspring is feasible almost surely, the complete probability mass lies inside the cone. Consequently, the second summand in Equation (G.347) vanishes. Additionally, the integral in the first summand yields the second (non-central) moment $\bar{r}^2 + \sigma_r^2$ because the bounds indicate the integration over the whole feasible region for the given area dq . In the feasible case, $I_2(q)$ therefore reads

$$I_{2\text{feas}}(q) = p_x(q)(\bar{r}^2 + \sigma_r^2). \quad (\text{G.349})$$

By insertion of $I_{2\text{feas}}(q)$ as $I_2(q)$ into Equation (G.346), use of $P_{Q\text{feas}}$ from Equation (D.116), and use of $p_x(x) = \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{x-x^{(g)}}{\sigma^{(g)}}\right)^2\right]$ one obtains

$$\begin{aligned} \mathbb{E}[q_{rm;\lambda\text{feas}}^2] &\approx (\bar{r}^2 + \sigma_r^2) \frac{\lambda!}{(\lambda-m)!(m-1)!} \int_{q=0}^{q=\infty} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)^2} \\ &\quad \times \left[1 - \Phi\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)\right]^{\lambda-m} \left[\Phi\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)\right]^{m-1} dq \end{aligned} \quad (\text{G.350})$$

$$= \bar{r}^2 + \sigma_r^2. \quad (\text{G.351})$$

In the last step the fact that the integral over the whole probability density function of the m -th order statistic yields 1 has been used. Both can be combined with the (approximate) feasibility probability

$$\mathbb{E}[q_{rm;\lambda}^2] \approx P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \mathbb{E}[q_{rm;\lambda\text{feas}}^2] + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] \mathbb{E}[q_{rm;\lambda\text{infeas}}^2]. \quad (\text{G.352})$$

Because $\mathbb{E}[q_{rm;\lambda\text{infeas}}^2] \approx \frac{1}{\xi} \mathbb{E}[q_{m;\lambda\text{infeas}}^2]$ (see Equation (G.348)) holds, $\mathbb{E}[q_{m;\lambda\text{infeas}}^2]$ needs to be determined next. For Equation (G.348), $\mathbb{E}[q_{m;\lambda\text{infeas}}^2]$ is derived and finally combined with an approximation for the feasibility probability. $\mathbb{E}[q_{m;\lambda}^2]$ writes⁴

$$\mathbb{E}[q_{m;\lambda}^2] = \frac{\lambda!}{(\lambda-m)!(m-1)!} \int_{q=0}^{q=\infty} q^2 p_Q(q) [1 - P_Q(q)]^{\lambda-m} [P_Q(q)]^{m-1} dq. \quad (\text{G.353})$$

Use of Equations (D.116) and (D.118) yields for the feasible case

$$\begin{aligned} \mathbb{E}[q_{m;\lambda\text{feas}}^2] &\approx \frac{\lambda!}{(\lambda-m)!(m-1)!} \int_{q=\bar{r}\sqrt{\xi}}^{q=\infty} q^2 \frac{1}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)^2} \\ &\quad \times \left[1 - \Phi\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)\right]^{\lambda-m} \left[\Phi\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)\right]^{m-1} dq. \end{aligned} \quad (\text{G.354})$$

The substitution

$$\frac{q-x^{(g)}}{\sigma^{(g)}} := -t \quad (\text{G.355})$$

is used. It follows that

$$q = -t\sigma^{(g)} + x^{(g)} \quad (\text{G.356})$$

and

$$dq = -\sigma^{(g)} dt. \quad (\text{G.357})$$

Using normalized quantities, t can be expressed as

$$t = \frac{-N(q-x^{(g)})}{\sigma^{(g)*} r^{(g)}}. \quad (\text{G.358})$$

Assuming $N \rightarrow \infty$ yields for the upper bound $t_u = -\infty$. For the lower bound it follows that

$$t_l = \frac{-N(\bar{r}\sqrt{\xi} - x^{(g)})}{\sigma^{(g)*} r^{(g)}} \quad (\text{G.359})$$

$$= -\frac{N}{\sigma^{(g)*}} \left(\frac{\bar{r}\sqrt{\xi}}{r^{(g)}} - \frac{x^{(g)}}{r^{(g)}} \right). \quad (\text{G.360})$$

Because $\bar{r} \simeq r^{(g)} \sqrt{1 + \frac{\sigma^{(g)*2}}{N}}$ (Equation (B.62)), for $N \rightarrow \infty$ and $\sigma^{(g)*2} \ll N$, $\bar{r}\sqrt{\xi} \simeq r^{(g)} \sqrt{\xi} \leq x^{(g)}$ follows. The last inequality follows from the fact that the parental individual is feasible (ensured by projection). Hence, it follows that for $N \rightarrow \infty$

$$t_l = \frac{N}{\sigma^{(g)*}} \underbrace{\left(\frac{x^{(g)}}{r^{(g)}} - \frac{\bar{r}\sqrt{\xi}}{r^{(g)}} \right)}_{\geq 0} \quad (\text{G.361})$$

$$\simeq \infty \quad (\text{G.362})$$

⁴Note that here only $\mathbb{E}[q_{m;\lambda\text{infeas}}^2]$ is necessary. In order to have an expression for the whole $\mathbb{E}[q_{m;\lambda}^2]$, the derivations for the feasible and infeasible cases are presented here for completeness.

holds. Applying the substitution results in

$$\begin{aligned} \mathbb{E}[q_{m;\lambda_{\text{feas}}}^2] &\approx -\frac{\lambda!}{(\lambda-m)!(m-1)!} \frac{1}{\sqrt{2\pi}} \int_{t=-\infty}^{t=\infty} (-\sigma^{(g)}t + x^{(g)})^2 e^{-\frac{1}{2}t^2} \\ &\quad \times [1 - \Phi(-t)]^{\lambda-m} [\Phi(-t)]^{m-1} dt \end{aligned} \quad (\text{G.363})$$

$$\begin{aligned} &= \frac{\lambda!}{(\lambda-m)!(m-1)!} \frac{1}{\sqrt{2\pi}} \int_{t=-\infty}^{t=\infty} \left(\sigma^{(g)2}t^2 - 2\sigma^{(g)}tx^{(g)} + x^{(g)2} \right) e^{-\frac{1}{2}t^2} \\ &\quad \times [\Phi(t)]^{\lambda-m} [1 - \Phi(t)]^{m-1} dt \end{aligned} \quad (\text{G.364})$$

$$= \sigma^{(g)2} e_{(m-1),\lambda}^{0,2} - 2\sigma^{(g)}x^{(g)} e_{(m-1),\lambda}^{0,1} + x^{(g)2} \quad (\text{G.365})$$

where the generalized progress coefficients have been used. Those are defined in [14, Eq. (5.112), p. 172]. They are defined because those integrals cannot be solved analytically for large λ and large μ . The generalized progress coefficients are defined as

$$e_{\mu,\lambda}^{\alpha,\beta} := \frac{\lambda - \mu}{(\sqrt{2\pi})^{\alpha+1}} \binom{\lambda}{\mu} \int_{t=-\infty}^{t=\infty} t^\beta e^{-\frac{\alpha+1}{2}t^2} [\Phi(t)]^{\lambda-\mu-1} [1 - \Phi(t)]^{\mu-\alpha} dt. \quad (\text{G.366})$$

Use of Equations (D.117) and (D.119) yields for the infeasible case

$$\begin{aligned} \mathbb{E}[q_{m;\lambda_{\text{infeas}}}^2] &\approx \frac{\lambda!}{(\lambda-m)!(m-1)!} \int_{q=0}^{q=\bar{r}\sqrt{\xi}} q^2 \left(\frac{(1+1/\xi)}{\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi}} \right) \frac{1}{\sqrt{2\pi}} \\ &\quad \times \exp \left[-\frac{1}{2} \left(\frac{(1+1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi}} \right)^2 \right] \\ &\quad \times \left[1 - \Phi \left(\frac{(1+1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi}} \right) \right]^{\lambda-m} \\ &\quad \times \left[\Phi \left(\frac{(1+1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi}} \right) \right]^{m-1} dq. \end{aligned} \quad (\text{G.367})$$

The substitution

$$\frac{(1+1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi}} := -t \quad (\text{G.368})$$

is used. It follows that

$$q = \frac{1}{(1+1/\xi)} \left(-\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi} t + x^{(g)} + \bar{r}/\sqrt{\xi} \right) \quad (\text{G.369})$$

and

$$dq = -\frac{\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi}}{(1+1/\xi)} dt. \quad (\text{G.370})$$

Using the normalized $\sigma^{(g)*}$, $\sigma_r \simeq \sigma^{(g)}$ for $N \rightarrow \infty$, and $\sigma^{(g)*} \ll N$ (derived from Equation (B.62)), t can be expressed as

$$t = -\frac{(1+1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\frac{\sigma^{(g)*2} r^{(g)2}}{N^2} + \frac{\sigma^{(g)*2} r^{(g)2}}{N^2 \xi}}} \quad (\text{G.371})$$

$$= -\frac{(1+1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\frac{1}{N\sqrt{\xi}} \sqrt{\xi \sigma^{(g)*2} r^{(g)2} + \sigma^{(g)*2} r^{(g)2}}} \quad (\text{G.372})$$

$$= -N\sqrt{\xi} \left[\frac{(1+1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\xi \sigma^{(g)*2} r^{(g)2} + \sigma^{(g)*2} r^{(g)2}}} \right] \quad (\text{G.373})$$

$$= -N\sqrt{\xi} \left[\frac{(1+1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sigma^{(g)*} r^{(g)} \sqrt{\xi + 1}} \right]. \quad (\text{G.374})$$

