Analysis of a Repair Mechanism for the $(1, \lambda)$-ES Applied to a Simple Constrained Problem

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

ABSTRACT

We study the behaviour of a $(1, \lambda)$-ES that applies a simple repair mechanism to infeasible candidate solutions for the problem of maximising a linear function with a single linear constraint. Integral expressions that describe the strategy's one-generation behaviour are derived and used in a simple zeroth order model for the steady state of the strategy. Applied to the analysis of cumulative step size adaptation, the approach provides an intuitive explanation for the algorithm's behaviour as well as a condition on the setting of its parameters. A comparison with the strategy that resamples infeasible candidate solutions rather than repairing them is drawn, and the qualitatively different behaviour is explained.

Categories and Subject Descriptors

I.2.8 [**Problem Solving, Control Methods and Search**]; G.1.6 [**Optimization**]: Constrained Optimization

General Terms

Algorithms

Keywords

Constrained optimisation, evolution strategy, step size adaptation

1. INTRODUCTION

A multitude of constraint handling techniques have been proposed for use in evolutionary algorithms. The range of strategies includes penalty approaches, repair mechanisms, and algorithms based on ideas from multi-objective optimisation. A survey of techniques has been compiled by Coello Coello [9]. Approaches used in the context of evolution strategies include those described by Oyman et al. [14], Runarsson and Yao [16, 17], Mezura-Montes and Coello Coello [12], and Kramer and Schwefel [10].

GECCO'11, July 12–16, 2011, Dublin, Ireland.

Knowledge of the relative capabilities and limitations of the approaches is most often gained from comparing their performances on large and diverse sets of test functions, such as the benchmark set compiled for the *CEC 2010 Special Session on Constrained Real-Parameter Optimization* [11]. The evaluation criteria used are often relatively complex and involve various parameters, such as the number of function evaluations and different quality thresholds. As a result, the observed outcomes are not always easy to interpret, and if a mechanism fails it often remains unclear why. Likewise, it is not always straightforward what "good" performance for a problem means.

This paper pursues a complementary approach in which the behaviour of an algorithm is not studied on a large test bed, but instead 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. Of particular significance is the interaction of constraint handling technique and step size adaptation approach. Most of the latter have been developed with unconstrained problems in mind, and it has been seen in previous work that the presence of constraints may lead step size adaptation to fail even on the simplest problems. As it appears likely that robust performance on complex problems requires an algorithm to be able to succeed on simple ones, work of the nature described here presents a concrete challenge to the algorithm designer to devise such strategies.

While a sizable number of studies have focused on the performance of evolution strategies for simple classes of unconstrained problems, little such work has been published in constrained settings. Among the sparse references is the early work by Rechenberg [15], who studies the performance of the $(1 + 1)$-ES[1] for the axis-aligned corridor model, and upon which the formulation of the $1/5$th success rule for step size adaptation is partly based. Schwefel [18] considers the performance of the $(1, \lambda)$-ES in the same environment. Beyer [7] 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.

Two more recent studies consider constrained problems that do not have that specific property. Arnold and Brauer [5] study the behaviour of the $(1 + 1)$-ES for a linear

[1]See Beyer and Schwefel [8] for an overview of evolution strategy terminology.

problem with a single linear constraint of general orientation. They describe the distance of the parental candidate solution from the constraint plane using a Markov chain approach and investigate the limit behaviour of the chain using two simple statistical models. Schwefel [19, page 116f] points out that using the 1/5th rule, the presence of constraints may lead to 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 non-singular point. The findings derived in [5] provide a quantitative confirmation of this. Arnold [3] studies the behaviour of a $(1, \lambda)$-ES that samples offspring candidate solutions until λ feasible ones have been generated in the same environment, using a similar approach. He finds that cumulative step size adaptation may fail in scenarios similar to those where the success probability based mechanism fails, though for different reasons.

The algorithm considered in [3] does not make any assumptions with regard to the constraints other than the availability of a black-box function that can be used to determine whether a candidate solution is feasible or not. If a candidate solution is infeasible, it is resampled. The goal of this paper is to investigate the consequences of being able to "repair" an infeasible candidate solution by projecting it orthogonally onto the boundary of the feasible region. It will be seen that the ability to repair infeasible offspring can result in qualitatively different behaviour of the strategy.

The remainder of this paper is organised as follows. Section 2 briefly formalises the algorithm and problem and introduces notational conventions used throughout the paper. Section 3 considers the single-step behaviour of the strategy and derives probability distributions that characterise the offspring candidate solutions as well as an expression for the progress rate of the strategy. Section 4 investigates the multi-step limit behaviour of the strategy for a fixed step size employing a simple zeroth-order model for the distribution of distances from the constraint plane. Section 5 considers cumulative step size adaptation and compares with the behaviour of the strategy studied in [3]. Section 6 concludes with a brief discussion and goals for future research.

