# Analysis of the $(1,\lambda)$-$\sigma$-Self-Adaptation Evolution Strategy with Repair by Projection Applied to a Conically Constrained Problem

Patrick Spettel[a,*], Hans-Georg Beyer[a]

[a] *Research Center Process and Product Engineering*
*Vorarlberg University of Applied Sciences*
*Hochschulstr. 1, 6850 Dornbirn, Austria*

---

**Abstract**

A thorough theoretical analysis of evolution strategies with constraint handling is important for the understanding of the inner workings of evolution strategies applied to constrained problems. Simple problems are of interest for the first analyses. To this end, the behavior of the $(1,\lambda)$-$\sigma$-Self-Adaptation Evolution Strategy applied to a conically constrained problem is analyzed. For handling infeasible offspring, a repair approach that projects infeasible offspring onto the boundary of the feasibility region is considered. Closed-form approximations are derived for the expected changes of an individual's parameter vector and mutation strength from one generation to the next. For analyzing the strategy's behavior over multiple generations, deterministic evolution equations are derived. It is shown that those evolution equations together with the approximate one-generation expressions allow to approximately predict the evolution dynamics using closed-form approximations. Those derived approximations are compared to simulations in order to visualize the approximation quality.

*Keywords:* Evolution strategies, Repair by projection, Conically constrained problem

---

## 1. Introduction

As constrained optimization is important in many practical applications, the theoretical understanding of evolution strategies (ESs) with constraint handling is of interest in current research. In [1], the $(1,\lambda)$-ES with constraint handling by resampling is analyzed for a single linear constraint. This is extended in [2] with the analysis of repair by projection for a single linear constraint. In [3], the repair approach analyzed in [2] is compared with an approach that reflects infeasible

*Corresponding author
*Email addresses:* Patrick.Spettel@fhv.at (Patrick Spettel), Hans-Georg.Beyer@fhv.at (Hans-Georg Beyer)

points into the feasible region and an approach that truncates infeasible points. A variant of the $(1, \lambda)$-ES for a conically constrained problem is considered in [4]. There, offspring are discarded until feasible are obtained. In [5], premature convergence for a $(1 + 1)$-ES with mutation strength adaptation by a variant of the 1/5th rule was proven. A $(1 + 1)$-ES with augmented Lagrangian constraint handling is presented in [6]. The one-generation behavior of the proposed ES is analyzed on the sphere model with one linear inequality constraint.

This algorithm has been extended to a multi-recombinative variant for a single linear constraint in [7] and multiple linear constraints in [8]. For both cases, linear convergence behavior was investigated by the use of Markov chains.

The goal of this paper is to extend the analysis for the local progress measures in [4] with an analysis for a repair approach based on projection. The repair method is performed by projecting infeasible offspring onto the boundary of the feasible region by minimizing the Euclidean distance to the constraint boundary. For the mutation strength control, $\sigma$-Self-Adaptation is considered. The remainder of the paper is organized as follows. The optimization problem and the algorithm under consideration are described in Section 2 and Section 3, respectively. Next, the theoretical results are presented. To this end, the theoretical analysis method used is explained in Section 4.1. This is followed by the investigation of the algorithm's microscopic aspects (one-generation behavior) in Section 4.2 and then by the analysis of the algorithm's macroscopic behavior (multi-generation behavior, i.e., the evolution dynamics) in Section 4.3. For both, the microscopic and the macroscopic behavior, closed-form approximations under asymptotic assumptions are derived. The approximations are compared to simulations. For readability purposes, most of the longer derivations are provided in the sections of the appendix. Moreover, where appropriate and interesting, further investigations have been performed. The outcomes of those investigations and further details for the derivations are provided in the technical report [9] accompanying this paper. Relevant cross-references are provided throughout the text.

*Notations:* Boldface $\mathbf{x} \in \mathbb{R}^N$ is a column vector with $N$ real-valued components. $\mathbf{x}^T$ is its transpose. $x_k$ and equivalently $(\mathbf{x})_k$ denote the $k$-th element of a vector $\mathbf{x}$. $\mathbf{x}_{m;\lambda}$ is the order statistic notation, i.e., it denotes the $m$-th best (with respect to fitness) of a list of $\lambda$ elements. $||\mathbf{x}|| = \sqrt{\sum_{k=1}^{N} {x_k}^2}$ denotes the Euclidean norm. $\mathbf{0}$ is the vector with all elements equal to zero. $\mathbf{I}$ is the identity matrix. $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ denotes the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. $\mathcal{N}(\mu, \sigma^2)$ is written for the normal distribution with mean $\mu$ and variance $\sigma^2$. The symbol $\sim$ means "distributed according to", $\gg$ "much greater than", $\simeq$ "asymptotically equal", and $\approx$ "approximately equal". A superscript $\mathbf{x}^{(g)}$ stands for the element in the $g$-th generation.

## 2. The Problem

The optimization problem under consideration is

$$f(\mathbf{x}) = x_1 \to \min! \tag{1}$$

subject to constraints

$$x_1^2 - \xi \sum_{k=2}^{N} x_k^2 \geq 0 \tag{2}$$

$$x_1 \geq 0 \tag{3}$$

where $\mathbf{x} = (x_1, \ldots, x_N)^T \in \mathbb{R}^N$ and $\xi > 0$.

Due to symmetry considerations, a point in the search space can be uniquely described by its distance $x$ from 0 in $x_1$ direction (cone axis) and the distance $r$ from the cone's axis. In the further analyses, this is called the $(x, r)^T$-space. It is visualized in Fig. 1. Because the distance from the cone's axis is positive, only half of the cone needs to be considered. The cone boundary and the projection lines are visualized. The equation for the cone boundary $r = \frac{x_1}{\sqrt{\xi}}$ is a direct consequence of the problem definition (Eq. (2)). The equation for the projection line can be derived by considering a direction vector of the cone boundary $\left(1, \frac{1}{\sqrt{\xi}}\right)^T$ and its counterclockwise rotation by 90 degrees $\left(-\frac{1}{\sqrt{\xi}}, 1\right)^T$. Using those direction vectors, at a given value $q$ (as shown in Fig. 1), an equation for the projection line reads $r = -\sqrt{\xi}x_1 + q\left(\sqrt{\xi} + \frac{1}{\sqrt{\xi}}\right)$. For $q = 0$, this results in $-\sqrt{\xi}x_1$. Additionally, in Fig. 1, an offspring $\tilde{\mathbf{x}}$ with its parent $\mathbf{x}$ and the corresponding mutation vector $\mathbf{z}$ scaled by the mutation strength $\tilde{\sigma}$ are shown. The offspring's $x_1$ and $r$ values after the projection step are indicated by $q$ and $q_r$, respectively.

## 3. The Algorithm

The considered algorithm is a $(1, \lambda)$-$\sigma$-Self-Adaptation-ES. The pseudo code is shown in Alg. 1. After initialization (Lines 1 to 2), the generation loop is entered. In Lines 6 to 15, $\lambda$ offspring are generated. For every offspring, its mutation strength $\tilde{\sigma}_l$ is determined by mutating the parental mutation strength $\sigma^{(g)}$ using a log-normal distribution (Line 7). This mutation strength is then used to sample the offspring's parameter vector from a multivariate normal distribution with mean $\mathbf{x}^{(g)}$ and standard deviation $\tilde{\sigma}_l$ (Line 8). Then, it is determined whether repair is necessary. Repair is necessary if the offspring is infeasible. If the generated offspring is infeasible, its parameter vector is projected onto the point on the boundary of the feasible region that minimizes the Euclidean distance to the offspring point (Lines 9 to 11). That is, if the offspring is not feasible, the optimization problem

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}'} \|\mathbf{x}' - \mathbf{x}\|^2 \text{ such that } {x'_1}^2 - \xi \sum_{k=2}^{N} {x'_k}^2 \geq 0, x'_1 \geq 0 \tag{4}$$

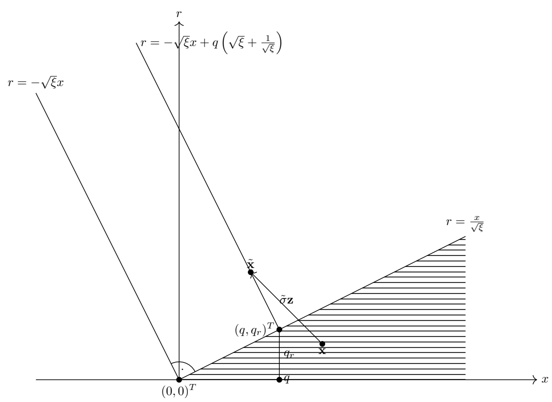


Figure 1: The conically constrained optimization problem in $N$ dimensions shown in the $(x, r)^T$-space.

must be solved. In (4), $\mathbf{x}$ is the individual to be projected. For this, a convenience function

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

is introduced, returning $\hat{\mathbf{x}}$ of the problem (4). Appendix A presents a geometrical approach for deriving a closed-form solution to the projection optimization problem (4). The values $x^{(g)}$, $r^{(g)}$, $q_l$, $q_{r\,l}$, $q_{1;\lambda}$, and $q_{r\,1;\lambda}$ are only needed in the theoretical analysis and can be removed in practical applications of the ES. They are indicated in the algorithm in Lines 4, 5, 13, 14, 19, and 20, respectively. In Line 12, the offspring's fitness is determined. After the procreation step, 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 set to the corresponding values of the offspring with the smallest objective function value. The update of the generation counter ends one iteration of the generation loop. The generation loop is quit if the defined termination criteria are met (for example if a maximum number of generations is reached, if a sigma threshold is reached, etc.). Fig. 2 shows an example of the $x$- and $r$-dynamics generated by running Alg. 1 (solid line) in comparison with the iteration of the closed-form approximation iterative system (dotted line) that is derived in the following sections.

## 4. The Theoretical Analysis

### *4.1. The Dynamical Systems Approach*

For the analysis of the $(1, \lambda)$-ES (Alg. 1), the $(x, r)^T$-modeling described in Section 2, is used. The goal is to compute the evolution dynamics of the $(1, \lambda)$-ES. For doing this, the dynamical systems approach presented in [10] is used. This approach models specific state variables of the strategy over time. For the problem under consideration, there are three random variables that describe the system (assuming constant exogenous parameters). The $x$ and $r$ values of the

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

```
Initialize x^(0), σ^(0), τ, λ
g ← 0
repeat
    x^(g) = (x^(g))_1
    r^(g) = sqrt( Σ_{k=2}^{N} (x^(g))_k^2 )
    for l ← 1 to λ do
        σ̃_l ← σ^(g) e^{τ N(0,1)}
        x̃_l ← x^(g) + σ̃_l N(0, I)
        if not isFeasible(x̃_l) then                ▷ see Algorithm 2
            x̃_l ← projectOntoCone(x̃_l)             ▷ see Eqs. (4) and (5)
        end if
        f̃_l ← f(x̃_l) = (x̃_l)_1
        q_l = (x̃_l)_1
        q_{r l} = sqrt( Σ_{k=2}^{N} (x̃_l)_k^2 )
    end for
    Sort offspring according to f̃_l in ascending order
    x^(g+1) ← x̃_{1;λ}
    σ^(g+1) ← σ̃_{1;λ}
    q_{1;λ} = (x^(g+1))_1
    q_{r 1;λ} = sqrt( Σ_{k=2}^{N} (x^(g+1))_k^2 )
    g ← g + 1
until termination criteria are met
```

**Algorithm 2** Feasibility check

```
function isFeasible(x)
    return(x_1 ≥ 0 ∧ x_1^2 − ξ Σ_{k=2}^{N} x_k^2 ≥ 0)
end function
```

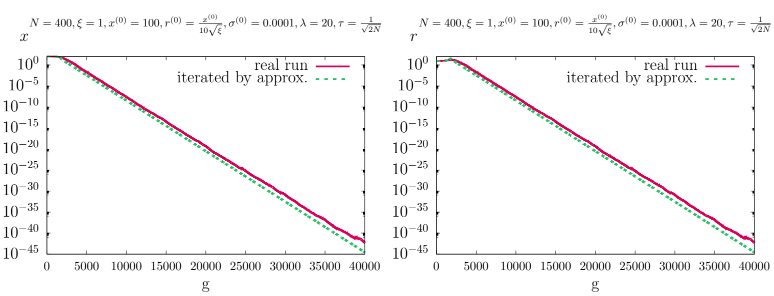


Figure 2: Comparison of the $x$- and $r$-dynamics of a real ES run (solid line) with the iteration of the closed-form approximation of the iterative system (dotted line) that is derived in the following sections.

current parental individual and the current mutation strength $\sigma$. The transition between consecutive states of the evolution strategy can then be modeled as a Markov process. Deriving closed-form expressions for the transition equations (so-called Chapman-Kolmogorov equations, see, e.g. [11]) is often not possible. Instead, approximate equations for the change of all the state variables are usually derived. The change of the random state variables is expressed in two parts. The first part is the expected change and the second part comprises the stochastic fluctuations. Making use of functions that express the change $(\varphi_x, \varphi_r, \psi)$, so-called evolution equations describe the system. They read

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

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

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

and are stochastic difference equations. The expected changes in the parameter space are expressed as so-called progress rates. They are defined as

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

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

Normalized variants are introduced as $\varphi_x^*(\cdot) := \frac{N\varphi_x(\cdot)}{x^{(g)}}$ and $\varphi_r^*(\cdot) := \frac{N\varphi_r(\cdot)}{r^{(g)}}$. Similarly, a normalized variant of the mutation strength is introduced as $\sigma^* := \frac{N\sigma}{r^{(g)}}$. Evolution of the mutation strength is performed by multiplication with a log-normally distributed random variable. Therefore, its relative expected change is of interest. This yields a slightly different (in comparison to the progress rates in the parameter space) progress measure for the mutation strength, the so-called self-adaptation response (SAR). It is defined as

$$\psi(x^{(g)}, r^{(g)}, \sigma^{(g)}) := \mathrm{E}\left[\frac{\sigma^{(g+1)} - \sigma^{(g)}}{\sigma^{(g)}} \,\middle|\, x^{(g)}, r^{(g)}, \sigma^{(g)}\right]. \tag{11}$$

The fluctuations are represented as random variables $\epsilon_x$, $\epsilon_r$, and $\epsilon_\sigma$ with necessarily $\mathrm{E}[\epsilon_x] = 0$, $\mathrm{E}[\epsilon_r] = 0$, and $\mathrm{E}[\epsilon_\sigma] = 0$. Fluctuation terms are treated in detail in [12]. For the analysis in this paper, it is assumed that it is sufficient to consider the dynamics without fluctuations (i.e., $\epsilon_x = 0$, $\epsilon_r = 0$, and $\epsilon_\sigma = 0$) in order to approximately model the evolution dynamics of the strategy in the asymptotic limit case $N \to \infty$. Such evolution equations without fluctuation terms are referred to as deterministic evolution equations (or mean value evolution equations). They are usually less complex to deal with in theoretical analyses. As it turns out, it is possible to approximately predict the mean value evolution dynamics of the evolution strategy using these evolution equations. The next step is to derive expressions for the functions $\varphi_x$, $\varphi_r$, and $\psi$.

*4.2. The Microscopic Aspects*

The microscopic aspects deal with the local analysis of the behavior of the evolution strategy. This means that, given the current parental individual and the current mutation strength, a step from one generation to the next is considered. The progress rates express the expected change of the parental individual in the parameter space. And the expected relative change of the mutation strength is given as the SAR. Those are derived in the next sections for the algorithm and the problem under consideration.

*4.2.1. Derivation of the x Progress Rate*

From the definition of the progress rate (Eq. (9)) and the pseudo-code of the ES (Alg. 1, Lines 4 and 19) it follows that

$$\varphi_x(x^{(g)}, r^{(g)}, \sigma^{(g)}) = x^{(g)} - \mathrm{E}[x^{(g+1)} \,|\, x^{(g)}, r^{(g)}, \sigma^{(g)}] \tag{12}$$

$$= x^{(g)} - \mathrm{E}[q_{1;\lambda} \,|\, x^{(g)}, r^{(g)}, \sigma^{(g)}]. \tag{13}$$

This means that the expectation of $q_{1;\lambda}$, i.e., the $x$ value after (possible) projection of the best offspring, is needed for the derivation of the progress rate in $x$ direction

$$\mathrm{E}[q_{1;\lambda} \,|\, x^{(g)}, r^{(g)}, \sigma^{(g)}] := \mathrm{E}[q_{1;\lambda}] = \int_{q=0}^{q=\infty} q\, p_{q_{1;\lambda}}(q)\, \mathrm{d}q \tag{14}$$

$$= \lambda \int_{q=0}^{q=\infty} q\, p_Q(q)[1 - P_Q(q)]^{\lambda-1}\, \mathrm{d}q. \tag{15}$$

Eq. (14) follows directly from the definition of expectation where

$$p_{q_{1;\lambda}}(q) := p_{q_{1;\lambda}}(q \,|\, x^{(g)}, r^{(g)}, \sigma^{(g)})$$

indicates the probability density function of the best (offspring with smallest $q$ value) offspring's $q$ value. The random variable $Q$ denotes the random $x$ values after projection. The step to Eq. (15) follows from the calculation of $p_{q_{1;\lambda}}(q)$. Because the objective function (Eq. (1)) is defined to return an individual's $x$ value, $p_{q_{1;\lambda}}(q)$ is the probability density function of the minimal $q$ value among $\lambda$ values. This calculation is well-known in order statistics (see, e.g., [13]). A short derivation is presented here for the case under consideration. One arbitrarily selected mutation out of the total $\lambda$ mutations has probability density $p_Q(q) := p_Q(q \,|\, x^{(g)}, r^{(g)}, \sigma^{(g)})$ for its projected value $q$. For this particular $q$ value to be the smallest, all other $\lambda-1$ values must have larger $q$ values. Because they are statistically independent, this results in the probability $[1-P_Q(q)]^{\lambda-1}$ of all other mutations being larger than $q$. $P_Q(q) := P_Q(q \,|\, x^{(g)}, r^{(g)}, \sigma^{(g)})$ denotes the cumulative distribution function of the random variable of the $q$ values $Q$, i.e., $P_Q(q) = \Pr[Q \le q]$. The density for the mutation considered is thus $p_Q(q)[1-P_Q(q)]^{\lambda-1}$. Since there are $\lambda$ possibilities for best $q$ values, one obtains $p_{q_{1;\lambda}}(q) = \lambda p_Q(q)[1 - P_Q(q)]^{\lambda-1}$ where $P_Q(q) = \Pr[Q \le q] = \int_{q'=0}^{q'=q} p_Q(q')\, \mathrm{d}q'$. Note that the lower bound is 0 because the $q$ values are defined to be the $x$ values after projection. Now, to proceed further with Eq. (15), the cumulative distribution function $P_Q(q)$ and the corresponding probability density function $p_Q(q)$ need to be derived.