The lower bound in the transformed integral therefore follows assuming $\xi \gg 1$, $N \rightarrow \infty$, using $\infty > \bar{r} \simeq r^{(g)} \geq 0$, and knowing that $0 \leq x^{(g)} < \infty$ as

$$t_l = -N\sqrt{\xi} \left[\frac{(1 + 1/\xi)0 - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sigma^{(g)*} r^{(g)} \sqrt{\xi + 1}} \right] \quad (\text{G.375})$$

$$= -N\sqrt{\xi} \left[\frac{-x^{(g)} - \bar{r}/\sqrt{\xi}}{\sigma^{(g)*} r^{(g)} \sqrt{\xi + 1}} \right] \quad (\text{G.376})$$

$$\simeq -N \underbrace{\left[\frac{-x^{(g)} - \bar{r}/\sqrt{\xi}}{\sigma^{(g)*} r^{(g)}} \right]}_{\leq 0} \quad (\text{G.377})$$

$$\simeq \infty. \quad (\text{G.378})$$

Similarly, the upper bound follows with the same assumptions and using the fact that the case under consideration is the infeasible case, i.e., $x^{(g)} \leq \bar{r}\sqrt{\xi}$

$$t_u = -N\sqrt{\xi} \left[\frac{(1 + 1/\xi)\bar{r}\sqrt{\xi} - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sigma^{(g)*} r^{(g)} \sqrt{\xi + 1}} \right] \quad (\text{G.379})$$

$$= -N \underbrace{\left[\frac{\bar{r}\sqrt{\xi} - x^{(g)}}{\sigma^{(g)*} r^{(g)}} \right]}_{\geq 0} \quad (\text{G.380})$$

$$\simeq -\infty. \quad (\text{G.381})$$

Actually applying the substitution leads to

$$\begin{aligned} \mathbb{E}[q_{m;\lambda_{\text{infeas}}}^2] &\approx -\frac{\lambda!}{(\lambda - m)!(m - 1)!} \\ &\times \int_{t=-\infty}^{t=-\infty} \left[\frac{1}{(1 + 1/\xi)} \left(-\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi} t + x^{(g)} + \bar{r}/\sqrt{\xi} \right) \right]^2 \\ &\times \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} [1 - \Phi(-t)]^{\lambda-m} [\Phi(-t)]^{m-1} dt \\ &= \frac{\lambda!}{(\lambda - m)!(m - 1)!} \end{aligned} \quad (\text{G.382})$$

$$\begin{aligned} &\times \int_{t=-\infty}^{t=\infty} \left[\frac{1}{(1 + 1/\xi)} \left(-\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi} t + x^{(g)} + \bar{r}/\sqrt{\xi} \right) \right]^2 \\ &\times \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} [\Phi(t)]^{\lambda-m} [1 - \Phi(t)]^{m-1} dt \\ &= \frac{(\sigma^{(g)2} + \sigma_r^2/\xi)}{(1 + 1/\xi)^2} e_{(m-1),\lambda}^{0,2} - \frac{\sqrt{\sigma^{(g)2} + \sigma_r^2/\xi}}{(1 + 1/\xi)^2} 2(x^{(g)} + \bar{r}/\sqrt{\xi}) e_{(m-1),\lambda}^{0,1} \\ &\quad + \frac{(x^{(g)} + \bar{r}/\sqrt{\xi})^2}{(1 + 1/\xi)^2}. \end{aligned} \quad (\text{G.383})$$

Both can be combined with the (approximate) feasibility probability

$$\mathbb{E}[q_{m;\lambda}^2] \approx P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \mathbb{E}[q_{m;\lambda_{\text{feas}}}^2] + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] \mathbb{E}[q_{m;\lambda_{\text{infeas}}}^2]. \quad (\text{G.384})$$

$$\mathbb{E}[q_{m;\lambda}^2] \approx P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \mathbb{E}[q_{m;\lambda_{\text{feas}}}^2] + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] \mathbb{E}[q_{m;\lambda_{\text{infeas}}}^2]. \quad (\text{G.385})$$

APPENDIX H
DERIVATION OF $E[\langle q^2 \rangle]$

$$E[\langle q^2 \rangle] = E\left[\frac{1}{\mu} \sum_{m=1}^{\mu} q_{m;\lambda}^2\right] = \frac{1}{\mu} \sum_{m=1}^{\mu} E[q_{m;\lambda}^2] \quad (\text{H.386})$$

follows by expanding the notation for the centroid computation and linearity of expectation. An approximation for $E[q_{m;\lambda}^2]$ has been derived in Appendix G as Equation (G.385) together with Equation (G.365) and Equation (G.384). Using Equations (G.365), (G.384), and (G.385),

$$\frac{1}{\mu} \sum_{m=1}^{\mu} E[q_{m;\lambda}^2] \approx \frac{1}{\mu} \sum_{m=1}^{\mu} \left[P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) E[q_{m;\lambda}^2]_{\text{feas}} \right. \quad (\text{H.387})$$

$$\begin{aligned} & \left. + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] E[q_{m;\lambda}^2]_{\text{infeas}} \right] \\ & \approx P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \left(\sigma^{(g)^2} \left[\frac{1}{\mu} \sum_{m=1}^{\mu} e_{(m-1),\lambda}^{0,2} \right] \right. \\ & \quad \left. - 2\sigma^{(g)} x^{(g)} \left[\frac{1}{\mu} \sum_{m=1}^{\mu} e_{(m-1),\lambda}^{0,1} \right] + x^{(g)^2} \right) \\ & + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] \left(\frac{(\sigma^{(g)^2} + \sigma_r^2/\xi)}{(1 + 1/\xi)^2} \left[\frac{1}{\mu} \sum_{m=1}^{\mu} e_{(m-1),\lambda}^{0,2} \right] \right. \\ & \quad \left. - \frac{\sqrt{\sigma^{(g)^2} + \sigma_r^2/\xi}}{(1 + 1/\xi)^2} 2(x^{(g)} + \bar{r}/\sqrt{\xi}) \left[\frac{1}{\mu} \sum_{m=1}^{\mu} e_{(m-1),\lambda}^{0,1} \right] \right. \\ & \quad \left. + \frac{(x^{(g)} + \bar{r}/\sqrt{\xi})^2}{(1 + 1/\xi)^2} \right) \end{aligned} \quad (\text{H.388})$$

can be written.

An expression for $\frac{1}{\mu} \sum_{m=1}^{\mu} e_{(m-1),\lambda}^{0,2}$ can be derived using the definition of the generalized progress coefficients (Equation (G.366)). By this definition,

$$\begin{aligned} \frac{1}{\mu} \sum_{m=1}^{\mu} e_{(m-1),\lambda}^{0,2} &= \frac{1}{\mu} \sum_{m=1}^{\mu} \left[\frac{\lambda - (m-1)}{\sqrt{2\pi}} \binom{\lambda}{m-1} \right. \\ & \quad \left. \times \int_{t=-\infty}^{t=\infty} t^2 e^{-\frac{1}{2}t^2} [\Phi(t)]^{\lambda-m} [1 - \Phi(t)]^{m-1} dt \right] \end{aligned} \quad (\text{H.389})$$

follows. Rewriting this results in

$$\frac{1}{\mu} \sum_{m=1}^{\mu} e_{(m-1),\lambda}^{0,2} = \frac{1}{\mu} \frac{\lambda!}{\sqrt{2\pi}} \int_{t=-\infty}^{t=\infty} t^2 e^{-\frac{1}{2}t^2} \sum_{m=1}^{\mu} \frac{[\Phi(t)]^{\lambda-m} [1 - \Phi(t)]^{m-1}}{(\lambda - m)!(m - 1)!} dt. \quad (\text{H.390})$$

Substituting $s := -t$ and using the identity $\Phi(-s) = 1 - \Phi(s)$ yields further

$$\frac{1}{\mu} \sum_{m=1}^{\mu} e_{(m-1),\lambda}^{0,2} = \frac{1}{\mu} \frac{\lambda!}{\sqrt{2\pi}} \int_{s=-\infty}^{s=\infty} s^2 e^{-\frac{1}{2}s^2} \sum_{m=1}^{\mu} \frac{[1 - \Phi(s)]^{\lambda-m} [\Phi(s)]^{m-1}}{(\lambda - m)!(m - 1)!} ds. \quad (\text{H.391})$$

Now, with the same arguments that were used to get from Equation (D.84) to Equation (D.89),

$$\frac{1}{\mu} \sum_{m=1}^{\mu} e_{(m-1),\lambda}^{0,2} = (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=-\infty}^{y=\infty} \phi(y) [1 - \Phi(y)]^{\lambda-\mu-1} [\Phi(y)]^{\mu-1} \int_{s=-\infty}^{s=y} s^2 \phi(s) ds dy \quad (\text{H.392})$$

can be derived. Use of the first identity of [14, Eq. (A.17), p. 331]

$$\int_{-\infty}^x t^2 e^{-\frac{1}{2}t^2} dt = \sqrt{2\pi} \Phi(x) - x e^{-\frac{1}{2}x^2} \quad (\text{H.393})$$

results in $\Phi(y) - y\phi(y)$ for the inner integral. Consequently,

$$\frac{1}{\mu} \sum_{m=1}^{\mu} e_{(m-1),\lambda}^{0,2} = (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=-\infty}^{y=\infty} \phi(y) [1 - \Phi(y)]^{\lambda-\mu-1} [\Phi(y)]^{\mu-1} (\Phi(y) - y\phi(y)) dy \quad (\text{H.394})$$

$$\begin{aligned} &= (\lambda - \mu) \binom{\lambda}{\mu} \int_{y=-\infty}^{y=\infty} \phi(y) [1 - \Phi(y)]^{\lambda-\mu-1} [\Phi(y)]^{\mu} dy \\ &\quad - \frac{(\lambda - \mu)}{2\pi} \binom{\lambda}{\mu} \int_{y=-\infty}^{y=\infty} y e^{-y^2} [1 - \Phi(y)]^{\lambda-\mu-1} [\Phi(y)]^{\mu-1} dy \end{aligned} \quad (\text{H.395})$$

follows. For the first summand, the same arguments leading to Equation (D.130) reveal that it is the integration over the density of the $(\mu + 1)$ -th order statistic of independently standard normally distributed variables. Hence, it integrates to 1. By substituting $s := -y$, the second summand is recognized as one of the generalized progress coefficients (Equation (G.366)). Thus, one gets

$$\frac{1}{\mu} \sum_{m=1}^{\mu} e_{(m-1),\lambda}^{0,2} = 1 + e_{\mu,\lambda}^{1,1}. \quad (\text{H.396})$$

