On the Behaviour of the $(1, \lambda)$-σSA-ES for a Constrained Linear Problem

Dirk V. Arnold

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

Abstract. This paper analyses the behaviour of the $(1, \lambda)$-σSA-ES with deterministic two-point rule when applied to a linear problem with a single linear constraint. Equations that describe the single-step behaviour of the strategy are derived and then used to predict the strategy's multi-step behaviour. The findings suggest that mutative self-adaptation will result in convergence of the $(1, \lambda)$-ES to non-stationary points if the angle between the gradient vector of the objective function and the normal vector of the constraint plane is small. Comparisons with the behaviour of evolution strategies that employ other step size adaptation mechanisms are drawn.

1 Introduction

Step size adaptation mechanisms and constraint handling techniques are important components of evolutionary algorithms (EAs) for constrained real valued optimisation. Most step size adaptation mechanisms have been devised with unconstrained optimisation in mind. Conversely, constraint handling techniques are often designed without much thought to their impact on step size adaptation. Schwefel [16] as early as the 1970s showed that a commonly employed step size adaptation mechanism may result in convergence to non-stationary points in an environment as simple as a linear problem with a single linear constraint.

An understanding of the interaction between step size adaptation mechanisms and constraint handling techniques is crucial for the design of EAs for constrained real valued optimisation. The *Handbook of Evolutionary Computation* [5, page B2.4:**11**f] lists a small number of studies that consider the behaviour of evolution strategies applied to simple constrained problems. Rechenberg [14] studies the performance of the $(1+1)$-ES[1] for the axis-aligned corridor model. Schwefel [16] considers the performance of the $(1, \lambda)$-ES in the same environment. Beyer [6] analyses the performance of the $(1 + 1)$-ES for a constrained, discus-like function. All of those have in common that the constraint planes are oriented such that their normal vectors are perpendicular to the gradient vector of the objective function. In contrast, Schwefel's work [16] suggests that convergence to

[1] See [9] for an explanation of the $(\mu/\rho \overset{+}{,} \lambda)$ terminology.

C.A. Coello Coello et al. (Eds.): PPSN 2012, Part I, LNCS 7491, pp. 82–91, 2012.

non-stationary points may occur in situations where the angle between those vectors, which we refer to as the constraint angle, is small. Studying the behaviour of EAs applied to a linear problem with a linear constraint of general orientation is fundamental as owing to Taylor's theorem, any smooth problem will appear increasingly linear as the step size of the strategy decreases. Arnold and Brauer [3] derive analytical results for the $(1+1)$-ES with success probability based step size adaptation and provide a quantitative confirmation of Schwefel's findings. More recent work [2, 1] analyses the behaviour of the $(1, \lambda)$-ES with cumulative step size adaptation for the constrained linear problem and compares two constraint handling techniques. It is found that convergence to non-stationary points in the face of small constraint angles is not unique to success probability based step size adaptation mechanisms.

The goal of this paper is to study the behaviour of the $(1, \lambda)$-σSA-ES, i.e., the $(1, \lambda)$-ES that employs mutative self-adaptation [16, 13] for step size control, when applied to a linear problem with a single linear constraint of general orientation. We assume that constraints are handled by resampling infeasible offspring candidate solutions. The work complements prior research that analyses the behaviour of mutative self-adaptation in unconstrained settings, including that by Hansen [10] who considers unconstrained linear problems, Beyer [7, 8] who considers spherically symmetric functions, and Meyer-Nieberg and Beyer [12] and Arnold and MacLeod [4] who consider ridge functions.

The remainder of this paper is organised as follows. Section 2 briefly describes the problem and the evolution strategy considered. Section 3 derives equations describing the single-step behaviour of the strategy. Section 4 considers multiple time steps and employs the balance criterion proposed by Lunacek and Whitley [11] in order to predict whether the strategy converges to a non-stationary point of the objective function. Section 5 concludes with a brief discussion of the findings and contrasts them with corresponding results for other step size adaptation mechanisms.

2 Problem and Algorithm

As in [3, 2, 1], 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 referred to as the constraint angle and denoted by θ as illustrated in Fig. 1. Constraint angles of interest are in the open interval $(0, \pi/2)$. The unit normal vector of the constraint plane expressed in the chosen

[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.