2. PROBLEM AND ALGORITHM

Throughout this paper, we consider the problem of maximising[2] a linear function $f : \mathbb{R}^N \to \mathbb{R}$, $N \geq 2$, 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 Fig. 1, 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 $\mathbf{n} = \langle \cos\theta, \sin\theta, 0, \ldots, 0 \rangle$. The signed distance of a point $\mathbf{x} = \langle x_1, x_2, \ldots, x_N \rangle \in \mathbb{R}^N$ from the constraint plane

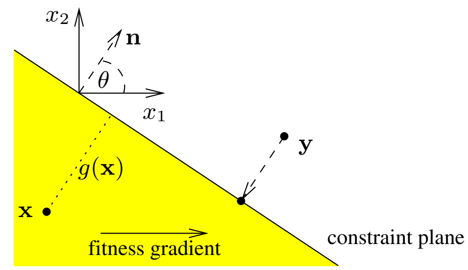

Figure 1: Linear objective function with a single linear constraint. The subspace spanned by the x_1- and x_2-axes is shown. The shaded area is the feasible region. The parental candidate solution x of the $(1, \lambda)$-ES is at a distance $g(\mathbf{x})$ from the constraint plane. Infeasible candidate solution y is projected orthogonally onto 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

$$\text{maximise} \quad f(\mathbf{x}) = cx_1 \quad \text{subject to} \quad g(\mathbf{x}) \geq 0$$

for some constant $c > 0$. Notice that due to the choice of coordinate system, variables $x_3, x_4, \ldots, x_N$ enter neither the objective function nor the constraint inequality.

Assuming a feasible initial candidate solution $\mathbf{x} \in \mathbb{R}^N$, the $(1, \lambda)$-ES generates a sequence of further candidate solutions by iterating the following steps:

1. For $i = 1, \ldots, \lambda$
   (a) Generate offspring candidate solution $\mathbf{y}^{(i)} \in \mathbb{R}^N$ by sampling an N-dimensional normal distribution with mean $\mathbf{x}$ and covariance matrix $\sigma^2 \mathbf{1}_{N\times N}$, where $\mathbf{1}_{N\times N}$ is the $N \times N$ identity matrix and $\sigma \in \mathbb{R}$ is referred to as the mutation strength.
   (b) If $\mathbf{n} \cdot \mathbf{y}^{(i)} > 0$ then project $\mathbf{y}^{(i)}$ onto the constraint plane by letting
   $$\mathbf{y}^{(i)} \leftarrow \mathbf{y}^{(i)} - (\mathbf{n} \cdot \mathbf{y}^{(i)})\mathbf{n}. \qquad (1)$$
   (c) Compute $f(\mathbf{y}^{(i)})$.
2. Replace the parental candidate solution with the offspring candidate solution that has the largest objective function value.
3. Modify the mutation strength.

Termination criteria are irrelevant in the context of this study. As a performance measure, we consider the expected per step distance covered in the direction of the gradient of the objective function, which is referred to as the progress rate.

3. SINGLE-STEP BEHAVIOUR

This section first describes the operation of the strategy as a Markov process. It then provides a characterisation of the distribution of offspring candidate solutions and finally considers the expected step made in a single iteration of the algorithm. In all of what follows, superscripts refer to iteration number rather than the numbering of offspring. Values associated with the offspring candidate solution that replaces the parent are indicated with a hat.

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

3.1 Markov Process

The generation of offspring candidate solutions is a two-step process that involves random sampling and potentially projection. The sampling of an offspring candidate solution $\mathbf{y}$ in step 1(a) of the algorithm in Section 2 involves generating standard normally distributed mutation vector $\mathbf{z} = \langle z_1, z_2, \ldots, z_N\rangle = (\mathbf{y} - \mathbf{x})/\sigma$. If

$$z_1 \cos\theta + z_2 \sin\theta > \delta \tag{2}$$

where $\delta = g(\mathbf{x})/\sigma$ is the normalised distance of the parental candidate solution from the constraint plane, then $\mathbf{y}$ is infeasible and projected orthogonally onto that plane. Let $\mathbf{w} = \langle w_1, w_2, \ldots, w_N\rangle = (\mathbf{y} - \mathbf{x})/\sigma$, where $\mathbf{y}$ refers to the offspring after step 1(b), denote the mutation vector after projection. From Eq. (1), its first two components are

$$w_1 = \begin{cases} \delta\cos\theta + (z_1\cos\theta - z_2\sin\theta)\sin\theta & \text{if Eq. (2) holds} \\ z_1 & \text{otherwise} \end{cases}$$

and

$$w_2 = \begin{cases} \delta\sin\theta - (z_1\sin\theta - z_2\cos\theta)\cos\theta & \text{if Eq. (2) holds} \\ z_2 & \text{otherwise} \end{cases}$$

All other components of $\mathbf{w}$ equal those of $\mathbf{z}$.

Until Section 5, we consider the behaviour of the strategy for fixed mutation strength. In that case, its state is completely described by the normalised distance δ of the parental candidate solution from the constraint plane. Simple trigonometry reveals that the evolution of δ is described by

$$\delta^{(t+1)} = \delta^{(t)} - \hat{w}_1\cos\theta - \hat{w}_2\sin\theta \tag{3}$$

where $\hat{\mathbf{w}} = \langle \hat{w}_1, \hat{w}_2, \ldots, \hat{w}_N\rangle$ is the mutation vector after projection corresponding to the offspring candidate solution with the largest objective function value.