Analogously,

$$\frac{1}{\mu} \sum_{m=1}^{\mu} e_{(m-1),\lambda}^{0,1} = e_{\mu,\lambda}^{1,0} = c_{\mu/\mu,\lambda} \quad (\text{H.397})$$

can be derived. Reinsertion of these results into Equation (H.388) yields

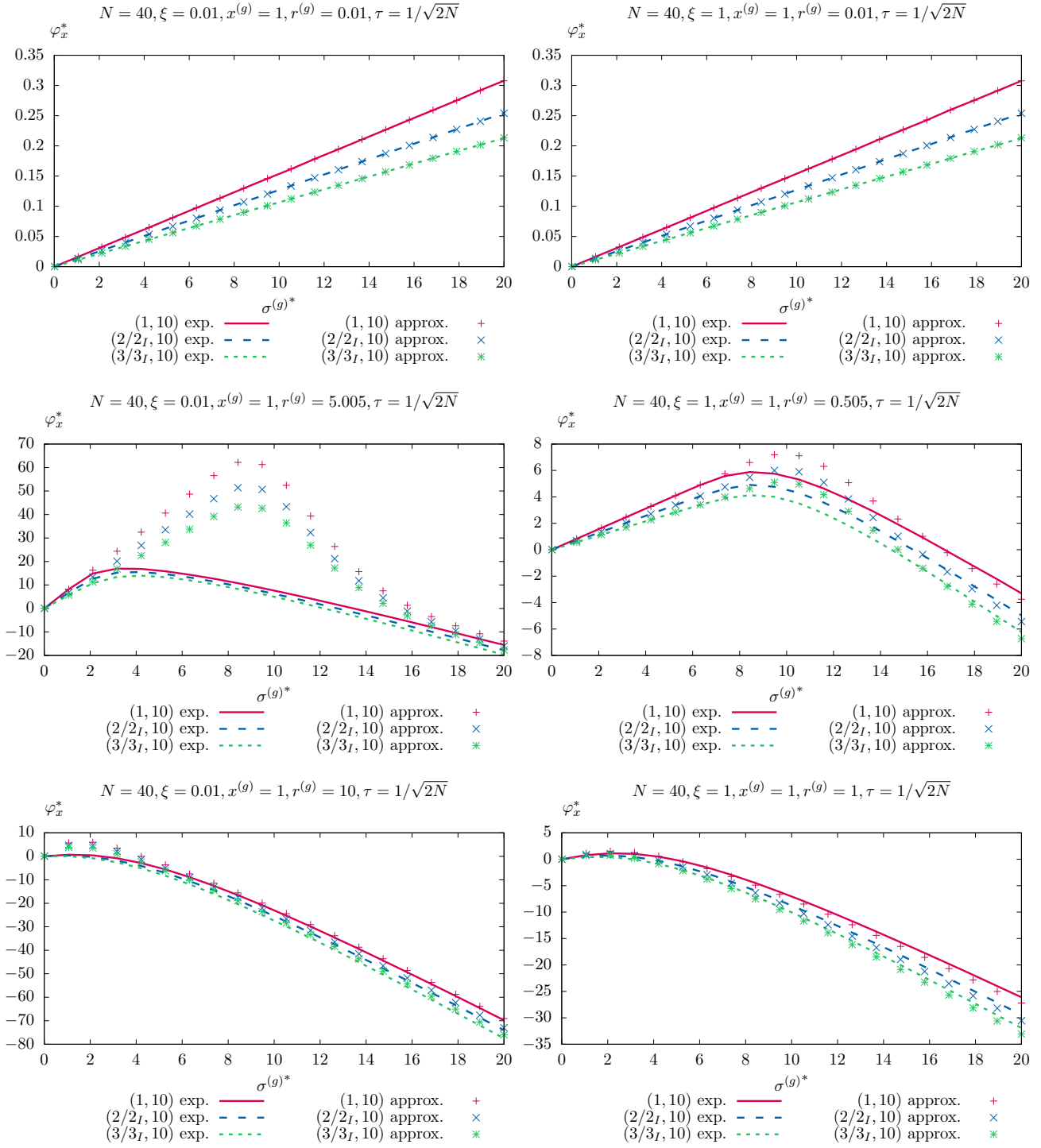
$$\begin{aligned} \frac{1}{\mu} \sum_{m=1}^{\mu} \mathbb{E}[q_{m;\lambda}^2] &\approx P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \left(\sigma^{(g)^2} \left[1 + e_{\mu,\lambda}^{1,1} \right] - 2\sigma^{(g)} x^{(g)} c_{\mu/\mu,\lambda} + x^{(g)^2} \right) \\ &\quad + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] \left(\frac{(\sigma^{(g)})^2 + \sigma_r^2/\xi}{(1 + 1/\xi)^2} \left[1 + e_{\mu,\lambda}^{1,1} \right] \right. \\ &\quad \left. - \frac{\sqrt{\sigma^{(g)^2 + \sigma_r^2/\xi}}}{(1 + 1/\xi)^2} 2(x^{(g)} + \bar{r}/\sqrt{\xi}) c_{\mu/\mu,\lambda} + \frac{(x^{(g)} + \bar{r}/\sqrt{\xi})^2}{(1 + 1/\xi)^2} \right) \end{aligned} \quad (\text{H.398})$$

