
Resampling versus Repair in Evolution Strategies Applied to a Constrained Linear Problem

Dirk V. Arnold

dirk@cs.dal.ca

Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia,
Canada B3H 4R2

doi:10.1162/EVCO_a_00088

Abstract

We study the behaviour of multi-recombination evolution strategies for the problem of maximising a linear function with a single linear constraint. Two variants of the algorithm are considered: a strategy that resamples infeasible candidate solutions and one that applies a simple repair mechanism. Integral expressions that describe the strategies' one-generation behaviour are derived and used in a simple zeroth order model for the steady state attained when operating with constant step size. Applied to the analysis of cumulative step size adaptation, the approach provides an intuitive explanation for the qualitative difference in the algorithm variants' behaviour. The findings have implications for the design of constraint handling techniques to be used in connection with cumulative step size adaptation.

Keywords

Evolution strategies, constrained optimisation, cumulative step size adaptation.

1 Introduction

The number of constraint handling techniques used in evolutionary algorithms is large and growing. Approaches range from penalty methods and repair mechanisms to algorithms based on ideas from multi-objective optimisation, and they include those described by Oyman et al. (1999), Runarsson and Yao (2000, 2005), Mezura-Montes and Coello Coello (2005), and Kramer and Schwefel (2006). A survey of techniques was compiled by Coello Coello (2008). Knowledge with regard to the relative capabilities of the approaches is almost always gained by comparing their performances on sets of test functions that are believed to be difficult. The influence of the choice of those functions as well as of parameters implicit in the performance criteria is often not obvious. The results are not easy to interpret. If a mechanism fails, it often remains unclear why. Likewise, it is not always straightforward what "good" performance for an involved problem means. Test function generators such as the one proposed by Michalewicz et al. (2000) that automatically generate test functions with parameters controlling properties such as the connectedness of the feasible region and the degree of multimodality of the objective provide some extra insights with regard to scaling properties. However, the functions generated are too complex to afford an understanding of algorithm behaviour on a microscopic level.

The approach pursued here is complementary to benchmarking studies on large testbeds. We study the behaviour of algorithms on a very simple class of test functions

for which analytically based findings can be derived. The results offer the advantage of being easy to interpret and reveal scaling properties and the influence of parameters on optimisation performance. It has been known at least since the early work of Schwefel (1975) that the presence of constraints can cause step size adaptation mechanisms to fail even for the simplest of problems. Understanding the interplay of step size adaptation mechanisms and constraint handling techniques on a microscopic level will benefit the developer of evolutionary algorithms for constrained optimisation.

A sizable number of studies pursue a similar approach in the realm of unconstrained evolutionary optimisation. Much less such work has been published in constrained settings. Among the sparse references is the early work by Rechenberg (1973), who studies the performance of the $(1 + 1)$ -ES¹ for the axis-aligned corridor model, and upon which the one-fifth success rule for step size adaptation is partly based. Schwefel (1975) considers the performance of the $(1, \lambda)$ -ES in the same environment. Beyer (1989) analyses the behaviour of the $(1 + 1)$ -ES for a constrained discus-like function. All of those analyses have in common that the normal vectors of the constraint planes are oriented such that they are perpendicular to the gradient vector of the objective function.

More recent work considers constrained problems that do not have that specific property. Arnold and Brauer (2008) study the behaviour of the $(1 + 1)$ -ES for a linear problem with a single linear constraint of general orientation. They describe the distance of the parental candidate solution from the constraint plane and investigate its limit behaviour using two simple statistical models. Schwefel (1975) points out that using the one-fifth rule, the presence of constraints may result in the step size being reduced in situations where the angle between the gradient direction and the normal vector of the active constraint is small, leading to convergence to a nonsingular point. The findings derived by Arnold and Brauer (2008) provide a quantitative confirmation of this. Arnold (2011b) and Arnold (2011a) study the behaviour of the $(1, \lambda)$ -ES with cumulative step size adaptation in the same environment. The former reference considers the case that infeasible candidate solutions are resampled until they are feasible; the latter one uses a simple repair mechanism that orthogonally projects infeasible candidate solutions onto the boundary of the feasible region. It is found that cumulative step size adaptation may fail in scenarios similar to those where the success probability based mechanism fails, though for different reasons, and that employing the repair mechanism results in more robust performance in the sense that convergence to a nonsingular point is avoided, provided that the cumulation parameter used in the step size adaptation mechanism is set appropriately.

In contrast to the work in Arnold (2011a, 2011b), this paper considers the behaviour of evolution strategies employing nonsingleton populations and recombination. The results derived earlier appear as special cases. Being able to study the influence of selection pressure (quantified by the truncation ratio) on algorithm performance adds a new dimension to the analysis. The remainder of the paper is organised as follows. Section 2 briefly formalises the algorithms and problem and introduces notational conventions used throughout the paper. Section 3 investigates the behaviour of the strategy for fixed step size employing a simple zeroth-order model for the distribution of distances from the constraint plane. Section 4 considers cumulative step size adaptation. Section 5 concludes with a brief discussion and directions for future research. Calculations that

¹See Beyer and Schwefel (2002) for an overview of evolution strategy terminology.

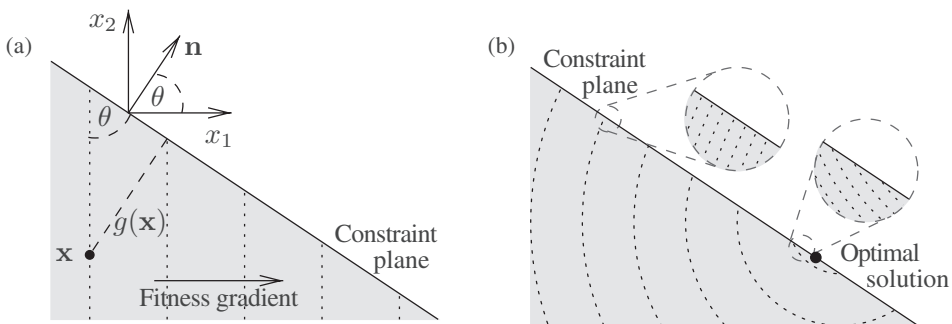


Figure 1: (a) Linear objective function with a single linear constraint. The subspace spanned by the x_1 and x_2 axes is shown. The shaded area is the feasible region. The dotted lines are contour lines of the objective function. The centroid \mathbf{x} of the parental population is at a distance $g(\mathbf{x})$ from the constraint plane. (b) Contour lines of a constrained quadratic function. The local constraint angle decreases as the optimal solution is approached.

underlie Sections 3 and 4 but that are not essential to the understanding of the analysis have been relegated to Appendices A and B in order to improve readability.

2 Problem and Algorithms

Throughout this paper, we consider the problem of maximising² a linear function $f : \mathbb{R}^N \rightarrow \mathbb{R}$, $N \geq 2$, with a single linear constraint. We assume that the gradient vector of the objective function forms an acute angle with the normal vector of the constraint plane. Without loss of generality, we choose a Euclidean coordinate system with its origin located on the constraint plane, and with its axes oriented such that the x_1 axis coincides with the gradient direction ∇f , and the x_2 axis lies in the two-dimensional plane spanned by the gradient vector and the normal vector of the constraint plane. The angle between those two vectors is denoted by θ as illustrated in Figure 1(a), and it is referred to as the constraint angle. Constraint angles of interest are in $(0, \pi/2)$. The unit normal vector of the constraint plane expressed in the coordinate system described above is $\mathbf{n} = (\cos \theta, \sin \theta, 0, \dots, 0)$. The signed distance of a point $\mathbf{x} = \langle x_1, x_2, \dots, x_N \rangle \in \mathbb{R}^N$ from the constraint plane is thus $g(\mathbf{x}) = -\mathbf{n} \cdot \mathbf{x} = -x_1 \cos \theta - x_2 \sin \theta$, resulting in the optimisation problem

$$\text{maximise } f(\mathbf{x}) = \alpha x_1 \text{ subject to } g(\mathbf{x}) \geq 0$$

for some constant $\alpha > 0$. Notice that due to the choice of coordinate system, variables x_3, x_4, \dots, x_N enter neither the objective function nor the constraint inequality. While simple, the linear environment models other constrained optimisation problems in the immediate vicinity of the constraint plane as illustrated in Figure 1(b).

The state of the $(\mu/\mu, \lambda)$ -ES with cumulative step size adaptation is described by population centroid $\mathbf{x} \in \mathbb{R}^N$, mutation strength $\sigma \in \mathbb{R}$, and search path $\mathbf{s} \in \mathbb{R}^N$. The latter is an exponentially fading record of recent steps taken by the strategy. As either

²Strictly speaking, the task is one of amelioration rather than maximisation, as a finite maximum does not exist. We do not make that distinction here.