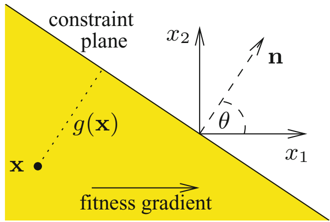

Fig. 1. Linear objective function with a single linear constraint. The subspace spanned by the x_1- and x_2-axes is shown. The shaded area is the feasible region. The parental candidate solution $\mathbf{x}$ of the $(1, \lambda)$-ES is at a distance $g(\mathbf{x})$ from the constraint plane.

coordinate system 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 is thus $g(\mathbf{x}) = -\mathbf{n}\cdot\mathbf{x} = -x_1\cos\theta - x_2\sin\theta$, resulting in the optimisation problem

$$\begin{aligned} &\text{maximise} \quad f(\mathbf{x}) = x_1 \\ &\text{subject to} \ \ g(\mathbf{x}) = -x_1\cos\theta - x_2\sin\theta \geq 0 \ . \end{aligned}$$

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 $\mathbf{x} \in \mathbb{R}^n$ and initial step size parameter $\sigma > 0$, the $(1, \lambda)$-σSA-ES generates a sequence of further candidate solutions by iterating the following three steps [16]:

1. Generate λ feasible offspring candidate solutions $\mathbf{y}^{(i)} = \mathbf{x} + \sigma\mathbf{z}^{(i)}$, $i = 1, \ldots, \lambda$, where the $\mathbf{z}^{(i)} \in \mathbb{R}^n$ are vectors with components drawn independently from normal distributions with mean zero and offspring dependent standard deviation ξ_i.
2. Evaluate $f(\mathbf{x}^{(i)})$ for $i = 1\ldots, \lambda$ and let $(1;\lambda)$ denote index of the offspring candidate solution with the largest objective function value.
3. Replace the parental candidate solution and update the step size parameter according to

$$\begin{aligned} \mathbf{x} &\leftarrow \mathbf{y}^{(1;\lambda)} \\ \sigma &\leftarrow \sigma\xi_{1;\lambda} \ . \end{aligned}$$

Vectors $\mathbf{z}^{(i)}$ are referred to as mutation vectors, step size parameter σ is referred to as the mutation strength, and the ξ_i are referred to as step size modifiers. Notice that Step 1 may require generating more than λ offspring as infeasible candidate solutions are rejected immediately. However, for the problem under consideration on average no more than 2λ offspring need to be sampled per iteration.

The expected length of mutation vector $\mathbf{z}^{(i)}$ is proportional to step size modifier ξ_i. The underlying proposition of mutative self-adaptation is that if offspring candidate solutions are generated with differing expected lengths of their mutation vectors, then selection of appropriate step sizes becomes a by-product of evolution. Common choices for the distribution of the ξ_i include [15]:

log-normal: $\xi_i = \exp(\tau\mathcal{N}(0,1))$ where $\mathcal{N}(0,1)$ denotes a standard normally distributed random variate sampled anew for each i

two-point: $\xi_i = \beta > 1$ with probability one half and $\xi_i = 1/\beta$ otherwise

deterministic two-point: $\xi_i = \beta > 1$ if $1 \leq i \leq \lambda/2$ and $\xi_i = 1/\beta$ otherwise.

Constants τ (for log-normal) and β (for two-point and deterministic two-point) need to be chosen large enough to result in meaningful differences between the distributions of the offspring they control while being small enough not to render step size control excessively noisy. Rechenberg [15] recommends $\beta = 1.3$ for the two-point rule. As shown in the context of spherically symmetric functions, the log-normal and two-point operators can be made to behave very similarly if the parameters τ and β are chosen appropriately [8]. For simplicity, in this paper only the deterministic two-point operator is considered.

If the $(1,\lambda)$-ES is run on the constrained linear problem described above and the mutation strength σ is held constant, then the distance of the parental candidate solution from the constraint plane will assume a time-invariant distribution. Larger mutation strengths will result in faster progress. If the mutation strength is not fixed but instead allowed to vary under the control of some step size adaptation mechanism, then step sizes will either increase or decrease indefinitely. Decreasing step sizes result in convergence to a non-stationary point; increasing step sizes result in continually accelerating progress and are thus desirable.