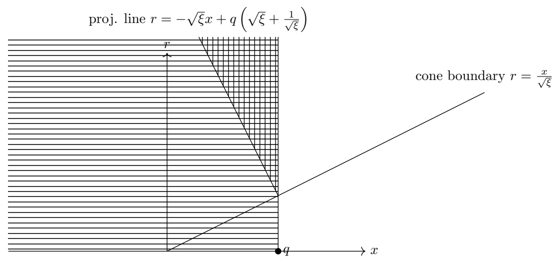


Figure 3: Visualization of the integration for the calculation of $P_Q(q) = \Pr[Q \le q]$. Note that $\Pr[Q \le q]$ is the area of the horizontally hatched region.

*The Exact $P_Q(q)$ and $p_Q(q)$ Functions.* The cumulative distribution function $P_Q(q)$ and its corresponding probability density function $p_Q(q)$ of projected values in direction of the cone axis are derived in this section. The approach followed is to compute $\Pr[Q \le q]$ by integration. By doing this, $P_Q(q) = \Pr[Q \le q]$ is derived. Then, $p_Q(q)$ follows by computing the derivative using the fact that $p_Q(q) = \frac{\mathrm{d}}{\mathrm{dq}} P_Q(q)$. For computing $\Pr[Q \le q]$, consider the visualization in Fig. 3. The probability $\Pr[Q \le q]$ expressed in terms of the offspring before projection is the area of the horizontally hatched region. The area of the horizontally hatched region can be obtained by integration. To this end, the area of the doubly hatched region is subtracted from the area of the horizontally hatched and doubly hatched regions. Formally, this reads

$$P_Q(q) = \int_{x=-\infty}^{x=q} \int_{r=0}^{r=\infty} p_{1;1}(x,r)\,\mathrm{d}r\,\mathrm{d}x - \int_{r=q/\sqrt{\xi}}^{r=\infty} \int_{x=-(1/\sqrt{\xi})r+(1+1/\xi)q}^{x=q} p_{1;1}(x,r)\,\mathrm{d}x\,\mathrm{d}r \tag{16}$$

$$= \int_{x=-\infty}^{x=q} \int_{r=0}^{r=\infty} p_{1;1}(x,r)\,\mathrm{d}r\,\mathrm{d}x - \int_{x=-\infty}^{x=q} \int_{r=-\sqrt{\xi}x+(\sqrt{\xi}+1/\sqrt{\xi})q}^{r=\infty} p_{1;1}(x,r)\,\mathrm{d}r\,\mathrm{d}x. \tag{17}$$

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. It is derived from Alg. 1. Lines 7 and 8 outline the way a single descendant before projection is generated. The ES is at the parental state $(x^{(g)}, r^{(g)}, \sigma^{(g)})^T$. First, $\sigma^{(g)}$ is mutated. This occurs with the conditional density (log-normal mutation operator)

$$p_\sigma(\sigma \,|\, \sigma^{(g)}) = \frac{1}{\sqrt{2\pi}\tau} \frac{1}{\sigma} \exp\left[-\frac{1}{2}\left(\frac{\ln(\sigma/\sigma^{(g)})}{\tau}\right)^2\right]. \tag{18}$$

The $\sigma$ value obtained serves as the mutation strength in mutating $x$ and $r$. This occurs with the conditional probability density $p_{x,r}(x, r \,|\, x^{(g)}, r^{(g)}, \sigma)$. The expression $p_{x,r}(x, r \,|\, x^{(g)}, r^{(g)}, \sigma)$ denotes the joint probability density function of offspring before projection. The parameter $\sigma$ occurs with conditional probability $p_\sigma(\sigma \,|\, \sigma^{(g)})\,\mathrm{d}\sigma$. The conditional probability density of a single descendant is derived by integrating over all possible values for $\sigma$

$$p_{1;1}(x, r \,|\, x^{(g)}, r^{(g)}, \sigma^{(g)}) = \int_{\sigma=0}^{\sigma=\infty} p_{x,r}(x, r \,|\, x^{(g)}, r^{(g)}, \sigma) p_\sigma(\sigma \,|\, \sigma^{(g)})\,\mathrm{d}\sigma. \quad (19)$$

Insertion of Eq. (19) into Eq. (16) or Eq. (17) results in the exact $P_Q(q)$ function. The resulting expression is difficult to deal with analytically. Numerical integration and approximation are two possibilities to proceed further. Because closed-form solutions are arguably preferable, the latter option is followed.

*The Approximation of the $P_Q(q)$ and $p_Q(q)$ Functions.* In the investigations of the following sections, $\tau$ is assumed to be small. By the properties of the log-normal distribution this leads to a small variance of possible mutated $\sigma$ values $\tilde{\sigma}_l$ of the offspring. The $\tilde{\sigma}_l$ values therefore tend to the expected value of the above defined log-normally distributed mutation strengths. This expected value approaches $\sigma^{(g)}$ for small $\tau$. The assumption allows to assume $\tilde{\sigma}_l \approx \sigma^{(g)}$ which simplifies the analysis because the $\sigma$ mutation can be ignored. Together with further assumptions, this makes closed-form results of the theoretical analysis possible. More formally, this can be shown by an expansion analogously to [10, Section 7.3.2.4]. Consequently, the following probability density for a single descendant before projection is used in the further analysis

$$p_{1;1}(x, r \,|\, x^{(g)}, r^{(g)}, \sigma^{(g)}) \approx p_{x,r}(x, r \,|\, x^{(g)}, r^{(g)}, \sigma^{(g)}) = p_x(x) p_r(r) \quad (20)$$

for sufficiently small $\tau$. The last equality in Eq. (20) follows by statistical independence of the mutation in parameter space. From Line 8 of Alg. 1 the probability density functions $p_x(x)$ and $p_r(r)$ follow.

*The Offspring Density Before Projection in $x$ Direction.* The offspring density of an offspring in $x$ direction before projection follows directly from the offspring generation in Alg. 1 (Line 8). As every component of the offspring's parameter vector $\tilde{\mathbf{x}}$ is independently and identically distributed (i.i.d.) according to a normal distribution it follows together with the assumption $\tilde{\sigma}_l \approx \sigma^{(g)}$ that

$$p_x(x) \approx \frac{1}{\sigma^{(g)}} \phi\left(\frac{x - x^{(g)}}{\sigma^{(g)}}\right) = \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{x - x^{(g)}}{\sigma^{(g)}}\right)^2\right] \quad (21)$$

holds. $\phi(x)$ denotes the probability density function of the standard normal distribution.

*The Offspring Density Before Projection in $r$ Direction and its Normal Approximation.* The offspring density of an offspring in $r$ direction before projection can be derived from Alg. 1 (Line 8). In Section 2 it has already been explained 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, a simplifying assumption can be made due to the problem's symmetry. Because the problem's dimensions $2, \ldots, N$ represent an $N-1$ dimensional sphere, the coordinate system can be rotated without loss of generality. For a particular individual $\tilde{\mathbf{x}}$ the coordinate system is rotated such that $x_1$ points in direction of the individual's $x_1$ direction, i.e., $x_1$ points in the direction of the vector $(\tilde{x}_1, 0, \ldots, 0)^T$ and $x_2$ points in the direction of the individual's other components, i.e., $x_2$ points in the direction of the vector $(0, \tilde{x}_2, \ldots, \tilde{x}_N)^T$. This rotation of the coordinate system allows to conveniently represent an individual. Note that the length in the direction of $x_2$ is exactly the individual's distance from the cone axis. With this observation, it immediately follows 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 $(\tilde{x}_1, \tilde{r}, 0, \ldots, 0)^T$ 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 $\chi$ distribution ([14]) with $N-1$ degrees of freedom ((positive) square root of a sum of $N-1$ normally i.i.d. variables). As, 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 $k \in \{3, \ldots, N\}$, it is the (positive) square root of a sum of squared normally distributed random variables. Therefore, $(r/\sigma^{(g)})^2 \sim \chi^2_{N-1,(r^{(g)}/\sigma^{(g)})^2}$ and consequently $r \sim \sigma^{(g)} \chi_{N-1, r^{(g)}/\sigma^{(g)}}$. The expressions $\chi^2_{df,nc}$ and $\chi_{df,nc}$ denote the $\chi^2$ and $\chi$ distributions, respectively, with $df$ degrees of freedom and non-centrality parameter $nc$. The cumulative distribution function and the probability density function of this scaled distribution can be expressed using the corresponding functions of the non-central $\chi$ distribution. The cumulative distribution function of the non-central $\chi$ distribution is denoted by $P_{\chi_{df,nc}}$. And the corresponding probability density function is denoted by $p_{\chi_{df,nc}}$. For the cumulative distribution function the scaled version reads $P_{\sigma^{(g)}\chi_{df,nc}}(x) = P_{\chi_{df,nc}}\left(\frac{x}{\sigma^{(g)}}\right)$. And for the probability density function the scaled version reads $p_{\sigma^{(g)}\chi_{df,nc}}(x) = \frac{1}{\sigma^{(g)}} p_{\chi_{df,nc}}\left(\frac{x}{\sigma^{(g)}}\right)$. Hence,

$$p_r(r) = \frac{1}{\sigma^{(g)}} p_{\chi_{N-1, r^{(g)}/\sigma^{(g)}}}\left(\frac{r}{\sigma^{(g)}}\right) \tag{22}$$

follows. In order to get tractable expressions in analysis steps that follow, $p_r(r)$ is approximated by a normal distribution

$$p_r(r) \approx \frac{1}{\sigma_r}\phi\left(\frac{r-\bar{r}}{\sigma_r}\right) = \frac{1}{\sqrt{2\pi}\sigma_r}\exp\left[-\frac{1}{2}\left(\frac{r-\bar{r}}{\sigma_r}\right)^2\right]. \tag{23}$$

The detailed derivation of the mean $\bar{r}$ and the standard deviation $\sigma_r$ is presented in Appendix B. Now that expressions for $p_x(x)$ and $p_r(r)$ have been derived, the exact $P_Q$ expression can be treated further to arrive at closed-form $P_Q$ approximations. Those derivations are described in detail in Appendix C.

*The $P_Q(q)$ Approximation.* Taking the upper bound from Eq. (C.15) for the case that the probability of generating feasible offspring tends to 0 and Eq. (C.24) for the other case, this allows for the definition of the approximated $P_Q(q)$ function

$$P_Q(q) \approx \begin{cases} P_{Q\text{feas}}(q) := \Phi\left(\dfrac{q - x^{(g)}}{\sigma^{(g)}}\right), \text{ for } q > \bar{r}\sqrt{\xi} & (24) \\ P_{Q\text{infeas}}(q) := \Phi\left(\dfrac{(1+1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{{\sigma^{(g)}}^2 + \sigma_r^2/\xi}}\right), \text{ otherwise.} & (25) \end{cases}$$

Taking the derivative with respect to $q$ yields

$$p_Q(q) \approx \begin{cases} p_{Q\text{feas}}(q) = \dfrac{1}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2}\left(\frac{q - x^{(g)}}{\sigma^{(g)}}\right)^2}, \text{ for } q > \bar{r}\sqrt{\xi} & (26) \\ p_{Q\text{infeas}}(q) = \left(\dfrac{(1+1/\xi)}{\sqrt{{\sigma^{(g)}}^2 + \sigma_r^2/\xi}}\right) & \\ \times \dfrac{1}{\sqrt{2\pi}} \exp\left[-\dfrac{1}{2}\left(\dfrac{(1+1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{{\sigma^{(g)}}^2 + \sigma_r^2/\xi}}\right)^2\right], & (27) \\ \text{otherwise.} & \end{cases}$$

They represent approximations for the cases that the offspring generation step leads to feasible offspring with high probability and that the offspring generation step leads to infeasible offspring with high probability, respectively. Because those are disjoint events in the limit case, they complement each other.

Now, the expected value defined in Eq. (15) can be approximated for both cases. They are denoted as $\mathrm{E}[q_{1;\lambda_{\text{feas}}}]$ and $\mathrm{E}[q_{1;\lambda_{\text{infeas}}}]$, respectively. Derivations for $\mathrm{E}[q_{1;\lambda_{\text{feas}}}]$ and $\mathrm{E}[q_{1;\lambda_{\text{infeas}}}]$ are presented in Appendices D and E, respectively. Consequently, approximate expressions for the progress rate follow with insertion into Eq. (13) for both cases.

*The Approximate $x$ Progress Rate in the case that the probability of feasible offspring tends to* 1. With Eq. (D.5), the asymptotic normalized $x$ progress rate for the case of feasible offspring reads

$$\varphi_{x\text{feas}}^* = \frac{N\left(x^{(g)} - \mathrm{E}[q_{1;\lambda_{\text{feas}}}]\right)}{x^{(g)}} = \frac{N\left(x^{(g)} - x^{(g)} + \sigma^{(g)} c_{1,\lambda}\right)}{x^{(g)}} \tag{28}$$

$$= \frac{N}{x^{(g)}}\sigma^{(g)} c_{1,\lambda} = \frac{N}{x^{(g)}}{\sigma^{(g)}}^* \frac{r^{(g)}}{N} c_{1,\lambda} = \frac{r^{(g)}}{x^{(g)}}{\sigma^{(g)}}^* c_{1,\lambda}. \tag{29}$$

The coefficient $c_{1,\lambda}$ has been introduced in [15] as the so-called progress coefficient. It is also defined in [10, Equation 3.100]. Its definition reads

$$c_{1,\lambda} := \frac{\lambda}{\sqrt{2\pi}} \int_{t=-\infty}^{t=\infty} t e^{-\frac{1}{2}t^2} [\Phi(t)]^{\lambda-1} \, \mathrm{d}t. \tag{30}$$

*The Approximate x Progress Rate in the case that the probability of feasible offspring tends to* 0. The normalized $x$ progress rate for the second case can be expressed using Eq. (E.9)

$$\varphi_{x\,\text{infeas}}^{*} = \frac{N\left(x^{(g)} - \mathrm{E}[q_{1;\lambda_{\text{infeas}}}]\right)}{x^{(g)}} \tag{31}$$

$$= \frac{N}{x^{(g)}}\left(\frac{1+\xi}{1+\xi}x^{(g)} - \mathrm{E}[q_{1;\lambda_{\text{infeas}}}]\right) \tag{32}$$

$$\approx \frac{N}{x^{(g)}}\left[\frac{x^{(g)} + \xi x^{(g)} - \xi x^{(g)}}{1+\xi} - \frac{\sqrt{\xi}\bar{r}}{1+\xi} + \frac{\xi}{1+\xi}\left(\sqrt{{\sigma^{(g)}}^2 + \sigma_r^2/\xi}\right)c_{1,\lambda}\right]. \tag{33}$$

This can further be simplified using $\sigma_r$ of the normal approximation (Eq. (B.6)) together with $\sigma$-normalization yielding

$$\begin{aligned}\varphi_{x\,\text{infeas}}^{*} = \frac{N}{1+\xi}&\left[1 - \sqrt{\xi}\frac{r^{(g)}}{x^{(g)}}\sqrt{1 + \frac{{\sigma^{(g)}}^{*2}}{N}\left(1 - \frac{1}{N}\right)}\right.\\ &\left.+\xi\frac{r^{(g)}}{x^{(g)}}\sqrt{\frac{{\sigma^{(g)}}^{*2}}{N^2} + \frac{1}{\xi}\frac{{\sigma^{(g)}}^{*2}}{N^2}\frac{1 + \frac{{\sigma^{(g)}}^{*2}}{2N}(1-\frac{1}{N})}{1 + \frac{{\sigma^{(g)}}^{*2}}{N}(1-\frac{1}{N})}}c_{1,\lambda}\right]\end{aligned} \tag{34}$$

$$\begin{aligned}\simeq \frac{N}{1+\xi}&\left[1 - \sqrt{\xi}\frac{r^{(g)}}{x^{(g)}}\sqrt{1 + \frac{{\sigma^{(g)}}^{*2}}{N}}\right.\\ &\left.+\xi\frac{r^{(g)}}{x^{(g)}}\frac{{\sigma^{(g)}}^{*}}{N}\sqrt{1 + \frac{1}{\xi}\frac{1 + \frac{{\sigma^{(g)}}^{*2}}{2N}}{1 + \frac{{\sigma^{(g)}}^{*2}}{N}}}c_{1,\lambda}\right].\end{aligned} \tag{35}$$

