
On the Behaviour of the $(1, \lambda)$ -ES for Conically Constrained Linear Problems

Dirk V. Arnold

dirk@cs.dal.ca

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

doi:10.1162/EVCO_a_00125

Abstract

We study the behaviour of a $(1, \lambda)$ -ES that handles constraints by resampling infeasible candidate solutions for linear optimization problems with a conically constrained feasible region. The analysis generalizes prior work in that no particular orientation of the cone relative to the gradient of the objective function is assumed. Expressions that describe the strategy's single-step behaviour are derived. Assuming that the mutation strength is adapted in a scale-invariant manner, a simple zeroth-order model is used to determine the speed of convergence of the strategy. We then derive expressions that approximately characterize the average step size and convergence rate attained when using cumulative step size adaptation and compare the values with optimal ones.

Keywords

Evolution strategies, constraint handling, cumulative step size adaptation.

1 Introduction

Analytical studies of evolution strategies in select test environments contribute to the understanding of the dependence of algorithm behaviour on strategy as well as problem parameters. The ranges of strategy variants and test function classes that have been considered in analytical studies of evolution strategies for constrained optimization are more limited than in the unconstrained case. An improved analytically based understanding of the behaviour of evolution strategies for a wider range of constrained test function classes will be useful for interpreting experimental results, selecting and designing useful test functions for empirical comparisons of algorithms, and developing and improving constraint handling techniques.

Early work that analyzes the performance of evolution strategies for constrained optimization includes that of Rechenberg (1973), who studies the behaviour of the $(1 + 1)$ -ES¹ for the axis-aligned corridor model. Schwefel (1981) considers the performance of the $(1, \lambda)$ -ES in the same environment. Beyer (1989) analyzes the behaviour of the $(1 + 1)$ -ES for a constrained discus-like function. All of these studies have in common that the normal vectors of the constraint planes are orthogonal to the gradient vector of the objective function.

More recently, Arnold and Brauer (2008) consider a linear problem with a single linear constraint of general orientation. They provide a quantitative confirmation of the observation made by Schwefel (1995, p. 116f) that when using the one-fifth rule, the presence of constraints may lead to the step size being systematically reduced in

¹See Beyer and Schwefel (2002) for an introduction to evolution strategy terminology.

situations where the angle between the gradient direction and the normal vector of the active constraint, which we refer to as the constraint angle, is small. Arnold (2011a, 2011b) studies the behaviour of the $(1, \lambda)$ -ES applied to the same problem and compares constraint handling through resampling of infeasible candidate solutions with constraint handling through a simple repair mechanism that projects infeasible candidate solutions onto the feasible region. He finds that cumulative step size adaptation may fail in the face of small constraint angles, and that using the repair mechanism makes it possible to choose the algorithm's parameters such that stagnation can be avoided for any constraint angle. Arnold (2013b) extends the analysis of the linearly constrained problem to consider the $(\mu/\mu, \lambda)$ -ES with cumulative step size adaptation. Arnold (2012) considers the $(1, \lambda)$ -ES with mutative self-adaptation in the same environment and finds that constraint handling through resampling of infeasible candidate solutions results in stagnation in the face of small constraint angles.

Common benchmark problems, such as those collected by Michalewicz and Schoenauer (1996), have optima located on the boundary of the feasible region. A linear problem with a single linear constraint does not possess a finite optimum and does not model that situation well. Rather than increasing the step size indefinitely, an evolution strategy applied to a constrained problem with the optimal solution located on the boundary of the feasible region needs to reduce its step size as it approaches the optimal solution while at the same time avoiding premature stagnation. Arnold (2013a) studies the behaviour of a $(1, \lambda)$ -ES that resamples infeasible candidate solutions for a linear problem with a feasible region delimited by a right circular cone, where the optimal solution lies at the cone's apex and the gradient direction coincides with the cone's central axis. He finds that cumulative step size adaptation enables the strategy to converge linearly for the entire range of parameter settings considered. Assuming that computational costs are dominated by objective function evaluations, the step sizes generated a yield close to optimal rates of convergence. Using smaller values for the number of feasible offspring generated per time step results in the strategy operating at a greater distance from the boundary of the feasible region and thus in fewer infeasible offspring that need to be resampled.

This paper extends the analysis presented by Arnold (2013a) to linear optimization problems with a conically constrained feasible region of general orientation. That is, the assumption that the gradient direction coincides with the cone's central axis is dropped, removing some of the artificial symmetry of the original problem. The remainder of the paper is organized as follows. Section 2 briefly describes the optimization problem and algorithm under consideration. Section 3 analyzes the single-step behaviour of the strategy and derives expressions for the expected step that it takes. Section 4 investigates the multi-step limit behaviour of the strategy assuming that the mutation strength is adapted such that the step size is proportional to the distance from the axis of the cone that delimits the feasible region. A simple zeroth-order model is employed, and its accuracy is evaluated using computer experiments. Section 5 employs the same approach for the analysis of the behaviour of cumulative step size adaptation. Expressions characterizing the step size as well as the convergence rate achieved with cumulative step size adaptation are derived and compared with optimal ones. Section 6 concludes with a brief discussion and goals for future research.

2 Problem and Algorithm

Throughout this paper we consider the problem of minimizing a linear objective subject to the constraint that feasible solutions lie in or on the surface of an infinite right circular

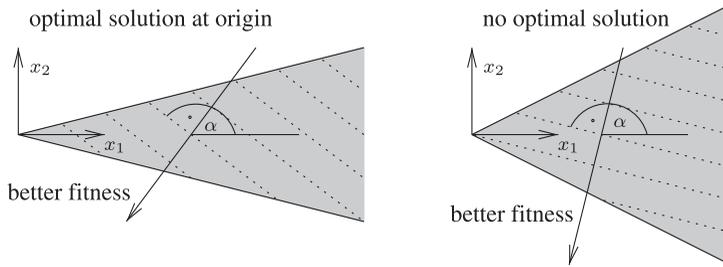


Figure 1: Conically constrained optimization problems in two dimensions. The shaded area is the feasible region. Dotted lines indicate lines of equal fitness. The problem depicted on the left has $\xi > \tan^2 \alpha$ and an optimal solution at the origin; that on the right has $\xi < \tan^2 \alpha$ and does not possess an optimal solution.

cone. By choosing a coordinate system such that the cone's apex is at the origin, the x_1 axis coincides with the cone's central axis, and the x_2 axis is oriented such that the span of the x_1 and x_2 axes includes the gradient direction, the problem can without loss of generality be stated as that of minimizing

$$f(\mathbf{x}) = x_1 \cos \alpha + x_2 \sin \alpha \tag{1}$$

subject to constraints

$$x_1^2 - \xi \sum_{i=2}^N x_i^2 \geq 0 \tag{2}$$

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

where $\mathbf{x} = \langle x_1, \dots, x_N \rangle \in \mathbb{R}^N$, $\xi > 0$, and $\alpha \in [0, \pi/2)$. The problem considered by Arnold (2013a) is recovered for $\alpha = 0$. Figure 1 illustrates the setup. Larger values of ξ result in a narrower cone of feasible solutions. If $\xi > \tan^2 \alpha$, then the optimal solution to the constrained optimization problem lies at the cone's apex and has objective function value zero. If $\xi = \tan^2 \alpha$, then all points $\langle t, -t/\sqrt{\xi}, 0, \dots, 0 \rangle$ with $t \geq 0$ are optimal. If $\xi < \tan^2 \alpha$, then no optimal solution exists and the objective function value can be decreased indefinitely.

The optimization strategy considered throughout this paper is the same as that employed by Arnold (2013a). Assuming a feasible initial candidate solution $\mathbf{x} \in \mathbb{R}^N$, the $(1, \lambda)$ -ES iteratively generates a sequence of further candidate solutions by performing the following four steps per iteration.

1. Generate λ feasible offspring candidate solutions $\mathbf{x}^{(i)} = \mathbf{x} + \sigma \mathbf{z}^{(i)}$, $i = 1, \dots, \lambda$. Vectors $\mathbf{z}^{(i)}$ are generated by repeatedly sampling mutation vectors \mathbf{w} from an N -dimensional normal distribution with mean zero and unit covariance matrix. The first vector \mathbf{w} for which $\mathbf{x} + \sigma \mathbf{w}$ is feasible is labelled $\mathbf{z}^{(i)}$. Step size parameter $\sigma > 0$ is referred to as the mutation strength.
2. Evaluate the $\mathbf{x}^{(i)}$ for $i = 1, \dots, \lambda$ and let $\hat{\mathbf{z}}$ denote the mutation vector corresponding to the offspring candidate solution with the smallest objective function value. Any ties are broken arbitrarily.
3. Update the parental candidate solution according to $\mathbf{x} \leftarrow \mathbf{x} + \sigma \hat{\mathbf{z}}$.
4. Modify the mutation strength.

We defer the discussion of the mutation strength adaptation component of the algorithm until Section 5. The approach of handling constraints through the resampling of infeasible candidate solutions has been discussed by Oyman et al. (1999).