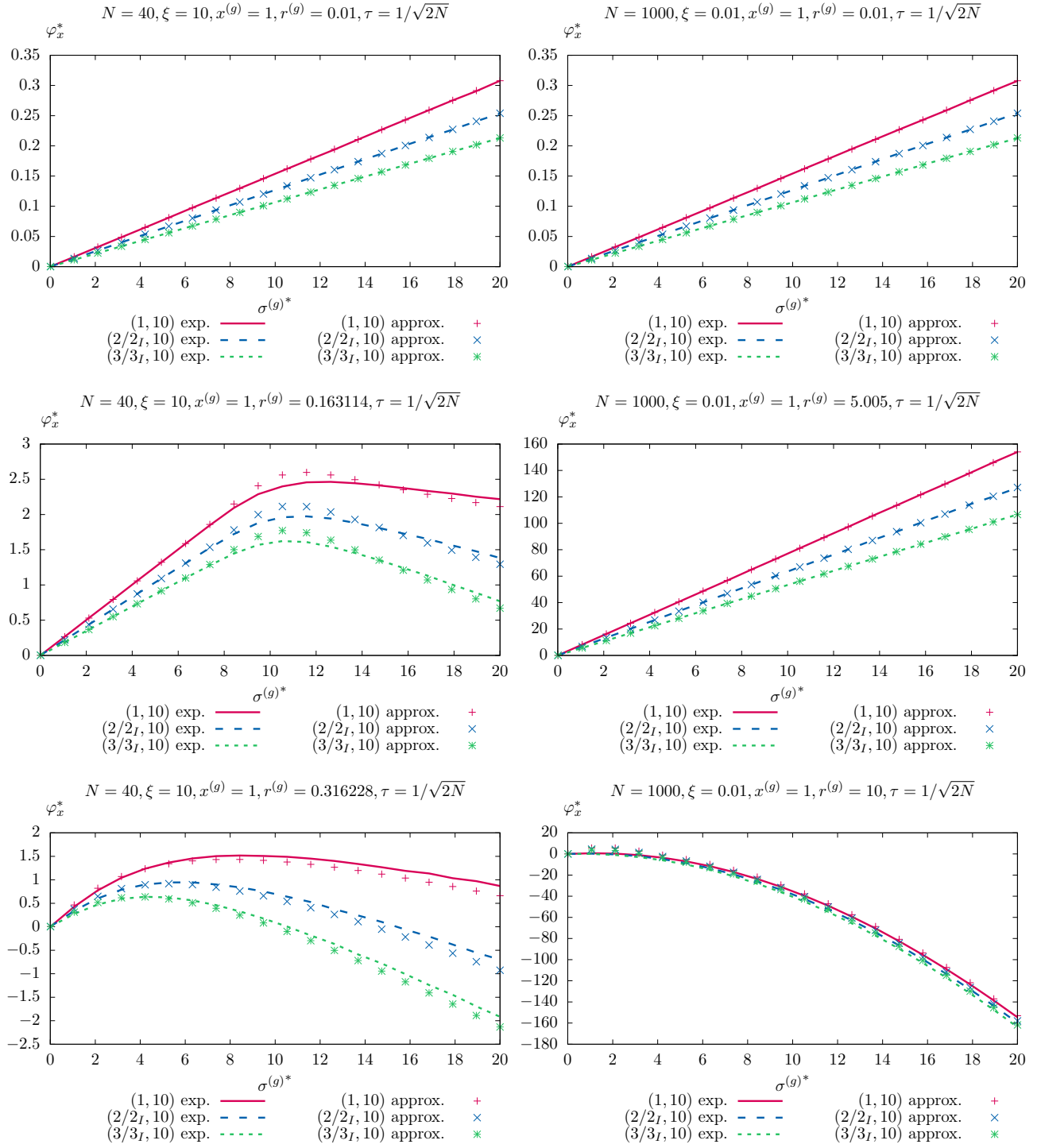
APPENDIX I

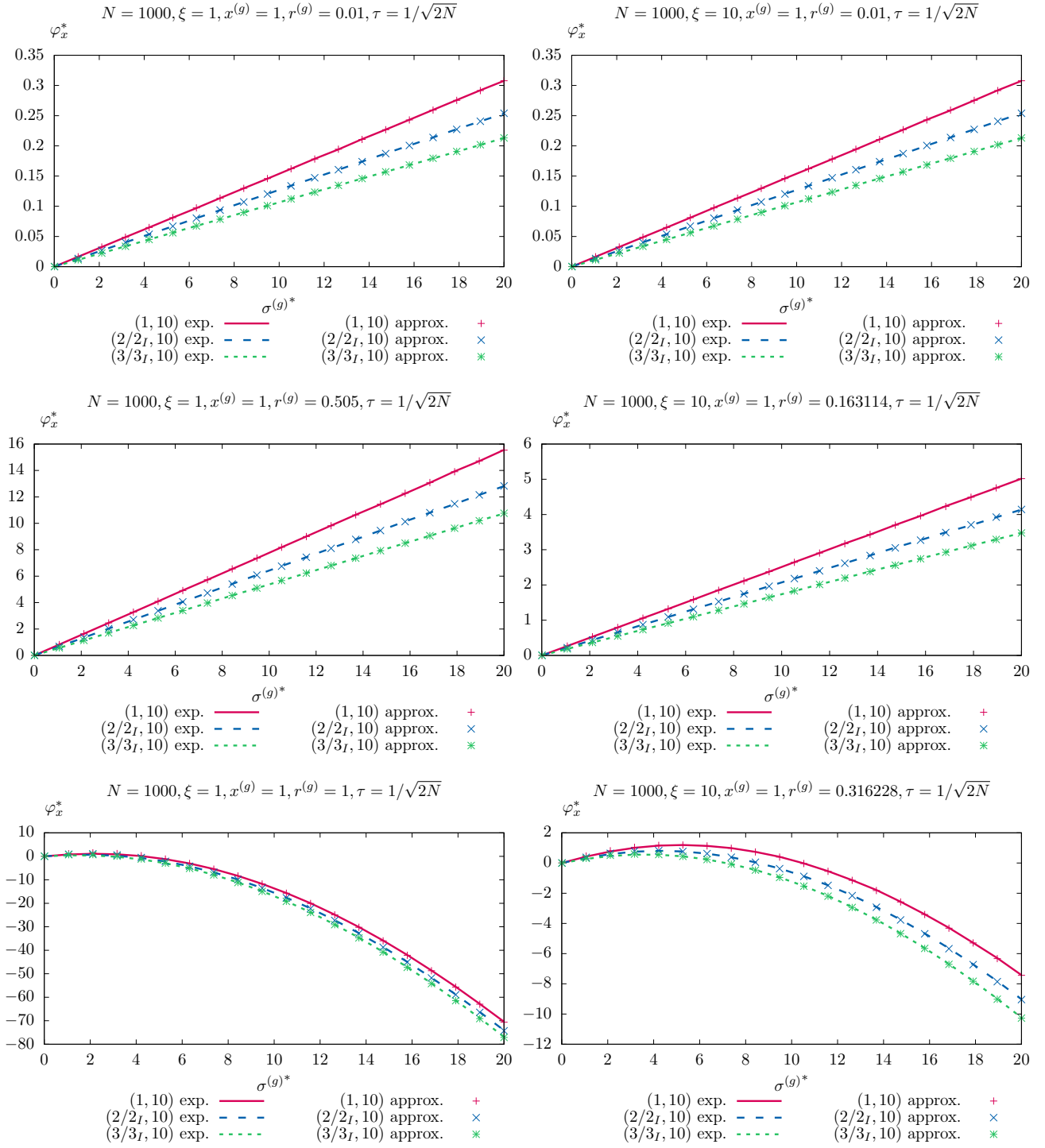
ADDITIONAL PLOTS COMPARING DERIVED APPROXIMATIONS WITH EXPERIMENTS

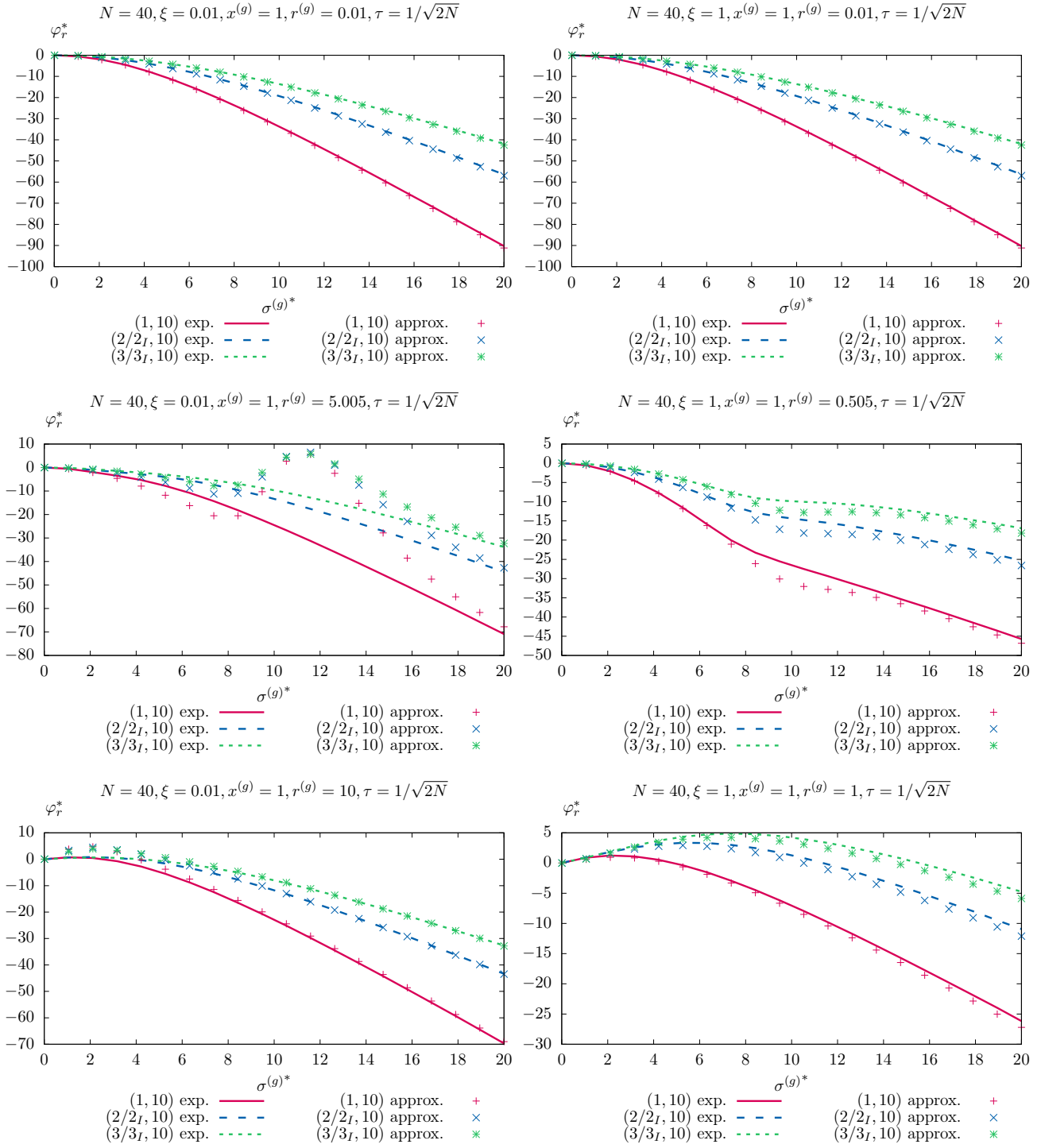
Figs. 11 to 19 show plots comparing the derived closed-form approximation for φ_x^* , φ_r^* , and ψ , respectively, with one-generation experiments. The pluses, crosses, and stars have been calculated by evaluating Equation (18) with Equations (15), (C.66), and (C.73), Equation (21) with Equations (15) and (C.73), and Equation (22) with Equation (15), respectively. The solid, dashed, and dotted lines have been generated by one-generation experiments. For this, the generational loop has been run 10^5 times for a fixed parental individual and constant parameters. The experimentally determined values for φ_x^* , φ_r^* , and ψ from those 10^5 runs have been averaged. The top row in every set of 3×2 plots shows the case that the parental individual is near the cone axis (note that here the feasible case dominates). The bottom row in every set of 3×2 plots shows the case that the parental individual is in the vicinity of the cone boundary (note that here the infeasible case dominates). And the middle row shows a case in between.

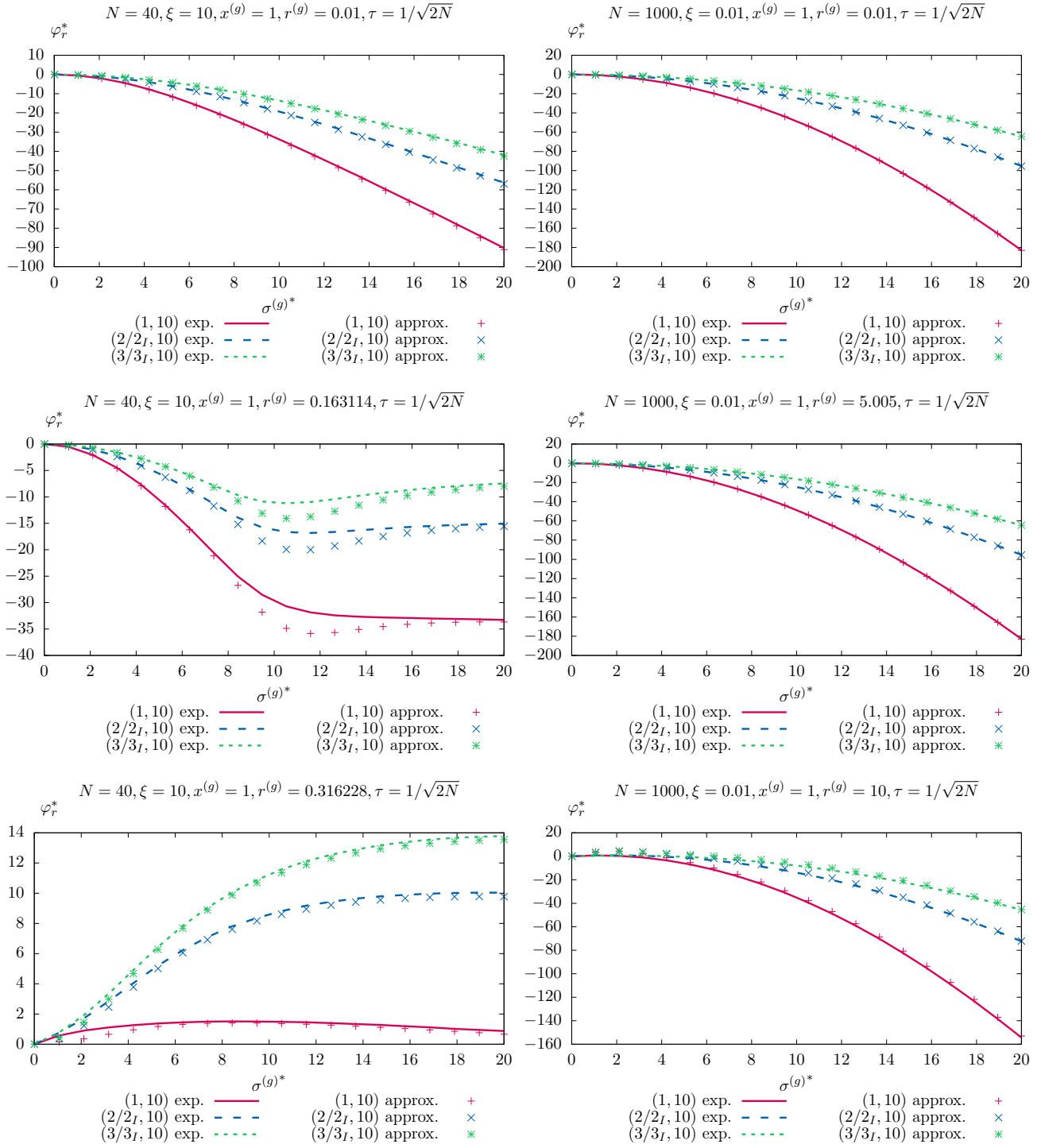
Figs. 20 to 25 show the mean value dynamics of the $(3/3_I, 10)$ -ES applied to the conically constrained problem with different parameters as indicated in the title of the subplots. The plots are organized into three rows and two columns. The first two rows show the x (first row, first column), r (first row, second column), σ (second row, first column), and σ^* (second row, second column) dynamics. The third row shows x and r converted into each other by $\sqrt{\xi}$. The third row shows that after some initial phase, the ES transitions into a steady state. In this steady state, the ES moves near the cone boundary. This becomes clear in the plots because the equation for the cone boundary is $r = x/\sqrt{\xi}$ or equivalently $x = r\sqrt{\xi}$. Furthermore, in this steady state, the normalized mutation strength is constant on average. The lines for the real runs have been generated by averaging 100 real runs of the ES. The fluctuations of those runs are visualized by the shaded areas. As the standard deviation around the mean can result in negative values, those show the range of the minimal and maximal values reached in the 100 runs. This allows using a logarithmic scale. Note that for the σ^* plot, the visualization of the fluctuations has been clipped. The lines for the iteration with one-generation experiments have been determined by iterating the mean value iterative system with one-generation experiments for $\varphi_x^{(g)*}$, $\varphi_r^{(g)*}$, and $\psi^{(g)}$. The lines for the iteration by approximation have been computed by iterating the mean value iterative system with the derived approximations in Section IV-B for $\varphi_x^{(g)*}$, $\varphi_r^{(g)*}$, and $\psi^{(g)}$. Note that due to the approximations used it is possible that in a generation g the iteration of the mean value iterative system yields infeasible $(x^{(g)}, r^{(g)})^T$. In such cases, the particular $(x^{(g)}, r^{(g)})^T$ values have been projected back and projected values used in the further iterations.