Algorithm 1 Single iteration of the $(\mu/\mu, \lambda)$ -ES with cumulative step size adaptation.

input: population centroid $\mathbf{x}^{(t)}$, mutation strength $\sigma^{(t)}$, search path $\mathbf{s}^{(t)}$

output: population centroid $\mathbf{x}^{(t+1)}$, mutation strength $\sigma^{(t+1)}$, search path $\mathbf{s}^{(t+1)}$

```

1 for  $i \leftarrow 1$  to  $\lambda$  do
2   if repair then
3      $\mathbf{y}^{(i,t)} \leftarrow \text{sampleNormal}(\mathbf{x}^{(t)}, \sigma^{(t)^2} \mathbf{I}_{N \times N});$ 
4     if  $\mathbf{n} \cdot \mathbf{y}^{(i,t)} > 0$  then
5        $\mathbf{y}^{(i,t)} \leftarrow \mathbf{y}^{(i,t)} - (\mathbf{n} \cdot \mathbf{y}^{(i,t)}) \mathbf{n};$  // project
6     end
7   end
8   else // resample
9     repeat
10       $\mathbf{y}^{(i,t)} \leftarrow \text{sampleNormal}(\mathbf{x}^{(t)}, \sigma^{(t)^2} \mathbf{I}_{N \times N});$ 
11      until  $\mathbf{n} \cdot \mathbf{y}^{(i,t)} \leq 0;$ 
12    end
13    evaluate  $f(\mathbf{y}^{(i,t)})$ ;
14  end
15  order the  $\mathbf{y}^{(i,t)}$  such that  $(k; \lambda)$  is the index of the candidate solution with the
     $k$ th largest objective function value;
16   $\mathbf{x}^{(t+1)} \leftarrow \frac{1}{\mu} \sum_{k=1}^{\mu} \mathbf{y}^{((k;\lambda),t)}$ ;
17   $\mathbf{s}^{(t+1)} \leftarrow (1-c)\mathbf{s}^{(t)} + \sqrt{\mu c(2-c)} \frac{\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}}{\sigma^{(t)}};$ 
18   $\sigma^{(t+1)} \leftarrow \sigma^{(t)} \exp\left(c \frac{\|\mathbf{s}^{(t+1)}\|^2 - N}{2DN}\right);$ 

```

only individual steps or the limit state attained after a large number of iterations is considered, neither initialisation conditions nor termination criteria are relevant for the analysis that follows. The state of the strategy is updated by iterating the steps detailed in Algorithm 1. Superscript t denotes iteration number, while index i numbers

offspring. The loop in lines 1 through 14 implements mutation, resampling, and repair. `sampleNormal(x, Σ)` samples from an N -dimensional normal distribution with mean \mathbf{x} and covariance matrix Σ , and $\mathbf{I}_{N \times N}$ denotes the $N \times N$ identity matrix. Selection and recombination are implemented in lines 15 and 16. Lines 17 and 18 update the search path and mutation strength, respectively, and implement cumulative step size adaptation as proposed by Ostermeier et al. (1994).³ Constant $c \in (0, 1)$ is referred to as the cumulation parameter and determines the effective length of the memory implemented in the search path, and $D > 0$ is a damping constant that determines the magnitude of the mutation strength updates. Clearly, the sign of $\|\mathbf{s}\|^2 - N$ determines whether the mutation strength is increased or decreased. The underlying idea is that long search paths indicate that the steps made by the strategy point predominantly in one direction and could beneficially be replaced with fewer but longer steps. A short search path indicates that the strategy steps back and forth and that it should operate with a smaller step size. The search path has an expected squared length of N if in the unconstrained setting offspring candidate solutions are selected at random (i.e., the ordering of offspring candidate solutions in line 15 of Algorithm 1 is random), in which case no change in mutation strength is effected. It is not unexpected that in constrained settings cumulative step size adaptation may fail under some conditions. When combined with the adaptive encoding mechanism described by Hansen (2008), the algorithm considered here in the unconstrained case in essence yields the CMA-ES as proposed by Hansen and Ostermeier (2001).

3 Behaviour for Fixed Mutation Strength

In this section, we assume that the mutation strength σ of the strategy is fixed. We refer to $\mathbf{z}^{(i,t)} = \langle z_1^{(i,t)}, z_2^{(i,t)}, \dots, z_N^{(i,t)} \rangle = (\mathbf{y}^{(i,t)} - \mathbf{x}^{(t)})/\sigma^{(t)}$ as the mutation vector associated with the i th offspring candidate solution after line 12 in Algorithm 1 (i.e., after resampling or repair). We are interested in the evolution of the normalised distance $\delta = g(\mathbf{x})/\sigma$ of the population centroid from the constraint plane. The update of that distance resulting from a single iteration of Algorithm 1 is described by

$$\delta^{(t+1)} = \delta^{(t)} - z_1^{(\text{avg},t)} \cos \theta - z_2^{(\text{avg},t)} \sin \theta \quad (1)$$

where, letting $(k; \lambda)$ denote the index of the k th best of the λ offspring generated,

$$\mathbf{z}^{(\text{avg},t)} = \frac{1}{\mu} \sum_{k=1}^{\mu} \mathbf{z}^{((k;\lambda),t)}$$

is the average of the mutation vectors corresponding to the offspring candidate solutions selected for survival. From Equation (1),

$$\begin{aligned} \text{Prob}[\delta^{(t+1)} \leq x \mid \delta^{(t)} = y] &= \text{Prob}[y - x \leq z_1^{(\text{avg},t)} \cos \theta + z_2^{(\text{avg},t)} \sin \theta] \\ &= \int_{-\infty}^{\infty} \int_{(y-x-z \cos \theta)/\sin \theta}^{\infty} p_{1,2}^{(\text{avg})}(z, w \mid y) \, dw \, dz \end{aligned}$$

³The update in Algorithm 1 differs from the original prescription in that Ostermeier et al. (1994) adapt the mutation strength based on the length of the search path rather than on its squared length. With appropriately chosen parameters, both variants behave similarly. Basing the update on the squared length simplifies the analysis.

where $p_{1,2}^{(avg)}$ is the (time independent) joint density of the first and second components $z_1^{(avg,t)}$ and $z_2^{(avg,t)}$ of $\mathbf{z}^{(avg,t)}$ conditional on the normalised distance of the population centroid from the constraint plane. Using Leibniz's integral rule to compute the derivative with respect to x yields

$$p_\delta(x | y) = \frac{d}{dx} \text{Prob}[\delta^{(t+1)} \leq x | \delta^{(t)} = y] \\ = \frac{1}{\sin \theta} \int_{-\infty}^{\infty} p_{1,2}^{(avg)} \left(z, \frac{y - x - z \cos \theta}{\sin \theta} \right) dz$$

for the probability density of the normalised distance from the constraint plane after a step conditional on that distance before the step. The probability density of the normalised distance from the constraint plane at time step $t + 1$ is thus

$$p_\delta^{(t+1)}(x) = \int_0^\infty p_\delta^{(t)}(y) p_\delta(x | y) dy \tag{2}$$

where $p_\delta^{(t)}$ is the density of δ values at time step t . As a result of iterating the update in Equation (1), the distribution of normalised distances from the constraint plane approaches a stationary limit distribution the density of which is a fixed point of the mapping in Equation (2).

An approach to computing approximations to such limit distributions that has been used extensively by Beyer (2001) is to expand the unknown distribution in terms of its moments, compute the resulting moments after an update, and impose stationarity conditions demanding that the moments remain unchanged, resulting in a system of as many equations as there are moments included in the calculations. A zeroth-order approximation considers the first moment of the distribution only, effectively approximating it with a Dirac distribution. Solving the corresponding stationarity condition

$$E[\delta^{(t+1)}] = \delta^{(t)} \tag{3}$$

for $\delta = \delta^{(t)}$ yields an approximation to the average normalised distance of the population centroid from the constraint plane. Better approximations can be obtained by considering higher-order moments. Arnold and Brauer (2008) use an exponential model for the distribution in the case of the (1 + 1)-ES and find that it improves the quality of the approximation. However, the Dirac model will be seen to be sufficient for explaining the qualitative difference in the behaviours of the two constraint handling mechanisms for the $(\mu/\mu, \lambda)$ -ES considered here.

Taking the expected value of Equation (1) and using the Dirac stationarity condition from Equation (3) yields

$$E[z_1^{(avg,t)}] \cos \theta + E[z_2^{(avg,t)}] \sin \theta = 0. \tag{4}$$

Expressions for the expected values that make it possible to numerically solve for $\delta = \delta^{(t)}$ are computed in Appendices A and B, with the outcome of the calculations given in Equations (19) and (20). Figure 2 illustrates the results. Parts (a) and (b) of the figure plot the average normalised distance of the population centroid from the constraint plane against the constraint angle. The lines were obtained by solving Equation (4) for δ . Larger constraint angles result in the constraint plane being tracked more loosely. Unsurprisingly, the strategy that repairs infeasible candidate solutions operates closer to the constraint plane than the one that resamples them. Comparing the lines that

