Comparison of Constraint Handling Mechanisms for the (1, $\lambda$)-ES on a Simple Constrained Problem

Michael Hellwig  
michael.hellwig@fhv.at  
Research Center Process and Product Engineering, Vorarlberg University of Applied Sciences, Dornbirn, 6850, Austria

Dirk V. Arnold  
dirk@cs.dal.ca  
Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4R2

Abstract  
This paper investigates constraint handling techniques used in non-elitist single parent evolution strategies for the problem of maximizing a linear function with a single linear constraint. Two repair mechanisms are considered and the analytical results are compared to those of earlier repair approaches in the same fitness environment. The first algorithm variant applies reflection to initially infeasible candidate solutions while the second repair method uses truncation to generate feasible solutions from infeasible ones. The distributions describing the strategies’ one-generation behavior are calculated and used in a zeroth-order model for the steady state attained when operating with fixed step size. Considering cumulative step size adaptation the qualitative differences in the behavior of the algorithm variants can be explained. The approach extends the theoretical knowledge of constraint handling methods in the field of evolutionary computation and has implications for the design of constraint handling techniques in connection with cumulative step size adaptation.

Keywords  
Evolution Strategies, Constrained Optimization, Cumulative Step Size Adaptation, Constraint Handling Techniques

1 Introduction

In the field of evolutionary algorithms approaches to constrained optimization problems make use of different types of constraint handling techniques. These include penalty methods, repair mechanisms, as well as methods based on principles of multi-objective optimization. In this context the work of Oyman et al. (1999), Runarsson and Yao (2000), Mezura-Montes and Coello Coello (2005), and Kramer and Schwefel (2006), should be referenced. An overview of constraint handling techniques is provided in Mezura-Montes and Coello Coello (2011).

Different approaches are usually evaluated by direct comparison of the performances on test functions considered to be difficult. As a disadvantage the effects of the test environment on the performance of the algorithm are often difficult to interpret. Also the configuration of specific strategy parameters is not obvious. That is, complex test functions do not contribute to insight into the behavior of the algorithm on a microscopic level. In contrast to benchmark studies on large testbeds this work concentrates on the analysis of the behavior of algorithms on a very simple class of test functions. Simple test environments often provide a description of local regions within more complex environments. This way, analytical results can be obtained allowing for a deeper understanding of the influence of strategy parameters on algorithm performance.
Early contributions to constrained optimization with evolution strategies include the work of Rechenberg (1973), who investigated the performance of the \((1 + 1)\)-ES for the axis-aligned corridor model. The same environment has been studied by Schwefel examining the behavior of the \((1, \lambda)\)-ES in Schwefel (1975), and in Beyer (1989) the dynamics of the \((1 + 1)\)-ES for a constrained discuss-like function have been investigated.

Recent work, see Arnold (2011a,b), and Arnold (2013), examines the behavior of non-elitist ES using cumulative step size adaptation (CSA) on a linear problem with a single linear constraint. Two constraint handling techniques are compared. It is found that the resampling of infeasible candidate solutions results in premature convergence of the strategy in the face of small constraint angles, i.e., small angles between the objective function’s gradient direction and the normal vector of the constraint plane. This is due to short search paths that result in a systematic reduction of the step size. Handling constraints by projecting infeasible solutions onto the boundary of the feasible region allows setting the parameters of CSA in such a way that premature convergence is avoided for any constraint angle. This is due to successful projected candidate solutions predominantly lying in one direction, resulting in long search paths.

Further repair mechanisms exist; Helwig et al. (2013) among other methods suggest a reflection method, as well as an hyperbolic approach, in conjunction with particle swarm optimization in order to deal with infeasible candidate solutions. These approaches serve as the basis of the repair methods considered here. The first one reflects initially infeasible candidate solutions into the feasible region. The other repair mechanism truncates the mutation vector of infeasible offspring in such a way that the offspring is then located on the boundary of the feasible region.

The goal of this paper is to analyze these constraint handling techniques in the context of the linear optimization problem suggested by Arnold (2013) and to examine their performance relative to those repair mechanisms already considered in that respective paper. Thereby the theoretical analysis of the behavior of CSA-ES is extended by the analysis of two additional repair mechanisms. This reveals further insight into the potential of constraint handling techniques to maintain a sufficient population diversity in order to prevent premature convergence when approaching a constraint boundary.

The paper is organized as follows: In Section 2 the optimization problem, the algorithm, as well as the two suggested constraint handling techniques are described. Section 3 investigates the behavior of the ES assuming that the mutation strength is kept constant. A simple zeroth-order model is established to characterize the average distance from the constraint plane. Section 4 considers cumulative step size adaptation within the strategy. The paper concludes with a discussion of the results and suggestions for future research in Section 5. For reasons of clarity and comprehensibility calculations that underlie Section 3 are presented in Appendix A.

2 Problem Formulation and Algorithm

Throughout this paper the problem of maximizing a linear function

\[ F : \mathbb{R}^N \rightarrow \mathbb{R}, \quad N \geq 2, \]

with a single linear constraint is investigated. More precisely, since there exists no finite maximum the task is one of amelioration rather than maximization. We assume that the gradient vector of the objective function forms an acute angle with the normal vector \( \mathbf{n} \) of the constraint plane. In that setting, the challenge for the evolution strategy is to increase its mutation strength,
Comparison of Constraint Handling Mechanisms

thereby accelerating long term progress. Systematically reducing the mutation strength, as observed for the (1 + 1)-ES with success probability based step size adaptation as early as the 1970s (Schwefel, 1975), results in eventual stagnation of the strategy. Thus, the interplay of step size adaptation mechanism and constraint handling technique is of central significance.

Without loss of generality, the coordinate system can be chosen in such a way that the origin lies on the constraint plane, and that the x_1-axis matches with the gradient direction ∇F. Additionally, the x_2-axis should lie in the two-dimensional subspace spanned by the gradient vector and the normal vector. The corresponding angle is referred to as the constraint angle \( \theta = \angle(∇F, n) \). Constraint angles of interest are in the interval \((0, \pi/2)\). The unit normal vector of the constraint plane in terms of the considered coordinate system is \( n = (\cos \theta, \sin \theta, 0, \ldots, 0)^T \in \mathbb{R}^N \). The signed distance of a point \( x = (x_1, x_2, \ldots, x_N)^T \in \mathbb{R}^N \) from the constraint plane is \( g(x) = -n^T x = -x_1 \cos \theta - x_2 \sin \theta \). An illustration of the problem definition can be found in Fig. 1. The resulting optimization problem reads

\[
\max_{x} F(x) = \alpha x_1 \\
\text{s.t. } g(x) \geq 0 \tag{P}
\]

for some constant \( \alpha > 0 \). Due to the choice of the coordinate system, the variables \( x_3, \ldots, x_N \) have no effect on either the objective function or the constraint inequality. Despite its simplicity, the linear environment serves as a model for other constrained optimization problems in the immediate vicinity of the constraint plane, see Arnold (2013).