3.2 Offspring Distribution

The joint distribution of the w_1- and w_2-components of the mutation vectors after projection is a mixture of the distribution of the corresponding components of mutation vectors of immediately feasible offspring and those projected onto the constraint plane. Let us consider the marginal distribution of the w_1-components. The contribution to the density $p_1(x)$ of that distribution from the immediately feasible offspring is obtained by integrating the joint density for $w_1 = x$ up to the constraint plane and thus equals

$$\frac{1}{2\pi}\int_{-\infty}^{(\delta - x\cos\theta)/\sin\theta} \mathrm{e}^{-\frac{1}{2}(x^2+y^2)}\mathrm{d}y$$

as the plane with $w_1 = x$ intersects the constraint plane at $w_2 = (\delta - x\cos\theta)/\sin\theta$. The contribution from the initially infeasible and thus projected offspring candidate solutions is obtained by integration along the line with $x = \delta\cos\theta + (w_1\sin\theta - w_2\cos\theta)\sin\theta$, on which all of the points projected onto the constraint plane at location $w_1 = x$ lie, and it equals

$$\frac{1}{2\pi\sin\theta\cos\theta}\int_x^\infty \mathrm{e}^{-\frac{1}{2}y^2} \exp\left(-\frac{1}{2}\left(y\tan\theta - \frac{x - \delta\cos\theta}{\sin\theta\cos\theta}\right)^2\right)\mathrm{d}y.$$

Adding both contributions and solving the integrals yields the marginal density function

$$p_1(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x^2}\Phi\left(\frac{\delta - x\cos\theta}{\sin\theta}\right) + \frac{1 - \Phi(\delta)}{\sqrt{2\pi}\sin\theta}\exp\left(-\frac{1}{2}\left(\frac{x - \delta\cos\theta}{\sin\theta}\right)^2\right) \tag{4}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. Integration of the density yields the cumulative distribution function

$$P_1(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x \mathrm{e}^{-\frac{1}{2}y^2}\Phi\left(\frac{\delta - y\cos\theta}{\sin\theta}\right)\mathrm{d}y + (1 - \Phi(\delta))\,\Phi\left(\frac{x - \delta\cos\theta}{\sin\theta}\right) \tag{5}$$

of the w_1-components of mutation vectors after projection.

Next, consider the distribution of the w_2-components of mutation vectors after projection conditional on $w_1 = x$. The probability of an offspring being infeasible and thus in need of repair conditional on $w_1 = x$ is obtained by computing the relative weight of the second term in the marginal density $p_1(x)$ in Eq. (4) and equals

$$\begin{aligned} P_{\text{inf}}(x) &= \text{Prob}\left[z_1\cos\theta + z_2\sin\theta > \delta | w_1 = x\right] \\ &= \frac{1 - \Phi(\delta)}{\sqrt{2\pi}p_1(x)\sin\theta}\exp\left(-\frac{1}{2}\left(\frac{x - \delta\cos\theta}{\sin\theta}\right)^2\right). \end{aligned}$$

All initially infeasible candidate solutions after projection have w_2-coordinate $(\delta - x\cos\theta)/\sin\theta$ and thus

$$\mathrm{E}\left[w_2 | w_1 = x \text{ and } z_1\cos\theta + z_2\sin\theta > \delta\right] = \frac{\delta - x\cos\theta}{\sin\theta}.$$

Initially feasible offspring candidate solutions are not projected and have expected value

$$\begin{aligned} &\mathrm{E}\left[w_2 | w_1 = x \text{ and } z_1\cos\theta + z_2\sin\theta \leq \delta\right] \\ &= \frac{1}{\sqrt{2\pi}\Phi((\delta - x\cos\theta)/\sin\theta)}\int_{-\infty}^{(\delta - x\cos\theta)/\sin\theta} y\mathrm{e}^{-\frac{1}{2}y^2}\mathrm{d}y \\ &= \frac{-1}{\sqrt{2\pi}\Phi((\delta - x\cos\theta)/\sin\theta)}\exp\left(-\frac{1}{2}\left(\frac{\delta - x\cos\theta}{\sin\theta}\right)^2\right). \end{aligned}$$

Weighting both terms with the probability of their occurrence yields

$$\begin{aligned} &\mathrm{E}\left[w_2 | w_1 = x\right] \\ &= (1 - P_{\text{inf}}(x))\mathrm{E}\left[w_2 | w_1 = x \text{ and } z_1\cos\theta + z_2\sin\theta \leq \delta\right] \\ &\quad + P_{\text{inf}}(x)\mathrm{E}\left[w_2 | w_1 = x \text{ and } z_1\cos\theta + z_2\sin\theta > \delta\right] \end{aligned}$$

Combining the results and simplifying allows deriving the expression

$$\begin{aligned} &\mathrm{E}\left[w_2 | w_1 = x\right] \\ &= \frac{1}{p_1(x)}\left[-\frac{1}{2\pi}\mathrm{e}^{-\frac{1}{2}x^2}\exp\left(-\frac{1}{2}\left(\frac{\delta - x\cos\theta}{\sin\theta}\right)^2\right)\right. \\ &\quad \left. + \frac{1 - \Phi(\delta)}{\sqrt{2\pi}\sin\theta}\frac{\delta - x\cos\theta}{\sin\theta}\exp\left(-\frac{1}{2}\left(\frac{x - \delta\cos\theta}{\sin\theta}\right)^2\right)\right] \end{aligned} \tag{6}$$

for the expected value of the w_2-component of a mutation vector after projection conditional on $w_1 = x$.

