Steady State Analysis of a Multi-Recombinative Meta-ES on a Conically Constrained Problem with Comparison to σSA and CSA

Patrick Spettel
Vorarlberg University of Applied Sciences, Research Center PPE
Dornbirn, Austria
patrick.spettel@fhv.at

Hans-Georg Beyer
Vorarlberg University of Applied Sciences, Research Center PPE
Dornbirn, Austria
hans-georg.beyer@fhv.at

Michael Hellwig
Vorarlberg University of Applied Sciences, Research Center PPE
Dornbirn, Austria
michael.hellwig@fhv.at

ABSTRACT

This paper concerns the theoretical analysis of a multi-recombinative meta-ES with repair by projection applied to a conically constrained problem. Using theoretical results for the mean value dynamics and steady state considerations of the inner ES, approximate closed-form expressions for the mean value dynamics and the steady state behavior of the outer ES are derived. The approximation quality is shown by comparison with real meta-ES runs using isolation periods larger than one. The theoretical results are compared to known theoretical results of the multi-recombinative ES with σ-Self-Adaptation and Cumulative Step-Size adaptation. It is shown that the meta-ES achieves the largest steady state progress for the considered problem at the cost of twice the function evaluations compared to the other variants.

CCS CONCEPTS

• **Theory of computation → Optimization with randomized search heuristics**; **Bio-inspired optimization**; **Theory of randomized search heuristics**.

KEYWORDS

Meta-Evolution Strategy, Constrained Optimization, Theoretical Analysis

ACM Reference Format:
Patrick Spettel, Hans-Georg Beyer, and Michael Hellwig. 2019. Steady State Analysis of a Multi-Recombinative Meta-ES on a Conically Constrained Problem with Comparison to σSA and CSA. In *Foundations of Genetic Algorithms XV (FOGA '19), August 27–29, 2019, Potsdam, Germany*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3299904.3340306

1 INTRODUCTION

Adaptive parameter control is important for the effectiveness of Evolution Strategies (ESs). Besides the 1/5th rule [15], σ-Self-Adaption (σSA) [15], and Cumulative Step-Size Adaptation (CSA) [13], the meta-ES is an alternative approach for adapting the mutation strength throughout the evolution. Specifically, meta-ESs are nested ESs that control the mutation strength through their hierarchical organization [12, 15].

Meta-ESs represent an alternative approach for mutation strength adaptation. Theoretical investigations on the Sharp Ridge function revealed its effectiveness and the influence of certain strategy parameters on the meta-ES behavior [9]. In that respective work, it has been shown that meta-ESs are capable of avoiding premature convergence. Since the Sharp Ridge problem shares similarities with simple constrained problems, those results indicate that meta-ESs can be a suitable option for constrained problems as well. Therefore, by extension of the results in [11], this work provides a theoretical analysis of a meta-ES variant with multi-recombinative inner ES in the context of a conically constrained problem. On the other hand, the present paper compares the derived performance results of the meta-ES to those of two other multi-recombinative ES variants that apply different mutation strength control mechanisms, namely σSA-ES and CSA-ES, respectively.

Theoretical analyses of ESs on constraint problems were done for simple linear and conical constraints.

ESs applied to a linear objective function with a single linear constraint have been analyzed in prior work. The main motivation for a linear optimization problem with (a single) linear constraint(s) is to model a real-world situation for the problem of premature convergence. For the case that the descent direction of the objective function's gradient points toward the hyperplane spanned by the linear constraint, the ESs naturally move in the direction of the boundary. In order to be effective on such problems, the ES has to avoid stagnation. In [2], resampling has been considered. Repair by projection has been analyzed in [1]. Reflection and truncation have been considered in [10], and those approaches have been compared to projection and resampling in that work. It has been shown for the particular problem under consideration that repair by projection can avoid premature convergence even for small constraint angles. In contrast, other repair approaches (especially resampling) exhibit premature convergence. Further, Lagrangian based approaches for linearly constrained problems have been investigated in [5–7].

As a next step after considering linear constraints, a conically constrained problem has been proposed in [4], where the objective function is linear and the constraint is a cone. The idea here is to model a situation where the optimum lies on a "corner" of the feasibility region, which aggravates the problem of possible premature convergence. A simplification of this problem is to treat the case that the gradient of the linear objective function coincides with the

FOGA '19, August 27–29, 2019, Potsdam, Germany

ACM ISBN 978-1-4503-6254-2/19/08…$15.00
https://doi.org/10.1145/3299904.3340306

cone axis. That results in a linear optimization problem constrained by a cone such that the optimum is at the cone's apex. It has been presented in [3] and simplifies theoretical analyses (in comparison to the more general conical problem). The $(1, \lambda)$-CSA-ES with discarding infeasible offspring has been analyzed for this problem in [3], analogously, the $(\mu/\mu_I, \lambda)$-CSA-ES has been analyzed in [14]. As a further step, repair by projection has been considered: The $(1, \lambda)$-σSA-ES has been analyzed in [16], the $(\mu/\mu_I, \lambda)$-σSA-ES in [17], the $(\mu/\mu_I, \lambda)$-CSA-ES in [18], and the $[1, 2(1, \lambda)^1]$-meta-ES[1] in [11].

The goal of this work is the extension of the $[1, 2(1, \lambda)^\gamma]$-meta-ES analysis with $\gamma = 1$ as presented in [11] to the $[1, 2(\mu/\mu_I, \lambda)^\gamma]$-meta-ES with repair by projection applied to the conically constrained problem with the objective function gradient along the cone's axis. It presents a further step toward a more thorough theoretical understanding of the working principles of meta-ESs applied to constrained optimization problems. Expressions are derived for describing the algorithm's dynamics and the algorithm's steady state behavior. The results are compared to σSA and CSA. The comparison shows that the steady state progress rate achieved by the meta-ES is larger than that of the σSA and CSA. However, one has to note that the meta-ES runs two inner ESs in parallel and therefore uses twice the amount of function evaluations than the σSA and CSA variants compared.

The work is organized as follows: In Section 2, the problem is introduced and the considered meta-ES is presented and explained. This is followed by the theoretical analysis of the algorithm in Section 3: Section 3.1 briefly recaps results that have already been derived in [11]. Based on those results, Section 3.2 analyzes the multi-recombinative $[1, 2(\mu/\mu_I, \lambda)^\gamma]$-meta-ES in this work. After that, the derived results are compared to the σSA and CSA variants for the mutation strength control in Section 4. Finally, the work is concluded in Section 5.

Contribution: Based on prior results from [17, 18], steady state expressions for a $[1, 2(\mu/\mu_I, \lambda)^\gamma]$-meta-ES are derived and compared to those of σSA-ES and CSA-ES. A closed-form expression for the steady state normalized mutation strength is derived (Equation (37)). The ES exhibits a cyclic behavior around this value. It is shown that it can be bracketed (Equation (54)) with the corresponding interval for the steady state normalized progress (Equation (55)). Those derived expressions are compared to simulations of real meta-ES runs (Figures 6 and 7) and to the theoretical results of σSA and CSA (Figures 8 to 10) derived in prior work.

[1]The notation $[1, 2(\mu/\mu_I, \lambda)^\gamma]$-meta-ES is used throughout this paper. It denotes a meta-ES that runs two inner $(\mu/\mu_I, \lambda)$-ESs (μ parental individuals, λ offspring, and intermediate recombination) in parallel for γ generations (isolation period γ) with different strategy parameters. After the isolation period, the outer ES selects the population and the strategy parameters of the better (w.r.t fitness) inner ES for the next outer ES iteration.

2 PROBLEM AND ALGORITHM

The problem under consideration can be stated as

$$\text{minimize } F(\mathbf{x}) = x_1 \tag{1}$$

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

$$\text{and } x_1 \geq 0, \tag{3}$$

where $\mathbf{x} \in \mathbb{R}^N$ and $\xi > 0$. The constraints (2) and (3) define a second-order cone and the gradient of the objective function (1) coincides with the cone's axis. It is important to note that since the cone axis coincides with the first dimension, the constraint in the $(N-1)$-dimensional orthogonal subspace is symmetric and represents a hypersphere. Consequently, a point in the search space can be uniquely described by the distance x from 0 in x_1-direction and the distance r from the cone axis in the $(N-1)$-dimensional space, resulting in what is referred to as the $(x, r)^T$-space. Since in the following isotropic mutations are considered, this further allows rotating the coordinate system such that a point in the search space $(\tilde{x}, \tilde{r}, 0, \ldots, 0)^T$ can be represented as $(\tilde{x}, \tilde{r})^T$.

Figure 1 visualizes the problem in the $(x, r)^T$-space. The feasible region is shaded. Its limiting equation $r = x/\sqrt{\xi}$ follows directly from inequality (2). The optimum $\mathbf{x}_{\text{opt}}$ is at the apex of the cone. The shown offspring individual $\tilde{\mathbf{x}}$ (with parental individual $\mathbf{x}$ and mutation $\tilde{\sigma}\mathbf{z}$) is infeasible and consequently repaired by projection onto the cone boundary at $(q, q_r)^T$ along the dashed projection line. The vectors $\mathbf{e_r}$, $\mathbf{e_c}$, and $\mathbf{e}_1$ denote unit vectors in direction of the r-axis, the cone boundary, and the x-axis, respectively. Infeasible offspring are repaired by projection onto the cone boundary. The projection can be stated as

$$\text{project}(\tilde{\mathbf{x}}_{\text{infeas}}) := \begin{cases} (\mathbf{e}_c^T \tilde{\mathbf{x}}_{\text{infeas}})\mathbf{e}_c & \text{if } \mathbf{e}_c^T \tilde{\mathbf{x}}_{\text{infeas}} > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where

$$\mathbf{e}_c = \frac{1}{\sqrt{1 + 1/\xi}} \left(\mathbf{e}_1 + \frac{\mathbf{e}_r}{\sqrt{\xi}} \right). \tag{5}$$

By insertion of the unit vectors

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

and

$$\mathbf{e}_r = \left(0, \frac{\tilde{x}_2}{||\tilde{\mathbf{r}}||}, \ldots, \frac{\tilde{x}_N}{||\tilde{\mathbf{r}}||} \right)^T,$$

one can derive

$$(\mathbf{e}_c^T \tilde{\mathbf{x}}_{\text{infeas}})\mathbf{e}_c = \frac{1}{1 + 1/\xi} \left(\tilde{x}_1 + \frac{||\tilde{\mathbf{r}}||}{\sqrt{\xi}} \right) \left(1, \frac{\tilde{x}_2}{\sqrt{\xi}||\tilde{\mathbf{r}}||}, \ldots, \frac{\tilde{x}_N}{\sqrt{\xi}||\tilde{\mathbf{r}}||} \right)^T \tag{6}$$

and

$$\mathbf{e}_c^T \tilde{\mathbf{x}}_{\text{infeas}} = \sqrt{\frac{1}{1 + 1/\xi}} \left(\tilde{x}_1 + \frac{||\tilde{\mathbf{r}}||}{\sqrt{\xi}} \right), \tag{7}$$

where $\tilde{x}_1 = (\tilde{\mathbf{x}}_{\text{infeas}})_1, \tilde{x}_2 = (\tilde{\mathbf{x}}_{\text{infeas}})_2, \ldots, \tilde{x}_N = (\tilde{\mathbf{x}}_{\text{infeas}})_N$, and $||\tilde{\mathbf{r}}|| = \sqrt{\sum_{k=2}^{N} (\tilde{\mathbf{x}}_{\text{infeas}})_k^2}$.