From Eq. (34) to Eq. (35), $\frac{1}{N}$ has been neglected compared to 1 as $N \to \infty$. It can be rewritten further yielding

$$\begin{aligned}\varphi_{x\,\text{infeas}}^{*} = \frac{N}{1+\xi}&\left(1 - \frac{\sqrt{\xi}r^{(g)}}{x^{(g)}}\sqrt{1 + \frac{{\sigma^{(g)}}^{*2}}{N}}\right)\\ &+ \frac{\sqrt{\xi}}{1+\xi}\frac{\sqrt{\xi}r^{(g)}}{x^{(g)}}{\sigma^{(g)}}^{*}c_{1,\lambda}\sqrt{1 + \frac{1}{\xi}\frac{1 + \frac{{\sigma^{(g)}}^{*2}}{2N}}{1 + \frac{{\sigma^{(g)}}^{*2}}{N}}}.\end{aligned} \tag{36}$$

*The Approximate x Progress Rate - Combination Using the Best Offspring's Feasibility Probability.* In order to get an approximate combined value $\varphi_x^*$, the expected values of both cases are weighted by their probability of occurrence. Those weighted values are then added yielding

$$\varphi_x^* \approx P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})\varphi_{x\,\text{feas}}^{*} + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})]\varphi_{x\,\text{infeas}}^{*}. \tag{37}$$

The simplified version is applicable if the probability of generating feasible offspring tends to 1 in the asymptotic case. In the opposite case, the result derived

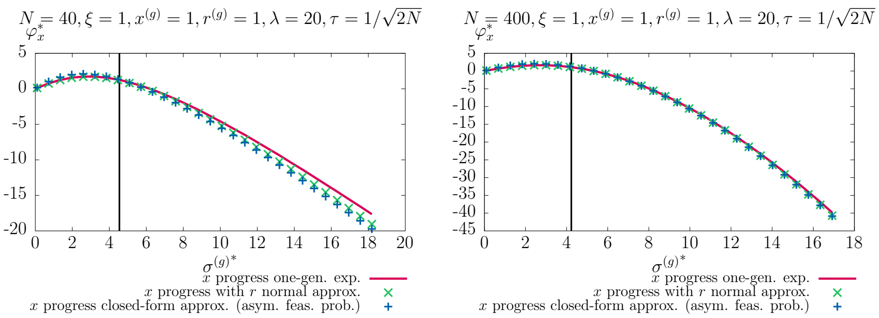


Figure 4: Comparison of the $x$ progress rate approximation with simulations.

from the $P_Q(q)$ upper bound is used. For the approximated combined value, $P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})$ is considered to denote the probability that the best offspring after projection has been feasible before projection. A derivation for an approximation of $P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})$ is shown in Appendix F. Insertion of Eqs. (29) and (36) into Eq. (37) finally results in the approximate normalized $x$ progress rate

$$\begin{aligned}\varphi_x^* \approx\ & P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \frac{r^{(g)}}{x^{(g)}} {\sigma^{(g)}}^* c_{1,\lambda} \\ & + \left(1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})\right) \left[ \frac{N}{1+\xi} \left( 1 - \frac{\sqrt{\xi} r^{(g)}}{x^{(g)}} \sqrt{1 + \frac{{\sigma^{(g)}}^{*2}}{N}} \right) \right. \\ & \left. + \frac{\sqrt{\xi}}{1+\xi} \frac{\sqrt{\xi} r^{(g)}}{x^{(g)}} {\sigma^{(g)}}^* c_{1,\lambda} \sqrt{1 + \frac{1}{\xi} \frac{1 + \frac{{\sigma^{(g)}}^{*2}}{2N}}{1 + \frac{{\sigma^{(g)}}^{*2}}{N}}} \right].\end{aligned} \tag{38}$$

Fig. 4 shows two example plots comparing the derived closed-form approximation for $\varphi_x^*$ with one-generation experiments and numerical integration results. For more detailed plots it is referred to [9, Fig. 3.5, pp. 50-52]. The solid line has been generated by one-generation experiments. The crosses have been computed by numerical integration (using the normal approximation for $r$). For this computation, Eqs. (13) and (15) have been numerically computed and subsequently normalized using Eqs. (17), (20), (21), and (23). The pluses have been calculated by evaluating Eq. (38) with Eq. (F.6). The vertical black line indicates the value of the normalized mutation strength in the steady state for the given parameters. It has been calculated using Eq. (73) (see Section 4.3.2 for steady state investigations). As one can see, the approximation quality of $\varphi_x^*$ in the vicinity of the steady state is rather good. Therefore, it can be used to investigate the dynamical behavior of the ES in the vicinity of the steady state.

*4.2.2. Derivation of the r Progress Rate*

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

$$\varphi_r(x^{(g)}, r^{(g)}, \sigma^{(g)}) = r^{(g)} - \mathrm{E}[r^{(g+1)} \,|\, x^{(g)}, r^{(g)}, \sigma^{(g)}] \tag{39}$$

$$= r^{(g)} - \mathrm{E}[q_{r\,1;\lambda} \,|\, x^{(g)}, r^{(g)}, \sigma^{(g)}]. \tag{40}$$

This means that the expectation of the $r$ value after projection of the best offspring is needed for the derivation of the progress rate in $r$ direction

$$\mathrm{E}[q_{r\,1;\lambda} \,|\, x^{(g)}, r^{(g)}, \sigma^{(g)}] := \mathrm{E}[q_{r\,1;\lambda}].$$

Its derivation is presented in Appendix G. Similar to Section 4.2.1 for the $x$ progress rate, the normalized $r$ progress rate can now be formulated. Using the derived approximate relation $\mathrm{E}[q_{r\,1;\lambda}] \approx \frac{1}{\sqrt{\xi}}\mathrm{E}[q_{1;\lambda}]$ (Eq. (G.15)) it reads

$$\varphi_{r_{\text{infeas}}}^{*} = N\frac{\left(r^{(g)} - \mathrm{E}[q_{r\,1;\lambda_{\text{infeas}}}]\right)}{r^{(g)}} \approx N\frac{\left(r^{(g)} - \frac{1}{\sqrt{\xi}}\mathrm{E}[q_{1;\lambda_{\text{infeas}}}]\right)}{r^{(g)}}. \tag{41}$$

Using Eq. (E.9), Eq. (41) can further be rewritten using $\sigma_r \simeq \sigma^{(g)}$ and $\bar{r} \simeq r^{(g)}\sqrt{1 + \frac{\sigma^{(g)*2}}{N}}$ from Eq. (B.6) for $N \to \infty$ to get from Eq. (43) to Eq. (44) resulting in

$$\varphi_{r_{\text{infeas}}}^{*} = N\left(1 - \frac{1}{\sqrt{\xi}r^{(g)}}\left[\frac{\xi}{1+\xi}\left(x^{(g)} + \bar{r}/\sqrt{\xi}\right) - \frac{\xi}{1+\xi}\left(\sqrt{{\sigma^{(g)}}^2 + \sigma_r^2/\xi}\right)c_{1,\lambda}\right]\right) \tag{42}$$

$$= N\left(1 - \frac{\sqrt{\xi}}{(1+\xi)}\left[\frac{x^{(g)}}{r^{(g)}} + \frac{\bar{r}}{r^{(g)}\sqrt{\xi}} - \left(\frac{\sqrt{{\sigma^{(g)}}^2 + \sigma_r^2/\xi}}{r^{(g)}}\right)c_{1,\lambda}\right]\right) \tag{43}$$

$$\simeq N\left(1 - \frac{\sqrt{\xi}}{(1+\xi)}\left[\frac{x^{(g)}}{r^{(g)}} + \frac{\sqrt{1 + {\sigma^{(g)*}}^2/N}}{\sqrt{\xi}} - \left(\frac{\sigma^{(g)}}{r^{(g)}}\sqrt{1 + 1/\xi}\right)c_{1,\lambda}\right]\right). \tag{44}$$

The progress coefficient $c_{1,\lambda}$ (see Eq. (30)) has been used.

*The Approximate r Progress Rate in the case that the probability of feasible offspring tends to* 1. If the offspring is feasible almost surely in the asymptotic $N \to \infty$ case, Eq. (G.6) can be simplified further. The complete probability mass lies inside the cone. Consequently, the second summand vanishes. Additionally, the integral in the first summand yields the expected $r$ value $\bar{r}$ because the bounds indicate the integration over the whole feasible region for the given area d$q$. In the feasible case, $I(q)$ therefore reads $I_{\text{feas}}(q) = p_x(q)\bar{r}$. By insertion of $I_{\text{feas}}(q)$ as $I(q)$ into Eq. (G.4), use of $P_{Q_{\text{feas}}}$ from Eq. (24), and use of $p_x(x)$ from Eq. (21) one obtains

$$\mathrm{E}[q_{r1;\lambda_{\text{feas}}}] \approx \bar{r}\,\underbrace{\lambda \int_{q=0}^{q=\infty} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)^2} \left[1-\Phi\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)\right]^{\lambda-1} \mathrm{d}q}_{=1} = \bar{r}. \tag{45}$$

Similar to Section 4.2.1 for the $x$ progress rate, the normalized $r$ progress rate can now be formulated. It reads

$$\varphi_{r_{\text{feas}}}^{*} = N\frac{\left(r^{(g)} - \mathrm{E}[q_{r1;\lambda_{\text{feas}}}]\right)}{r^{(g)}} \approx N\frac{\left(r^{(g)} - \bar{r}\right)}{r^{(g)}} \simeq N\frac{\left(r^{(g)} - r^{(g)}\sqrt{1+\frac{{\sigma^{(g)}}^{*2}}{N}}\right)}{r^{(g)}} \tag{46}$$

$$= N\left(1-\sqrt{1+\frac{{\sigma^{(g)}}^{*2}}{N}}\right) \simeq N - N - \frac{N{\sigma^{(g)}}^{*2}}{2N} = -\frac{{\sigma^{(g)}}^{*2}}{2}. \tag{47}$$

In Eq. (47), Taylor expansion of the square root and cutoff after the linear term (with subsequent distribution of $N$ over the terms in the parentheses) have been performed.

*The Approximate r Progress Rate - Combination Using the Best Offspring's Feasibility Probability.* Analogously to Eq. (37), both cases are combined into

$$\varphi_r^* \approx P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})\varphi_{r_{\text{feas}}}^* + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})]\varphi_{r_{\text{infeas}}}^*. \tag{48}$$

Insertion of Eqs. (44) and (47) into Eq. (48) finally results in the approximate normalized $r$ progress rate

$$\begin{aligned}\varphi_r^* \approx &\left(1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})\right)\\ &\times N\Bigg\{1 - \frac{\sqrt{\xi}}{(1+\xi)}\Bigg[\left(\frac{x^{(g)}}{r^{(g)}} + \frac{\sqrt{1+{\sigma^{(g)}}^{*2}/N}}{\sqrt{\xi}}\right)\\ &\qquad - \left(\frac{\sigma^{(g)}}{r^{(g)}}\sqrt{1+1/\xi}\right)c_{1,\lambda}\Bigg]\Bigg\}\\ &- P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})\frac{{\sigma^{(g)}}^{*2}}{2}.\end{aligned} \tag{49}$$

Fig. 5 shows two example plots comparing the derived closed-form approximation for $\varphi_r^*$ with one-generation experiments and numerical integration results. For more detailed plots it is referred to [9, Fig. 3.6, pp. 59-61]. The solid line has been generated by one-generation experiments. The crosses have been computed by numerical integration (using the normal approximation for $r$). For

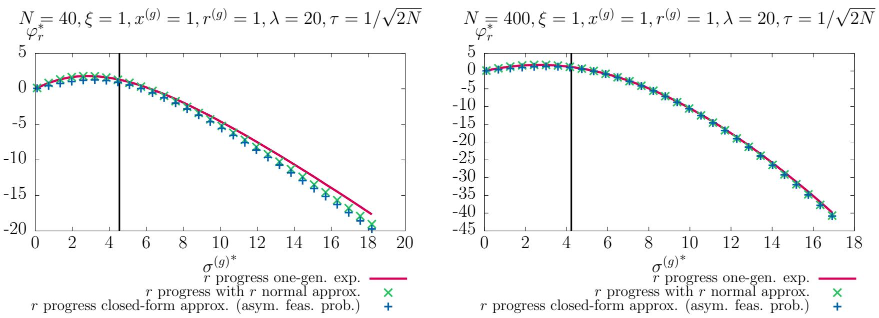


Figure 5: Comparison of the $r$ progress rate approximation with simulations.

this computation, Term (40) and Eq. (G.2) have been numerically computed and subsequently normalized using Eqs. (17), (20), (21), and (23). The pluses have been calculated by evaluating Eq. (48) with Eqs. (41), (47), and (F.6). The vertical black line indicates the value of the normalized mutation strength in the steady state for the given parameters. It has been calculated using Eq. (73) (see Section 4.3.2 for steady state investigations).

*4.2.3. Derivation of the SAR*

From the definition of the SAR (Eq. (11)) and the pseudo-code of the ES (Alg. 1, Line 18) it follows that

$$\psi(x^{(g)}, r^{(g)}, \sigma^{(g)}) = \mathrm{E}\left[\frac{\sigma^{(g+1)} - \sigma^{(g)}}{\sigma^{(g)}} \,\middle|\, x^{(g)}, r^{(g)}, \sigma^{(g)}\right] \tag{50}$$

$$= \mathrm{E}\left[\frac{\tilde{\sigma}_{1;\lambda} - \sigma^{(g)}}{\sigma^{(g)}} \,\middle|\, x^{(g)}, r^{(g)}, \sigma^{(g)}\right] \tag{51}$$

$$= \int_{\sigma=0}^{\sigma=\infty} \left(\frac{\sigma - \sigma^{(g)}}{\sigma^{(g)}}\right) p_{\tilde{\sigma}_{1;\lambda}}(\sigma)\,\mathrm{d}\sigma \tag{52}$$

where $p_{\tilde{\sigma}_{1;\lambda}}(\sigma) := p_{\tilde{\sigma}_{1;\lambda}}(\sigma \,|\, x^{(g)}, r^{(g)}, \sigma^{(g)})$ denotes the probability density function of the best offspring's mutation strength. Note that $\tilde{\sigma}_{1;\lambda}$ is not obtained by direct selection. It is the $\sigma$ value of the individual with the best objective function value that is selected among the $\lambda$ offspring. This probability density function can be derived with a similar argument from the area of order statistics as in Section 4.2.1 for the best $q$ value. It reads

$$p_{\tilde{\sigma}_{1;\lambda}}(\sigma) = p_\sigma(\sigma \,|\, \sigma^{(g)})\lambda \int_{q=0}^{q=\infty} p_Q(q \,|\, x^{(g)}, r^{(g)}, \sigma)[1 - P_Q(q)]^{\lambda-1}\,\mathrm{d}q \tag{53}$$

where $p_\sigma(\sigma \,|\, \sigma^{(g)})$ denotes the log-normal probability density function. This result can now be inserted into Eq. (52) resulting in

$$\psi(x^{(g)}, r^{(g)}, \sigma^{(g)}) = \int_{\sigma=0}^{\sigma=\infty} \left(\frac{\sigma - \sigma^{(g)}}{\sigma^{(g)}}\right) p_\sigma(\sigma \,|\, \sigma^{(g)}) \times \lambda \int_{q=0}^{q=\infty} p_Q(q \,|\, x^{(g)}, r^{(g)}, \sigma)[1 - P_Q(q)]^{\lambda-1} \,\mathrm{d}q \,\mathrm{d}\sigma. \tag{54}$$

Due to difficulties in analytically solving Eq. (54), an approximate solution is derived. To this end, the approach used in [10, Section 7.3.2.4] is followed. For this, Eq. (54) is written as

$$\psi(x^{(g)}, r^{(g)}, \sigma^{(g)}) = \int_{\sigma=0}^{\sigma=\infty} f(\sigma) p_\sigma(\sigma \,|\, \sigma^{(g)}) \,\mathrm{d}\sigma \tag{55}$$

$$= \int_{\sigma=0}^{\sigma=\infty} f(\sigma) \frac{1}{\sqrt{2\pi}\tau} \frac{1}{\sigma} \exp\left[-\frac{1}{2}\left(\frac{\ln(\sigma/\sigma^{(g)})}{\tau}\right)^2\right] \mathrm{d}\sigma \tag{56}$$

with

$$f(\sigma) = \left(\frac{\sigma - \sigma^{(g)}}{\sigma^{(g)}}\right) \lambda \int_{q=0}^{q=\infty} p_Q(q \,|\, x^{(g)}, r^{(g)}, \sigma)[1 - P_Q(q)]^{\lambda-1} \,\mathrm{d}q. \tag{57}$$

Now, $\tau$ is assumed to be small. Therefore, Eq. (56) can be expanded into a Taylor series at $\tau = 0$. Further, as $\tau$ 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 [10, Section 7.3.2.4] for all the details) one obtains

$$\psi(x^{(g)}, r^{(g)}, \sigma^{(g)}) = f(\sigma^{(g)}) + \frac{\tau^2}{2}\sigma^{(g)} \left.\frac{\partial f}{\partial \sigma}\right|_{\sigma=\sigma^{(g)}} + \frac{\tau^2}{2}{\sigma^{(g)}}^2 \left.\frac{\partial^2 f}{\partial \sigma^2}\right|_{\sigma=\sigma^{(g)}} + O(\tau^4). \tag{58}$$

To proceed further, the expressions $f(\sigma^{(g)})$, $\left.\frac{\partial f}{\partial \sigma}\right|_{\sigma=\sigma^{(g)}}$, and $\left.\frac{\partial^2 f}{\partial \sigma^2}\right|_{\sigma=\sigma^{(g)}}$ need to be evaluated for the feasible and the infeasible cases. These calculations are described in more detail in Appendix H.