At this point the general algorithm of the (1, \( \lambda \))-CSA-ES dealing with problem (P) is introduced. It replaces infeasible offspring candidate solutions with feasible ones applying a specific repair mechanism. The two repair mechanisms investigated in this paper are explained in the corresponding subsections. Beginning with a feasible candidate solution \( x \in \mathbb{R}^N \) the (1, \( \lambda \))-ES generates \( \lambda \) offspring by performing the following steps per iteration:

1. For \( j = 1, \ldots, \lambda \)
   
   (a) Generate offspring candidate solution \( y^{(j)} \in \mathbb{R}^N \) by multiplying an \( N \)-dimensional, independent and identically distributed vector \( z \) drawn from a standard normal distribution with mean zero and unit covariance matrix with the mutation strength \( \sigma \) and adding it to the parental candidate solution \( x \):
   
   \[ y^{(j)} \leftarrow x + \sigma z^{(j)}. \]

   (b) If \( n^T y^{(j)} > 0 \), then repair the infeasible offspring candidate
   
   \[ y^{(j)} \leftarrow \text{REPAIR}(y^{(j)}) \]

   (c) Evaluate \( F(y^{(j)}) \) for \( j = 1, \ldots, \lambda \) and let \( \hat{z} \) denote the (possibly repaired) mutation vector of the best offspring, i.e., of the offspring with the largest objective function value.

2. Replace the parental candidate solution \( x \) with the best offspring candidate solution according to \( x \leftarrow x + \sigma \hat{z} \).

3. Modify the mutation strength \( \sigma \) using cumulative step size adaptation.
The algorithm generates an offspring candidate solution by random sampling and, where necessary additional repair. The sampling of an offspring candidate solution \( y \) in step 1(a) involves generating a mutation vector \( z = (y - x)/\sigma \in \mathbb{R}^N \) with standard normally distributed components. Introducing the normalized distance of the parental candidate solution \( x \) from the constraint plane

\[
\delta = \frac{g(x)}{\sigma}
\]

the admissibility of an offspring \( y^{(j)} \) can be described in terms of the corresponding mutation vector \( z^{(j)} \). That is, if the condition

\[
\mathbf{n}^\top z = z_1 \cos \theta + z_2 \sin \theta > \delta
\]

holds, then \( y \) is infeasible and repaired by the strategy to provide a candidate solution within the feasible region of the search space. Repair is performed in step 1(b) of the algorithm and will be described concentrating on the respective constraint handling mechanisms within the subsequent subsections.

In step 3 the algorithm adapts the strategy’s step size using cumulative step size adaptation. This mutation strength control mechanism has been introduced in Ostermeier et al. (1994) and is popular due to its use in the CMA-ES proposed by Hansen and Ostermeier (2001). Cumulative step size adaptation compiles a search path \( s \in \mathbb{R}^N \) starting from an initial \( s^{(0)} = 0 \) by implementing a fading record of past steps taken by the strategy according to

\[
s^{(t+1)} = (1 - c) s^{(t)} + \sqrt{c(2 - c)} \mathbf{z}.
\]

The length of its memory is determined by the choice of the constant \( c \in (0, 1) \) which is referred to as the cumulation parameter. CSA updates the mutation strength from generation \( t \) to \( (t + 1) \) conforming to the rule

\[
\sigma^{(t+1)} = \sigma^{(t)} \exp \left( c \frac{||s^{(t+1)}||^2 - N^2}{2DN} \right)
\]

with damping parameter \( D > 0 \) that scales the magnitude of the updates. Notice, update rule (3) is different from the original mutation strength update in Hansen and Ostermeier (2001), in the way that it adapts \( \sigma \) based on the squared length of the search path rather than based on search path’s length. With appropriately chosen parameters both variants often show a similar behavior. In accordance with Arnold (2013) this work uses the mutation strength update (3) as it simplifies the CSA analysis.

The decision whether the mutation strength is decreased or increased depends on the sign of the numerator \( ||s||^2 - N \) in (3). The basic idea is that long search paths indicate that selected mutation vectors predominantly point in one direction and could be replaced with fewer but longer steps. A short search path suggests that the strategy is moving back and forth and thus should benefit from smaller step sizes. If the unconstrained setting with randomly selected offspring candidate solutions (i.e., the ordering in step 2. of the algorithm is random) is considered the search path has the expected squared length of \( N \). In this case the mutation strength performs a random walk on log scale. That is, in expectation the logarithm of the mutation strength does not change at all. It is not unexpected that CSA in constrained settings may fail under some conditions.
Comparison of Constraint Handling Mechanisms

Figure 1: Linear objective function with a single constraint displayed within the subspace spanned by the \( x_1 \)- and \( x_2 \)-axes. The shaded area is the feasible region. The parental candidate solution \( x \) is at a distance \( g(x) \) from the constraint plane. The infeasible candidate solution \( y \) is reflected into the feasible region along the normal vector \( n \) of the constraint plane resulting in the new point \( y' \).

2.1 Repair by Reflection

In this section the repair mechanism denoted as reflection is considered. It enables the \((1, \lambda)\)-ES to deal with a single linear constraint and is illustrated in Fig. 1. Initially infeasible offspring candidate solutions are mirrored into the feasible region of the search space. The repair step within the algorithm above reads

\[
1(b) \quad \text{If } n^\top y^{(j)} > 0, \text{ then repair the infeasible offspring candidate solution by reflection} \\
\quad y^{(j)} \leftarrow y^{(j)} - 2(n^\top y^{(j)})n.
\]

Reflection of an initially infeasible offspring candidate solution transforms the first two components of the corresponding mutation vector \( z \) in the following way

\[
z_1 \leftarrow \begin{cases} 
z_1 + 2\delta \cos \theta - 2z_1 \cos^2 \theta - 2z_2 \sin \theta \cos \theta , & \text{if Eq. (1) holds} \\
z_1 , & \text{otherwise,}
\end{cases}
\]

and

\[
z_2 \leftarrow \begin{cases} 
z_2 + 2\delta \sin \theta - 2z_2 \sin^2 \theta - 2z_1 \cos \theta \sin \theta , & \text{if Eq. (1) holds} \\
z_2 , & \text{otherwise.}
\end{cases}
\]

All other components of \( z \) remain unchanged. Notice, the feasible point after reflection resides at the same distance from the constraint plane as the initially infeasible solution.