The considered meta-ES algorithm, which is applied to the presented conically constrained problem, is shown in Algorithm 1. In the beginning, the initial parental individual $\mathbf{x}^{(0)}$, the initial

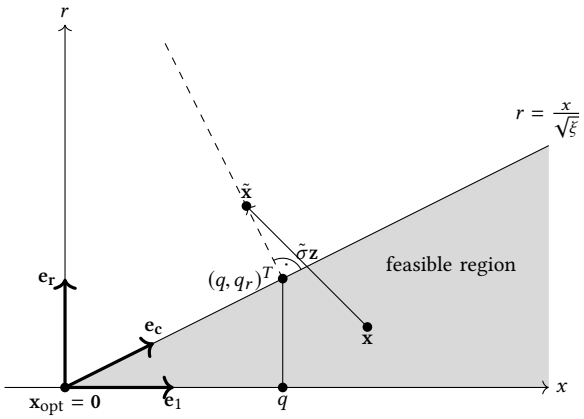

Figure 1: The conically constrained optimization problem in N dimensions shown in the $(x, r)^T$-space. The feasible region is shaded. The indicated offspring individual $\tilde{\mathbf{x}}$ (with parental individual $\mathbf{x}$ and mutation $\tilde{\sigma}\mathbf{z}$) is infeasible and therefore projected onto the cone boundary at $(q, q_r)^T$ along the dashed projection line.

mutation strength $\sigma^{(0)}$, the number of offspring λ, the parental population size μ, the length of the isolation period γ, and the isolation period counter t are initialized in Line 1 and Line 2, respectively. After that, the outer ES loop (Lines 3 to 11) runs as long as the termination criterion is not met: The increased mutation strength σ_1 and the decreased mutation strength σ_2 are computed in Line 4 and Line 5, respectively. Then, the inner ES is called for both mutation strength settings in Line 6 and Line 7. After that, the parental individual and the mutation strength for the next generation are set to the results of the inner ES that resulted in the better objective function value (Line 8 and Line 9)[2]. The update of the isolation period counter (Line 10) ends one iteration of the loop.

The inner ES is a $(\mu/\mu_I, \lambda)$-ES with repair by projection and constant mutation strength. Its pseudo-code is shown in Algorithm 2. After initialization in Line 1 and Line 2, the generational loop (Lines 3 to 14) is entered. It is run for the given isolation period γ. In every generation, λ offspring are created in Lines 4 to 10. For the creation of each offspring, a mutation vector is sampled from an isotropic multivariate normal distribution, scaled with σ, and added to the parental individual in Line 5. In case that the generated offspring is infeasible, it is repaired by projection (Lines 6 to 8). After that, its fitness is determined in Line 9. After the procreation step, the offspring are sorted in Line 11 in ascending order. The parental individual for the next generation is set to the centroid of the μ best offspring in Line 12. The update of the generation counter (Line 13) ends one iteration of the generational loop.

As can be seen, the inner ES uses constant σ and allows the outer ES to control the mutation strength. The outer ES always chooses the setting with the better result of the two inner mutation strength choices.

[2]The order statistic notation is used, where more generally "$m; \lambda$" denotes the m-th best (w.r.t. fitness) out of λ values.

Algorithm 1 Pseudo-code of the $[1, 2(\mu/\mu_I, \lambda)^\gamma]$-meta-ES.

1: Initialization of $\mathbf{x}^{(0)}, \sigma^{(0)}, \lambda, \mu, \gamma, \alpha$, where $\mu \geq 1, \lambda > \mu, \sigma^{(0)} > 0$, $\gamma > 0, \alpha > 1$ must hold
2: $t \leftarrow 0$
3: **repeat**
4: $\quad \sigma_1 \leftarrow \sigma^{(t)}\alpha$
5: $\quad \sigma_2 \leftarrow \sigma^{(t)}/\alpha$
6: $\quad (\mathbf{x}_1, F(\mathbf{x}_1)) \leftarrow \text{ES}(\mathbf{x}^{(t)}, \sigma_1, \lambda, \mu, \gamma)$ ▷ see Algorithm 2
7: $\quad (\mathbf{x}_2, F(\mathbf{x}_2)) \leftarrow \text{ES}(\mathbf{x}^{(t)}, \sigma_2, \lambda, \mu, \gamma)$ ▷ see Algorithm 2
8: $\quad \mathbf{x}^{t+1} \leftarrow \mathbf{x}_{1;2}$ ▷ $\mathbf{x}_{1;2}$: value of the fitter inner ES result
9: $\quad \sigma^{t+1} \leftarrow \sigma_{1;2}$ ▷ $\sigma_{1;2}$: value of the fitter inner ES result
10: $\quad t \leftarrow t + 1$
11: **until** termination criterion

Algorithm 2 Pseudo-code of the $(\mu/\mu_I, \lambda)$-ES with repair by projection and constant mutation strength.

1: **function** ES($\mathbf{x}^{(0)}$, σ, λ, μ, γ)
2: $\quad g \leftarrow 0$
3: $\quad$ **while** $g < \gamma$ **do**
4: $\quad\quad$ **for** $l \leftarrow 1$ **to** λ **do**
5: $\quad\quad\quad \tilde{\mathbf{x}}_l \leftarrow \mathbf{x}^{(g)} + \sigma\mathcal{N}(\mathbf{0}, \mathbf{I})$ ▷ Mutation
6: $\quad\quad\quad$ **if not** isFeasible($\tilde{\mathbf{x}}_l$) **then** ▷ see (2) and (3)
7: $\quad\quad\quad\quad \tilde{\mathbf{x}}_l \leftarrow$ project($\tilde{\mathbf{x}}_l$) ▷ Repair (see (4) and (5))
8: $\quad\quad\quad$ **end if**
9: $\quad\quad\quad \tilde{f}_l \leftarrow F(\tilde{\mathbf{x}}_l)$ ▷ Evaluation
10: $\quad\quad$ **end for**
11: $\quad\quad$ Sort offspring according to $\tilde{f}_l$ in ascending order
12: $\quad\quad \mathbf{x}^{(g+1)} \leftarrow \frac{1}{\mu}\sum_{m=1}^{\mu} \tilde{\mathbf{x}}_{m;\lambda}$ ▷ Selection and recombination
13: $\quad\quad g \leftarrow g + 1$
14: $\quad$ **end while**
15: $\quad$ **return** $\left(\mathbf{x}^{(\gamma)}, F(\mathbf{x}^{(\gamma)})\right)$
16: **end function**

3 ANALYSIS

For analyzing the dynamics and the steady state behavior[3] of the meta-ES, the inner ES must be analyzed as an initial step. The goal is to have closed-form expressions for predicting the objective function value of the inner ES after γ inner generations. This allows theoretically reasoning about the outer ES. The following subsections first treat the inner ES (Section 3.1) and then the outer ES (Section 3.2).

3.1 Inner Dynamics

Since the inner ES is a $(\mu/\mu_I, \lambda)$-ES with repair by projection, the analysis performed in [17] builds the starting point of the meta-ES investigations. The results are briefly recapped and explained here for the reader, but it is important to note that this section does not present new contributions. The main results of [17] are closed-form expressions considering the $(\mu/\mu_I, \lambda)$-σSA-ES. As a step toward those results, progress rates for a given parental state have been derived for the asymptotic limit case $N \to \infty$. The assumption

[3]Steady state in this context characterizes the ES behavior in which there is (in expectation) a constant descent rate of the outer ES toward the optimizer. In this situation, the expected normalized mutation strength is constant.

$N \to \infty$ was necessary in order to derive closed-form expressions. Even though $N \to \infty$ was assumed, those expressions can still serve as qualitative approximations for finite values of N. The progress rate approximations are used in this work for theoretically modeling the behavior of the inner ES. Only the mean value dynamics are considered (the fluctuations are assumed to be negligible). This allows stating a system of mean value evolution equations for x and r from a generation g to the next generation $g+1$ as

$$x^{(g+1)} = x^{(g)} \left(1 - \frac{\varphi_x^{(g)^*}}{N}\right) \tag{8}$$

$$r^{(g+1)} = r^{(g)} \left(1 - \frac{\varphi_r^{(g)^*}}{N}\right). \tag{9}$$

The quantities $\varphi_x^{(g)^*}$ and $\varphi_r^{(g)^*}$ are normalized progress rates in x and r direction, see Equation (13) and Equation (16) below. For their derivation, the cases of being far from the cone boundary and being in the vicinity of the cone boundary have been distinguished. The main idea for that has been that for $N \to \infty$ the probability of generating infeasible offspring tends to 1 in the vicinity of the cone boundary, and it tends to 0 being far from the cone boundary. Hence, the cases are denoted using subscripts named "infeas" and "feas", respectively. The feasibility probability has been derived as [17]

$$P_{\text{feas}} := P_{\text{feas}}(x^{(g)}, r^{(g)}, \sigma) \simeq \Phi\left[\frac{1}{\sigma}\left(\frac{x^{(g)}}{\sqrt{\xi}} - \bar{r}\right)\right], \tag{10}$$

with

$$\bar{r} = r^{(g)} \sqrt{1 + \frac{{\sigma^{(g)}}^{*2}}{N}\left(1 - \frac{1}{N}\right)}. \tag{11}$$

Here, Φ denotes the cumulative distribution function of the standard normal distribution, $\bar{r}$ is the expected value of a normal approximation of the distance from the cone axis r, and ${\sigma^{(g)}}^*$ is the normalized mutation strength ${\sigma^{(g)}}^* := \frac{N}{r^{(g)}}\sigma$. Equation (10) is accurate for sufficiently large ξ and is an approximation for smaller values of ξ. Note that since σ is constant for the inner ES, the superscript for denoting the generation for the normalized mutation strength is due to $r^{(g)}$. Using the feasibility probability, the derivations for the feasible and infeasible cases of the normalized progress rate derivations can be combined yielding [17]

$$\varphi_x^{*(g)} := \varphi_x^*(x^{(g)}, r^{(g)}, \sigma) \tag{12}$$

$$= \frac{N}{x^{(g)}} \mathrm{E}[x^{(g)} - x^{(g+1)} \mid x^{(g)}, r^{(g)}, \sigma] \tag{13}$$

$$\approx P_{\text{feas}} \frac{r^{(g)}}{x^{(g)}} {\sigma^{(g)}}^* c_{\mu/\mu,\lambda} + [1 - P_{\text{feas}}] \varphi_{x\,\text{infeas}}^* \tag{14}$$

and

$$\varphi_r^{*(g)} := \varphi_r^*(x^{(g)}, r^{(g)}, \sigma) \tag{15}$$

$$= \frac{N}{r^{(g)}} \mathrm{E}[r^{(g)} - r^{(g+1)} \mid x^{(g)}, r^{(g)}, \sigma] \tag{16}$$

$$\approx P_{\text{feas}} N \left(1 - \sqrt{1 + \frac{{\sigma^{(g)}}^{*2}}{\mu N}}\right) + [1 - P_{\text{feas}}] N \left(1 - \frac{x^{(g)}}{\sqrt{\xi} r^{(g)}}\left(1 - \frac{\varphi_{x\,\text{infeas}}^{*(g)}}{N}\right)\sqrt{\frac{1 + \frac{{\sigma^{(g)}}^{*2}}{\mu N}}{1 + \frac{{\sigma^{(g)}}^{*2}}{N}}}\right) \tag{17}$$

with

$$\varphi_{x\,\text{infeas}}^{*(g)} \approx \frac{N}{1+\xi}\left(1 - \frac{\sqrt{\xi} r^{(g)}}{x^{(g)}}\sqrt{1 + \frac{{\sigma^{(g)}}^{*2}}{N}}\right) + \frac{\sqrt{\xi}}{1+\xi}\frac{\sqrt{\xi} r^{(g)}}{x^{(g)}} {\sigma^{(g)}}^* c_{\mu/\mu,\lambda} \sqrt{1 + \frac{1}{\xi}\frac{1 + \frac{{\sigma^{(g)}}^{*2}}{2N}}{1 + \frac{{\sigma^{(g)}}^{*2}}{N}}}. \tag{18}$$

$c_{\mu/\mu,\lambda} = e_{\mu,\lambda}^{1,0}$ is one of the so-called generalized progress coefficients [8, Eq. (5.112), p. 172]. Their definition reads

$$e_{\mu,\lambda}^{\alpha,\beta} := \frac{\lambda - \mu}{(\sqrt{2\pi})^{\alpha+1}} \binom{\lambda}{\mu} \int_{t=-\infty}^{t=\infty} t^{\beta} e^{-\frac{\alpha+1}{2}t^2} \times [\Phi(t)]^{\lambda-\mu-1} [1 - \Phi(t)]^{\mu-\alpha} \, \mathrm{d}t. \tag{19}$$

The main assumptions used for the derivation of the recapped expressions are sufficiently large values of the dimensionality N and of the cone parameter ξ, respectively. Consequently, the approximations get better the larger the values of ξ and N get (with $\xi \ll N$). Nevertheless, they can be used for qualitative predictions for smaller values of ξ and N.