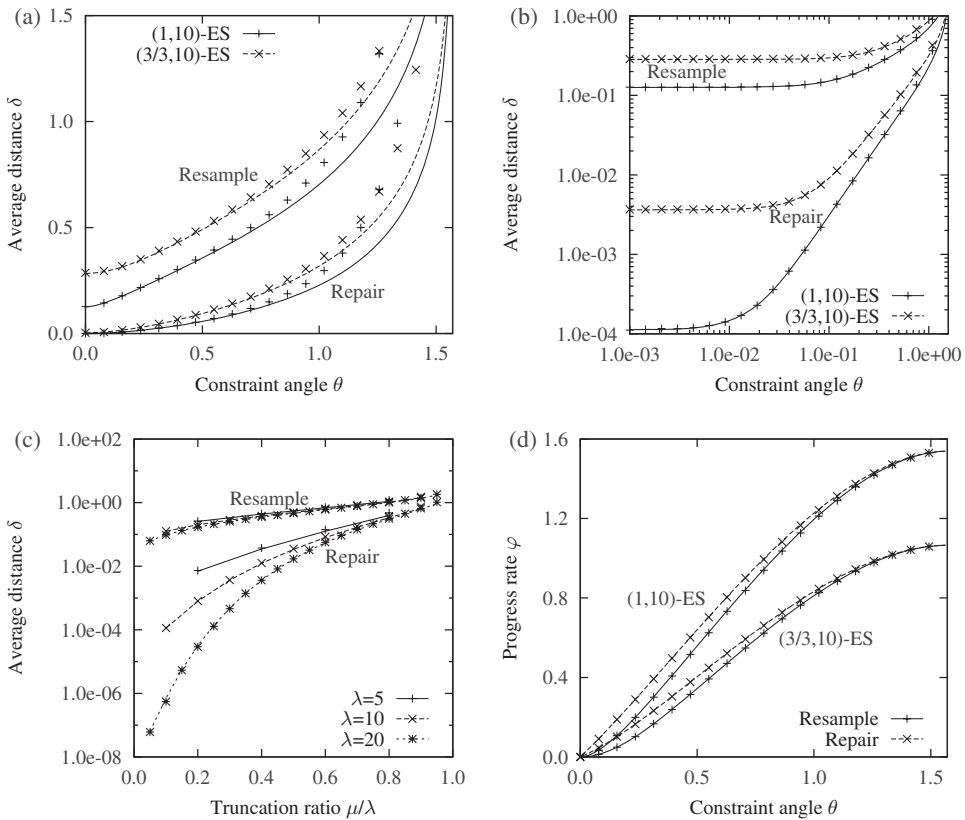


Figure 2: Behaviour of $(\mu/\mu, \lambda)$ -ES for fixed step size. (a) Average normalised distance δ of the population centroid from the constraint plane plotted against the constraint angle θ . (b) The same information plotted using logarithmic scales for both axes. (c) Average normalised distance δ from the constraint plane plotted against the truncation ratio μ/λ for $\theta = 0$. (d) Normalised progress rate φ plotted against constraint angle θ . All lines have been obtained using the Dirac model. All points mark measurements obtained by averaging over 10^7 steps in runs of the respective evolution strategies with fixed mutation strength.

were obtained using the Dirac model with the points that mark measurements made in runs of the respective evolution strategies with fixed mutation strength, it can be seen that the quality of the approximation deteriorates with increasing constraint angle. With increasing distance from the constraint plane, the variance of δ values increases, rendering the Dirac model, which ignores those variations, less appropriate. Figure 2(c) considers the values attained for constraint angle $\theta = 0$ and shows how the average normalised distance from the constraint plane varies with population size parameters μ and λ in that case. Increasing μ for λ fixed results in the constraint plane being tracked more loosely while increasing λ for fixed truncation ratio μ/λ leads to the strategy operating closer to the constraint plane. Finally, Figure 2(d) plots the normalised progress rate

$$\varphi = E[z_1^{(\text{avg})}]$$

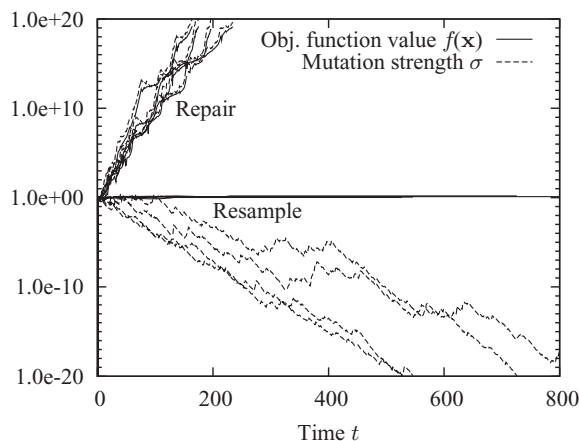


Figure 3: Objective function value $f(\mathbf{x})$ of the population centroid and mutation strength σ plotted against time step t . Shown are four runs of the (3/3, 10)-ES with cumulative step size adaptation for each of the two constraint handling approaches.

which quantifies the average size of the step in the direction of the gradient of the objective function per iteration of the strategy, against the constraint angle. It can be seen (1) that the normalised progress rate increases with increasing constraint angle, (2) that the strategy that selects only the best of the offspring progresses faster than the one that recombines several best, and (3) that the strategy that resamples infeasible candidate solutions is somewhat slower than the one that employs repair.

4 Mutation Strength Adaptation

We now abandon the assumption that the mutation strength of the strategy is fixed. For the constrained linear environment considered here, the $(\mu/\mu, \lambda)$ -ES with cumulative step size adaptation does not assume a stationary limit state. The mutation strength is either increased or decreased indefinitely. As decreasing mutation strengths lead to stagnation and convergence to a nonsingular point, increasing σ is desirable. Figure 3 shows four runs of the (3/3, 10)-ES with cumulative step size adaptation for each of the two constraint handling approaches. In all cases $\theta = 0.0873 = 5^\circ$, $N = 2$, $c = 0.5$, $\mathbf{s}^{(0)} = \mathbf{0}$, and $\delta^{(0)} = \sigma^{(0)} = D = \alpha = 1.0$. While the mutation strength and objective function value of the population centroid increase exponentially quickly for the strategy that uses the repair mechanism, the strategy that employs the resampling approach decreases its mutation strength exponentially quickly and objective function values stagnate. It will be seen below that a lower setting for the cumulation parameter would enable the resampling strategy to avoid stagnation.

The $(\mu/\mu, \lambda)$ -ES with cumulative step size adaptation applied to the simple constrained linear optimisation problem considered here forms a nonlinear, stochastic dynamic system the state of which includes δ , σ , and variables describing the search path \mathbf{s} (at least $\|\mathbf{s}\|^2$ and components s_1 and s_2). Rather than attempting to solve the difficult problem of examining its dynamic behaviour, we consider the logarithmic adaptation response

$$\Delta_{\sigma}^{(t+1)} = \frac{2DN}{c} \log \left(\frac{\sigma^{(t+1)}}{\sigma^{(t)}} \right)$$

of the strategy operating out of a stationary state. A positive logarithmic adaptation response indicates that the mutation strength is increased; a negative logarithmic adaptation response implies a decrease in step size. The expected logarithmic adaptation response depends on the normalised distance of the parental centroid from the constraint plane as well as on the search path. In order to predict whether the strategy will converge or diverge, we assume that it has been run with a fixed step size until time step t , where t is large enough for the strategy to operate in the limit state considered in Section 3, and we compute the expected logarithmic adaptation response at time step $t + 1$. Experiments show that the resulting predictions of the strategy's dynamic behaviour are accurate across a wide range of parameter settings.

Assuming that the search path is initialised as $\mathbf{s}^{(0)} = \mathbf{0}$, from line 17 in Algorithm 1 it follows that the search path at time step $t + 1$ can be expressed as

$$\mathbf{s}^{(t+1)} = \sqrt{\mu c(2 - c)} \sum_{j=0}^t (1 - c)^j \mathbf{z}^{(\text{avg}, t-j)}.$$

Its expected squared length

$$E[\|\mathbf{s}^{(t+1)}\|^2] = \sum_{i=1}^N E[s_i^{(t+1)2}] \tag{5}$$

can be computed from the squares

$$s_i^{(t+1)2} = \mu c(2 - c) \left[\sum_{j=0}^t (1 - c)^{2j} z_i^{(\text{avg}, t-j)2} + 2 \sum_{j=1}^t \sum_{k=0}^{j-1} (1 - c)^{j+k} z_i^{(\text{avg}, t-j)} z_i^{(\text{avg}, t-k)} \right]$$

of its components. Due to the assumption that the strategy is operating out of a stationary state, the distribution of the $z_i^{(\text{avg}, t)}$ is independent of time. Omitting time superscripts, using the identities

$$\sum_{j=0}^t a^{2j} = \frac{1 - a^{2(t+1)}}{1 - a^2}$$

and

$$\begin{aligned} \sum_{j=1}^t \sum_{k=0}^{j-1} a^{j+k} &= \sum_{j=1}^t a^j \frac{1 - a^j}{1 - a} \\ &= \frac{1}{1 - a} \left[\sum_{j=0}^t a^j - 1 - \sum_{j=0}^t a^{2j} + 1 \right] \\ &= \frac{1 - a^{t+1}}{(1 - a)^2} - \frac{1 - a^{2(t+1)}}{(1 - a)(1 - a^2)} \end{aligned}$$

for the geometric series, and for brevity writing

$$e_{i,j} = E[z_i^{(\text{avg})j}]$$

for the j th moment about zero of $z_i^{(\text{avg})}$ yields

$$s_i^2 = \mu e_{i,2} + 2\mu \frac{1 - c}{c} e_{i,1}^2$$

in the limit of large t . Components z_3, \dots, z_N of mutation vectors are independently standard normally distributed and thus the $z_i^{(avg)}$ for $i > 2$ are independently normally distributed with mean zero and variance $1/\mu$. Thus, $e_{i,1} = 0$ and $e_{i,2} = 1/\mu$ for $i = 3, 4, \dots, N$. As a result, from Equation (5) it follows that

$$E[\Delta_\sigma^{(t+1)}] = \mu (e_{1,2} + e_{2,2}) + 2\mu \frac{1-c}{c} (e_{1,1}^2 + e_{2,1}^2) - 2. \tag{6}$$

Note that this result depends on the search space dimensionality only through the cumulation parameter c , which is usually set to an N -dependent value.