3 Single-Step Behaviour

Let $\delta = g(\mathbf{x})/\sigma$ denote the normalised distance of the parental candidate solution $\mathbf{x}$ from the constraint plane. As infeasible offspring are resampled, the probability distribution of the z_1- and z_2-components of mutation vectors of feasible offspring candidate solutions generated with step size modifier ξ is a truncated normal distribution with joint density

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

where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution. The normalising term $\Phi(\delta/\xi)$ equals the probability that a randomly generated offspring candidate solution is feasible. The marginal density of the z_1-component is

$$\begin{aligned} p_1(x \mid \xi) &= \int_{-\infty}^{\infty} p_{1,2}(x, y \mid \xi)\,\mathrm{d}y \\ &= \frac{1}{\sqrt{2\pi}\xi\Phi(\delta/\xi)}\mathrm{e}^{-\frac{1}{2}(x/\xi)^2}\Phi\left(\frac{\delta - x\cos\theta}{\xi\sin\theta}\right) \ . \end{aligned} \tag{2}$$

We write $P_1(x \mid \xi, \delta)$ for the corresponding cumulative distribution function. The density of the z_2-component conditional on the value of the z_1-component is

$$p_2(y \mid z_1 = x, \xi) = \frac{p_{1,2}(x, y \mid \xi)}{p_1(x \mid \xi)} \ . \tag{3}$$

Integration of the probability density yields

$$P_2(y \mid z_1 = x, \xi) = \begin{cases} \dfrac{\Phi(y/\xi)}{\Phi((\delta - x\cos\theta)/(\xi\sin\theta))} & \text{if } y < \dfrac{\delta - x\cos\theta}{\sin\theta} \\ 1 & \text{otherwise} \end{cases} \tag{4}$$

for the conditional cumulative distribution function of the z_2-component.

An important quantity to consider is the probability P_+ that the offspring candidate solution that is selected to replace the parent is one generated with step size modifier $\xi = \beta$ (as opposed to $\xi = 1/\beta$). That probability is of course also the probability that the step size of the strategy increases in the present step. As selection is based purely on the z_1-components of the mutation vectors, the cumulative distribution function of the z_1-component of the best of the $\lambda/2$ offspring candidate solutions generated with step size modifier ξ is

$$Q_\xi(x) = P_1^{\lambda/2}(x \mid \xi) \ .$$

The corresponding probability density function is

$$q_\xi(x) = \frac{\mathrm{d}}{\mathrm{d}x} Q_\xi(x) = \frac{\lambda}{2} p_1(x \mid \xi) P_1^{\lambda/2-1}(x \mid \xi) \ .$$

Probability P_+ is obtained by integrating the probability that the best offspring candidate solution generated with step size modifier β is superior to the best one generated with step size modifier $1/\beta$ and thus equals

$$P_+ = \int_{-\infty}^{\infty} q_\beta(x) Q_{1/\beta}(x)\, \mathrm{d}x \ . \tag{5}$$

Figure 2 illustrates how this probability depends on the normalised parental distance from the constraint plane and on the magnitude of the step size modifier. The plots have been generated from Eq. (5) and are for $\beta = 1.3$ in the left hand graph and for $\theta = \pi/8$ in the right hand one. If the parental candidate solution is far from the constraint plane, then the probability of generating an infeasible candidate solution that needs to be resampled is small and P_+ is independent of the constraint angle and exceeds one half. With decreasing distance from the constraint plane, P_+ decreases. Depending on the values of λ and θ it may either decrease below one half or remain above. Larger values of λ generally result in larger values of P_+. The curves in the right hand graph are monotonic and start at a value of one half for $\beta = 1$, suggesting that the choice of the step size modifier does not impact whether P_+ exceeds one half or not.