For justifying the use of the approximations (Equation (14) and Equation (17)), the system stated as Equation (8) and Equation (9) is iterated using them and the predicted dynamics are compared to mean value dynamics of real inner ES (Algorithm 2) runs. The results are shown in Figure 2. For the parameter ξ, different values have been chosen and are shown in comparison. The runs have been initialized with $\mathbf{x}^{(0)} = (\sqrt{N}, 10^{-3}, 0, \ldots, 0)^T$, $N = 400$, and constant $\sigma = 1$. This results in dynamics that show the behavior starting near the cone axis and moving toward the cone boundary transitioning into a steady state. As it turns out, in the steady state the ES moves in the vicinity of the cone boundary with a finite normalized mutation strength and hence for $N \to \infty$ the case that the created offspring are infeasible overwhelms. The observed deviations for $\xi = 0.1$ stem from the above-mentioned assumptions in the derivations for the progress rate expressions. For the parameters considered, $\xi = 0.1$ is not large enough such that the approximations are as accurate as for the larger values of ξ considered. In Figure 3, the experiments are started on the cone boundary, i.e., $\mathbf{x}^{(0)} = (\sqrt{N}, \sqrt{N}/\sqrt{\xi}, 0, \ldots, 0)^T$. In this way, the comparison starts almost in the steady state. The chosen parameters are again $\sigma = 1$, $N = 400$, and different values of ξ are considered and compared. It is important to note that this initialization can result in deviations for the first number of

generations because $r = x/\sqrt{\xi}$ only holds in the steady state for the case $\mu = 1$. For $\mu > 1$, one would need to know the steady state normalized mutation strength to initialize further in the steady state (according to Equation (22)).

Assuming the ES to be sufficiently near to the cone boundary (and therefore in the steady state), simplifications are possible because $P_{\text{feas}} = 0$ can be assumed. One gets using (18)

$$\begin{aligned}\varphi_x^{*(g)} &\approx \varphi_{x\,\text{infeas}}^{*(g)} \\ &\approx \frac{N}{1+\xi}\left(1 - \frac{\sqrt{\xi}r^{(g)}}{x^{(g)}}\sqrt{1+\frac{\sigma^{(g)*2}}{N}}\right) + \frac{\sigma^{(g)*}c_{\mu/\mu,\lambda}}{\sqrt{1+\xi}}\frac{\sqrt{\xi}r^{(g)}}{x^{(g)}}\end{aligned} \tag{20}$$

by making use of

$$\frac{1}{\xi}\frac{1+\frac{\sigma^{*(g)^2}}{2N}}{1+\frac{\sigma^{*(g)^2}}{N}} \simeq \frac{1}{\xi} \tag{21}$$

that holds for $\sigma^{*(g)^2} \ll N$ and sufficiently large N. Since the steady state behavior is assumed, the steady state expression

$$\left(\frac{\sqrt{\xi}r}{x}\right)_{ss} = \sqrt{\frac{1+\frac{\sigma_{ss}^{*\,2}}{\mu N}}{1+\frac{\sigma_{ss}^{*\,2}}{N}}}. \tag{22}$$

derived in [17] can be used for predicting the mean value dynamics of the inner ES in the steady state. Insertion of Equation (22) into Equation (20) results in

$$\begin{aligned}\varphi_x^{*(g)}(\sigma^{(g)*}) \approx\ & \frac{N}{1+\xi}\left(1 - \sqrt{1+\frac{\sigma^{(g)*2}}{\mu N}}\right) \\ & + \frac{\sigma^{(g)*}c_{\mu/\mu,\lambda}}{\sqrt{1+\xi}}\sqrt{\frac{1+\frac{\sigma^{(g)*2}}{\mu N}}{1+\frac{\sigma^{(g)*2}}{N}}}\end{aligned} \tag{23}$$

and with the insertion of Equation (23) into Equation (17) as $\varphi_{x\,\text{infeas}}^{*(g)}$ one gets

$$\varphi_r^{*(g)}(\sigma^{(g)*}) \approx \varphi_x^{*(g)}(\sigma^{(g)*}) \tag{24}$$

for the steady state case.

3.2 Outer Dynamics

Now that suitable approximate expressions for the inner ES have been stated, the goal is to make use of them for the analysis of the $[1, 2(\mu/\mu_I, \lambda)^\gamma]$-meta-ES that controls the mutation strength (Algorithm 1). Figure 4 shows an example of its mean value dynamics with $\gamma = 1$ in comparison to the theoretical iteration of Equation (8) and Equation (9) using Equation (23) and Equation (24).

It is important to note that Equation (14) and Equation (17) are difficult to treat further for the steady state considerations because of the weighted sum involving the feasibility probability. Therefore, for the analysis of the outer ES, it is again assumed that $P_{\text{feas}} = 0$, i.e., it is not the aim to handle the transient phase theoretically. This is justified since one observes in simulations that indeed the ES transitions into this steady state behavior. The mutation strength control of Algorithm 1 can be modeled by considering the difference of the objective function values returned by both inner ESs. Since the objective function only regards the first component of an individual's parameter vector, the x progress rate can be used for modeling the change of σ from one isolation period to the next. Notice that t is used to indicate the isolation period of the outer ES and g is used to indicate the generation of the inner ES. Hence, a value of x before isolation at isolation period t and initial inner generation counter g_0 corresponds to $x^{(t)} = x^{(g_0)}$. The corresponding value after isolation corresponds to $x^{(t+1)} = x^{(g_0+\gamma)}$. Because for the analysis it is assumed that the inner ES operates in the steady state, $\varphi_x^{*(g)}(\sigma^{(g)*})$ is used that only depends on $\sigma^{(g)*}$ but not on the position in the parameter space. For small values of γ, $\varphi_x^{(t)*} \approx constant$ is assumed. With this assumption, Equation (8) can be recursively applied starting at $x^{(g_0)}$ to compute $x^{(g_0+\gamma)}$ with $\varphi_x^{(g)*} = \varphi_x^{(t)*}$ for all the γ inner iterations with generation numbers $g \in \{g_0, g_0+1, \ldots, g_0+\gamma-1\}$. One obtains

$$F_1^{(t+1)} := F(\mathbf{x}_1^{(t+1)}) \approx x^{(t)}\left(1 - \frac{\varphi_x^{(t)*}(\alpha\sigma^{(t)*})}{N}\right)^\gamma \tag{25}$$

$$\simeq x^{(t)}\left(1 - \gamma\frac{\varphi_x^{(t)*}(\alpha\sigma^{(t)*})}{N}\right) \tag{26}$$

$$F_2^{(t+1)} := F(\mathbf{x}_2^{(t+1)}) \approx x^{(t)}\left(1 - \frac{\varphi_x^{(t)*}(\sigma^{(t)*}/\alpha)}{N}\right)^\gamma \tag{27}$$

$$\simeq x^{(t)}\left(1 - \gamma\frac{\varphi_x^{(t)*}(\sigma^{(t)*}/\alpha)}{N}\right). \tag{28}$$

With the assumption of small γ as stated above, the steps from Equation (25) to Equation (26) and Equation (27) to Equation (28) follow by the use of the Taylor expansion

$$\left(1 - \frac{\varphi_x^{(t)*}(\sigma^{(t)*})}{N}\right)^\gamma \simeq \left(1 - \gamma\frac{\varphi_x^{(t)*}(\sigma^{(t)*})}{N}\right)$$

with cut-off after the linear term under the assumption

$$\varphi_x^{(t)*}(\sigma^{(t)*})/N \ll 1.$$

The mutation strength update can be modeled by the difference $D^{(t+1)} := F_1^{(t+1)} - F_2^{(t+1)}$ as

$$\sigma^{(t+1)} = \begin{cases}\alpha\sigma^{(t)} & \text{if } D^{(t+1)} < 0, \\ \sigma^{(t)}/\alpha & \text{if } D^{(t+1)} > 0, \\ \sigma^{(t)} & \text{otherwise.}\end{cases} \tag{29}$$

Using Equation (23) with the assumption that

$$\sqrt{\frac{1+\frac{\sigma^{(g)*2}}{\mu N}}{1+\frac{\sigma^{(g)*2}}{N}}} \simeq 1 \tag{30}$$

for $N \to \infty$, one obtains

$$\begin{aligned}D^{(t+1)} = \frac{\gamma x^{(t)}}{1+\xi}\Bigg(&\sqrt{1+\frac{\alpha^2\sigma^{(t)*2}}{\mu N}} - \frac{\alpha\sigma^{(t)*}c_{\mu/\mu,\lambda}\sqrt{1+\xi}}{N} \\ &- \sqrt{1+\frac{\sigma^{(t)*2}}{\alpha^2\mu N}} + \frac{\sigma^{(t)*}c_{\mu/\mu,\lambda}\sqrt{1+\xi}}{\alpha N}\Bigg).\end{aligned} \tag{31}$$