2.2 Repair by Truncation

We now consider the second repair mechanism which truncates the mutation vector of an initially infeasible offspring candidate solution at the edge of the feasible region of the search
space. In the following we refer to this method as truncation. As well as repair by reflection truncation enables the (1,λ)-ES to deal with the previously introduced single linear constraint. The procedure is illustrated in Fig. 2. Accordingly, the repair step 1(b) within the algorithm can be formulated as

1(b) If \( n^\top y(j) > 0 \), then repair the infeasible offspring candidate by applying

\[
y'(j) = y(j) - \frac{n^\top y(j)}{n^\top z(j)} z(j).
\]

An offspring candidate solution is generated by random sampling and, if applicable, repair by truncation. Truncation modifies the standard normally distributed mutation vector of an initially infeasible offspring in the following way:

\[
z_1 = \begin{cases} \delta \frac{z_1 \cos \theta + z_2 \sin \theta}{z_1} & \text{if Eq. (1) holds} \\ z_1 & \text{otherwise,} \end{cases}
\]  \( (6) \)

and

\[
z_2 = \begin{cases} \delta \frac{z_1 \cos \theta + z_2 \sin \theta}{z_2} & \text{if Eq. (1) holds} \\ z_2 & \text{otherwise.} \end{cases}
\]  \( (7) \)

Again, all other components of \( z \) are not affected by the repair step.
3 Behavior for Fixed Mutation Strength

In the first step, the strategy is investigated considering a fixed mutation strength $\sigma$. That is, the CSA step within the algorithm is omitted during this section. The probability of an offspring candidate solution being feasible, as well as the expected improvement in the objective function value, depend on the distance $g(x)$ of the parental candidate solution from the constraint plane. Hence, the ES behavior can be completely described by means of the normalized distance $\delta = g(x)/\sigma$. According to Arnold (2013) the evolution of $\delta$ is then characterized by the difference equation

$$\delta(t+1) = \delta(t) - \hat{z}_1 \cos \theta - \hat{z}_2 \sin \theta$$

where superscripts on $\delta$ indicate time, or the number of generations. We refer to $\hat{z} = (\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_N)^T$ as the mutation vector after repair associated with the best offspring candidate solution in generation $t$.

In the following, a characterization of the distribution of offspring candidate solutions conditional on $\delta$ is provided and the expected step made in a single iteration of the algorithm is calculated considering both repair mechanisms. Finally, the average normalized distance from the constraint plane realized by the respective repair methods when operating with constant mutation strength is investigated.

3.1 Single Step Behavior: Mutation

The variation step of the evolution strategy consists of mutation, and when applicable repair according to the specific constraint handling approach used. Since initially infeasible offspring candidate solutions are repaired the strategy affects the distribution of the original normally distributed offspring. The marginal distribution of the $z_1$ components, as well as the distribution of the $z_2$ components conditional on $z_1$, after repair are derived in appendix A for both repair mechanisms. Taking into account Eq. (34), in the case of reflection the marginal density of the $z_1$ components after reflection reads

$$p_{1\text{reflection}}(x) = \frac{e^{-\frac{1}{2}(2\delta \cos \theta - x)^2}}{\sqrt{2\pi}} \Phi(h(x)) + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \Phi\left(\frac{\delta - x \cos \theta}{\sin \theta}\right).$$

Notice, in this context $\Phi(x) = \int_{-\infty}^{x} e^{-\frac{1}{2}z^2} dz/\sqrt{2\pi}$ denotes the cumulative distribution function of the standard normal distribution and the abbreviation

$$h(x) := \frac{\delta(1 - 2 \sin^2 \theta) - x \cos \theta}{\sin \theta}$$

is used. According to Eq. (36), the corresponding cumulative distribution function is

$$P_{1\text{reflection}}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{h(x)} \Phi(y) e^{-\frac{1}{2}(2\delta \cos \theta - y)^2} dy + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{h(x)} e^{-\frac{1}{2}y^2} \Phi\left(\frac{\delta - y \cos \theta}{\sin \theta}\right) dy.$$ (11)

Considering repair by truncation the respective marginal density of the $z_1$ components is derived in Eq. (55) and reads

$$p_{1\text{truncation}}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \Phi\left(\frac{\delta - x \cos \theta}{\sin \theta}\right) + \frac{\delta}{2\pi \sin \theta k(x)} e^{-\frac{1}{2}k(x)}.$$ (12)
The corresponding cumulative distribution function is found in Eq. (56) as

\[
p_{\text{truncation}}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} \Phi \left( \frac{\delta - y \cos \theta}{\sin \theta} \right) dy + \frac{\delta}{2\pi \sin \theta} \int_{-\infty}^{\infty} \frac{1}{k(y)} e^{-\frac{y^2}{2}} dy.
\]

(13)

In the case of truncation the abbreviation

\[k(x) := \delta^2 + \left( \frac{x - \delta \cos \theta}{\sin \theta} \right)^2\]

(14)
is used to improve the readability of the results.

### 3.2 Single Step Behavior: Selection

The purpose of this section is the characterization of the expected step made by the strategy for both repair mechanisms. To this end, the first two components of the mutation vector \( \hat{z} \) which belong to the best offspring candidate solution are considered and their first two moments about zero

\[e_{i,j} = E \left[ \hat{z}_i^j \right], \quad i, j \in \{1, 2\}\]

(15)

are computed. Since the following formulas hold for both repair cases superscripts indicating the constraint handling method are omitted. Depending on the repair method applied, the first component \( \hat{z}_1 = z_i^{1,1} \) is the \( \lambda \)-th order statistic of a sample of \( \lambda \) independent realizations of a random variable distributed according to \( p_{\text{reflection}}(x) \) in (9) in case of reflection, or according to \( p_{\text{truncation}}(x) \) in (12) respectively in case of truncation. The probability density function of the \( \hat{z}_1 \) component is thus

\[\hat{p}_1(x) = \lambda p_1(x)P_{1}^{i-1}(x).\]

(16)

By integration of Eq. (16) the means of \( \hat{z}_i \), \( i = 1, 2 \) can be computed as

\[e_{1,1} = \int_{-\infty}^{\infty} x \hat{p}_1(x) dx = \lambda \int_{-\infty}^{\infty} x p_1(x)P_{1}^{i-1}(x) dx,
\]

(17)

and

\[e_{2,1} = \int_{-\infty}^{\infty} E \left[ z_2 | z_1 = x \right] \hat{p}_1(x) dx = \lambda \int_{-\infty}^{\infty} E \left[ z_2 | z_1 = x \right] p_1(x)P_{1}^{i-1}(x) dx.
\]

(18)

Similarly, the second moments about zero of \( \hat{z}_i \) for \( i = 1, 2 \) equal

\[e_{1,2} = \int_{-\infty}^{\infty} x^2 \hat{p}_1(x) dx = \lambda \int_{-\infty}^{\infty} x^2 p_1(x)P_{1}^{i-1}(x) dx,
\]

(19)

and

\[e_{2,2} = \int_{-\infty}^{\infty} E \left[ z_2^2 | z_1 = x \right] \hat{p}_1(x) dx = \lambda \int_{-\infty}^{\infty} E \left[ z_2^2 | z_1 = x \right] p_1(x)P_{1}^{i-1}(x) dx.
\]

(20)

The moments \( E \left[ z_2^j | z_1 = x \right], \quad j = 1, 2 \) conditional on \( z_1 \) are presented in Appendix A for both constraint handling methods. In the case of reflection they can be found in Eqs. (40) and (41). Considering repair by truncation the respective expressions are derived in Eqs. (60) and (61).
3.3 Steady State Behavior

All results derived up to this point are concerned with the algorithm’s behavior in single time steps and are conditional on the normalized distance \( \delta \) from the constraint plane. Due to the assumption that the mutation strength is held constant the distribution of \( \delta \) values approaches a stationary limit distribution when the algorithm is iterated. According to Meyer-Nieberg and Beyer (2012), an approximation of this limit distribution can be computed by expanding the unknown distribution in terms of its moments, calculating the resulting moments after an update, and imposing stationarity conditions demanding that the moments remain unchanged. This approach results in a system of as many equations as there are moments included in the computation. Making use of a zeroth-order approximation, this limit distribution is well characterized by its mean neglecting higher order moments. The limit distribution can thus be modeled as a (shifted) Dirac delta function and by solving the corresponding stationarity condition

\[
E[\delta^{(t+1)}] = \delta^{(t)}
\]  

for \( \delta = \delta^{(t)} \) one obtains an approximation of the average distance of the parental candidate solution from the constraint plane. Though considering higher-order moments may provide better approximations the Dirac model turns out to provide a quite accurate description of the behavior of the constraint handling mechanisms.