Equations for $e_{i,j}$ for $i, j \in \{1, 2\}$ are derived in Appendices A and B. Numerically evaluating Equations (19), (20), (21), and (22) using the Dirac model for the average distance of the population centroid from the constraint plane yields the curves in Figure 4. Figure 4(a) shows the same results as Figure 2(d), but with logarithmic scales for both axes. Comparison of the curves with the values obtained experimentally in runs of the respective evolution strategies suggests good visual agreement for the first moments $e_{1,1}$ and $e_{2,1}$. The quality of the approximation relying on the Dirac model is not as good for moments $e_{1,2}$ when constraint angles are highly acute. However, the values in that scenario are generally small and have limited impact on the results below. The agreement of values for $e_{2,2}$ derived based on the Dirac model with experimental results deteriorates with increasing constraint angle but appears visually good for small values of θ .

The expected logarithmic adaptation response computed in Equation (6) decreases with increasing cumulation parameter c . For a given constraint angle θ and values of c close enough to zero, the term that involves the squared first moments is large enough for the sum of the first two terms to outweigh the negative third term, ensuring positive expected logarithmic adaptation response and thus an expected increase in the step size. As c approaches 1.0, however, the second term in Equation (6) decreases to zero. Depending on the magnitude of the first term, the expected logarithmic adaptation response may be negative, resulting in decreasing step sizes. Demanding that $E[\Delta_\sigma]$ be nonnegative yields the condition

$$c \leq \frac{e_{1,1}^2 + e_{2,1}^2}{1/\mu + e_{1,1}^2 + e_{2,1}^2 - (e_{1,2} + e_{2,2})/2} \tag{7}$$

for the cumulation parameter. The maximum value of c for which convergence to a nonstationary point is avoided (obtained by considering equality in Equation (7), where the $e_{i,j}$ are computed using the Dirac model for the average distance from the constraint plane) is illustrated in Figure 5. It can be seen from Figures 5(a) that for the strategy that resamples infeasible offspring candidate solutions, that value decreases to zero as the constraint angle becomes increasing acute. As a result, the strategy will converge to a nonstationary limit point, no matter what the value of c is. The strategy that repairs infeasible candidate solutions avoids this type of behaviour for any $\theta > 0$, as the maximum value of c for which convergence is avoided tends toward a finite limit value. That limit value is shown in Figure 5(b) for several values of the population size parameters μ and λ . As $c \in (0, 1]$, values exceeding 1.0 are not shown. It can be seen that for a fixed truncation ratio, larger values of λ allow larger values for c and thus faster adaptation. Only for truncation ratios exceeding one-half do admissible maximum values of c become very small.

An intuitive explanation for the failure of cumulative step size adaptation for small constraint angles in the strategy that resamples infeasible offspring candidate solutions

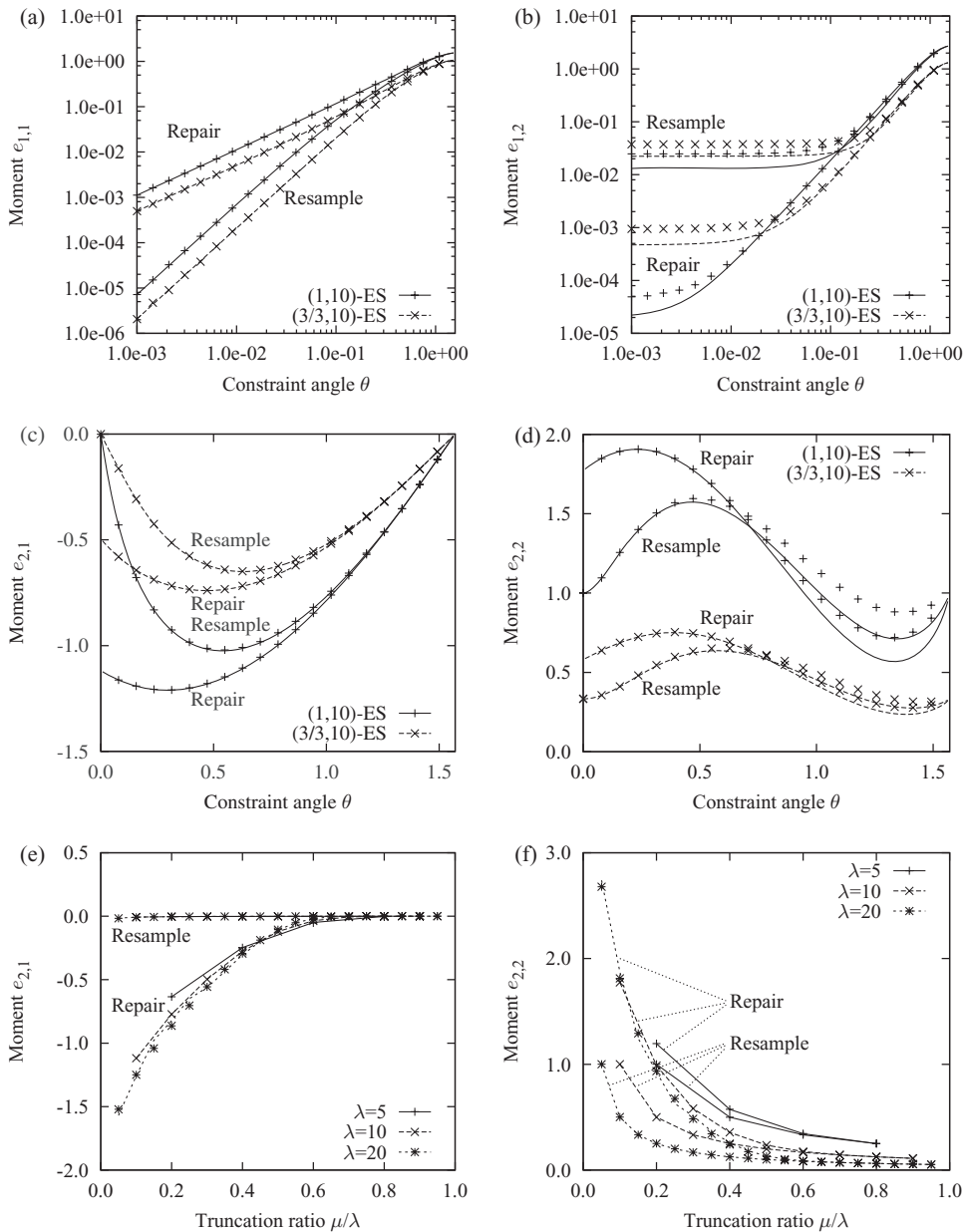


Figure 4: Moments of the first and second components of $\mathbf{z}^{(avg)}$. (a) through (d) plot $e_{i,j}$ against the constraint angle θ for $i, j \in \{1, 2\}$. (e) and (f) plot the limit values of $e_{2,j}$ for $\theta \rightarrow 0$ against the truncation ratio μ/λ for $j \in \{1, 2\}$ (with the discrete data points connected by lines for clarity). All points mark measurements obtained by averaging over 10^7 steps in runs of the respective evolution strategies with fixed mutation strength. The runs depicted in (e) and (f) use $\theta = 10^{-3}$ and are visually indistinguishable from runs with smaller constraint angles.

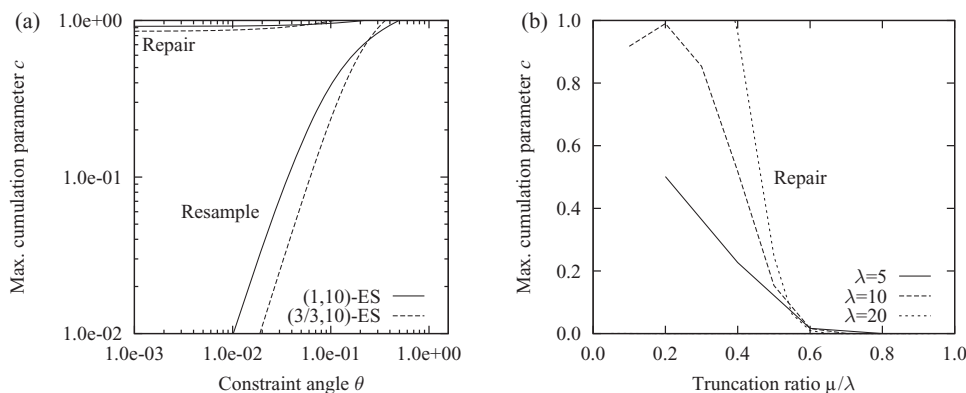


Figure 5: (a) Maximum value of the cumulation parameter c for which the expected logarithmic adaptation response is positive plotted against the constraint angle θ . (b) Limit values for $\theta \rightarrow 0$ of the maximum value of the cumulation parameter c for which the expected logarithmic adaptation response is positive plotted against the truncation ratio μ/λ . For clarity, lines have been added to connect the discrete data points. Data points for the strategy that resamples infeasible offspring are identically zero.

can be derived from Figure 4. In the limit of very small constraint angles, both $e_{1,1}$ and $e_{1,2}$ assume very small values for both strategies as both stay in the immediate vicinity of the constraint plane (compare Figures 2(b) and 4(a) and 4(b)). In order to have positive expected logarithmic adaptation response, the terms in Equation (6) that involve the moments $e_{2,1}$ and $e_{2,2}$ of the step in the direction of the x_2 axis must be large enough to compensate for the lack of a contribution to the expected squared length of the search path from the components in the direction of the x_1 axis. The strategy that resamples infeasible offspring in the limit of very small constraint angles performs an almost random walk in the direction of the x_2 axis, resulting in $e_{2,1} \approx 0.0$ and $e_{2,2} \approx 1/\mu$ (compare Figures 4(e) and 4(f)), which is insufficient to ensure positive expected logarithmic adaptation response unless c is small enough to sufficiently amplify the term involving $e_{1,1}^2$ in Equation (6).