*The Approximate SAR in the case that the probability of feasible offspring tends to* 0. The approximate SAR expression can be expressed for the infeasible case by insertion of Eqs. (H.2), (H.4), and (H.16) into Eq. (58). It yields

$$\psi_{\text{infeas}} \approx 0 + \frac{\tau^2}{2} + \frac{\tau^2}{2}{\sigma^{(g)}}^2 \left[\frac{2}{{\sigma^{(g)}}^2}\left(d^{(2)}_{1,\lambda} - 1\right) - \frac{2}{{\sigma^{(g)}}^2} \frac{\sigma^{(g)} N}{\sqrt{1 + \frac{1}{\xi}}\sqrt{\xi} r^{(g)}} c_{1,\lambda}\right] \tag{59}$$

$$= \tau^2 \left[\left(d^{(2)}_{1,\lambda} - \frac{1}{2}\right) - \frac{c_{1,\lambda}{\sigma^{(g)}}^*}{\sqrt{1 + \xi}}\right] \tag{60}$$

where the definition of the higher-order progress coefficients [10, Equation 4.41]

$$d_{1,\lambda}^{(k)} := \frac{\lambda}{\sqrt{2\pi}} \int_{t=-\infty}^{t=\infty} t^k e^{-\frac{1}{2}t^2} [\Phi(t)]^{\lambda-1} \, dt \tag{61}$$

and the definition of $c_{1,\lambda}$ from Eq. (30) have been used.

*The Approximate SAR in the case that the probability of feasible offspring tends to* 1. The approximate SAR expression can be expressed for the feasible case by insertion of Eqs. (H.2), (H.4), and (H.21) into Eq. (58). It yields

$$\psi_{\text{feas}} \approx 0 + \frac{\tau^2}{2} + \frac{\tau^2}{2} {\sigma^{(g)}}^2 \frac{2}{{\sigma^{(g)}}^2} \left( d_{1,\lambda}^{(2)} - 1 \right) = \tau^2 \left( d_{1,\lambda}^{(2)} - \frac{1}{2} \right). \tag{62}$$

*The Approximate SAR - Combination Using the Best Offspring's Feasibility Probability.* Similarly to Eq. (37), both cases are combined into

$$\psi \approx P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \psi_{\text{feas}} + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] \psi_{\text{infeas}}. \tag{63}$$

Insertion of Eqs. (60) and (62) into Eq. (63) finally results in the approximate SAR

$$\psi \approx \tau^2 \left[ \left( d_{1,\lambda}^{(2)} - \frac{1}{2} \right) - \left( 1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \right) \frac{c_{1,\lambda} {\sigma^{(g)}}^*}{\sqrt{1+\xi}} \right]. \tag{64}$$

Fig. 6 shows two example plots comparing the derived closed-form approximation for $\psi$ to one-generation experiments and numerical integration results. For more detailed plots it is referred to [9, Fig. 3.7, pp. 70-72]. The solid line has been generated by one-generation experiments. The crosses have been computed by numerical integration (using the normal approximation for $r$). For this computation, Eq. (52) has been numerically computed using Eqs. (17), (18), (20), (21), and (23). The pluses have been calculated by evaluating Eq. (64) with Eq. (F.6). The vertical black line indicates the value of the normalized mutation strength in the steady state for the given parameters. It has been calculated using Eq. (73) (see Section 4.3.2 for steady state investigations). Again, the approximation quality in the vicinity of the steady state is acceptable. However, due to the linear approximation used, $\psi$ must necessarily deviate for sufficiently large $\sigma^*$.

### *4.3. The Evolution Dynamics in the Deterministic Approximation*

#### *4.3.1. The Evolution Equations*

As already explained in Section 4.1, the evolution equations (Eqs. (6) to (8)) are difficult to treat analytically. Therefore, approximations are usually used. One such approximation is to ignore the stochastic fluctuations and only use the mean terms, i.e., using Eqs. (6) to (8) with $\epsilon_x = 0$, $\epsilon_r = 0$, and $\epsilon_\sigma = 0$.

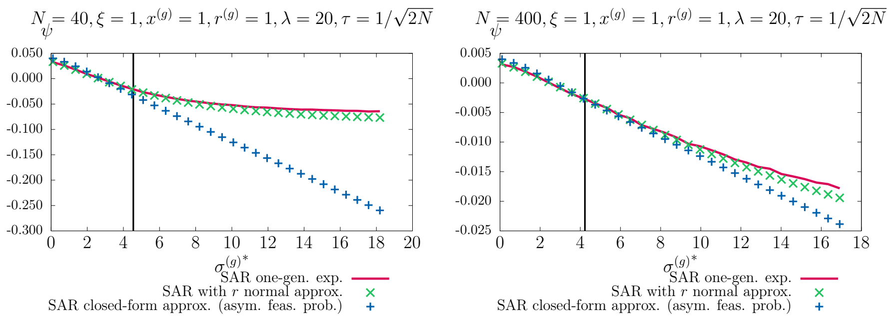


Figure 6: Comparison of the SAR approximation with simulations.

Expressing the $x$ and $r$ evolution equations using the respective normalized variants, the deterministic evolution equations read

$$x^{(g+1)} = x^{(g)} - \frac{x^{(g)}\varphi_x^{(g)*}}{N} = x^{(g)}\left(1 - \frac{\varphi_x^{(g)*}}{N}\right) \tag{65}$$

$$r^{(g+1)} = r^{(g)} - \frac{r^{(g)}\varphi_r^{(g)*}}{N} = r^{(g)}\left(1 - \frac{\varphi_r^{(g)*}}{N}\right) \tag{66}$$

$$\sigma^{(g+1)} = \sigma^{(g)} + \sigma^{(g)}\psi^{(g)} = \sigma^{(g)}\left(1 + \psi^{(g)}\right). \tag{67}$$

Eqs. (65) to (67) represent the so-called mean value iterative system. It can be iterated starting with initial $x^{(0)}$, $r^{(0)}$, and $\sigma^{(0)}$ values to predict the actual values of real ES runs. Fig. 7 shows the mean value dynamics of the $(1,\lambda)$-ES applied to the conically constrained problem for $N = 40$. For more detailed plots it is referred to [9, Fig. 3.8, pp. 74-85]. The plots show that the approximation deviates from the real dynamics for small $N$ (e.g., $N = 40$). The deviations from the real runs get smaller with larger $N$ (e.g., $N = 400$).

### *4.3.2. The ES in the Stationary State*

For constant exogenous parameters the state of the $(1,\lambda)$-ES on the conically constrained problem is completely described by $(x^{(g)}, r^{(g)}, \sigma^{(g)})^T$. The stationary state (also called steady state) is the state obtained for sufficiently large $g$, i.e., for large time scales. A correct working ES is expected to steadily move towards the optimizer. Consequently, for sufficiently large $g$, the normalized mutation strength is expected to be constant. In other words, the steady state normalized mutation strength $\sigma^*_{ss}$ is expected to be constant. More formally this reads $\sigma^*_{ss} := \lim_{g\to\infty} \sigma^{(g)*}$. This means that for sufficiently large $g$, $\sigma^{(g)*} = \sigma^{(g+1)*} = \sigma^*_{ss}$ should hold. In order to proceed further, Eq. (67) is expressed in normalized terms $\frac{r^{(g+1)}\sigma^{(g+1)*}}{N} = \frac{r^{(g)}\sigma^{(g)*}}{N}\left(1 + \psi^{(g)}\right)$. Requiring that $\sigma^{(g)*} = \sigma^{(g+1)*}$ one obtains further $r^{(g+1)} = r^{(g)}\left(1 + \psi^{(g)}\right)$. Insertion of

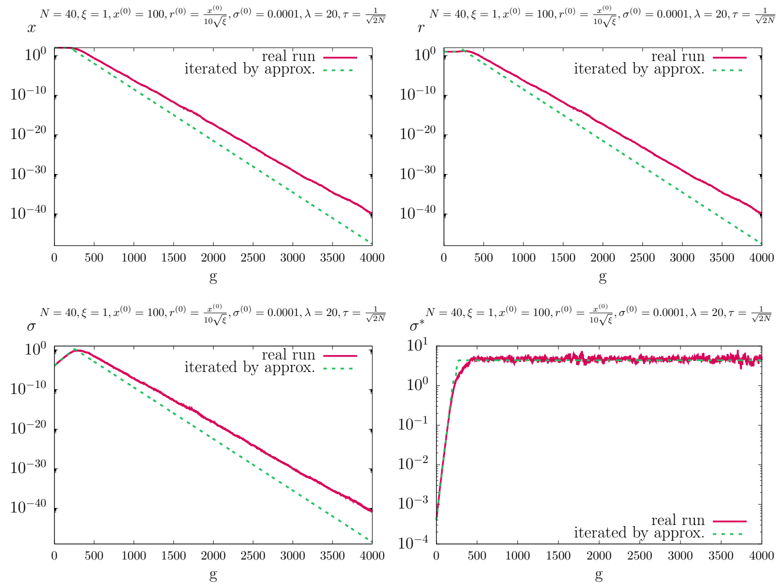


Figure 7: Mean value dynamics closed-form approximation and real-run comparison of the $(1, \lambda)$-ES with repair by projection applied to the conically constrained problem ($N = 40$).

Eq. (66) results in

$$r^{(g)}\left(1 - \frac{{\varphi_r^{(g)}}^*}{N}\right) = r^{(g)}\left(1 + \psi^{(g)}\right) \tag{68}$$

$$\frac{{\varphi_r^{(g)}}^*}{N} = -\psi^{(g)}. \tag{69}$$

For the infeasible case in the vicinity of the cone boundary, i.e., if $P_{\text{feas}} \approx 0$ and $\frac{x^{(g)}}{\sqrt{\xi}r^{(g)}} \approx 1$, one obtains using the derived approximate relation $\mathrm{E}[q_{r\,1;\lambda}] \approx \frac{1}{\sqrt{\xi}}\mathrm{E}[q_{1;\lambda}]$ (Eq. (G.15)) and Eq. (31)

$$\varphi_r^* = N\left(1 - \frac{1}{r^{(g)}}\mathrm{E}[q_{r\,1;\lambda}]\right) \approx N\left(1 - \frac{1}{r^{(g)}}\frac{1}{\sqrt{\xi}}\mathrm{E}[q_{1;\lambda}]\right) \tag{70}$$

$$= N\left(1 - \frac{x^{(g)}}{\sqrt{\xi}r^{(g)}}\left(1 - \frac{\varphi_x^*}{N}\right)\right) \approx N\left(1 - \left(1 - \frac{\varphi_x^*}{N}\right)\right) = \varphi_x^*. \tag{71}$$

Considering this case further ($P_{\text{feas}} \approx 0$ and $\frac{x^{(g)}}{\sqrt{\xi}r^{(g)}} \approx 1$), use of Eq. (36) and Eq. (64) together with Eq. (69) and Eq. (71) results after simplification for the asymptotic case $N \to \infty$ in

$$\frac{1 - \sqrt{1 + \frac{{\sigma^{(g)}}^{*2}}{N}}}{1 + \xi} + \frac{{\sigma^{(g)}}^* c_{1,\lambda}}{N\sqrt{1 + \xi}} = -\tau^2\left(d_{1,\lambda}^{(2)} - \frac{1}{2}\right) + \tau^2\frac{c_{1,\lambda}{\sigma^{(g)}}^*}{\sqrt{1 + \xi}}. \tag{72}$$

This leads to a quadratic equation. Solving it is not particularly difficult but long (use of a computer algebra system (CAS) can be advantageous). Hence, only the solution is given

$$\sigma_{ss}{}^* \approx \frac{c_{1,\lambda}\sqrt{1+\xi}(1-N\tau^2)(1+(d^{(2)}_{1,\lambda}-\frac{1}{2})(1+\xi)\tau^2)}{1-\frac{1+\xi}{N}c^2_{1,\lambda}(1-N\tau^2)^2} + \frac{\sqrt{(1+\xi)(c^2_{1,\lambda}(1-N\tau^2)^2+N\tau^2(d^{(2)}_{1,\lambda}-\frac{1}{2})(2+(d^{(2)}_{1,\lambda}-\frac{1}{2})(1+\xi)\tau^2))}}{1-\frac{1+\xi}{N}c^2_{1,\lambda}(1-N\tau^2)^2}. \tag{73}$$

Since $\sigma^* \geq 0$ does hold, only the positive root has been chosen.

Aiming for a simpler form, assuming further $N \gg {\sigma^{(g)}}^2$ allows simplifying Eq. (36) (using Taylor expansion and cutoff after the linear term for the first square root and assuming $\frac{1+\frac{{\sigma^{(g)}}^{*2}}{2N}}{1+\frac{{\sigma^{(g)}}^{*2}}{N}} \approx 1$) to

$$\varphi_x{}^* \simeq \frac{N}{1+\xi}\left(1-1-\frac{{\sigma^{(g)}}^{*2}}{2N}\right)+\frac{\sqrt{\xi}}{1+\xi}{\sigma^{(g)}}^* c_{1,\lambda}\sqrt{1+\frac{1}{\xi}} \tag{74}$$

$$= \frac{c_{1,\lambda}}{\sqrt{1+\xi}}{\sigma^{(g)}}^* - \frac{{\sigma^{(g)}}^{*2}}{2(1+\xi)} = c_{1,\lambda}\left(\frac{{\sigma^{(g)}}^*}{\sqrt{1+\xi}}\right)-\frac{1}{2}\left(\frac{{\sigma^{(g)}}^*}{\sqrt{(1+\xi)}}\right)^2. \tag{75}$$

Inserting this into the derived steady state condition Eq. (69) yields using Eqs. (64) and (75)

$$\varphi_x{}^* = -N\psi \tag{76}$$

$$c_{1,\lambda}\left(\frac{\sigma^*_{ss}}{\sqrt{1+\xi}}\right)-\frac{1}{2}\left(\frac{\sigma^*_{ss}}{\sqrt{(1+\xi)}}\right)^2 = -N\tau^2\left[\left(d^{(2)}_{1,\lambda}-\frac{1}{2}\right)-\frac{c_{1,\lambda}\sigma^*_{ss}}{\sqrt{1+\xi}}\right] \tag{77}$$

$$c_{1,\lambda}\sigma^*_{ss}-\frac{1}{2}{\sigma^*_{ss}}^2\frac{\sqrt{(1+\xi)}}{1+\xi} = -N\tau^2\left(d^{(2)}_{1,\lambda}-\frac{1}{2}\right)\sqrt{1+\xi} + N\tau^2 c_{1,\lambda}\sigma^*_{ss} \tag{78}$$

$$\frac{1}{2}{\sigma^*_{ss}}^2\frac{1}{\sqrt{(1+\xi)}}+\sigma^*_{ss}(N\tau^2 c_{1,\lambda}-c_{1,\lambda}) = N\tau^2\left(d^{(2)}_{1,\lambda}-\frac{1}{2}\right)\sqrt{1+\xi}. \tag{79}$$

Solving this quadratic equation results in

$$\sigma_{ss}{}^* \approx \left[c_{1,\lambda}(1-N\tau^2)+\sqrt{c^2_{1,\lambda}(1-N\tau^2)^2+N\tau^2\left(2d^{(2)}_{1,\lambda}-1\right)}\right]\sqrt{1+\xi}. \tag{80}$$

Here, only the positive root has been chosen since $\sigma^* \geq 0$ does hold. This is a remarkable result. One recognizes the equations for the sphere model (see [10, Page 301, Equation 7.171]), i.e., $\sigma^*_{ss} = \sqrt{1+\xi}\sigma^*_{ss_{\text{sphere}}}$. Additionally, insertion of $\sigma^*_{ss}$ into $\varphi^*_x$ yields $\varphi^*_{x\,ss} = \varphi^*_{ss_{\text{sphere}}}$. Therefore, in the steady state, the

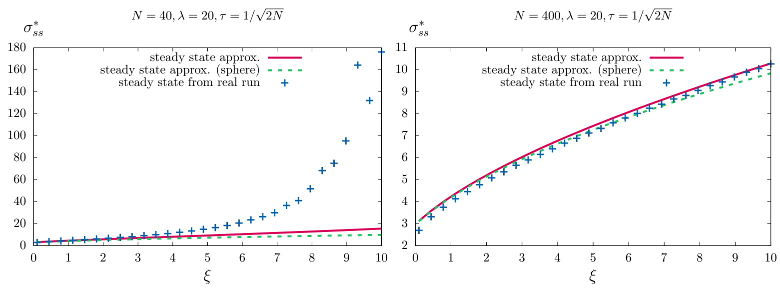


Figure 8: Steady state closed-form approximation and real-run comparison of the $(1, \lambda)$-ES with repair by projection applied to the conically constrained problem.