3 Single-Step Behaviour

The analysis of the single-step behaviour of the strategy is simplified by virtue of the symmetries inherent in the constraint functions. Specifically, the location of a candidate solution is adequately described by its x_1 and x_2 coordinates, and the distance $R = \sqrt{\sum_{i=3}^N x_i^2}$ from the origin in the space spanned by the x_3 through x_N axes. When considering single iterations of the $(1, \lambda)$ -ES with parental candidate solution \mathbf{x} we can choose a coordinate system such that $\mathbf{x} = \langle x_1, x_2, R, 0, \dots, 0 \rangle$, where $R > 0^2$. Moreover, in our calculations we make simplifications assuming that the dimension N of the search space tends to ∞ .

3.1 Normalization and Constraints

In preparation of the analysis below, we introduce variables

$$\vartheta_{\mathbf{x}} = \frac{x_2}{x_1} \tag{4}$$

and

$$\delta_{\mathbf{x}} = \frac{N}{R^2} (x_1^2 - \xi x_2^2 - \xi R^2) \tag{5}$$

that we refer to as the objective bias and the normalized amount of slack present in the constraint given by Equation (2), respectively. Moreover, we introduce normalized mutation strength

$$\sigma^* = \frac{N\sigma}{R}. \tag{6}$$

Assuming that \mathbf{x} is feasible, it follows from Equations (4) and (5) that

$$x_1 = R \sqrt{\frac{\xi + \delta_{\mathbf{x}}/N}{1 - \xi \vartheta_{\mathbf{x}}^2}}. \tag{7}$$

The purpose of describing the state of the algorithm in terms of $\vartheta_{\mathbf{x}}$, $\delta_{\mathbf{x}}$, and R rather than in terms of x_1 , x_2 , and R will become clear in Section 4, where we consider scale-invariant behaviour.

According to Equations (2) and (3), an offspring candidate solution $\mathbf{x} + \sigma \mathbf{w}$, where $\mathbf{w} = \langle w_1, \dots, w_N \rangle$ and the w_i are independent standard normal random variables, is feasible if and only if

$$x_1 + \sigma w_1 \geq 0$$

and

$$(x_1 + \sigma w_1)^2 - \xi \sum_{i=2}^N (x_i + \sigma w_i)^2 \geq 0.$$

Using Equations (4), (5), (6), and (7), these conditions can equivalently be written as

$$\sqrt{\frac{\xi + \delta_{\mathbf{x}}/N}{1 - \xi \vartheta_{\mathbf{x}}^2}} + \frac{\sigma^*}{N} w_1 \geq 0 \tag{8}$$

²While $R = 0$ is possible, after initialization the probability of generating such a candidate solution is vanishingly small.

and

$$\delta_x + 2\sqrt{\frac{\xi + \delta_x/N}{1 - \xi\vartheta_x^2}}\sigma^*(w_1 - \xi\vartheta_x w_2) - 2\xi\sigma^*w_3 + \frac{\sigma^{*2}w_1^2}{N} - \xi\frac{\sigma^{*2}}{N}\sum_{i=2}^N w_i^2 \geq 0. \quad (9)$$

For given σ^* , ϑ_x , and δ_x and $N \rightarrow \infty$, the left-hand side of Equation (8) converges in distribution to $\sqrt{\xi/(1 - \xi\vartheta_x^2)}$ and the condition will be satisfied with overwhelming probability. The left-hand side of Equation (9) converges to

$$\chi(w_1, w_2, w_3) = \delta_x + 2\sqrt{\frac{\xi}{1 - \xi\vartheta_x^2}}\sigma^*(w_1 - \xi\vartheta_x w_2) - 2\xi\sigma^*w_3 - \xi\sigma^{*2} \quad (10)$$

and the condition will be satisfied if $\chi(w_1, w_2, w_3) > 0$.

3.2 Distribution of Feasible Offspring

As $\chi(w_1, w_2, w_3)$ is normally distributed with mean $\delta_x - \xi\sigma^{*2}$ and variance $4\xi(1 + \xi)\sigma^{*2}/(1 - \xi\vartheta_x^2)$, it follows that for $N \rightarrow \infty$ the probability of candidate solution $\mathbf{x} + \sigma\mathbf{w}$ being feasible is

$$P_{\text{feasible}} = \Phi\left(\frac{\delta_x - \xi\sigma^{*2}}{2\sigma^*}\sqrt{\frac{1 - \xi\vartheta_x^2}{\xi(1 + \xi)}}\right) \quad (11)$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. Unsurprisingly, the probability of offspring candidate solution $\mathbf{x} + \sigma\mathbf{w}$ being feasible increases with an increasing amount of slack present in the constraint in Equation (2), and it decreases with increasing mutation strength.

From the considerations above, in the limit $N \rightarrow \infty$ the joint probability density of the first three components z_1, z_2 , and z_3 of mutation vector \mathbf{z} resulting in feasible offspring is

$$p_{1,2,3}(x, y, z) = \begin{cases} \frac{1}{\sqrt{2\pi^3} P_{\text{feasible}}} e^{-\frac{1}{2}(x^2+y^2+z^2)} & \text{if } \chi(x, y, z) > 0 \\ 0 & \text{otherwise} \end{cases}.$$

The marginal probability density of the z_1 and z_2 components is

$$\begin{aligned} p_{1,2}(x, y) &= \int_{-\infty}^{\infty} p_{1,2,3}(x, y, z) dz \\ &= \frac{1}{2\pi P_{\text{feasible}}} e^{-\frac{1}{2}(x^2+y^2)} \Phi\left(\frac{\delta_x - \xi\sigma^{*2} + 2\sqrt{\xi/(1 - \xi\vartheta_x^2)}\sigma^*(x - \xi\vartheta_x y)}{2\xi\sigma^*}\right) \end{aligned} \quad (12)$$

where Equation (10) has been used. As the z_i with $i \geq 4$ have no influence on whether an offspring candidate solution is feasible, they are independently distributed with standard normal density.

3.3 Normalized Fitness Difference

The objective function value of (feasible) offspring candidate solution $\mathbf{x} + \sigma\mathbf{z}$ is

$$\begin{aligned} f(\mathbf{x} + \sigma\mathbf{z}) &= (x_1 + \sigma z_1) \cos \alpha + (x_2 + \sigma z_2) \sin \alpha \\ &= f(\mathbf{x}) + \sigma(z_1 \cos \alpha + z_2 \sin \alpha). \end{aligned}$$

As within an iteration, all offspring are generated from the same parental centroid and use the same mutation strength, the offspring candidate solution with the smallest value of the normalized fitness difference

$$\frac{f(\mathbf{x} + \sigma \mathbf{z}) - f(\mathbf{x})}{\sigma} = z_1 \cos \alpha + z_2 \sin \alpha$$

is selected to replace the parent in the next iteration. The probability density of normalized fitness difference values of feasible offspring is

$$\begin{aligned} p_f(w) &= \frac{d}{dw} \text{Prob} [z_1 \cos \alpha + z_2 \sin \alpha < w] \\ &= \frac{d}{dw} \int_{-\infty}^{\infty} \int_{-\infty}^{(w-x \cos \alpha)/\sin \alpha} p_{1,2}(x, y) dy dx. \end{aligned}$$

Using Leibniz's integral rule yields

$$p_f(w) = \frac{1}{\sin \alpha} \int_{-\infty}^{\infty} p_{1,2} \left(x, \frac{w - x \cos \alpha}{\sin \alpha} \right) dx \tag{13}$$

$$\begin{aligned} &= \frac{1}{\sqrt{2\pi} P_{\text{feasible}}} e^{-\frac{1}{2}w^2} \\ &\times \Phi \left(\frac{(\delta_x - \xi \sigma^{*2})\sqrt{(1 - \xi \vartheta_x^2)/\xi}/(2\sigma^*) + w(\cos \alpha - \xi \vartheta_x \sin \alpha)}{\sqrt{\xi(1 - \xi \vartheta_x^2) + (\sin \alpha + \xi \vartheta_x \cos \alpha)^2}} \right) \end{aligned} \tag{14}$$

where the detailed derivation is relegated to Appendix A. We write $P_f(\cdot)$ for the corresponding cumulative distribution function.