4 Multi-Step Behaviour

The results derived up to this point depend on the normalised distance δ of the parental candidate solution from the constraint plane. Assuming for now that

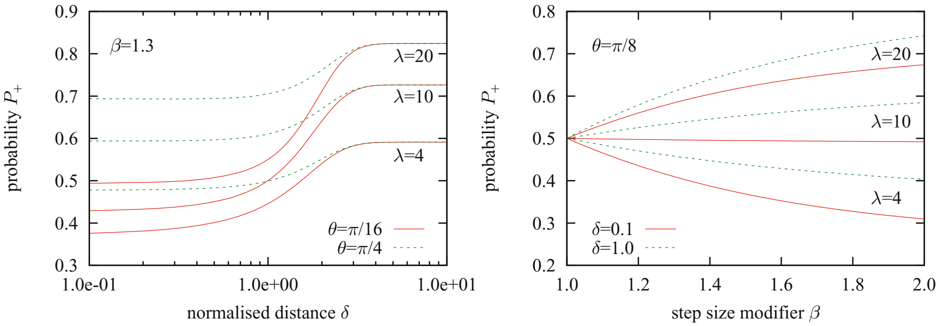

Fig. 2. Probability P_+ that a candidate solution generated with mutation strength modifier $\xi = \beta$ is selected as the next parental candidate solution plotted against the normalised parental distance δ from the constraint plane and against the magnitude of the step size modifier β

the mutation strength is fixed, if the strategy is iterated the normalised distance from the constraint plane evolves according to

$$\delta^{(t+1)} = \delta^{(t)} - z_1^{(1;\lambda)} \cos\theta - z_2^{(1;\lambda)} \sin\theta \tag{6}$$

where superscripts on δ denote time and those on z_1 and z_2 indicate the offspring candidate solution selected to replace the parent. The cumulative distribution function of $\delta^{(t+1)}$ conditional on $\delta^{(t)} = \delta$ is obtained using Eq. (6) by integrating the probability that $\delta^{(t+1)} < y$, yielding

$$\begin{aligned}
P_\delta^{(t+1)}(y \mid \delta^{(t)} = \delta) &= \mathrm{Prob}\left[\delta^{(t+1)} < y \,\middle|\, \delta^{(t)} = \delta\right] \\
&= \int_{-\infty}^{\infty} q_\beta(x) Q_{1/\beta}(x) \left[1 - P_2\left(\frac{\delta - y - x\cos\theta}{\sin\theta} \,\middle|\, z_1 = x, \xi = \beta\right)\right] \mathrm{d}x \\
&\quad + \int_{-\infty}^{\infty} q_{1/\beta}(x) Q_\beta(x) \left[1 - P_2\left(\frac{\delta - y - x\cos\theta}{\sin\theta} \,\middle|\, z_1 = x, \xi = 1/\beta\right)\right] \mathrm{d}x
\end{aligned}$$

where conditional probability $P_2(\cdot|\cdot)$ is given in Eq. (4). Computing the derivative with respect to y yields conditional probability density $p_\delta^{(t+1)}(y \mid \delta^{(t)} = \delta)$.

For fixed mutation strength σ, the distance δ of the parental candidate solution from the constraint plane assumes a time-invariant limit distribution the density of which satisfies the evolution equation

$$p_\delta(y) = \int_0^\infty p_\delta(x) p_\delta(y|x) \, \mathrm{d}x \tag{7}$$

where the conditional density is that derived above. An approximation to the stationary limit distribution can be derived using the approach pursued by Beyer [8] in a different context: Expand the unknown distribution at time step t into a Gram-Charlier series with unknown cumulants. Then determine cumulants at time step $t+1$ using Eq. (7). Considering cumulants up to the kth and imposing equality constraints on the cumulants yields a system of k equations in the k unknown cumulants. In the simplest case, only a single cumulant (the mean) is considered, and the equality constraint is

$$\mathrm{E}\left[\delta^{(t+1)} \,\middle|\, \delta^{(t)} = \delta\right] = \delta \tag{8}$$

which can be solved for the approximate average distance δ of the parental candidate solution from the constraint plane.