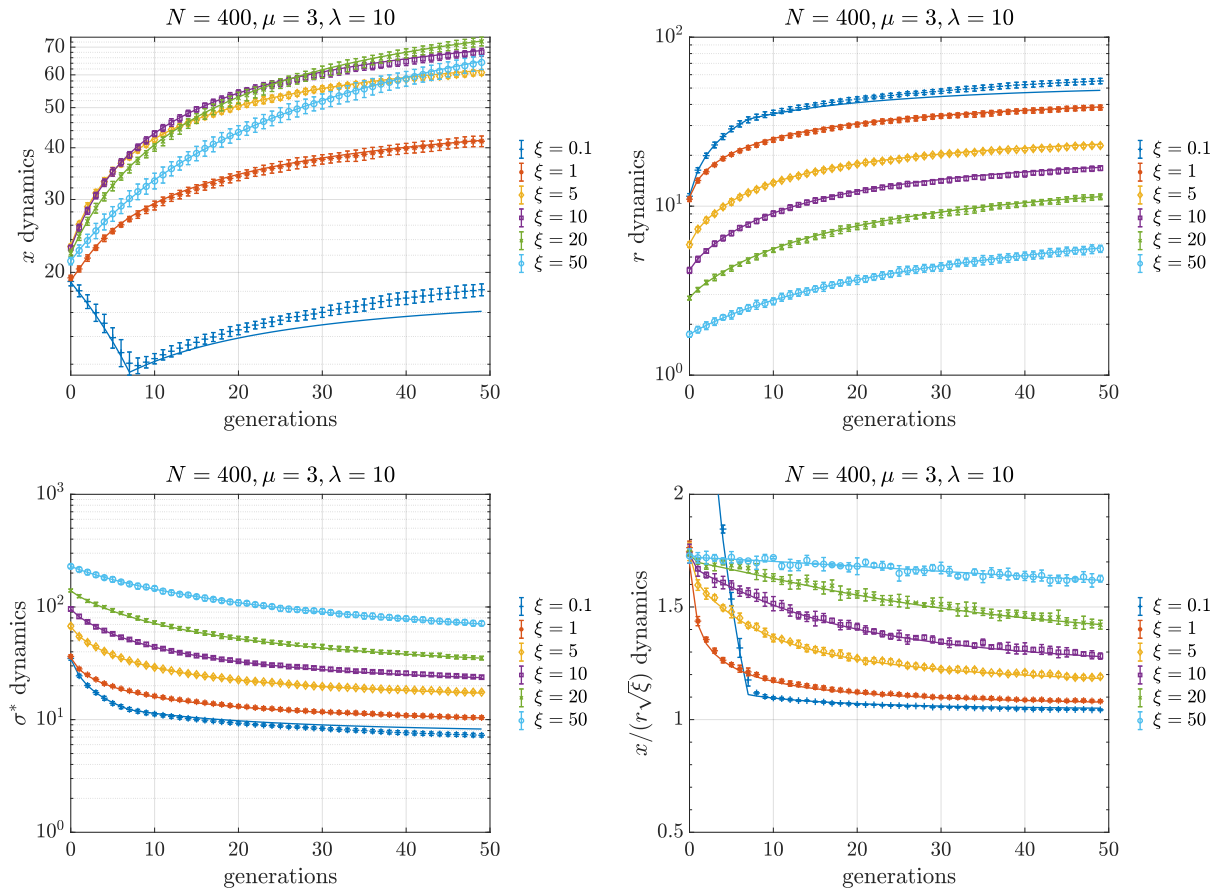

Figure 2: Theoretical predictions (solid lines) compared with the mean value dynamics for the inner $(3/3_I, 10)$-ES (mean value data points are indicated by the markers with corresponding error bars showing the standard deviation). The dimension $N = 400$ has been considered and the constant mutation strength $\sigma = 1$ has been used. The data points for the experiments show averages of 20 independent ES runs. The predictions have been computed by iterating Equation (8) and Equation (9) using Equation (14) and Equation (17), respectively. The initial parental individual has been set such that $x^{(0)} = \sqrt{N}$ and $r^{(0)} = 10^{-3}$, i.e., the initial parental individual is near the cone axis.

Equation (31) can be simplified for the asymptotic limit case $N \to \infty$ by Taylor expansion of $\sqrt{1 + \frac{\sigma^{(t)*2}}{\mu N}}$ around zero and cutoff after the linear term, i.e., $\sqrt{1 + \frac{\sigma^{(t)*2}}{\mu N}} \approx 1 + \frac{\sigma^{(t)*2}}{2\mu N}$ for $\sigma^{*(t)2} \ll N$. With this and additional rewriting, one gets

$$D^{(t+1)} = \frac{\gamma c_{\mu/\mu,\lambda} \sigma^{(t)*} x^{(t)} (\alpha^2 - 1)}{\alpha N \sqrt{1+\xi}} \underbrace{\left(\frac{\sigma^{(t)*}}{2\mu\sqrt{1+\xi} c_{\mu/\mu,\lambda}} \left(\alpha + \frac{1}{\alpha} \right) - 1 \right)}_{=:\Delta(\sigma^{(t)*})}. \tag{32}$$

Because $\alpha > 1$ is required, only the sign of $\Delta(\sigma^{(t)*})$ is relevant for the update of σ, i.e.,

$$\sigma^{(t+1)} = \sigma^{(t)} \alpha^{-\operatorname{sign}\left(\Delta(\sigma^{(t)*})\right)}. \tag{33}$$

Notice that the σ update is independent of the isolation period parameter γ, since the steady state has been assumed. The normalized variant of Equation (33) can be expressed as

$$\frac{\sigma^{(t+1)*} r^{(t+1)}}{N} = \frac{\sigma^{(t)*} r^{(t)}}{N} \alpha^{-\operatorname{sign}\left(\Delta(\sigma^{(t)*})\right)} \tag{34}$$

$$\sigma^{(t+1)*} = \frac{\sigma^{(t)*} r^{(t)}}{r^{(t+1)}} \alpha^{-\operatorname{sign}\left(\Delta(\sigma^{(t)*})\right)}. \tag{35}$$

Equation (35) can be rewritten using Equation (24) and the same argument leading to Equation (26) for the r case using Equation (9) resulting in

$$\sigma^{(t+1)*} = \sigma^{(t)*} \frac{\alpha^{-\operatorname{sign}\left(\Delta(\sigma^{(t)*})\right)}}{1 - \frac{\gamma}{N} \varphi_x^* \left(\sigma^{(t)*} \alpha^{-\operatorname{sign}(\Delta(\sigma^{(t)*}))} \right)}, \tag{36}$$

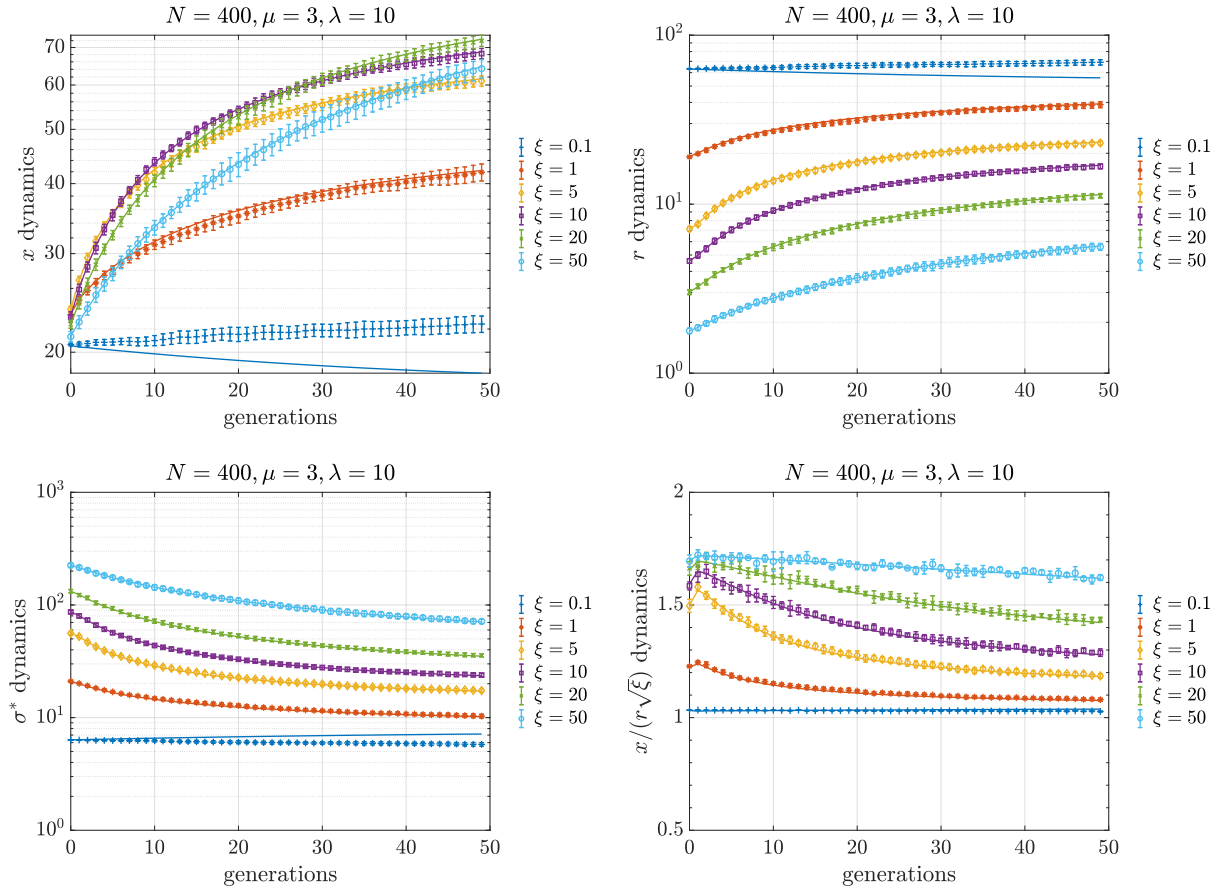

Figure 3: Theoretical predictions (solid lines) compared with the mean value dynamics for the inner $(3/3_I, 10)$-ES (mean value data points are indicated by the markers with corresponding error bars showing the standard deviation). The dimension $N = 400$ has been considered and the constant mutation strength $\sigma = 1$ has been used. The data points for the experiments show averages of 20 independent ES runs. The predictions have been computed by iterating Equation (8) and Equation (9) using Equation (14) and Equation (17), respectively. The initial parental individual has been set such that $x^{(0)} = \sqrt{N}$ and $r^{(0)} = \sqrt{N}/\sqrt{\xi}$, i.e., the initial parental individual is on the cone boundary.

which is an iterative mapping of the form

$$\sigma^{(t+1)^*} = f_\sigma(\sigma^{(t)^*}; \mu, \lambda, \gamma, \alpha, N).$$

Studying the influence of the parameters on Equation (36) is a rather difficult task because the dynamics of such a recurrence equation can have stable fixed points, limit cycles, and chaotic behaviors.

A discontinuity can be observed (cf. Figure 5) that represents a change from increasing σ to decreasing σ. It can be computed by setting the discriminant function $\Delta(\sigma^*)$ from Equation (32) to zero, i.e., $\Delta(\sigma_0^*) = 0$. Straightforward rewriting leads to

$$\sigma_0^* = 2\mu\sqrt{1+\xi}c_{\mu/\mu,\lambda}\frac{\alpha}{\alpha^2+1}. \tag{37}$$

The reason that the discontinuity shown in the left plot of Figure 5 does not coincide completely with σ_0^* for $N = 1000$ is due to the simplifications made for $N \to \infty$ such as (30).

As one can see, σ_0^* depends on the parameters ξ, α, μ, and λ. Influenced by α, different behaviors can be observed. If the fixed point σ_f^* of Equation (36) is at the point $\sigma_f^* = \sigma_0^*$, it is not a stable one as can be seen in the figure. The other fixed point at $\sigma_f^* = 0$ can also not be a stable fixed point because $\frac{\mathrm{d}f_\sigma(\sigma^*)}{\mathrm{d}\sigma^*} > 1$ (as one can conclude from the figure).

