
Analysis of the $(\mu/\mu_I, \lambda)$ -CSA-ES with Repair by Projection Applied to a Conically Constrained Problem

Patrick Spettel

patrick.spettel@fhv.at

Research Center Process and Product Engineering, Vorarlberg University of Applied Sciences, Dornbirn, 6850, Austria

Hans-Georg Beyer

hans-georg.beyer@fhv.at

Department of Computer Science, Research Center Process and Product Engineering, Vorarlberg University of Applied Sciences, Dornbirn, 6850, Austria

Abstract

Theoretical analyses of evolution strategies are indispensable for gaining a deep understanding of their inner workings. For constrained problems, rather simple problems are of interest in the current research. This work presents a theoretical analysis of a multi-recombinative evolution strategy with cumulative step size adaptation applied to a conically constrained linear optimization problem. The state of the strategy is modeled by random variables and a stochastic iterative mapping is introduced. For the analytical treatment, fluctuations are neglected and the mean value iterative system is considered. Non-linear difference equations are derived based on one-generation progress rates. Based on that, expressions for the steady state of the mean value iterative system are derived. By comparison with real algorithm runs, it is shown that for the considered assumptions, the theoretical derivations are able to predict the dynamics and the steady state values of the real runs.

Keywords

Evolution strategy, constraint handling, repair by projection, cumulative step size adaptation, conically constrained problem.

1 Introduction

Thorough theoretical investigations of evolution strategies (ESs) are necessary for gaining a deep understanding of how they work. A lot of research has been done for analyzing ESs applied to unconstrained problems. For the constrained setting, there are still many aspects for which a deep theoretical understanding is missing. Theoretical comparisons of different infeasibility handling techniques are of interest. In addition, theoretical investigations of parameter choices have not yet been deeply investigated for constrained problems. Furthermore, theoretical comparisons of mutation strength control mechanisms are not fully explored. Moreover, as all of those heavily depend on the problem under consideration, more problems that model specific parts of practical constrained optimization problems need to be considered. As a step in that direction, this work theoretically analyzes a $(\mu/\mu_I, \lambda)$ -ES with cumulative step size adaptation (CSA) applied to a conically constrained linear problem.

Regarding related work, a $(1, \lambda)$ -ES with constraint handling by discarding infeasible offspring has been analyzed by Arnold (2011b) for a single linear constraint. Repair

by projection has been considered (Arnold, 2011a) and a comparison with repair by reflection and repair by truncation has been performed by Hellwig and Arnold (2016). In all of that work, analysis of the respective algorithms' one-generation and steady state behavior considering CSA is performed. Expressions with integrals are numerically integrated. Based on Lagrangian constraint handling, Arnold and Porter (2015) presented a $(1 + 1)$ -ES applied to a single linear inequality constraint with the sphere model. The one-generation behavior has been analyzed in that work. The insights gained from that analysis have been used to develop an algorithm that is experimentally evaluated.

A theoretical investigation based on Markov chains for a multi-recombinative variant with Lagrangian constraint handling has been presented by Atamna et al. (2016). Investigation of a single linear constraint in that work has been extended to multiple linear constraints (Atamna et al., 2017).

The simple linearly constrained optimization problem with a single linear constraint used in Arnold (2011b), Arnold (2011a), and Hellwig and Arnold (2016) has been used as one step toward a better understanding of ESs for constrained optimization. However, as it does not have an optimal solution, it does only model parts of the difficulties in handling constrained optimization problems. In particular, if the linear objective function gradient points toward the constraint, the ES has to avoid premature convergence. As a step further, Arnold (2013) has considered a conically constrained linear problem in which the linear objective function gradient coincides with the cone axis. As an extension, Arnold (2014) has proposed the more general problem that does not assume that the objective function gradient points in direction of the cone axis. However, in the work presented in this paper, the choice of the objective function gradient in direction of the cone axis is used because it simplifies the theoretical analysis. This problem models the situation of the optimum lying on the boundary of the feasibility region in addition to the premature convergence difficulty. Hence, the ES has to avoid premature convergence but at the same time the steps need to get smaller and smaller during the process of approaching the optimum. Arnold (2013) has considered a $(1, \lambda)$ -ES applied to that problem by discarding infeasible offspring. Integral expressions for the one-generation behavior have been derived. Numerical integration has been used to analyze the steady state behavior. For the σ control strategy, CSA has been used. Spettel and Beyer (2018a) have considered the same problem and have analyzed a $(1, \lambda)$ - σ -Self-Adaptation ES (σ SA-ES). It has been extended to the multi-recombinative $(\mu/\mu_I, \lambda)$ variant (Spettel and Beyer, 2018b). In contrast to the work by Arnold, closed-form expressions have been derived under asymptotic assumptions for the one-generation behavior, the mean value dynamics, and the steady state behavior. This enables to directly analyze the influence of the different algorithmic parameters for the algorithm behavior.

The contribution of this paper is the analysis considering CSA instead of σ SA for the mutation strength control mechanism in the $(\mu/\mu_I, \lambda)$ -ES with repair by projection applied to the conically constrained linear problem. Based on the derivations for the microscopic behavior in Spettel and Beyer (2018b), approximate closed-form expressions are derived for predicting the mean value dynamics and the steady state behavior of the multi-recombinative CSA-ES with projection. Multi-recombinative ESs with projection have already shown their suitability for constrained problems in real-world applications. Beyer and Finck (2012) have considered repair by projection in the context of a portfolio optimization problem. This design gave rise to a more principled design using projection for optimization problems with linear constraints (Spettel et al., 2018). In line with those empirical investigations, the theoretical analysis in this work concerns

multi-recombinative ESs. In order to simplify the analysis, isotropic mutations are considered. This removes the need to analyze the covariance matrix adaptation. Again, the aim is to derive closed-form expressions under asymptotic considerations for predicting the mean value dynamics and the steady state behavior. Because the CSA is a completely different approach for adapting the mutation strength σ in comparison to σ SA, one main interest is the comparison of those two σ control mechanisms. CSA uses information gained from the observed steps taken in the search space, whereas σ SA uses the information from the objective function space. Furthermore, it is not immediately apparent on how the cumulation parameter in the CSA should be chosen. The theoretical analysis gives insights in this regard. A particular focus of the analysis lies on the difference between the choice of $c = \frac{1}{\sqrt{N}}$ proposed in early publications (Hansen and Ostermeier, 1997) on the covariance matrix adaptation evolution strategy (CMA-ES) and $c = \Theta\left(\frac{1}{N}\right)$ that is proposed in newer publications (Hansen and Ostermeier, 2001; Hansen, 2016).

The remainder of the paper is organized as follows. Section 2 introduces the optimization problem under consideration and describes the algorithm that is analyzed. Section 3 concerns the theoretical analysis. First, a mean value iterative system that models the dynamics of the ES is derived in Section 3.1. Second, steady state considerations are shown in Section 3.2. For the theoretical considerations, plots comparing them to results of real ES runs are presented for showing the approximation quality. Section 4 discusses the results and Section 5 concludes the paper.

2 Problem and Algorithm

Minimization of

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

subject to constraints

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

$$x_1 \geq 0 \quad (3)$$

is considered in this work ($\mathbf{x} = (x_1, \dots, x_N)^T \in \mathbb{R}^N$ and $\xi > 0$).

The state of an ES individual can be uniquely described in the $(x, r)^T$ -space. It consists of x , the distance from 0 in x_1 -direction (cone axis), and r , the distance from the cone axis. Because isotropic mutations are considered in the ES, the coordinate system can be rotated (w.l.o.g.) such that $(\tilde{x}, \tilde{r})^T$ corresponds to $(\tilde{x}, \tilde{r}, 0, \dots, 0)^T$ in the parameter space. Figure 1 visualizes the problem. The equation for the cone boundary is $r = \frac{x}{\sqrt{\xi}}$, which follows from Equation (2). The projection line can be derived using the cone direction vector $\left(1, \frac{1}{\sqrt{\xi}}\right)^T$ and its counterclockwise rotation by 90 degrees $\left(-\frac{1}{\sqrt{\xi}}, 1\right)^T$ yielding $r = -\sqrt{\xi}x + q\left(\sqrt{\xi} + \frac{1}{\sqrt{\xi}}\right)$. The values \mathbf{x} and $\tilde{\mathbf{x}}$ denote a parental individual and an offspring individual, respectively. The corresponding mutation is indicated as $\tilde{\sigma}\mathbf{z}$. The values of x and r after projection are denoted by q and q_r , respectively.

As motivated in the introduction, the algorithm to be analyzed is a $(\mu/\mu_I, \lambda)$ -CSA-ES with repair by projection applied to the problem introduced above. Its pseudo code is shown in Algorithm 1. In the beginning, the parameters are initialized (Lines 1 to 2).

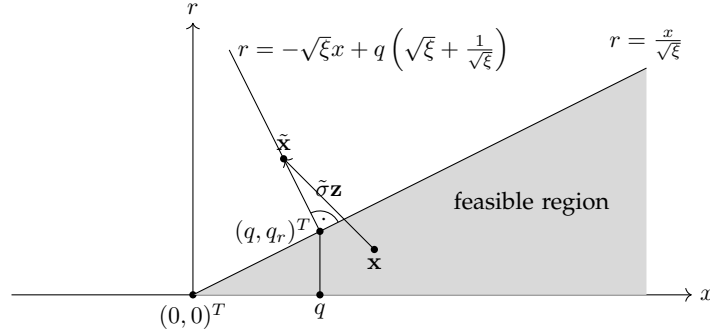


Figure 1: The conically constrained optimization problem in N dimensions shown in the $(x, r)^T$ -space. As shown in the picture, the offspring individual $\tilde{\mathbf{x}}$ is infeasible and therefore projected onto the cone boundary at $(q, q_r)^T$.

In the generational loop, λ offspring are created (Lines 6 to 16). Each offspring's parameter vector is sampled from a multivariate normal distribution with mean $\mathbf{x}^{(g)}$ and standard deviation $\sigma^{(g)}$ in Lines 7 and 8. If the generated offspring is infeasible ($\text{isFeasible}(\mathbf{x}) = x_1 \geq 0 \wedge x_1^2 - \xi \sum_{k=2}^N x_k^2 \geq 0$), its parameter vector is projected onto the point on the boundary of the feasible region that minimizes the Euclidean distance to the offspring point. The corresponding mutation vector leading to this repaired point is calculated back (Lines 9 to 12). Projection means solving the optimization problem

$$\begin{aligned} \hat{\mathbf{y}} &= \arg \min_{\mathbf{y}} \|\mathbf{y} - \mathbf{x}\|^2 \\ \text{s.t. } &y_1^2 - \xi \sum_{k=2}^N y_k^2 \geq 0 \\ &y_1 \geq 0 \end{aligned} \quad (4)$$

where \mathbf{x} is the individual to be projected. The function

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

is introduced, which returns $\hat{\mathbf{y}}$ of the problem stated in (4). Appendix A in the supplementary material of Spettel and Beyer (2018b) shows a geometrical approach for deriving a closed-form solution to the projection optimization problem (4) that this work builds on. Given an infeasible individual \mathbf{x} , it reads

$$\hat{\mathbf{y}} = \begin{cases} \frac{\xi}{\xi+1} \left(x_1 + \frac{\|\mathbf{r}\|}{\sqrt{\xi}} \right) \left(1, \frac{x_2}{\sqrt{\xi}\|\mathbf{r}\|}, \dots, \frac{x_N}{\sqrt{\xi}\|\mathbf{r}\|} \right)^T & \text{if } \sqrt{\frac{\xi}{\xi+1}} \left(x_1 + \frac{\|\mathbf{r}\|}{\sqrt{\xi}} \right) > 0 \\ \mathbf{0} & \text{otherwise,} \end{cases} \quad (6)$$

where $\|\mathbf{r}\| = \sqrt{\sum_{k=2}^N x_k^2}$. After possible repair, the offspring's fitness is determined in Line 13. The next generation's parental individual $\mathbf{x}^{(g+1)}$ (Line 18) and the next generation's mutation strength $\sigma^{(g+1)}$ (Line 20) are computed next. The next generation's parental parameter vector is set to the mean of the μ best (w.r.t. fitness) offspring parameter vectors¹. For the mutation strength update, first the cumulative s-vector is

¹Note that the order statistic notation $m; \lambda$ is used to denote the m -th best (w.r.t. fitness) out of λ values. The notation $(\mathbf{x})_k$ is used to denote the k -th element of a vector \mathbf{x} . It is equivalent to writing x_k .

updated. The cumulation parameter c determines the fading strength. The mutation strength is then updated using this \mathbf{s} -vector. The parameter D acts as a damping factor. If the squared length of the \mathbf{s} -vector is smaller than N , the step size is decreased. Otherwise, the step size is increased. Intuitively, this means that multiple correlated steps allow a larger step size and vice versa. The update of the generation counter ends one iteration of the generation loop. The values $x^{(g)}$, $r^{(g)}$, q_l , \mathbf{q}'_l , $\langle q \rangle$, and $\langle q_r \rangle$ 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, 14, 15, 21 and 22, respectively.

Figure 2 shows an example of the x - and r -dynamics of Algorithm 1 (solid line) in comparison with results of the closed-form approximate iterative system (dotted line) that is derived in the sections that follow. As one can see, the theoretical prediction comes close to the real dynamics for the case shown.

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

```

1: Initialize  $\mathbf{x}^{(0)}, \mathbf{s}^{(0)}, \sigma^{(0)}, c, D, \lambda, \mu$ 
2:  $g \leftarrow 0$ 
3: repeat
4:    $x^{(g)} = (\mathbf{x}^{(g)})_1$ 
5:    $r^{(g)} = \sqrt{\sum_{k=2}^N (\mathbf{x}^{(g)})_k^2}$ 
6:   for  $l \leftarrow 1$  to  $\lambda$  do
7:      $\tilde{\mathbf{z}}_l \leftarrow \mathcal{N}_l(\mathbf{0}, \mathbf{I})$  ▷ sample from normal distribution
8:      $\tilde{\mathbf{x}}_l \leftarrow \mathbf{x}^{(g)} + \sigma^{(g)} \tilde{\mathbf{z}}_l$ 
9:     if not isFeasible( $\tilde{\mathbf{x}}_l$ ) then
10:       $\tilde{\mathbf{x}}_l \leftarrow \text{repair}(\tilde{\mathbf{x}}_l)$  ▷ see Equations (4) and (5)
11:       $\tilde{\mathbf{z}}_l \leftarrow (\tilde{\mathbf{x}}_l - \mathbf{x}^{(g)})/\sigma^{(g)}$ 
12:     end if
13:      $\tilde{f}_l \leftarrow f(\tilde{\mathbf{x}}_l) = (\tilde{\mathbf{x}}_l)_1$  ▷ determine fitness of the offspring
14:      $q_l = (\tilde{\mathbf{x}}_l)_1$ 
15:      $\mathbf{q}'_l = \tilde{\mathbf{x}}_l$ 
16:   end for
17:   Sort offspring according to  $\tilde{f}_l$  in ascending order
18:    $\mathbf{x}^{(g+1)} \leftarrow \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\mathbf{x}}_{m;\lambda}$  ▷ compute centroid of the  $\mu$  best offspring
19:    $\mathbf{s}^{(g+1)} \leftarrow (1-c)\mathbf{s}^{(g)} + \sqrt{\mu c(2-c)} \underbrace{\left( \frac{1}{\mu} \sum_{m=1}^{\mu} \tilde{\mathbf{z}}_{m;\lambda} \right)}_{\langle \tilde{\mathbf{z}} \rangle}$  ▷ compute next  $\mathbf{s}$ 
20:    $\sigma^{(g+1)} \leftarrow \sigma^{(g)} \exp\left(\frac{\|\mathbf{s}\|^2 - N}{2DN}\right)$  ▷ compute next  $\sigma$ 
21:    $\langle q \rangle = (\mathbf{x}^{(g+1)})_1$ 
22:    $\langle q_r \rangle = \sqrt{\sum_{k=2}^N (\mathbf{x}^{(g+1)})_k^2}$ 
23:    $g \leftarrow g + 1$ 
24: until termination criteria are met

```

3 Theoretical Analysis

To completely describe the state of the ES, the random variables σ , \mathbf{s} , and the squared length $\|\mathbf{s}\|^2$ need to be modeled in addition to the variables for the position in the parameter space, x and r . The random vector \mathbf{s} is decomposed into its component along the cone axis $s_1^{(g)}$ and the remaining part in direction of the parental individual's

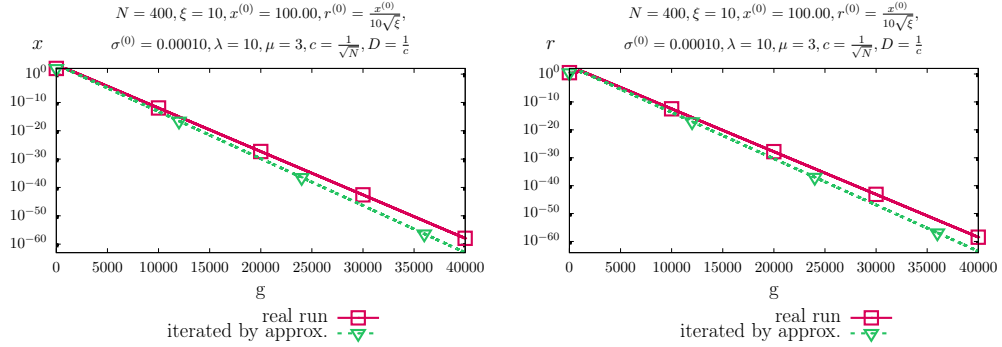


Figure 2: Real $(3/3_I, 10)$ -CSA-ES run mean value dynamics (solid line) in comparison to the iteration of the closed-form (approximate) iterative system (dotted line).

2..N components

$$s_{\odot}^{(g)} := \frac{1}{r^{(g)}} \sum_{k=2}^N (\mathbf{x}^{(g)})_k (s^{(g)})_k. \quad (7)$$

3.1 Derivation of a Mean Value Iterative System for Modeling the Dynamics of the ES

The aim of this section is to derive (approximate) closed-form expressions for predicting the mean value dynamics of the ES. To make the distinction between the random variable and its mean value in the iterative system clear, $\bar{z} := E[z]$ is used to denote the expected value of a random variate z . Thus, the iterative system can be written as

$$\begin{pmatrix} \overline{x^{(g+1)}} \\ \overline{r^{(g+1)}} \\ \overline{s_1^{(g+1)}} \\ \overline{s_{\odot}^{(g+1)}} \\ \overline{\|\mathbf{s}^{(g+1)}\|^2} \\ \overline{\sigma^{(g+1)}} \end{pmatrix} \leftarrow \begin{pmatrix} \overline{x^{(g)}} \\ \overline{r^{(g)}} \\ \overline{s_1^{(g)}} \\ \overline{s_{\odot}^{(g)}} \\ \overline{\|\mathbf{s}^{(g)}\|^2} \\ \overline{\sigma^{(g)}} \end{pmatrix} + \begin{pmatrix} \epsilon_{x^{(g)}} \\ \epsilon_{r^{(g)}} \\ \epsilon_{s_1^{(g)}} \\ \epsilon_{s_{\odot}^{(g)}} \\ \epsilon_{\|\mathbf{s}^{(g)}\|^2} \\ \epsilon_{\sigma^{(g)}} \end{pmatrix}, \quad (8)$$

where $\overline{x^{(g)}}$, $\overline{r^{(g)}}$, $\overline{s_1^{(g)}}$, $\overline{s_{\odot}^{(g)}}$, $\overline{\|\mathbf{s}^{(g)}\|^2}$, and $\overline{\sigma^{(g)}}$ represent the expected value parts and $\epsilon_{x^{(g)}}$, $\epsilon_{r^{(g)}}$, $\epsilon_{s_1^{(g)}}$, $\epsilon_{s_{\odot}^{(g)}}$, $\epsilon_{\|\mathbf{s}^{(g)}\|^2}$, and $\epsilon_{\sigma^{(g)}}$ represent the fluctuation parts. Assuming that the fluctuation parts are negligible, the mean value iterative system

$$\begin{pmatrix} \overline{x^{(g+1)}} \\ \overline{r^{(g+1)}} \\ \overline{s_1^{(g+1)}} \\ \overline{s_{\odot}^{(g+1)}} \\ \overline{\|\mathbf{s}^{(g+1)}\|^2} \\ \overline{\sigma^{(g+1)}} \end{pmatrix} \leftarrow \begin{pmatrix} \overline{x^{(g)}} \\ \overline{r^{(g)}} \\ \overline{s_1^{(g)}} \\ \overline{s_{\odot}^{(g)}} \\ \overline{\|\mathbf{s}^{(g)}\|^2} \\ \overline{\sigma^{(g)}} \end{pmatrix} \quad (9)$$

can be stated. This section presents derivations of difference equations for the system (9). Using those derived difference equations allows to approximately predict the

mean value behavior by iterating the mean value system. In Section 3.1.1, difference equations are presented for expressing $\overline{x^{(g+1)}}$ with $\overline{x^{(g)}}$ and $\overline{r^{(g+1)}}$ with $\overline{r^{(g)}}$ by using the respective local progress rates. Section 3.1.2, Section 3.1.3, and Section 3.1.4 deal with the derivation of difference equations for $\overline{s_1}$, $\overline{s_\odot}$, and $\overline{\|s^{(g)}\|^2}$, respectively. They are derived from the corresponding steps of Algorithm 1 and they also make use of the local progress rates. Finally, the difference equation for $\overline{\sigma}$ is stated in Section 3.1.5, the derived system of equations is summarized, and it is compared to real ES runs in Section 3.1.6.