3.4 Expected Step

The expected values of the $z_1, z_2,$ and z_3 components of the mutation vector corresponding to the offspring candidate solution selected to replace the parent can be computed using elementary results from the field of order statistics (Balakrishnan and Rao, 1998). Specifically, as there are λ possible outcomes of selection and the selected candidate solution has a smaller objective function value than the remaining $\lambda - 1$ candidate solutions, the expected values can be obtained from integration over the densities computed above as

$$E[\hat{z}_1] = \lambda \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x p_{1,2}(x, y) [1 - P_f(x \cos \alpha + y \sin \alpha)]^{\lambda-1} dy dx \tag{15}$$

$$E[\hat{z}_2] = \lambda \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y p_{1,2}(x, y) [1 - P_f(x \cos \alpha + y \sin \alpha)]^{\lambda-1} dy dx \tag{16}$$

$$E[\hat{z}_3] = \lambda \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z p_{1,2,3}(x, y, z) [1 - P_f(x \cos \alpha + y \sin \alpha)]^{\lambda-1} dz dy dx. \tag{17}$$

Letting

$$I_1 = \lambda \int_{-\infty}^{\infty} w p_f(w) [1 - P_f(w)]^{\lambda-1} dw \tag{18}$$

$$I_2 = \lambda(\lambda - 1) \int_{-\infty}^{\infty} p_f^2(w) [1 - P_f(w)]^{\lambda-2} dw \tag{19}$$

calculations detailed in Appendix B derive expressions

$$E[\hat{z}_1] = \frac{1}{\cos \alpha - \xi \vartheta_x \sin \alpha} I_1 + \left(\frac{1}{\cos \alpha - \xi \vartheta_x \sin \alpha} - \cos \alpha \right) I_2 \quad (20)$$

$$E[\hat{z}_2] = -\frac{\xi \vartheta_x}{\cos \alpha - \xi \vartheta_x \sin \alpha} I_1 - \left(\frac{\xi \vartheta_x}{\cos \alpha - \xi \vartheta_x \sin \alpha} + \sin \alpha \right) I_2 \quad (21)$$

$$E[\hat{z}_3] = -\frac{\sqrt{\xi(1 - \xi \vartheta_x^2)}}{\cos \alpha - \xi \vartheta_x \sin \alpha} (I_1 + I_2) \quad (22)$$

allowing the straightforward numerical computation of the expected values.

3.5 Change of Normalized Slack

The normalized amount of slack $\delta_{x+\sigma z}$ present in feasible candidate solution $\mathbf{x} + \sigma \mathbf{z}$ can be computed from Equation (5) as

$$\begin{aligned} \delta_{x+\sigma z} &= N \left[\frac{(x_1 + \sigma z_1)^2 - \xi(x_2 + \sigma z_2)^2}{\sum_{i=3}^N (x_i + \sigma z_i)^2} - \xi \right] \\ &= N \left[\frac{x_1^2 - \xi x_2^2 + 2\sigma x_1 z_1 - 2\xi \sigma x_2 z_2 + \sigma^2 z_1^2 - \xi \sigma^2 z_2^2}{R^2 + 2R\sigma z_3 + \sigma^2 \sum_{i=3}^N z_i^2} - \xi \right]. \end{aligned}$$

Using Equations (4), (5), (6), and (7) and cancelling terms results in

$$\delta_{x+\sigma z} = \frac{N\xi + \delta_x + 2\sqrt{(\xi + \delta_x/N)/(1 - \xi \vartheta_x^2)} \sigma^* (z_1 - \xi \vartheta_x z_2) + \sigma^{*2} (z_1^2 - \xi z_2^2) / N}{1 + 2\sigma^* z_3 / N + \sigma^{*2} \sum_{i=3}^N z_i^2 / N^2} - N\xi$$

where δ_x is the normalized amount of slack present in the parent. Performing a Taylor expansion of the fraction yields

$$\begin{aligned} \delta_{x+\sigma z} &= \left(N\xi + \delta_x + 2\sqrt{\frac{\xi + \delta_x/N}{1 - \xi \vartheta_x^2}} \sigma^* (z_1 - \xi \vartheta_x z_2) + \frac{\sigma^{*2}}{N} (z_1^2 - \xi z_2^2) \right) \\ &\quad \times \left(1 - \frac{2\sigma^*}{N} z_3 - \frac{\sigma^{*2}}{N^2} \sum_{i=3}^N z_i^2 + \dots \right) - N\xi \\ &= \delta_x + 2\sqrt{\frac{\xi + \delta_x/N}{1 - \xi \vartheta_x^2}} \sigma^* (z_1 - \xi \vartheta_x z_2) - 2\xi \sigma^* z_3 - \frac{\xi \sigma^{*2}}{N} \sum_{i=3}^N z_i^2 + \dots \end{aligned}$$

where the dots represent terms of quadratic and higher order. Dropping the terms that disappear in the limit $N \rightarrow \infty$ and replacing the term involving the sum of squared components of the mutation vector with its limit value finally yields

$$\delta_{x+\sigma z} = \delta_x + 2\sqrt{\frac{\xi}{1 - \xi \vartheta_x^2}} \sigma^* (z_1 - \xi \vartheta_x z_2) - 2\xi \sigma^* z_3 - \xi \sigma^{*2} \quad (23)$$

for the normalized amount of slack present in $\mathbf{x} + \sigma \mathbf{z}$.

3.6 Change of Objective Bias

Using Equations (4), (5), (6), and (7), the objective bias of candidate solution $\mathbf{x} + \sigma \mathbf{z}$ can be computed similarly as

$$\begin{aligned} \vartheta_{\mathbf{x}+\sigma\mathbf{z}} &= \frac{x_2 + \sigma z_2}{x_1 + \sigma z_1} \\ &= \frac{\vartheta_{\mathbf{x}} + \sqrt{(1 - \xi \vartheta_{\mathbf{x}}^2)/(\xi + \delta_{\mathbf{x}}/N)} \sigma^* z_2/N}{1 + \sqrt{(1 - \xi \vartheta_{\mathbf{x}}^2)/(\xi + \delta_{\mathbf{x}}/N)} \sigma^* z_1/N}. \end{aligned}$$

Performing a Taylor expansion of the fraction yields

$$\begin{aligned} \vartheta_{\mathbf{x}+\sigma\mathbf{z}} &= \left(\vartheta_{\mathbf{x}} + \sqrt{\frac{1 - \xi \vartheta_{\mathbf{x}}^2}{\xi + \delta_{\mathbf{x}}/N}} \frac{\sigma^*}{N} z_2 \right) \left(1 - \sqrt{\frac{1 - \xi \vartheta_{\mathbf{x}}^2}{\xi + \delta_{\mathbf{x}}/N}} \frac{\sigma^*}{N} z_1 + \dots \right) \\ &= \vartheta_{\mathbf{x}} + \sqrt{\frac{1 - \xi \vartheta_{\mathbf{x}}^2}{\xi}} \frac{\sigma^*}{N} (z_2 - \vartheta_{\mathbf{x}} z_1) + \dots \end{aligned} \tag{24}$$

where the dots represent terms of quadratic and higher order.

4 Steady State Behaviour

In this section we consider the behaviour of the $(1, \lambda)$ -ES with scale-invariant mutation strength. That is, we assume that σ is adapted such that the normalized mutation strength σ^* is constant. While this does not constitute a viable algorithm, the results will form a stepping stone toward the analysis of the behaviour of cumulative step size adaptation in Section 5. Moreover, they will allow deriving optimal parameter settings that values generated by any step size adaptation mechanism can be compared against.

Assuming scale-invariance, running the $(1, \lambda)$ -ES according to Equations (23) and (24) amounts to iterating update equations

$$\delta^{(t+1)} = \delta^{(t)} + 2 \sqrt{\frac{\xi}{1 - \xi \vartheta^{(t)2}}} \sigma^* \left(\hat{z}_1^{(t)} - \xi \vartheta^{(t)} \hat{z}_2^{(t)} \right) - 2 \xi \sigma^* \hat{z}_3^{(t)} - \xi \sigma^{*2} \tag{25}$$

$$\vartheta^{(t+1)} = \vartheta^{(t)} + \sqrt{\frac{1 - \xi \vartheta^{(t)2}}{\xi}} \frac{\sigma^*}{N} \left(\hat{z}_2^{(t)} - \vartheta^{(t)} \hat{z}_1^{(t)} \right) \tag{26}$$

where superscripts indicate iteration number, and thus generates a time-homogeneous Markov chain with state variables δ and ϑ . We employ the dynamical systems approach described by Meyer-Nieberg and Beyer (2012) and compute an approximation to the mean values of normalized slack and objective bias after initialization effects have faded by solving

$$E[\delta^{(t+1)}] = \delta^{(t)} \tag{27}$$

$$E[\vartheta^{(t+1)}] = \vartheta^{(t)} \tag{28}$$

for $\delta = \delta^{(t)}$ and $\vartheta = \vartheta^{(t)}$. Presumably, a better approximation to the steady state of the Markov chain could be obtained by employing an exponential model for the distribution of normalized slack values, as done in Arnold and Brauer (2008) for the case of a linear problem with a single linear constraint. However, it will be seen that the