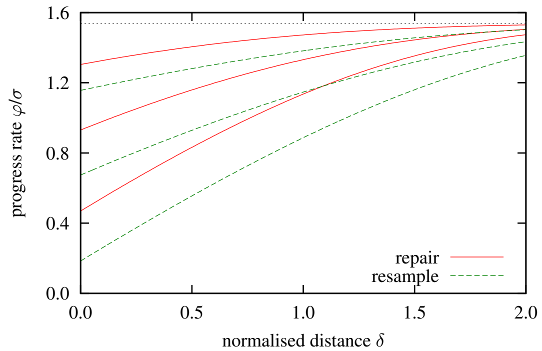

Figure 2: Progress rate φ divided by the mutation strength σ plotted against the normalised distance δ of the parent from the constraint plane for $\lambda = 10$ and, from bottom to top, $\theta \in \{\pi/8, \pi/4, 3\pi/8\}$.

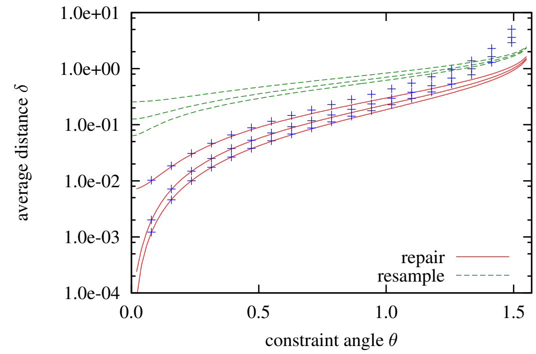

Figure 3: Average normalised distance δ from the constraint plane plotted against the constraint angle θ for, from top to bottom, $\lambda \in \{5, 10, 20\}$.

3.3 Expected Step

The w_1-component of the mutation vector after projection that results in the best offspring candidate solution is the λth order statistic of the sample of w_1-values of the entire set of offspring. Its probability density is

$$\hat{p}_1(x) = \lambda p_1(x) P_1^{\lambda-1}(x).$$

The expected value of the w_1-component of the selected offspring candidate solution is thus

$$\begin{aligned} \mathrm{E}[\hat{w}_1] &= \int_{-\infty}^{\infty} x\hat{p}_1(x)\,\mathrm{d}x \\ &= \lambda \int_{-\infty}^{\infty} x p_1(x) P_1^{\lambda-1}(x)\,\mathrm{d}x \qquad (7) \end{aligned}$$

and can be computed numerically using Eqs. (4) and (5). The expected value of the w_2-component of the selected offspring is

$$\begin{aligned} \mathrm{E}[\hat{w}_2] &= \int_{-\infty}^{\infty} \mathrm{E}[w_2|w_1 = x]\,\hat{p}_1(x)\,\mathrm{d}x \\ &= \lambda \int_{-\infty}^{\infty} \mathrm{E}[w_2|w_1 = x]\,p_1(x) P_1^{\lambda-1}(x)\,\mathrm{d}x \qquad (8) \end{aligned}$$

and can be computed numerically using Eqs. (4), (5) and (6).

The progress rate

$$\varphi = \mathrm{E}\left[x_1^{(t+1)} - x_1^{(t)}\right]$$

of the strategy, i.e., the expected step taken in the direction of the gradient of the objective function, is the product of the mutation strength and $\mathrm{E}[\hat{w}_1]$. It is shown as a function of the normalised distance of the parental candidate solution from the constraint plane in Fig. 2. The solid lines in that figure have been obtained from Eq. (7) for $\lambda = 10$ and three different values of the constraint angle. The dashed lines depict corresponding results for the strategy that resamples infeasible offspring instead of repairing them and have been derived in [3]. It can be seen that for any value of δ, the strategy that repairs infeasible offspring makes greater progress in the direction of the gradient of the objective function. At the same time, progress rates increase with increasing distance from the constraint plane. For large δ, all curves approach the same value attained in the absence of constraints.

4. STEADY STATE BEHAVIOUR

The results derived thus far consider single time steps only and are conditional on the distance of the parental candidate solution from the constraint plane. Assuming for now that the mutation strength σ is held constant, when the algorithm is iterated, the distribution of δ-values tends to a stationary limit distribution. As done in [3] for the strategy that resamples instead of repairing infeasible candidate solutions, as a zeroth order approximation let us assume that the limit distribution is well characterised by its mean, and that higher order moments can be neglected. That is, we model the limit distribution as a (shifted) Dirac delta function and compute its mode by requiring that

$$\mathrm{E}\left[\delta^{(t+1)}\right] = \delta^{(t)}.$$

According to Eq. (3), this results in stationarity condition

$$\mathrm{E}[\hat{w}_1]\cos\theta + \mathrm{E}[\hat{w}_2]\sin\theta = 0.$$

Using Eqs. (7) and (8) and solving numerically for δ yields the solid curves shown in Fig. 3. The dashed lines again represent results from [3] for the strategy that resamples infeasible offspring rather than repairing them. The points mark measurements obtained by averaging over 10^6 steps of the strategy described in Section 2. Comparing the solid lines and points in the figure, the quality of the zeroth order approximation provided by the delta Dirac model appears visually good for small constraint angles, but it deteriorates markedly with increasing θ. That is, in situations where the strategy closely tracks the constraint plane the behaviour of the algorithm is quite well described by the model that does not include variations of that distance. As the constraint angle becomes less acute and the strategy tracks the constraint plane at a greater distance, variations in that distance become significant and the delta Dirac model is inappropriate. The same effect has been observed in [3] for the strategy that resamples infeasible offspring rather than repairing them. Presumably, using an exponential model for the distribution of distances as employed in [5] for the $(1+1)$-ES would allow deriving more accurate values.