Taking the expected value of Eq. (8) and using the stationary condition (21) yields

\[
\epsilon_{1,1} \cos \theta + \epsilon_{2,1} \sin \theta = 0.
\]  

Integral representations of the two moments have been derived in Section 3.2. Thus equation (22) can be solved numerically for \( \delta = \delta^{(t)} \). The results are illustrated in Fig. 3 for both constraint handling techniques. The average normalized distance of the parental centroid from the constraint plane of each repair mechanism is compared to previous results obtained in Arnold (2013) using resampling, as well as repair by projection. In this context, resampling refers to the constraint handling technique which continuously resamples infeasible candidate solutions until \( \lambda \) feasible offspring candidate solutions have been generated in each iteration of the algorithm. The repair mechanism referred to as projection simply translates initially infeasible candidate solutions along the normal vector of the constraint plane onto the boundary of the feasible region. In part (a) of Fig. 3 the average normalized distance, realized by the \((1, \lambda)\)-ES that repairs initially infeasible candidate solutions by reflection, is plotted against the constraint angle \( \theta \). Part (c) presents the respective results for the ES that truncates the mutation vector of an initially infeasible candidate solution at the boundary of the feasible set. The parts (b) and (d) illustrate the same information making use of logarithmic scales. For both strategies, i.e., reflection as well as truncation, the average normalized distance increases with increasing \( \theta \). Since it truncates infeasible candidate solutions at the boundary of the feasible region the strategy using truncation tracks the constraint plane more closely than the strategy that uses reflection. The observed distances resemble those obtained by projection in terms of the realized average normalized distances for small constraint angles. The points in Fig. 3 represent experimental data which are obtained by averaging over \( 10^5 \) steps, i.e., generations, of the respective evolution strategies described in Section 2 with fixed mutation strength \( \sigma \). Comparing these data points to the solid lines, the quality of the zeroth order approximation provided by the Dirac model reveals a good agreement for small constraint angles. Considering larger \( \theta \) values results in deteriorations of the approximation quality. As the constraint angle becomes less acute the strategy tracks the constraint plane at a greater distance. Thus variations in that distance become significant. In this situation the Dirac model is inappropriate. Making use of an exponential model, like the
Figure 3: The average normalized distance $\delta$ from the constraint plane is plotted against the constraint angle $\theta$ for the (1, 10)-ES. (a) The numerical results of reflection are displayed by the solid line. (b) Plot of the same information using logarithmic scales for both axes. (c) Illustration of the numerically generated average distance realized by truncation displayed by the solid line. (d) Truncation results in logarithmic axes. In each part the predictions are compared to experimental results obtained by averaging over $10^5$ steps of the corresponding evolution strategies, as well as to numerical results of repair by projection (dash-dotted line) and resampling (dashed line).

one proposed in Arnold and Brauer (2008), would likely increase the accuracy of the distance distribution.

4 Mutation Strength Control

In this section we drop the assumption that the mutation strength of the strategy is fixed. That is, the strategy’s step size is adapted using cumulative step size adaptation as described in Section 2. This affects the longterm behavior of the (1, $\lambda$)-ES in the constrained linear environment considered in such a way that the strategy does either increase or decrease the mutation strength $\sigma$ on average. With respect to the optimization problem (P) increasing $\sigma$ is required because decreasing $\sigma$ would lead to premature convergence.

The behavior of the (1, $\lambda$)-ES with cumulative step size adaptation applied to the constrained linear problem can be described by a nonlinear, stochastic dynamic system which depends on $\sigma$, $\delta$, as well as the search path $s$. Rather than trying to solve the complex problem of investigating the dynamics of the stochastic process we follow the approach in Arnold (2013) where the logarithmic adaptation response of the strategy operating out of a stationary state is
Comparison of Constraint Handling Mechanisms

The logarithmic adaptation response of the \((1, \lambda)-\)ES applying CSA is defined as

\[
\Delta^{(t+1)} = \frac{2DN}{c} \log \left( \frac{\sigma^{(t+1)}}{\sigma^{(t)}} \right).
\] (23)

The sign of \(\Delta^{(t+1)}\) indicates whether the mutation strength is increased or decreased, i.e., if the logarithmic adaptation response yields positive values then \(\sigma\) is increased and vice versa. Experiments indicate that the resulting predictions of the strategy’s dynamic behavior are accurate for a broad range of parameter settings.

Assume that the strategy has been iterated with a fixed mutation strength for a sufficiently long time until time step \(t\) is computed. Then the expected logarithmic adaptation response at time step \((t + 1)\) is computed. The derivation of the expected value has also been completed in Arnold (2013) resulting in

\[
E \left[ \Delta^{(t+1)} \right] = (e_{1,2} + e_{2,2}) + 2 \frac{(1 - c)}{c} \left( e_{1,1}^2 + e_{2,1}^2 \right) - 2.
\] (24)

This equation depends on the first two moments of the first and second component of the mutation vector that provides the best offspring candidate solution in the respective generation. The moments \(e_{i,j}, i, j = 1, 2\) are specified in Eqs. (17) to (20) for both repair mechanisms considered. Notice, Eq. (24) depends on the search space dimension only through the choice of the cumulation parameter \(c\), which is usually specified as an \(N\)-dependent value.

The theoretical results of the moments \(e_{i,j}, i, j = 1, 2\) can be validated by comparing them to measurements from experimental runs of the evolution strategies. Using the Dirac model to compute the average distance of the parental candidate solution from the constraint plane by solving (22) the moments \(e_{i,j}, i, j = 1, 2\) can be evaluated numerically. For reflection, and truncation respectively, the resulting curves are plotted against the constraint angle \(\theta\) in Fig. 4. The results of repair by reflection are represented by the solid lines while the moments obtained by truncation are illustrated as dashed lines. The comparison of the curves with the corresponding values obtained experimentally in runs of the respective \((1, \lambda)\)-ES with fixed mutation strength shows a good visual agreement for the first moments \(e_{1,1}\) and \(e_{2,1}\). The approximation quality of the Dirac model deteriorates for the second moments \(e_{1,2}\), and \(e_{2,2}\). In the first case, the approximation in case of highly acute constraint angles exhibits small deviations. However, the values in that scenario are generally small and have limited impact on the results below. Regarding \(e_{2,2}\) the deviations occur with increasing values of \(\theta\) but appear to be visually good for small constraint angles.