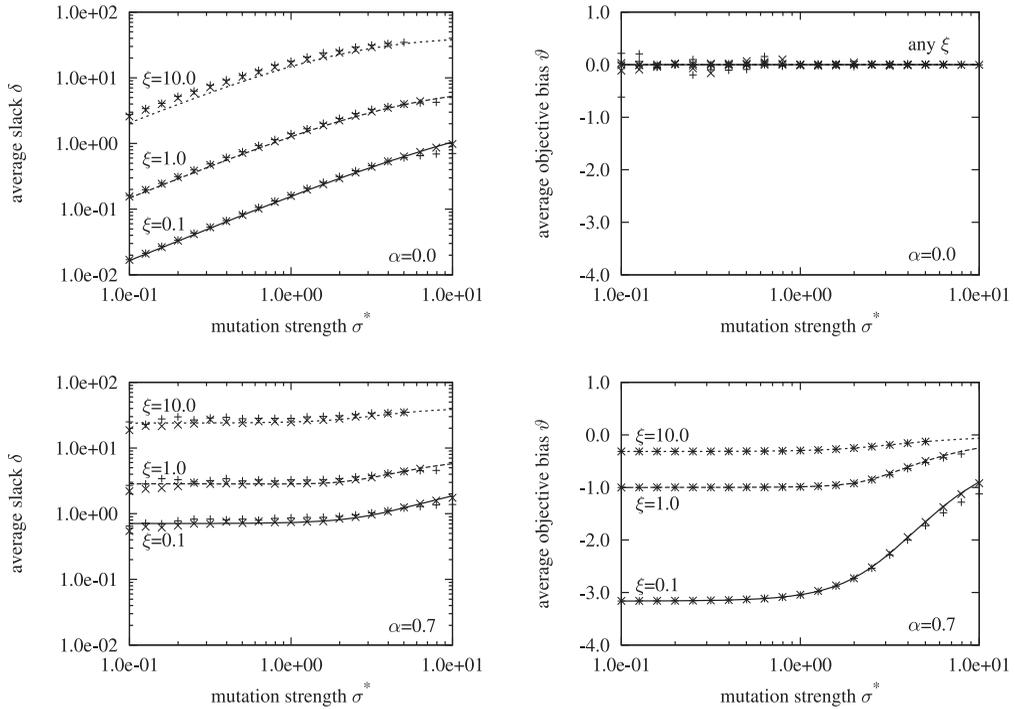


Figure 2: Average normalized slack δ (left) and average objective bias ϑ (right) of the $(1, 10)$ -ES with scale-invariant mutation strength plotted against normalized mutation strength σ^* . The data shown represent results for $\alpha = 0.0$ (top) and $\alpha = 0.7$ (bottom). Points mark data measured in runs of the strategy for $N = 40$ (+) and $N = 400$ (x).

results derived on the basis of the simple zeroth-order model are sufficient to quite adequately characterize the behaviour of the strategy for the range of normalized mutation strengths of interest.

Figure 2 plots the average normalized slack δ and average objective bias ϑ in runs of the $(1, \lambda)$ -ES with $\lambda = 10$ and scale-invariant step size against the normalized mutation strength σ^* . The curves have been obtained by numerical root finding using Equations (27) and (28) with Equations (20), (21), (22), (25), and (26). The points mark average measurements made in runs of the strategy for search space dimensions $N \in \{40, 400\}$. Each run is initialized with the parental candidate solution on the boundary of the feasible region. Mixing times are very long for small values of the normalized mutation strength. In order to avoid initialization effects from impacting the measurements, the strategy is run with decreasing normalized mutation strength for $400N$ iterations in which no measurements are made. The data points shown in the figure are the result of averaging the normalized slack and objective bias over the following 20,000 time steps. A run is terminated prematurely and no data point is included in the figure if the generation of an offspring results in at least 1,000 infeasible candidate solutions before sampling a feasible one.

It can be seen from the figure that the average normalized slack increases with increasing normalized mutation strength as well as with increasing ξ . In contrast to the case where $\alpha = 0.0$ (shown at the top and considered previously by Arnold, 2013a)

where the normalized amount of slack appears to tend to zero as normalized mutation strengths decrease, for $\alpha = 0.7$ the curves are nearly flat for small values of σ^* . At the same time, the objective bias, which is zero for $\alpha = 0.0$, takes on negative values for $\alpha = 0.7$. For small values of the normalized mutation strength, average values of ϑ are close to $-1/\sqrt{\xi}$, indicating that the strategy tracks the boundary of the feasible region near the bottom of the cone shown in Figure 1. The accuracy of the predictions made on the basis of the zeroth-order model and the assumption of infinite search space dimensionality is quite good throughout. It visibly decreases for larger values of σ^* or ξ , where a model that includes fluctuations of δ would be required for more accurate results. Deviations of the measured values from the analytically obtained curves can also be observed for small values of the normalized mutation strength, where the long mixing times are responsible and higher accuracy would be observed if more iterations were discarded as well as averaged over in the experimental runs.

In the case that $\xi > \tan^2 \alpha$ (i.e., that an optimal solution exists and has objective function value zero), depending on the value of σ^* , either linear convergence or linear divergence of the strategy is observed. Two important quantities that determine the computational cost of a run are the average probability P_{feasible} of a random candidate solution being feasible and the normalized per iteration quality gain³

$$\Delta_x = -NE \left[\log \left(\frac{f(\mathbf{x} + \sigma \hat{\mathbf{z}})}{f(\mathbf{x})} \right) \right]. \tag{29}$$

The former determines the number of candidate solutions that need to be sampled on average (and thereby the number of constraint function evaluations) in order to generate a feasible offspring candidate solution. The latter coincides with the rate of convergence (or divergence, if negative) of the strategy, scaled with the search space dimension N . From Equation (1) with Equations (4), (6), and (7), the normalized quality gain of the strategy equals

$$\begin{aligned} \Delta_x &= -NE \left[\log \left(1 + \sigma \frac{\hat{z}_1 \cos \alpha + \hat{z}_2 \sin \alpha}{x_1 \cos \alpha + x_2 \sin \alpha} \right) \right] \\ &= -NE \left[\log \left(1 + \frac{\sigma^*}{N} \sqrt{\frac{1 - \xi \vartheta_x^2}{\xi + \delta_x/N}} \frac{\hat{z}_1 \cos \alpha + \hat{z}_2 \sin \alpha}{\cos \alpha + \vartheta_x \sin \alpha} \right) \right]. \end{aligned}$$

Performing a Taylor expansion of the logarithm and dropping terms that disappear in the limit $N \rightarrow \infty$ yields

$$\Delta_x = -\sigma^* \sqrt{\frac{1 - \xi \vartheta_x^2}{\xi}} \frac{E[\hat{z}_1] \cos \alpha + E[\hat{z}_2] \sin \alpha}{\cos \alpha + \vartheta_x \sin \alpha} \tag{30}$$

as an approximation for the normalized quality gain of the strategy.

In the case that $\xi < \tan^2 \alpha$ (i.e., an optimal solution does not exist), the strategy can be observed to diverge linearly. In contrast to the former case, once negative objective function values of candidate solutions have been achieved, solution quality is improved by increasing absolute objective function values. We thus use the negative of Equation (30) to estimate the normalized rate of divergence. Accordingly, we measure the normalized quality gain in runs of the evolution strategy as the average observed

³We employ a logarithmic definition of the quality gain in order to avoid the pitfalls of the definition based on expected relative improvement pointed out by Auger and Hansen (2006).

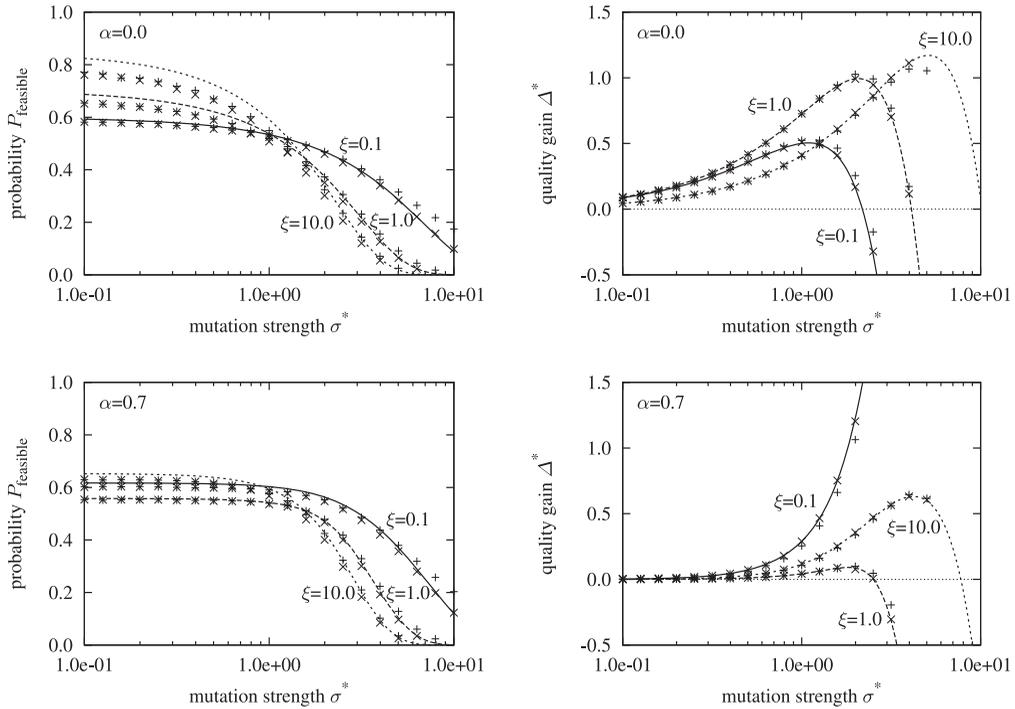


Figure 3: Average probability P_{feasible} of generating feasible offspring (left) and normalized per iteration quality gain Δ^* (right) of the $(1, 10)$ -ES with scale-invariant mutation strength plotted against normalized mutation strength σ^* . The data shown represent results for $\alpha = 0.0$ (top) and $\alpha = 0.7$ (bottom). Points mark data measured in runs of the strategy for $N = 40$ (+) and $N = 400$ (x).

value of $N \log(f(\mathbf{x} + \sigma \hat{\mathbf{z}})/f(\mathbf{x}))$. After initialization effects have faded, we do not observe changes in sign of the parental objective function value, ensuring that the logarithm is real valued.