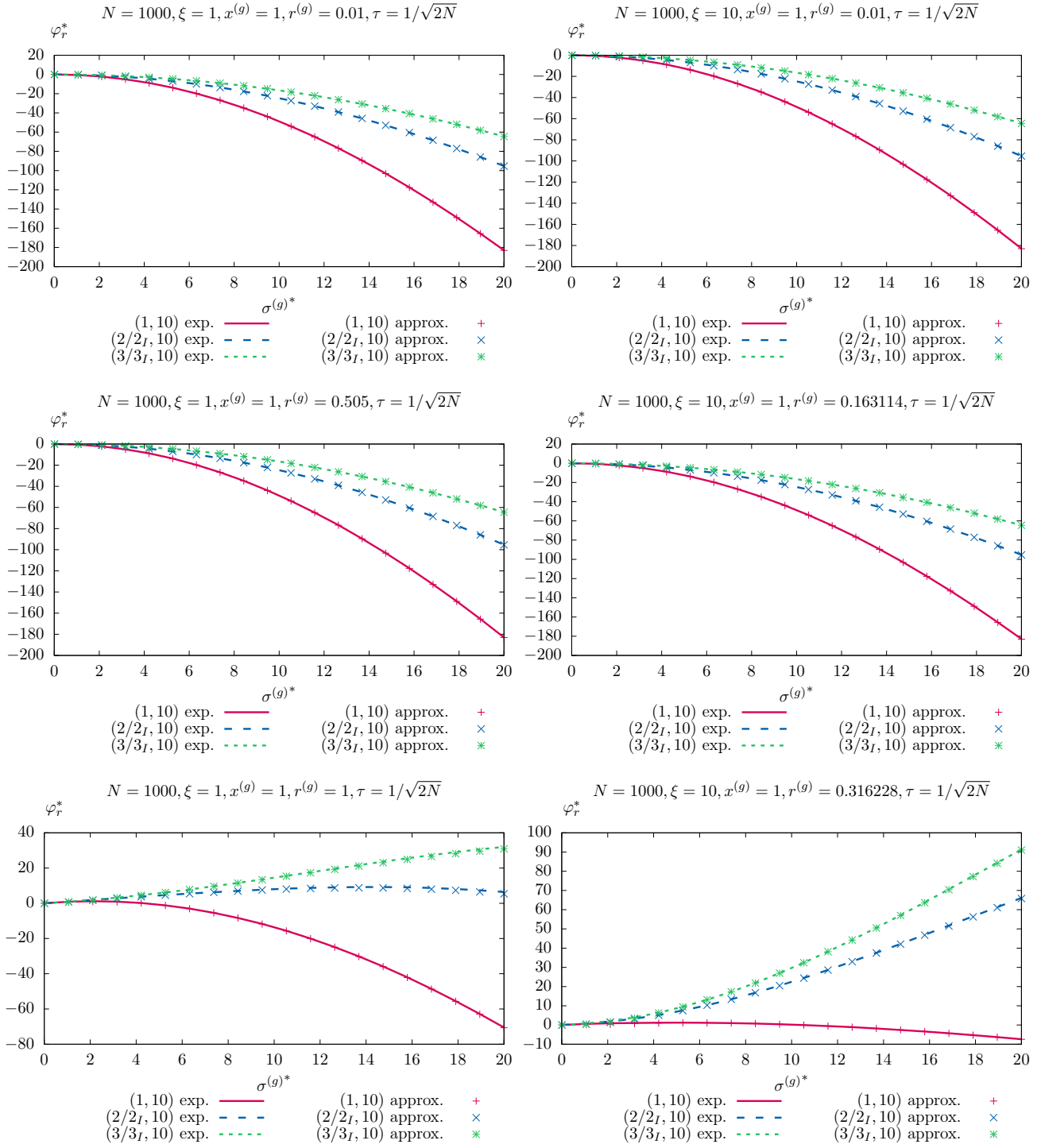
Fig. 11. Comparison of the x progress rate approximation with simulations. (Part 1)

Fig. 12. Comparison of the x progress rate approximation with simulations. (Part 2)

Fig. 13. Comparison of the x progress rate approximation with simulations. (Part 3)

Fig. 14. Comparison of the r progress rate approximation with simulations. (Part 1)

Fig. 15. Comparison of the r progress rate approximation with simulations. (Part 2)

Fig. 16. Comparison of the r progress rate approximation with simulations. (Part 3)

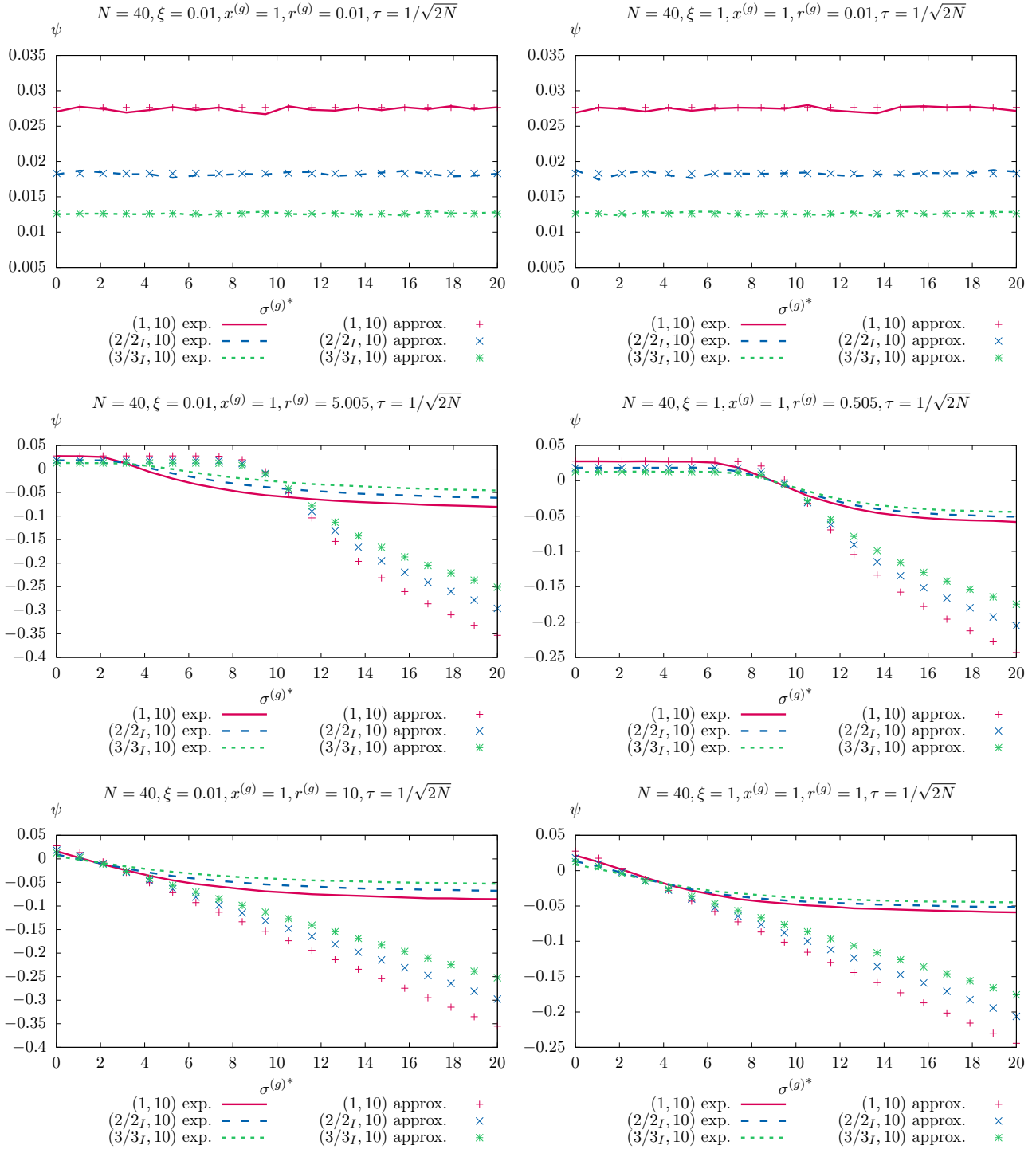


Fig. 17. Comparison of the SAR approximation with simulations. (Part 1)