It can be seen from Fig. 3 that for both strategies, the average distance from the constraint plane decreases with increasing λ, and it increases with increasing constraint angle θ. Unsurprisingly, as its candidate solutions will often lie

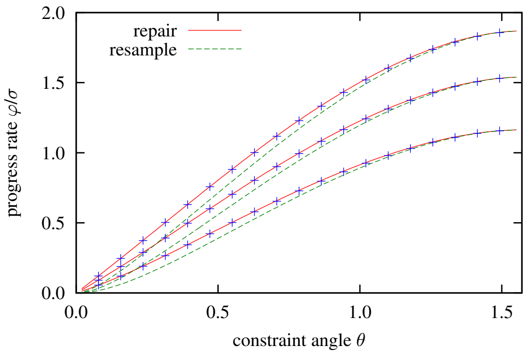

Figure 4: Progress rate φ divided by the mutation strength σ plotted against the constraint angle θ for, from bottom to top, $\lambda \in \{5, 10, 20\}$.

on the constraint plane, the strategy that projects infeasible offspring onto that plane tracks it much more closely than the one that resamples.

The solid curves in Fig. 4 have been obtained by using the δ-value predicted by the delta Dirac model in Eq. (7) in order to compute the progress rate of the $(1, \lambda)$-ES. As above, the dashed lines represent results from [3] for the strategy that resamples infeasible offspring candidate solutions rather than repairing them. The points mark measurements averaged over 10^6 iterations of the strategy described in Section 2. As observed in [3], the quality of the approximation of the progress rate provided by the delta Dirac model is visually good for the entire range of constraint angles. While for larger values of θ, Fig. 3 has shown that the approximation of the average value of δ is very inaccurate, the influence of the distances from the constraint plane on the progress rate becomes weaker as the slope of the curves in Fig. 2 decreases with increasing δ.

It can be seen from Fig. 4 that the progress rate increases with increasing θ and λ. Differences in performance between the strategy that repairs infeasible offspring and the strategy that resamples are significant mostly for small constraint angles, where progress is slow in either case. Even though the strategy that repairs infeasible candidate solutions tracks the constraint plane more closely (and thus operates farther to the left hand side in the graph in Fig. 2), it generally exhibits larger progress rates than the strategy that resamples. The difference appears quantitative rather than qualitative.

5. MUTATION STRENGTH ADAPTATION

The mutation strength of the $(1, \lambda)$-ES can be adapted using one of several mechanisms, including mutative self-adaptation, two-point adaptation, and cumulative step size adaptation. The latter mechanism has been introduced by Ostermeier et al. [13] and is popular due to its use in the CMA-ES. It has been studied for the sphere model in [1, 4], for a class of further convex quadratic functions in [2], and for ridge functions in [6]. In the case of the $(1, \lambda)$-ES, cumulative step size adaptation employs a search path $\mathbf{s} \in \mathbb{R}^N$ defined by $\mathbf{s}^{(0)} = \mathbf{0}$ and

$$\mathbf{s}^{(t+1)} = (1 - c)\mathbf{s}^{(t)} + \sqrt{c(2 - c)}\hat{\mathbf{w}} \tag{9}$$

that implements an exponentially fading record of past steps taken by the strategy. 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. The mutation strength is updated according to[3]

$$\sigma^{(t+1)} = \sigma^{(t)} \exp\left(c\frac{\|\mathbf{s}^{(t+1)}\|^2 - N}{2DN}\right) \tag{10}$$

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 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 a squared length of N if consecutive steps are perpendicular on average, in which case no change in mutation strength is effected.

For the linear, constrained environment considered here, the $(1, \lambda)$-ES with cumulative step size adaptation does not assume a stationary limit state. The mutation strength is either increased or decreased on average. As the progress rate of the strategy is positive and proportional to the mutation strength, increasing σ is desirable while decreasing it leads to stagnation and convergence to a non-singular point. Similarly to [1] we define the logarithmic adaptation response of the strategy as

$$\Delta_\sigma^{(t+1)} = D \log\left(\frac{\sigma^{(t+1)}}{\sigma^{(t)}}\right).$$

Rather than attempting to solve the difficult problem of examining the dynamic behaviour of the non-linear stochastic process generated by the algorithm, we consider the logarithmic adaptation response of the $(1, \lambda)$-ES operating out of a stationary state. That is, we assume that the strategy has been run with a fixed mutation strength until time step t, where t is large, and we compute the expected logarithmic adaptation response at time step $t + 1$. The same approach has been pursued in [3] for the strategy that resamples infeasible offspring candidate solutions rather than repairing them. The derivation in that reference holds without changes for the algorithm considered here and yields expression

$$\mathrm{E}\left[\Delta_\sigma\right] = \frac{c}{2N}\left[e_{1,2} + e_{2,2} + 2\frac{1 - c}{c}\left(e_{1,1}^2 + e_{2,1}^2\right) - 2\right] \tag{11}$$

where

$$e_{i,j} = \mathrm{E}\left[\hat{w}_i^j\right]$$

for the expected logarithmic adaptation response. Expressions for $e_{1,1}$ and $e_{2,1}$ have been derived in Eqs. (7) and (8), respectively. Similarly,

$$e_{1,2} = \mathrm{E}\left[\hat{w}_1^2\right] = \lambda \int_{-\infty}^{\infty} x^2 p_1(x) P_1^{\lambda-1}(x)\,\mathrm{d}x$$

[3]Notice that this update differs from the original prescription in that [13] adapts 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. Basing the update of the mutation strength on the squared length of the search path simplifies the analysis.