Figure 3 plots the average probability P_{feasible} and the normalized quality gain Δ^* from runs of the $(1, \lambda)$ -ES with $\lambda = 10$ and scale-invariant step size against the normalized mutation strength σ^* . The lines have been obtained from Equations (11) and (30) with Equations (20) and (21), with the average normalized slack and objective bias obtained from the zeroth-order model based on Equations (27) and (28). The points mark measurements made in the same runs of the evolution strategy used to generate Figure 2. The higher accuracy of the quality gain predictions compared to those of P_{feasible} is a result of the skewness of the distribution of normalized slack values, which is not considered in the simple zeroth-order model, and the differing slopes of the functions in Equations (11) and (30) in the vicinity of the normalized slack and objective bias values typically encountered.

It can be seen from the figure that the average probability P_{feasible} decreases with increasing normalized mutation strength and thus that larger mutation strengths result in more constraint function evaluations per iteration. Moreover, in contrast to the case of a linear function with a single linear constraint (as considered in Arnold and Brauer, 2008; Arnold, 2011a, 2011b, 2012, 2013b), P_{feasible} can drop significantly below one-half

for the conically constrained problem. In five of the six cases considered, the normalized quality gain shows the usual unimodal behaviour, where the quality gain first increases with increasing normalized mutation strength σ^* , but then starts decreasing and eventually turns negative (signifying divergence) as σ^* increases further. Normalized quality gain values are larger for $\alpha = 0.0$ than for $\alpha = 0.7$. The sixth case ($\alpha = 0.7$, $\xi = 0.1$) is the only one where $\xi < \tan^2 \alpha$ and thus where there is no optimal solution at the cone's apex. In that case, larger normalized mutation strengths result in more rapid divergence and increasingly larger quality gain values.

Figure 4 plots the optimal normalized mutation strength σ^* (i.e., the normalized mutation strength that maximizes the normalized quality gain), the corresponding average probability P_{feasible} , and the resulting normalized quality gain Δ^* against both ξ and α . The curves have been obtained by numerically optimizing the normalized quality gain computed from Equation (30) with the average normalized slack and objective bias obtained from the zeroth-order model based on Equations (27) and (28). It can be seen that where the problem parameters are such that an optimal solution exists, optimal normalized mutation strengths increase with increasing values of ξ . The larger normalized mutation strengths for larger values of ξ result in faster convergence of the $(1, \lambda)$ -ES. Figure 4 also shows that the faster convergence in the face of larger values of ξ comes at the cost of small values of P_{feasible} and thus larger numbers of constraint function evaluations.

5 Mutation Strength Adaptation

Achieving scale-invariant step size of course requires information generally not available to the algorithm. Instead, evolution strategies employ one of several mechanisms for adapting their mutation strength. Among the techniques used, mutative self-adaptation (Schwefel, 1981) and cumulative step size adaptation (Ostermeier et al., 1994) are the two most common, with the latter forming the global step size adaptation component of the CMA-ES (Hansen and Ostermeier, 2001).

Cumulative step size adaptation for the $(1, \lambda)$ -ES employs a search path $\mathbf{s} \in \mathbb{R}^N$ defined by $\mathbf{s}^{(0)} = \mathbf{0}$ and

$$\mathbf{s}^{(t+1)} = (1 - c)\mathbf{s}^{(t)} + \sqrt{c(2 - c)}\mathbf{z}^{(t)} \tag{31}$$

where superscripts indicate iteration number, that implements an exponentially fading record of past steps taken by the strategy. Constant $c \in (0, 1)$, which goes to zero as $N \rightarrow \infty$, is referred to as the cumulation parameter and determines the effective length of the memory implemented in the search path. The mutation strength is updated according to⁴

$$\sigma^{(t+1)} = \sigma^{(t)} \exp\left(c \frac{\|\mathbf{s}^{(t+1)}\|^2 - N}{2DN}\right) \tag{32}$$

where $D > 0$ is a damping constant that scales the magnitude of the 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

⁴Notice that this update 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 often behave similarly. As in previous work analyzing cumulative step size adaptation, basing the update of the mutation strength on the squared length of the search path simplifies the analysis.

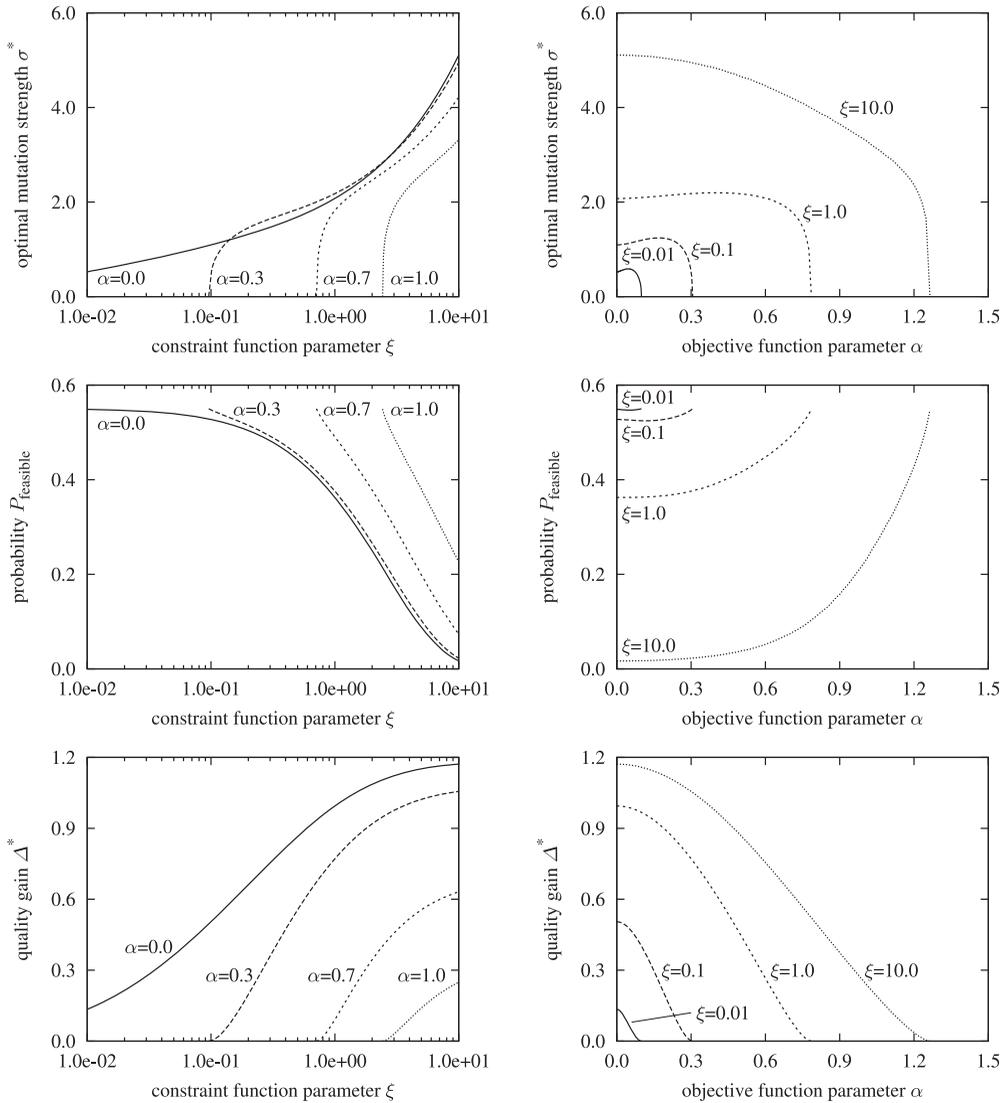


Figure 4: Optimal normalized mutation strength σ^* , corresponding average probability P_{feasible} of generating feasible offspring, and normalized quality gain Δ^* of the $(1, 10)$ -ES with scale-invariant mutation strength plotted against constraint and objective function parameters ξ (left) and α (right).

longer steps. Short search paths indicate 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 consecutive steps are perpendicular on average, in which case no change in mutation strength is effected.