3.1.1 Derivation of Mean Value Difference Equations for x and r

The starting points for the derivation of mean value difference equations for x and r are the progress rates in x and r direction. Their definitions read

$$\varphi_x(\overline{x^{(g)}}, \overline{r^{(g)}}, \overline{\sigma^{(g)}}) := \mathbb{E}[\overline{x^{(g)}} - x^{(g+1)} \mid \overline{x^{(g)}}, \overline{r^{(g)}}, \overline{\sigma^{(g)}}] \quad (10)$$

$$\varphi_r(\overline{x^{(g)}}, \overline{r^{(g)}}, \overline{\sigma^{(g)}}) := \mathbb{E}[\overline{r^{(g)}} - r^{(g+1)} \mid \overline{x^{(g)}}, \overline{r^{(g)}}, \overline{\sigma^{(g)}}]. \quad (11)$$

They describe the one-generation expected change in the parameter space. The normalizations

$$\varphi_x^* := \frac{N\varphi_x}{\overline{x^{(g)}}}, \quad (12)$$

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

and

$$\sigma^* := \frac{N\sigma}{\overline{r^{(g)}}} \quad (14)$$

are introduced in order to have quantities that are independent of the position in the search space. Using Equation (10) with Equation (12) and Equation (11) with Equation (13), the equations

$$\overline{x^{(g+1)}} = \overline{x^{(g)}} - \frac{\overline{x^{(g)}}\varphi_x^{(g)*}}{N} = \overline{x^{(g)}} \left(1 - \frac{\varphi_x^{(g)*}}{N}\right) \quad (15)$$

$$\overline{r^{(g+1)}} = \overline{r^{(g)}} - \frac{\overline{r^{(g)}}\varphi_r^{(g)*}}{N} = \overline{r^{(g)}} \left(1 - \frac{\varphi_r^{(g)*}}{N}\right) \quad (16)$$

follow. Approximations for $\varphi_x^{(g)*}$ and $\varphi_r^{(g)*}$ have already been derived by Spettel and Beyer (2018b, Equations (37) and (38)). They are recapped here to make this work self-containing. Note that in this regard the present work builds on Spettel and Beyer (2018b), in which expressions for $\varphi_x^{(g)*}$ and $\varphi_r^{(g)*}$ have been derived under the asymptotic assumptions of sufficiently large values of ξ and N . In those derivations, two cases have been distinguished. If one considers the ES being far from the cone boundary, offspring are feasible with overwhelming probability. The opposite case of being in the vicinity of the cone boundary results in infeasible offspring almost surely. These observations allow simplifications for the former case because the projection can be ignored. Both cases are combined into single equations by weighting the feasible and infeasible cases with an approximation for the offspring feasibility and offspring infeasibility

probability, respectively. The r -distribution in those derivations has been approximated by a normal distribution $\mathcal{N}(\bar{r}, \sigma_r^2)$ where

$$\bar{r} = \overline{r^{(g)}} \sqrt{1 + \frac{\overline{\sigma^{(g)*^2}}}{N} \left(1 - \frac{1}{N}\right)} \quad (17)$$

and

$$\sigma_r = \overline{r^{(g)}} \frac{\overline{\sigma^{(g)*}}}{N} \sqrt{\frac{1 + \frac{\overline{\sigma^{(g)*^2}}}{2N} \left(1 - \frac{1}{N}\right)}{1 + \frac{\overline{\sigma^{(g)*^2}}}{N} \left(1 - \frac{1}{N}\right)}} \quad (18)$$

(it is referred to Appendix B in the supplementary material of Spettel and Beyer (2018b) for the detailed derivation). The results that build the basis for the following CSA analysis are briefly recapped here in order to make the proposed work self-containing.² The expression for $\varphi_x^{(g)*}$ has been derived as

$$\begin{aligned} \varphi_x^{(g)*} &\approx P_{\text{feas}}(\overline{x^{(g)}}, \overline{r^{(g)}}, \overline{\sigma^{(g)}}) \left[\frac{\overline{r^{(g)}}}{\overline{x^{(g)}}} \overline{\sigma^{(g)*}} c_{\mu/\mu, \lambda} \right] + [1 - P_{\text{feas}}(\overline{x^{(g)}}, \overline{r^{(g)}}, \overline{\sigma^{(g)}})] \\ &\times \underbrace{\left[\frac{N}{1 + \xi} \left(1 - \frac{\sqrt{\xi} \overline{r^{(g)}}}{\overline{x^{(g)}}} \sqrt{1 + \frac{\overline{\sigma^{(g)*^2}}}{N}} \right) + \frac{\sqrt{\xi}}{1 + \xi} \frac{\sqrt{\xi} \overline{r^{(g)}}}{\overline{x^{(g)}}} \overline{\sigma^{(g)*}} c_{\mu/\mu, \lambda} \sqrt{1 + \frac{1}{\xi} \frac{1 + \frac{\overline{\sigma^{(g)*^2}}}{2N}}{1 + \frac{\overline{\sigma^{(g)*^2}}}{N}}} \right]}_{=:\varphi_{x \text{ infeas}}^{*(g)}} \end{aligned} \quad (19)$$

in Spettel and Beyer (2018b) and the one for $\varphi_r^{(g)*}$ has been derived as

$$\begin{aligned} \varphi_r^{(g)*} &\approx P_{\text{feas}}(\overline{x^{(g)}}, \overline{r^{(g)}}, \overline{\sigma^{(g)}}) N \left(1 - \sqrt{1 + \frac{\overline{\sigma^{(g)*^2}}}{\mu N}} \right) \\ &+ [1 - P_{\text{feas}}(\overline{x^{(g)}}, \overline{r^{(g)}}, \overline{\sigma^{(g)}})] N \left(1 - \frac{\overline{x^{(g)}}}{\sqrt{\xi} \overline{r^{(g)}}} \left(1 - \frac{\varphi_{x \text{ infeas}}^{*(g)}}{N} \right) \sqrt{\frac{1 + \frac{\overline{\sigma^{(g)*^2}}}{\mu N}}{1 + \frac{\overline{\sigma^{(g)*^2}}}{N}}} \right) \end{aligned} \quad (20)$$

in Spettel and Beyer (2018b). An approximation for the offspring feasibility probability has been derived in Spettel and Beyer (2018b) as

$$P_{\text{feas}}(\overline{x^{(g)}}, \overline{r^{(g)}}, \overline{\sigma^{(g)}}) \simeq \Phi \left[\frac{1}{\overline{\sigma^{(g)}}} \left(\frac{\overline{x^{(g)}}}{\sqrt{\xi}} - \bar{r} \right) \right], \quad (21)$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. $\varphi_{x \text{ infeas}}^{*(g)}$ denotes the infeasible part of Equation (19). The constant $c_{\mu/\mu, \lambda}$ is a so-called progress coefficient. A definition is given in (Beyer, 2001, Eq. 6.102, p. 247). It reads

$$c_{\mu/\mu, \lambda} := \frac{\lambda - \mu}{2\pi} \binom{\lambda}{\mu} \int_{t=-\infty}^{t=\infty} e^{-t^2} [\Phi(t)]^{\lambda-\mu-1} [1 - \Phi(t)]^{\mu-1} dt. \quad (22)$$

²In the further considerations, the symbols “ \simeq ” and “ \approx ” are used. Expressions in the form of lhs \simeq rhs denote that lhs is asymptotically equal to rhs for given asymptotical assumptions (e.g. $N \rightarrow \infty$). The particular assumptions are stated explicitly for every use of “ \simeq ”. That is, in the limit case of the given assumptions, lhs is equal to rhs. The form lhs \approx rhs is used for cases where rhs is an approximation for lhs with given assumptions that are not of asymptotical nature. In this sense, “ \approx ” is weaker than “ \simeq ”.

3.1.2 Derivation of a Mean Value Difference Equation for s_1

For s_1 , a mean value difference equation can be derived using the update rule from Line 19 of Algorithm 1. Computation of the expected value with $\overline{s_1^{(g+1)}} := E[s_1^{(g+1)}]$ directly yields

$$\overline{s_1^{(g+1)}} = (1 - c)\overline{s_1^{(g)}} + \sqrt{\mu c(2 - c)}E[(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_1]. \quad (23)$$

$E[(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_1]$ can be expressed with the progress rate in x -direction φ_x . From the definition of the x progress rate,

$$\varphi_x^{(g)} = E[x^{(g)} - x^{(g+1)}] = E[x^{(g)} - (x^{(g)} + \sigma^{(g)}(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_1)] = -\overline{\sigma^{(g)}}E[(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_1] \quad (24)$$

follows. Therefore,

$$E[(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_1] = -\frac{\varphi_x^{(g)}}{\overline{\sigma^{(g)}}} \quad (25)$$

holds. Using Equation (25) and Equation (14),

$$\overline{s_1^{(g+1)}} = (1 - c)\overline{s_1^{(g)}} + \sqrt{\mu c(2 - c)} \left(-\frac{N\varphi_x^{(g)}}{\overline{\sigma^{(g)}}^* \overline{r^{(g)}}} \right) \quad (26)$$

follows.

3.1.3 Derivation of a Mean Value Difference Equation for s_{\odot}

The update rule from Line 19 of Algorithm 1 with consideration of Equation (7) can be used to derive a mean value difference equation for s_{\odot} . The detailed derivation is provided in the supplementary material (Appendix A) leading to

$$\begin{aligned} \overline{s_{\odot}^{(g+1)}} &\simeq (1 - c) \left(1 + \frac{\overline{\sigma^{(g)}}^*}{N} \left(-\frac{N\varphi_{r,2}^{(g)}}{2\overline{\sigma^{(g)}}^* \overline{r^{(g)}}^2} - \frac{\overline{\sigma^{(g)}}^*}{2\mu} \right) \right) \overline{s_{\odot}^{(g)}} \\ &\quad + \sqrt{\mu c(2 - c)} \left(-\frac{N\varphi_{r,2}^{(g)}}{2\overline{\sigma^{(g)}}^* \overline{r^{(g)}}^2} + \frac{\overline{\sigma^{(g)}}^*}{2\mu} \right). \end{aligned} \quad (27)$$

3.1.4 Derivation of a Mean Value Difference Equation for $\|\mathbf{s}\|^2$

Again, using the update rule from Line 19 of Algorithm 1,

$$\|\mathbf{s}^{(g+1)}\|^2 = \|(1 - c)\mathbf{s}^{(g)} + \sqrt{\mu c(2 - c)}\langle \tilde{\mathbf{z}}^{(g)} \rangle\|^2 \quad (28)$$

$$= (1 - c)^2 \|\mathbf{s}^{(g)}\|^2 + 2(1 - c)\sqrt{\mu c(2 - c)}\mathbf{s}^{(g)T} \langle \tilde{\mathbf{z}}^{(g)} \rangle + \mu c(2 - c) \|\langle \tilde{\mathbf{z}}^{(g)} \rangle\|^2 \quad (29)$$

can be derived. For treating $\mathbf{s}^{(g)T} \langle \tilde{\mathbf{z}}^{(g)} \rangle$, the vector $\mathbf{s}^{(g)}$ can be decomposed into a sum of vectors in direction of the cone axis $\mathbf{e}_1^{(g)}$, in direction of the parental individual's 2..N components $\mathbf{e}_{\odot}^{(g)}$, and in a direction $\mathbf{e}_{\ominus}^{(g)}$ that is orthogonal to $\mathbf{e}_1^{(g)}$ and $\mathbf{e}_{\odot}^{(g)}$. Formally, this can be written as

$$\mathbf{s}^{(g)} = s_1^{(g)} \mathbf{e}_1^{(g)} + s_{\odot}^{(g)} \mathbf{e}_{\odot}^{(g)} + s_{\ominus}^{(g)} \mathbf{e}_{\ominus}^{(g)} \quad (30)$$

where $\|\mathbf{e}_1^{(g)}\| = \|\mathbf{e}_{\odot}^{(g)}\| = \|\mathbf{e}_{\ominus}^{(g)}\| = 1$ and $\mathbf{e}_1^{(g)T} \mathbf{e}_{\odot}^{(g)} = \mathbf{e}_1^{(g)T} \mathbf{e}_{\ominus}^{(g)} = \mathbf{e}_{\odot}^{(g)T} \mathbf{e}_{\ominus}^{(g)} = 0$. $s_1^{(g)}$, $s_{\odot}^{(g)}$, and $s_{\ominus}^{(g)}$ denote the corresponding projections in those directions. Consequently,

$$\mathbf{s}^{(g)T} \langle \tilde{\mathbf{z}}^{(g)} \rangle = s_1^{(g)} (\langle \tilde{\mathbf{z}}^{(g)} \rangle)_1 + s_{\odot}^{(g)} z_{\odot}^{(g)} + s_{\ominus}^{(g)} \mathbf{e}_{\ominus}^{(g)T} \langle \tilde{\mathbf{z}}^{(g)} \rangle \quad (31)$$

and subsequently

$$\mathbb{E}[\mathbf{s}^{(g)T} \langle \tilde{\mathbf{z}}^{(g)} \rangle] = \mathbb{E}[s_1^{(g)} (\langle \tilde{\mathbf{z}}^{(g)} \rangle)_1] + \mathbb{E}[s_{\odot}^{(g)} z_{\odot}^{(g)}] + \mathbb{E}[s_{\ominus}^{(g)} \mathbf{e}_{\ominus}^{(g)T} \langle \tilde{\mathbf{z}}^{(g)} \rangle] \quad (32)$$

follow. Taking into account the statistical independence between the cumulation path and a particular generation's mutations allows writing

$$\mathbb{E}[\mathbf{s}^{(g)T} \langle \tilde{\mathbf{z}}^{(g)} \rangle] = \mathbb{E}[s_1^{(g)}] \mathbb{E}[(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_1] + \mathbb{E}[s_{\odot}^{(g)}] \mathbb{E}[z_{\odot}^{(g)}] + \mathbb{E}[s_{\ominus}^{(g)}] \mathbb{E}[\mathbf{e}_{\ominus}^{(g)T} \langle \tilde{\mathbf{z}}^{(g)} \rangle] \quad (33)$$

$$\simeq \mathbb{E}[s_1^{(g)}] \mathbb{E}[(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_1] + \mathbb{E}[s_{\odot}^{(g)}] \mathbb{E}[z_{\odot}^{(g)}]. \quad (34)$$

Again, $\mathbb{E}[\mathbf{e}_{\ominus}^{(g)T} \langle \tilde{\mathbf{z}}^{(g)} \rangle]$ vanishes because those mutations are selectively neutral and isotropic. Taking expectation of Equation (29), considering Equations (14), (25), (34) and (A.25), and using $\mathbb{E}[\|\langle \tilde{\mathbf{z}}^{(g)} \rangle\|^2] \simeq \frac{N}{\mu}$,

$$\begin{aligned} \overline{\|\mathbf{s}^{(g+1)}\|^2} &\simeq (1-c)^2 \overline{\|\mathbf{s}^{(g)}\|^2} + 2(1-c) \sqrt{\mu c(2-c)} \\ &\quad \times \left(\overline{s_1^{(g)}} \mathbb{E}[(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_1] + \overline{s_{\odot}^{(g)}} \mathbb{E}[z_{\odot}^{(g)}] \right) + c(2-c)N \end{aligned} \quad (35)$$

$$\begin{aligned} &\simeq (1-c)^2 \overline{\|\mathbf{s}^{(g)}\|^2} + 2(1-c) \sqrt{\mu c(2-c)} \\ &\quad \times \left(\overline{s_1^{(g)}} \left(-\frac{N \varphi_x^{(g)}}{\sigma^{(g)*} r^{(g)}} \right) + \overline{s_{\odot}^{(g)}} \left(-\frac{N \varphi_{r^2}^{(g)}}{2\sigma^{(g)*} r^{(g)2}} - \frac{\sigma^{(g)*}}{2\mu} \right) \right) + c(2-c)N \end{aligned} \quad (36)$$

follows.

3.1.5 Derivation of a Mean Value Difference Equation for σ

From the update rule of σ in Line 20 of Algorithm 1, $\sigma^{(g+1)} = \sigma^{(g)} \exp\left(\frac{\|\mathbf{s}^{(g+1)}\|^2 - N}{2DN}\right)$ follows for the update of the mutation strength. Taking expected values and knowing that $\sigma^{(g)}$ is constant w.r.t. $\|\mathbf{s}^{(g+1)}\|^2$, this writes $\overline{\sigma^{(g+1)}} = \overline{\sigma^{(g)}} \mathbb{E}\left[\exp\left(\frac{\|\mathbf{s}^{(g+1)}\|^2 - N}{2DN}\right)\right]$. Assuming that the fluctuations of $\|\mathbf{s}^{(g+1)}\|^2$ around its expected value are sufficiently small, the expected value can be pulled into the exponential function yielding

$$\overline{\sigma^{(g+1)}} \simeq \overline{\sigma^{(g)}} \exp\left(\frac{\overline{\|\mathbf{s}^{(g+1)}\|^2} - N}{2DN}\right). \quad (37)$$

3.1.6 Summary of the Mean Value Difference Equations

$$\overline{x^{(g+1)}} = \overline{x^{(g)}} \left(1 - \frac{\varphi_x^{(g)*}}{N} \right) \quad (38)$$

$$\overline{r^{(g+1)}} = \overline{r^{(g)}} \left(1 - \frac{\varphi_r^{(g)*}}{N} \right) \quad (39)$$

$$\overline{s_1^{(g+1)}} \simeq (1-c)\overline{s_1^{(g)}} + \sqrt{\mu c(2-c)} \left(-\frac{N\varphi_x^{(g)}}{\sigma^{(g)*} r^{(g)}} \right) \quad (40)$$

$$\begin{aligned} \overline{s_\odot^{(g+1)}} \simeq (1-c) \left(1 + \frac{\overline{\sigma^{(g)*}}}{N} \left(-\frac{N\varphi_{r^2}^{(g)}}{2\sigma^{(g)*} r^{(g)2}} - \frac{\overline{\sigma^{(g)*}}}{2\mu} \right) \right) \overline{s_\odot^{(g)}} \\ + \sqrt{\mu c(2-c)} \left(-\frac{N\varphi_{r^2}^{(g)}}{2\sigma^{(g)*} r^{(g)2}} + \frac{\overline{\sigma^{(g)*}}}{2\mu} \right) \end{aligned} \quad (41)$$

$$\begin{aligned} \|\overline{\mathbf{s}^{(g+1)}}\|^2 \simeq (1-c)^2 \|\overline{\mathbf{s}^{(g)}}\|^2 + 2(1-c)\sqrt{\mu c(2-c)} \\ \times \left(\overline{s_1^{(g)}} \left(-\frac{N\varphi_x^{(g)}}{\sigma^{(g)*} r^{(g)}} \right) + \overline{s_\odot^{(g)}} \left(-\frac{N\varphi_{r^2}^{(g)}}{2\sigma^{(g)*} r^{(g)2}} - \frac{\overline{\sigma^{(g)*}}}{2\mu} \right) \right) + c(2-c)N \end{aligned} \quad (42)$$

$$\overline{\sigma^{(g+1)}} \simeq \overline{\sigma^{(g)}} \exp \left(\frac{\|\overline{\mathbf{s}^{(g+1)}}\|^2 - N}{2DN} \right) \quad (43)$$

$$\overline{\sigma^{(g+1)*}} = \frac{N\overline{\sigma^{(g+1)}}}{\overline{r^{(g+1)}}} \quad (44)$$

The mean value dynamics of the $(3/3_I, 10)$ -CSA-ES on the conically constrained problem are shown in Figure 3 for $N = 400$, $\xi = 10$, $c = \frac{1}{\sqrt{N}}$, and $D = \frac{1}{c}$. The agreement of the simulations and the derived expressions is not perfect due to the asymptotic assumptions used. But the prediction comes near to the real behavior. In particular, one observes that the lines of the iteration with one-generation experiments are very similar to the lines generated by real ES runs. Consequently, the modeling of the system with Equations (38) to (44) is appropriate and the deviations for the theoretically derived expressions are mainly due to approximations in the derivations of the local progress rates. For this, it is referred to the additional figures provided in the supplementary material (Appendix B). They show a larger deviation for smaller values of ξ and smaller values of N . But notice that in those figures the iteration with one-generation experiments for the local progress measures coincides well with the results of real ES runs. This again shows the appropriateness of the modeling in Equations (38) to (44). The deviations for small N stem from asymptotic assumptions using $N \rightarrow \infty$. They help simplifying expressions resulting in a theoretical analysis that is tractable. The deviations for small ξ are due to approximations in the derivation of the offspring cumulative distribution function after the projection step in x_1 -direction $P_Q(q)$ (for the details, it is referred to Section 3.1.2.1.2.3 in Spettel and Beyer (2018c), in particular to the step from Equation (3.73) to Equation (3.74)).

For the figures, results of 100 real runs of the ES have been averaged for generating the solid lines. The lines for the iteration by approximation have been computed by iterating the mean value iterative system (Equations (38) to (44)) with Equations (19), (20)

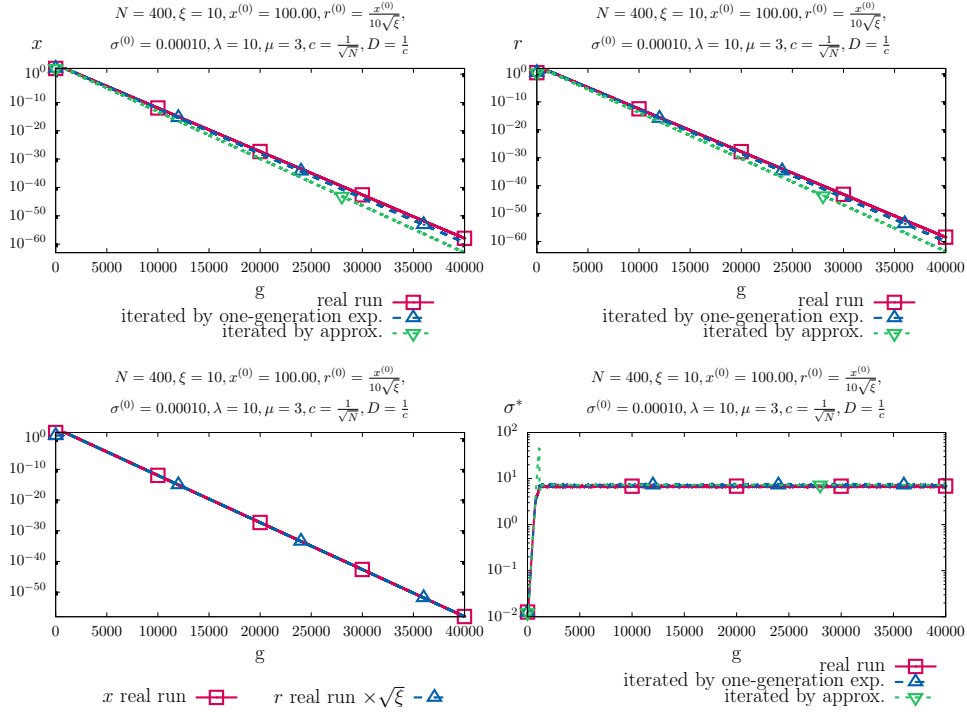


Figure 3: Real run and approximation comparison of the $(3/3_I, 10)$ -CSA-ES mean value dynamics ($N = 400, \xi = 10$). The agreement of the iteration with the theoretically derived expressions and the real ES runs is not perfect due to the asymptotic assumptions used. But the prediction comes near to the real behavior. In addition, the iteration with the one-generation experiments for the local progress rates is very similar to the mean value dynamics of the real ES runs. Consequently, the modeling of the system with Equations (38) to (44) is appropriate.

and (A.21) for $\varphi_x^{(g)}$ (and $\varphi_x^{(g)*}$), $\varphi_r^{(g)}$ (and $\varphi_r^{(g)*}$), and $\varphi_{r^2}^{(g)}$, respectively. The lines for the iteration with one-generation experiments have been generated by iterating the system (Equations (38) to (44)) and simulating $\varphi_x^{(g)}$ (and $\varphi_x^{(g)*}$), $\varphi_r^{(g)}$ (and $\varphi_r^{(g)*}$), and $\varphi_{r^2}^{(g)}$. It can happen that in a generation of iterating the system (Equations (38) to (44)), infeasible $(x^{(g)}, r^{(g)})^T$ are created. In such circumstances, the corresponding $(x^{(g)}, r^{(g)})^T$ have been projected back.

3.2 Behavior of the ES in the Steady State

The goal of this section is to analyze the steady state behavior of the mean value iterative system that is summarized in Section 3.1.6. A working ES should steadily decrease x and r (Equation (38) and Equation (39), respectively) in order to move toward the optimizer. For determining the steady state normalized mutation strength value, the fixed point of the system of non-linear equations (Equations (40) to (44)) is to be computed.

Section 3.2.1 comprises a first step toward closed-form approximations for the steady state values of the system summarized in Section 3.1.6. Expressions are derived that finally lead to a steady state equation for the normalized mutation strength.

A closed-form solution of that derived equation (Equation (62)) is not apparent. Hence, further assumptions for different cases are considered in Section 3.2.2 and Section 3.2.3.

Section 3.2.2 considers $c = 1/\sqrt{N}$ that has been proposed in early CMA-ES papers such as Hansen and Ostermeier (1997). Section 3.2.3 uses $c = \Theta(\frac{1}{N})$ proposed in newer publications (see (Hansen and Ostermeier, 2001; Hansen, 2016)). Again, to arrive at closed-form expressions, asymptotic considerations ($N \rightarrow \infty$) are taken into account.

3.2.1 Derivations Toward Closed-Form Steady State Expressions

To compute the fixed point of the system described by Equations (40) to (44), stationary state expressions φ_{xss}^* , φ_{rss}^* , $\left(-\frac{N\varphi_x^{(g)}}{\sigma^{(g)*}r^{(g)}}\right)_{ss}$, and $\left(-\frac{N\varphi_{r2}^{(g)}}{2\sigma^{(g)*}r^{(g)2}}\right)_{ss}$ for $\varphi_x^{(g)*}$, $\varphi_r^{(g)*}$, $-\frac{N\varphi_x^{(g)}}{\sigma^{(g)*}r^{(g)}}$, and $-\frac{N\varphi_{r2}^{(g)}}{2\sigma^{(g)*}r^{(g)2}}$, respectively, need to be derived first because they are dependent on the position in the parameter space. The bottom left subplot of Figure 3 shows that the ES moves in the vicinity of the cone boundary in the steady state. This can be seen because the dynamics of x and r are plotted by converting them into each other for the cone boundary case. Notice that those lines coincide in the steady state. In this situation, $P_{\text{feas}} \simeq 0$ for $N \rightarrow \infty$. This follows from Equation (21). By the cone boundary equation (Equation (2)), a parental individual $(x^{(g)}, r^{(g)})^T$ is on the cone boundary for $r^{(g)} = x^{(g)}/\sqrt{\xi}$. Using this together with Equation (14) and Equation (21) yields

$$P_{\text{feas}} \simeq \Phi \left[N \left(\frac{\sigma^{(g)*}}{r^{(g)}} (r^{(g)} - \bar{r}) \right) \right]. \quad (45)$$

By taking into account Equation (17),

$$P_{\text{feas}} \simeq \Phi \left[N \sigma^{(g)*} \left(1 - \sqrt{1 + \frac{\sigma^{(g)*2}}{N} \left(1 - \frac{1}{N} \right)} \right) \right] \quad (46)$$

follows. If $N\sigma^{(g)*}$ is sufficiently large, $P_{\text{feas}} \simeq 0$.