The strategy that repairs infeasible offspring does not perform a random walk in the direction of the x_2 axis. In the limit of very small θ , on average, half of the offspring are infeasible and will be projected onto the constraint plane. On average, half of those that are projected have a negative z_2 component and are likely to prevail under selection as their z_1 components are positive. Unless λ is very small or the truncation ratio is large, the majority of mutation vectors being averaged through recombination will thus point in the direction of the negative x_2 axis and lead to relatively large values of $e_{2,1}^2$ and $e_{2,2}$ that ensure positive expected logarithmic adaptation response (compare Figures 4(e) and 4(f)). It is important to note that this is not a consequence of a bias built into the step size adaptation mechanism as a result of changing the length of individual steps through projection. It is easy to see that projection of mutation vectors never increases their length, and that thus any bias as a result of projection will be toward smaller step sizes, not toward larger ones. Positive expected logarithmic adaptation response is a result of the coherence of the selected steps in the direction of the negative x_2 axis, not the length of the mutation vectors.

Finally, the accuracy of the predictions made on the basis of the Dirac model remain to be verified. The lines in Figure 6 were obtained from Equation (7) for several

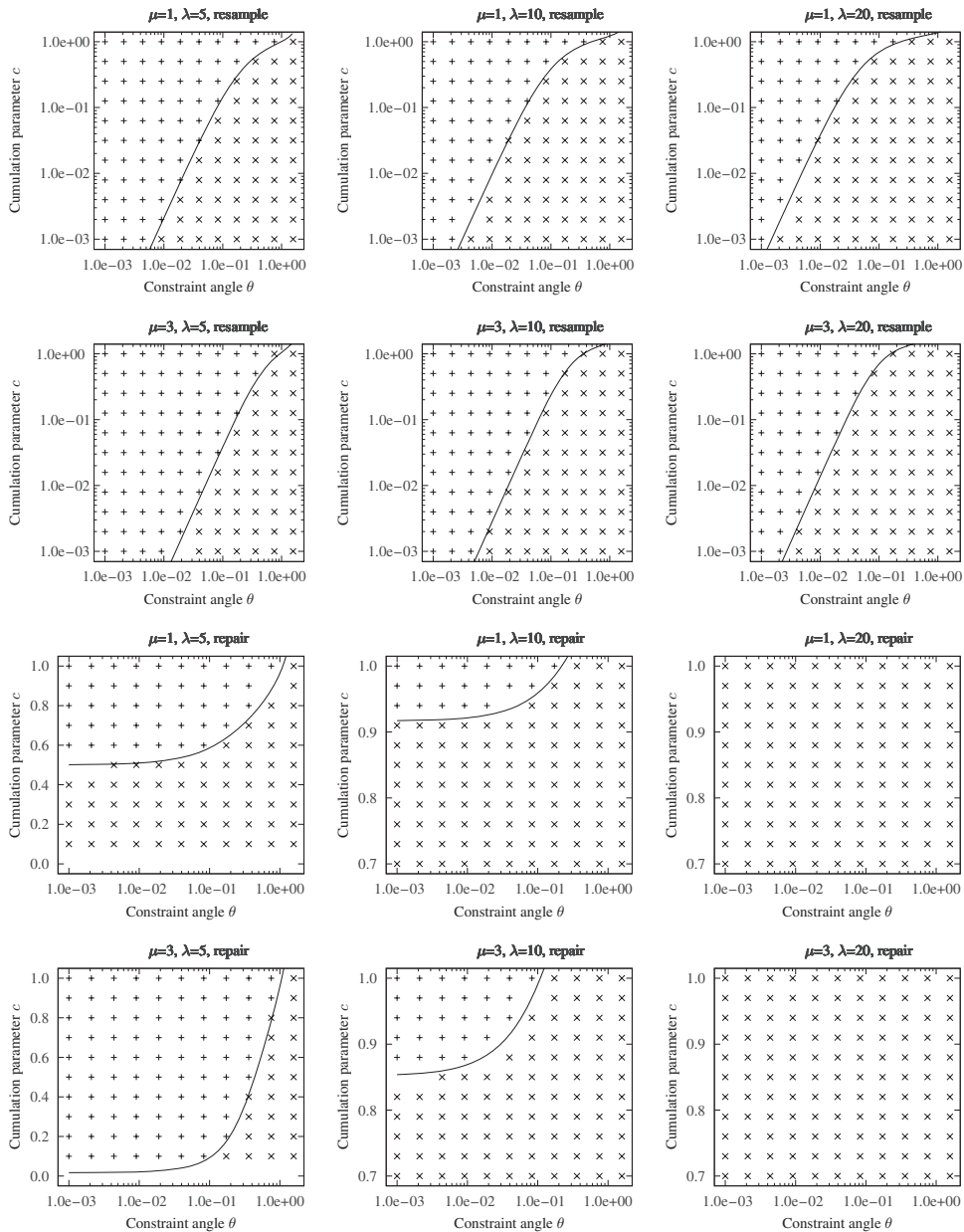


Figure 6: Experimental verification of the convergence behaviour. The lines represent analytically obtained maximum values of the cumulation parameter c for which convergence to a nonstationary point is avoided; plotted against the constraint angle θ . The dots were obtained from runs of the respective $(\mu/\mu, \lambda)$ -ES with cumulative step size adaptation as described in the text. The + symbol marks the cases where the mutation strength is decreased and the \times symbol marks those cases where it is increased.

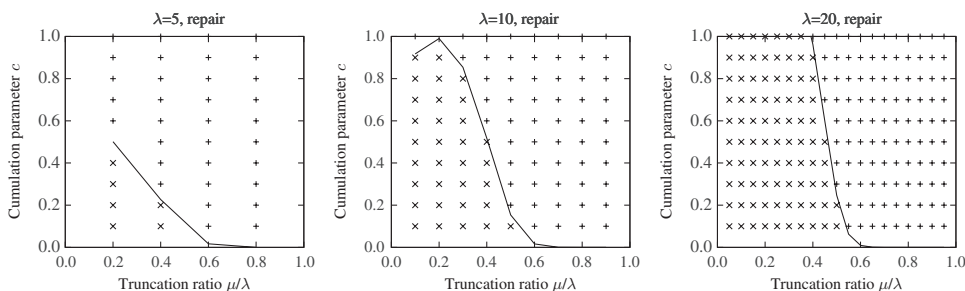


Figure 7: Experimental verification of the convergence behaviour. The lines represent analytically obtained maximum values of the cumulation parameter c for which convergence to a nonstationary point is avoided in the limit $\theta \rightarrow 0$ plotted against the truncation ratio μ/λ . The dots were obtained from runs of the respective $(\mu/\mu, \lambda)$ -ES with cumulative step size adaptation and $\theta = 10^{-3}$ as described in the text. The + symbol marks the cases where the mutation strength is decreased and the \times symbol marks those cases where it is increased.

population size parameter settings. For $\lambda = 20$ and constraint handling by repair, the values computed consistently exceed 1.0, resulting in no curves in the respective plots. Figure 7 shows limit values for $\theta \rightarrow 0$ obtained from the same equation. Each dot in either figure marks the results from 10 independent runs of the respective evolution strategies with cumulative step size adaptation. Each run was initialised with $\delta^{(0)} = 1.0$, $\sigma^{(0)} = 1.0$, and $\mathbf{s}^{(0)} = \mathbf{0}$ and uses $N = 2$ and $D = 1.0$. We ran further experiments with other settings for the search space dimensionality and damping parameter and observed virtually identical results. A run was terminated when the mutation strength reached a value either smaller than 10^{-20} or greater than 10^{20} . The very small step sizes in the former case suggest that stagnation is likely to ensue while the large step sizes in the latter case indicate continued progress at increasingly faster rates. The + symbol in the plots indicates that at least nine out of the 10 runs terminated with $\sigma < 10^{-20}$; the \times symbol indicates that $\sigma > 10^{20}$ upon termination in at least nine out of the 10 runs. If neither symbol is present at a grid location in the plots, then either termination criterion was satisfied in at least two out of the 10 runs. It can be seen that the agreement of predictions based on Equation (7) with experimentally observed values is very good across the entire range of population size parameters and constraint angles considered.

5 Summary and Conclusions

We studied the behaviour of the $(\mu/\mu, \lambda)$ -ES with cumulative step size adaptation applied to a linear problem with a single linear inequality constraint. Two constraint handling techniques were considered: the resampling of infeasible candidate solutions, and their repair through projection onto the constraint plane. We have argued that while it is simple, the scenario of a linear function with a single linear constraint is an interesting one to consider as it reflects microscopic properties of more general constrained, differentiable problems.

Using a simple model, we derived an approximation to the average distance of the population centroid from the constraint plane if the strategy is run with fixed mutation strength. We then used that approximation in the context of adaptive step sizes and derived an expression for the expected logarithmic adaptation response of the strategy

if it operates out of a stationary state. Experiments showed that predictions of whether the strategy is capable of sustained progress or converges to a nonsingular point can be made using those results across a wide range of population size parameter settings and constraint angles. Most interestingly, the findings revealed a qualitative difference between the behaviours resulting from the two constraint handling mechanisms for highly acute constraint angles. While the strategy that resamples infeasible candidate solutions will converge to a nonsingular point for any setting of the cumulation parameter c , the strategy that employs a simple repair mechanism is capable of sustained progress as long as that parameter is set to not too large a value as governed by Equation (7). The difference in the behaviour of the two strategy variants has been explained by considering the components of the search path in the two-dimensional subspace spanned by the gradient of the objective function and the normal vector of the constraint plane. For small constraint angles, both constraint handling mechanisms result in very small steps in the direction of the gradient vector being made. However, while for the strategy that resamples infeasible offspring steps in the perpendicular direction become increasingly random as θ decreases, they tend to point in the same direction for any $\theta > 0$ for the strategy that applies a repair mechanism, resulting in longer search paths and consequently in a systematic increase of the step size.