$(1, \lambda)$-ES with repair by projection applied to the conically constrained problem approaches the optimizer with the same rate as if a sphere were to be optimized *irrespectively of the value of* $\xi$. That is, due to the definition of $\varphi^*$ one observes $x^{((g))}$ and $r^{(g)}$ dynamics proportional to $\exp\left(-\frac{\varphi^*}{N} g\right)$ in the steady state. This implies linear convergence order with a convergence rate of $\frac{\varphi^*}{N}$. This result can be interpreted in the sense that the optimal repair has transformed the conical constraint into a sphere model.

Fig. 8 shows plots of the steady state computations. The derived closed-form approximation has been compared to real ES runs. For $N = 40$, the approximation for $\sigma^*_{ss}$ is good for small values of $\xi$. For higher values of $\xi$, the deviations of the approximation from the real runs increase. For $N = 400$, the approximation for $\sigma^*_{ss}$ comes close to the results of the real runs for the values of $\xi$ under consideration. More plots can be found in [9, Fig. 3.9, pp. 89-90].

## 5. Conclusion

An optimization problem with a conically shaped feasibility region has been presented. A $(1, \lambda)$-ES that can be applied to the presented problem has then been described. Whereas prior work dealt with resampling, the repair approach in this work is based on projection, i.e., the minimal repair principle has been applied. The derivation of closed-form approximations has been shown. First, approximations for the one-generation behavior of the algorithm have been shown. Second, those approximate expressions have been used in derived deterministic evolution equations. As a result, closed-form approximations have been derived for predicting the evolution dynamics of the ES. Plots comparing the derived approximations to simulations have been presented to show the quality of the approximations.

As a remarkable result it is to be mentioned that the optimal repair by projection results in a linear convergence order of the evolution process. That is, the conically constrained linear optimization problem has been turned into a spherical unconstrained problem.

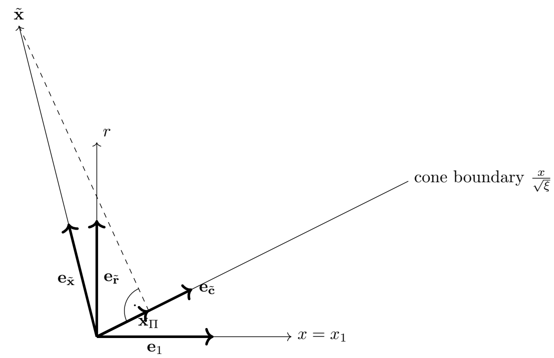


Figure 9: 2D-plane spanned by an offspring $\tilde{\mathbf{x}}$ and the $x_1$ coordinate axis. Vectors introduced for expressing the projection onto the feasible region of Fig. 1 are visualized. The vectors $\mathbf{e}_1$, $\mathbf{e}_{\tilde{\mathbf{x}}}$, $\mathbf{e}_{\tilde{\mathbf{r}}}$, and $\mathbf{e}_{\tilde{\mathbf{c}}}$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

Analysis of a multi-recombinative ES applied to the presented problem is a topic for future work. Moreover, it is of interest to analyze cumulative step size adaptation as the mutation strength control.

## Appendix

The sections in this appendix present detailed derivations. They are referenced at the appropriate places in the main text.

## 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 described. 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 (Eq. (2)). The projected vector is indicated by $\mathbf{x}_\Pi$. 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 $\mathbf{x}_\Pi$. By inspecting Fig. 9 one can see that the projected vector $\mathbf{x}_\Pi$ can be expressed as

$$\mathbf{x}_\Pi = \begin{cases} (\mathbf{e}_{\tilde{\mathbf{c}}}^T \tilde{\mathbf{x}})\mathbf{e}_{\tilde{\mathbf{c}}} & \text{if } \mathbf{e}_{\tilde{\mathbf{c}}}^T \tilde{\mathbf{x}} > 0 \\ \mathbf{0} & \text{otherwise.} \end{cases} \tag{A.1}$$

From Eq. (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

$$\mathbf{e}_{\tilde{\mathbf{c}}} = \frac{\mathbf{e}_1 + \frac{1}{\sqrt{\xi}}\mathbf{e}_{\tilde{\mathbf{r}}}}{\sqrt{1 + \frac{1}{\xi}}}. \tag{A.2}$$

The unit vectors $\mathbf{e}_1 = (1, 0, \ldots, 0)^T$ 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.,

$$\mathbf{e}_{\tilde{\mathbf{r}}} = (0, (\tilde{\mathbf{x}})_2, \ldots, (\tilde{\mathbf{x}})_N)^T / ||\tilde{\mathbf{r}}||. \tag{A.3}$$

Inserting Eq. (A.3) into Eq. (A.2) results in

$$\mathbf{e}_{\tilde{\mathbf{c}}} = \left(1, \frac{(\tilde{\mathbf{x}})_2}{\sqrt{\xi}||\tilde{\mathbf{r}}||}, \ldots, \frac{(\tilde{\mathbf{x}})_N}{\sqrt{\xi}||\tilde{\mathbf{r}}||}\right)^T \Bigg/ \sqrt{1 + \frac{1}{\xi}}. \tag{A.4}$$

Using Eq. (A.4) it follows that

$$\begin{aligned}
\mathbf{e}_{\tilde{\mathbf{c}}}^T \tilde{\mathbf{x}} &= \left((\tilde{\mathbf{x}})_1 + \frac{(\tilde{\mathbf{x}})_2^2 + \cdots + (\tilde{\mathbf{x}})_N^2}{\sqrt{\xi}||\tilde{\mathbf{r}}||}\right) \Bigg/ \sqrt{1 + \frac{1}{\xi}} \\
&= \left((\tilde{\mathbf{x}})_1 + \frac{||\tilde{\mathbf{r}}||^2}{\sqrt{\xi}||\tilde{\mathbf{r}}||}\right) \Bigg/ \sqrt{1 + \frac{1}{\xi}} \\
&= \left((\tilde{\mathbf{x}})_1 + \frac{||\tilde{\mathbf{r}}||}{\sqrt{\xi}}\right) \Bigg/ \sqrt{1 + \frac{1}{\xi}} = \left((\tilde{\mathbf{x}})_1 \sqrt{\xi} + ||\tilde{\mathbf{r}}||\right) / \sqrt{\xi + 1}.
\end{aligned} \tag{A.5}$$

Use of Eqs. (A.1), (A.4), and (A.5) yields for the case $\mathbf{e}_{\tilde{\mathbf{c}}}^T \tilde{\mathbf{x}} > 0$

$$\begin{aligned}
\mathbf{x}_\Pi &= (\mathbf{e}_{\tilde{\mathbf{c}}}^T \tilde{\mathbf{x}}) \mathbf{e}_{\tilde{\mathbf{c}}} \\
&= \frac{1}{1 + \frac{1}{\xi}} \left((\tilde{\mathbf{x}})_1 + \frac{(\tilde{\mathbf{x}})_2^2 + \cdots + (\tilde{\mathbf{x}})_N^2}{\sqrt{\xi}||\tilde{\mathbf{r}}||}\right) \left(1, \frac{(\tilde{\mathbf{x}})_2}{\sqrt{\xi}||\tilde{\mathbf{r}}||}, \ldots, \frac{(\tilde{\mathbf{x}})_N}{\sqrt{\xi}||\tilde{\mathbf{r}}||}\right)^T.
\end{aligned} \tag{A.6}$$

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

$$q = (\mathbf{x}_\Pi)_1 = \frac{1}{1 + \frac{1}{\xi}} \left((\tilde{\mathbf{x}})_1 + \frac{(\tilde{\mathbf{x}})_2^2 + \cdots + (\tilde{\mathbf{x}})_N^2}{\sqrt{\xi}||\tilde{\mathbf{r}}||}\right) = \frac{\xi}{\xi + 1} \left((\tilde{\mathbf{x}})_1 + \frac{||\tilde{\mathbf{r}}||}{\sqrt{\xi}}\right). \tag{A.7}$$

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

$$(\mathbf{x}_\Pi)_k = \frac{\xi}{\xi + 1} \left((\tilde{\mathbf{x}})_1 + \frac{||\tilde{\mathbf{r}}||}{\sqrt{\xi}}\right) \frac{(\tilde{\mathbf{x}})_k}{\sqrt{\xi}||\tilde{\mathbf{r}}||} = \frac{\xi}{\xi + 1} \left(\frac{(\tilde{\mathbf{x}})_1}{\sqrt{\xi}||\tilde{\mathbf{r}}||} + \frac{1}{\xi}\right) (\tilde{\mathbf{x}})_k. \tag{A.8}$$

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

$$\begin{aligned}
q_r = ||\mathbf{r}_\Pi|| &= \sqrt{\sum_{k=2}^{N} (\mathbf{x}_\Pi)_k^2} = \sqrt{\sum_{k=2}^{N} \left(\frac{\xi}{\xi + 1} \left(\frac{(\tilde{\mathbf{x}})_1}{\sqrt{\xi}||\tilde{\mathbf{r}}||} + \frac{1}{\xi}\right)\right)^2 (\tilde{\mathbf{x}})_k^2} \\
&= \left(\frac{\xi}{\xi + 1} \left(\frac{(\tilde{\mathbf{x}})_1}{\sqrt{\xi}||\tilde{\mathbf{r}}||} + \frac{1}{\xi}\right)\right) ||\tilde{\mathbf{r}}||.
\end{aligned} \tag{A.9}$$

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.

## B. Derivation of the Normal Approximation for the Offspring Density in $r$ Direction

From the offspring generation (Lines 7 and 8) it follows that the distance from the cone's axis of the offspring is

$$\tilde{r} = \sqrt{{r^{(g)}}^2 + 2\sigma^{(g)}r^{(g)}z_2 + {\sigma^{(g)}}^2 z_2^2 + {\sigma^{(g)}}^2 \sum_{k=3}^{N} z_k^2}. \tag{B.1}$$

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 $z_k$, $\mathrm{E}[z_k] = 0$, and $\mathrm{E}[z_k^2] = 1$, the expected values $\mathrm{E}\left[2\sigma^{(g)}r^{(g)}z_2 + {\sigma^{(g)}}^2 z_2^2\right] = {\sigma^{(g)}}^2$ and $\mathrm{E}\left[{\sigma^{(g)}}^2 \sum_{k=3}^{N} z_k^2\right] = {\sigma^{(g)}}^2(N-2)$ follow. The variances are computed as

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

and

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

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

With these results, the normal approximation follows as

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

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

$$\tilde{r} \approx r^{(g)}\sqrt{1+\frac{{\sigma^{(g)}}^{*2}}{N}\left(1-\frac{1}{N}\right)}\sqrt{1+\frac{2\sigma^{(g)^*}}{N}\frac{\sqrt{1+\frac{{\sigma^{(g)}}^{*2}}{2N}\left(1-\frac{1}{N}\right)}}{1+\frac{{\sigma^{(g)}}^{*2}}{N}\left(1-\frac{1}{N}\right)}\mathcal{N}(0,1)}. \tag{B.5}$$

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

$$\tilde{r} \approx \underbrace{r^{(g)}\sqrt{1+\frac{{\sigma^{(g)}}^{*2}}{N}\left(1-\frac{1}{N}\right)}}_{\bar{r}} + \underbrace{r^{(g)}\frac{\sigma^{(g)^*}}{N}\sqrt{\frac{1+\frac{{\sigma^{(g)}}^{*2}}{2N}\left(1-\frac{1}{N}\right)}{1+\frac{{\sigma^{(g)}}^{*2}}{N}\left(1-\frac{1}{N}\right)}}}_{\sigma_r}\mathcal{N}(0,1). \tag{B.6}$$

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

$$p_r(r) \approx \frac{1}{\sigma_r}\phi\left(\frac{r-\bar{r}}{\sigma_r}\right) = \frac{1}{\sqrt{2\pi}\sigma_r}\exp\left[-\frac{1}{2}\left(\frac{r-\bar{r}}{\sigma_r}\right)^2\right]. \tag{B.7}$$

## C. Derivation of a $P_Q(q)$ Approximation

In the following, asymptotic lower and upper bounds for $P_Q(q)$ are derived. Use of the approximations in Eqs. (20), (21), and (23), and subsequent insertion into Eq. (16) yield

$$\begin{aligned} P_Q(q) \approx & \int_{x=-\infty}^{x=q} p_x(x) \underbrace{\int_{r=0}^{r=\infty} p_r(r)\,\mathrm{d}r}_{=1}\,\mathrm{d}x \\ & - \int_{r=q/\sqrt{\xi}}^{r=\infty}\left[\int_{x=-(1/\sqrt{\xi})r+(1+1/\xi)q}^{x=q} p_x(x)\,\mathrm{d}x\right]p_r(r)\,\mathrm{d}r \end{aligned} \tag{C.1}$$

$$\begin{aligned} \approx & \int_{x=-\infty}^{x=q}\frac{1}{\sqrt{2\pi}\sigma^{(g)}}\exp\left[-\frac{1}{2}\left(\frac{x-x^{(g)}}{\sigma^{(g)}}\right)^2\right]\mathrm{d}x \\ & - \int_{r=q/\sqrt{\xi}}^{r=\infty}\left\{\int_{x=-(1/\sqrt{\xi})r+(1+1/\xi)q}^{x=q}\right. \\ & \left.\frac{1}{\sqrt{2\pi}\sigma^{(g)}}\exp\left[-\frac{1}{2}\left(\frac{x-x^{(g)}}{\sigma^{(g)}}\right)^2\right]\mathrm{d}x\right\} \\ & \times\frac{1}{\sqrt{2\pi}\sigma_r}\exp\left[-\frac{1}{2}\left(\frac{r-\bar{r}}{\sigma_r}\right)^2\right]\mathrm{d}r. \end{aligned} \tag{C.2}$$

Expressing the integrals where possible using the cumulative distribution function of the standard normal distribution $\Phi(\cdot)$, Eq. (C.2) can be turned into

$$
\begin{aligned}
P_Q(q) \approx\ & \Phi\left(\frac{q - x^{(g)}}{\sigma^{(g)}}\right) \\
& - \int_{r=q/\sqrt{\xi}}^{r=\infty} \left[\Phi\left(\frac{q - x^{(g)}}{\sigma^{(g)}}\right) - \Phi\left(\frac{-(1/\sqrt{\xi})r + (1+1/\xi)q - x^{(g)}}{\sigma^{(g)}}\right)\right] \\
& \qquad \times \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left[-\frac{1}{2}\left(\frac{r - \bar{r}}{\sigma_r}\right)^2\right] \mathrm{d}r.
\end{aligned}
\tag{C.3}
$$

Eq. (C.3) can further be simplified to yield

$$
\begin{aligned}
P_Q(q) \approx\ & \Phi\left(\frac{q - x^{(g)}}{\sigma^{(g)}}\right) \Phi\left(\frac{q/\sqrt{\xi} - \bar{r}}{\sigma_r}\right) \\
& + \int_{r=q/\sqrt{\xi}}^{r=\infty} \Phi\left(\frac{-(1/\sqrt{\xi})r + (1+1/\xi)q - x^{(g)}}{\sigma^{(g)}}\right) \\
& \qquad \times \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left[-\frac{1}{2}\left(\frac{r - \bar{r}}{\sigma_r}\right)^2\right] \mathrm{d}r.
\end{aligned}
\tag{C.4}
$$

Because $\Phi(\cdot)$ and $\frac{1}{\sigma_r}\phi\left(\frac{r-\bar{r}}{\sigma_r}\right) = \frac{1}{\sqrt{2\pi}\sigma_r}\exp\left[-\frac{1}{2}\left(\frac{r-\bar{r}}{\sigma_r}\right)^2\right]$ are non-negative functions and the second summand is non-negative, Eq. (C.4) can be lower bounded

$$
P_Q(q) \geq \Phi\left(\frac{q - x^{(g)}}{\sigma^{(g)}}\right) \Phi\left(\frac{q/\sqrt{\xi} - \bar{r}}{\sigma_r}\right). \tag{C.5}
$$

To arrive at an upper bound, the integration order in the subtrahend of Eq. (16) can be reversed yielding Eq. (17). Analogously to the derivation of the lower bound, Eqs. (20), (21), (23), and (C.2), and insertion into Eq. (17)