The approach to the analysis of cumulative step size adaptation is the same as that used in Section 4 to study the behaviour of the scale-invariant strategy. We identify a set of variables that describes the state of the algorithm such that the update equations for those variables that are induced by the strategy behaviour are time-invariant. We derive

approximate values for the mean values of the joint limit distribution of the variables using a simple zeroth-order model. Imposing conditions akin to Equations (27) and (28) yields a system of as many equations as there are variables. The same approach has been used to analyze the behaviour of cumulative step size adaptation in other environments (Arnold, 2002; Arnold and Beyer, 2010; Arnold, 2013b). A better characterization of the joint limit distribution could be obtained by expanding it into a Gram-Charlier series with unknown cumulants as proposed by Beyer and Schwefel (2002). Considering cumulants beyond the means would allow deriving approximations for variances and covariances, and so forth.

While the normalized slack and objective bias are sufficient for describing the state of the scale-invariant strategy, a greater number of variables is required for the case of cumulative step size adaptation. A minimal set consists of the normalized slack δ , the objective bias ϑ , the normalized mutation strength σ^* , the components s_1 and s_2 of the search path in the directions of the x_1 and x_2 axes, the length

$$s_{\odot} = -\frac{1}{R} \sum_{i=3}^N s_i x_i \tag{33}$$

of the projection of the search path onto the vector pointing from the current candidate solution to the origin of the space spanned by the x_3 through x_N axes, and the deviation $\|\mathbf{s}\|^2 - N$ of the search path's squared length from the value where no change in mutation strength is effected.

According to Equation (31), the evolution of the search path in the subspace spanned by the x_1 axis is governed by

$$s_1^{(t+1)} = (1 - c)s_1^{(t)} + \sqrt{c(2 - c)}\hat{z}_1^{(t)}.$$

Imposing condition $E[s_1^{(t+1)}] = s_1^{(t)}$ and solving for $s_1 = s_1^{(t)}$ yields

$$s_1 = \sqrt{\frac{2 - c}{c}} E[\hat{z}_1] \tag{34}$$

as an approximation to the average value of s_1 . Proceeding equivalently in the subspace spanned by the x_2 axis results in

$$s_2 = \sqrt{\frac{2 - c}{c}} E[\hat{z}_2] \tag{35}$$

as an approximation to the average value of s_2 .

Similarly, from Equation (33) with Equation (31), the evolution of the component of the search path in direction of the origin of the space spanned by the x_3 through x_N axes is described by

$$\begin{aligned} s_{\odot}^{(t+1)} &= -\frac{1}{R^{(t+1)}} \sum_{i=3}^N s_i^{(t+1)} x_i^{(t+1)} \\ &= -\frac{\sum_{i=3}^N \left((1 - c)s_i^{(t)} + \sqrt{c(2 - c)}\hat{z}_i^{(t)} \right) \left(x_i^{(t)} + \sigma^{(t)}\hat{z}_i^{(t)} \right)}{\sqrt{\sum_{i=3}^N \left(x_i^{(t)} + \sigma^{(t)}\hat{z}_i^{(t)} \right)^2}} \\ &= \frac{(1 - c) \left(R^{(t)} s_{\odot}^{(t)} - \sigma^{(t)} \sum_{i=3}^N s_i^{(t)} \hat{z}_i^{(t)} \right) - \sqrt{c(2 - c)} \left(R^{(t)} \hat{z}_3^{(t)} + \sigma^{(t)} \sum_{i=3}^N \hat{z}_i^{(t)2} \right)}{\sqrt{R^{(t)2} + 2R^{(t)}\sigma^{(t)}\hat{z}_3^{(t)} + \sigma^{(t)2} \sum_{i=3}^N \hat{z}_i^{(t)2}}}. \end{aligned}$$

Using Equation (6), cancelling terms, expanding the fraction into a Taylor series, taking expected values, and omitting terms that disappear in the limit $N \rightarrow \infty$ yields

$$E[s_{\odot}^{(t+1)}] = (1 - c)s_{\odot}^{(t)} - \sqrt{c(2 - c)} (E[\hat{z}_3] + \sigma^*).$$

Imposing condition $E[s_{\odot}^{(t+1)}] = s_{\odot}^{(t)}$ and solving for $s_{\odot} = s_{\odot}^{(t)}$ yields

$$s_{\odot} = -\sqrt{\frac{2 - c}{c}} (E[\hat{z}_3] + \sigma^*) \tag{36}$$

as an approximation to the average value of s_{\odot} .

Considering the squared length of the search path, using Equation (31) yields

$$\begin{aligned} \|\mathbf{s}^{(t+1)}\|^2 &= \sum_{i=1}^N \left((1 - c)s_i^{(t)} + \sqrt{c(2 - c)}\hat{z}_i^{(t)} \right)^2 \\ &= (1 - c)^2 \|\mathbf{s}^{(t)}\|^2 + 2(1 - c)\sqrt{c(2 - c)}\mathbf{s}^{(t)} \cdot \hat{\mathbf{z}}^{(t)} + c(2 - c)\|\hat{\mathbf{z}}^{(t)}\|^2. \end{aligned}$$

Taking expected values results in

$$\begin{aligned} E[\|\mathbf{s}^{(t+1)}\|^2] &= (1 - c)^2 \|\mathbf{s}^{(t)}\|^2 + 2(1 - c)\sqrt{c(2 - c)} \left(s_1^{(t)}E[\hat{z}_1] + s_2^{(t)}E[\hat{z}_2] - s_{\odot}^{(t)}E[\hat{z}_3] \right) \\ &\quad + c(2 - c)E[\|\hat{\mathbf{z}}\|^2]. \end{aligned}$$

Imposing condition $E[\|\mathbf{s}^{(t+1)}\|^2] = \|\mathbf{s}^{(t)}\|^2$, using Equations (34), (35), and (36), and omitting terms that disappear in the limit $N \rightarrow \infty$ yields

$$\begin{aligned} \|\mathbf{s}\|^2 - N &= \frac{2(1 - c)}{\sqrt{c(2 - c)}} (s_1E[\hat{z}_1] + s_2E[\hat{z}_2] - s_{\odot}E[\hat{z}_3]) \\ &= \frac{2(1 - c)}{c} \left(E[\hat{z}_1]^2 + E[\hat{z}_2]^2 + E[\hat{z}_3]^2 + \sigma^*E[\hat{z}_3] \right) \end{aligned} \tag{37}$$

as an approximation for the average deviation of the squared length of the search path from N .

Finally, using Equations (6) and (32), the update of the normalized mutation strength is described by

$$\begin{aligned} \sigma^{*(t+1)} &= \frac{N\sigma^{(t+1)}}{R^{(t+1)}} \\ &= \frac{N\sigma^{(t)} \exp\left(c \frac{\|\mathbf{s}^{(t+1)}\|^2 - N}{2DN}\right)}{\sqrt{R^{(t)2} + 2R^{(t)}\sigma^{(t)}\hat{z}_3^{(t)} + \sigma^{(t)2} \sum_{i=3}^N \hat{z}_i^{(t)2}}}. \end{aligned}$$

Again using Equation (6), cancelling terms, and Taylor expanding both the fraction and the exponential function yields

$$\begin{aligned} \sigma^{*(t+1)} &= \sigma^{*(t)} \left(1 + c \frac{\|\mathbf{s}^{(t+1)}\|^2 - N}{2DN} + \dots \right) \left(1 - \frac{\sigma^{*(t)}}{N} \hat{z}_3^{(t)} - \frac{\sigma^{*(t)2}}{2N} + \dots \right) \\ &= \sigma^{*(t)} \left[1 - \frac{1}{N} \left(\sigma^{*(t)} \hat{z}_3^{(t)} + \frac{\sigma^{*(t)2}}{2} \right) + c \frac{\|\mathbf{s}^{(t+1)}\|^2 - N}{2DN} + \dots \right] \end{aligned}$$