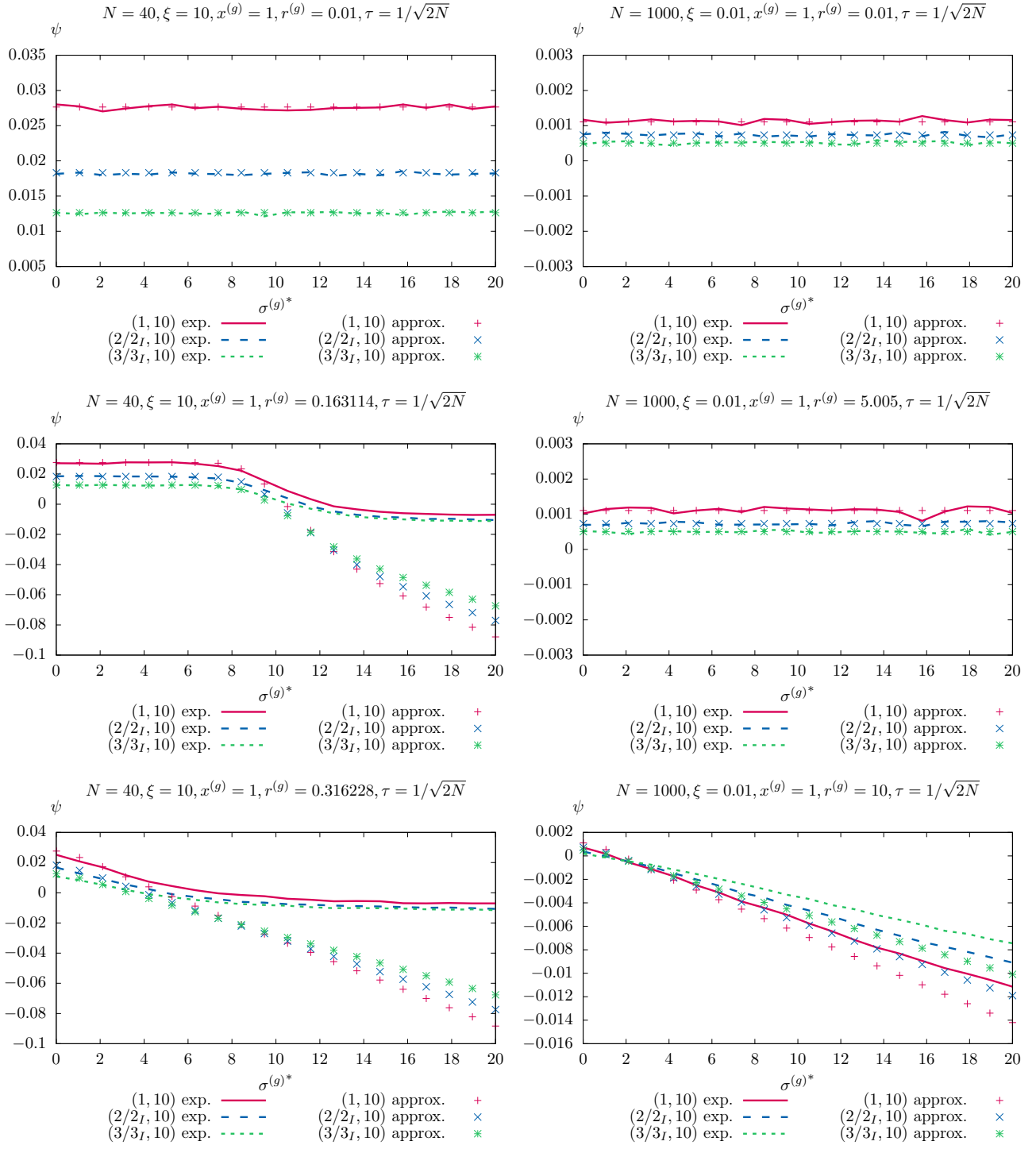


Fig. 18. Comparison of the SAR approximation with simulations. (Part 2)

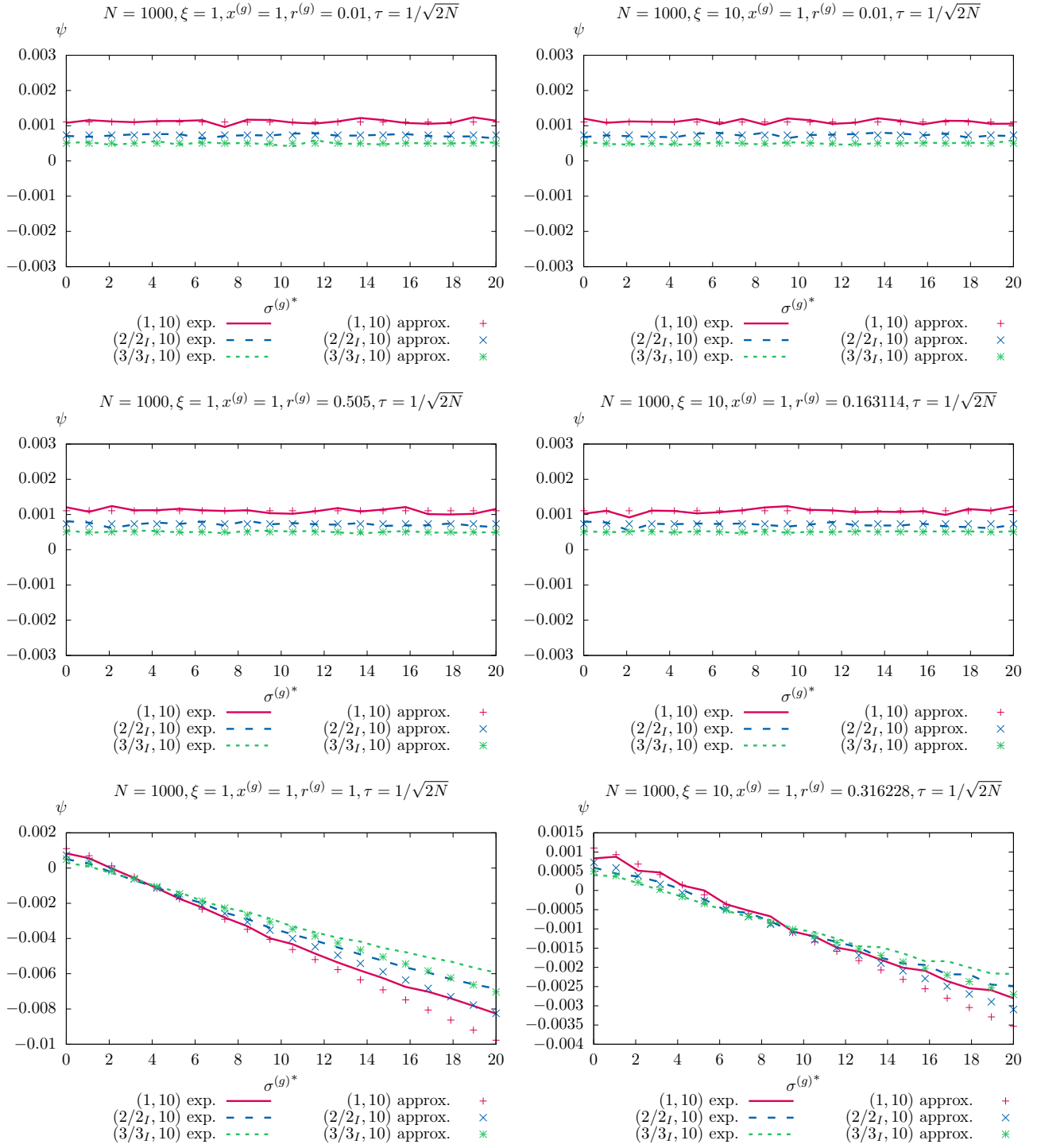


Fig. 19. Comparison of the SAR approximation with simulations. (Part 3)

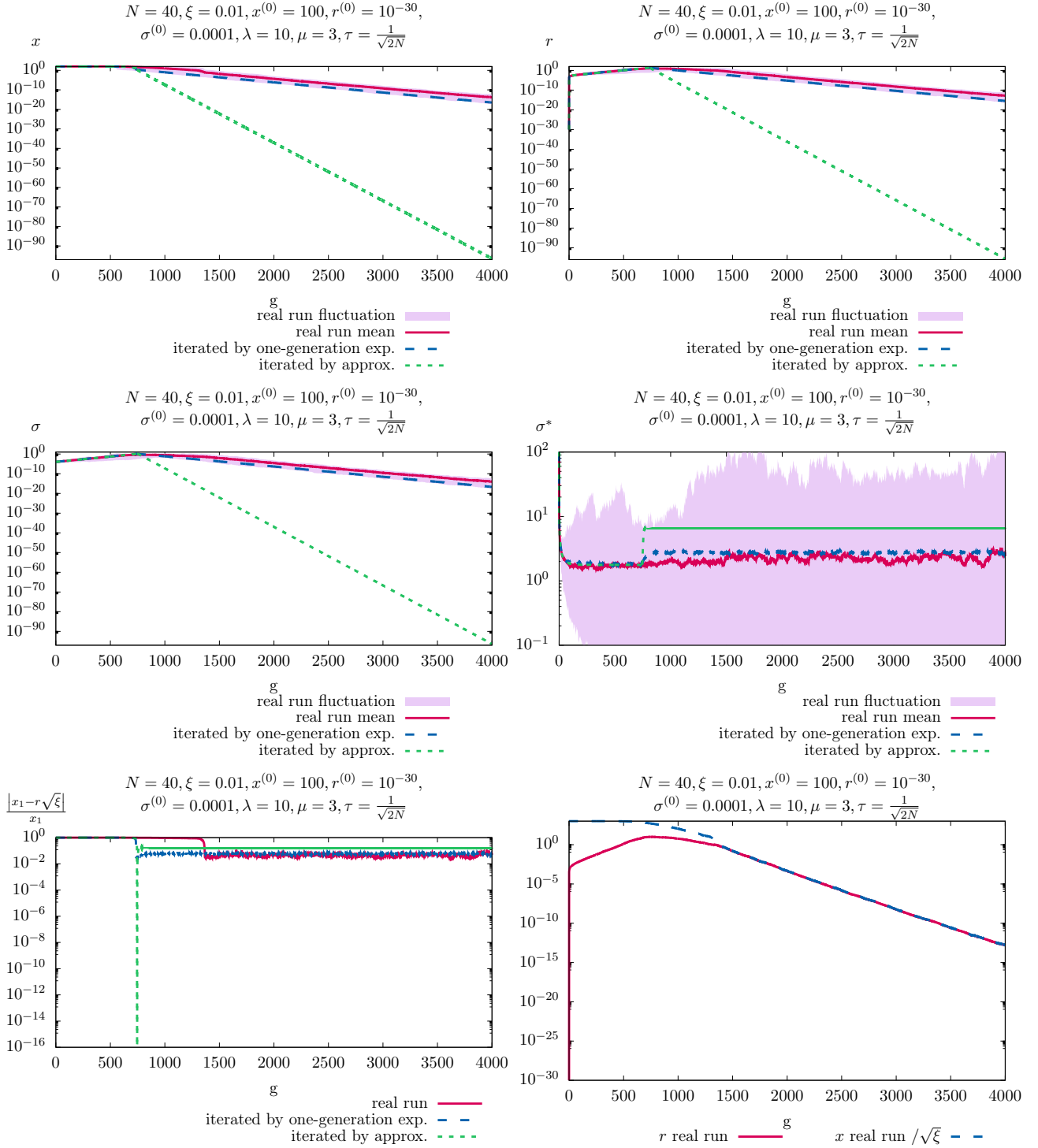


Fig. 20. Mean value dynamics closed-form approximation and real-run comparison of the $(3/3_I, 10)$ -ES with repair by projection applied to the conically constrained problem. (Part 1)