Clearly, these results have strong implications for the choice of constraint handling approach to be used in combination with cumulative step size adaptation. They also provide guidance to the algorithm engineer developing new constraint handling techniques, suggesting that if step size adaptation is to be based on the (squared) length of the search path, the impact of constraint handling on that path needs to be considered. It is desirable to extend the analyses presented here to investigate properties of further constraint handling techniques as well as of other step size adaptation mechanisms, such as mutative self-adaptation. Future work will also consider nonlinear objective functions where the local constraint angle is not constant, nonlinear constraints, and problems with multiple constraints.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

- Arnold, D. V. (2002). *Noisy optimization with evolution strategies*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Arnold, D. V. (2011a). Analysis of a repair mechanism for the $(1, \lambda)$ -ES applied to a simple constrained problem. In *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2011*, pp. 853–860.
- Arnold, D. V. (2011b). On the behaviour of the $(1, \lambda)$ -ES for a simple constrained problem. In H.-G. Beyer and W. B. Langdon (Eds.), *Proceedings of Foundations of Genetic Algorithms, FOGA 2011*, pp. 15–24.
- Arnold, D. V., and Brauer, D. (2008). On the behaviour of the $(1 + 1)$ -ES for a simple constrained problem. In G. Rudolph et al. (Eds.), *Parallel Problem Solving from Nature, PPSN X*, pp. 1–10. Berlin: Springer Verlag.
- Beyer, H.-G. (1989). *Ein Evolutionsverfahren zur mathematischen Modellierung stationärer Zustände in dynamischen Systemen*. PhD thesis, Hochschule für Architektur und Bauwesen, Weimar.

- Beyer, H.-G. (2001). *The theory of evolution strategies*. Berlin: Springer Verlag.
- Beyer, H.-G., and Schwefel, H.-P. (2002). Evolution strategies—A comprehensive introduction. *Natural Computing*, 1(1): 3–52.
- Coello Coello, C. A. (2008). Constraint-handling techniques used with evolutionary algorithms. In *Conference Companion to the Conference on Genetic and Evolutionary Computation, GECCO'08*, pp. 2445–2466.
- Hansen, N. (2008). Adaptive encoding: How to render search coordinate system invariant. In G. Rudolph et al. (Eds.), *Parallel Problem Solving from Nature, PPSN X*, pp. 205–214.
- Hansen, N., and Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2): 159–195.
- Kramer, O., and Schwefel, H.-P. (2006). On three new approaches to handle constraints within evolution strategies. *Natural Computing*, 5(4): 363–385.
- Mezura-Montes, E., and Coello Coello, C. A. (2005). A simple multi-membered evolution strategy to solve constrained optimization problems. *IEEE Transactions on Evolutionary Computation*, 9(1): 1–17.
- Michalewicz, Z., Deb, K., Schmidt, M., and Stidsen, T. J. (2000). Test-case generator for nonlinear continuous parameter optimization techniques. *IEEE Transactions on Evolutionary Computation*, 4(3): 197–215.
- Ostermeier, A., Gawelczyk, A., and Hansen, N. (1994). Step-size adaptation based on non-local use of selection information. In Y. Davidor et al. (Eds.), *Parallel Problem Solving from Nature, PPSN III*, pp. 189–198.
- Oyman, A. I., Deb, K., and Beyer, H.-G. (1999). An alternative constraint handling method for evolution strategies. In *Proceedings of the 1999 IEEE Congress on Evolutionary Computation*, pp. 612–619.
- Rechenberg, I. (1973). *Evolutionsstrategie—Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Friedrich Frommann Verlag.
- Runarsson, T. P., and Yao, X. (2000). Stochastic ranking for constrained evolutionary optimization. *IEEE Transactions on Evolutionary Computation*, 4(3): 274–283.
- Runarsson, T. P., and Yao, X. (2005). Search biases in constrained evolutionary optimization. *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews*, 35(2): 233–243.
- Schwefel, H.-P. (1975). *Evolutionsstrategie und numerische Optimierung*. Ph.D. Dissertation, Technische Universität Berlin, Fachbereich Verfahrenstechnik.

Appendix A Single Step Behaviour: Mutation

This appendix computes the marginal density of the z_1 components as well as the first two moments about zero of the z_2 components of mutation vectors for both constraint handling techniques. Both will be used in Appendix B in order to compute the corresponding moments after selection and averaging. Superscripts denoting time or indexing offspring candidate solutions are omitted as only individual offspring are considered.

A.1 Resampling

An offspring candidate solution $\mathbf{y} = \mathbf{x} + \sigma \mathbf{z}$ is feasible if and only if $\delta \geq z_1 \cos \theta + z_2 \sin \theta$, where $\delta = g(\mathbf{x})/\sigma$ refers to the normalised distance of the centroid of the parental population from the constraint plane. Thus, the joint probability density of the z_1 and

z_2 components for the strategy that resamples infeasible offspring is a truncated normal density described by

$$p_{1,2}^{\text{resample}}(x, y) = \begin{cases} \frac{1}{2\pi \Phi(\delta)} e^{-\frac{1}{2}(x^2+y^2)} & \text{if } \delta \geq x \cos \theta + y \sin \theta \\ 0 & \text{otherwise} \end{cases}$$

where $\Phi(x) = \int_{-\infty}^x e^{-z^2/2} dz / \sqrt{2\pi}$ denotes the cumulative distribution function of the standard normal distribution. The normalising factor in the denominator is the probability

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{(\delta-x \cos \theta) / \sin \theta} e^{-\frac{1}{2}(x^2+y^2)} dy dx = \Phi(\delta)$$

of an offspring being feasible. The marginal density of the z_1 component of feasible mutation vectors is thus

$$\begin{aligned} p_1^{\text{resample}}(x) &= \int_{-\infty}^{\infty} p_{1,2}^{\text{resample}}(x, y) dy \\ &= \frac{1}{\sqrt{2\pi} \Phi(\delta)} e^{-\frac{1}{2}x^2} \Phi\left(\frac{\delta - x \cos \theta}{\sin \theta}\right). \end{aligned} \tag{8}$$

The corresponding cumulative distribution function

$$P_1^{\text{resample}}(x) = \frac{1}{\sqrt{2\pi} \Phi(\delta)} \int_{-\infty}^x e^{-\frac{1}{2}y^2} \Phi\left(\frac{\delta - y \cos \theta}{\sin \theta}\right) dy \tag{9}$$

cannot be expressed in closed form.

The first two moments about zero of the z_2 component conditional on the z_1 component can be computed from conditional probability density

$$p_2^{\text{resample}}(y|z_1 = x) = \frac{p_{1,2}^{\text{resample}}(x, y)}{p_1^{\text{resample}}(x)}$$

as

$$\begin{aligned} E[z_2|z_1 = x] &= \int_{-\infty}^{\infty} y p_2^{\text{resample}}(y|z_1 = x) dy \\ &= \frac{1}{2\pi \Phi(\delta) p_1^{\text{resample}}(x)} e^{-\frac{1}{2}x^2} \int_{-\infty}^{(\delta-x \cos \theta) / \sin \theta} y e^{-\frac{1}{2}y^2} dy \\ &= -\frac{1}{2\pi \Phi(\delta) p_1^{\text{resample}}(x)} e^{-\frac{1}{2}x^2} \exp\left(-\frac{1}{2} \left(\frac{\delta - x \cos \theta}{\sin \theta}\right)^2\right) \end{aligned} \tag{10}$$

and

$$\begin{aligned} E[z_2^2|z_1 = x] &= \int_{-\infty}^{\infty} y^2 p_2^{\text{resample}}(y|z_1 = x) dy \\ &= \frac{1}{2\pi \Phi(\delta) p_1^{\text{resample}}(x)} e^{-\frac{1}{2}x^2} \int_{-\infty}^{(\delta-x \cos \theta) / \sin \theta} y^2 e^{-\frac{1}{2}y^2} dy \\ &= 1 - \frac{\delta - x \cos \theta}{2\pi \Phi(\delta) \sin \theta p_1^{\text{resample}}(x)} e^{-\frac{1}{2}x^2} \exp\left(-\frac{1}{2} \left(\frac{\delta - x \cos \theta}{\sin \theta}\right)^2\right) \end{aligned} \tag{11}$$

respectively.