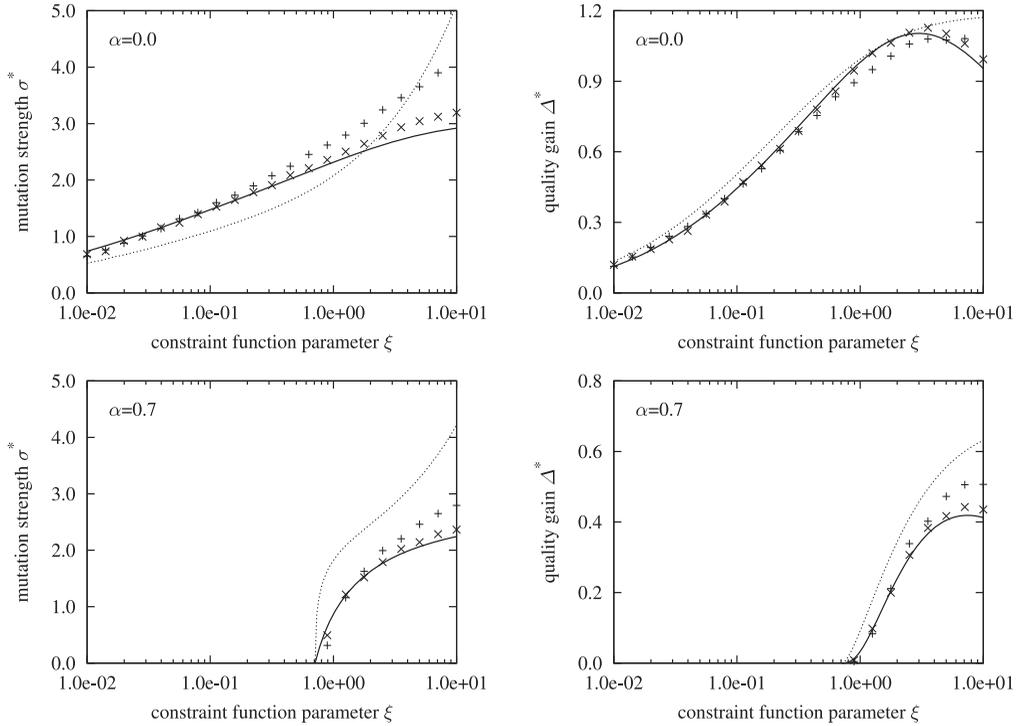


Figure 5: Average normalized mutation strength σ^* (left) and normalized quality gain Δ^* (right) of the $(1, \lambda)$ -ES with cumulative step size adaptation plotted against constraint function parameter ξ . The data shown represent results for $\alpha = 0.0$ (top) and $\alpha = 0.7$ (bottom). Points mark data measured in runs of the strategy for $N = 40$ (+) and $N = 400$ (x). The thin dotted lines indicate the values for optimal scale-invariant mutation strength shown in Figure 4.

where the dots represent terms of quadratic and higher order. Taking expected values and imposing condition $E[\sigma^{*(t+1)}] = \sigma^{*(t)}$ with Equation (37) yields

$$\sigma^* E[\hat{z}_3] + \frac{\sigma^{*2}}{2} = \frac{1 - c}{D} \left(E[\hat{z}_1]^2 + E[\hat{z}_2]^2 + E[\hat{z}_3]^2 + \sigma^* E[\hat{z}_3] \right).$$

As $c \rightarrow 0$ for $N \rightarrow \infty$, this finally yields condition

$$E[\hat{z}_1]^2 + E[\hat{z}_2]^2 + E[\hat{z}_3]^2 + (1 - D)\sigma^* E[\hat{z}_3] = D \frac{\sigma^{*2}}{2} \tag{38}$$

which, with Equations (20), (21), and (22) and the zeroth-order model for the normalized slack δ and objective bias ϑ , can be used to solve numerically for an approximation to the average normalized mutation strength generated by cumulative step size adaptation.

Figure 5 plots the average normalized mutation strength σ^* and the normalized quality gain Δ^* of the $(1, \lambda)$ -ES with cumulative step size adaptation against constraint function parameter ξ . The curves have been obtained by numerical root finding from Equations (38), (27), and (28) with Equations (20), (21), and (22) to compute the expected values. The thin dotted lines show the values for optimal scale-invariant mutation strength derived in Section 4 for comparison. The points mark averages measured in runs of the $(1, \lambda)$ -ES with cumulative step size adaptation with $c = 1/\sqrt{N}$, $D = 1$,

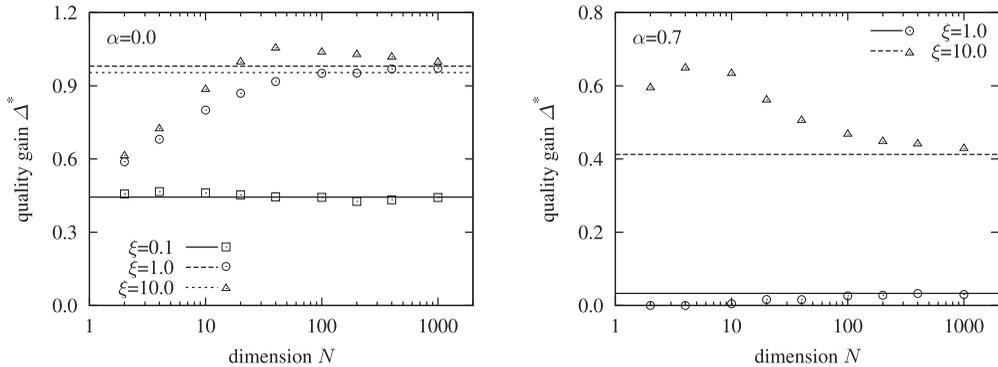


Figure 6: Normalized quality gain Δ^* of the $(1, 10)$ -ES with cumulative step size adaptation plotted against the search space dimension N . The data shown are for $\alpha = 0.0$ (left) and $\alpha = 0.7$ (right). The horizontal lines represent values computed for $N \rightarrow \infty$. The points mark measurements made in runs of the strategy.

and $N \in \{40, 400\}$. Each run is initialized with the parental candidate solution on the boundary of the feasible region and an initial mutation strength such that $\sigma^* = 1$. The first $40N$ iterations of each run are discarded in order to avoid initialization conditions from impacting the measurements. The data points shown in the figure are the result of averaging over the following 20,000 time steps, where again data points from runs in which at least 1,000 infeasible candidate solutions were sampled in the generation of a single offspring are excluded.

It can be seen that cumulative step size adaptation successfully controls the mutation strength of the strategy in that linear convergence (as signified by positive quality gain values) is achieved for all combinations of parameter values of α and ξ considered that admit an optimal solution. Step sizes generated usually differ from optimal ones. However, quality gain values observed are not far from optimal ones, and the lower mutation strengths employed for large values of ξ have the advantage of resulting in fewer infeasible candidate solutions being generated.

Figure 6 illustrates the magnitude of the error made in the derivation of the normalized quality gain. The lines have been obtained from Equation (30) using the simple zeroth-order model that neglects fluctuations and assumes $N \rightarrow \infty$. The points mark measurements made in runs of the $(1, 10)$ -ES with cumulative step size adaptation as described above, but without terminating a run once 1,000 infeasible candidate solutions have been sampled in the generation of a single offspring candidate solution. It can be seen that while for $\alpha = 0.0$ and $\xi = 0.1$ the quality of the approximation is good for N as small as 2, larger values of ξ as well as $\alpha = 0.7$ may result in significant amounts of error in low-dimensional search spaces.

6 Discussion and Conclusions

To conclude, we have presented an analysis of the behaviour of a $(1, \lambda)$ -ES that handles constraints by resampling infeasible candidate solutions applied to conically constrained linear problems. Conically constrained linear problems are among the easiest to analyze problems with optimal solutions located on the boundary of the feasible region. Assuming scale-invariant mutation strength, we have found that the probability of generating infeasible offspring and thus needing to resample repeatedly increases with

increasing step size. Larger values of the parameter ξ of the constraint function correspond to a more strongly constrained problem (i.e., one with a narrower feasible region), but at the same time make it possible to achieve faster rates of convergence. However, the faster rates of convergence, which are achieved with larger step sizes, come at the cost of generating more infeasible candidate solutions that need to be resampled.

We then investigated the performance of cumulative step size adaptation and found that it generates mutation strengths that enable the evolution strategy to converge linearly for the entire range of parameter settings considered that result in problems with the optimal solution located at the cone's apex. Quality gain values achieved using cumulative step size adaptation are close to optimal, and while the gap to optimal values widens for larger values of ξ , the mutation strengths generated by cumulative step size adaptation have the advantage of resulting in fewer infeasible candidate solutions than optimal settings do.

Future work to be conducted includes the extension of the analysis presented here to consider evolution strategies that employ non-singleton populations of parental candidate solutions, such as the $(\mu/\mu, \lambda)$ -ES. For the special case that α equals zero and the gradient direction is aligned with the cone's axis, this has been accomplished by Porter and Arnold (2014), who find that employing non-singleton populations with multi-recombination leads to the evolution strategy operating at a greater distance from the boundary of the infeasible region, resulting in fewer infeasible offspring being generated. This is advantageous in the case that the cost of constraint function evaluations is non-negligible, and it remains to be seen whether the same can be observed for non-zero values of α . It is furthermore desirable to conduct similar analyses for other step size adaptation and constraint handling techniques, as well as for further constrained problems with different characteristics.

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). The suggestion to consider a conically constrained linear problem of general orientation is due to one of the reviewers of Arnold (2013a). The author is grateful to one of the reviewers of the present paper for pointing out a major flaw in the original manuscript.