Let us study the case of continuous mutation strength reduction $\Delta(\sigma_f^*) > 0$ next. Note that this case holds for $\sigma_f^* > \sigma_0^*$. With $\Delta(\sigma_f^*) > 0$, consideration of the fixed point $\sigma^{(t+1)^*} = \sigma^{(t)^*}$ in Equation (36) results in

$$1 = \frac{1}{\alpha}\frac{1}{1-\frac{\gamma}{N}\varphi_x^*\left(\sigma_f^*/\alpha\right)}. \tag{38}$$

Using Equation (23) with the assumption stated in (30) for $N \to \infty$, one can write

$$1 = \frac{1}{\alpha}\frac{1}{1-\frac{\gamma}{N}\left[\frac{N}{1+\xi}\left(1-\sqrt{1+\frac{\sigma_f^{*2}}{\alpha^2\mu N}}\right)+\frac{\sigma_f^* c_{\mu/\mu,\lambda}}{\alpha\sqrt{1+\xi}}\right]} \tag{39}$$

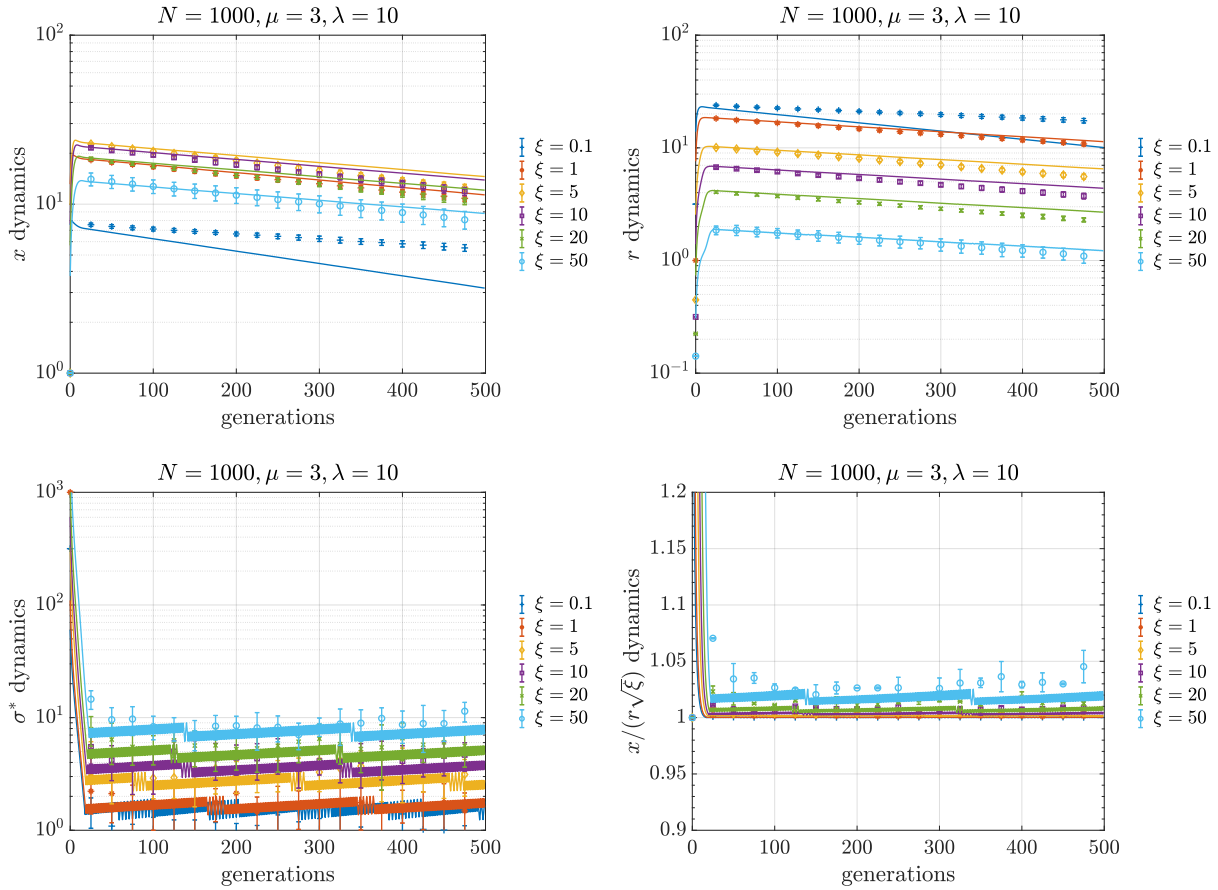

Figure 4: Theoretical predictions (solid lines) compared with the mean value dynamics for the $[1, 2(3/3_I, \lambda)^1]$-ES (mean value data points are indicated by the markers with corresponding error bars showing the standard deviation) for the dimension $N = 1000$ and $\alpha = 1.2$. The data points for the experiments show averages of 20 independent ES runs. The predictions have been computed by iterating Equation (8) and Equation (9) using Equation (23) and Equation (24), respectively. For the experimental runs and the theoretical iteration, σ has been updated as shown in Algorithm 1. The initial parental individual has been set such that $x^{(0)} = 1$ and $r^{(0)} = 1/\sqrt{\xi}$, i.e., the initial parental individual is on the cone boundary. The initial mutation strength has been set to $\sigma^{(0)} = 1$.

for Equation (38). Equation (39) can be simplified for the asymptotic limit case $N \to \infty$ by Taylor expansion of $\sqrt{1 + \frac{\sigma_f^{*2}}{\mu N}}$ around zero and cutoff after the linear term, i.e., $\sqrt{1 + \frac{\sigma_f^{*2}}{\mu N}} \approx 1 + \frac{\sigma_f^{*2}}{2\mu N}$ for $\sigma_f^{*2} \ll N$ yielding

$$1 = \frac{1}{\alpha} \frac{1}{1 + \frac{\gamma \sigma_f^{*2}}{2\alpha^2 \mu N(1+\xi)} - \frac{\gamma \sigma_f^* c_{\mu/\mu,\lambda}}{N\alpha\sqrt{1+\xi}}} \tag{40}$$

after straightforward simplifications. As one can see, Equation (40) represents a quadratic equation in σ_f^*. Rewriting it into standard form yields

$$\sigma_f^{*2} - 2\alpha\mu\sqrt{1+\xi}c_{\mu/\mu,\lambda}\sigma_f^* + \frac{2\alpha(\alpha-1)\mu N(1+\xi)}{\gamma} = 0. \tag{41}$$

Solving it results in

$$\sigma_{f\pm}^* = \alpha\mu\sqrt{1+\xi}c_{\mu/\mu,\lambda}\left(1 \pm \sqrt{1 - \frac{2N}{\mu c_{\mu/\mu,\lambda}^2 \gamma}\frac{\alpha-1}{\alpha}}\right) \tag{42}$$

after rewriting. The solution σ_{f-}^* is of relevance depending on the square root expression in Equation (42). If $\sigma_{f-}^* < \sigma_0^*$, it does not satisfy the assumption of continuous mutation strength decrease $\Delta(\sigma_f^*) > 0$. Hence, for that case, the fixed point is

$$\sigma_{f+}^* = \alpha\mu\sqrt{1+\xi}c_{\mu/\mu,\lambda}\left(1 + \sqrt{1 - \frac{2N}{\mu c_{\mu/\mu,\lambda}^2 \gamma}\frac{\alpha-1}{\alpha}}\right). \tag{43}$$

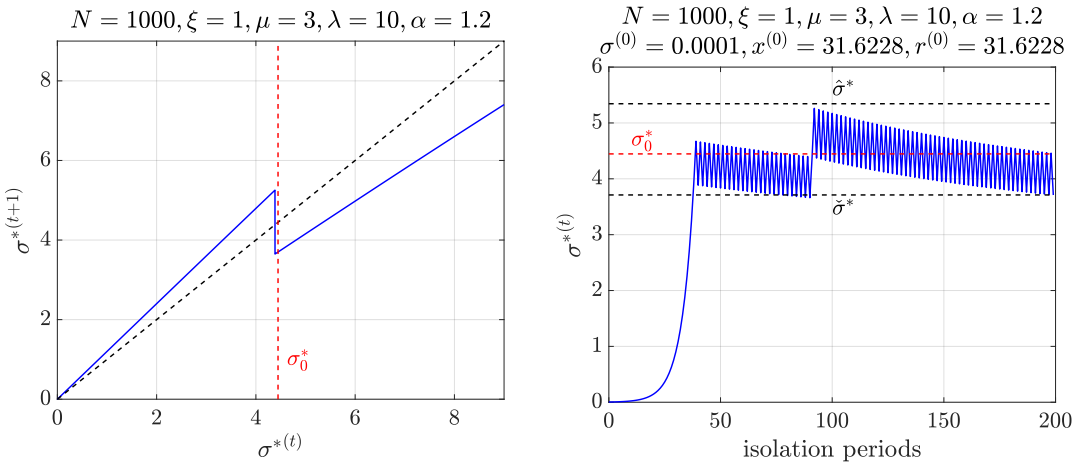

Figure 5: Visualization of the σ^*-behavior of the $[1, 2(3/3_I, 10)^1]$-meta-ES for $N = 1000$, $\xi = 1$, and $\alpha = 1.2$. The left plot shows an unstable fixed point of Equation (36), which results in a limit cycle as shown in the right plot. The discontinuity point σ_0^* is shown as the dashed red line and calculated using Equation (37). The upper and lower bounds of the normalized mutation strength bracketing the cycle have been computed using Equation (54) and are shown in the right plot as dashed black lines. The dashed black line in the left plot shows $\sigma^{*(t+1)} = \sigma^{*(t)}$. The initial parental individual for the creation of the right subplot has been set such that $x^{(0)} = \sqrt{N}$ and $r^{(0)} = \sqrt{N}/\sqrt{\xi}$.

For the case that $\sigma^*_{f_-} > \sigma_0^*$, the positive and negative solutions derived as Equation (42) are relevant. The stable attractor is the one that satisfies $\frac{\mathrm{d}f_\sigma(\sigma^*)}{\mathrm{d}\sigma^*} < 1$.

If the expression under the square root in Equation (43) is negative, no real solution exists. Computing the zero of that expression under the square root yields the limiting case for the existence of a stable fixed point with $\alpha_0 > 0$:

$$1 - \frac{2N}{\mu c^2_{\mu/\mu,\lambda}\gamma}\frac{\alpha_0 - 1}{\alpha_0} \geq 0 \tag{44}$$

$$\alpha_0 \mu c^2_{\mu/\mu,\lambda}\gamma \geq 2N\alpha_0 - 2N \tag{45}$$

$$\alpha_0 \leq \frac{2N}{2N - \mu c^2_{\mu/\mu,\lambda}\gamma}. \tag{46}$$

Consequently, for $\alpha < \alpha_0$ it is true that $\Delta(\sigma^*_f) > 0$ and the mutation strength is continuously decreased as can be seen from Equation (33). For the opposite case of $\alpha > \alpha_0$, the dynamics of σ^* fluctuate around σ_0^*.