Focusing on Eq. (24) the logarithmic adaptation response decreases with increasing cumulation parameter. As \(c\) approaches one the second term decreases to zero. That is, depending on the first term the expected logarithmic adaptation response may be negative, resulting in decreasing step sizes. In order to achieve a positive expected logarithmic response the negative third term in (24) has to be outweighed by the first two terms. This can be realized for \(c\) values that are close enough to zero. Requiring \(E [\Delta_c] \geq 0\) yields the condition

\[
c \leq \frac{e_{1,1}^2 + e_{2,1}^2}{1 + e_{1,1}^2 + e_{2,1}^2 - (e_{1,2} + e_{2,2})/2}
\] (25)

for the cumulation parameter. Condition (25) allows for the computation of the maximal \(c\) value for which stagnation, i.e. convergence to a non-stationary point, can be avoided. The calculation is performed by considering equality in (25), where the moments \(e_{i,j}\) are computed.
Figure 4: The moments $e_{i,j}$ of the first and second components of $\hat{z}$ are plotted against the constraint angle $\theta$, for $i, j \in \{1, 2\}$ and $\lambda = 10$. The discrete data points represent measurements obtained by averaging over $10^5$ runs of the $(1, 10)$-ES with fixed mutation strength $\sigma$.

by making use of the Dirac model for the average distance from the constraint plane. The results for the strategies using reflection, as well as truncation, are illustrated in Fig. 5 and compared to the corresponding results for projection and resampling obtained earlier in Arnold (2011a), and Arnold (2011b) respectively. It can be observed that for the strategy that reflects initially infeasible candidate solutions the maximal cumulation parameter $c$ capable of avoiding premature convergence decreases to zero as the constraint angle becomes increasingly acute. For the strategies applying truncation, as well as resampling, the same behavior can be deduced from Fig. 5. That is, provided that the constraint angle $\theta$ is small enough, all three strategies will converge to a non-stationary limit point independent of the choice of the cumulation parameter $c$. This behavior has already been investigated for the strategy that resamples infeasible offspring candidate solutions. The approach of the maximal cumulation parameter toward a finite limit value for decreasing $\theta$, as observed in Arnold (2013), is only realized by the strategy that repairs infeasible candidate solutions by projection. Thus, only projection is able to ensure continuous progress for arbitrarily small constraint angles. Note, as $c \in (0, 1\), values exceeding one are not shown in Fig. 5.

Recalling Fig. 4, the failure of cumulative step size adaptation in the limit of small constraint angles can be explained as follows: The moments $e_{1,1}$ and $e_{1,2}$ of the parental mutation vector’s first component describe the behavior in $x_1$ direction. Considering small values of $\theta$ both moments assume very small values near zero as both strategies operate in close vicinity to the constraint plane. In order to compensate for the small contribution of the moments $e_{1,1}$ and $e_{1,2}$ to the expected squared length of the search path $\|s\|^2$ on the $x_1$-axis the moments $e_{2,1}$...
and $e_{2,2}$ have to be large enough to ensure a positive expected logarithmic adaptation response $E[\Delta r]$. In terms of the first two moments of $z_1$ and $z_2$ components repair by reflection resembles the strategy that resamples infeasible solutions. In Arnold (2013) it was found that the behavior in $x_2$ direction almost performs a random walk, i.e., $e_{2,1} \approx 0$ and $e_{2,2} \approx 1$. The same observation can be made for the strategy that applies reflection (see Fig. 4). This can be clarified considering the way reflection acts on initially infeasible offspring candidate solutions. In the limit of very small $\theta$, on average, half of the offspring are infeasible and will be reflected into the feasible region. On average, half of those that are reflected have a positive $z_2$ component and the other half is reflected on a negative $z_2$ component. But offspring being reflected on a negative $z_2$ component are for small $\theta$ only slightly more likely to be selected. That is, in contrast to the observations made for projection, see Arnold (2013), the strategy using reflection does not exhibit a strong correlation between candidates being successful and a large negative value of its $z_2$ components. As a consequence reflection behaves similar to resampling.

Regarding the strategy that truncates the mutation vectors of initially infeasible candidate solutions at the boundary of the feasible area, $e_{1,1}$ and $e_{1,2}$ approach even smaller values. The observed moments $e_{2,1}$ and $e_{2,2}$ assume zero for decreasing constraint angles $\theta$. Thus they are not suitable to compensate for the low contribution of $e_{1,1}$ and $e_{1,2}$ on the $x_1$-axis. To prevent convergence to a non-stationary limit point the strategy using truncation requires even lower choices of the cumulation parameter $c$. In the limit of very small $\theta$, the probability of generating an initially infeasible offspring is again approximately 0.5. The area that provides positive $z_1$ components after truncation shrinks considerably with decreasing constraint angle. That is, the mutation vectors of successful candidate solutions after truncation are on average very small. We can infer the same correlation from truncation that was observed in the case of projection, i.e., that successful offspring after truncation tend to be associated with negative $z_2$ components. But this correlation is not able to compensate the drawback of small mutation vectors. As a consequence, using repair by truncation within the $(1, \lambda)$-CSA-ES is not a well suited constraint handling method, especially for rather small constraint angles $\theta$.

![Figure 5](image.png)

Figure 5: The maximum value of the cumulation parameter $c$ for which convergence is avoided plotted against the constraint angle $\theta$. The results for the repair mechanisms reflection, truncation, and projection, as well as constraint handling by resampling are displayed for $\lambda = 10$. 

The accuracy of the predictions based on the Dirac model is verified in Fig. 6. There, all lines are obtained from Eq. (25) for both strategies, i.e., reflection and truncation respectively, as well as for several offspring population sizes $\lambda$. Each dot in each figure represents experimental results obtained from 10 independent runs of the respective evolution strategies with cumulative step size adaptation initialized with $\delta^{(0)} = 1$, $\sigma^{(0)} = 1$, and $s^{(0)} = 0$. Further, search space dimensionality $N = 2$ and damping parameter $D = 1$ have been used. The runs were terminated after the mutation strength assumed a value either smaller than $10^{-10}$ or greater than $10^{10}$. The very small step sizes suggest that stagnation is likely to ensue while large step sizes point to continuing progress at increasing rates. The + symbol in Fig. 6 indicates that at least 9 out of the 10 runs terminated with $\sigma < 10^{-10}$; the $\times$ symbol indicates that termination took place with $\sigma > 10^{10}$ in at least 9 out of the 10 runs. Neither symbol is present at a grid location in the plots if either termination criterion was satisfied in at least 2 out of the 10 runs. Within the range of considered constraint angles and population size parameters the agreement of our predictions based on Eq. (25) with the experimentally generated results is very good. Only for small population sizes the strategy using truncation shows deviations from the theoretical predictions. This points towards the Dirac model providing an insufficient approximation of the evolution strategy’s behavior in the range of small constraint angles and small population sizes.
5 Summary and Conclusions

This paper investigated constraint handling techniques for the (1, λ)-ES using cumulative step size adaptation on a linear problem with a single linear inequality constraint. Two different repair mechanisms were considered: the reflection of infeasible points into the feasible region as well as the truncation of the mutation vectors at the boundary of the feasible area. As pointed out earlier, an interesting aspect of the simple scenario of a linear function with a single linear constraint is that it can be used to model microscopic properties of more general constrained problems.