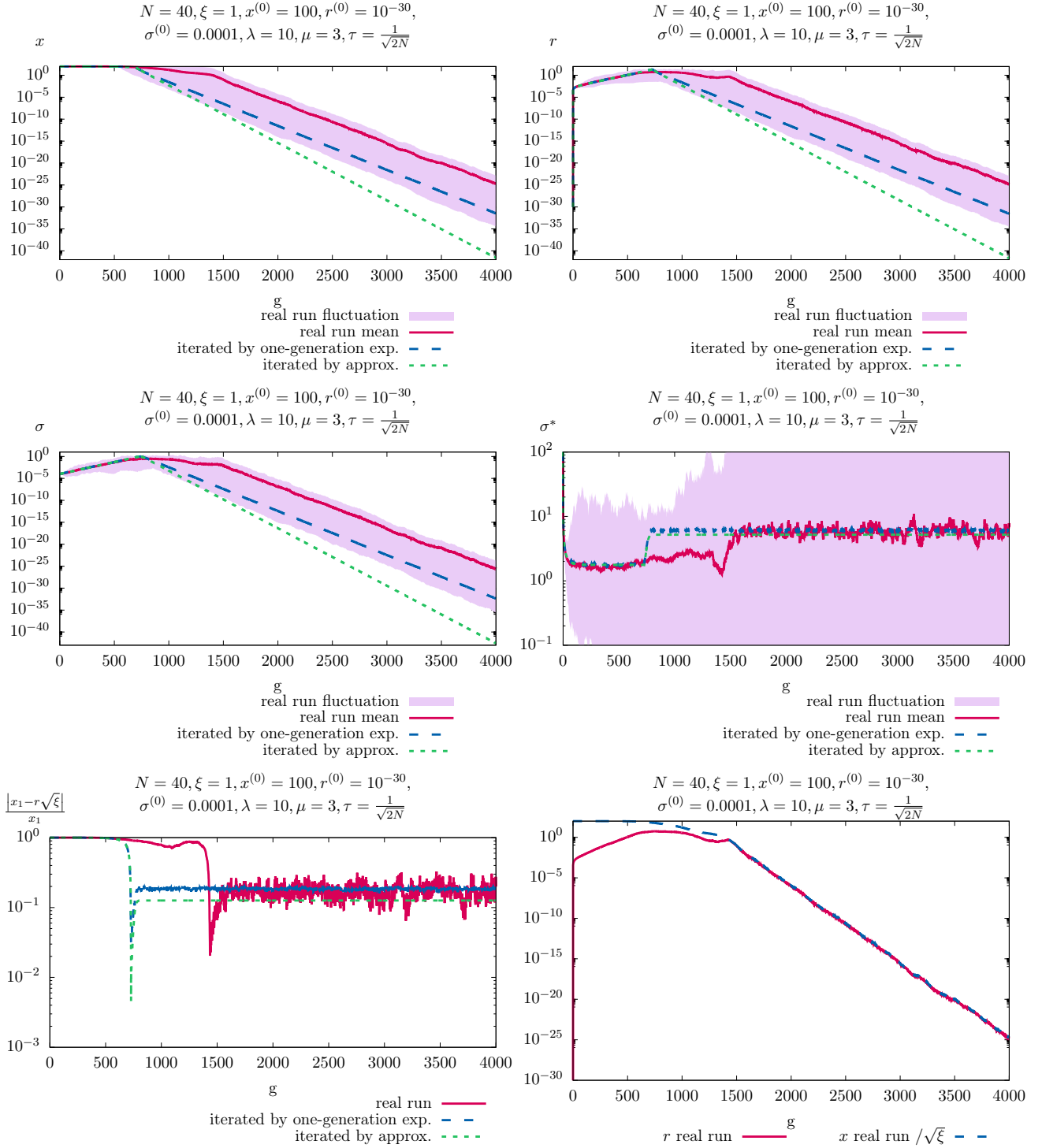


Fig. 21. Mean value dynamics closed-form approximation and real-run comparison of the $(3/3_I, 10)$ -ES with repair by projection applied to the conically constrained problem. (Part 2)

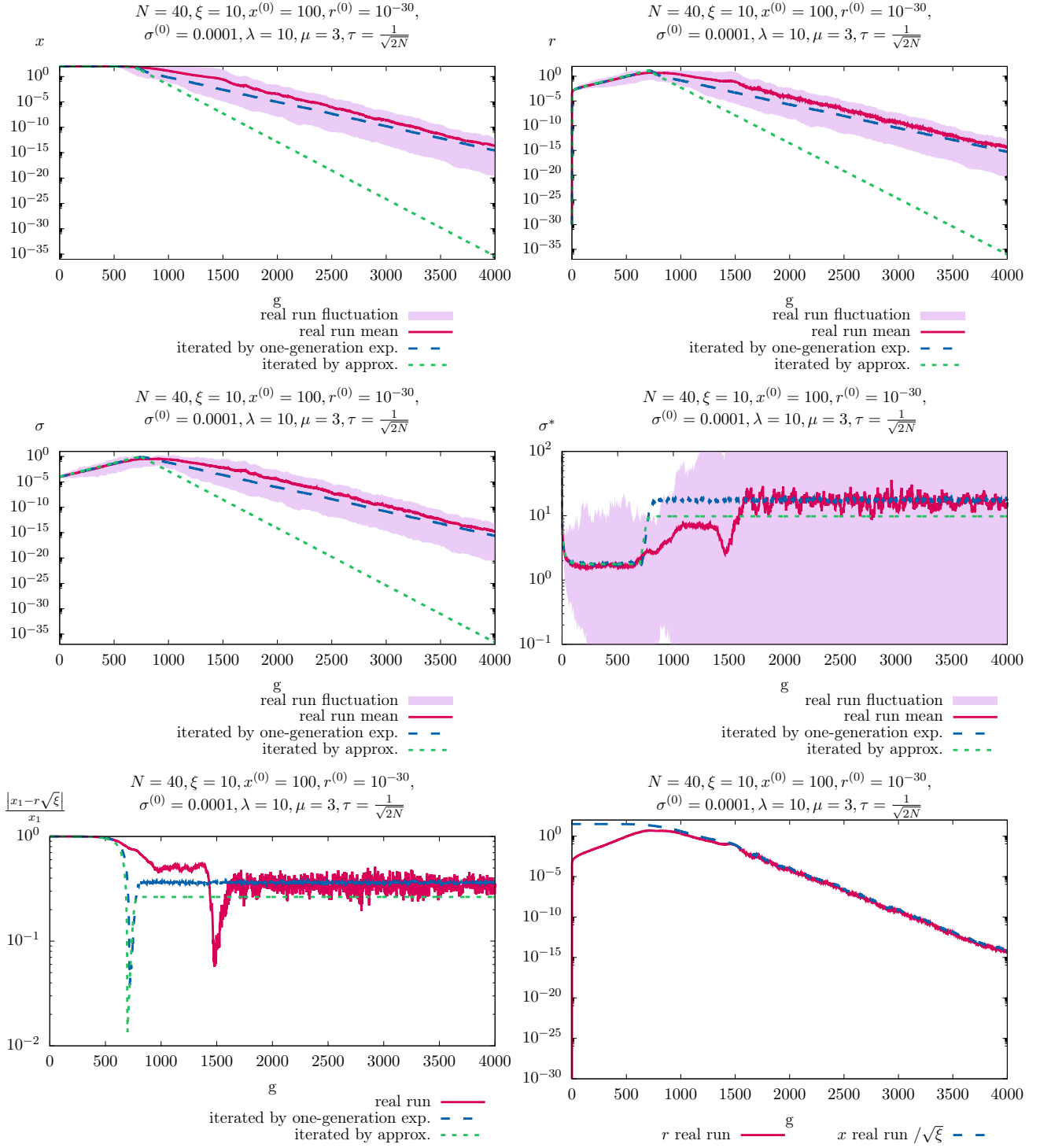


Fig. 22. Mean value dynamics closed-form approximation and real-run comparison of the $(3/3_I, 10)$ -ES with repair by projection applied to the conically constrained problem. (Part 3)