The lower and upper bounds for the observed σ^* values in the limit cycle for the case $\alpha > \alpha_0$ can be computed. They can be calculated as the left-sided and right-sided limits of Equation (36). By introducing the abbreviations $\varphi^*_+ := \varphi^*_x(\alpha\sigma_0^*)$ and $\varphi^*_- := \varphi^*_x(\sigma_0^*/\alpha)$, one can write

$$\hat{\sigma}^* := \lim_{\sigma^* \to \sigma_0^* - 0} \frac{\alpha\sigma^*}{1 - \frac{\gamma}{N}\varphi^*_x(\alpha\sigma^*)} \tag{47}$$

$$= \frac{\alpha\sigma_0^*}{1 - \frac{\gamma}{N}\varphi^*_+} \tag{48}$$

$$= 2\mu\sqrt{1+\xi}c_{\mu/\mu,\lambda}\frac{\alpha^2}{\alpha^2+1}\frac{1}{1 - \frac{\gamma}{N}\varphi^*_+} \tag{49}$$

and

$$\check{\sigma}^* := \lim_{\sigma^* \to \sigma_0^* + 0} \frac{\sigma^*/\alpha}{1 - \frac{\gamma}{N}\varphi^*_x(\sigma^*/\alpha)} \tag{50}$$

$$= \frac{\sigma_0^*/\alpha}{1 - \frac{\gamma}{N}\varphi^*_-} \tag{51}$$

$$= 2\mu\sqrt{1+\xi}c_{\mu/\mu,\lambda}\frac{1}{\alpha^2+1}\frac{1}{1 - \frac{\gamma}{N}\varphi^*_-}. \tag{52}$$

For φ^*_+ and φ^*_-, one derives

$$\frac{2\mu c^2_{\mu/\mu,\lambda}\alpha^2}{(\alpha^2+1)^2} \tag{53}$$

by consideration of Equation (23) with the assumption stated in (30) for $N \to \infty$ and a Taylor expansion of $\sqrt{1 + \frac{\sigma^{(g)*2}}{\mu N}}$ around zero and cutoff after the linear term for $\sigma^{(g)*2} \ll N$. These derivations lead to the interval for the expected normalized mutation strength

$$\sigma^* \in [\check{\sigma}^*, \hat{\sigma}^*] = [1, \alpha^2] \cdot \frac{2\mu\sqrt{1+\xi}c_{\mu/\mu,\lambda}}{\alpha^2+1}\frac{1}{1 - \frac{\gamma}{N}\varphi^*_+} \tag{54}$$

and the interval for the expected progress of the meta-ES

$$\tilde{\varphi}^*_x \in [\varphi^*_x(\check{\sigma}^*), \underbrace{\varphi^*_x(\sigma^*_{\mathrm{opt}})}_{=:\varphi^*_{x\,\mathrm{opt}}}] = \left[\left(\frac{2\alpha}{\alpha^2+1}\right), 1\right] \cdot \frac{\mu c^2_{\mu/\mu,\lambda}}{2}, \tag{55}$$

where $\frac{\gamma 2\mu c^2_{\mu/\mu,\lambda}\alpha^2}{N(\alpha^2+1)^2} \ll 1$ has been assumed for the calculation of $\varphi^*_x(\check{\sigma}^*)$ and $\varphi^*_{x\,\mathrm{opt}} = \frac{\mu c^2_{\mu/\mu,\lambda}}{2}$ has been derived as follows: Again, Equation (23) can be simplified using (30) for $N \to \infty$ and a Taylor

expansion of $\sqrt{1+\frac{\sigma^{(g)*2}}{\mu N}}$ around zero and cutoff after the linear term for $\sigma^{(g)*2} \ll N$ leading to

$$\varphi_x^{*(g)}(\sigma^{*(g)}) \simeq \frac{\sigma^{*(g)} c_{\mu/\mu,\lambda}}{\sqrt{1+\xi}} - \frac{\sigma^{*(g)2}}{2\mu(1+\xi)} \tag{56}$$

for the steady state case with $P_{\text{feas}} = 0$ and movement in the vicinity of the cone boundary. Calculating the derivative of Equation (56) results in

$$\frac{\mathrm{d}\,\varphi_x^{*(g)}(\sigma^{*(g)})}{\mathrm{d}\,\sigma^{*(g)}} = \frac{c_{\mu/\mu,\lambda}}{\sqrt{1+\xi}} - \frac{\sigma^{*(g)}}{\mu(1+\xi)}. \tag{57}$$

Setting Equation (57) to zero, one obtains

$$\sigma^*_{\text{opt}} = \mu c_{\mu/\mu,\lambda}\sqrt{1+\xi} \tag{58}$$

that maximizes φ_x^*. Back-insertion of Equation (58) into Equation (56) finally yields the maximal (optimal) progress as

$$\varphi^*_{x\,\text{opt}} = \frac{\mu c^2_{\mu/\mu,\lambda}}{2}. \tag{59}$$

Figures 6 and 7 show comparisons of the derived steady state expressions with real ES runs. One can see that except for $\xi = 0.1$, the normalized steady state mutation strength prediction comes near to the one determined from the experiments. The predicted progress is larger than the one observed in the experiments. However, for $\xi = 5$ and $\xi = 10$, the experimental results are in the theoretically determined bracket $[\varphi_x^*(\check{\sigma}^*), \varphi_x^*(\sigma^*_{\text{opt}})]$ for $\alpha \geq 1.2$.

4 COMPARISON WITH σSA AND CSA

The derived steady state normalized mutation strength and progress rate expressions for the $[1, 2(\mu/\mu_I, \lambda)^\gamma]$-meta-ES allow a theoretical comparison with other mutation strength control methods. Since the $(\mu/\mu_I, \lambda)$-ES applied to the conically constrained problem has already been theoretically analyzed with σSA [17] and CSA [18] as the mutation strength adaptation mechanism, the goal of this section is to compare the steady state behavior of the meta adaptation approach to σSA and CSA.

Figures 8 to 10 show the comparison of the theoretical steady state mutation strength and progress rate predictions for $\mu = 1$, $\mu = 2$, and $\mu = 3$, respectively. The parameter λ is set to 10 for all variants. The CSA parameters c and D are set to $c = 1/N$ and $D = 1/c$, respectively, the σSA parameter τ is set to $\tau = 1/\sqrt{2N}$, and the parameter α of the the meta-ES is set to $\alpha = 1.2$. The choice for the CSA cumulation parameter $c = 1/N$ has been used in the work that presents the theoretical analysis concerning CSA [18]. For σSA, the setting $\tau = 1/\sqrt{2N}$ has also been discussed as a recommended value in the work with the σSA theoretical analysis [17]. The choice of $\alpha = 1.2$ for the meta-ES stems from the observation that the predictions are near the real behavior in Figures 6 and 7. The lines for the meta-ES are determined by computing the steady state normalized mutation strength σ^*_{ss} using Equation (37). For σSA, [17, Eq. (59)] is used to compute σ^*_{ss}. And for CSA, [18, Eq. (118)] is used to compute σ^*_{ss}. The results are then inserted into Equation (23) using (30) for $N \to \infty$ for plotting the line showing $\varphi^*_{x\,ss}(\sigma^*_{ss})$.

Since it has been pointed out in the theoretical derivations for the meta, σSA, and CSA approaches that $N \to \infty$ has been used, the $N = 10000$ plots are the most accurate predictions of the real behavior. A further peculiarity in the derivations for all the approaches has been mentioned to be the necessity of assuming sufficiently large values of ξ in order to make the analyses tractable. This explains possible discrepancies for small ξ.

The plots show that the meta-ES achieves the largest steady state progress for all the values ξ considered in dimension 10000. Between the σSA and the meta-ES, the difference in the progress is about 0.1, whereas the CSA achieves less than half the progress of that of the meta-ES. Since the meta-ES runs two inner ESs in parallel, it uses twice the amount of function evaluations than the σSA-ES and CSA-ES variants compared.

Moreover, one observes that for larger N, the steady state progress for the range of ξ considered gets almost constant. For smaller N, it increases slightly with increasing values of ξ.

5 CONCLUSIONS AND OUTLOOK

This work extended the analysis of the $[1, 2(1, \lambda)^1]$-meta-ES with repair by projection applied to the conically constrained problem [11] to the $[1, 2(\mu/\mu_I, \lambda)^\gamma]$-meta-ES. The derived expressions show a similar prediction quality as the $\mu = 1$ case for the dynamics and the steady state considerations. The most important difference of the $\mu > 1$ in comparison to the $\mu = 1$ case is the steady state distance to the cone boundary. Whereas it is 0 in the $\mu = 1$ case, the ES in the $\mu > 1$ case evolves also along the cone boundary but at a certain distance from it. This has an important influence on the behavior of the meta-ES that makes the analysis for $\gamma = 1$ difficult, if not impossible. Therefore, the experimental results presented and compared with the steady state derivations have considered $\gamma > 1$. The reason for the difficulty is two-fold. First, treatment of only the infeasible case (being in the vicinity of the cone boundary) for the behavior of the inner ES led to tractable approximate expressions for the steady state analysis. For the experimental comparison with the theoretical results, this means that the experiments need to be started almost in the steady state, since the transient phase is not modeled for the theoretical considerations. However, the second difficulty is that for $\mu > 1$, the steady state distance from the cone boundary depends on the normalized steady state mutation strength. Therefore, with an isolation of $\gamma = 1$, the experiments cannot be initialized near enough the steady state, since the first inner iteration increases/decreases the mutation strength by α, which results in a situation in which the initial parental individual is no longer in the steady state. To overcome this, $\gamma > 1$ allows the inner ES to transition sufficiently into the steady state. It has been shown that with $\gamma = 30$ the theoretical predictions serve as usable approximations of the real behavior for larger values of N and ξ.

The comparison of the meta-ES with σSA and CSA led to an interesting finding. It turned out that the meta-ES achieves the largest steady state progress. However, in comparison to σSA and CSA it requires twice the number of function evaluations due to the two inner ESs running in parallel. An interesting topic for future work is to investigate whether the same observation also holds for the case that the ES simply discards infeasible offspring instead of repairing them by projection.

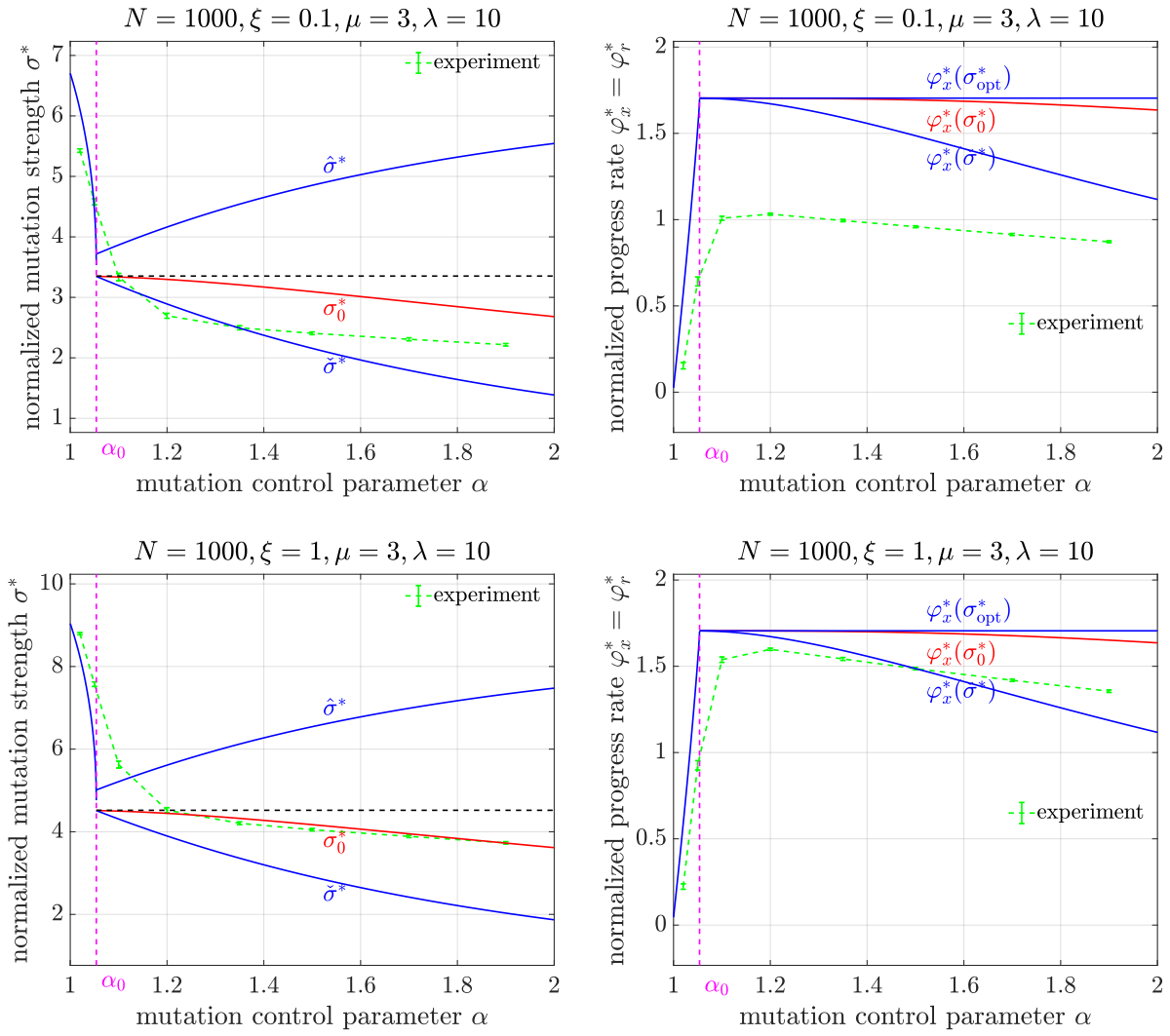

Figure 6: Comparison of the $[1, 2(3/3_I, 10)^{30}]$-meta-ES mutation strength and progress rate steady state theoretical results with experiments for different values of α. The lines with the error bars show the experimental results. Their data have been obtained by computing the mean and standard deviation of the last 60% of $2N$ isolation periods from 50 independent algorithm runs. The runs have been started in the vicinity of the cone boundary such that the steady state is reached after a relatively small number of isolation periods. The dashed horizontal lines in the left plots show σ_{opt}^* (Equation (58)). (Part 1/2)

ACKNOWLEDGMENTS

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

REFERENCES

[1] Dirk V. Arnold. 2011. Analysis of a Repair Mechanism for the $(1, \lambda)$-ES Applied to a Simple Constrained Problem. In *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*. ACM, ACM, New York, NY, USA, 853–860.