Steady state expressions for the distance ratio and the progress rates have already been derived in Spettel and Beyer (2018b). This work builds on those derived expressions. They are recapped here for making the present work self-containing. The steady state distance ratio reads

$$\left(\frac{x}{\sqrt{\xi}r} \right)_{ss} = \frac{1}{\sqrt{\frac{1 + \frac{\sigma_{ss}^{*2}}{\mu N}}{1 + \frac{\sigma_{ss}^{*2}}{N}}}} \quad (47)$$

and the steady state progress rates read

$$\varphi_{rss}^* = \varphi_{xss}^* \approx \frac{N}{1 + \xi} \left(-\frac{\sigma_{ss}^{*2}}{2\mu N} \right) + \frac{\sigma_{ss}^* c_{\mu/\mu, \lambda}}{\sqrt{1 + \xi}}. \quad (48)$$

A steady state expression for $-\frac{N\varphi_x^{(g)}}{\sigma^{(g)*}r^{(g)}}$ is derived next. With Equation (12) and Equation (48),

$$\left(-\frac{N\varphi_x}{\sigma^*r} \right)_{ss} = -\left(\frac{x}{r} \right)_{ss} \frac{\varphi_{xss}^*}{\sigma_{ss}^*} \quad (49)$$

can be derived. Use of Equation (47) for the fraction $\left(\frac{x}{r}\right)_{ss}$ results in

$$\left(-\frac{N\varphi_x}{\sigma^*r}\right)_{ss} = -\sqrt{\frac{\xi}{1+\frac{\sigma_{ss}^{*2}}{\mu N}}}\frac{\varphi_{xss}^*}{\sigma_{ss}^*}. \quad (50)$$

Similarly, a steady state expression for $-\frac{N\varphi_{r^2}^{(g)}}{2\sigma^{(g)*}r^{(g)^2}}$ can be derived. Considering the infeasible case (because in the steady state $P_{\text{feas}} \simeq 0$) of Equation (A.21), we have

$$\left(-\frac{N\varphi_{r^2}}{2\sigma^*r^2}\right)_{ss} = -\frac{N}{2\sigma_{ss}^*} \left[1 - \left(\frac{1}{\xi r^2}\right)_{ss} \text{E}[\langle q \rangle_{\text{infeas}}]^2 \left(\frac{1+\frac{\sigma_{ss}^{*2}}{\mu N}}{1+\frac{\sigma_{ss}^{*2}}{N}}\right)\right]. \quad (51)$$

According to Lines 4 and 21 of Algorithm 1, $\text{E}[\langle q \rangle] = \text{E}[x^{(g+1)}] = \overline{x^{(g+1)}}$. Hence, Equation (51) can be rewritten using Equation (15) for the infeasible $\langle q \rangle$ case $\text{E}[\langle q \rangle_{\text{infeas}}]$ and Equation (47) for $\left(\frac{x^2}{\xi r^2}\right)_{ss}$, resulting in

$$\left(-\frac{N\varphi_{r^2}}{2\sigma^*r^2}\right)_{ss} = -\frac{N}{2\sigma_{ss}^*} \left[1 - \left(1 - \frac{\varphi_{xss}^*}{N}\right)^2\right]. \quad (52)$$

Using Equation (50) and Equation (52), steady state expressions for Equations (40) to (44) can be derived. Requiring $s_1^{(g+1)} = s_1^{(g)} = s_{1ss}$ in Equation (40) using Equation (50) yields

$$s_{1ss} = -\frac{\sqrt{\mu c(2-c)}}{c} \left(\sqrt{\frac{\xi}{1+\frac{\sigma_{ss}^{*2}}{\mu N}}}\frac{\varphi_{xss}^*}{\sigma_{ss}^*} \right). \quad (53)$$

Analogously, requiring $s_{\odot}^{(g+1)} = s_{\odot}^{(g)} = s_{\odot ss}$ in Equation (41) using Equation (52) results in

$$s_{\odot ss} = \frac{\sqrt{\mu c(2-c)} \left(-\frac{N}{2\sigma_{ss}^*} \left[1 - \left(1 - \frac{\varphi_{xss}^*}{N}\right)^2\right] + \frac{\sigma_{ss}^*}{2\mu} \right)}{c - (1-c)\frac{\sigma_{ss}^*}{N} \left[-\frac{N}{2\sigma_{ss}^*} \left[1 - \left(1 - \frac{\varphi_{xss}^*}{N}\right)^2\right] - \frac{\sigma_{ss}^*}{2\mu} \right]}. \quad (54)$$

In the same way, setting $\|\mathbf{s}^{(g+1)}\|^2 = \|\mathbf{s}^{(g)}\|^2 = \|\mathbf{s}\|_{ss}^2$ in Equation (42) using Equation (50) and Equation (52) gives

$$\begin{aligned} \|\mathbf{s}\|_{ss}^2 &= N - \frac{2(1-c)\sqrt{\mu c(2-c)}}{-2c+c^2} \\ &\times \left[s_{1ss} \left(-\sqrt{\frac{\xi}{1+\frac{\sigma_{ss}^{*2}}{\mu N}}}\frac{\varphi_{xss}^*}{\sigma_{ss}^*} \right) + s_{\odot ss} \left(-\frac{N}{2\sigma_{ss}^*} \left[1 - \left(1 - \frac{\varphi_{xss}^*}{N}\right)^2\right] - \frac{\sigma_{ss}^*}{2\mu} \right) \right]. \end{aligned} \quad (55)$$

For the mutation strength,

$$\frac{\sigma^{(g+1)*}r^{(g+1)}}{N} = \frac{\sigma^{(g)*}r^{(g)}}{N} \exp\left(\frac{\|\mathbf{s}^{(g+1)}\|^2 - N}{2DN}\right) \quad (56)$$

follows from Line 20 of Algorithm 1 with the use of Equation (14). Rewriting Equation (56) and using Equation (A.8) together with Equation (A.9) for the fraction $r^{(g)}/r^{(g+1)}$, we have

$$\sigma^{(g+1)*} \simeq \sigma^{(g)*} \frac{1}{\sqrt{1 + \frac{2\sigma^{(g)*}}{N} z_{\odot}^{(g)} + \frac{\sigma^{(g)*2}}{\mu N}}} \exp\left(\frac{\|\mathbf{s}^{(g+1)}\|^2 - N}{2DN}\right) \quad (57)$$

$$\sigma^{(g+1)*2} \left(1 + \frac{2\sigma^{(g)*}}{N} z_{\odot}^{(g)} + \frac{\sigma^{(g)*2}}{\mu N}\right) \simeq \sigma^{(g)*2} \exp\left(\frac{\|\mathbf{s}^{(g+1)}\|^2 - N}{DN}\right). \quad (58)$$

Use of the Taylor expansion $\exp(x) \simeq 1 + x$ (around zero and neglecting terms of quadratic and higher order) results in

$$\sigma^{(g+1)*2} \left(1 + \frac{2\sigma^{(g)*}}{N} z_{\odot}^{(g)} + \frac{\sigma^{(g)*2}}{\mu N}\right) \simeq \sigma^{(g)*2} \left(1 + \frac{\|\mathbf{s}^{(g+1)}\|^2 - N}{DN}\right). \quad (59)$$

Computing the expectation of Equation (59) and requiring $\overline{\sigma^{(g+1)*}} = \overline{\sigma^{(g)*}} = \sigma_{ss}^*$, we get

$$\sigma_{ss}^{*2} \left[\frac{1}{2} + \frac{1}{N} \left(\sigma_{ss}^* \mathbb{E}[z_{\odot}] + \frac{\sigma_{ss}^{*2}}{2\mu} \right) \right] = \sigma_{ss}^{*2} \left(\frac{1}{2} + \frac{\|\mathbf{s}\|_{ss}^2 - N}{2DN} \right) \quad (60)$$

$$\sigma_{ss}^* \mathbb{E}[z_{\odot}] + \frac{\sigma_{ss}^{*2}}{2\mu} = \frac{\|\mathbf{s}\|_{ss}^2 - N}{2D}. \quad (61)$$

Usage of Equation (A.25) together with the steady state expression derived in Equation (52) for $\mathbb{E}[z_{\odot}]$ results in

$$\sigma_{ss}^* \left(-\frac{N}{2\sigma_{ss}^*} \left[1 - \left(1 - \frac{\varphi_{xss}^*}{N} \right)^2 \right] - \frac{\sigma_{ss}^*}{2\mu} \right) + \frac{\sigma_{ss}^{*2}}{2\mu} = \frac{\|\mathbf{s}\|_{ss}^2 - N}{2D}. \quad (62)$$

Consideration of Equations (48) and (53) to (55) allows numerically solving Equation (62) for σ_{ss}^* . By doing this, the influence of the cumulation parameter can be further investigated. To this end, Figure 4 shows the influence of the cumulation parameter for the considered (3/3_I, 10)-CSA-ES for the dimensions 40 and 400. Results for dimensions 1000 and 10000 are provided in the supplementary material (Appendix C, Figure 23). The steady state normalized x progress rate (Equation (48)) is plotted against the steady state normalized mutation strength for different values of ξ . Steady state values achieved for different settings of c with $D = \frac{1}{c}$ have been computed numerically (using Equation (62) with Equations (48) and (53) to (55)) and are shown with the filled black markers. A particular emphasis in the comparison is on the choices $c = \Theta\left(\frac{1}{\sqrt{N}}\right)$ and $c = \Theta\left(\frac{1}{N}\right)$ that have been proposed in CMA-ES papers (see Hansen and Ostermeier (1997) and Hansen and Ostermeier (2001); Hansen (2016), respectively).

Interestingly, the choice of $c = \Theta\left(\frac{1}{N}\right)$ results in steady state normalized x progress values that are independent of ξ . They stay almost constant with increasing ξ for the dimensions considered whereas for $c = \Theta\left(\frac{1}{\sqrt{N}}\right)$ the steady state normalized x progress values decrease with increasing ξ for larger dimensions. For the considered dimensions, only for $N = 40$ the choice $c = \frac{1}{\sqrt{N}}$ achieves constant progress with increasing ξ .

In addition, the choice of $c = \Theta\left(\frac{1}{N}\right)$ results in better steady state normalized x progress compared to $c = \Theta\left(\frac{1}{\sqrt{N}}\right)$ the higher N and ξ get. For example, for the configurations considered in the plots, the choice $c = \frac{\mu+2}{N+\mu+5}$ results in the highest steady state x progress for $N = 400$ and $\xi \geq 25$. For $N = 40$, the choices $c = \frac{1}{\sqrt{N}}$ and $c = \frac{2}{\sqrt{N}}$ for $\xi \geq 25$ and $\xi < 25$, respectively, result in the highest steady state x progress.

In order to analyze the influence of the parameters further theoretically, the aim of the following two sections is to derive closed-form expressions using asymptotic assumptions.

3.2.2 Derivation of Closed-Form Approximations for the Steady State with the Assumptions $c = \Theta\left(\frac{1}{\sqrt{N}}\right)$ and $N \rightarrow \infty$

The goal of this section is to simplify the expressions derived in Section 3.2.1 further using additional asymptotic assumptions in order to arrive at closed-form steady state approximations.

The expression derived for $\left(-\frac{N\varphi_{r^2}}{2\sigma_{ss}^*}\right)_{ss}$ as Equation (52) is simplified further yielding

$$\left(-\frac{N\varphi_{r^2}}{2\sigma_{ss}^*}\right)_{ss} = -\frac{N}{2\sigma_{ss}^*} \left[1 - \left(1 - \frac{\varphi_{xss}^*}{N}\right)^2\right] = -\frac{N}{2\sigma_{ss}^*} \left(2\frac{\varphi_{xss}^*}{N} - \frac{\varphi_{xss}^{*2}}{N^2}\right) \quad (63)$$

$$= -\frac{\varphi_{xss}^*}{\sigma_{ss}^*} + \frac{\varphi_{xss}^{*2}}{2\sigma_{ss}^*N} \simeq -\frac{\varphi_{xss}^*}{\sigma_{ss}^*}. \quad (64)$$

In Equation (64), $\sigma_{ss}^*N \gg \varphi_{xss}^{*2}$ has been assumed and therefore the second summand has been neglected.

Insertion of Equation (64) into Equation (62) replacing $-\frac{N}{2\sigma_{ss}^*} \left[1 - \left(1 - \frac{\varphi_{xss}^*}{N}\right)^2\right]$ yields (after simplification)

$$-\varphi_{xss}^* = \frac{\|\mathbf{s}\|_{ss}^2 - N}{2D} \quad (65)$$

for the steady state mutation strength equation. Equation (64) can also be inserted into Equation (54) replacing $-\frac{N}{2\sigma_{ss}^*} \left[1 - \left(1 - \frac{\varphi_{xss}^*}{N}\right)^2\right]$. This results in

$$s_{\odot ss} \simeq \frac{\sqrt{\mu c(2-c)} \left[-\frac{\varphi_{xss}^*}{\sigma_{ss}^*} + \frac{\sigma_{ss}^*}{2\mu}\right]}{c - (1-c)\frac{\sigma_{ss}^*}{N} \left[-\frac{\varphi_{xss}^*}{\sigma_{ss}^*} - \frac{\sigma_{ss}^*}{2\mu}\right]} = \frac{\sqrt{\mu c(2-c)} \left(-\frac{\varphi_{xss}^*}{\sigma_{ss}^*} + \frac{\sigma_{ss}^*}{2\mu}\right)}{c - (1-c) \left(-\frac{\varphi_{xss}^*}{N} - \frac{\sigma_{ss}^{*2}}{2\mu N}\right)}. \quad (66)$$

$$= \frac{\sqrt{\mu c(2-c)} \left(-\frac{\varphi_{xss}^*}{\sigma_{ss}^*} + \frac{\sigma_{ss}^*}{2\mu}\right)}{c + \frac{\varphi_{xss}^*}{N} + \frac{\sigma_{ss}^{*2}}{2\mu N} - \frac{c\varphi_{xss}^*}{N} - \frac{c\sigma_{ss}^{*2}}{2\mu N}}. \quad (67)$$

With the assumptions $N \rightarrow \infty$ and $c = \Theta\left(\frac{1}{\sqrt{N}}\right)$, the expression $\frac{\varphi_{xss}^*}{N} + \frac{\sigma_{ss}^{*2}}{2\mu N} - \frac{c\varphi_{xss}^*}{N} - \frac{c\sigma_{ss}^{*2}}{2\mu N}$ is an order of magnitude smaller than c and can therefore be neglected w.r.t. c . Hence, Equation (67) simplifies to

$$s_{\odot ss} \simeq \frac{\sqrt{\mu c(2-c)}}{c} \left(-\frac{\varphi_{xss}^*}{\sigma_{ss}^*} + \frac{\sigma_{ss}^*}{2\mu}\right). \quad (68)$$

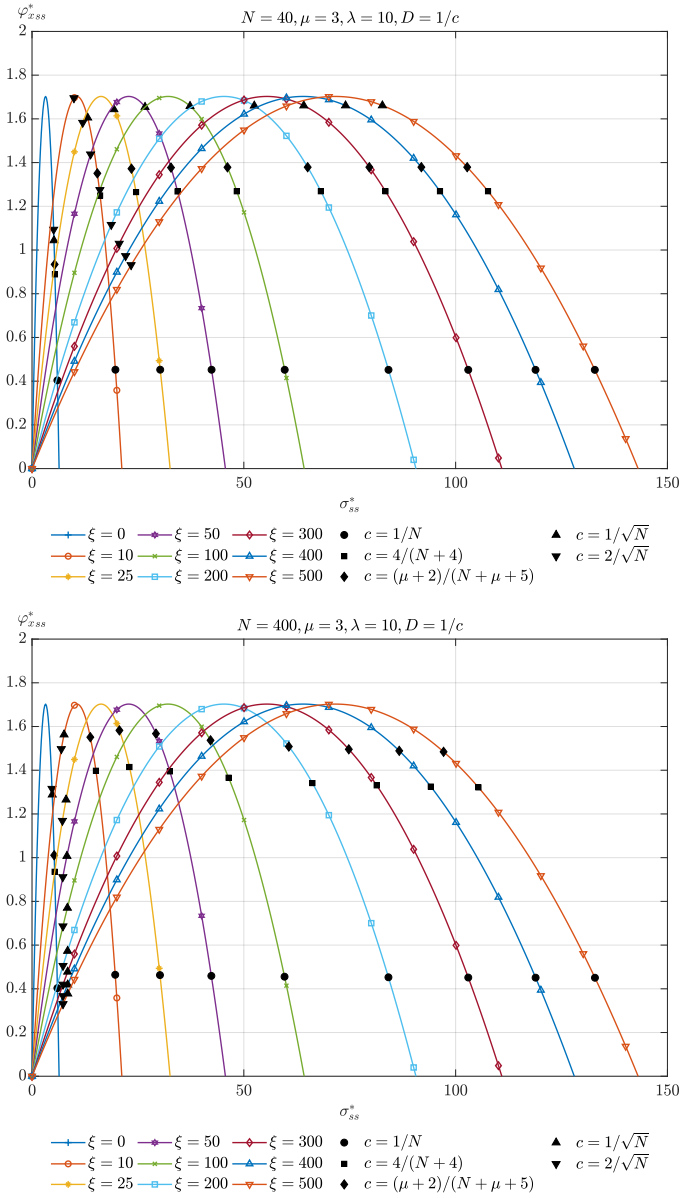


Figure 4: Influence of the cumulation parameter for the considered $(3/3_I, 10)$ -CSA-ES for $N = 40$ (top) and $N = 400$ (bottom). For different values of ξ , the steady state progress rate is plotted against the steady state normalized mutation strength. The filled black markers indicate the steady state values achieved by different c parameter settings with $D = 1/c$.

Similarly, Equation (64) inserted into Equation (55) replacing $-\frac{N}{2\sigma_{ss}^*} \left[1 - \left(1 - \frac{\varphi_{xss}^*}{N} \right)^2 \right]$ results in

$$\|s\|_{ss}^2 = N - \frac{2(1-c)\sqrt{\mu c(2-c)}}{-2c+c^2} \left[s_{1ss} \left(-\sqrt{\frac{\xi}{1+\frac{\sigma_{ss}^{*2}}{\mu N}}} \frac{\varphi_{xss}^*}{\sigma_{ss}^*} \right) + s_{\odot ss} \left(-\frac{\varphi_{xss}^*}{\sigma_{ss}^*} - \frac{\sigma_{ss}^{*2}}{2\mu} \right) \right]. \quad (69)$$

Insertion of Equation (68) and Equation (53) into Equation (69) yields (after straightforward simplification)

$$\|s\|_{ss}^2 \simeq N + \frac{2(1-c)\mu}{c} \left[\left(\xi \left(\frac{1+\frac{\sigma_{ss}^{*2}}{N}}{1+\frac{\sigma_{ss}^{*2}}{\mu N}} \right) + 1 \right) \frac{\varphi_{xss}^{*2}}{\sigma_{ss}^{*2}} - \frac{\sigma_{ss}^{*2}}{4\mu^2} \right]. \quad (70)$$

$\xi \left(\frac{1+\frac{\sigma_{ss}^{*2}}{N}}{1+\frac{\sigma_{ss}^{*2}}{\mu N}} \right) \simeq \xi$ for $N \rightarrow \infty$ allows writing

$$\|s\|_{ss}^2 \simeq N + \frac{2(1-c)\mu}{c} \left[(\xi + 1) \frac{\varphi_{xss}^{*2}}{\sigma_{ss}^{*2}} - \frac{\sigma_{ss}^{*2}}{4\mu^2} \right] \quad (71)$$

$$= N + \frac{2(1-c)\mu}{c} \left[\left(c_{\mu/\mu,\lambda}^2 - \frac{\sigma_{ss}^* c_{\mu/\mu,\lambda}}{\mu\sqrt{1+\xi}} + \frac{\sigma_{ss}^{*2}}{(1+\xi)4\mu^2} \right) - \frac{\sigma_{ss}^{*2}}{4\mu^2} \right]. \quad (72)$$

From Equation (71) to Equation (72), φ_{xss}^* has been substituted by Equation (48), its square has been calculated, and the resulting expression has been simplified.

With this, insertion of Equation (48) and Equation (72) into Equation (65) yields the quadratic equation

$$\frac{\sigma_{ss}^{*2}}{(1+\xi)2\mu} - \frac{\sigma_{ss}^* c_{\mu/\mu,\lambda}}{\sqrt{1+\xi}} = \frac{2(1-c)\mu}{2cD} \left(c_{\mu/\mu,\lambda}^2 - \frac{\sigma_{ss}^* c_{\mu/\mu,\lambda}}{\mu\sqrt{1+\xi}} + \frac{\sigma_{ss}^{*2}}{(1+\xi)4\mu^2} - \frac{\sigma_{ss}^{*2}}{4\mu^2} \right) \quad (73)$$

for the steady state normalized mutation strength equation. By solving Equation (73) for the positive root (because $\sigma_{ss}^* > 0$) with subsequent simplification of the result we get

$$\sigma_{ss}^* = \frac{2\mu\sqrt{\xi+1}c_{\mu/\mu,\lambda} \left((cD+c-1) + \sqrt{c^2(D^2+\xi+1) - 2c(\xi+1) + \xi+1} \right)}{2cD - c\xi + \xi} \quad (74)$$

as an asymptotic ($N \rightarrow \infty$) closed-form expression for the steady state normalized mutation strength. Insertion of $c = 1/\sqrt{N}$ and $D = 1/c = \sqrt{N}$ into Equation (74) results in the expression

$$\sigma_{ss}^* = \frac{2\mu\sqrt{\xi+1}c_{\mu/\mu,\lambda} \left(1 + \frac{1}{\sqrt{N}} - 1 + \sqrt{1 + \frac{\xi}{N} + \frac{1}{N} - \frac{2\xi}{\sqrt{N}} - \frac{2}{\sqrt{N}} + \xi + 1} \right)}{2 - \frac{\xi}{\sqrt{N}} + \xi}. \quad (75)$$

Assuming $N \rightarrow \infty$ and $\frac{\xi}{\sqrt{N}} \rightarrow 0$ allows a further asymptotic simplification of Equation (75) (neglecting $\frac{1}{\sqrt{N}}$, $\frac{\xi}{N}$, $\frac{1}{N}$, $\frac{2\xi}{\sqrt{N}}$, and $\frac{\xi}{\sqrt{N}}$) resulting in

$$\sigma_{ss}^* \simeq \frac{2\mu\sqrt{\xi+1}\sqrt{\xi+2}c_{\mu/\mu,\lambda}}{\xi+2} = \frac{2\mu\sqrt{\xi+1}c_{\mu/\mu,\lambda}}{\sqrt{\xi+2}}. \quad (76)$$

For sufficiently large ξ , $\sqrt{\xi+1} \simeq \sqrt{\xi+2}$, and Equation (76) writes $\sigma_{ss}^* \simeq 2\mu c_{\mu/\mu, \lambda}$. Back-insertion of Equation (74) (or Equation (76)) into Equations (47), (48) and (53) to (55) allows calculating the steady state distance from the cone boundary, the normalized steady state progress, s_{1ss} , $s_{\odot ss}$, and $\|s\|_{ss}^2$.

Figure 5 shows plots of the steady state computations. Results computed by Equation (74) have been compared to real $(\mu/\mu_I, \lambda)$ -CSA-ES runs for $\lambda = 10$ and $\mu \in \{1, 2, 3\}$. Further experimental results (including results for $\mu \in \{4, 6, 8\}$) are provided in the supplementary material (Figures 15 to 18 in Appendix B). The values for the points denoting the approximations have been determined by computing the normalized steady state mutation strength σ_{ss}^* using Equation (74) for different values of ξ . The results for φ_x^* and φ_r^* have been determined by using the computed steady state σ_{ss}^* values with Equation (48). The approximations for $\left(\frac{x}{\sqrt{\xi}r}\right)_{ss}$ have been determined by evaluating Equation (47). The values for the points denoting the experiments have been determined by computing the averages of the particular values in real ES runs. The figures show that the derived expressions get better for larger values of ξ and N . Again, the deviations for small ξ are due to approximations in the derivation of the local progress rates. The deviations for small N stem from the use of asymptotic assumptions $N \rightarrow \infty$.