The expected distance from the constraint plane after a time step conditional on that distance before the time step is

$$\begin{aligned}
\mathrm{E}\left[\delta^{(t+1)} \,\middle|\, \delta^{(t)} = \delta\right] &= \int_0^\infty y p_\delta^{(t+1)}(y \mid \delta^{(t)} = \delta)\, \mathrm{d}y \\
&= \frac{1}{\sin\theta} \int_{-\infty}^\infty q_\beta(x) Q_{1/\beta}(x) \int_0^\infty y p_2 \left(\frac{\delta - y - x\cos\theta}{\sin\theta} \,\middle|\, z_1 = x, \xi = \beta \right) \mathrm{d}y\, \mathrm{d}x \\
&\quad + \frac{1}{\sin\theta} \int_{-\infty}^\infty q_{1/\beta}(x) Q_\beta(x) \int_0^\infty y p_2 \left(\frac{\delta - y - x\cos\theta}{\sin\theta} \,\middle|\, z_1 = x, \xi = 1/\beta \right) \mathrm{d}y\, \mathrm{d}x
\end{aligned}$$

where the conditional density $p_2(\cdot|\cdot)$ is given in Eq. (3). Solving the inner integrals yields expression

$$\begin{aligned}
\mathrm{E}\left[\delta^{(t+1)} \,\middle|\, \delta^{(t)} = \delta\right] = \int_{-\infty}^\infty & \gamma_\beta(x) q_\beta(x) Q_{1/\beta}(x)\, \mathrm{d}x \\
&+ \int_{-\infty}^\infty \gamma_{1/\beta}(x) q_{1/\beta}(x) Q_\beta(x)\, \mathrm{d}x
\end{aligned} \tag{9}$$

with

$$\gamma_\xi(x) = \delta - x\cos\theta + \frac{\xi\sin\theta}{\sqrt{2\pi}} \frac{\exp(-((\delta - x\cos\theta)/(\xi\sin\theta))^2/2)}{\Phi((\delta - x\cos\theta)/(\xi\sin\theta))}$$

for the expected distance from the constraint plane after a time step conditional on that distance before the time step.

Figure 3 illustrates how the average normalised distance from the constraint plane and the probability P_+ that an offspring candidate solution generated with the larger step size modifier replaces the parent depend on the constraint angle. The curves have been obtained by using Eq. (9) in Eq. (8) for $\beta = 1.3$, solving for δ, and using the result in Eq. (5). The dots mark measurements made in runs of the $(1, \lambda)$-ES with fixed step size. The average distance at which the constraint plane is tracked decreases with decreasing constraint angle and with increasing λ. The probability that an offspring candidate solution generated with step size modifier $\xi = \beta$ is selected to replace the parental candidate solution decreases with decreasing constraint angle and with decreasing λ. It exceeds one half for large constraint angles, but is below one half for small ones. The accuracy of the predictions made based on the simple stationarity requirement that considers the mean of the distribution only appears visually good for small constraint angles.

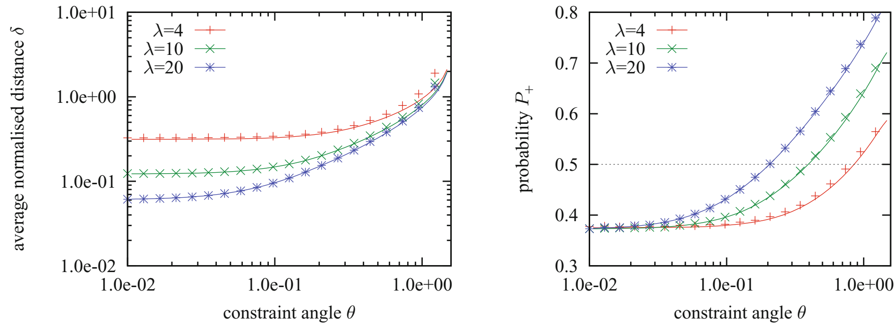

Fig. 3. Average normalised distance δ of the parental candidate solution from the constraint plane and probability P_+ of a candidate solution generated with step size modifier $\xi = \beta$ being selected to replace the parent plotted against constraint angle θ

If the mutation strength is not fixed but instead under the control of mutative self-adaptation, then, depending on the number of offspring λ generated per time step and the constraint angle θ, the strategy will either systematically reduce its step size and converge to a non-stationary point, or it will increase the step size and diverge. Clearly, for the constrained linear problem, which does not have a finite optimum, the latter is desirable. In order to establish whether convergence or divergence occurs, we employ the simple balance criterion proposed by Lunacek and Whitley [11] in the context of ridge functions. Specifically, we consider the probability P_+ that the mutation strength increases in the strategy's stationary state for fixed step size. If that probability exceeds one half, then divergence will occur; if it is below one half, then the strategy will converge.