[2] Dirk V. Arnold. 2011. On the Behaviour of the $(1, \lambda)$-ES for a Simple Constrained Problem. In *Proceedings of the 11th Workshop Proceedings on Foundations of Genetic Algorithms*. ACM, ACM, New York, NY, USA, 15–24.

[3] Dirk V. Arnold. 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*. ACM, New York, NY, USA, 423–430.

[4] Dirk V. Arnold. 2014. On the Behaviour of the $(1, \lambda)$-ES for Conically Constrained Linear Problems. *Evolutionary Computation* 22, 3 (2014), 503–523.

[5] Dirk V. Arnold and Jeremy Porter. 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)*. ACM, New York, NY, USA, 249–256. https://doi.org/10.1145/2739480.2754813

[6] Asma Atamna, Anne Auger, and Nikolaus Hansen. 2016. Augmented Lagrangian Constraint Handling for CMA-ES - Case of a Single Linear Constraint. In *International Conference on Parallel Problem Solving from Nature*. Springer, Springer, Cham, 181–191.

[7] Asma Atamna, Anne Auger, and Nikolaus Hansen. 2017. Linearly Convergent Evolution Strategies via Augmented Lagrangian Constraint Handling. In *Proceedings of the 14th ACM/SIGEVO Conference on Foundations of Genetic Algorithms*. ACM, ACM, New York, NY, USA, 149–161.

[8] Hans-Georg Beyer. 2001. *The Theory of Evolution Strategies*. Springer, Berlin Heidelberg.

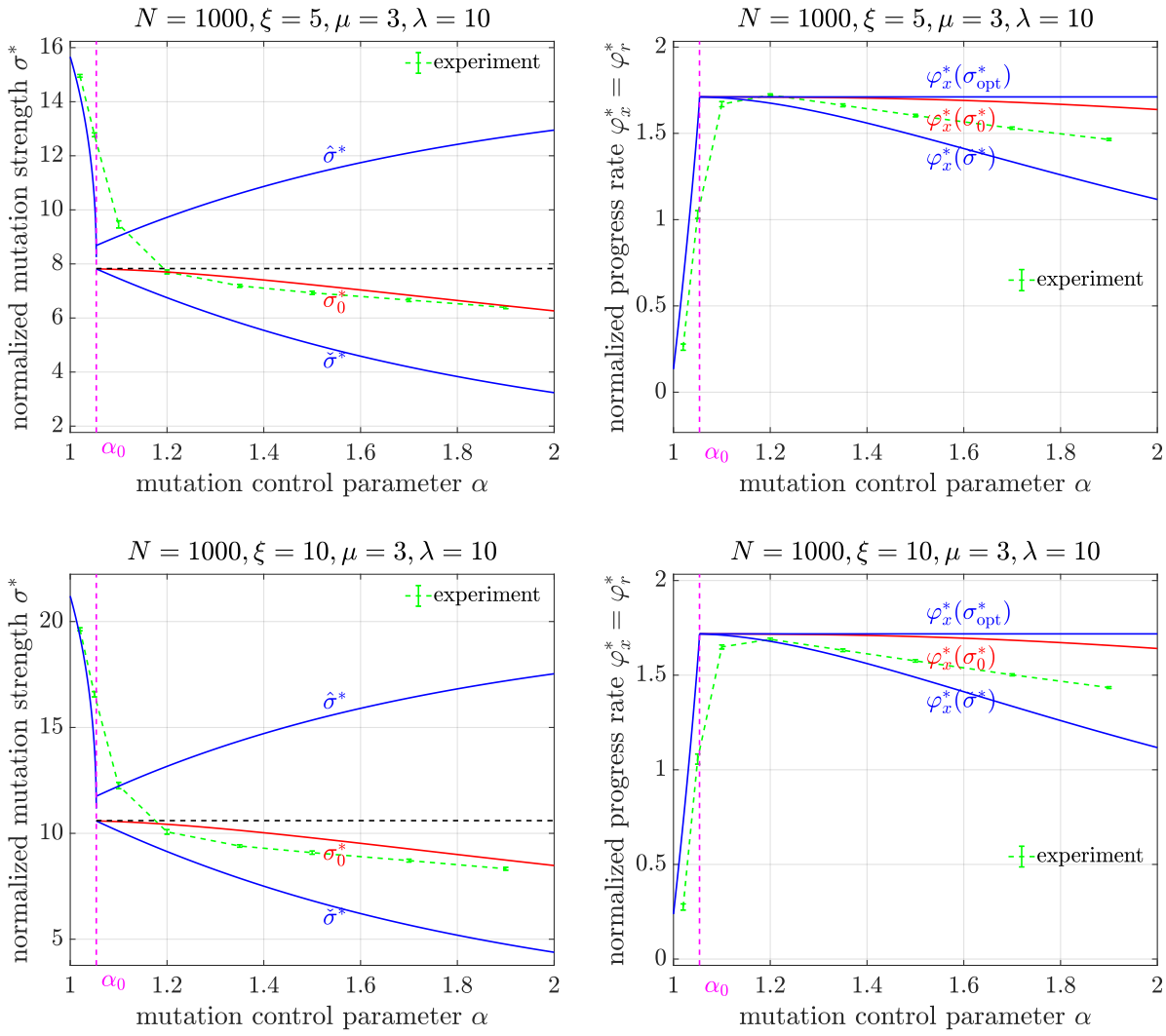

Figure 7: Comparison of the $[1, 2(3/3_I, 10)^{30}]$-meta-ES mutation strength and progress rate steady state theoretical results with experiments for different values of α. The lines with the error bars show the experimental results. Their data have been obtained by computing the mean and standard deviation of the last 60% of $2N$ isolation periods from 50 independent algorithm runs. The runs have been started in the vicinity of the cone boundary such that the steady state is reached after a relatively small number of isolation periods. The dashed horizontal lines in the left plots show σ_{opt}^* (Equation (58)). (Part 2/2)

[9] Hans-Georg Beyer and Michael Hellwig. 2012. Mutation Strength Control by Meta-ES on the Sharp Ridge. In *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation*. ACM, ACM, New York, NY, USA, 305–312.

[10] Michael Hellwig and Dirk V. Arnold. 2016. Comparison of Constraint-Handling Mechanisms for the (1, λ)-ES on a Simple Constrained Problem. *Evolutionary Computation* 24, 1 (2016), 1–23.

[11] Michael Hellwig and Hans-Georg Beyer. 2019. Analysis of a Meta-ES on a Conically Constrained Problem. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO '19)*. ACM, New York, NY, USA. Accepted. Available online: https://homepages.fhv.at/hgb/New-Papers/GECCO19_HB19.pdf (accessed Apr 17, 2019).

[12] Michael Herdy. 1992. Reproductive Isolation as Strategy Parameter in Hierarchically Organized Evolution Strategies. In *Parallel Problem Solving from Nature 2, PPSN-II, Brussels, Belgium, September 28-30, 1992*. North-Holland, Amsterdam, 209–220.

[13] Andreas Ostermeier, Andreas Gawelczyk, and Nikolaus Hansen. 1994. Step-size adaptation based on non-local use of selection information. In *Parallel Problem Solving from Nature — PPSN III*. Springer Berlin Heidelberg, Berlin, Heidelberg, 189–198.

[14] Jeremy Porter and Dirk V. Arnold. 2015. Analyzing the Behaviour of Multi-Recombinative Evolution Strategies Applied to a Conically Constrained Problem. In *Evolutionary Constrained Optimization*. Springer, New Delhi, 181–204.

[15] Ingo Rechenberg. 1994. *Evolutionsstrategie '94*. Werkstatt Bionik und Evolutionstechnik, Vol. 1. Frommann-Holzboog, Stuttgart.

[16] Patrick Spettel and Hans-Georg Beyer. 2018. Analysis of the (1, λ)-σ-Self-Adaptation Evolution Strategy with repair by projection applied to a conically constrained problem. *Theoretical Computer Science* in press (2018). https://doi.org/10.1016/j.tcs.2018.10.036 Available online: http://www.sciencedirect.com/science/article/pii/S0304397518306571 (accessed Dec 14, 2018).

[17] Patrick Spettel and Hans-Georg Beyer. 2018. Analysis of the (μ/μ_I, λ)-σ-Self-Adaptation Evolution Strategy with Repair by Projection Applied to a Conically