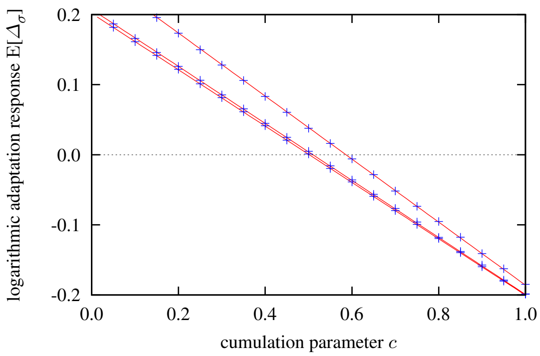

Figure 5: Expected logarithmic adaptation response $\mathrm{E}[\Delta_\sigma]$ plotted against cumulation parameter c for $N = 2$, $\lambda = 5$, and, from top to bottom, $\theta \in \{10^{-1}, 10^{-2}, 10^{-3}\}$.

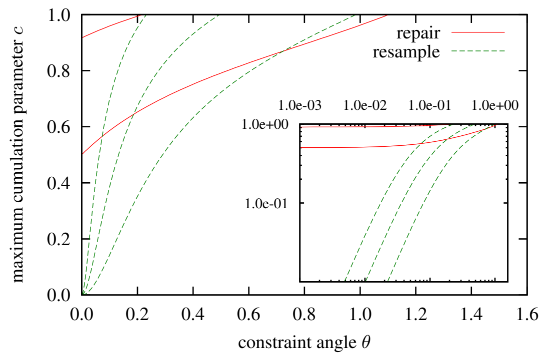

Figure 6: Maximum value of the cumulation parameter c for which the expected logarithmic adaptation response is positive plotted against the constraint angle θ for, from bottom to top, $\lambda \in \{5, 10, 20\}$. The inset shows a log-log plot of the same curves.

and

$$e_{2,2} = \mathrm{E}\left[\hat{w}_2^2\right] = \lambda \int_{-\infty}^{\infty} \mathrm{E}\left[w_2^2 | w_1 = x\right] p_1(x) P_1^{\lambda-1}(x)\,\mathrm{d}x$$

where

$$\begin{aligned} &\mathrm{E}\left[w_2^2 | w_1 = x\right] \\ &\quad = (1 - P_{\mathrm{inf}}(x))\mathrm{E}\left[w_2^2 | w_1 = x \text{ and } z_1 \cos\theta + z_2 \sin\theta \leq \delta\right] \\ &\qquad + P_{\mathrm{inf}}(x)\mathrm{E}\left[w_2^2 | w_1 = x \text{ and } z_1 \cos\theta + z_2 \sin\theta > \delta\right] \end{aligned}$$

with

$$\begin{aligned} &\mathrm{E}\left[w_2^2 | w_1 = x \text{ and } z_1 \cos\theta + z_2 \sin\theta \leq \delta\right] \\ &\quad = \frac{1}{\sqrt{2\pi}\Phi((\delta - x\cos\theta)/\sin\theta)} \int_{-\infty}^{(\delta - x\cos\theta)/\sin\theta} y^2 \mathrm{e}^{-\frac{1}{2}y^2} \mathrm{d}y \\ &\quad = 1 - \frac{1}{\sqrt{2\pi}} \frac{\delta - x\cos\theta}{\sin\theta} \frac{\exp(-((\delta - x\cos\theta)/\sin\theta)^2/2)}{\Phi((\delta - x\cos\theta)/\sin\theta)} \end{aligned}$$

and

$$\mathrm{E}\left[w_2^2 | w_1 = x \text{ and } z_1 \cos\theta + z_2 \sin\theta > \delta\right] = \left(\frac{\delta - x\cos\theta}{\sin\theta}\right)^2.$$

Both coefficients can be computed numerically using Eqs. (4) and (5).

The expected logarithmic adaptation response is plotted against the cumulation parameter c for $N = 2$, $\lambda = 5$, and several constraint angles in Fig. 5. The lines depict results from Eq. (11), where values from the Dirac delta model have been used for the average distance of the strategy from the constraint plane. The points mark measurements made in runs of the strategy in which δ is free to vary while σ remains fixed. It can be seen that the expected logarithmic adaptation response decreases with increasing cumulation parameter, and that it may assume negative values, signalling convergence to a non-stationary point, if c becomes too large.

Of particular interest is the zero crossing of the curves in Fig. 5 as it represents the maximum value of the cumulation parameter c for which the expected logarithmic adaptation response is positive. It can be computed from Eq. (11), which yields

$$c \leq \frac{e_{1,1}^2 + e_{2,1}^2}{1 + e_{1,1}^2 + e_{2,1}^2 - e_{1,2}/2 - e_{2,2}/2} \qquad (12)$$

when set to zero and solved for c. That relationship is illustrated with solid lines in Fig. 6. The dashed lines once again represent corresponding results from [3] for the strategy that resamples infeasible offspring rather than repairing them. Note that for the strategy that applies the repair mechanism and $\lambda = 20$, the value obtained from Eq. (12) exceeds 1.0 for any constraint angle, and that the corresponding curve therefore does not show up in the figure.

The qualitative difference between the curves for the strategy that resamples infeasible offspring and the one that applies the repair mechanism is striking. For the former, Fig. 6 illustrates that as the constraint angle becomes smaller and smaller, the range of cumulation parameter values for which convergence to a non-stationary point is avoided shrinks as well. For the strategy that repairs infeasible offspring on the other hand, Eq. (12) predicts that the value of the cumulation parameter that separates the two regimes approaches a finite (and not very small) limit value as θ decreases.