The use of the simple Dirac model allowed for approximating the average distance of the parental candidate solution from the constraint plane provided that the strategy runs with fixed mutation strength. Considering cumulative step size adaptation and assuming that the evolution strategy operates out of a stationary state, this approximation was then used to derive an expression for the expected logarithmic adaptation response. Hence, it was possible to calculate the maximal cumulation parameter up to which premature convergence can be avoided by the (1, λ)-ES. The theoretical predictions have been verified by experiments over a wide range of constraint angles with varying number of offspring. The validation revealed generally very good agreement.

While projection still allows for positive logarithmic adaptation response for arbitrarily small constraint angles, neither reflection nor truncation turned out to enable the strategy to make sustained progress by continually increasing the mutation strength for small θ. The reason for the better performance of projection is that it exhibits a strong correlation between an offspring being successful and a large negative value of its $z_2$ component. This is not true for repair by reflection since an initially infeasible offspring with negative $z_2$ component is not necessarily successful after reflection. Thus reflection shows a behavior similar to resampling. For truncation the strong correlation between a successful offspring and a negative value of its $z_2$ component can also be found. However, on average truncation generates short mutation vectors and consequently the correlation turns out to be insufficient to make up for the short length. There is also empirical evidence in Beyer and Finck (2012) that for more complex constraints repair by projection provides better performance compared to truncation and resampling.

The results obtained provide insight into the interactions of constraint handling techniques used in combination with cumulative step size adaptation. Possible future work includes the extension of the analysis to the more general multi-recombinant ($μ/μ, λ$)-ES. Furthermore, the interplay of repair by reflection with other step size adaptation methods, such as mutative self-adaptation, remains to be discussed. A theoretical attempt to address problems with a nonlinear constraint has been provided in Arnold (2014). There, a linear optimization problem with conically constrained feasible region was investigated by using resampling to deal with infeasible candidate solutions. A future analysis of this type of problem regarding repair by projection seems reasonable. As well, investigations regarding nonlinear objective functions and further nonlinear constraints are worth considering.

Acknowledgements

This work was supported by the Austrian Science Fund (FWF) under grant P22649-N23.
References


Comparison of Constraint Handling Mechanisms


Appendix

A Computation of the Distributions

This appendix computes the distribution of the mutation vectors’ first components \(z_1\) after the repair step for both constraint handling techniques, i.e., for repair by reflection as well as the repair method using truncation. Subsequently, the distribution of the mutation vectors’ \(z_2\) components conditional on \(z_1\) is computed for the respective strategies.

A.1 Reflection

In the first step the probability distribution function of the \(z_1\) components of the mutation vector after reflection is calculated. Afterwards we derive the first two moments about zero of its \(z_2\) components conditional on \(z_1\).

A.1.1 Distribution of the \(z_1\) components after reflection

The distribution of the \(z_1\) components after reflection is the sum of the distribution of the corresponding components from immediately feasible offspring candidate solutions and the distribution of those candidates that are initially infeasible and thus repaired by reflection into the feasible area. Due to Section 2.1, see Eqs. (4) and (5), reflection transforms the mutation vector of infeasible candidate solutions in the following way

\[
\begin{bmatrix}
    z_1 \\
    z_2
\end{bmatrix}
\leftarrow
\begin{bmatrix}
    z_1 - 2(z_1 \cos \theta + z_2 \sin \theta) \cos \theta + 2\delta \cos \theta \\
    z_2 - 2(z_1 \cos \theta + z_2 \sin \theta) \sin \theta + 2\delta \sin \theta
\end{bmatrix}.
\]

A feasible offspring candidate solution \(y = x + \sigma z\) has to satisfy the constraint condition \(\delta \geq z_1 \cos \theta + z_2 \sin \theta\). Here, \(\delta = g(x)/\sigma\) refers to the normalized distance of the parental candidate solution to the constraint plane and \(\theta\) denotes the constraint angle. The joint probability density of the initially feasible \(z_1\) and \(z_2\) components is the truncated normal density

\[
p_{z_1z_2}(x, y) = \begin{cases} 
  \frac{1}{2\pi \sigma^2} e^{-\frac{1}{2\sigma^2}(y^2 + z_1^2)}, & \text{if } x \cos \theta + y \sin \theta \leq \delta \\
  0, & \text{otherwise.}
\end{cases}
\]

Notice, the factor \(\Phi(\delta)\) in the denominator is the probability of an offspring being feasible

\[
\Phi(\delta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y^2 + z_1^2)} dy dz.
\]

Accordingly, the joint probability density of initially infeasible \(z_1\) and \(z_2\) components before reflection is a truncated normal density as well. It is described by

\[
p_{z_1z_2}(x, y) = \begin{cases} 
  \frac{1}{2\pi \sigma^2} e^{-\frac{1}{2\sigma^2}(y^2 + z_1^2)}, & \text{if } x \cos \theta + y \sin \theta > \delta \\
  0, & \text{otherwise.}
\end{cases}
\]

The strategy reflects infeasible candidate solutions into the feasible region of the search space. That is, a transformation of the random variables \(x\) and \(y\) according to the repair step (26) applied to Eq. (29) yields the joint probability density of initially infeasible \(z_1\) and \(z_2\) components after reflection. After simple rearrangements the transformation reads

\[
\begin{bmatrix}
    z_1 \\
    z_2
\end{bmatrix}
\leftarrow
\begin{bmatrix}
    2\delta \cos \theta - z_1(1 - 2 \sin^2 \theta) - 2z_2 \sin \theta \cos \theta \\
    2\delta \sin \theta - 2z_1 \sin \theta \cos \theta - z_2(1 - 2 \cos^2 \theta)
\end{bmatrix}.
\]

Computing the Jacobian matrix \(J\) of this transformation with determinant \(|\det(J)| = 1\) and applying some straightforward simplifications the resulting joint probability density of the initially infeasible components after reflection is

\[
p_{z_1z_2}(x, y) = \begin{cases} 
  \frac{1}{2\pi \sigma^2} e^{-\frac{1}{2\sigma^2}(y^2 + z_1^2)}, & \text{if } x \cos \theta + y \sin \theta \leq \delta \\
  0, & \text{otherwise.}
\end{cases}
\]
Weighting the joint densities \( p_{\text{reflection}}^{\text{feas}}(x,y) \) and \( p_{\text{reflection}}^{\text{infeas}}(x,y) \) with the probability of their occurrence \( \Phi(\delta) \), and \( (1 - \Phi(\delta)) \) respectively, the overall joint probability density of the \( z_1 \) and \( z_2 \) components after the repair step (26) is thus

\[
p_{\text{reflection}}^{\text{feas-infeas}}(x,y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} + \frac{1}{2\pi} e^{-\frac{1}{2}((2\cos \theta - x)^2 + (2\sin \theta - y)^2)}
\]

(32)

for those \( x \) and \( y \) that satisfy the condition \( x \cos \theta + y \sin \theta \leq \delta \) and zero otherwise.