(different order of integration in the subtrahend) yield

$$
\begin{aligned}
P_Q(q) \approx & \int_{x=-\infty}^{x=q} p_x(x) \underbrace{\int_{r=0}^{r=\infty} p_r(r)\,\mathrm{d}r}_{=1}\,\mathrm{d}x \\
& - \int_{x=-\infty}^{x=q} p_x(x) \left[ \int_{r=-\sqrt{\xi}x+(\sqrt{\xi}+1/\sqrt{\xi})q}^{r=\infty} p_r(r)\,\mathrm{d}r \right] \mathrm{d}x
\end{aligned} \tag{C.6}
$$

$$
\begin{aligned}
\approx & \int_{x=-\infty}^{x=q} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[ -\frac{1}{2} \left( \frac{x - x^{(g)}}{\sigma^{(g)}} \right)^2 \right] \mathrm{d}x \\
& - \int_{x=-\infty}^{x=q} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[ -\frac{1}{2} \left( \frac{x - x^{(g)}}{\sigma^{(g)}} \right)^2 \right] \\
& \times \left\{ \int_{r=-\sqrt{\xi}x+(\sqrt{\xi}+1/\sqrt{\xi})q}^{r=\infty} \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left[ -\frac{1}{2} \left( \frac{r - \bar{r}}{\sigma_r} \right)^2 \right] \mathrm{d}r \right\} \mathrm{d}x.
\end{aligned} \tag{C.7}
$$

Expressing the integrals where possible using the cumulative distribution function of the standard normal distribution $\Phi(\cdot)$, Eq. (C.7) can be turned into

$$
\begin{aligned}
P_Q(q) \approx & \Phi\left( \frac{q - x^{(g)}}{\sigma^{(g)}} \right) - \Phi\left( \frac{q - x^{(g)}}{\sigma^{(g)}} \right) \\
& + \int_{x=-\infty}^{x=q} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[ -\frac{1}{2} \left( \frac{x - x^{(g)}}{\sigma^{(g)}} \right)^2 \right] \\
& \times \Phi\left( \frac{-\sqrt{\xi}x + (\sqrt{\xi} + 1/\sqrt{\xi})q - \bar{r}}{\sigma_r} \right) \mathrm{d}x.
\end{aligned} \tag{C.8}
$$

Eq. (C.8) can be upper bounded yielding

$$
\begin{aligned}
P_Q(q) \leq & \int_{x=-\infty}^{x=\infty} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[ -\frac{1}{2} \left( \frac{x - x^{(g)}}{\sigma^{(g)}} \right)^2 \right] \\
& \times \Phi\left( \frac{-\sqrt{\xi}x + (\sqrt{\xi} + 1/\sqrt{\xi})q - \bar{r}}{\sigma_r} \right) \mathrm{d}x
\end{aligned} \tag{C.9}
$$

$$
= \Phi\left( \frac{(1 + 1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{{\sigma^{(g)}}^2 + \sigma_r^2/\xi}} \right). \tag{C.10}
$$

The bound 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 Eq. (C.8) has an upper bound of $q$ and a closed form solution is not apparent, the integral in Eq. (C.9) has an upper bound of $\infty$ and can be solved analytically. To this end, the identity

$$
\int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} \Phi(at + b) = \sqrt{2\pi}\Phi\left( \frac{b}{\sqrt{1 + a^2}} \right), \tag{C.11}
$$

which has been derived and proven in [10, Equation A.10], is used. The substitution $t := \frac{x-x^{(g)}}{\sigma^{(g)}}$ that implies $\mathrm{d}x = \sigma^{(g)}\,\mathrm{d}t$ yields for Eq. (C.9)

$$\int_{x=-\infty}^{x=\infty} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{x-x^{(g)}}{\sigma^{(g)}}\right)^2\right] \Phi\left(\frac{-\sqrt{\xi}x + (\sqrt{\xi}+1/\sqrt{\xi})q - \bar{r}}{\sigma_r}\right)\mathrm{d}x = \int_{t=-\infty}^{t=\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \Phi\left(\frac{-\sqrt{\xi}(\sigma^{(g)}t + x^{(g)}) + (\sqrt{\xi}+1/\sqrt{\xi})q - \bar{r}}{\sigma_r}\right)\mathrm{d}t. \tag{C.12}$$

In this form, Eq. (C.11) can directly be used for Eq. (C.12) with

$$b = \frac{-\sqrt{\xi}x^{(g)} + (\sqrt{\xi}+1/\sqrt{\xi})q - \bar{r}}{\sigma_r} \tag{C.13}$$

and

$$a = \frac{-\sqrt{\xi}\sigma^{(g)}}{\sigma_r}. \tag{C.14}$$

Insertion of Eq. (C.13) and Eq. (C.14) into Eq. (C.11) results, after simplification, in Eq. (C.10). Consequently, $P_Q(q)$ can be bounded in the asymptotic case considered, as

$$\Phi\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)\Phi\left(\frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r}\right) \le P_Q(q) \le \Phi\left(\frac{(1+1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{{\sigma^{(g)}}^2 + \sigma_r^2/\xi}}\right). \tag{C.15}$$

*C.1. The Derived $P_Q(q)$ Upper Bound as an Approximation for $P_Q(q)$*

The goal is to find a good approximation for $P_Q(q)$ for the asymptotic cases. As it turns out, the upper bound (Eq. (C.10)) is a good approximation for large $\xi$. It is referred to [9, Section 3.1.2.1.2.3, pp. 33-34] for an error analysis.

*C.2. The $P_Q(q)$ Approximation in the case that the probability of feasible offspring tends to* 1

By taking a close look at the integral in Eq. (C.8), one can see that the expression can be further simplified under certain conditions. Normalization of

$\sigma_r$ and multiplication of the argument to $\Phi(\cdot)$ by $\frac{1/\sqrt{\xi}}{1/\sqrt{\xi}}$ in Eq. (C.8) results in

$$\begin{aligned} P_Q(q) \approx & \int_{x=-\infty}^{x=q} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{x-x^{(g)}}{\sigma^{(g)}}\right)^2\right] \\ & \qquad \times \Phi\left(\frac{-x+(1+1/\xi)q-\bar{r}/\sqrt{\xi}}{\sigma_r^* r^{(g)}/(N\sqrt{\xi})}\right) \mathrm{d}x \end{aligned} \tag{C.16}$$

$$\begin{aligned} = & \int_{x=-\infty}^{x=q} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{x-x^{(g)}}{\sigma^{(g)}}\right)^2\right] \\ & \times \Phi\left(N\sqrt{\xi}\frac{-x/r^{(g)}+(1+1/\xi)q/r^{(g)}-\bar{r}/(r^{(g)}\sqrt{\xi})}{\sigma_r^*}\right) \mathrm{d}x. \end{aligned} \tag{C.17}$$

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

$$\Phi\left(N\sqrt{\xi}\frac{-\frac{x}{r^{(g)}}+(1+\frac{1}{\xi})\frac{q}{r^{(g)}}-\frac{\bar{r}}{r^{(g)}\sqrt{\xi}}}{\sigma_r^*}\right) \approx 1 \tag{C.18}$$

$$\iff \left(1+\frac{1}{\xi}\right)\frac{q}{r^{(g)}} - \frac{x}{r^{(g)}} - \frac{\bar{r}}{r^{(g)}\sqrt{\xi}} > 0 \tag{C.19}$$

$$\iff x < \left(1+\frac{1}{\xi}\right)q - \frac{\bar{r}}{\sqrt{\xi}}. \tag{C.20}$$

Making use of $x \le q$ that follows from the bounds of the integration, this can further be rewritten to

$$x < \left(1+\frac{1}{\xi}\right)q - \frac{\bar{r}}{\sqrt{\xi}} \iff q < \left(1+\frac{1}{\xi}\right)q - \frac{\bar{r}}{\sqrt{\xi}} \iff \frac{\bar{r}\sqrt{\xi}}{q} < 1. \tag{C.21}$$

The resulting condition (Eq. (C.21)) 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

$$\Pr\left[\sqrt{\xi}\tilde{r} \le |q|\right] = \int_{\tilde{r}=0}^{\tilde{r}=q/\sqrt{\xi}} p_r(\tilde{r})\,\mathrm{d}\tilde{r} \approx \Phi\left(\frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r}\right) = \Phi\left(N\frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r^* r^{(g)}}\right). \tag{C.22}$$

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

$$\Phi\left(N\frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r^* r^{(g)}}\right) \approx 1 \iff \frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r^* r^{(g)}} > 0 \iff \frac{\bar{r}\sqrt{\xi}}{q} < 1. \tag{C.23}$$

Under this condition, Eq. (C.16) can be simplified in the asymptotic case $N \to \infty$ to

$$P_Q(q) \approx \Phi\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right) \quad \text{for } q > \bar{r}\sqrt{\xi}. \tag{C.24}$$

with

$$p_Q(q) = \frac{\mathrm{d}}{\mathrm{d}q}P_Q(q) \approx \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)^2} \quad \text{for } q > \bar{r}\sqrt{\xi}. \tag{C.25}$$

## D. Derivation of an $\mathrm{E}[q_{1;\lambda_{\mathrm{feas}}}]$ Approximation

Insertion of Eqs. (24) and (26) into Eq. (15) yields for the feasible case

$$\mathrm{E}[q_{1;\lambda_{\mathrm{feas}}}] \approx \lambda \int_{q=0}^{q=\infty} q\, p_{Q_{\mathrm{feas}}}(q)[1 - P_{Q_{\mathrm{feas}}}(q)]^{\lambda-1}\,\mathrm{d}q \tag{D.1}$$

$$= \lambda \int_{q=\bar{r}\sqrt{\xi}}^{q=\infty} q \frac{1}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)^2} \left[1 - \Phi\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)\right]^{\lambda-1} \mathrm{d}q. \tag{D.2}$$

The substitution $\frac{q-x^{(g)}}{\sigma^{(g)}} := -t$ is used. It follows that $q = -t\sigma^{(g)} + x^{(g)}$ and $\mathrm{d}q = -\sigma^{(g)}\,\mathrm{d}t$. Using normalized quantities, $t$ can be expressed as $t = \frac{-N(q-x^{(g)})}{\sigma^{(g)*} r^{(g)}}$. Assuming $N \to \infty$ yields for the upper bound $t_u = -\infty$ and for the lower bound $t_l = \infty$. The derivation of the these bounds is provided in more detail in [9, Section 3.1.2.1.2.6, pp. 40-41]. Applying the substitution results in

$$\mathrm{E}[q_{1;\lambda_{\mathrm{feas}}}] \approx -\lambda \int_{t=\infty}^{t=-\infty} (x^{(g)} - t\sigma^{(g)}) \frac{\sigma^{(g)}}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2}t^2} \Bigg[\underbrace{1 - \Phi(-t)}_{=\Phi(t)}\Bigg]^{\lambda-1} \mathrm{d}t \tag{D.3}$$

$$= x^{(g)} \frac{\lambda}{\sqrt{2\pi}} \int_{t=-\infty}^{t=\infty} e^{-\frac{1}{2}t^2} \Phi(t)^{\lambda-1}\,\mathrm{d}t - \sigma^{(g)} \frac{\lambda}{\sqrt{2\pi}} \int_{t=-\infty}^{t=\infty} t e^{-\frac{1}{2}t^2} \Phi(t)^{\lambda-1}\,\mathrm{d}t \tag{D.4}$$

$$= x^{(g)} - \sigma^{(g)} c_{1,\lambda}. \tag{D.5}$$

In the step from Eq. (D.3) to Eq. (D.4), the fact that taking the negative of an integral can be expressed by exchanging the lower and upper bounds of the integral, the symmetry of the standard normal distribution have been used, and the fact that the integral of a sum can be expressed as a sum of integrals have been used. In Eq. (D.4), the integrals contain the probability density function of the order statistic of standard normals. The first integral is an integration over all possible values. Hence, this equals 1. The second integral is its expected value and is the so-called progress coefficient $c_{1,\lambda}$ (see Eq. (30)).

## E. Derivation of an $\mathrm{E}[q_{1;\lambda_{\text{infeas}}}]$ Approximation

Analogously to the feasible case, insertion of Eqs. (25) and (27) into Eq. (15) yields for the other case

$$\mathrm{E}[q_{1;\lambda_{\text{infeas}}}] \approx \lambda \int_{q=0}^{q=\infty} q\, p_{Q_{\text{infeas}}}(q)[1-P_{Q_{\text{infeas}}}(q)]^{\lambda-1}\,\mathrm{d}q \tag{E.1}$$

$$\begin{aligned} &= \lambda \int_{q=0}^{q=\bar{r}\sqrt{\xi}} q \left(\frac{(1+1/\xi)}{\sqrt{{\sigma^{(g)}}^2+\sigma_r^2/\xi}}\right)\frac{1}{\sqrt{2\pi}} \\ &\quad\times \exp\left[-\frac{1}{2}\left(\frac{(1+1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{{\sigma^{(g)}}^2+\sigma_r^2/\xi}}\right)^2\right] \\ &\quad\times \left[1-\Phi\left(\frac{(1+1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{{\sigma^{(g)}}^2+\sigma_r^2/\xi}}\right)\right]^{\lambda-1} \mathrm{d}q. \end{aligned} \tag{E.2}$$

The substitution

$$\frac{(1+1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{{\sigma^{(g)}}^2+\sigma_r^2/\xi}} := -t \tag{E.3}$$

is used. It follows that

$$q = \frac{1}{(1+1/\xi)}\left(-\sqrt{{\sigma^{(g)}}^2+\sigma_r^2/\xi}\,t + x^{(g)} + \bar{r}/\sqrt{\xi}\right) \tag{E.4}$$

and

$$\mathrm{d}q = -\frac{\sqrt{{\sigma^{(g)}}^2+\sigma_r^2/\xi}}{(1+1/\xi)}\,\mathrm{d}t. \tag{E.5}$$

Using the normalized ${\sigma^{(g)}}^*$, $\sigma_r \simeq \sigma^{(g)}$ for $N \to \infty$, and ${\sigma^{(g)}}^* \ll N$ (derived from Eq. (B.6)), $t$ can be expressed as

$$t = -\frac{(1+1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{\sqrt{\frac{{\sigma^{(g)}}^{*2}{r^{(g)}}^2}{N^2}+\frac{{\sigma^{(g)}}^{*2}{r^{(g)}}^2}{N^2\xi}}} = -N\sqrt{\xi}\left[\frac{(1+1/\xi)q - x^{(g)} - \bar{r}/\sqrt{\xi}}{{\sigma^{(g)}}^* r^{(g)}\sqrt{\xi+1}}\right]. \tag{E.6}$$

The lower bound in the transformed integral therefore follows assuming $\xi \gg 1$, $N \to \infty$, using $\infty > \bar{r} \simeq r^{(g)} \geq 0$, and knowing that $0 \leq x^{(g)} < \infty$ as $t_l \simeq \infty$. Similarly, the upper bound follows with the same assumptions $t_u \simeq -\infty$. The derivation of the these bounds is provided in more detail [9, Section 3.1.2.1.2.7, pp. 41-44]. Actually applying the substitution to Eq. (E.2) leads to

$$\begin{aligned} \mathrm{E}[q_{1;\lambda_{\text{infeas}}}] = \frac{\lambda}{\sqrt{2\pi}}\left(\frac{1}{1+1/\xi}\right)\int_{t=-\infty}^{t=\infty}\left(-\sqrt{{\sigma^{(g)}}^2+\sigma_r^2/\xi}\,t + x^{(g)} + \bar{r}/\sqrt{\xi}\right) \\ \times e^{-\frac{1}{2}t^2}\Phi(t)^{\lambda-1}\,\mathrm{d}t \end{aligned} \tag{E.7}$$

where the fact that taking the negative of an integral can be expressed by exchanging the lower and upper bounds of the integral and the symmetry of the standard normal distribution have been used. This can be further treated yielding

$$\begin{aligned}
\mathrm{E}[q_{1;\lambda_{\text{infeas}}}] = &-\frac{\sqrt{{\sigma^{(g)}}^2+\sigma_r^2/\xi}}{1+1/\xi}\underbrace{\frac{\lambda}{\sqrt{2\pi}}\int_{t=-\infty}^{t=\infty} te^{-\frac{1}{2}t^2}\Phi(t)^{\lambda-1}\,\mathrm{d}t}_{=c_{1,\lambda}} \\
&+\frac{x^{(g)}+\bar{r}/\sqrt{\xi}}{1+1/\xi}\underbrace{\frac{\lambda}{\sqrt{2\pi}}\int_{t=-\infty}^{t=\infty} e^{-\frac{1}{2}t^2}\Phi(t)^{\lambda-1}\,\mathrm{d}t}_{=1}
\end{aligned} \tag{E.8}$$

$$= \frac{\xi}{1+\xi}\left(x^{(g)}+\bar{r}/\sqrt{\xi}\right)-\frac{\xi}{1+\xi}\left(\sqrt{{\sigma^{(g)}}^2+\sigma_r^2/\xi}\right)c_{1,\lambda}. \tag{E.9}$$

In Eq. (E.8), the integrals in the first and second summands contain the probability density function of the order statistic of standard normals. The second integral is an integration over all possible values. Hence, this equals 1. The second integral is its expected value and is the so-called progress coefficient $c_{1,\lambda}$ (see Eq. (30)).