Figure 4 illustrates how the minimum number of offspring required to avoid convergence to a non-stationary point depends on the constraint angle. The solid lines in both plots have been obtained by using Eq. (8) with $\beta = 1.3$ to determine the stationary δ, using Eq. (5) to obtain the corresponding P_+, and then determining the smallest λ such that P_+ exceeds one half. The points in the left hand graph mark measurements made in runs of the $(1, \lambda)$-σSA-ES. For each combination of λ and θ values considered, 100 runs of the strategy (initialised with the parental candidate solution on the constraint plane and $\sigma = 1$) were conducted until the mutation strength reached a value of either 10^{-20} (which is taken to be indicative of convergence) or 10^{20} (which is taken to be indicative of divergence). If at least 90 of the 100 runs yielded the same result, the location was marked with $\times$ (indicating convergence) or $+$ (indicating divergence). The quality of the predictions made on the basis of the simple stationarity and balance criteria is excellent.

The right hand graph in Fig. 4 contrasts results for the $(1, \lambda)$-σSA-ES with corresponding results for the $(1 + 1)$-ES with $1/5$th-success rule [3] and the $(1, \lambda)$-ES with cumulative step size adaptation [2]. The latter curves correspond to, from top to bottom, values of the cumulation parameter of $c = 0.5$, 0.05, and 0.005. In contrast to the $(1+1)$-ES, the $(1, \lambda)$-σSA-ES is capable of avoiding

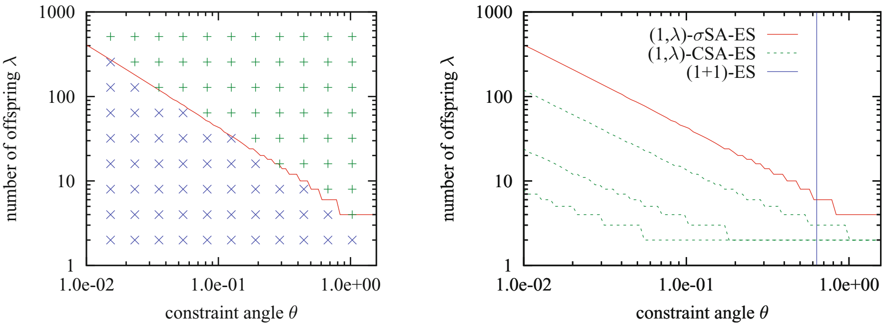

Fig. 4. Number of offspring λ per time step required to avoid convergence plotted against the constraint angle θ. The identical solid curve in both plots represents results for the $(1, \lambda)$-σSA-ES. The dashed lines in the right hand graph represent results for the $(1, \lambda)$-ES with cumulative step size adaptation and several values of the cumulation parameter. The vertical line in the right hand graph marks the constraint angle below which the $(1 + 1)$-ES converges to a non-stationary point.

convergence for any value of θ. However, the number of offspring per time step that is needed becomes very large as constraint angles become increasingly acute. (The plots suggest that the value of λ required is inversely proportional to θ.) In comparison, the $(1, \lambda)$-ES with cumulative step size adaptation manages to avoid convergence using significantly smaller values of λ.

5 Discussion and Future Work

To conclude, we have analysed the behaviour of the $(1, \lambda)$-σSA-ES for a linear problem with a single linear constraint of general orientation. A simple stationarity requirement has been used to approximate the average distance of the strategy from the constraint plane if the step size is fixed. The balance condition proposed by Lunacek and Whitley [11] has then been used to establish whether the adaptive strategy will converge to a non-stationary point or diverge. It has been found that divergence, which is the desirable behaviour, for increasingly acute constraint angles requires increasingly larger numbers of offspring generated per time step. Compared to the $(1, \lambda)$-ES with cumulative step size adaptation, for a given value of the constraint angle, the number of offspring required by the $(1, \lambda)$-σSA-ES is much higher.