3.2.3 Derivation of Closed-Form Approximations for the Steady State with the Assumptions $c = \Theta\left(\frac{1}{N}\right)$ and $N \rightarrow \infty$

In Section 3.2.2 it has been assumed that $c = \Theta\left(\frac{1}{\sqrt{N}}\right)$ from Equation (67) to Equation (68). This section presents a derivation for the case $c = \Theta\left(\frac{1}{N}\right)$. To this end, Equation (66) is rewritten to

$$s_{\odot ss} \simeq \frac{\sqrt{\mu c(2-c)}}{c} \frac{1}{\sigma_{ss}^*} \frac{\left(-\varphi_{xss}^* + \frac{\sigma_{ss}^{*2}}{2\mu}\right)}{\left(1 + \frac{1}{N}\left(\frac{1}{c} - 1\right)\left(\varphi_{xss}^* + \frac{\sigma_{ss}^{*2}}{2\mu}\right)\right)}. \quad (77)$$

With $c = \Theta\left(\frac{1}{N}\right)$, we have $\frac{1}{c} \gg 1$. This together with the assumption $|\varphi_{xss}^*| \ll \frac{\sigma_{ss}^{*2}}{2\mu}$ allows rewriting Equation (77) to

$$s_{\odot ss} \simeq \frac{\sqrt{\mu c(2-c)}}{c} \frac{\sigma_{ss}^*}{2\mu + \frac{\sigma_{ss}^{*2}}{cN}}. \quad (78)$$

With the additional assumption $2\mu \ll \frac{\sigma_{ss}^{*2}}{cN}$, Equation (78) simplifies to

$$s_{\odot ss} \simeq \frac{\sqrt{\mu c(2-c)}}{c} \frac{cN}{\sigma_{ss}^*}. \quad (79)$$

Insertion of Equation (79) and Equation (53) into Equation (69) yields

$$\|s\|_{ss}^2 \simeq N - \frac{2(1-c)\mu c(2-c)}{(-2c+c^2)c} \left[\xi \left(\frac{1 + \frac{\sigma_{ss}^{*2}}{N}}{1 + \frac{\sigma_{ss}^{*2}}{\mu N}} \right) \frac{\varphi_{xss}^{*2}}{\sigma_{ss}^{*2}} + \frac{cN}{\sigma_{ss}^*} \left(-\frac{\varphi_{xss}^*}{\sigma_{ss}^*} - \frac{\sigma_{ss}^*}{2\mu} \right) \right] \quad (80)$$

$$= N + \frac{2(1-c)\mu}{c} \left(\xi \frac{\varphi_{xss}^{*2}}{\sigma_{ss}^{*2}} - \frac{cN\varphi_{xss}^*}{\sigma_{ss}^{*2}} - \frac{cN}{2\mu} \right). \quad (81)$$

In the step from Equation (80) to Equation (81), $\xi \left(\frac{1 + \frac{\sigma_{ss}^{*2}}{N}}{1 + \frac{\sigma_{ss}^{*2}}{\mu N}} \right) \simeq \xi$ for $N \rightarrow \infty$ has been used.

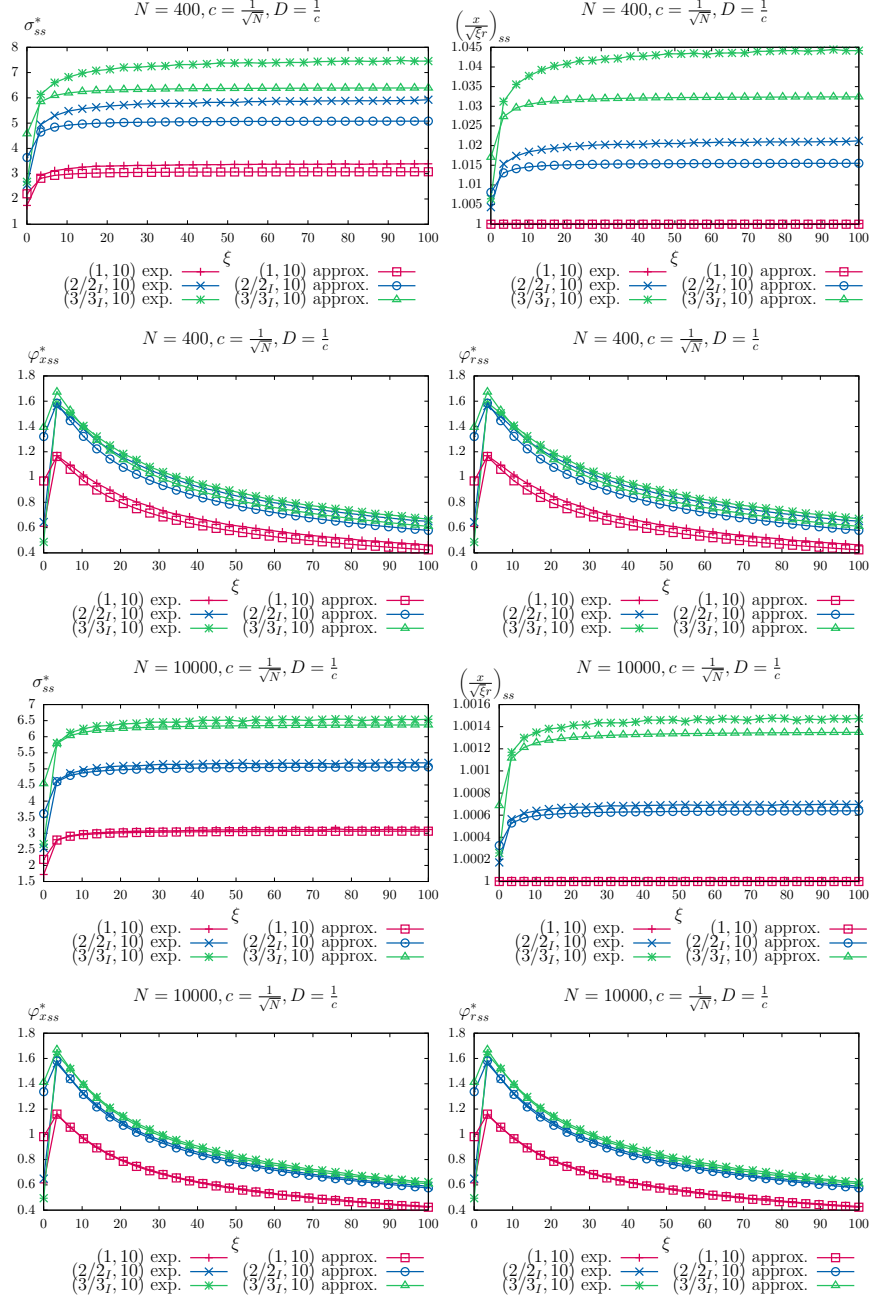


Figure 5: Steady state closed-form approximation and real run comparison of the $(\mu/\mu_I, \lambda)$ -CSA-ES with repair by projection applied to the conically constrained problem for $c = \frac{1}{\sqrt{N}}$, $D = \frac{1}{c}$, $\lambda = 10$, $\mu \in \{1, 2, 3\}$.

Insertion of Equation (48) and Equation (81) into Equation (65) yields

$$\frac{\sigma_{ss}^{*2}}{(1+\xi)2\mu} - \frac{\sigma_{ss}^* c_{\mu/\mu,\lambda}}{\sqrt{1+\xi}} = \frac{2(1-c)\mu}{2cD} \left[\xi \left(\frac{c_{\mu/\mu,\lambda}^2}{1+\xi} - \frac{\sigma_{ss}^* c_{\mu/\mu,\lambda}}{\mu\sqrt{1+\xi}(1+\xi)} + \frac{\sigma_{ss}^{*2}}{(1+\xi)^2 4\mu^2} \right) - \frac{cN c_{\mu/\mu,\lambda}}{\sqrt{1+\xi} \sigma_{ss}^*} + \frac{cN}{(1+\xi)2\mu} - \frac{cN}{2\mu} \right] \quad (82)$$

for the steady state mutation strength equation. By assuming $c \ll 1$, Equation (82) simplifies. Together with grouping the powers of σ_{ss}^* it writes

$$\begin{aligned} \frac{\sigma_{ss}^{*2}}{(1+\xi)2\mu} - \frac{\mu\xi\sigma_{ss}^{*2}}{cD(1+\xi)^2 4\mu^2} - \frac{\sigma_{ss}^* c_{\mu/\mu,\lambda}}{\sqrt{1+\xi}} + \frac{\mu\xi\sigma_{ss}^* c_{\mu/\mu,\lambda}}{cD\mu\sqrt{1+\xi}(1+\xi)} - \frac{\mu\xi c_{\mu/\mu,\lambda}^2}{cD(1+\xi)} \\ - \frac{cN}{2cD(1+\xi)} + \frac{cN}{2cD} + \frac{\mu cN c_{\mu/\mu,\lambda}}{cD\sqrt{1+\xi}\sigma_{ss}^*} = 0. \end{aligned} \quad (83)$$

Introducing common denominators allows rewriting Equation (83) to

$$\begin{aligned} \sigma_{ss}^{*2} \left(\frac{2cD(1+\xi) - \xi}{cD(1+\xi)^2 4\mu} \right) + \sigma_{ss}^{*1} \left(\frac{\xi c_{\mu/\mu,\lambda} - c_{\mu/\mu,\lambda} cD(1+\xi)}{cD\sqrt{1+\xi}(1+\xi)} \right) \\ + \sigma_{ss}^{*0} \left(\frac{-2\mu\xi c_{\mu/\mu,\lambda}^2 - cN + cN(1+\xi)}{2cD(1+\xi)} \right) + \sigma_{ss}^{*-1} \left(\frac{\mu cN c_{\mu/\mu,\lambda}}{cD\sqrt{1+\xi}} \right) = 0. \end{aligned} \quad (84)$$

Simplification of Equation (84) using $cD = 1$ and $cN \simeq 1$ results in

$$\sigma_{ss}^{*2} \left(\frac{2+\xi}{(1+\xi)^2 4\mu} \right) + \sigma_{ss}^* \left(\frac{-c_{\mu/\mu,\lambda}}{\sqrt{1+\xi}(1+\xi)} \right) + \left(\frac{\xi - 2\mu\xi c_{\mu/\mu,\lambda}^2}{2(1+\xi)} \right) + \frac{1}{\sigma_{ss}^*} \left(\frac{\mu c_{\mu/\mu,\lambda}}{\sqrt{1+\xi}} \right) = 0. \quad (85)$$

Note that multiplying Equation (85) by σ_{ss}^* results in a cubic equation that can be solved. However, the expressions for the closed-form solutions are rather long. Hence, a quadratic equation is aimed for. To this end, Equation (85) is approximated quadratically. Neglecting $\sigma_{ss}^{*-1} \left(\frac{\mu c_{\mu/\mu,\lambda}}{\sqrt{1+\xi}} \right)$ and $\sigma_{ss}^* \left(\frac{-c_{\mu/\mu,\lambda}}{\sqrt{1+\xi}(1+\xi)} \right)$ in Equation (85) results in

$$\sigma_{ss}^{*2} \left(\frac{2+\xi}{(1+\xi)^2 4\mu} \right) + \left(\frac{\xi - 2\mu\xi c_{\mu/\mu,\lambda}^2}{2(1+\xi)} \right) = 0. \quad (86)$$

Solving Equation (86) for the positive root with subsequent simplification yields

$$\sigma_{ss}^* \approx \sqrt{\frac{2\mu\xi(2\mu c_{\mu/\mu,\lambda}^2 - 1)(1+\xi)}{2+\xi}}. \quad (87)$$

Figure 6 shows plots of the steady state computations comparing the $(\mu/\mu_I, \lambda)$ -CSA-ES for $\lambda = 10$ and $\mu \in \{1, 2, 3\}$. Further experimental results (including results for $\mu \in \{4, 6, 8\}$) are provided in the supplementary material (Figures 19 to 22 in Appendix B). Results computed by Equation (87) have been compared to real ES runs. The values for the points denoting the approximations have been determined by computing the normalized steady state mutation strength σ_{ss}^* using Equation (87) for different values of ξ . The results for φ_x^* and φ_r^* have been determined by using the computed

steady state σ_{ss}^* values with Equation (48). The approximations for $\left(\frac{x}{\sqrt{\xi}r}\right)_{ss}$ have been determined by evaluating Equation (47). The values for the points denoting the experiments have been determined by computing the averages of the particular values in real ES runs. The rather large deviations for $\varphi_{x_{ss}}^*$ and $\varphi_{r_{ss}}^*$ for values of ξ ranging from around 3 to 30 stem from the σ_{ss}^* approximation (Equation (87)). A further investigation resulted in the plots shown in Figure 13 and Figure 14 in the supplementary material. They compare the numerical solution of Equation (62) with real ES runs and show a better agreement.

4 Discussions

The comparison of the mean value iterative system summarized in Section 3.1.6 with real ES runs shows that the theoretical prediction is not completely perfect because of the asymptotic assumptions used in the derivations (see Figure 3). For large ξ and large N the prediction comes near to the real behavior. The deviations for small N are due to the asymptotic assumptions $N \rightarrow \infty$ that are used in the derivations of the microscopic and macroscopic aspects of the ES. They are used to simplify the expressions and thus make a theoretical analysis tractable. The deviations for small ξ stem from the derivation of the offspring cumulative distribution function after the projection step in x_1 -direction $P_Q(q)$ (for the details, it is referred to Section 3.1.2.1.2.3 in Spettel and Beyer (2018c), in particular to the step from Equation (3.73) to Equation (3.74)). The same observations regarding the deviations can be made for the derived steady state expressions (see Figures 5 and 6).

For the steady state derivations, it is of particular interest to compare the results obtained in this work for the CSA-ES with the results obtained for the σ SA-ES. The $(\mu/\mu_I, \lambda)$ - σ SA-ES has been theoretically analyzed by Spettel and Beyer (2018b) applied to the same conically constrained problem. In that work, the microscopic and macroscopic aspects of the $(\mu/\mu_I, \lambda)$ - σ SA-ES have been investigated. For the microscopic aspects, expressions for the local progress for x and r and the self-adaptation response (SAR) function have been derived using asymptotic assumptions. Those results have then been used for the macroscopic analysis. The mean value dynamics generated by iteration using those local measures have been compared to real runs. In addition, steady state expressions have been derived and discussed. They show that the σ SA-ES is able to achieve sufficiently high mutation strengths to keep the progress almost constant for increasing ξ . Surprisingly, for the CSA-ES, the choice of the cumulation parameter c has a qualitative influence on the behavior. As detailed in the following paragraphs, of the two cases $c = 1/\sqrt{N}$ and $c = \Theta\left(\frac{1}{N}\right)$ considered, only the choice of $c = \Theta\left(\frac{1}{N}\right)$ results in a qualitative behavior of the CSA-ES that is similar to the σ SA-ES.

Considering the choice of $c = 1/\sqrt{N}$ proposed in early publications on CMA-ES (Hansen and Ostermeier, 1997), the steady state mutation strengths attained flatten with increasing ξ . As a consequence, the steady state progress decreases with higher values of ξ . This can be seen by considering Equation (76) that leads to $\sigma_{ss}^* \simeq 2\mu c_{\mu/\mu, \lambda}$ for sufficiently large ξ . For $\mu = 3$ and $\lambda = 10$, this results in a steady state normalized mutation strength of approximately 6.39. Note that this value corresponds to the approximations for the larger values of ξ shown in Figure 5 (subplot showing σ_{ss}^* for $N = 10000$). Equation (76) can be inserted into the steady state progress rate (Equation (48)) yielding

$$\varphi_{r_{ss}}^* = \varphi_{x_{ss}}^* \approx \frac{2\mu\sqrt{\xi+1}c_{\mu/\mu, \lambda}^2}{\sqrt{\xi+1}\sqrt{\xi+2}} - \frac{4\mu^2(\xi+1)c_{\mu/\mu, \lambda}^2}{(\xi+2)(\xi+1)2\mu} = \frac{2\mu c_{\mu/\mu, \lambda}^2}{\sqrt{\xi+2}} - \frac{2\mu c_{\mu/\mu, \lambda}^2}{\xi+2}. \quad (88)$$

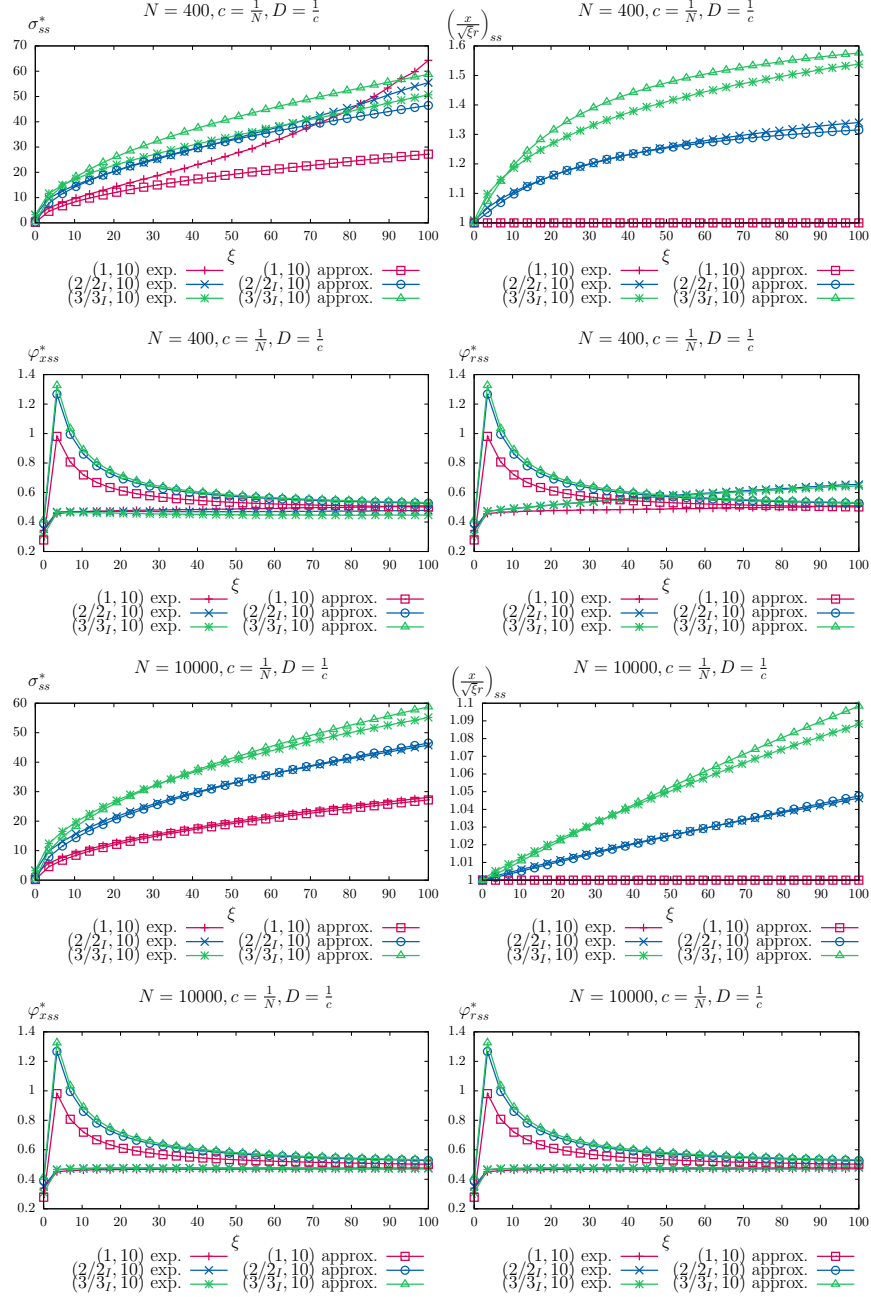


Figure 6: Steady state closed-form approximation and real run comparison of the $(\mu/\mu_I, \lambda)$ -CSA-ES with repair by projection applied to the conically constrained problem for $c = \frac{1}{N}$, $D = \frac{1}{c}$, $\lambda = 10$, $\mu \in \{1, 2, 3\}$.

From the simplified result of Equation (88) one immediately notices that $\varphi_{ss}^* \rightarrow 0$ for $\xi \rightarrow \infty$ (respecting $\frac{\xi}{\sqrt{N}} \rightarrow 0$ that was used in the derivations leading to Equation (76)). This is exactly what one sees in Figure 5, the stationary state progress decreases with increasing ξ .

In contrast, for the case $c = \Theta(\frac{1}{N})$ that is proposed in newer publications ($\frac{4}{N+4}$ by Hansen and Ostermeier (2001) or $\frac{\mu+2}{N+\mu+5}$ by Hansen (2016), both of which are in $\Theta(\frac{1}{N})$ for μ independent of N), the steady state mutation strength increases with increasing ξ . It is therefore able to achieve a constant progress rate for increasing ξ . The steady state progress is less than that of the σ SA-ES. Due to the increase of σ_{ss}^* with increasing ξ , the increasing deviations of the approximation from the simulations can be explained: In the derivations leading to Equation (A.10), it has been assumed that $\mu N \gg \sigma^{(g)*2}$. As the steady state σ^* increases with ξ , N must be increased in order to have the same approximation quality for higher values of ξ . This can be explained more formally. Assuming large ξ , $(1 + \xi)/(2 + \xi) \simeq 1$ holds in Equation (87). Hence,

$$\sigma_{ss}^* \simeq \sqrt{\xi} \sqrt{2\mu(2\mu c_{\mu/\mu, \lambda}^2 - 1)} \quad (89)$$

follows, which - for large ξ - is of the same order as the one of the σ SA-ES (see Eq. (62) in Spettel and Beyer (2018b)). While it is common practice to use $c = \Theta(\frac{1}{N})$ since the seminal CMA-ES paper (Hansen and Ostermeier, 2001), this is the first theoretical result that shows an advantage of the $\Theta(\frac{1}{N})$ choice theoretically.

A further aspect for discussion is the special case of the non-recombinative $(1, \lambda)$ -CSA-ES that is contained in the derivations for the $(\mu/\mu_I, \lambda)$ -CSA-ES. The iterative system for the non-recombinative ($\mu = 1$) case can be derived analogously to the multi-recombinative ($\mu > 1$) case. The resulting equations differ in the expressions for $\varphi_x^{(g)}$, $\varphi_r^{(g)}$, and $\varphi_{r^2}^{(g)}$. It has been investigated further (not shown here) and for $N^2(1 + 1/\xi) \gg \sigma^{(g)*2}$ the mean value iterative systems of the $(1, \lambda)$ -CSA-ES and the $(\mu/\mu_I, \lambda)$ -CSA-ES agree. An interesting observation between the case $\mu = 1$ and the case $\mu > 1$ is the evolution near the cone boundary. Whereas the CSA-ES with $\mu = 1$ evolves on the boundary, the CSA-ES with $\mu > 1$ attains a certain steady state distance from the boundary (refer to the subplots of Figure 5 and Figure 6 showing $(\frac{x}{\sqrt{\xi r}})_{ss}$). This can be explained: Considering a parental individual on the cone boundary, the offspring are infeasible with overwhelming probability for sufficiently large N . Hence, they are repaired by projection and are on the boundary after projection. In particular, the best of them is on the boundary. Therefore, for $\mu = 1$, the ES evolves on the boundary. For $\mu > 1$, the centroid computation after projection results in offspring that are inside the feasible region due to its convexity.

5 Conclusions

In this work, the $(\mu/\mu_I, \lambda)$ -CSA-ES has been theoretically analyzed. For this, a mean value iterative system has been introduced and compared to real ES runs. Based on this derived system, steady state expressions have been derived and compared to ES simulations. It has been shown that for $c = \Theta(\frac{1}{N})$ a behavior that is similar to the σ SA can be achieved. For $c = \frac{1}{\sqrt{N}}$, the mutation strength achieved is not high enough to reach a similar steady state progress as for σ SA. Similar to the σ SA-ES, the CSA-ES evolves on the boundary for the non-recombinative case and it evolves at a certain distance from the boundary for the multi-recombinative case.

To conclude the paper, topics for future work are outlined. In addition to the σ SA and the CSA for the σ control mechanism, it is of interest to investigate the behavior of Meta-ESs applied to the conically constrained problem. Meta-ESs can be used as a further approach to control the mutation strength. The main idea is to evolve the mutation strength by competition of two ESs running with different mutation strengths for an isolation period. A Meta-ES controls the mutation strength evolution by picking and mutating the mutation strength of the better performing ES.

Comparison of the repair by projection approach with other repair methods is another topic for further research. Analysis of ESs applied to other constrained problems is another research direction for the future. A possibility for a direct next step is the analysis of the conically constrained linear optimization problem for which the objective function gradient does not coincide with the cone axis (Arnold, 2014).