The different behaviours of the two strategies can be explained from Eq. (11). In the limit of very small constraint angles, $e_{1,1}$ and $e_{1,2}$ assume very small values for both strategies as both stay in immediate vicinity of the constraint plane (compare Figs. 3 and 4). In order to have a positive logarithmic adaptation response, the terms involving $e_{2,1}$ and $e_{2,2}$ in Eq. (11) must be large enough to compensate for the lack of a contribution to the 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 $\hat{w}_2$, resulting in $e_{2,1} \approx 0.0$ and $e_{2,2} \approx 1.0$, which is insufficient to ensure positive logarithmic adaptation response unless c is small enough in order to sufficiently amplify the term involving $e_{1,1}^2$ in Eq. (11). The strategy that repairs infeasible offspring on the other hand does not perform a random walk in $\hat{w}_2$. In the limit of very small θ, on average half of the offspring are infeasible and will be projected onto the constraint plane. In the mean, half of those that are projected have a negative w_2-component and are likely to prevail under selection as their w_1-components are positive. Unless λ is very small, the majority of steps 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 logarithmic adaptation response. 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. Projection of mutation vectors never increases their length (i.e., $\|\mathbf{w}\| \leq \|\mathbf{z}\|$), and thus any bias as a result of projection is toward smaller step sizes, not toward larger ones. The positive 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, to confirm the validity of the results that have been derived, Fig. 7 shows observations of runs of $(1, \lambda)$-ES in search spaces with constraint angle $\theta = 10^{-6}$, $N \in \{2, 10\}$, $\lambda \in \{5, 10, 20\}$, $D = 1.0$, and different settings of the cumulation parameter c. The constraint angle has deliberately been chosen small in order to test the strategy in those situations where the algorithm that resamples infeasible offspring is known to fail unless c is chosen very small. All curves have been obtained by averaging mutation strengths geometrically over 100 independent runs. For $\lambda = 5$, Eq. (12) suggests that the point that separates the runs that result in increasing mutation strength from those where the step size is decreased is approximately 0.5. Indeed, it can be observed that for $c = 0.6$ convergence to a non-stationary point occurs, while for $c = 0.4$ the step size is increased as desired. The slope of the curves for $N = 2$ is steeper than the slope of those for $N = 10$ due to the presence of N in the denominator in Eq. (11). The same behaviour is observed for $\lambda = 10$ with $c = 0.85$ and $c = 0.95$, where according to Eq. (12) the point separating the two regimes should be approximately 0.9. For $\lambda = 20$, Eq. (12) predicts that increasing step sizes will be observed even for $c = 1.0$, and that behaviour is indeed observed in the experiments.

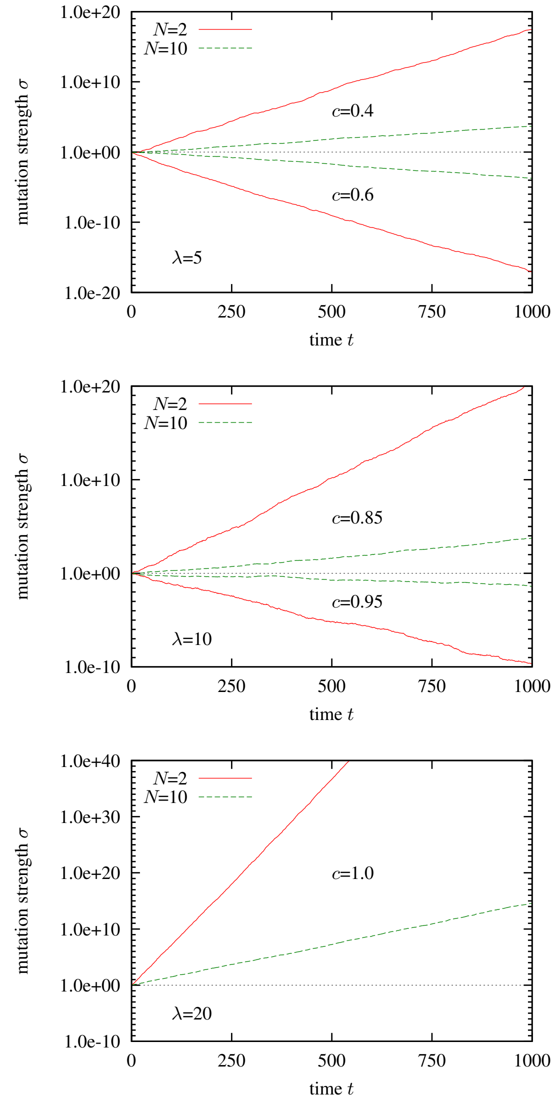

Figure 7: Mutation strength σ plotted against time for runs of the $(1, \lambda)$-ES with cumulative step size adaptation and different combinations of the number λ of offspring generated per time step, search space dimension N, and the cumulation parameter c.