There are multiple opportunities for further improving the understanding of the interaction of step size adaptation mechanisms and constraint handling techniques using the approach pursued here. Obvious extensions include considering the log-normal and two-point rules in place of the deterministic two-point rule, but differences are likely to be quantitative rather than qualitative. Regarding the $(\mu/\mu, \lambda)$-σSA-ES, it may be expected that the bias toward larger step sizes that results from the arithmetic averaging of mutation strengths has a beneficial impact on the performance of the strategy, but the magnitude of that effect remains to be seen. Further future work includes the consideration of other constraint handling approaches, such as the simple repair mechanism previously considered for the $(1, \lambda)$-ES with cumulative step size adaptation [1], and of further constrained test problems.

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

References

[1] Arnold, D.V.: Analysis of a repair mechanism for the $(1, \lambda)$-ES applied to a simple constrained problem. In: Genetic and Evolutionary Computation Conference — GECCO 2011, pp. 853–860. ACM Press (2011)
[2] Arnold, D.V.: On the behaviour of the $(1, \lambda)$-ES for a simple constrained problem. In: Foundations of Genetic Algorithms 11, pp. 15–24. ACM Press (2011)
[3] Arnold, D.V., Brauer, D.: On the Behaviour of the (1+1)-ES for a Simple Constrained Problem. In: Rudolph, G., Jansen, T., Lucas, S., Poloni, C., Beume, N. (eds.) PPSN X. LNCS, vol. 5199, pp. 1–10. Springer, Heidelberg (2008)
[4] Arnold, D.V., MacLeod, A.: Step length adaptation on ridge functions. Evolutionary Computation 16(2), 151–184 (2008)
[5] Bäck, T., Fogel, D.B., Michalewicz, Z.: Handbook of Evolutionary Computation. Oxford University Press (1997)
[6] Beyer, H.-G.: Ein Evolutionsverfahren zur mathematischen Modellierung stationärer Zustände in dynamischen Systemen. PhD thesis, Hochschule für Architektur und Bauwesen, Weimar (1989)
[7] Beyer, H.-G.: Toward a theory of evolution strategies: Self-adaptation. Evolutionary Computation 3(3), 311–347 (1996)
[8] Beyer, H.-G.: The Theory of Evolution Strategies. Springer (2001)
[9] Beyer, H.-G., Schwefel, H.-P.: Evolution strategies — A comprehensive introduction. Natural Computing 1(1), 3–52 (2002)
[10] Hansen, N.: An analysis of mutative σ-self-adaptation on linear fitness functions. Evolutionary Computation 14(3), 255–275 (2006)
[11] Lunacek, M., Whitley, L.D.: Searching for Balance: Understanding Self-adaptation on Ridge Functions. In: Runarsson, T.P., Beyer, H.-G., Burke, E.K., Merelo-Guervós, J.J., Whitley, L.D., Yao, X. (eds.) PPSN IX. LNCS, vol. 4193, pp. 82–91. Springer, Heidelberg (2006)
[12] Meyer-Nieberg, S., Beyer, H.-G.: Mutative Self-adaptation on the Sharp and Parabolic Ridge. In: Stephens, C.R., Toussaint, M., Whitley, L.D., Stadler, P.F. (eds.) FOGA 2007. LNCS, vol. 4436, pp. 70–96. Springer, Heidelberg (2007)
[13] Meyer-Nieberg, S., Beyer, H.-G.: Self-adaptation in Evolutionary Algorithms. In: Lobo, F.G., Lima, C.F., Michalewicz, Z. (eds.) Parameter Setting in Evolutionary Algorithms. SCI, vol. 54, pp. 47–74. Springer, Heidelberg (2007)
[14] Rechenberg, I.: Evolutionsstrategie — Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Friedrich Frommann Verlag (1973)
[15] Rechenberg, I.: Evolutionsstrategie 94. Friedrich Frommann Verlag (1994)
[16] Schwefel, H.-P.: Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie. Birkhäuser Verlag (1977)