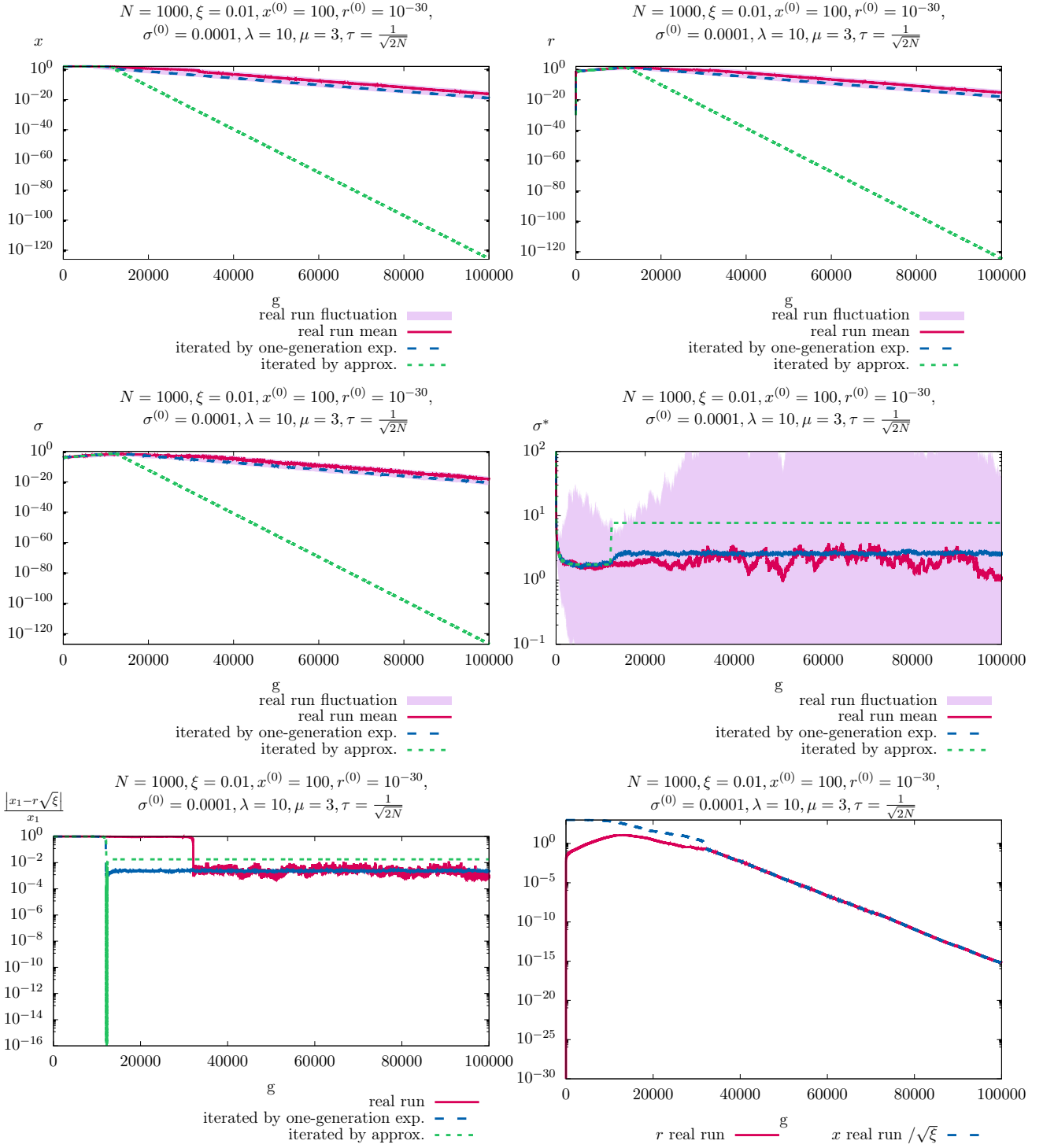


Fig. 23. Mean value dynamics closed-form approximation and real-run comparison of the $(3/3_I, 10)$ -ES with repair by projection applied to the conically constrained problem. (Part 4)

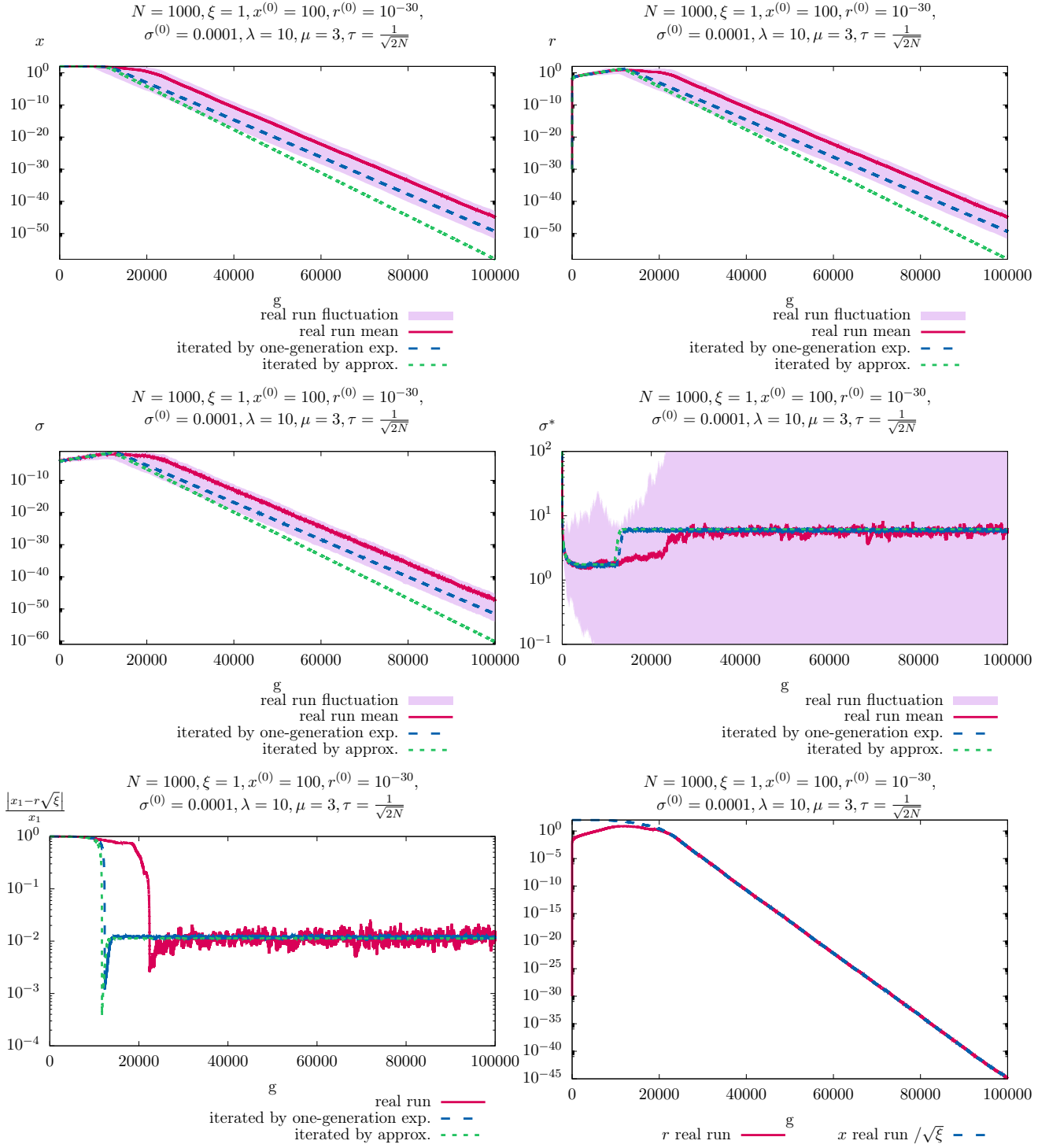


Fig. 24. Mean value dynamics closed-form approximation and real-run comparison of the $(3/3_I, 10)$ -ES with repair by projection applied to the conically constrained problem. (Part 5)

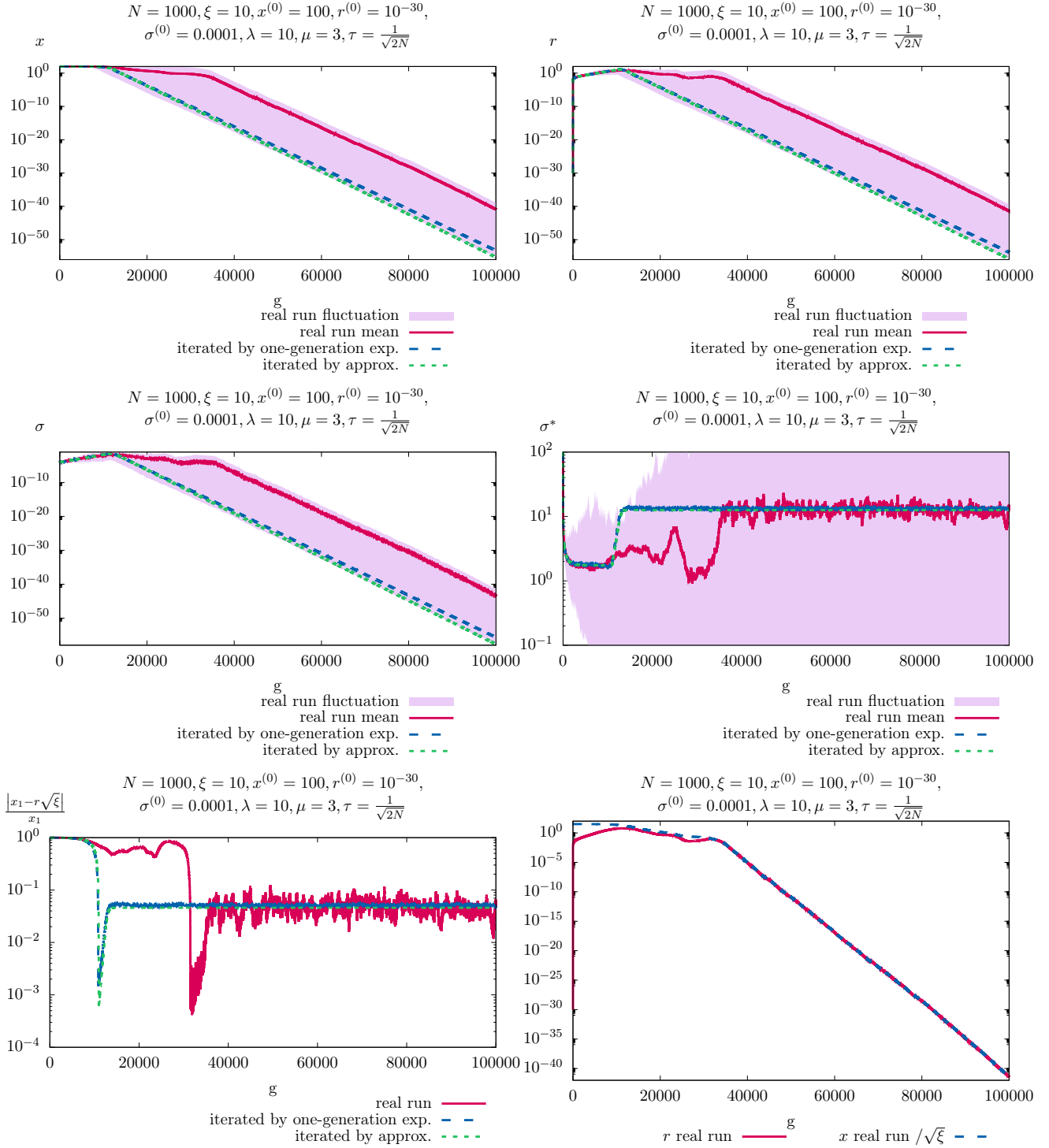


Fig. 25. Mean value dynamics closed-form approximation and real-run comparison of the $(3/3_I, 10)$ -ES with repair by projection applied to the conically constrained problem. (Part 6)

REFERENCES FOR THE APPENDICES

- [13] P. Spettel and H.-G. Beyer, "Technical report: Analysis of the $(1, \lambda)$ - σ -self-adaptation evolution strategy with repair by projection applied to a conically constrained problem," Vorarlberg University of Applied Sciences, Tech. Rep. TR-SAESCONE-18, 2018. [Online]. Available: https://www.fhv.at/fileadmin/user_upload/fhv/files/forschung/ppe/working-papers/Analysis_oncommalambdasigmaSA-ES_Projection_Cone_TR.pdf
- [14] H.-G. Beyer, *The Theory of Evolution Strategies*, ser. Natural Computing Series. Springer, 2001.
- [15] B. Arnold, N. Balakrishnan, and H. Nagaraja, *A First Course in Order Statistics*, ser. Wiley Series in Probability and Statistics. Wiley, 1992.