The marginal density of the \( z_1 \) components after reflection is derived by integration over all feasible \( z_2 \) values

\[
p_{1,\text{reflection}}(x) = \int_{-\infty}^{\infty} p_{\text{reflection}}^{\text{feas-infeas}}(x,y) dy.
\]

(33)

Solving the integral yields

\[
p_{1,\text{reflection}}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\delta - x \cos \theta)^2} \Phi\left(\frac{\delta - x \cos \theta}{\sin \theta}\right) + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(2\cos \theta - x)^2} \Phi\left(\frac{2(1 - 2\sin^2 \theta) - x \cos \theta}{\sin \theta}\right)
\]

(34)

By introducing the abbreviation

\[h(x) := \frac{\delta (1 - 2\sin^2 \theta) - x \cos \theta}{\sin \theta} \]

(35)

and integrating the marginal density with respect to \( x \) the corresponding cumulative distribution function of the \( z_1 \) components after reflection is

\[
P_{1,\text{reflection}}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \Phi(\delta(y)) e^{-\frac{1}{2}(2\cos \theta - y)^2} dy + \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{1}{2}(2\cos \theta - y)^2} \Phi\left(\frac{\delta - y \cos \theta}{\sin \theta}\right) dy.
\]

(36)

### A.1.2 Distribution of the \( z_2 \) components after reflection

The next step considers the distribution of the \( z_2 \) components after reflection conditional on \( z_1 = x \). On the basis of this distribution the first two moments about zero can be computed. The probability \( P_{\text{ref}}(x) = \text{Prob}\{y \text{ has been reflected} | z_1 = x\} \) of an offspring being infeasible and thus being in need of repair conditional on \( z_1 = x \) is obtained by computation of the relative weight of the second term in the marginal density \( p_{1,\text{reflection}}^{\text{feas-infeas}}(x) \) in Eq. (34) and reads

\[
P_{\text{ref}}(x) = \frac{1}{P_{1,\text{reflection}}^{\text{feas-infeas}}(x)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(2\cos \theta - x)^2} \Phi(\delta(x)).
\]

(37)

The contributions to the first two moments after reflection conditional on \( z_1 = x \) which arise from initially feasible candidate solutions, can be computed from the conditional probability density

\[
p_{2,\text{feas}}^{\text{feas}}(y|z_1 = x) = p_{1,\text{reflection}}^{\text{feas-infeas}}(x,y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x \cos \theta)^2}
\]

(38)

where

\[
p_{1,\text{reflection}}(x) = \frac{1}{\sqrt{2\pi} \Phi(\delta)} e^{-\frac{1}{2}(x \cos \theta)^2} \Phi\left(\frac{\delta - x \cos \theta}{\sin \theta}\right)
\]

(39)

denotes the marginal density function of the \( z_1 \) components of initially feasible candidate solutions according to Eq. (27). Taking into account Eq. (38) the contributions to the first two moments equal

\[
E[z_2|z_1 = x \land y \text{ has not been reflected}] = \frac{1}{\sqrt{2\pi} \Phi\left(\frac{\delta - x \cos \theta}{\sin \theta}\right)} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y \cos \theta)^2} \Phi\left(\frac{\delta - x \cos \theta}{\sin \theta}\right) dy
\]

(40)

\[
= \frac{-1}{\sqrt{2\pi} \Phi\left(\frac{\delta - x \cos \theta}{\sin \theta}\right)} \exp \left( -\frac{1}{2} \left(\frac{\delta - x \cos \theta}{\sin \theta}\right)^2 \right).
\]
and

\[
E \left[ z_2 | z_1 \right] = x \land y \text{ has not been reflected} \\
= \frac{1}{\sqrt{2 \pi} \Phi \left( \frac{\delta - \cos \theta}{\sin \theta} \right)} \int_{-\infty}^{\infty} y^2 e^{-\frac{y^2}{2}} dy \\
= 1 - \frac{1}{\sqrt{2 \pi} \Phi \left( \frac{\delta - \cos \theta}{\sin \theta} \right)} e^{\frac{1}{2} \left( \frac{\delta - \cos \theta}{\sin \theta} \right)^2}.
\]

(41)

The next step is concerned with the calculation of the contributions caused by initially infeasible candidate solutions after reflection. Therefore, the marginal density function of the \( z_1 \) components of initially infeasible candidate solutions is derived from the joint density in Eq. (31) yielding

\[
p_1^{ref}(x) = \frac{1}{\sqrt{2 \pi} (1 - \Phi(\delta))} e^{-\frac{1}{2} \delta \cos \theta \cdot x^2} \Phi(h(x)).
\]

(42)

Thus the conditional density of the \( z_2 \) components reads

\[
p_2^{ref}(y|z_1 = x) = \frac{p_1^{ref}(x, y)}{p_1^{ref}(x)} = \frac{1}{\sqrt{2 \pi} \Phi(h(x))} e^{-\frac{1}{2} \delta \sin \theta \cdot y^2},
\]

(43)

and the contributions from initially infeasible and thus reflected offspring to the first moments of \( z_2 \) can be computed as

\[
E \left[ z_2 | z_1 \right] = x \land y \text{ has been reflected} \\
= \frac{1}{\sqrt{2 \pi} \Phi(h(x))} \int_{-\infty}^{\infty} ye^{-\frac{1}{2} \delta \sin \theta \cdot y^2} dy \\
= 2\delta \sin \theta \Phi(h(x)) e^{\frac{1}{2} h(x)^2}.
\]

(44)

and

\[
E \left[ z_2^2 | z_1 \right] = x \land y \text{ has been reflected} \\
= \frac{1}{\sqrt{2 \pi} \Phi(h(x))} \int_{-\infty}^{\infty} y^2 e^{-\frac{1}{2} \delta \sin \theta \cdot y^2} dy \\
= \left( 4\delta^2 \sin^2 \theta + 1 \right) - \frac{4\delta \sin \theta \cdot h(x)}{\sqrt{2 \pi} \Phi(h(x))} e^{\frac{1}{2} h(x)^2}.
\]

(45)

Weighting the terms with the probabilities of their occurrence and applying some simplifications one obtains

\[
E \left[ z_2 | z_1 \right] = x \land y \text{ has not been reflected} \\
+ P_{ref}(x) E \left[ z_2 | z_1 \right] = x \land y \text{ has been reflected} \\
= \frac{1}{p_1^{reflection}(x)} \left[ \frac{1}{2 \pi} e^{-\frac{1}{2} x^2} e^{\frac{1}{2} \frac{\delta \cos \theta}{\sin \theta} y^2} \right] \\
+ \frac{2\delta \sin \theta}{\sqrt{2 \pi} \Phi(h(x))} e^{\frac{1}{2} \delta \cos \theta \cdot y^2} \Phi(h(x)) \\
- \frac{1}{2 \pi} e^{-\frac{1}{2} \delta \sin \theta \cdot y^2} e^{\frac{1}{2} \delta \sin \theta \cdot x^2}.
\]

(46)
and

\[ E[z_2^2 | z_1 = x] = (1 - P_{ref}(x)) E[z_2^2 | z_1 = x \land y \text{ has not been reflected}] + P_{ref}(x) E[z_2^2 | z_1 = x \land y \text{ has been reflected}] \]

\[ = \frac{1}{p_{ref}(x)} \left[ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \delta^2 \phi} \left( \frac{\delta - x \cos \theta}{\sin \theta} \right) \right. \]

\[ \left. - \frac{1}{2\pi} \left( \frac{\delta - x \cos \theta}{\sin \theta} \right) e^{-\frac{1}{2} \left( \frac{\delta - x \cos \theta}{\sin \theta} \right)^2} \right] \]

(47)

for the first two moments about zero of the \( z_2 \) component of mutation vectors after reflection conditional on \( z_1 = x \).