Acknowledgments

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

References

- Arnold, D. V. (2011a). 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*, pages 853–860. ACM.
- Arnold, D. V. (2011b). 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*, pages 15–24. ACM.
- Arnold, D. V. (2013). 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*, pages 423–430. ACM.
- Arnold, D. V. (2014). On the behaviour of the $(1, \lambda)$ -ES for conically constrained linear problems. *Evolutionary Computation*, 22(3):503–523.
- Arnold, D. V. and Porter, J. (2015). 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*, pages 249–256, New York, NY, USA. ACM.
- Atamna, A., Auger, A., and Hansen, N. (2016). Augmented Lagrangian constraint handling for CMA-ES - case of a single linear constraint. In *International Conference on Parallel Problem Solving from Nature*, pages 181–191. Springer.
- Atamna, A., Auger, A., and Hansen, N. (2017). Linearly convergent evolution strategies via augmented Lagrangian constraint handling. In *Proceedings of the 14th ACM/SIGEVO Conference on Foundations of Genetic Algorithms*, pages 149–161. ACM.
- Beyer, H.-G. (2001). *The Theory of Evolution Strategies*. Natural Computing Series. Springer.
- Beyer, H.-G. and Finck, S. (2012). On the design of constraint covariance matrix self-adaptation evolution strategies including a cardinality constraint. *IEEE Transactions on Evolutionary Computation*, 16(4):578–596.
- Hansen, N. (2016). The CMA evolution strategy: A tutorial. arXiv:1604.00772 [cs.LG]. Available online: <http://arxiv.org/abs/1604.00772> (accessed Jun 07, 2017).
- Hansen, N. and Ostermeier, A. (1997). Convergence properties of evolution strategies with the derandomized covariance matrix adaptation: The $(\mu/\mu_I, \lambda)$ -CMA-ES. *Eufit*, 97:650–654.
- Hansen, N. and Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195.

- Hellwig, M. and Arnold, D. V. (2016). Comparison of constraint-handling mechanisms for the $(1, \lambda)$ -ES on a simple constrained problem. *Evolutionary Computation*, 24(1):1–23.
- Spettel, P. and Beyer, H.-G. (2018a). Analysis of the $(1, \lambda)$ - σ -self-adaptation evolution strategy with repair by projection applied to a conically constrained problem. *Theoretical Computer Science*. Available online: <http://www.sciencedirect.com/science/article/pii/S0304397518306571> (accessed Dec 14, 2018).
- Spettel, P. and Beyer, H.-G. (2018b). Analysis of the $(\mu/\mu_I, \lambda)$ - σ -self-adaptation evolution strategy with repair by projection applied to a conically constrained problem. arXiv:1812.06300 [cs.NE]. Available online: <https://arxiv.org/abs/1812.06300> (accessed Dec 18, 2018).
- Spettel, P. and Beyer, H.-G. (2018c). Technical report: Analysis of the $(1, \lambda)$ - σ -self-adaptation evolution strategy with repair by projection applied to a conically constrained problem. Technical Report TR-SAESCONe-18, Vorarlberg University of Applied Sciences. Available online: https://www.fhv.at/fileadmin/user_upload/fhv/files/forschung/ppe/working-papers/Analysis_onecommalambdasigmaSA-ES_Projection_Cone_TR.pdf (accessed Dec 14, 2018).
- Spettel, P., Beyer, H.-G., and Hellwig, M. (2018). A covariance matrix self-adaptation evolution strategy for optimization under linear constraints. *IEEE Transactions on Evolutionary Computation*, pages 1–1.

Analysis of the $(\mu/\mu_I, \lambda)$ -CSA-ES with Repair by Projection Applied to a Conically Constrained Problem

Patrick Spettel

patrick.spettel@fhv.at

Research Center Process and Product Engineering, Vorarlberg University of Applied Sciences, Dornbirn, 6850, Austria

Hans-Georg Beyer

hans-georg.beyer@fhv.at

Department of Computer Science, Research Center Process and Product Engineering, Vorarlberg University of Applied Sciences, Dornbirn, 6850, Austria

SUPPLEMENTARY MATERIAL

Appendix

This appendix contains material supplementing the main text. Appendix A contains a detailed derivation for the s_{\odot} mean value difference equation. Additional figures comparing the derived approximations with simulations for the iterative system and steady state considerations are presented in Appendix B. Appendix C provides plots of further cumulation parameter influence investigations. Finally, Appendix D shows plots that visualize the influences of the damping parameter.

A Derivation of a Mean Value Difference Equation for s_{\odot}

For s_{\odot} , a mean value difference equation can be derived using the update rule from Line 19 of Algorithm 1 and considering Equation (7). To begin with,

$$s_{\odot}^{(g+1)} = \frac{1}{r^{(g+1)}} \sum_{k=2}^N (\mathbf{x}^{(g+1)})_k (\mathbf{s}^{(g+1)})_k \quad (\text{A.1})$$

$$= \frac{1}{r^{(g+1)}} \sum_{k=2}^N \left[(\mathbf{x}^{(g)})_k + \sigma^{(g)} (\langle \tilde{\mathbf{z}}^{(g)} \rangle)_k \right] \left[(1-c)(\mathbf{s}^{(g)})_k + \sqrt{\mu c(2-c)} (\langle \tilde{\mathbf{z}}^{(g)} \rangle)_k \right] \quad (\text{A.2})$$

$$\begin{aligned} &= \frac{1}{r^{(g+1)}} \sum_{k=2}^N (1-c) \left[(\mathbf{x}^{(g)})_k (\mathbf{s}^{(g)})_k + \sigma^{(g)} (\langle \tilde{\mathbf{z}}^{(g)} \rangle)_k (\mathbf{s}^{(g)})_k \right] \\ &\quad + \frac{1}{r^{(g+1)}} \sum_{k=2}^N \sqrt{\mu c(2-c)} \left[(\mathbf{x}^{(g)})_k (\langle \tilde{\mathbf{z}}^{(g)} \rangle)_k + \sigma^{(g)} (\langle \tilde{\mathbf{z}}^{(g)} \rangle)_k (\langle \tilde{\mathbf{z}}^{(g)} \rangle)_k \right] \end{aligned} \quad (\text{A.3})$$

can be derived. Equation (A.3) can further be rewritten by the introduction of $z_{\odot}^{(g)} := \frac{1}{r^{(g)}} \sum_{k=2}^N (\mathbf{x}^{(g)})_k (\langle \tilde{\mathbf{z}}^{(g)} \rangle)_k$ (similar to Equation (7)) and use of Equation (14) resulting in

$$\begin{aligned} s_{\odot}^{(g+1)} &= \frac{r^{(g)}}{r^{(g+1)}} (1-c) \left[s_{\odot}^{(g)} + \frac{\sigma^{(g)*}}{N} (\langle \tilde{\mathbf{z}}^{(g)} \rangle)_{2..N}^T (\mathbf{s}^{(g)})_{2..N} \right] \\ &\quad + \frac{r^{(g)}}{r^{(g+1)}} \sqrt{\mu c(2-c)} \left[z_{\odot}^{(g)} + \frac{\sigma^{(g)*}}{N} \|(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_{2..N}\|^2 \right]. \end{aligned} \quad (\text{A.4})$$

For the fraction $r^{(g)}/r^{(g+1)}$, $r^{(g+1)}$ has to be derived. From the offspring generation and selection steps it follows that

$$r^{(g+1)} = \sqrt{\sum_{k=2}^N ((\mathbf{x}^{(g)})_k + \sigma^{(g)} (\langle \tilde{\mathbf{z}}^{(g)} \rangle)_k)^2} \quad (\text{A.5})$$

$$= \sqrt{\sum_{k=2}^N \left((\mathbf{x}^{(g)})_k^2 + 2\sigma^{(g)} (\mathbf{x}^{(g)})_k (\langle \tilde{\mathbf{z}}^{(g)} \rangle)_k + \sigma^{(g)2} (\langle \tilde{\mathbf{z}}^{(g)} \rangle)_k^2 \right)} \quad (\text{A.6})$$

$$= \sqrt{r^{(g)2} + 2 \frac{\sigma^{(g)*}}{N} r^{(g)2} z_{\odot}^{(g)} + \frac{\sigma^{(g)*2}}{N^2} r^{(g)2} \|(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_{2..N}\|^2} \quad (\text{A.7})$$

holds. Using the result from Equation (A.7),

$$\frac{r^{(g)}}{r^{(g+1)}} = \sqrt{\frac{1}{1 + \frac{2\sigma^{(g)*}}{N} z_{\odot}^{(g)} + \frac{\sigma^{(g)*2}}{N^2} \|(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_{2..N}\|^2}} \quad (\text{A.8})$$

can be derived. For further simplification of Equation (A.8), asymptotic assumptions are made for $N \rightarrow \infty$. Because the mutation vector is corrected in case of projection (Line 11 in Algorithm 1), $\langle \tilde{\mathbf{z}}^{(g)} \rangle$ denotes the centroid of the μ best (w.r.t. fitness)

offspring mutation vectors after the projection step. Approximation of $\langle \tilde{\mathbf{z}}^{(g)} \rangle$ for the asymptotic case by its value before projection and selection yields a normal distribution for $(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_k = \frac{1}{\mu} \sum_{m=1}^{\mu} (\tilde{\mathbf{z}}_{m;\lambda})_k \sim \mathcal{N}(0, \frac{1}{\mu}) = \frac{1}{\sqrt{\mu}} \mathcal{N}(0, 1)$, which follows by the properties of a sum of normal distributed random variables.

Hence, $\|(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_{2..N}\|^2$ can be approximated by a χ^2 distribution with $N-1$ degrees of freedom. As the expected value of the χ^2 distribution corresponds to its number of degrees of freedom,

$$\frac{\|(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_{2..N}\|^2}{N} \simeq \frac{1}{\mu} \quad (\text{A.9})$$

follows for $N \rightarrow \infty$ by the law of large numbers. With Equation (A.9) and the assumptions $N \gg 2\sigma^{(g)*} z_{\odot}^{(g)}$ and $\mu N \gg \sigma^{(g)*2}$,

$$\frac{r^{(g)}}{r^{(g+1)}} \simeq 1 \quad (\text{A.10})$$

follows. Making use of Equation (A.10) and Equation (A.9), Equation (A.4) can be simplified for the asymptotic case $N \rightarrow \infty$ yielding

$$\begin{aligned} s_{\odot}^{(g+1)} &\simeq (1-c)s_{\odot}^{(g)} + (1-c)\frac{\sigma^{(g)*}}{N}(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_{2..N}^T (\mathbf{s}^{(g)})_{2..N} \\ &\quad + \sqrt{\mu c(2-c)} z_{\odot}^{(g)} + \sqrt{\mu c(2-c)} \frac{\sigma^{(g)*}}{\mu}. \end{aligned} \quad (\text{A.11})$$

Taking expected values of Equation (A.11) with $E[s_{\odot}^{(g+1)}] := \overline{s_{\odot}^{(g+1)}}$ results in

$$\begin{aligned} \overline{s_{\odot}^{(g+1)}} &\simeq (1-c)\overline{s_{\odot}^{(g)}} + (1-c)\frac{\overline{\sigma^{(g)*}}}{N}E[(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_{2..N}^T (\mathbf{s}^{(g)})_{2..N}] \\ &\quad + \sqrt{\mu c(2-c)}E[z_{\odot}^{(g)}] + \sqrt{\mu c(2-c)}\frac{\overline{\sigma^{(g)*}}}{\mu}. \end{aligned} \quad (\text{A.12})$$

To treat Equation (A.12) further, $E[(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_{2..N}^T (\mathbf{s}^{(g)})_{2..N}]$ and $E[z_{\odot}^{(g)}]$ need to be derived.

For $E[(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_{2..N}^T (\mathbf{s}^{(g)})_{2..N}]$, $(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_{2..N}$ is decomposed into a vector in direction of the parental individual's $2..N$ components $\mathbf{e}_{\odot}^{(g)}$ and in a direction $\mathbf{e}_{\ominus}^{(g)}$ that is orthogonal to $\mathbf{e}_{\odot}^{(g)}$, i.e., $\mathbf{e}_{\odot}^{(g)T} \mathbf{e}_{\ominus}^{(g)} = 0$. Further, in the following the assumption is made that those direction vectors are unit vectors, i.e., $\|\mathbf{e}_{\odot}^{(g)}\| = 1$ and $\|\mathbf{e}_{\ominus}^{(g)}\| = 1$. Therefore, $(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_{2..N}$ can be written as

$$(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_{2..N} = z_{\odot}^{(g)} \mathbf{e}_{\odot}^{(g)} + z_{\ominus}^{(g)} \mathbf{e}_{\ominus}^{(g)}, \quad (\text{A.13})$$

where $z_{\odot}^{(g)}$ and $z_{\ominus}^{(g)}$ are the projections of the mutation vector in direction of $\mathbf{e}_{\odot}^{(g)}$ and $\mathbf{e}_{\ominus}^{(g)}$, respectively. Using Equation (A.13),

$$(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_{2..N}^T (\mathbf{s}^{(g)})_{2..N} = z_{\odot}^{(g)} \underbrace{\mathbf{e}_{\odot}^{(g)T} (\mathbf{s}^{(g)})_{2..N}}_{=s_{\odot}^{(g)}} + z_{\ominus}^{(g)} \mathbf{e}_{\ominus}^{(g)T} (\mathbf{s}^{(g)})_{2..N} \quad (\text{A.14})$$

follows. Note that $\mathbf{e}_{\odot}^{(g)T} (\mathbf{s}^{(g)})_{2..N}$ corresponds to the definition in Equation (7). Taking into account the statistical independence of the cumulated path vector and the mutation

in the current generation, taking expectation results in

$$\mathbb{E}[(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_{2..N}^T (\mathbf{s}^{(g)})_{2..N}] = \mathbb{E}[z_{\odot}^{(g)}] \mathbb{E}[s_{\odot}^{(g)}] + \mathbb{E}[z_{\odot}^{(g)}] \mathbb{E}[\mathbf{e}_{\ominus}^{(g)T} (\mathbf{s}^{(g)})_{2..N}] \quad (\text{A.15})$$

$$= \mathbb{E}[z_{\odot}^{(g)}] \overline{s_{\odot}^{(g)}}. \quad (\text{A.16})$$

Note that $\mathbb{E}[z_{\odot}^{(g)}]$ vanishes because the mutations in direction $\mathbf{e}_{\ominus}^{(g)}$ are isotropic and selectively neutral. Hence, the second summand of Equation (A.15) is 0 in expectation.

$\mathbb{E}[z_{\odot}^{(g)}]$ can be calculated from the progress rate of the quadratic distance from the cone axis. It writes

$$\varphi_{r^2}^{(g)} := \mathbb{E}[r^{(g)^2} - r^{(g+1)^2}] = r^{(g)^2} - \mathbb{E}[r^{(g+1)^2}] = r^{(g)^2} - \mathbb{E}[\langle q_r \rangle^2] \quad (\text{A.17})$$

$$\approx r^{(g)^2} - \left\{ P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \mathbb{E}[\langle q_r \rangle_{\text{feas}}^2] + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] \mathbb{E}[\langle q_r \rangle_{\text{infeas}}^2] \right\} \quad (\text{A.18})$$

where $\langle q_r \rangle$ denotes the distance from the cone boundary of the centroid after projection (cf. Line 22 of Algorithm 1). Expressions for $\mathbb{E}[\langle q_r \rangle_{\text{feas}}^2]$ and $\mathbb{E}[\langle q_r \rangle_{\text{infeas}}^2]$ have already been derived in Appendix D in the supplementary material of Spettel and Beyer (2018b). The used Taylor approximation in Equation (D.157) of that work allows using the square of Equation (D.165) for the feasible case yielding

$$\mathbb{E}[\langle q_r \rangle_{\text{feas}}^2] \approx r^{(g)^2} + \frac{\sigma^{(g)^2}}{\mu} (N - 1). \quad (\text{A.19})$$

Similarly, the Taylor expansion used in Equation (D.172) of that work allows using the square of Equation (D.217) as an approximation for the infeasible case. It reads

$$\mathbb{E}[\langle q_r \rangle_{\text{infeas}}^2] \approx \frac{\mathbb{E}[\langle q \rangle_{\text{infeas}}]^2}{\xi} \left(\frac{1 + \frac{\sigma^{(g)*2}}{\mu N}}{1 + \frac{\sigma^{(g)*2}}{N}} \right). \quad (\text{A.20})$$

Using Equation (A.19) and Equation (A.20),

$$\begin{aligned} \varphi_{r^2}^{(g)} \approx & r^{(g)^2} - \left\{ P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)}) \left[r^{(g)^2} + \frac{\sigma^{(g)^2}}{\mu} (N - 1) \right] \right. \\ & \left. + [1 - P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma^{(g)})] \left[\frac{\mathbb{E}[\langle q \rangle_{\text{infeas}}]^2}{\xi} \left(\frac{1 + \frac{\sigma^{(g)*2}}{\mu N}}{1 + \frac{\sigma^{(g)*2}}{N}} \right) \right] \right\} \end{aligned} \quad (\text{A.21})$$

follows, where a closed-form approximation

$$\mathbb{E}[\langle q \rangle_{\text{infeas}}] \approx \frac{\xi}{1 + \xi} \left(x^{(g)} + \bar{r} / \sqrt{\xi} \right) - \frac{\xi}{1 + \xi} \sqrt{\sigma^{(g)^2} + \sigma_r^2 / \xi c_{\mu/\mu, \lambda}} \quad (\text{A.22})$$

has been derived in Spettel and Beyer (2018b) as well (refer to the derivations leading to Equation (C.149) in Appendix C in the supplementary of that work for the details). With Equation (A.21) and Equation (A.22), $\varphi_{r^2}^{(g)}$ can be computed for a given state of the system. The goal is now to express $\varphi_{r^2}^{(g)}$ in terms of $\mathbb{E}[z_{\odot}^{(g)}]$. Subsequently solving

for $E[z_{\odot}^{(g)}]$ allows then to compute its value. Using Equation (A.17) and Equation (A.7) with Equation (A.9), $\varphi_{r^2}^{(g)}$ can alternatively be written as

$$\varphi_{r^2}^{(g)} = \overline{r^{(g)}}^2 - E \left[\overline{r^{(g)}}^2 + 2\overline{\sigma^{(g)}} \overline{r^{(g)}} z_{\odot}^{(g)} + \overline{\sigma^{(g)}}^2 \|(\langle \tilde{\mathbf{z}}^{(g)} \rangle)_{2..N}\|^2 \right] \quad (\text{A.23})$$

$$\simeq -2\overline{\sigma^{(g)}} \overline{r^{(g)}} E \left[z_{\odot}^{(g)} \right] - \overline{\sigma^{(g)}}^2 \frac{N}{\mu}. \quad (\text{A.24})$$

Equation (A.24) can be solved for $E[z_{\odot}^{(g)}]$ yielding

$$E[z_{\odot}^{(g)}] \simeq - \left(\frac{\varphi_{r^2}^{(g)} + \overline{\sigma^{(g)}}^2 \frac{N}{\mu}}{2\overline{\sigma^{(g)}} \overline{r^{(g)}}} \right) = - \frac{\varphi_{r^2}^{(g)}}{2\overline{\sigma^{(g)}} \overline{r^{(g)}}} - \frac{\overline{\sigma^{(g)}}^2 N}{2\overline{\sigma^{(g)}} \overline{r^{(g)}} \mu} = - \frac{N\varphi_{r^2}^{(g)}}{2\overline{\sigma^{(g)}}^* \overline{r^{(g)}}^2} - \frac{\overline{\sigma^{(g)}}^*}{2\mu}. \quad (\text{A.25})$$

Reinsertion of Equation (A.16) and Equation (A.25) into Equation (A.12) yields

$$\overline{s_{\odot}^{(g+1)}} \simeq (1-c) \left(\overline{s_{\odot}^{(g)}} + \frac{\overline{\sigma^{(g)}}^*}{N} E[z_{\odot}^{(g)}] \overline{s_{\odot}^{(g)}} \right) + \sqrt{\mu c(2-c)} \left(E[z_{\odot}^{(g)}] + \frac{\overline{\sigma^{(g)}}^*}{\mu} \right) \quad (\text{A.26})$$

$$\begin{aligned} &\simeq (1-c) \left(1 + \frac{\overline{\sigma^{(g)}}^*}{N} \left(- \frac{N\varphi_{r^2}^{(g)}}{2\overline{\sigma^{(g)}}^* \overline{r^{(g)}}^2} - \frac{\overline{\sigma^{(g)}}^*}{2\mu} \right) \right) \overline{s_{\odot}^{(g)}} \\ &\quad + \sqrt{\mu c(2-c)} \left(- \frac{N\varphi_{r^2}^{(g)}}{2\overline{\sigma^{(g)}}^* \overline{r^{(g)}}^2} + \frac{\overline{\sigma^{(g)}}^*}{2\mu} \right). \end{aligned} \quad (\text{A.27})$$

B Additional Results Comparing the Derived Approximations with Simulations

Figures 7 to 12 show the mean value dynamics of the $(3/3_I, 10)$ -CSA-ES applied to the conically constrained problem with different parameters as indicated in the title of the subplots. The plots are organized into three rows and two columns. The first two rows show the x (first row, first column), r (first row, second column), σ (second row, first column), and σ^* (second row, second column) dynamics. The third row shows x and r converted into each other by $\sqrt{\xi}$. The third row shows that after some initial phase, the ES transitions into a stationary state. In this steady state, the ES moves along the cone boundary. This becomes clear in the plots because the equation for the cone boundary is $r = x/\sqrt{\xi}$ or equivalently $x = r\sqrt{\xi}$. In the first two rows, the solid line has been generated by averaging 100 real runs of the ES. The dashed line has been determined by iterating the mean value iterative system of Section 3.1.6 with one-generation experiments for $\varphi_x^{(g)}$, $\varphi_x^{(g)*}$, $\varphi_r^{(g)}$, $\varphi_r^{(g)*}$, and $\varphi_{r^2}^{(g)}$. The dotted lines have been computed by iterating the mean value iterative system with the derived approximations as indicated in the derivations leading to the equations in Section 3.1.6 for $\varphi_x^{(g)}$, $\varphi_x^{(g)*}$, $\varphi_r^{(g)}$, $\varphi_r^{(g)*}$, and $\varphi_{r^2}^{(g)}$. Due to the approximations used it is possible that in a generation g , the iteration of the mean value iterative system yields infeasible $(x^{(g)}, r^{(g)})^T$. In such cases, the particular $(x^{(g)}, r^{(g)})^T$ have been projected back and projected values have been used in the further iterations.