## F. Derivation of an Approximation for the Offspring Feasibility Probability

An offspring with values $(\tilde{x},\tilde{r})^T$ is feasible if the condition $\tilde{r} \le \tilde{x}/\sqrt{\xi}$ holds. For a particular value $q=\tilde{x}$, this probability is denoted by

$$\Pr\left[\sqrt{\xi}\tilde{r} \le |q|\right] = \int_{\tilde{r}=0}^{\tilde{r}=q/\sqrt{\xi}} p_r(\tilde{r})\,\mathrm{d}\tilde{r} \approx \Phi\left(\frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r}\right) \simeq \Phi\left(\frac{q/\sqrt{\xi}-\bar{r}}{\sigma^{(g)}}\right). \tag{F.1}$$

It is computed by integrating from the cone axis up to the constraint boundary at the given $q$ value. The last equality results from Eq. (B.6) ($\sigma_r \approx \sigma^{(g)}$ for $N \to \infty$). This particular $q$ is the best projected value among $\lambda$ with probability density $p_{1;\lambda}(q) = \lambda p_Q(q)[1-P_Q(q)]^{\lambda-1}$. As an approximation for the best offspring probability, the probability for a single offspring to be feasible is computed and considered here. In [9, Section 3.1.2.1.2.8, pp. 44-47] it is shown in more detail that this single offspring probability can be derived from the best offspring feasibility in the asymptotic case $N \to \infty$. For the single offspring case, integrating Eq. (F.1) over all possible values of $q$ results in

$$P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \simeq \int_{q=-\infty}^{q=\infty} \Phi\left[\frac{1}{\sigma^{(g)}}\left(\frac{q}{\sqrt{\xi}}-\bar{r}\right)\right] p_x(q)\,\mathrm{d}q. \tag{F.2}$$

with $p_x(q) = \frac{1}{\sqrt{2\pi}\sigma^{(g)}}\exp\left[-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)^2\right]$ from Eq. (21). The substitution $t := \frac{q-x^{(g)}}{\sigma^{(g)}}$ is used. It implies $q = \sigma^{(g)}t + x^{(g)}$ and $\mathrm{d}q = \sigma\mathrm{d}t$. With normalization it follows further that $t = N\left(\frac{q - x^{(g)}}{\sigma^{(g)*} r^{(g)}}\right)$. This implies $t_l = -\infty$ and $t_u = \infty$. Application of the substitution yields

$$\int_{q=0}^{q=\infty} \Phi\left[\frac{1}{\sigma^{(g)}}\left(\frac{q}{\sqrt{\xi}} - \bar{r}\right)\right] \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{q - x^{(g)}}{\sigma^{(g)}}\right)^2\right] \mathrm{d}q \tag{F.3}$$

$$= \int_{t=-\infty}^{t=\infty} \Phi\left[\frac{1}{\sigma^{(g)}}\left(\frac{\sigma^{(g)} t + x^{(g)}}{\sqrt{\xi}} - \bar{r}\right)\right] \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \,\mathrm{d}t. \tag{F.4}$$

Now, Eq. (C.11) can be applied with $a = \frac{1}{\sqrt{\xi}}$ and $b = \frac{x^{(g)}}{\sqrt{\xi}\sigma^{(g)}} - \frac{\bar{r}}{\sigma^{(g)}}$ resulting in

$$\int_{t=-\infty}^{t=\infty} \Phi\left[\frac{1}{\sigma^{(g)}}\left(\frac{\sigma^{(g)} t + x^{(g)}}{\sqrt{\xi}} - \bar{r}\right)\right] \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \,\mathrm{d}t \tag{F.5}$$

$$= \Phi\left[\frac{1}{\sigma^{(g)}\sqrt{1 + \frac{1}{\xi}}}\left(\frac{x^{(g)}}{\sqrt{\xi}} - \bar{r}\right)\right] \simeq \Phi\left[\frac{1}{\sigma^{(g)}}\left(\frac{x^{(g)}}{\sqrt{\xi}} - \bar{r}\right)\right]. \tag{F.6}$$

The last step was derived under the assumption $\xi \to \infty$.

## G. Derivation of an $\mathrm{E}[q_{r\,1;\lambda_{\mathrm{infeas}}}]$ Approximation

This section presents the derivation of an approximation for the expectation of the $r$ value after projection of the best offspring in $r$ direction

$$\mathrm{E}[q_{r\,1;\lambda} \,|\, x^{(g)}, r^{(g)}, \sigma^{(g)}] := \mathrm{E}[q_{r\,1;\lambda}] = \int_{q_r=0}^{q_r=\infty} q_r \, p_{q_{r\,1;\lambda}}(q_r) \,\mathrm{d}q_r. \tag{G.1}$$

Eq. (G.1) follows directly from the definition of expectation where

$$p_{q_{r\,1;\lambda}}(q_r) := p_{q_{r\,1;\lambda}}(q_r \,|\, x^{(g)}, r^{(g)}, \sigma^{(g)})$$

indicates the probability density function of the best (offspring with smallest $q$ value) offspring's $q_r$ value. Considering an arbitrary offspring with $(\tilde{x}, \tilde{r})^T$ values before projection and $(q, q_r)^T$ values after projection. The $q_r$ value is the best among the $\lambda$ values if its corresponding $q$ value is the smallest among the $\lambda$ offspring's $q$ values. This follows from the definition of the objective function (Eq. (1)). One arbitrarily selected mutation out of the total $\lambda$ mutations has joint probability density $p_{Q,Q_r}(q, q_r) := p_{Q,Q_r}(q, q_r \,|\, x^{(g)}, r^{(g)}, \sigma^{(g)})$ for its projected values $q$ and $q_r$. With a similar order statistic argument as for the $x$ progress case in Section 4.2.1, one obtains $\lambda p_{Q,Q_r}(q, q_r)[1 - P_Q(q)]^{\lambda-1}$. Integration over all possible values of $q$ yields

$$p_{q_{r\,1;\lambda}}(q_r) = \lambda \int_{q=0}^{q=\infty} p_{Q,Q_r}(q, q_r)[1 - P_Q(q)]^{\lambda-1} \,\mathrm{d}q. \tag{G.2}$$

Insertion of Eq. (G.2) into Eq. (G.1) results in

$$\mathrm{E}[q_{r\,1;\lambda}] = \int_{q_r=0}^{q_r=\infty} q_r\,\lambda \int_{q=0}^{q=\infty} p_{Q,Q_r}(q,q_r)[1-P_Q(q)]^{\lambda-1}\,\mathrm{d}q\,\mathrm{d}q_r \tag{G.3}$$

$$= \lambda \int_{q=0}^{q=\infty} \underbrace{\left[\int_{q_r=0}^{q_r=\infty} q_r p_{Q,Q_r}(q,q_r)\,\mathrm{d}q_r\right]}_{=:I(q)}[1-P_Q(q)]^{\lambda-1}\,\mathrm{d}q. \tag{G.4}$$

The integral in $I(q)$ in Eq. (G.4) can be expressed in terms of the values before the projection. Because the value $q_r$ is an individual's $r$ value after projection, it only takes values in the interval $[0, x/\sqrt{\xi}]$ for $x \in \mathbb{R}, x \geq 0$. All values outside this interval are projected onto the cone boundary. The integral in Eq. (G.4) is therefore split into multiple parts. Considering an infinitesimally small $\mathrm{d}q$,

$$I(q)\mathrm{d}q = \int_{\tilde{r}=0}^{\tilde{r}=q/\sqrt{\xi}} \tilde{r} p_{1;1}(q,\tilde{r})\,\mathrm{d}\tilde{r}\,\mathrm{d}q + \frac{q}{\sqrt{\xi}} \underbrace{\left(p_Q(q)\,\mathrm{d}q - \int_{\tilde{r}=0}^{\tilde{r}=q/\sqrt{\xi}} p_{1;1}(q,\tilde{r})\,\mathrm{d}\tilde{r}\,\mathrm{d}q\right)}_{=\,\mathrm{d}P_{\text{line}}} \tag{G.5}$$

is derived. The first summand in Eq. (G.5) corresponds to the part inside the cone ($\tilde{r} \leq q/\sqrt{\xi}$). There, no projection occurs. Consequently, this part is equal to the first summand in $I(q)$. The second summand corresponds to the projected part. Everything on the projection line resulting in particular $(q, q_r)^T$ values, is projected to $q_r = q/\sqrt{\xi}$. It can be expressed in terms of $Q$. The probability for an offspring to fall onto the infinitesimally small area at $q$ around the projection line, equals $\mathrm{d}P_{\text{line}}$. The summand in $\mathrm{d}P_{\text{line}}$ corresponds to the probability that the projected $x$ value equals $q$. It includes the feasible case. The subtrahend subtracts the probability for an offspring to fall onto the infinitesimally small area at $q$ around the line from the cone axis to the cone boundary. As described in Section 4.2.1, $p_{1;1}(x,r)$ denotes the joint probability density of an offspring's $(x,r)^T$ values before projection. The same argument leading to Eq. (20) is used. That is, assuming a sufficiently small value for $\tau$, the log-normally distributed offspring mutation strength $\tilde{\sigma}_l$ tends to the parental mutation strength $\sigma^{(g)}$. Under this assumption that $\tilde{\sigma}_l \approx \sigma^{(g)}$, Eq. (G.5) can be simplified yielding

$$\begin{aligned} I(q)\,\mathrm{d}q = p_x(q)\,\mathrm{d}q \int_{\tilde{r}=0}^{\tilde{r}=q/\sqrt{\xi}} \tilde{r} p_r(\tilde{r})\,\mathrm{d}\tilde{r} \\ + \frac{q}{\sqrt{\xi}}\left(p_Q(q)\,\mathrm{d}q - p_x(q)\,\mathrm{d}q \int_{\tilde{r}=0}^{\tilde{r}=q/\sqrt{\xi}} p_r(\tilde{r})\,\mathrm{d}\tilde{r}\right) \end{aligned} \tag{G.6}$$

$$I(q) = p_x(q) \int_{\tilde{r}=0}^{\tilde{r}=q/\sqrt{\xi}} \tilde{r} p_r(\tilde{r})\,\mathrm{d}\tilde{r} + \frac{q}{\sqrt{\xi}}\left(p_Q(q) - p_x(q) \int_{\tilde{r}=0}^{\tilde{r}=q/\sqrt{\xi}} p_r(\tilde{r})\,\mathrm{d}\tilde{r}\right). \tag{G.7}$$

Insertion of Eqs. (21) and (23) into Eq. (G.7) yields

$$\begin{aligned}I(q) \approx \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)^2\right] \\ \int_{\tilde{r}=0}^{\tilde{r}=q/\sqrt{\xi}} \tilde{r}\frac{1}{\sqrt{2\pi}\sigma_r} \exp\left[-\frac{1}{2}\left(\frac{\tilde{r}-\bar{r}}{\sigma_r}\right)^2\right] \mathrm{d}\tilde{r} \\ + \frac{q}{\sqrt{\xi}}\left(p_Q(q) - \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)^2\right]\right. \\ \left.\times \int_{\tilde{r}=0}^{\tilde{r}=q/\sqrt{\xi}} \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left[-\frac{1}{2}\left(\frac{\tilde{r}-\bar{r}}{\sigma_r}\right)^2\right] \mathrm{d}\tilde{r}\right).\end{aligned} \quad \text{(G.8)}$$

The integral

$$\int_{\tilde{r}=0}^{\tilde{r}=q/\sqrt{\xi}} \tilde{r}\frac{1}{\sqrt{2\pi}\sigma_r} \exp\left[-\frac{1}{2}\left(\frac{\tilde{r}-\bar{r}}{\sigma_r}\right)^2\right] \mathrm{d}\tilde{r} \quad \text{(G.9)}$$

can be solved by substituting $t := \frac{\tilde{r}-\bar{r}}{\sigma_r}$, which implies $\mathrm{d}\tilde{r} = \sigma_r \,\mathrm{d}t$ and $\tilde{r} = \sigma_r t + \bar{r}$. The upper bound $t_u = \frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r}$ and lower bound $t_l = -\frac{\bar{r}}{\sigma_r}$ follow. The lower bound $t_l$ tends to $-\infty$ for $N \to \infty$ because normalization yields $t_l = -N\frac{\bar{r}}{\sigma_r^* r^{(g)}}$. The application of the substitution results in

$$\begin{aligned}\int_{t=-\infty}^{t=\frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r}} (\sigma_r t + \bar{r})\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}\,\mathrm{d}t &= \sigma_r\frac{1}{\sqrt{2\pi}}\int_{t=-\infty}^{t=\frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r}} te^{-\frac{1}{2}t^2}\,\mathrm{d}t \\ &\quad + \bar{r}\int_{t=-\infty}^{t=\frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r}} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}\,\mathrm{d}t \quad \text{(G.10)} \\ &= -\sigma_r\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r}\right)^2\right] \\ &\quad + \bar{r}\Phi\left(\frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r}\right).\end{aligned} \quad \text{(G.11)}$$

In the last step, the identity $\int_{t=-\infty}^{t=x} te^{-\frac{1}{2}t^2}\mathrm{d}t = -e^{-\frac{1}{2}x^2}$ from [10, Equation A.16] and the cumulative distribution function of the standard normal distribution have been used. Insertion of Eq. (G.11) into Eq. (G.8) results in

$$
\begin{aligned}
I(q) \approx{}& \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)^2\right] \\
&\times \left\{-\sigma_r \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r}\right)^2\right] + \bar{r}\Phi\left(\frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r}\right)\right\} \\
&\quad + \frac{q}{\sqrt{\xi}}\left(p_Q(q) - \frac{1}{\sqrt{2\pi}\sigma^{(g)}} \exp\left[-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)^2\right]\right. \\
&\qquad \left. \times \int_{\tilde{r}=0}^{\tilde{r}=q/\sqrt{\xi}} \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left[-\frac{1}{2}\left(\frac{\tilde{r}-\bar{r}}{\sigma_r}\right)^2\right] \mathrm{d}\tilde{r}\right)
\end{aligned}
\tag{G.12}
$$

$$
\begin{aligned}
={}& \frac{1}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)^2} \\
&\times \left[\left(\bar{r} - \frac{q}{\sqrt{\xi}}\right)\Phi\left(\frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r}\right) - \frac{\sigma_r}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r}\right)^2}\right] \\
&+ \frac{q}{\sqrt{\xi}} p_Q(q)
\end{aligned}
\tag{G.13}
$$

after simplification.

*The Approximate $r$ Progress Rate in the case that the probability of feasible offspring tends to* 0. Insertion of Eq. (G.13) into Eq. (G.4) results in the expected $r$ value after projection of the best offspring for the infeasible case

$$
\begin{aligned}
\mathrm{E}[q_{r\,1;\lambda_{\text{infeas}}}] \approx{}& \lambda \int_{q=0}^{q=\infty} \frac{1}{\sqrt{2\pi}\sigma^{(g)}} e^{-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)^2} \\
&\times \left[\left(\bar{r} - \frac{q}{\sqrt{\xi}}\right)\Phi\left(\frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r}\right) - \frac{\sigma_r}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{q/\sqrt{\xi}-\bar{r}}{\sigma_r}\right)^2}\right] \\
&\quad [1-P_Q(q)]^{\lambda-1}\,\mathrm{d}q \\
&\quad + \frac{1}{\sqrt{\xi}} \underbrace{\lambda \int_{q=0}^{q=\infty} q\, p_Q(q)[1-P_Q(q)]^{\lambda-1}\,\mathrm{d}q}_{=\mathrm{E}[q_{1;\lambda}]}\,.
\end{aligned}
\tag{G.14}
$$

$$
\approx \frac{1}{\sqrt{\xi}}\mathrm{E}[q_{1;\lambda}]. \tag{G.15}
$$

In the last step the remaining integral is assumed to be negligible in the asymptotic case.