A.2 Repair

The distribution of offspring for the strategy that repairs infeasible candidate solutions rather than resampling them has two components: one from offspring that are immediately feasible and one from those that are initially infeasible and repaired by projection onto the constraint plane. Considering the latter, repair as described in line 5 of Algorithm 1 results in a transformation of mutation vectors according to

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \leftarrow \delta \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} + \begin{bmatrix} \sin^2 \theta & -\sin \theta \cos \theta \\ -\sin \theta \cos \theta & \cos^2 \theta \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}.$$

The probability of an initially infeasible candidate solution having a z_1 component less than some x is obtained by integration of the initial density as

$$\begin{aligned} \text{Prob}[z_1 < x] &= \int_{-\infty}^{\infty} \int_{(\delta - y \cos \theta) / \sin \theta}^{\infty} \frac{e^{-\frac{1}{2}(y^2 + z^2)}}{2\pi(1 - \Phi(\delta))} \mathbb{1}_{\delta \cos \theta + y \sin^2 \theta - z \sin \theta \cos \theta < x} \, dz \, dy \\ &= \int_{-\infty}^x \int_{(\delta - y \cos \theta) / \sin \theta}^{\infty} \frac{e^{-\frac{1}{2}(y^2 + z^2)}}{2\pi(1 - \Phi(\delta))} \, dz \, dy \\ &\quad + \int_x^{\infty} \int_{(\delta \cos \theta + y \sin^2 \theta - x) / (\sin \theta \cos \theta)}^{\infty} \frac{e^{-\frac{1}{2}(y^2 + z^2)}}{2\pi(1 - \Phi(\delta))} \, dz \, dy \\ &= \frac{1}{\sqrt{2\pi}(1 - \Phi(\delta))} \int_{-\infty}^x e^{-\frac{1}{2}y^2} \left[1 - \Phi\left(\frac{\delta - y \cos \theta}{\sin \theta}\right) \right] \, dy \\ &\quad + \frac{1}{\sqrt{2\pi}(1 - \Phi(\delta))} \int_x^{\infty} e^{-\frac{1}{2}y^2} \left[1 - \Phi\left(\frac{\delta \cos \theta + y \sin^2 \theta - x}{\sin \theta \cos \theta}\right) \right] \, dy \end{aligned}$$

where $\mathbb{1}_{\text{cond}}$ equals one if cond holds true and zero otherwise. Computing the derivative with respect to x yields after some straightforward simplifications

$$\frac{d}{dx} \text{Prob}[z_1 < x] = \frac{1}{\sqrt{2\pi} \sin \theta} \exp\left(-\frac{1}{2} \left(\frac{\delta \cos \theta - x}{\sin \theta}\right)^2\right).$$

Together with the contribution $\Phi(\delta)p_1^{\text{resample}}(x)$ from the initially feasible offspring, the marginal density of z_1 components is thus

$$p_1^{\text{repair}}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \Phi\left(\frac{\delta - x \cos \theta}{\sin \theta}\right) + \frac{1 - \Phi(\delta)}{\sqrt{2\pi} \sin \theta} \exp\left(-\frac{1}{2} \left(\frac{\delta \cos \theta - x}{\sin \theta}\right)^2\right). \quad (12)$$

Integrating with respect to x yields

$$P_1^{\text{repair}}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}y^2} \Phi\left(\frac{\delta - y \cos \theta}{\sin \theta}\right) \, dy + [1 - \Phi(\delta)] \left[1 - \Phi\left(\frac{\delta \cos \theta - x}{\sin \theta}\right) \right] \quad (13)$$

as the corresponding cumulative distribution function.

Next, consider the distribution of the z_2 components of mutation vectors after repair conditional on $z_1 = x$. The probability of an offspring being infeasible and thus in need of repair conditional on $z_1 = x$ is obtained by computing the relative weight of the second term in the marginal density $p_1^{\text{repair}}(x)$ in Equation (12) and equals

$$\begin{aligned} P_{\text{proj}}(x) &= \text{Prob}[\mathbf{y} \text{ has been projected} \mid z_1 = x] \\ &= \frac{1 - \Phi(\delta)}{\sqrt{2\pi} p_1^{\text{repair}}(x) \sin \theta} \exp\left(-\frac{1}{2} \left(\frac{\delta \cos \theta - x}{\sin \theta}\right)^2\right). \end{aligned}$$

Initially feasible candidate solutions are not projected and have moments described by Equations (10) and (11). Initially infeasible candidate solutions after projection have z_2 component $(\delta - x \cos \theta) / \sin \theta$ and thus

$$E[z_2 | z_1 = x \text{ and } \mathbf{y} \text{ has been projected}] = \frac{\delta - x \cos \theta}{\sin \theta}$$

and

$$E[z_2^2 | z_1 = x \text{ and } \mathbf{y} \text{ has been projected}] = \left(\frac{\delta - x \cos \theta}{\sin \theta} \right)^2.$$

Weighting terms with the probabilities of their occurrence yields after some straightforward simplifications

$$\begin{aligned} E[z_2 | z_1 = x] &= P_{\text{proj}}(x) E[z_2 | z_1 = x \text{ and } \mathbf{y} \text{ has been projected}] \\ &\quad + (1 - P_{\text{proj}}(x)) E[z_2 | z_1 = x \text{ and } \mathbf{y} \text{ has not been projected}] \\ &= \frac{1}{\sqrt{2\pi} p_1^{\text{repair}}(x)} \left[\frac{-1}{\sqrt{2\pi}} e^{-\frac{1}{2}\delta^2} + \frac{1 - \Phi(\delta)}{\sin \theta} \frac{\delta - x \cos \theta}{\sin \theta} \right] \exp \left(-\frac{1}{2} \left(\frac{\delta \cos \theta - x}{\sin \theta} \right)^2 \right) \end{aligned} \tag{14}$$

and

$$\begin{aligned} E[z_2^2 | z_1 = x] &= P_{\text{proj}}(x) E[z_2^2 | z_1 = x \text{ and } \mathbf{y} \text{ has been projected}] \\ &\quad + (1 - P_{\text{proj}}(x)) E[z_2^2 | z_1 = x \text{ and } \mathbf{y} \text{ has not been projected}] \\ &= 1 - \frac{1}{\sqrt{2\pi} p_1^{\text{repair}}(x)} \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\delta^2} \frac{\delta - x \cos \theta}{\sin \theta} \right. \\ &\quad \left. + \frac{1 - \Phi(\delta)}{\sin \theta} \left(1 - \left(\frac{\delta - x \cos \theta}{\sin \theta} \right)^2 \right) \right] \exp \left(-\frac{1}{2} \left(\frac{\delta \cos \theta - x}{\sin \theta} \right)^2 \right) \end{aligned} \tag{15}$$

for the first two moments about zero of the z_2 component of mutation vectors after repair conditional on $z_1 = x$.

Appendix B Single Step Behaviour: Selection and Recombination

This appendix strives to characterise the step made by the strategy by computing the first two moments about zero of $z_i^{(\text{avg})}$ for $i \in \{1, 2\}$ and both the cases of resampling and repair. Due to the choice of coordinate system, $z_1^{(k;\lambda)}$ is the $(\lambda + 1 - k)$ th order statistic of a sample of λ independent realisations of a random variable distributed according to Equation (9) for the case of resampling and according to Equation (13) for the case of repair. The probability density function of the z_1 component of the k th best of the λ offspring is thus

$$p_1^{(k;\lambda)}(x) = \frac{\lambda!}{(\lambda - k)!(k - 1)!} p_1(x) [P_1(x)]^{\lambda - k} [1 - P_1(x)]^{k - 1} \tag{16}$$

where $\lambda \geq k \geq 1$ and superscripts indicating the type of constraint handling have been omitted as the formula holds in both cases. The joint density of the z_1 components of

the k th and l th best of the offspring is

$$p_1^{(k,l;\lambda)}(x, y) = \frac{\lambda!}{(\lambda - l)!(l - k - 1)!(k - 1)!} p_1(x)p_1(y) [P_1(x)]^{\lambda-l} [P_1(y) - P_1(x)]^{l-k-1} [1 - P_1(y)]^{k-1} \tag{17}$$

where $\lambda \geq l > k \geq 1$ and $y \geq x$.

B.1 Computing First Moments

Using Equation (16), the mean of $z_1^{(avg)}$ equals

$$\begin{aligned} E[z_1^{(avg)}] &= \frac{1}{\mu} \sum_{k=1}^{\mu} \int_{-\infty}^{\infty} x p_1^{(k;\lambda)}(x) dx \\ &= \frac{\lambda!}{\mu} \int_{-\infty}^{\infty} x p_1(x) \sum_{k=1}^{\mu} \frac{[P_1(x)]^{\lambda-k} [1 - P_1(x)]^{k-1}}{(\lambda - k)!(k - 1)!} dx. \end{aligned}$$

Identity

$$\begin{aligned} &\sum_{i_1=v}^{\mu} \sum_{i_2=v-1}^{i_1-1} \dots \sum_{i_{v-1}=1}^{i_{v-2}-1} \prod_{k=0}^v \frac{[Q_{k+1} - Q_k]^{i_k - i_{k+1} - 1}}{(i_k - i_{k+1} - 1)!} \\ &= \frac{1}{(\lambda - \mu - 1)!(\mu - v)!} \int_0^{Q_1} z^{\lambda-\mu-1} (1 - z)^{\mu-v} dz \end{aligned} \tag{18}$$

where formally $Q_0 = 0$, $Q_{v+1} = 1$, $i_0 = \lambda + 1$, and $i_{v+1} = 0$, holds for integer $v > 0$ and for any integers μ and λ with $v \leq \mu < \lambda$ and real numbers $Q_k, k = 1, 2, \dots, v$ (see Arnold, 2002, p. 113). Using it with $v = 1$ to replace the summation with an integration yields

$$E[z_1^{(avg)}] = (\lambda - \mu) \binom{\lambda}{\mu} \int_{-\infty}^{\infty} x p_1(x) \int_0^{P_1(x)} z^{\lambda-\mu-1} (1 - z)^{\mu-1} dz dx.$$

A change of variables by defining $z = P_1(y)$ and subsequently exchanging the order of the integrations results in

$$\begin{aligned} E[z_1^{(avg)}] &= (\lambda - \mu) \binom{\lambda}{\mu} \int_{-\infty}^{\infty} x p_1(x) \int_{-\infty}^x p_1(y) [P_1(y)]^{\lambda-\mu-1} [1 - P_1(x)]^{\mu-1} dy dx \\ &= (\lambda - \mu) \binom{\lambda}{\mu} \int_{-\infty}^{\infty} p_1(y) [P_1(y)]^{\lambda-\mu-1} [1 - P_1(y)]^{\mu-1} \underbrace{\int_y^{\infty} x p_1(x) dx}_{I_1(y)} dy. \end{aligned} \tag{19}$$

Computing the inner integral $I_1(y)$ using Equation (8) for the case of resampling and Equation (12) for the case of repair is straightforward. The results of the calculations are given in Figure B1.