Figure 13 and Figure 14 show plots of the steady state computations using the numerical solution of Equation (62) for σ_{ss}^* , which has been compared to real ES runs. The results for φ_x^* and φ_r^* have been determined by using the numerically determined

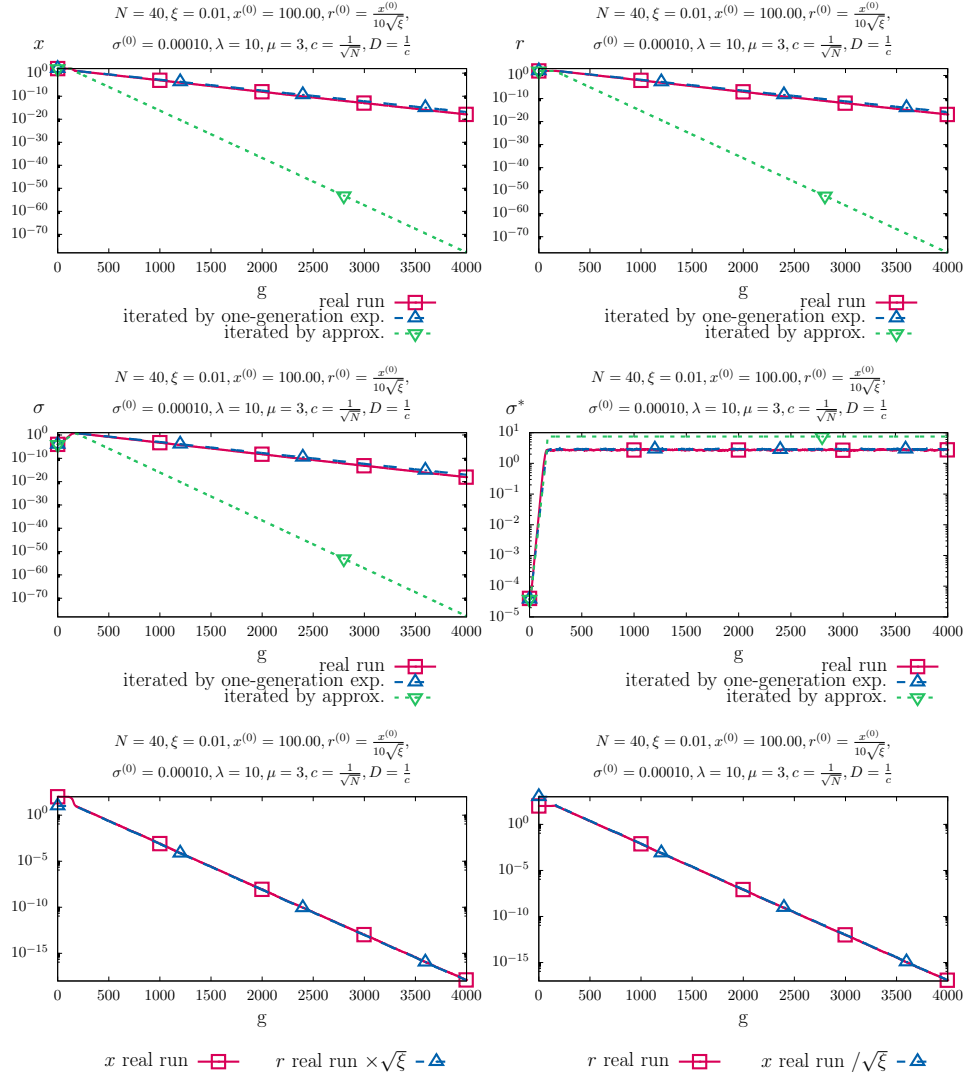


Figure 7: Mean value dynamics closed-form approximation and real-run comparison of the $(3/3_I, 10)$ -CSA-ES with repair by projection applied to the conically constrained problem. (Part 1)

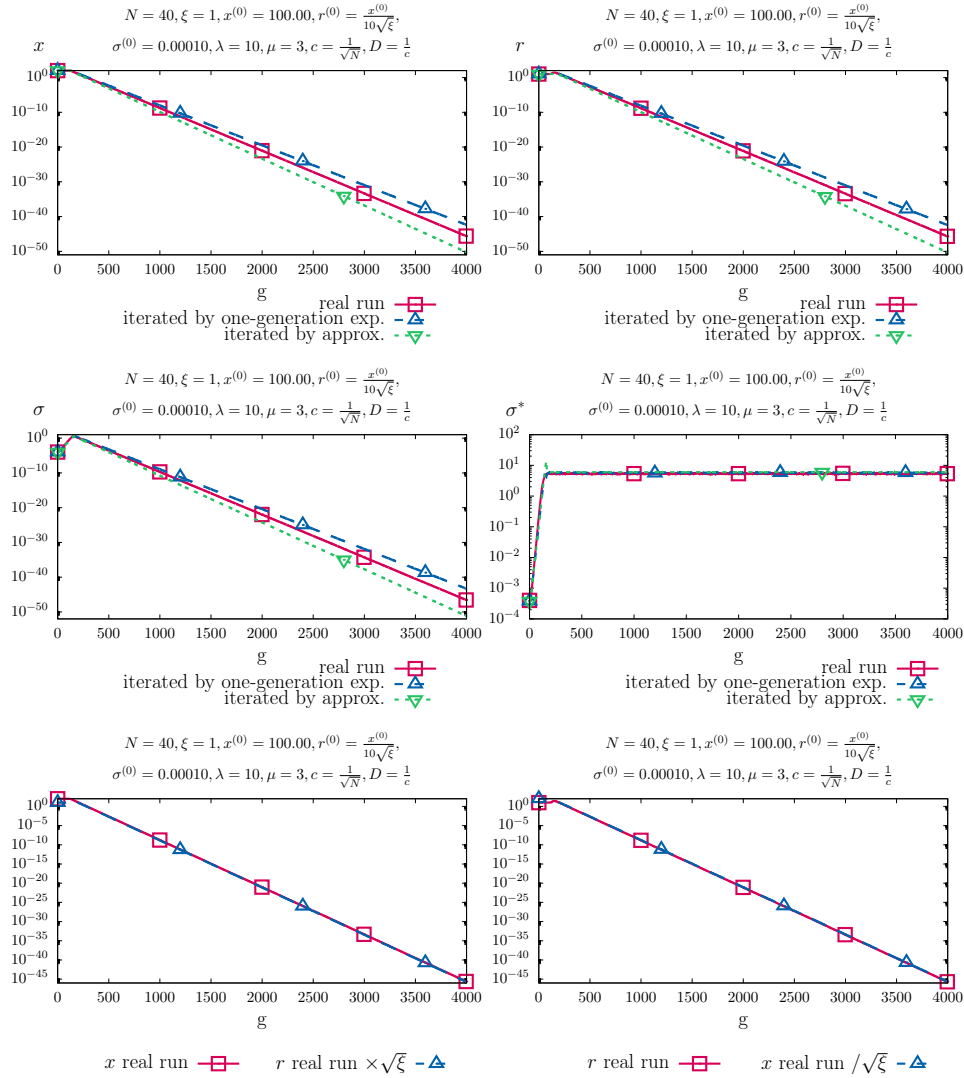


Figure 8: Mean value dynamics closed-form approximation and real-run comparison of the (3/3_I, 10)-CSA-ES with repair by projection applied to the conically constrained problem. (Part 2)

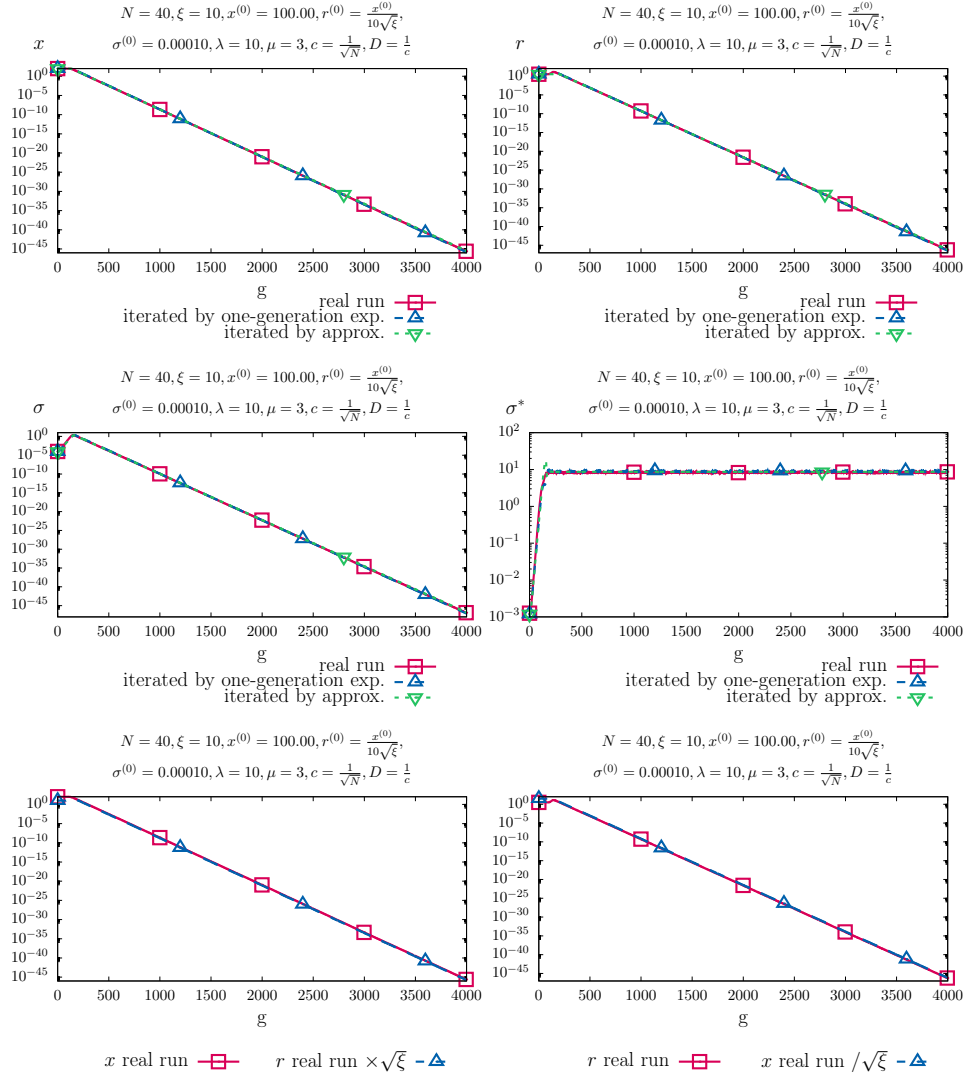


Figure 9: Mean value dynamics closed-form approximation and real-run comparison of the $(3/3_I, 10)$ -CSA-ES with repair by projection applied to the conically constrained problem. (Part 3)

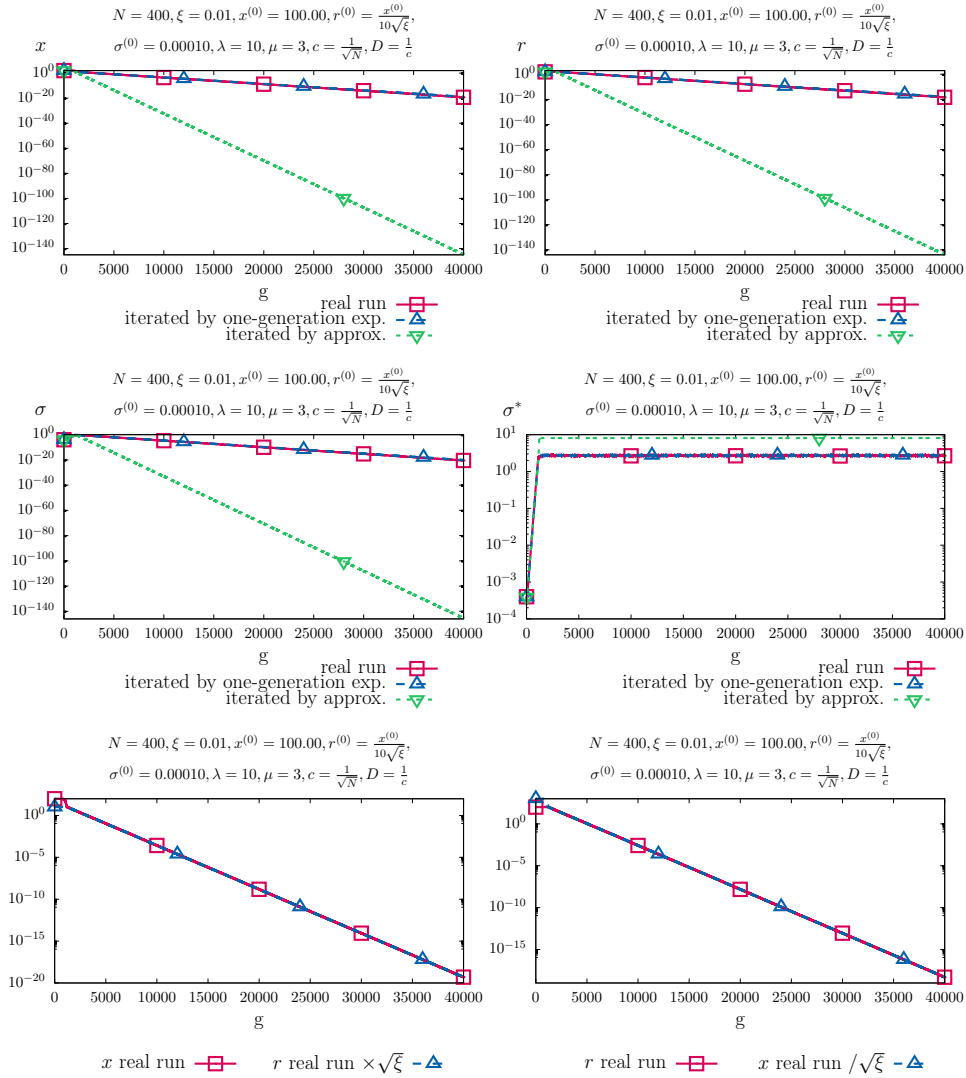


Figure 10: Mean value dynamics closed-form approximation and real-run comparison of the $(3/3_I, 10)$ -CSA-ES with repair by projection applied to the conically constrained problem. (Part 4)

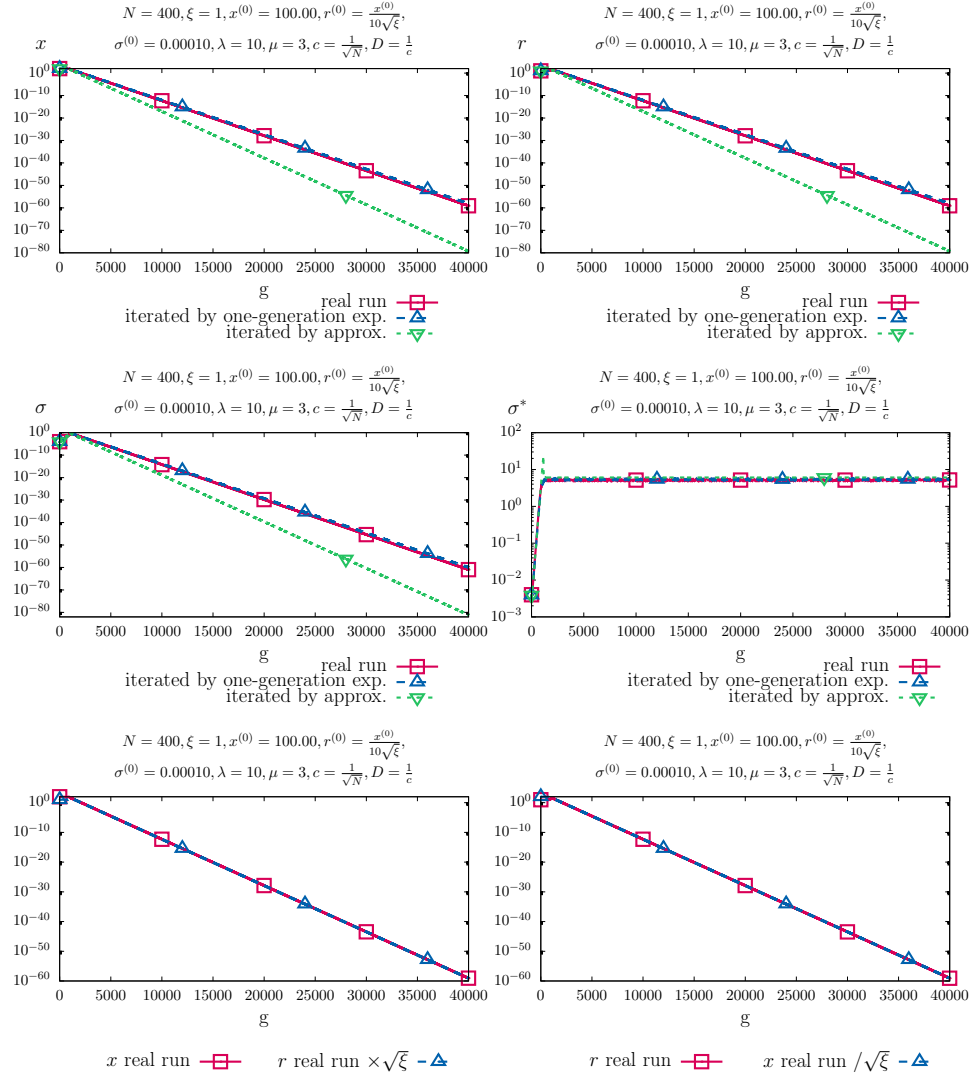


Figure 11: Mean value dynamics closed-form approximation and real-run comparison of the $(3/3_I, 10)$ -CSA-ES with repair by projection applied to the conically constrained problem. (Part 5)

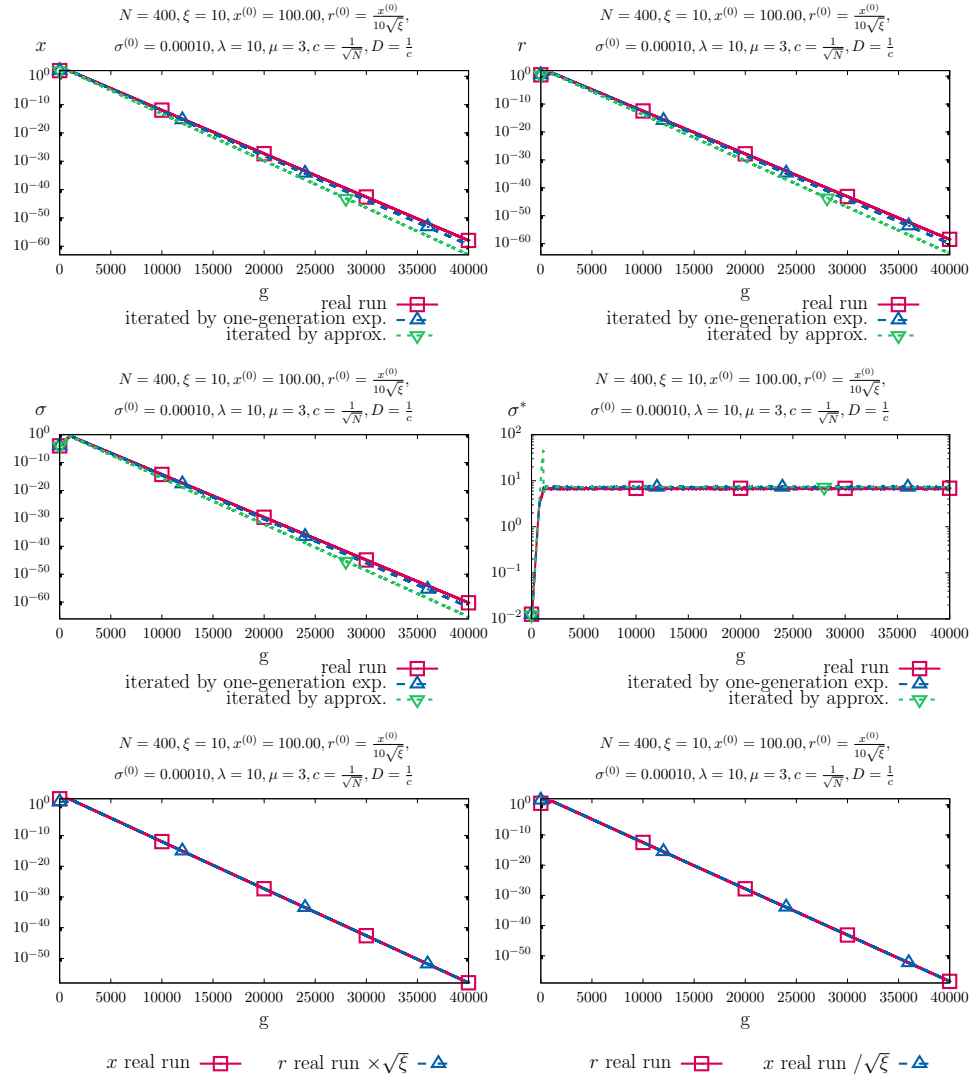


Figure 12: Mean value dynamics closed-form approximation and real-run comparison of the $(3/3_I, 10)$ -CSA-ES with repair by projection applied to the conically constrained problem. (Part 6)

steady state σ_{ss}^* values with Equation (48). The approximations for $\left(\frac{x}{\sqrt{\xi}r}\right)_{ss}$ have been determined by evaluating Equation (47). The values for the points denoting the experiments have been determined by computing the averages of the particular values in real ES runs.

Figures 15 to 22 show plots of the steady state computations. Results computed by Equation (74) for Figures 15 to 18 and Equation (87) for Figures 19 to 22 have been compared to real $(\mu/\mu_I, \lambda)$ -CSA-ES runs for $\lambda = 10$. Different sets of μ values are considered. Figures 15, 16, 19 and 20 show the results for $\mu \in \{1, 2, 3\}$ and Figures 17, 18, 21 and 22 show the results for $\mu \in \{4, 6, 8\}$. The values for the points denoting the approximations have been determined by computing the normalized steady state mutation strength σ_{ss}^* using Equation (74) for Figures 15 to 18 and Equation (87) for Figures 19 to 22, for different values of ξ . The results for φ_x^* and φ_r^* have been determined by using the computed steady state σ_{ss}^* values with Equation (48). The approximations for $\left(\frac{x}{\sqrt{\xi}r}\right)_{ss}$ have been determined by evaluating Equation (47). The values for the points denoting the experiments have been determined by computing the averages of the particular values in real ES runs. The figures show that the derived expressions get better for larger values of ξ and N . Again, the deviations for small ξ are due to approximations in the derivation of the local progress rates. The deviations for large N stem from the use of asymptotic assumptions $N \rightarrow \infty$.

C Additional Results of the Cumulation Parameter's Influence

Figure 23 shows the influence of the cumulation parameter for the considered $(3/3_I, 10)$ -CSA-ES for the dimensions 1000 and 10000. They extend the results shown in the main text. The steady state normalized x progress rate (Equation (48)) is plotted against the steady state normalized mutation strength for different values of ξ . Steady state values achieved for different settings of c with $D = \frac{1}{c}$ have been computed numerically (using Equation (62) with Equations (48) and (53) to (55)) and are shown with the filled black markers.

D Additional Investigations of the Damping Parameter

Figures 24 to 27 show the influence of the damping parameter for the considered $(3/3_I, 10)$ -CSA-ES for the dimensions 40, 400, 1000, and 10000. The steady state normalized x progress rate (Equation (48)) is plotted against the steady state normalized mutation strength for different values of ξ . Steady state values achieved for different settings of D with $c = \frac{1}{\sqrt{N}}$ (Figures 24 and 25) and $c = \frac{1}{N}$ (Figures 26 and 27) have been computed numerically (using Equation (62) with Equations (48) and (53) to (55)) and are shown with the filled black markers.

The plots show that for $c = 1/\sqrt{N}$ the damping parameters considered only have an effect for dimension 40. In contrast to that, for $c = 1/N$, the damping parameters considered can increase the steady state x progress for all the dimensions and values of ξ shown.