### A.2 Truncation

Now considering repair by truncation, the probability distribution function of the \( z_1 \) components of the mutation vector and the first two moments about zero of its \( z_2 \) components conditional on \( z_1 \) are computed.

#### A.2.1 Distribution of the \( z_1 \) components after truncation

The distribution after truncation also has two components: one from offspring that are initially feasible and one resulting from initially infeasible offspring candidate solutions. Truncation repairs these infeasible offspring by cutting off the mutation vector at the edge of the feasible region of the search space. According to Section 2.2 this results in a transformation of mutation vectors described by

\[
\begin{bmatrix}
  z_1 \\
  z_2
\end{bmatrix} \leftarrow \frac{\delta}{\cos \theta + \sin \theta} \begin{bmatrix}
  z_1 \\
  z_2
\end{bmatrix}.
\]

Notice, the contribution from initially feasible candidate solutions to the distribution of the \( z_1 \) components is independent of the repair method and thus equals the first addend in Eq. (34)

\[ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \delta^2 \phi} \left( \frac{\delta - x \cos \theta}{\sin \theta} \right). \]

(49)

That is, only the derivation of the contribution from truncated offspring to the marginal density of the \( z_1 \) components has to be carried out. The probability of an initially infeasible candidate solution having a \( z_1 \) component less than some \( x \) is the sum

\[ \text{Prob}[z_1 < x] = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{1}{2} (x-y\cos \theta)^2} \, dy \, dx \]

(50)

\[ + \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{1}{2} (y\sin \theta - \delta)^2} \, dx \, dy \]

\[ = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2} \frac{y^2}{\sin \theta}} \phi \left( \frac{y \cos \theta - \delta}{\sin \theta} \right) \, dy \]

(51)

Computing the derivative with respect to \( x \) and applying some straightforward simplifications yields

\[ \frac{d}{dx} \text{Prob}[z_1 < x] = \frac{\delta}{2\pi \sin \theta} \frac{\delta}{x^2} \int_{-\infty}^{+\infty} ye^{-\frac{1}{2} \left( \frac{y - \delta \cos \theta}{\sin \theta} \right)^2} \, dy. \]

(52)

After solving the integral in (52) and introducing the abbreviation

\[ k(x) \coloneqq \left( \delta^2 + \left( \frac{x - \delta \cos \theta}{\sin \theta} \right)^2 \right) \]

(53)
the contribution of initially infeasible offspring after truncation is obtained by
\[ p_{1}^{\text{trunc}}(x) = \frac{\delta}{2\pi \sin \theta} \frac{1}{k(x)} e^{\frac{-x^2}{2\delta}}. \quad (54) \]

Finally, combining (49) and (54) the marginal density of the \( z_1 \) component of mutation vectors after repair by truncation is
\[ p_{1}^{\text{trunc}}(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \Phi \left( \frac{\delta - x \cos \theta}{\sin \theta} \right) + \frac{\delta}{2\pi \sin \theta} \frac{1}{k(x)} e^{\frac{-x^2}{2\delta}}. \quad (55) \]

By integration of (55) with respect to \( x \) and weighting both terms with the probability of their occurrence yields the moments of the contribution of initially infeasible offspring after truncation conditional on \( z_1 = x \) is regarded. The probability of an offspring being in need of repair conditional on \( z_1 = x \) is the relative weight of \( p_{1}^{\text{trunc}}(x) \) in the marginal density (55) and equals
\[ P_{\text{unn}}(x) = \text{Prob} \{ x \text{ has been truncated} | z_1 = x \} \]
\[ = \frac{p_{1}^{\text{trunc}}(x)}{p_{1}^{\text{trunc}}(x)} \]
\[ = \frac{1}{p_{1}^{\text{trunc}}(x)} \frac{\delta}{2\pi \sin \theta} \frac{1}{k(x)} \exp \left( -\frac{1}{2} \right). \quad (57) \]

Initially feasible offspring are not truncated. The moments of their \( z_2 \) components are described by Equations (40) and (41). Initially infeasible candidate solutions after truncation have the \( z_2 \) component \( (\delta - x \cos \theta) / \sin \theta \) and thus
\[ E \{ z_2 | z_1 = x \ \text{and} z \text{has been truncated} \} = \frac{\delta - x \cos \theta}{\sin \theta}, \quad (58) \]
and
\[ E \{ z_2^2 | z_1 = x \ \text{and} z \text{has been truncated} \} = \left( \frac{\delta - x \cos \theta}{\sin \theta} \right)^2. \quad (59) \]

Weighting both terms with the probability of their occurrence yields the moments of the \( z_2 \) components of mutation vectors after truncation conditional on \( z_1 = x \). After some straightforward simplifications these moments are
\[ E \{ z_2 | z_1 = x \} = (1 - P_{\text{unn}}(x))E \{ z_2 | z_1 = x \ \text{and} z \text{has not been truncated} \}
+ P_{\text{unn}}(x)E \{ z_2 | z_1 = x \ \text{and} z \text{has been truncated} \} \]
\[ = \frac{1}{p_{1}^{\text{trunc}}(x)} \left[ \frac{\delta}{2\pi \sin \theta} \frac{1}{k(x)} e^{\frac{-x^2}{2\delta}} \right. \]
\[ - \frac{1}{2\pi} e^{\frac{-x^2}{2}} e^{\frac{-x^2}{2\sin \theta}} \left( \frac{\delta - x \cos \theta}{\sin \theta} \right)^2 \right]. \quad (60) \]
and
\[ E \{ z_2^2 | z_1 = x \} = (1 - P_{\text{unn}}(x))E \{ z_2^2 | z_1 = x \ \text{and} z \text{has not been truncated} \}
+ P_{\text{unn}}(x)E \{ z_2^2 | z_1 = x \ \text{and} z \text{has been truncated} \} \]
\[ = \frac{1}{p_{1}^{\text{trunc}}(x)} \left[ \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \Phi \left( \frac{\delta - x \cos \theta}{\sin \theta} \right) \right. \]
\[ - \frac{1}{2\pi} \left( \frac{\delta - x \cos \theta}{\sin \theta} \right)^2 e^{\frac{-x^2}{2\sin \theta}} e^{\frac{-x^2}{2\sin \theta}} \left( \frac{\delta - x \cos \theta}{\sin \theta} \right)^2 \]
\[ + \frac{\delta}{2\pi \sin \theta} \frac{1}{k(x)} e^{\frac{-x^2}{2\delta}} \right]. \quad (61) \]