6. SUMMARY AND CONCLUSIONS

To conclude, in this paper we have studied the behaviour of a $(1, \lambda)$-ES with cumulative step size adaptation that repairs infeasible candidate solutions by orthogonally projecting them onto the boundary of the feasible region for a simple linear problem with a single linear constraint. For fixed mutation strength, the normalised distance of the parental candidate solution from the constraint plane is the only state variable of the stochastic process that describes the operation of the strategy. It has been possible to derive exact expressions that characterise the single step behaviour of the algorithm and are conditional on the normalised distance from the constraint plane. The multi-step behaviour forms a non-linear stochastic process and is not as easily determined. Under the assumption of constant mutation strength, a simple zeroth order model has been used to compute the approximate average normalised distance of the strategy from the constraint plane. That value has been used to compute approximate values of the progress rate of the strategy as well as its expected logarithmic adaptation response when operating out of the stationary state.

The most interesting finding of this paper is that the strategy that repairs infeasible candidate solutions exhibits qualitatively differently behaviour from the one that resamples infeasible offspring. For the latter strategy, it has been found in [3] that for any value of the cumulation parameter, cumulative step size adaptation will systematically reduce the mutation strength if the constraint angle becomes too small, resulting in convergence to a non-stationary point. When repairing infeasible candidate solutions on the other hand, it has been possible to derive approximate values for the cumulation parameter that guarantee that convergence to a non-stationary point does not occur, no matter how small the constraint angle. The difference in the behaviour of the two strategies 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.

Opportunities for extending this work are manifold. The exponential model in combination with the Kullback-Leibler divergence used for modelling the steady state of the $(1+1)$-ES in [5] can be used in an attempt to obtain a better approximation of the distribution of normalised distances from the constraint plane. Further future work includes the extension of the results of the paper to the more general $(\mu/\mu, \lambda)$-ES, which performs recombination of several selected candidate solutions. It also remains to consider other step size adaptation mechanisms, such as mutative self-adaptation, and non-linear objective functions where the local constraint angle is not constant.

ACKNOWLEDGEMENTS

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Foundation for Innovation (CFI).

7. REFERENCES

[1] D. V. Arnold. *Noisy Optimization with Evolution Strategies*. Kluwer Academic Publishers, 2002.

[2] D. V. Arnold. On the use of evolution strategies for optimising certain positive definite quadratic forms. In *Genetic and Evolutionary Computation Conference — GECCO 2007*, pages 634–641. ACM Press, 2007.

[3] D. V. Arnold. On the behaviour of the $(1, \lambda)$-ES for a simple constrained problem. In H.-G. Beyer and W. B. Langdon, editors, *Foundations of Genetic Algorithms — FOGA 2011*, pages 15–24. ACM Press, 2011.

[4] D. V. Arnold and H.-G. Beyer. Performance analysis of evolutionary optimization with cumulative step length adaptation. *IEEE Transactions on Automatic Control*, 49(4):617–622, 2004.

[5] D. V. Arnold and D. Brauer. On the behaviour of the $(1 + 1)$-ES for a simple constrained problem. In G. Rudolph et al., editors, *Parallel Problem Solving from Nature — PPSN X*, pages 1–10. Springer Verlag, 2008.

[6] D. V. Arnold and A. MacLeod. Step length adaptation on ridge functions. *Evolutionary Computation*, 16(2):151–184, 2008.

[7] H.-G. Beyer. *Ein Evolutionsverfahren zur mathematischen Modellierung stationärer Zustände in dynamischen Systemen*. PhD thesis, Hochschule für Architektur und Bauwesen, Weimar, 1989.

[8] H.-G. Beyer and H.-P. Schwefel. Evolution strategies — A comprehensive introduction. *Natural Computing*, 1(1):3–52, 2002.

[9] C. A. Coello Coello. Constraint-handling techniques used with evolutionary algorithms. In *Proceedings of the 2008 GECCO Conference Companion on Genetic and Evolutionary Computation*, pages 2445–2466. ACM Press, 2008.

[10] O. Kramer and H.-P. Schwefel. On three new approaches to handle constraints within evolution strategies. *Natural Computing*, 5(4):363–385, 2006.

[11] R. Mallipeddi and P. N. Suganthan. Problem definitions and evaluation criteria for the CEC 2010 Competition on Constrained Real-Parameter Optimization. Technical report, Nanyang Technological University, Singapore, 2010.

[12] E. Mezura-Montes and C. A. Coello Coello. A simple multi-membered evolution strategy to solve constrained optimization problems. *IEEE Transactions on Evolutionary Computation*, 9(1):1–17, 2005.

[13] A. Ostermeier, A. Gawelczyk, and N. Hansen. Step-size adaptation based on non-local use of selection information. In Y. Davidor et al., editors, *Parallel Problem Solving from Nature — PPSN III*, pages 189–198. Springer Verlag, 1994.

[14] A. I. Oyman, K. Deb, and H.-G. Beyer. An alternative constraint handling method for evolution strategies. In *Proc. of the 1999 IEEE Congress on Evolutionary Computation*, pages 612–619. IEEE Press, 1999.

[15] I. Rechenberg. *Evolutionsstrategie — Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Friedrich Frommann Verlag, 1973.

[16] T. P. Runarsson and X. Yao. Stochastic ranking for constrained evolutionary optimization. *IEEE Transactions on Evolutionary Computation*, 4(3):274–283, 2000.

[17] T. P. Runarsson and X. Yao. Search biases in constrained evolutionary optimization. *IEEE Transactions on Systems, Man and Cybernetics — Part C: Applications and Reviews*, 35(2):233–243, 2005.

[18] H.-P. Schwefel. *Numerical Optimization of Computer Models*. Wiley, 1981.

[19] H.-P. Schwefel. *Evolution and Optimum Seeking*. Wiley, 1995.