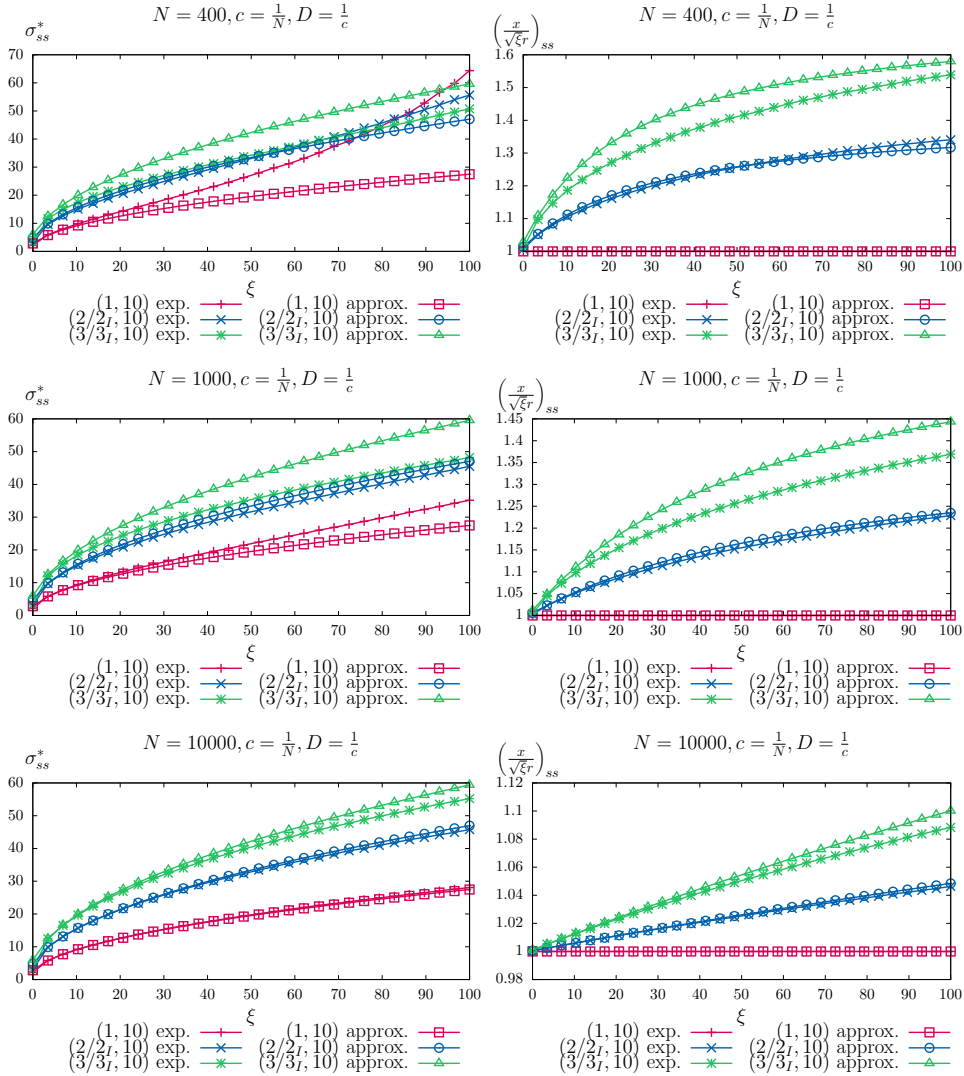


Figure 13: Steady state comparison of the $(\mu/\mu_I, \lambda)$ -CSA-ES with repair by projection applied to the conically constrained problem for $c = \frac{1}{N}$, $D = \frac{1}{c}$, $\lambda = 10$, $\mu \in \{1, 2, 3\}$. The numerical solution of Equation (62) for σ_{ss}^* has been considered for these subplots. (Part 1)

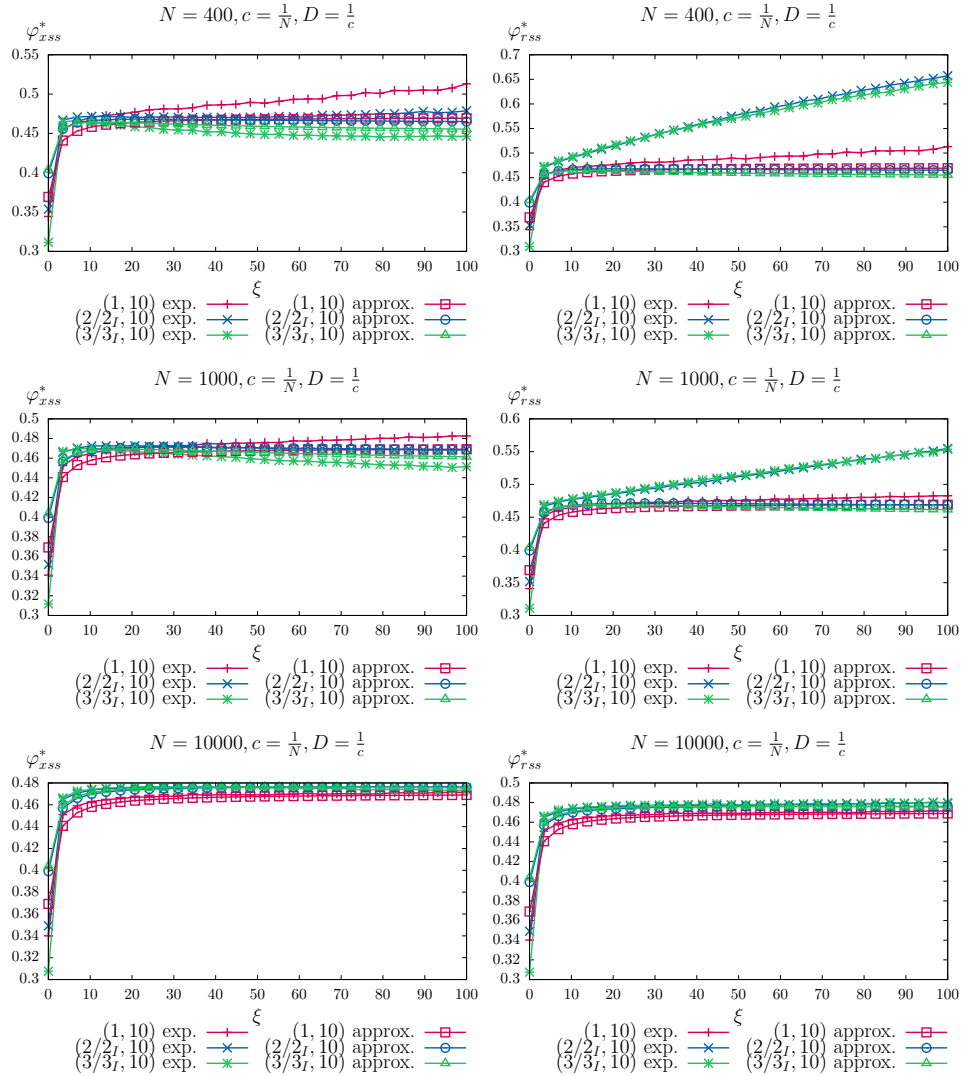


Figure 14: Steady state comparison of the $(\mu/\mu_I, \lambda)$ -CSA-ES with repair by projection applied to the conically constrained problem for $c = \frac{1}{N}$, $D = \frac{1}{c}$, $\lambda = 10$, $\mu \in \{1, 2, 3\}$. The numerical solution of Equation (62) for σ_{ss}^* has been considered for these subplots. (Part 2)

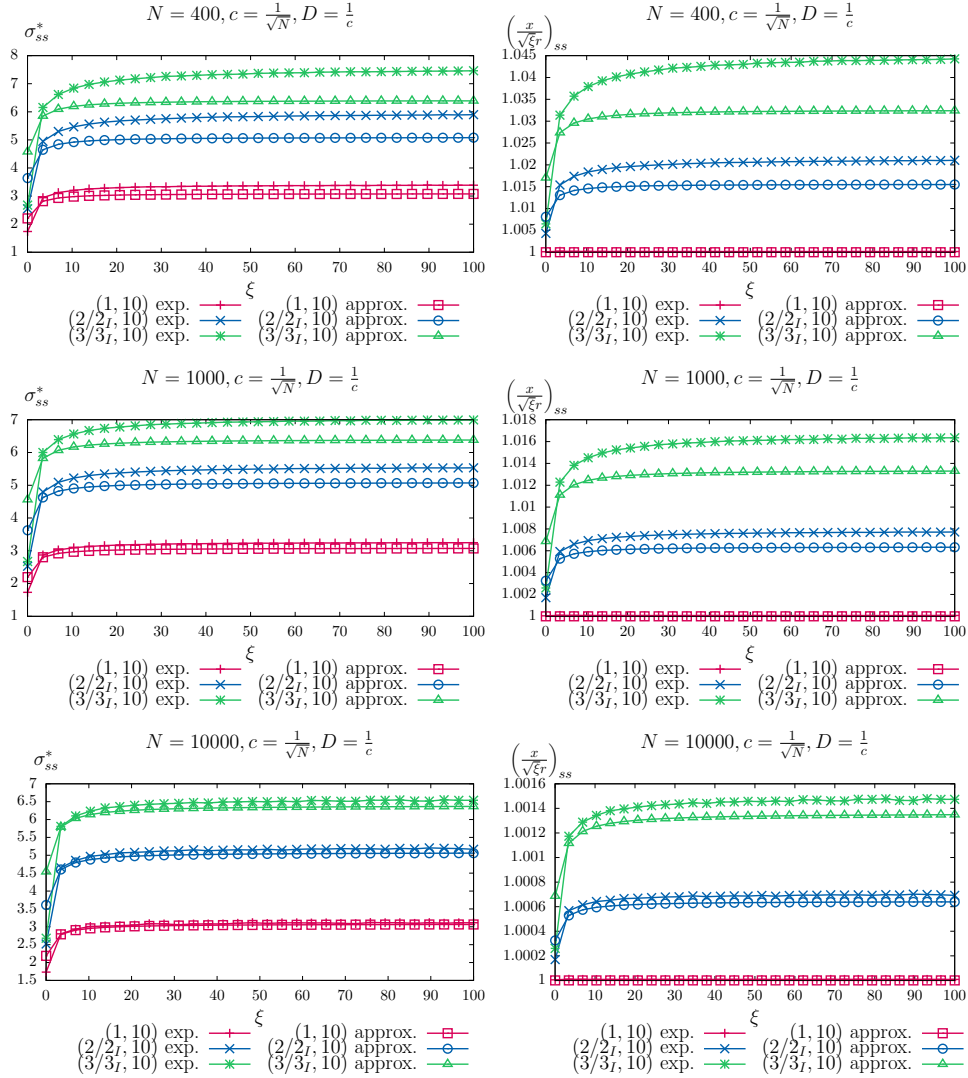


Figure 15: Steady state closed-form approximation and real run comparison of the $(\mu/\mu_I, \lambda)$ -CSA-ES with repair by projection applied to the conically constrained problem for $c = \frac{1}{\sqrt{N}}, D = \frac{1}{c}, \lambda = 10, \mu \in \{1, 2, 3\}$. (Part 1)

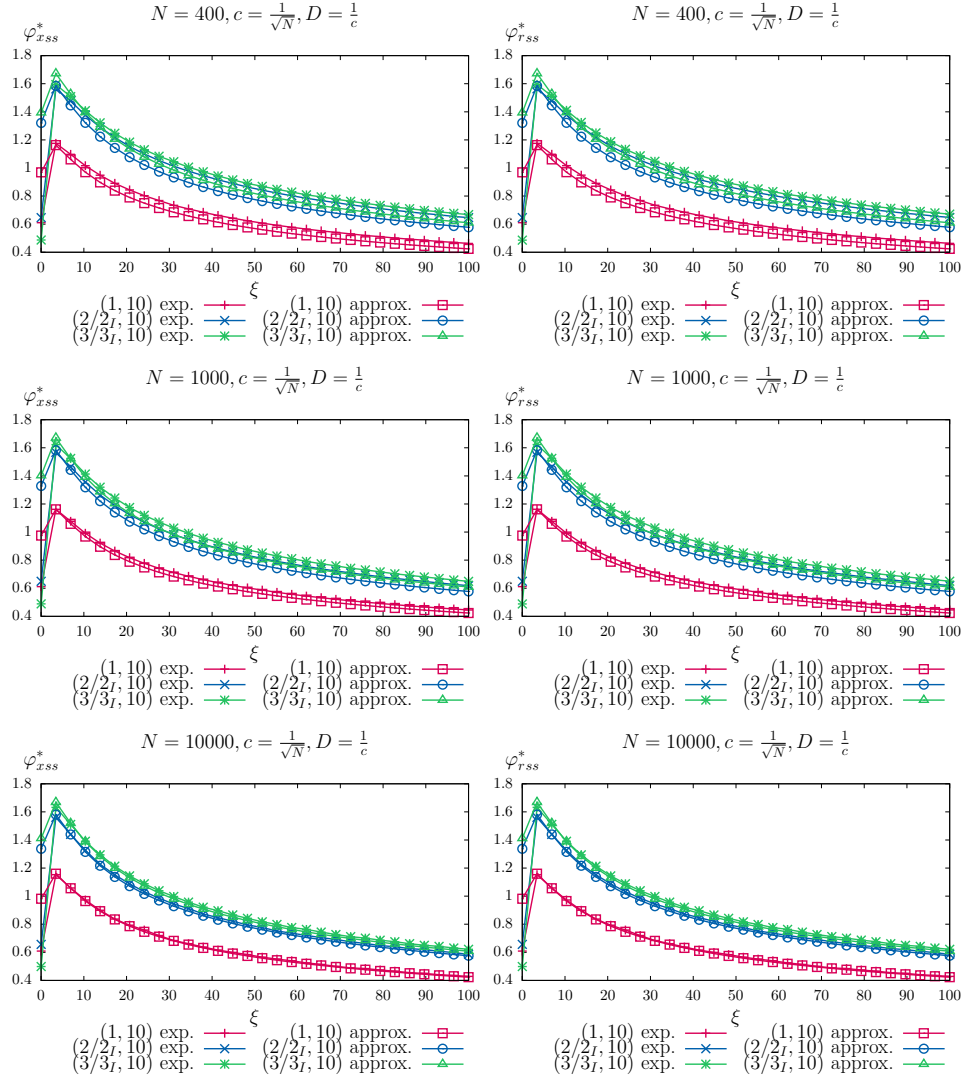


Figure 16: Steady state closed-form approximation and real run comparison of the $(\mu/\mu_I, \lambda)$ -CSA-ES with repair by projection applied to the conically constrained problem for $c = \frac{1}{\sqrt{N}}$, $D = \frac{1}{c}$, $\lambda = 10$, $\mu \in \{1, 2, 3\}$. (Part 2)

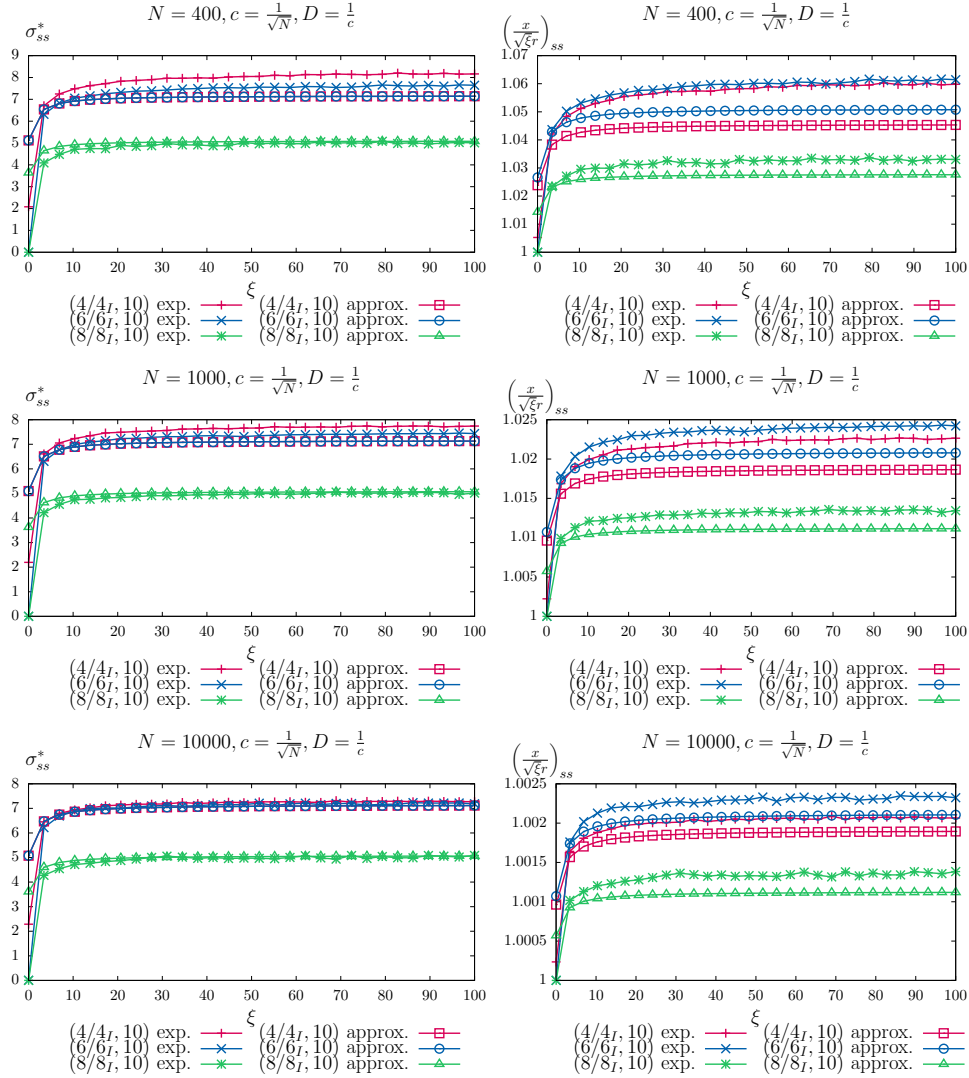


Figure 17: Steady state closed-form approximation and real run comparison of the $(\mu/\mu_I, \lambda)$ -CSA-ES with repair by projection applied to the conically constrained problem for $c = \frac{1}{\sqrt{N}}$, $D = \frac{1}{c}$, $\lambda = 10$, $\mu \in \{4, 6, 8\}$. (Part 1)

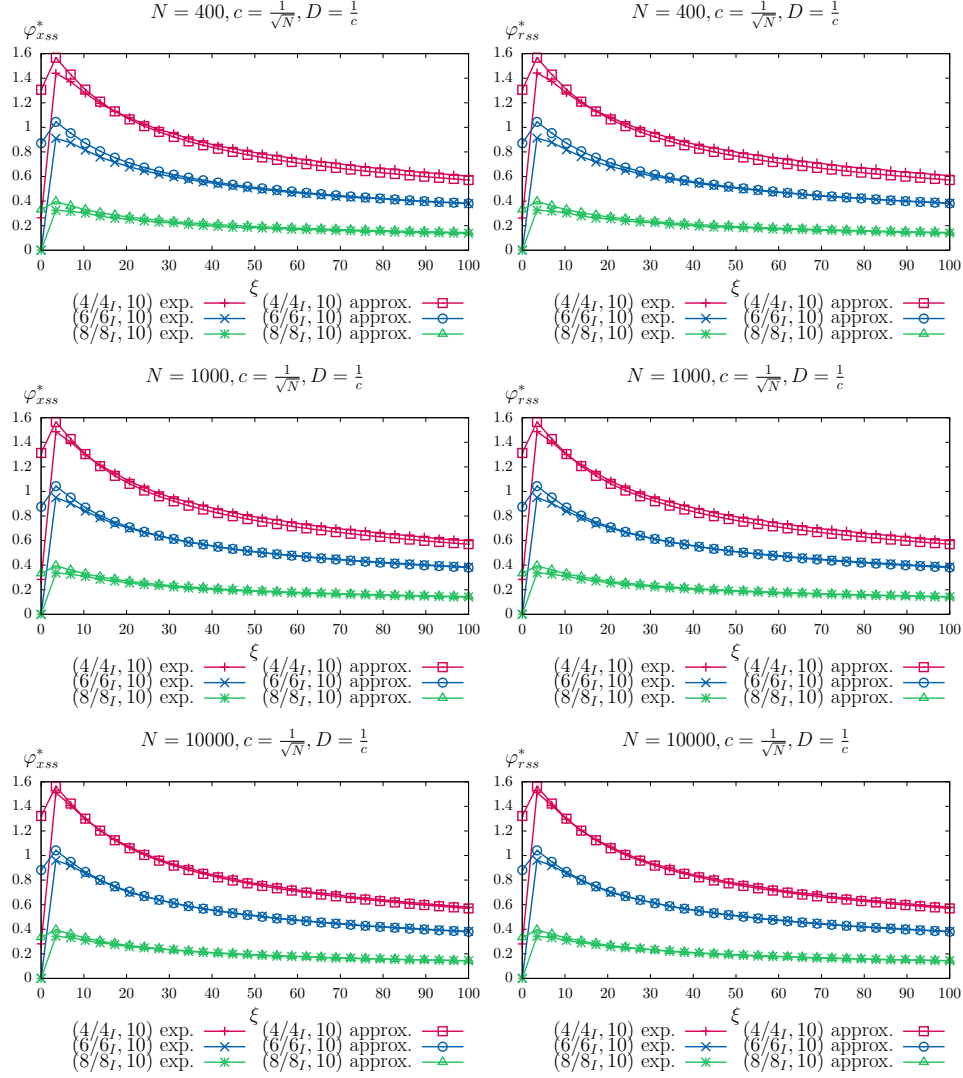


Figure 18: Steady state closed-form approximation and real run comparison of the $(\mu/\mu_I, \lambda)$ -CSA-ES with repair by projection applied to the conically constrained problem for $c = \frac{1}{\sqrt{N}}$, $D = \frac{1}{c}$, $\lambda = 10$, $\mu \in \{4, 6, 8\}$. (Part 2)

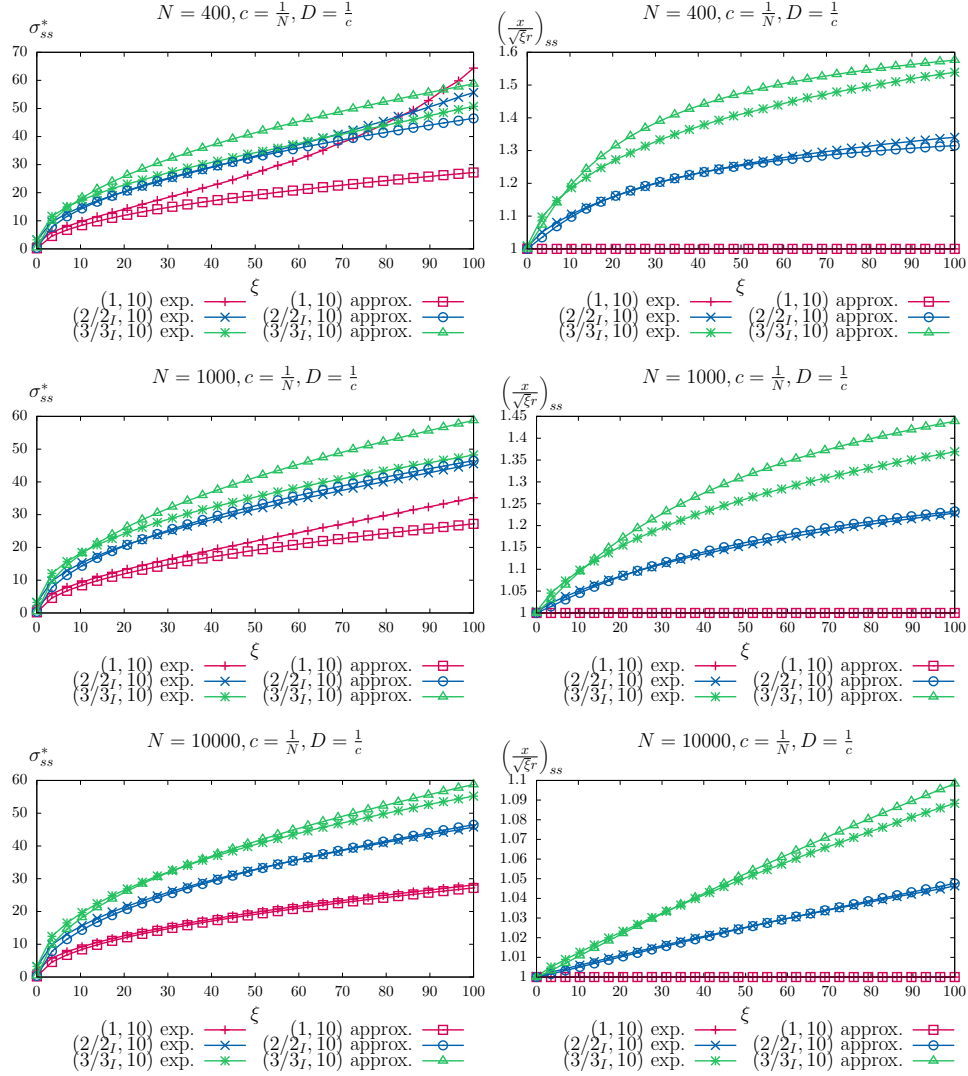


Figure 19: Steady state closed-form approximation and real run comparison of the $(\mu/\mu_I, \lambda)$ -CSA-ES with repair by projection applied to the conically constrained problem for $c = \frac{1}{N}$, $D = \frac{1}{c}$, $\lambda = 10$, $\mu \in \{1, 2, 3\}$. (Part 1)