Similarly, the mean of $z_2^{(avg)}$ can be computed as

$$\begin{aligned} E[z_2^{(avg)}] &= \frac{1}{\mu} \sum_{k=1}^{\mu} \int_{-\infty}^{\infty} E[z_2 | z_1 = x] p_1^{(k;\lambda)}(x) dx \\ &= (\lambda - \mu) \binom{\lambda}{\mu} \int_{-\infty}^{\infty} p_1(y) [P_1(y)]^{\lambda-\mu-1} [1 - P_1(y)]^{\mu-1} \\ &\quad \times \underbrace{\int_y^{\infty} E[z_2 | z_1 = x] p_1(x) dx}_{I_2(y)} dy. \end{aligned} \tag{20}$$

<p>Abbreviations:</p> $Q(y) = 1 - \Phi\left(\frac{\delta \cos \theta - y}{\sin \theta}\right) \quad q(y) = \frac{1}{\sqrt{2\pi} \sin \theta} \exp\left(-\frac{1}{2} \left(\frac{\delta \cos \theta - y}{\sin \theta}\right)^2\right)$
<p>Resampling:</p> $I_1^{\text{resample}}(y) = p_1^{\text{resample}}(y) - (1 - Q(y)) \frac{\cos \theta}{\sqrt{2\pi} \Phi(\delta)} e^{-\frac{1}{2} \delta^2}$ $I_2^{\text{resample}}(y) = -(1 - Q(y)) \frac{\sin \theta}{\sqrt{2\pi} \Phi(\delta)} e^{-\frac{1}{2} \delta^2}$ $J_1^{\text{resample}}(y) = 1 - P_1^{\text{resample}}(y) + y p_1^{\text{resample}}(y) - [(1 - Q(y)) \delta \cos \theta + q(y) \sin^2 \theta] \frac{\cos \theta}{\sqrt{2\pi} \Phi(\delta)} e^{-\frac{1}{2} \delta^2}$ $J_2^{\text{resample}}(y) = 1 - P_1^{\text{resample}}(y) - [(1 - Q(y)) \delta - q(y) \cos \theta] \frac{\sin^2 \theta}{\sqrt{2\pi} \Phi(\delta)} e^{-\frac{1}{2} \delta^2}$
<p>Repair:</p> $I_1^{\text{repair}}(y) = \Phi(\delta) I_1^{\text{resample}}(y) + (1 - \Phi(\delta)) [(1 - Q(y)) \delta \cos \theta + q(y) \sin^2 \theta]$ $I_2^{\text{repair}}(y) = \Phi(\delta) I_2^{\text{resample}}(y) + (1 - \Phi(\delta)) [(1 - Q(y)) \delta \sin \theta - q(y) \cos \theta \sin \theta]$ $J_1^{\text{repair}}(y) = \Phi(\delta) J_1^{\text{resample}}(y) + (1 - \Phi(\delta)) [(1 - Q(y)) (\delta^2 \cos^2 \theta + \sin^2 \theta) + q(y) \sin^2 \theta (y + \delta \cos \theta)]$ $J_2^{\text{repair}}(y) = \Phi(\delta) J_2^{\text{resample}}(y) + (1 - \Phi(\delta)) [(1 - Q(y)) (\delta^2 \sin^2 \theta + \cos^2 \theta) + q(y) \cos \theta (y \cos \theta - \delta (1 + \sin^2 \theta))]$

Figure B1: Solutions for the inner integrals in Equations (19), (20), (21), and (22).

The solutions for the inner integral $I_2(y)$ obtained using Equations (10) and (14) for the cases of resampling and repair, respectively, are given in Figure B1.

B.2 Computing Second Moments

As

$$z_1^{(\text{avg})^2} = \left(\frac{1}{\mu} \sum_{k=1}^{\mu} z_1^{(k;\lambda)} \right)^2 = \frac{1}{\mu^2} \sum_{k=1}^{\mu} z_1^{(k;\lambda)^2} + \frac{2}{\mu^2} \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} z_1^{(l;\lambda)} z_1^{(k;\lambda)}$$

using Equations (16) and (17) the second moment about zero of $z_1^{(\text{avg})}$ equals

$$\begin{aligned} E[z_1^{(\text{avg})^2}] &= \frac{1}{\mu^2} \sum_{k=1}^{\mu} \int_{-\infty}^{\infty} x^2 p_1^{(k;\lambda)}(x) dx + \frac{2}{\mu^2} \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} \int_{-\infty}^{\infty} \int_x^{\infty} xy p_1^{(k,l;\lambda)}(x, y) dy dx \\ &= \frac{\lambda!}{\mu^2} \int_{-\infty}^{\infty} x^2 p_1(x) \sum_{k=1}^{\mu} \frac{[P_1(x)]^{\lambda-k} [1 - P_1(x)]^{k-1}}{(\lambda - k)!(k - 1)!} dx \\ &\quad + \frac{2\lambda!}{\mu^2} \int_{-\infty}^{\infty} \int_x^{\infty} xy p_1(x) p_1(y) \\ &\quad \times \sum_{l=2}^{\mu} \sum_{k=1}^{l-1} \frac{[P_1(x)]^{\lambda-l} [P_1(y) - P_1(x)]^{l-k-1} [1 - P_1(y)]^{k-1}}{(\lambda - l)!(l - k - 1)!(k - 1)!} dy dx. \end{aligned}$$

Using Equation (18) with $\nu = 1$ and $\nu = 2$ yields

$$\begin{aligned} E[z_1^{(\text{avg})^2}] &= \frac{\lambda - \mu}{\mu} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} x^2 p_1(x) \int_0^{P_1(x)} z^{\lambda-\mu-1} (1 - z)^{\mu-1} dz dx \\ &\quad + 2 \frac{\mu - 1}{\mu} (\lambda - \mu) \binom{\lambda}{\mu} \int_{-\infty}^{\infty} x p_1(x) \int_x^{\infty} y p_1(y) \\ &\quad \times \int_0^{P_1(x)} z^{\lambda-\mu-1} (1 - z)^{\mu-2} dz dy dx. \end{aligned}$$

Changing variables and exchanging the order of the integrations analogously to Appendix B.1 results in

$$\begin{aligned} E[z_1^{(\text{avg})^2}] &= \frac{\lambda - \mu}{\mu} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} p_1(y) [P_1(y)]^{\lambda-\mu-1} [1 - P_1(y)]^{\mu-1} \underbrace{\int_y^{\infty} x^2 p_1(x) dx}_{J_1(y)} dy \\ &\quad + 2 \frac{\mu - 1}{\mu} (\lambda - \mu) \binom{\lambda}{\mu} \int_{-\infty}^{\infty} p_1(y) [P_1(y)]^{\lambda-\mu-1} [1 - P_1(y)]^{\mu-2} \\ &\quad \times \underbrace{\int_y^{\infty} x p_1(x) \int_x^{\infty} z p_1(z) dz dx}_{K_1(y)} dy. \end{aligned} \tag{21}$$

Considering $K_1(y)$, exchanging the order of the integrations and subsequently renaming the variables yields

$$\begin{aligned} K_1(y) &= \int_y^{\infty} x p_1(x) \int_x^{\infty} z p_1(z) dz dx \\ &= \int_y^{\infty} z p_1(z) \int_y^z x p_1(x) dx dz \\ &= \int_y^{\infty} x p_1(x) \int_y^x z p_1(z) dz dx. \end{aligned}$$

Adding the integral expressions in the first and third rows results in

$$2K_1(y) = \int_y^{\infty} x p_1(x) \int_y^{\infty} z p_1(z) dz dx = \left[\int_y^{\infty} x p_1(x) dx \right]^2$$

and thus $K_1(y) = (I_1(y))^2/2$. Similar calculations for the second moment about zero of $z_2^{(\text{avg})}$ yield

$$\begin{aligned}
 E[z_2^{(\text{avg})^2}] &= \frac{\lambda - \mu}{\mu} \binom{\lambda}{\mu} \int_{-\infty}^{\infty} p_1(y) [P_1(y)]^{\lambda-\mu-1} [1 - P_1(y)]^{\mu-1} \\
 &\quad \times \underbrace{\int_y^{\infty} E[z_2^2 | z_1 = x] p_1(x) dx}_{J_2(y)} dy \\
 &\quad + 2 \frac{\mu - 1}{\mu} (\lambda - \mu) \binom{\lambda}{\mu} \int_{-\infty}^{\infty} p_1(y) [P_1(y)]^{\lambda-\mu-1} [1 - P_1(y)]^{\mu-2} \\
 &\quad \times \underbrace{\int_y^{\infty} E[z_2 | z_1 = x] p_1(x) \int_x^{\infty} E[z_2 | z_1 = z] p_1(z) dz dx}_{K_2(y)} dy \quad (22)
 \end{aligned}$$

where analogously to the above $K_2(y) = (I_2(y))^2/2$. The solutions for the inner integrals $J_i(y)$ for $i \in \{1, 2\}$ are again collected in Figure B1.