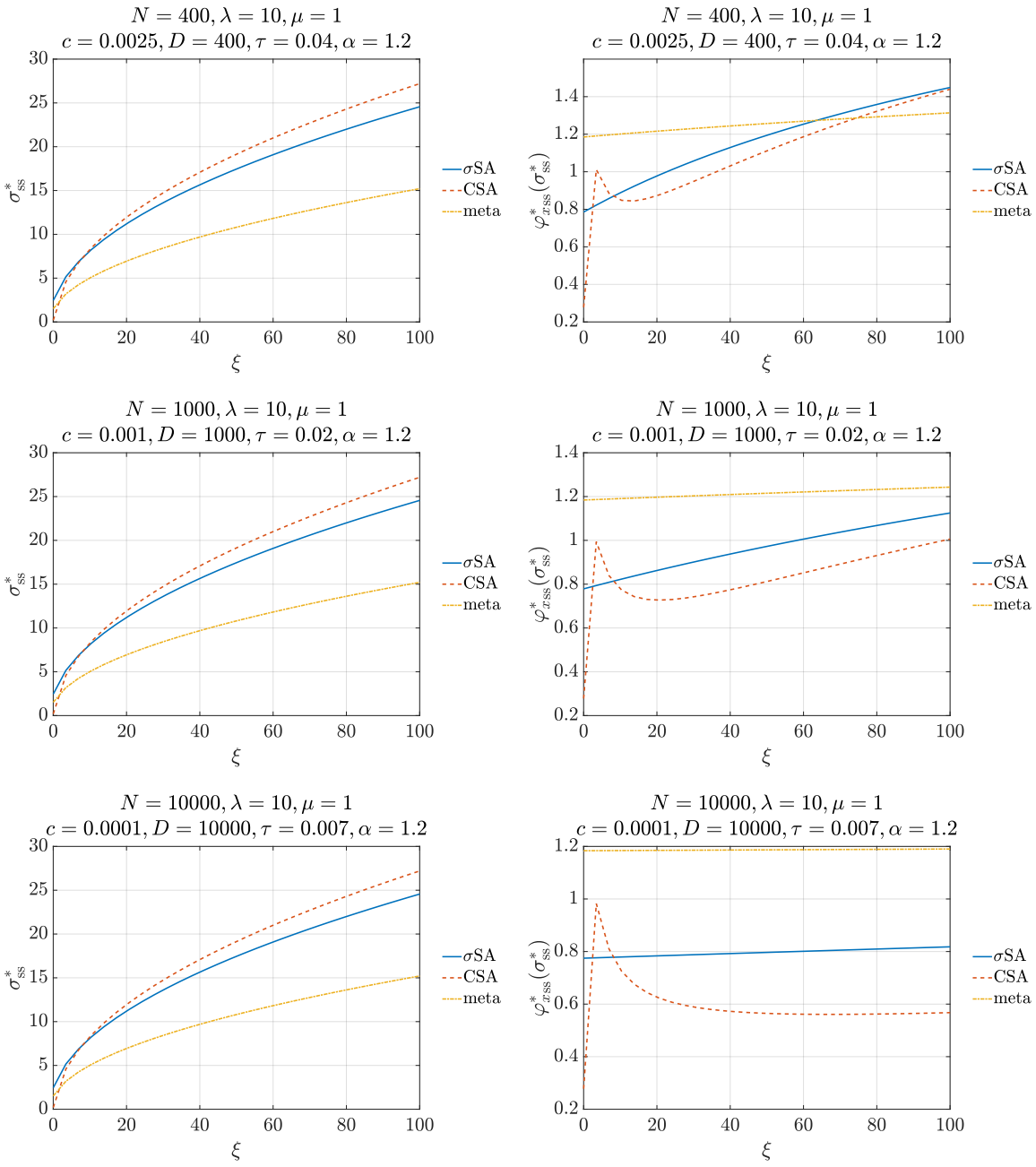

Figure 8: Comparison of the theoretical steady state normalized mutation strength and steady state progress rate results of the mutation strength control mechanisms σSA, CSA, and the meta approach. The considered dimensions are $N = 400$ (top), $N = 1000$ (middle), and $N = 10000$ (bottom). The parameters are set to the ones discussed in the corresponding analyses: $c = 1/N$, $D = 1/c$, $\tau = 1/\sqrt{2N}$, $\alpha = 1.2$, $\lambda = 10$, and $\mu = 1$.

Constrained Problem. *CoRR* arXiv:1812.06300 [cs.NE] (12 2018). Available online: https://arxiv.org/abs/1812.06300 (accessed Dec 18, 2018).

[18] Patrick Spettel and Hans-Georg Beyer. 2019. Analysis of the $(\mu/\mu_I, \lambda)$-CSA-ES with Repair by Projection Applied to a Conically Constrained Problem. *CoRR* arXiv:1901.07871 [cs.NE] (2019). Available online: https://arxiv.org/abs/1901.07871 (accessed Apr 06, 2019).

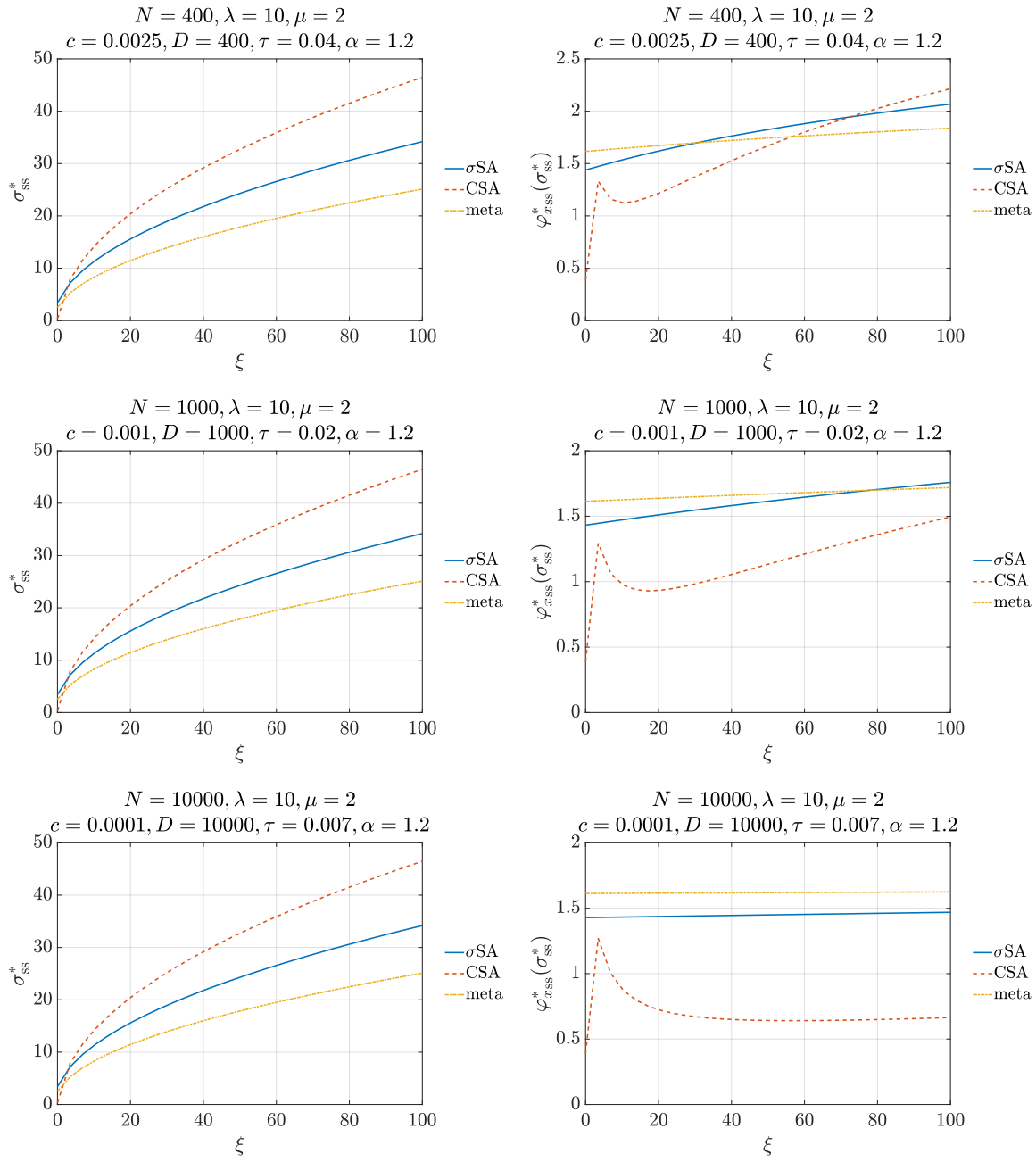

Figure 9: Comparison of the theoretical steady state normalized mutation strength and steady state progress rate results of the mutation strength control mechanisms σSA, CSA, and the meta approach. The considered dimensions are $N = 400$ (top), $N = 1000$ (middle), and $N = 10000$ (bottom). The parameters are set to the ones discussed in the corresponding analyses: $c = 1/N$, $D = 1/c$, $\tau = 1/\sqrt{2N}$, $\alpha = 1.2$, $\lambda = 10$, and $\mu = 2$.

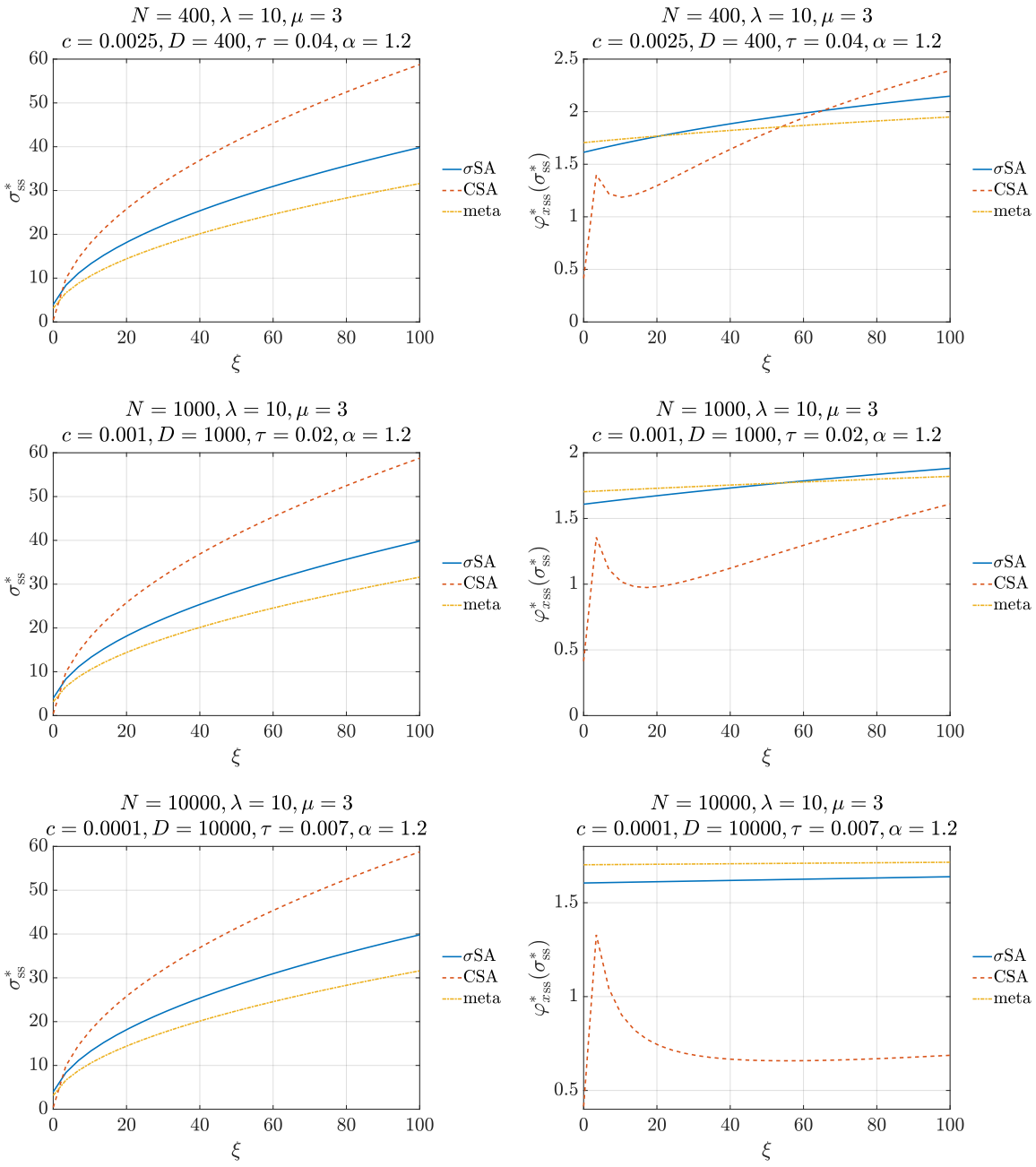

Figure 10: Comparison of the theoretical steady state normalized mutation strength and steady state progress rate results of the mutation strength control mechanisms σSA, CSA, and the meta approach. The considered dimensions are $N = 400$ (top), $N = 1000$ (middle), and $N = 10000$ (bottom). The parameters are set to the ones discussed in the corresponding analyses: $c = 1/N$, $D = 1/c$, $\tau = 1/\sqrt{2N}$, $\alpha = 1.2$, $\lambda = 10$, and $\mu = 3$.