## H. Derivations for Expressions leading to the SAR Approximation

In this section, the expressions $f(\sigma^{(g)})$, $\frac{\partial f}{\partial \sigma}\Big|_{\sigma=\sigma^{(g)}}$, and $\frac{\partial^2 f}{\partial \sigma^2}\Big|_{\sigma=\sigma^{(g)}}$ are calculated where

$$f(\sigma) = \left(\frac{\sigma - \sigma^{(g)}}{\sigma^{(g)}}\right) \lambda \int_{q=0}^{q=\infty} p_Q(q \,|\, x^{(g)}, r^{(g)}, \sigma)[1 - P_Q(q)]^{\lambda-1} \,\mathrm{d}q. \tag{H.1}$$

Because $f(\sigma^{(g)}) = \frac{\sigma^{(g)} - \sigma^{(g)}}{\sigma^{(g)}} \times \cdots = 0$,

$$f(\sigma^{(g)}) = 0 \tag{H.2}$$

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

$$\begin{aligned}\frac{\partial f}{\partial \sigma} &= \frac{1}{\sigma^{(g)}} \lambda \int_{q=0}^{q=\infty} p_Q(q \,|\, x^{(g)}, r^{(g)}, \sigma)[1 - P_Q(q)]^{\lambda-1} \,\mathrm{d}q \\ &\quad + \frac{\sigma - \sigma^{(g)}}{\sigma^{(g)}} \lambda \int_{q=0}^{q=\infty} \frac{\partial}{\partial \sigma} p_Q(q \,|\, x^{(g)}, r^{(g)}, \sigma)[1 - P_Q(q)]^{\lambda-1} \,\mathrm{d}q.\end{aligned} \tag{H.3}$$

Therefore, it follows that

$$\begin{aligned}\frac{\partial f}{\partial \sigma}\Big|_{\sigma=\sigma^{(g)}} &= \frac{1}{\sigma^{(g)}} \underbrace{\lambda \int_{q=0}^{q=\infty} p_Q(q \,|\, x^{(g)}, r^{(g)}, \sigma^{(g)})[1 - P_Q(q)]^{\lambda-1} \,\mathrm{d}q}_{=1} \\ &\quad + \underbrace{\frac{\sigma^{(g)} - \sigma^{(g)}}{\sigma^{(g)}}}_{=0} \lambda \int_{q=0}^{q=\infty} \frac{\partial}{\partial \sigma} p_Q(q \,|\, x^{(g)}, r^{(g)}, \sigma) \\ &\quad \times [1 - P_Q(q)]^{\lambda-1} \,\mathrm{d}q\Big|_{\sigma=\sigma^{(g)}} = \frac{1}{\sigma^{(g)}}.\end{aligned} \tag{H.4}$$

Computing the derivative with respect to $\sigma$ of Eq. (H.3) using the product rule for the second summand results in

$$\begin{aligned}\frac{\partial^2 f}{\partial \sigma^2} &= \frac{2}{\sigma^{(g)}} \lambda \int_{q=0}^{q=\infty} \frac{\partial}{\partial \sigma} p_Q(q \,|\, x^{(g)}, r^{(g)}, \sigma)[1 - P_Q(q)]^{\lambda-1} \,\mathrm{d}q \\ &\quad + \frac{\sigma - \sigma^{(g)}}{\sigma^{(g)}} \lambda \int_{q=0}^{q=\infty} \frac{\partial^2}{\partial \sigma^2} p_Q(q \,|\, x^{(g)}, r^{(g)}, \sigma)[1 - P_Q(q)]^{\lambda-1} \,\mathrm{d}q.\end{aligned} \tag{H.5}$$

This implies that

$$\frac{\partial^2 f}{\partial \sigma^2}\Big|_{\sigma=\sigma^{(g)}} = \frac{2}{\sigma^{(g)}} \lambda \int_{q=0}^{q=\infty} \frac{\partial}{\partial \sigma} p_Q(q \,|\, x^{(g)}, r^{(g)}, \sigma)[1 - P_Q(q)]^{\lambda-1} \,\mathrm{d}q\Big|_{\sigma=\sigma^{(g)}}. \tag{H.6}$$

Hence, $\frac{\partial}{\partial \sigma} p_Q(q \,|\, x^{(g)}, r^{(g)}, \sigma)$ has to be derived next. To this end, Eqs. (24) and (25) are revisited. The infeasible case is treated first followed by the feasible case.

*H.1. The case that the probability of feasible offspring tends to* 0

For the derivation in the case that the probability of feasible offspring tends to 0, Eq. (25) is simplified further using $\sqrt{1+x} \simeq 1 + \frac{x}{2} + O(x^2)$

$$\bar{r} \simeq \sqrt{1 + \frac{{\sigma^{(g)}}^2 N}{{r^{(g)}}^2}} \simeq 1 + \frac{{\sigma^{(g)}}^2 N}{2{r^{(g)}}^2} \tag{H.7}$$

and (for $N \to \infty$) $\sigma^{(g)} \approx \sigma_r$ yielding

$$P_{Q_{\text{infeas}}}(q \,|\, x^{(g)}, r^{(g)}, \sigma) \approx \Phi\left(\frac{\left(1+\frac{1}{\xi}\right) q - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}}\left(1 + \frac{\sigma^2 N}{2{r^{(g)}}^2}\right)}{\sigma\sqrt{1+\frac{1}{\xi}}}\right) \tag{H.8}$$

and

$$p_{Q_{\text{infeas}}}(q \,|\, x^{(g)}, r^{(g)}, \sigma) = \frac{\mathrm{d}}{\mathrm{d}q} P_{Q_{\text{infeas}}}(q) \tag{H.9}$$

$$\approx \frac{1}{\sqrt{2\pi}} \frac{\sqrt{1+\frac{1}{\xi}}}{\sigma} \exp\left[-\frac{1}{2}\left(\frac{\left(1+\frac{1}{\xi}\right) q - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}}\left(1 + \frac{\sigma^2 N}{2{r^{(g)}}^2}\right)}{\sigma\sqrt{1+\frac{1}{\xi}}}\right)^2\right]. \tag{H.10}$$

Taking the derivative of Eq. (H.10) with respect to $\sigma$ one obtains

$$\begin{aligned}
&\frac{\partial}{\partial\sigma} p_{Q_{\text{infeas}}}(q \,|\, x^{(g)}, r^{(g)}, \sigma) \\
&= \frac{1}{\sqrt{2\pi}} \frac{\sqrt{1+\frac{1}{\xi}}}{\sigma^2} \left\{-1 + \left[\frac{\left(1+\frac{1}{\xi}\right) q - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}}}{\sigma\sqrt{1+\frac{1}{\xi}}} + \frac{\sigma N}{2\sqrt{\xi} r^{(g)} \sqrt{1+\frac{1}{\xi}}}\right]\right. \\
&\qquad \left. \times \left[\frac{\left(1+\frac{1}{\xi}\right) q - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}}\left(1 + \frac{\sigma^2 N}{2{r^{(g)}}^2}\right)}{\sigma\sqrt{1+\frac{1}{\xi}}}\right]\right\} \\
&\qquad \times \exp\left[-\frac{1}{2}\left(\frac{\left(1+\frac{1}{\xi}\right) q - x^{(g)} - \frac{r^{(g)}}{\sqrt{\xi}}\left(1 + \frac{\sigma^2 N}{2{r^{(g)}}^2}\right)}{\sigma\sqrt{1+\frac{1}{\xi}}}\right)^2\right].
\end{aligned} \tag{H.11}$$

Addition of the term $\frac{\sigma N}{2\sqrt{\xi} r^{(g)} \sqrt{1+\frac{1}{\xi}}} - \frac{\sigma N}{2\sqrt{\xi} r^{(g)} \sqrt{1+\frac{1}{\xi}}}$, which is equal to 0, to the expressions inside the first pair of square brackets, regrouping, and subsequent insertion of the resulting expression and Eq. (H.8) into Eq. (H.6) result in

$$\begin{aligned}\frac{\partial^2 f_{\text{infeas}}}{\partial\sigma^2}\bigg|_{\sigma=\sigma^{(g)}} &= \frac{2}{\sigma^{(g)}}\lambda\int_{q=0}^{q=\bar{r}\sqrt{\xi}}\frac{1}{\sqrt{2\pi}}\frac{\sqrt{1+\frac{1}{\xi}}}{\sigma^{(g)2}}\Bigg\{-1\\
&+\left[\frac{\left(1+\frac{1}{\xi}\right)q-x^{(g)}-\frac{r^{(g)}}{\sqrt{\xi}}\left(1+\frac{\sigma^{(g)2}N}{2r^{(g)2}}\right)}{\sigma^{(g)}\sqrt{1+\frac{1}{\xi}}}\right.\\
&\left.+\frac{\sigma^{(g)}N}{\sqrt{\xi}r^{(g)}\sqrt{1+\frac{1}{\xi}}}\right]\left[\frac{\left(1+\frac{1}{\xi}\right)q-x^{(g)}-\frac{r^{(g)}}{\sqrt{\xi}}\left(1+\frac{\sigma^{(g)2}N}{2r^{(g)2}}\right)}{\sigma^{(g)}\sqrt{1+\frac{1}{\xi}}}\right]\Bigg\}\\
&\times\exp\left[-\frac{1}{2}\left(\frac{\left(1+\frac{1}{\xi}\right)q-x^{(g)}-\frac{r^{(g)}}{\sqrt{\xi}}\left(1+\frac{\sigma^{(g)2}N}{2r^{(g)2}}\right)}{\sigma^{(g)}\sqrt{1+\frac{1}{\xi}}}\right)^2\right]\\
&\times\left[1-\Phi\left(\frac{\left(1+\frac{1}{\xi}\right)q-x^{(g)}-\frac{r^{(g)}}{\sqrt{\xi}}\left(1+\frac{\sigma^{(g)2}N}{2r^{(g)2}}\right)}{\sigma^{(g)}\sqrt{1+\frac{1}{\xi}}}\right)\right]^{\lambda-1}\mathrm{d}q.\end{aligned} \tag{H.12}$$

For solving this integral, $-t := \frac{\left(1+\frac{1}{\xi}\right)q-x^{(g)}-\frac{r^{(g)}}{\sqrt{\xi}}\left(1+\frac{\sigma^{(g)2}N}{2r^{(g)2}}\right)}{\sigma^{(g)}\sqrt{1+\frac{1}{\xi}}}$ is substituted. It further implies $\mathrm{d}q = -\frac{\sigma^{(g)}}{\sqrt{1+1/\xi}}\mathrm{d}t$. Expressing $t$ with normalized $\sigma^{(g)} = \frac{r^{(g)}\sigma^{(g)*}}{N}$ one obtains $t = -N\frac{\left(1+\frac{1}{\xi}\right)q-x^{(g)}-\frac{r^{(g)}}{\sqrt{\xi}}\left(1+\frac{\sigma^{(g)*2}}{2N}\right)}{\sigma^{(g)*}r^{(g)}\sqrt{1+\frac{1}{\xi}}}$. For $N\to\infty$, the integration bounds after substitution follow as $t_l=\infty$ and $t_u=-\infty$. The resulting integral reads

$$\frac{\partial^2 f_{\text{infeas}}}{\partial\sigma^2}\bigg|_{\sigma=\sigma^{(g)}} \tag{H.13}$$

$$= -\frac{2}{\sigma^{(g)2}}\frac{\lambda}{\sqrt{2\pi}}\int_{t=\infty}^{t=-\infty}\left\{-1+t^2-\frac{\sigma^{(g)}N}{\sqrt{1+\frac{1}{\xi}}\sqrt{\xi}r^{(g)}}t\right\}e^{-\frac{1}{2}t^2}[\underbrace{1-\Phi(-t)}_{=\Phi(t)}]^{\lambda-1}\,\mathrm{d}t \tag{H.14}$$

$$= \frac{2}{\sigma^{(g)2}}\frac{\lambda}{\sqrt{2\pi}}\int_{t=-\infty}^{t=\infty}\left\{-1+t^2-\frac{\sigma^{(g)}N}{\sqrt{1+\frac{1}{\xi}}\sqrt{\xi}r^{(g)}}t\right\}e^{-\frac{1}{2}t^2}[\Phi(t)]^{\lambda-1}\,\mathrm{d}t \tag{H.15}$$

$$= \frac{2}{\sigma^{(g)2}}\left(d_{1,\lambda}^{(2)}-1\right)-\frac{2}{\sigma^{(g)2}}\frac{\sigma^{(g)}N}{\sqrt{1+\frac{1}{\xi}}\sqrt{\xi}r^{(g)}}d_{1,\lambda}^{(1)}. \tag{H.16}$$

In the step from Eq. (H.14) to Eq. (H.15), the fact that negating an integral is equivalent to exchanging the upper and lower bounds and the identity 1 −

$\Phi(-t) = \Phi(t)$ have been used. In the step from Eq. (H.15) to Eq. (H.16) the higher-order progress coefficients (see Eq. (61)) $d_{1,\lambda}^{(1)}$ and $d_{1,\lambda}^{(2)}$ have been used to express the integrals. It holds that $d_{1,\lambda}^{(1)} = c_{1,\lambda}$.

*H.2. The case that the probability of feasible offspring tends to* 1

Now, the case that the probability of feasible offspring tends to 1 is considered in Eq. (H.6). Note that Eqs. (H.2) and (H.4) are the same for both cases, the infeasible case and the feasible case. To treat Eq. (H.6) further for the feasible case,

$$\frac{\partial}{\partial\sigma} p_{Q_{\text{feas}}}(q \,|\, x^{(g)}, r^{(g)}, \sigma) = \frac{\partial}{\partial\sigma}\left[\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma}\right)^2}\right] \tag{H.17}$$

has to be derived. With the use of the product rule and subsequent simplification one obtains

$$\frac{\partial}{\partial\sigma}\left[\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma}\right)^2}\right] = \frac{1}{\sqrt{2\pi}\sigma^2}\left[-1 + \left(\frac{q-x^{(g)}}{\sigma}\right)^2\right] e^{-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma}\right)^2}. \tag{H.18}$$

Insertion of this result and Eq. (25) into Eq. (H.6) results in

$$\left.\frac{\partial^2 f_{\text{feas}}}{\partial\sigma^2}\right|_{\sigma=\sigma^{(g)}} = \frac{2}{\sigma^{(g)}}\lambda \int_{q=\bar{r}\sqrt{\xi}}^{q=\infty} \frac{1}{\sqrt{2\pi}{\sigma^{(g)}}^2}\left[-1 + \left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)^2\right] e^{-\frac{1}{2}\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)^2} \times [1 - \Phi\left(\frac{q-x^{(g)}}{\sigma^{(g)}}\right)]^{\lambda-1}\,\mathrm{d}q. \tag{H.19}$$

The substitution $-t := \frac{q-x^{(g)}}{\sigma^{(g)}}$, which implies $\mathrm{d}q = -\sigma^{(g)}\mathrm{d}t$, and therefore for $N \to \infty$ lower bound $t_l = \infty$ and upper bound $t_u = -\infty$, yields

$$\left.\frac{\partial^2 f_{\text{feas}}}{\partial\sigma^2}\right|_{\sigma=\sigma^{(g)}} = \frac{2}{{\sigma^{(g)}}^2}\frac{\lambda}{\sqrt{2\pi}}\int_{t=-\infty}^{t=\infty}[-1+t^2]e^{-\frac{1}{2}t^2}[\Phi(t)]^{\lambda-1}\,\mathrm{d}t \tag{H.20}$$

$$= \frac{2}{{\sigma^{(g)}}^2}\left(d_{1,\lambda}^{(2)} - 1\right). \tag{H.21}$$

**Acknowledgments**

This work was supported by the Austrian Science Fund FWF under grant P29651-N32.

## References


[1] D. V. Arnold, On the behaviour of the (1, $\lambda$)-ES for a simple constrained problem, in: Proceedings of the 11th Workshop Proceedings on Foundations of Genetic Algorithms, ACM, 2011, pp. 15–24.

[2] D. V. Arnold, Analysis of a repair mechanism for the (1, $\lambda$)-ES applied to a simple constrained problem, in: Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation, ACM, 2011, pp. 853–860.

[3] M. Hellwig, D. V. Arnold, Comparison of constraint-handling mechanisms for the (1, $\lambda$)-ES on a simple constrained problem, Evolutionary Computation 24 (1) (2016) 1–23.

[4] D. V. Arnold, On the behaviour of the (1, $\lambda$)-ES for a conically constrained problem, in: Proceedings of the 15th Annual Conference on Genetic and Evolutionary Computation, ACM, 2013, pp. 423–430.

[5] O. Kramer, Premature convergence in constrained continuous search spaces, in: International Conference on Parallel Problem Solving from Nature, Springer, 2008, pp. 62–71.

[6] D. V. Arnold, J. Porter, Towards an augmented lagrangian constraint handling approach for the (1+1)-ES, in: Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation, GECCO '15, ACM, New York, NY, USA, 2015, pp. 249–256. doi:10.1145/2739480.2754813. URL http://doi.acm.org/10.1145/2739480.2754813

[7] A. Atamna, A. Auger, 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.

[8] A. Atamna, A. Auger, N. Hansen, Linearly convergent evolution strategies via augmented lagrangian constraint handling, in: Proceedings of the 14th ACM/SIGEVO Conference on Foundations of Genetic Algorithms, ACM, 2017, pp. 149–161.

[9] P. Spettel, H.-G. Beyer, Technical report: Analysis of the $(1,\lambda)$-$\sigma$-self-adaptation evolution strategy with repair by projection applied to a conically constrained problem, Tech. rep., Vorarlberg University of Applied Sciences (2018). URL https://www.fhv.at/fileadmin/user_upload/fhv/files/forschung/ppe/working-papers/Analysis_onecommalambdasigmaSA-ES_Projection_Cone_TR.pdf

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

[11] M. Fisz, Wahrscheinlichkeitsrechnung und mathematische Statistik, Hochschulbücher für Mathematik, Berlin: VEB Deutscher Verlag der Wissenschaften, 1971.

[12] S. Meyer-Nieberg, Self-adaptation in evolution strategies, Ph.D. thesis, Dortmund University of Technology (12 2007).

[13] B. Arnold, N. Balakrishnan, H. Nagaraja, A First Course in Order Statistics, Wiley Series in Probability and Statistics, Wiley, 1992.

[14] C. Forbes, M. Evans, N. Hastings, B. Peacock, Statistical Distributions, Wiley, 2011.

[15] I. Rechenberg, The Evolution Strategy. A Mathematical Model of Darwinian Evolution, Springer Berlin Heidelberg, Berlin, Heidelberg, 1984.