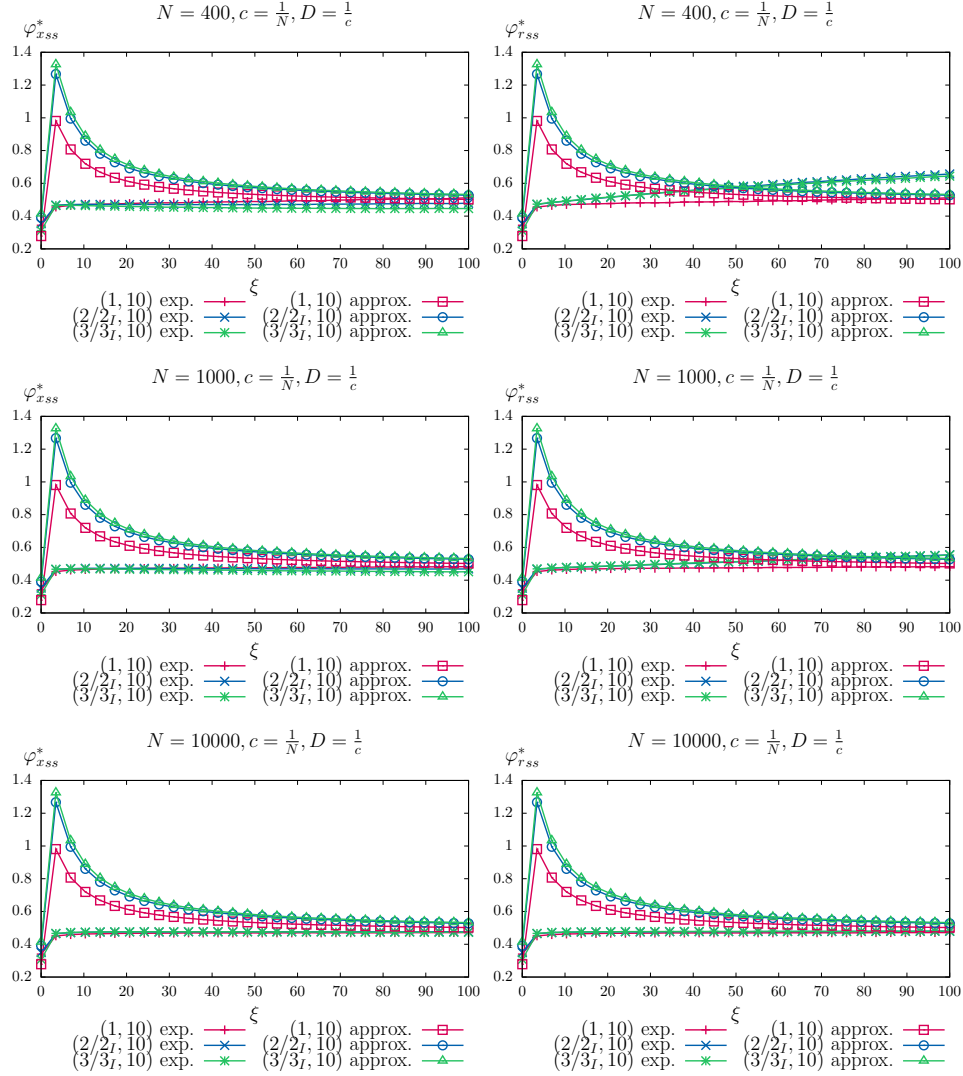


Figure 20: Steady state closed-form approximation and real run comparison of the $(\mu/\mu_I, \lambda)$ -CSA-ES with repair by projection applied to the conically constrained problem for $c = \frac{1}{N}$, $D = \frac{1}{c}$, $\lambda = 10$, $\mu \in \{1, 2, 3\}$. (Part 2)

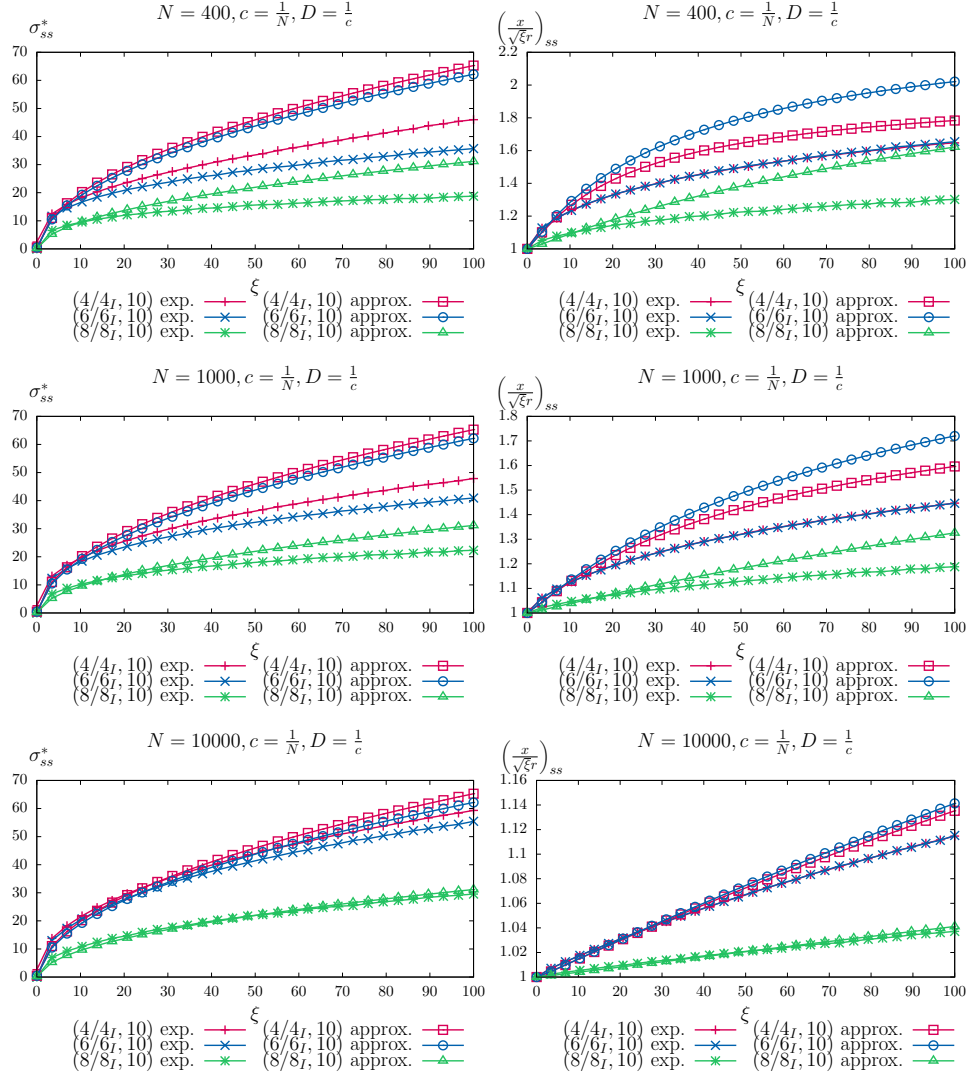


Figure 21: Steady state closed-form approximation and real run comparison of the $(\mu/\mu_I, \lambda)$ -CSA-ES with repair by projection applied to the conically constrained problem for $c = \frac{1}{N}$, $D = \frac{1}{c}$, $\lambda = 10$, $\mu \in \{4, 6, 8\}$. (Part 1)

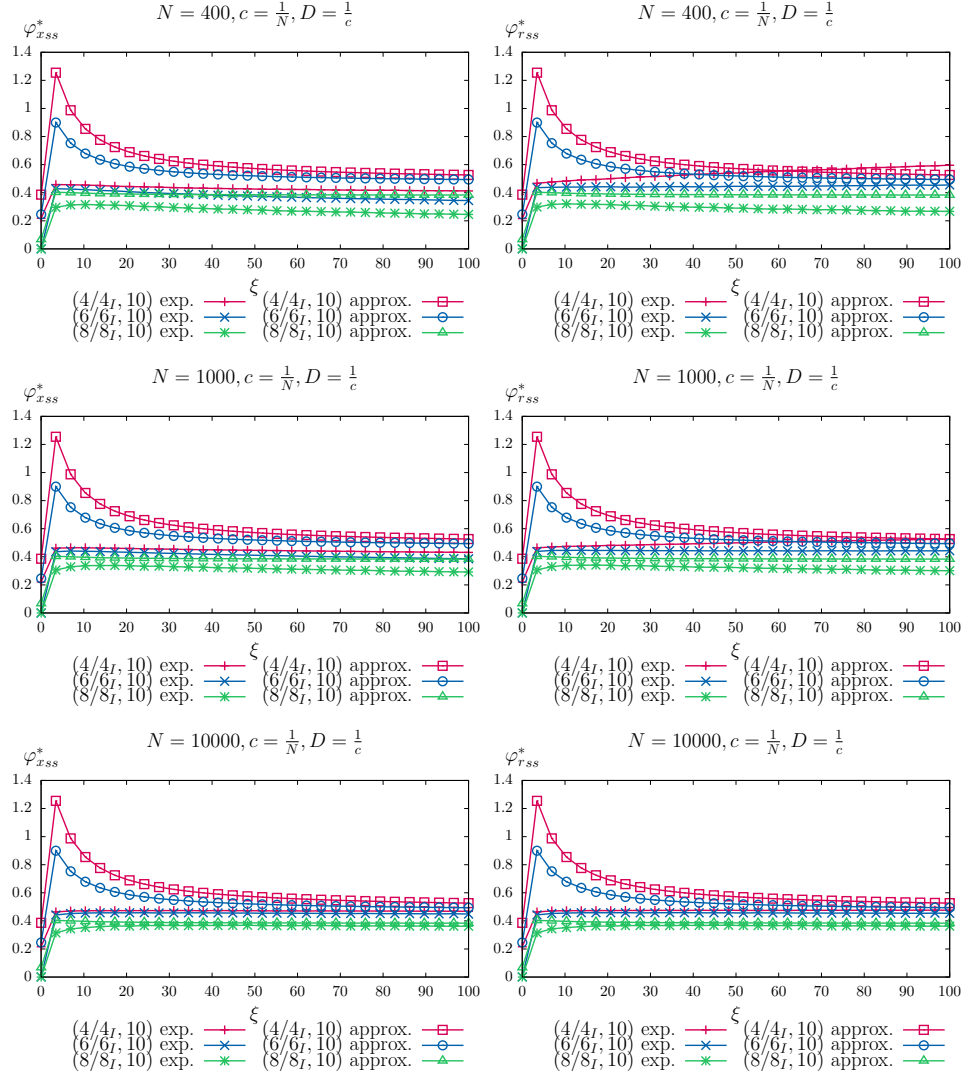


Figure 22: Steady state closed-form approximation and real run comparison of the $(\mu/\mu_I, \lambda)$ -CSA-ES with repair by projection applied to the conically constrained problem for $c = \frac{1}{N}$, $D = \frac{1}{c}$, $\lambda = 10$, $\mu \in \{4, 6, 8\}$. (Part 2)

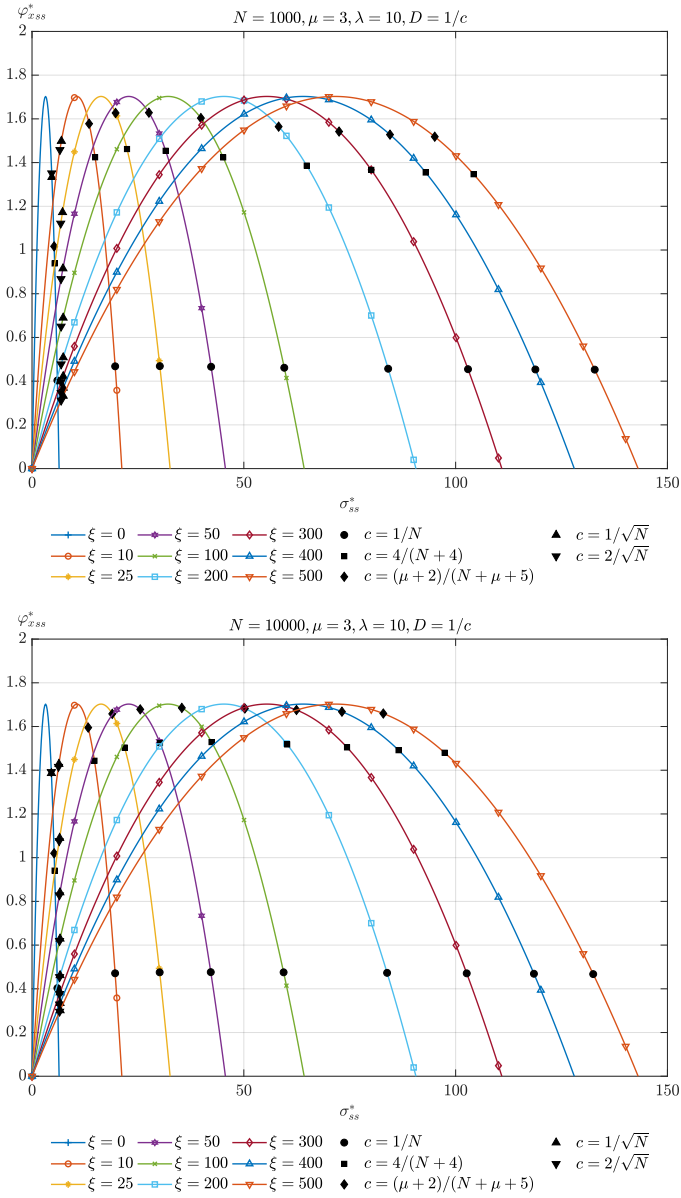


Figure 23: Influence of the cumulation parameter for the considered $(3/3_I, 10)$ -CSA-ES for $N = 1000$ (top) and $N = 10000$ (bottom). For different values of ξ , the steady state progress rate is plotted against the steady state normalized mutation strength. The filled black markers indicate the steady state values achieved by different c parameter settings with $D = 1/c$.

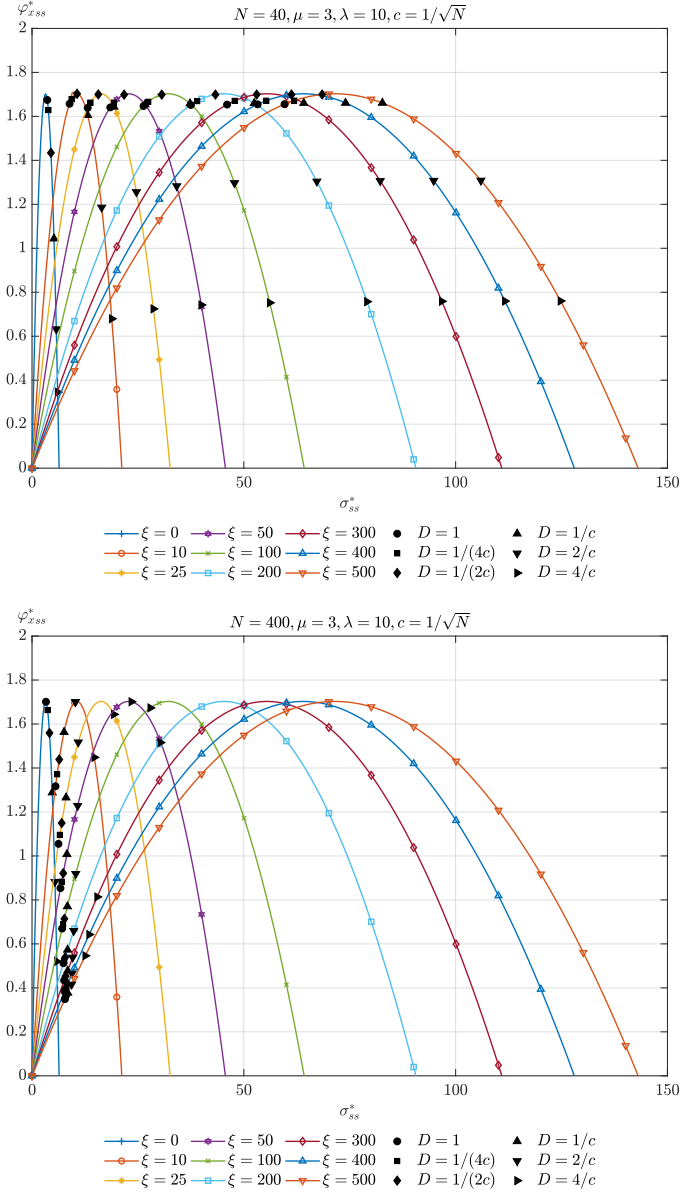


Figure 24: Influence of the damping parameter for the considered $(3/3_I, 10)$ -CSA-ES for $N = 40$ (top) and $N = 400$ (bottom). For different values of ξ , the steady state progress rate is plotted against the steady state normalized mutation strength. The filled black markers indicate the steady state values achieved by different D parameter settings with $c = 1/\sqrt{N}$.

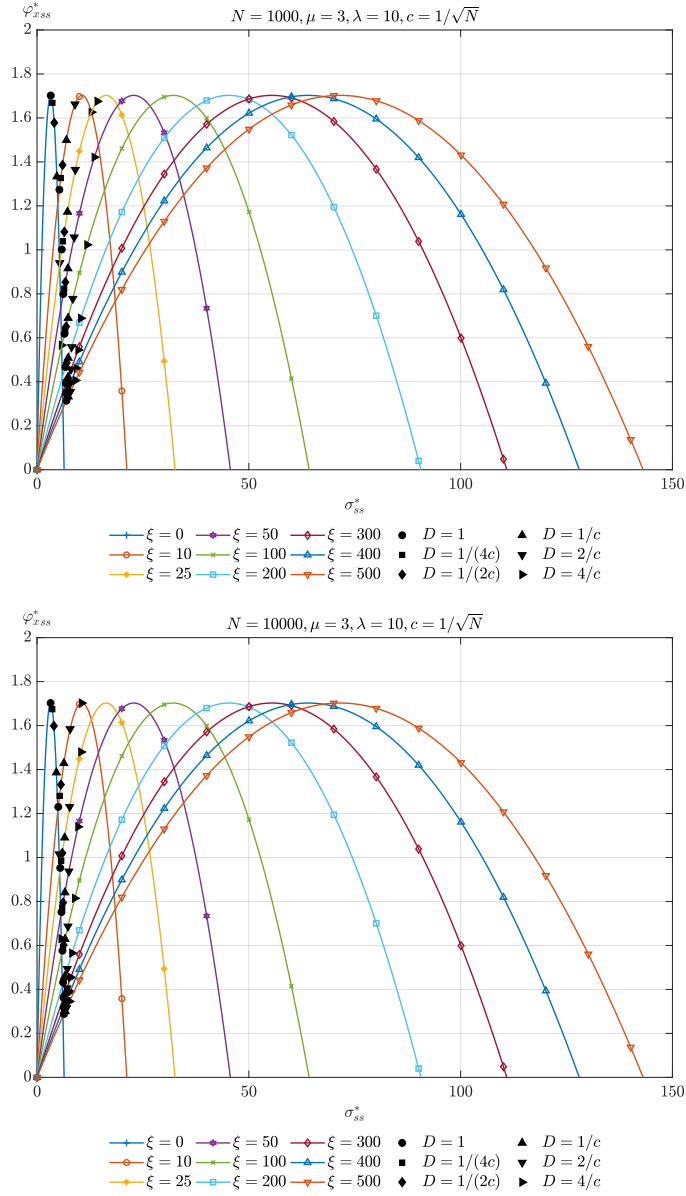


Figure 25: Influence of the damping parameter for the considered $(3/3_I, 10)$ -CSA-ES for $N = 1000$ (top) and $N = 10000$ (bottom). For different values of ξ , the steady state progress rate is plotted against the steady state normalized mutation strength. The filled black markers indicate the steady state values achieved by different D parameter settings with $c = 1/\sqrt{N}$.

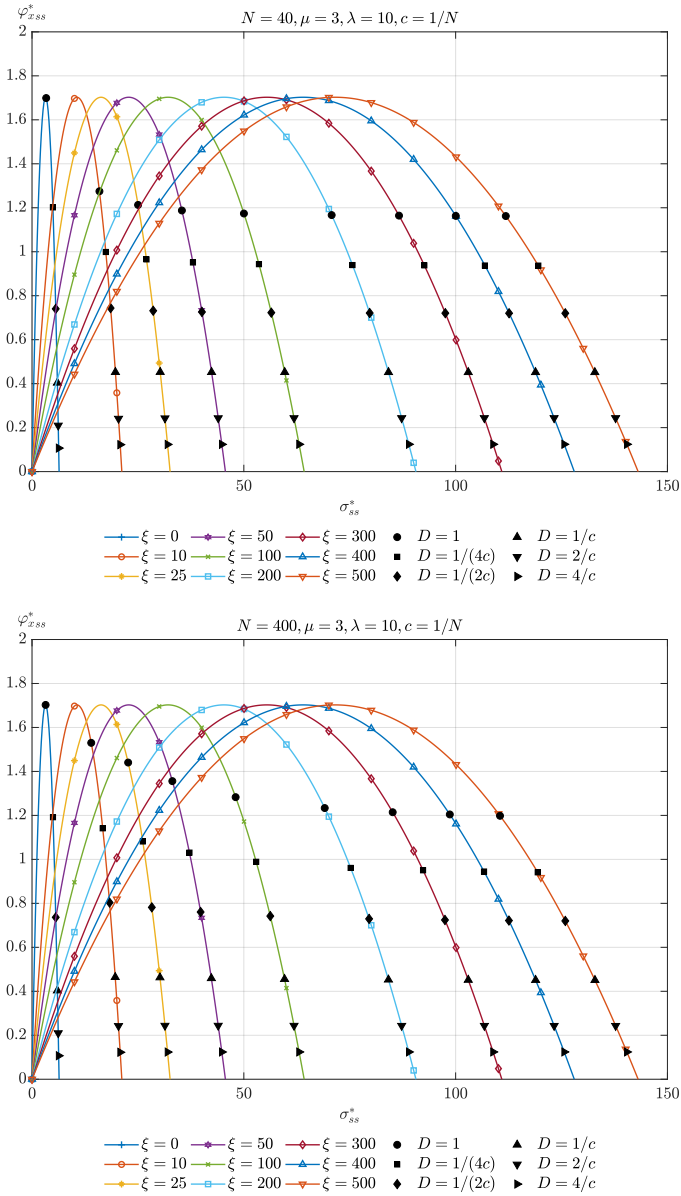


Figure 26: Influence of the damping parameter for the considered $(3/3_I, 10)$ -CSA-ES for $N = 40$ (top) and $N = 400$ (bottom). For different values of ξ , the steady state progress rate is plotted against the steady state normalized mutation strength. The filled black markers indicate the steady state values achieved by different D parameter settings with $c = 1/N$.

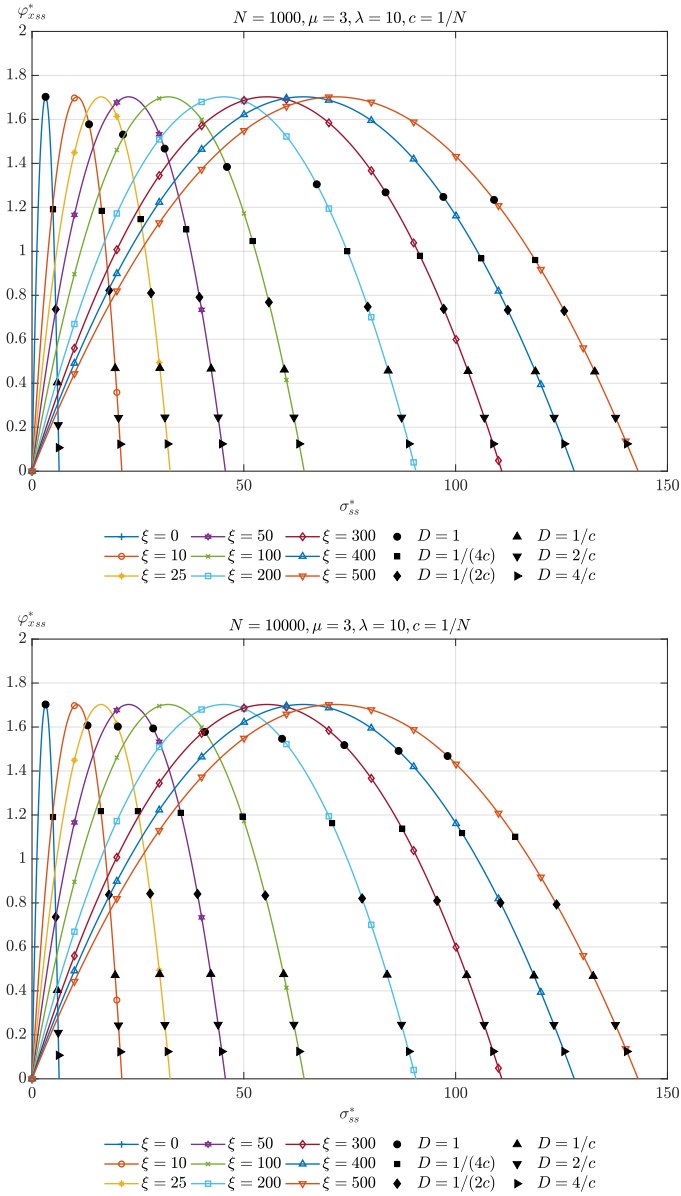


Figure 27: Influence of the damping parameter for the considered $(3/3_I, 10)$ -CSA-ES for $N = 1000$ (top) and $N = 10000$ (bottom). For different values of ξ , the steady state progress rate is plotted against the steady state normalized mutation strength. The filled black markers indicate the steady state values achieved by different D parameter settings with $c = 1/N$.