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. *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.), *Foundations of Genetic Algorithms, FOGA 2011*, pp. 15–24.
- Arnold, D. V. (2012). On the behaviour of the $(1, \lambda)$ - σ SA-ES for a constrained linear problem. *Parallel problem solving from nature PPSN XII*, pp. 82–91.
- Arnold, D. V. (2013a). On the behaviour of the $(1, \lambda)$ -ES for a conically constrained problem. In *Genetic and Evolutionary Computation Conference, GECCO 2013*, pp. 423–430.
- Arnold, D. V. (2013b). Resampling versus repair in evolution strategies applied to a constrained linear problem. *Evolutionary Computation*, 21(3):389–411.

- Arnold, D. V., and Beyer, H.-G. (2010). On the behaviour of evolution strategies optimising cigar functions. *Evolutionary Computation*, 18:661–682.
- Arnold, D. V., and Brauer, D. (2008). On the behaviour of the $(1 + 1)$ -ES for a simple constrained problem. In *Parallel Problem Solving from Nature, PPSN X*, pp. 1–10.
- Auger, A., and Hansen, N. (2006). Reconsidering the progress rate theory for evolution strategies in finite dimensions. In *Genetic and Evolutionary Computation Conference, GECCO 2006*, pp. 445–452.
- Balakrishnan, N., and Rao, C. R. (1998). Order statistics: An introduction. In *Handbook of statistics*, Vol. 16 (pp. 3–24). Amsterdam: Elsevier.
- Beyer, H.-G. (1989). *Ein Evolutionsverfahren zur mathematischen Modellierung stationärer Zustände in dynamischen Systemen*. Ph.D. thesis, Hochschule für Architektur und Bauwesen, Weimar.
- Beyer, H.-G., and Schwefel, H.-P. (2002). Evolution strategies—A comprehensive introduction. *Natural Computing*, 1(1):3–52.
- Hansen, N., and Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2): 159–195.
- Meyer-Nieberg, S., and Beyer, H.-G. (2012). The dynamical systems approach—Progress measures and convergence properties. In *Handbook of Natural Computing* (pp. 741–814). Berlin: Springer-Verlag.
- Michalewicz, Z., and Schoenauer, M. (1996). Evolutionary algorithms for constrained parameter optimization problems. *Evolutionary Computation*, 4(1): 1–32.
- Ostermeier, A., Gawelczyk, A., and Hansen, N. (1994). Step-size adaptation based on non-local use of selection information. In *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.
- Porter, J., and Arnold, D. V. (2014). Analyzing the behaviour of multi-recombinative evolution strategies applied to a conically constrained problem. In R. Datta and K. Deb (Eds.), *Evolutionary Constrained Optimization*. Berlin: Springer-Verlag. To appear.
- Rechenberg, I. (1973). *Evolutionsstrategie—Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Friedrich Frommann Verlag.
- Schwefel, H.-P. (1981). *Numerical optimization of computer models*. New York: Wiley.
- Schwefel, H.-P. (1995). *Evolution and optimum seeking*. New York: Wiley.

Appendix A Distribution of Normalized Fitness Difference Values

From Equation (13) with Equation (12), the probability density of normalized fitness difference values is

$$\begin{aligned}
 p_f(w) &= \frac{1}{\sin \alpha} \int_{-\infty}^{\infty} p_{1,2} \left(x, \frac{w - x \cos \alpha}{\sin \alpha} \right) dx \\
 &= \frac{1}{2\pi P_{\text{feasible}} \sin \alpha} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} \exp \left(-\frac{1}{2} \left(\frac{w - x \cos \alpha}{\sin \alpha} \right)^2 \right) \\
 &\quad \times \Phi \left(\frac{\delta_x - \xi \sigma^{*2} + 2\sqrt{\xi/(1 - \xi \vartheta_x^2)} \sigma^* (x - \xi \vartheta_x (w - x \cos \alpha) / \sin \alpha)}{2\xi \sigma^*} \right) dx.
 \end{aligned}$$

Performing a change of variable with $y = (x - w \cos \alpha) / \sin \alpha$ and rearranging terms results in

$$p_f(w) = \frac{1}{2\pi P_{\text{feasible}}} e^{-\frac{1}{2}w^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} \Phi\left(\frac{y-E}{D}\right) dy$$

with $E = -((\delta_x - \xi \sigma^{*2})\sqrt{(1 - \xi \vartheta_x^2)/\xi}/(2\sigma^*) + w(\cos \alpha - \xi \vartheta_x \sin \alpha))/(\sin \alpha + \xi \vartheta_x \cos \alpha)$ and $D = \sqrt{\xi(1 - \xi \vartheta_x^2)}/(\sin \alpha + \xi \vartheta_x \cos \alpha)$. Finally, using identity

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} \Phi\left(\frac{y-E}{D}\right) dy = \Phi\left(\frac{-E}{\sqrt{1+D^2}}\right) \tag{39}$$

(cf. Arnold, 2002, p. 117) gives Equation (14).

Appendix B Computation of the Expected Values

B.1 Computation of $E[\hat{z}_1]$

Performing a change of variable $w = x \cos \alpha + y \sin \alpha$ in Equation (15) yields

$$E[\hat{z}_1] = \frac{\lambda}{\sin \alpha} \int_{-\infty}^{\infty} [1 - P_f(w)]^{\lambda-1} \underbrace{\int_{-\infty}^{\infty} x p_{1,2}\left(x, \frac{w - x \cos \alpha}{\sin \alpha}\right) dx}_{J_1(w)} dw.$$

It is straightforward to show that the inner integral can be solved using Equation (39) and written as

$$J_1(w) = \frac{\sin \alpha}{\cos \alpha - \xi \vartheta_x \sin \alpha} [w p_f(w) + \sin \alpha (\sin \alpha + \xi \vartheta_x \cos \alpha) p'_f(w)]$$

resulting in

$$E[\hat{z}_1] = \frac{\lambda}{\cos \alpha - \xi \vartheta_x \sin \alpha} \int_{-\infty}^{\infty} w p_f(w) [1 - P_f(w)]^{\lambda-1} dw + \frac{\lambda \sin \alpha (\sin \alpha + \xi \vartheta_x \cos \alpha)}{\cos \alpha - \xi \vartheta_x \sin \alpha} \int_{-\infty}^{\infty} p'_f(w) [1 - P_f(w)]^{\lambda-1} dw.$$

Using integration by parts for the second integral and the abbreviations from Equations (18) and (19) gives Equation (20).

B.2 Computation of $E[\hat{z}_2]$

Performing a change of variable $w = x \cos \alpha + y \sin \alpha$ in Equation (16) yields

$$E[\hat{z}_2] = \frac{\lambda}{\sin \alpha} \int_{-\infty}^{\infty} [1 - P_f(w)]^{\lambda-1} \underbrace{\int_{-\infty}^{\infty} \frac{w - x \cos \alpha}{\sin \alpha} p_{1,2}\left(x, \frac{w - x \cos \alpha}{\sin \alpha}\right) dx}_{J_2(w)} dw.$$

It is straightforward to show that the inner integral can be solved using Equation (39) and written as

$$J_2(w) = \frac{\sin \alpha}{\cos \alpha - \xi \vartheta_x \sin \alpha} [-\xi \vartheta_x w p_f(w) - \cos \alpha (\sin \alpha + \xi \vartheta_x \cos \alpha) p'_f(w)]$$

resulting in

$$E[\hat{z}_2] = \frac{-\lambda \xi \vartheta_x}{\cos \alpha - \xi \vartheta_x \sin \alpha} \int_{-\infty}^{\infty} w p_f(w) [1 - P_f(w)]^{\lambda-1} dw - \frac{\lambda \cos \alpha (\sin \alpha + \xi \vartheta_x \cos \alpha)}{\cos \alpha - \xi \vartheta_x \sin \alpha} \int_{-\infty}^{\infty} p'_f(w) [1 - P_f(w)]^{\lambda-1} dw.$$

Using integration by parts for the second integral and the abbreviations from Equations (18) and (19) gives Equation (21).

B.3 Computation of $E[\hat{z}_3]$

Performing a change of variable $w = x \cos \alpha + y \sin \alpha$ in Equation (17) yields

$$E[\hat{z}_3] = \frac{\lambda}{\sin \alpha} \int_{-\infty}^{\infty} [1 - P_f(w)]^{\lambda-1} \underbrace{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z p_{1,2,3} \left(x, \frac{w - x \cos \alpha}{\sin \alpha}, z \right) dz dx}_{J_3(w)} dw.$$

It is straightforward to show that the inner integrals can be solved using Equation (39) and written as

$$J_3(w) = \frac{-\sin \alpha \sqrt{\xi(1 - \xi \vartheta_x^2)}}{\cos \alpha - \xi \vartheta_x \sin \alpha} [w p_f(w) + p'_f(w)]$$

resulting in

$$E[\hat{z}_3] = \frac{-\sqrt{\xi(1 - \xi \vartheta_x^2)}}{\cos \alpha - \xi \vartheta_x \sin \alpha} \left[\int_{-\infty}^{\infty} (w p_f(w) + p'_f(w)) [1 - P_f(w)]^{\lambda-1} dw \right].$$

Using integration by parts and the abbreviations from Equations (18) and (19) gives Equation (22).

