On the Use of Evolution Strategies for Optimization on Spherical Manifolds

Dirk V. Arnold

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

Abstract. We study the behaviour of evolution strategies applied to a simple class of unimodal optimization problems on spherical manifolds. The techniques used are the same as those commonly employed for the analysis of the behaviour of evolution strategies in Euclidean search spaces. However, we find that there are significant differences in strategy behaviour unless the vicinity of an optimal solution has been reached. Experiments with cumulative step size adaptation reveal the existence of metastable states associated with large step sizes, which can preclude reaching optimal solutions.

1 Introduction

The vast majority of work on real-valued evolutionary optimization is concerned with Euclidean search spaces. However, there are important applications where the search domain is not Euclidean but a more general Riemannian manifold instead. See [6] for an introduction to Riemannian geometry. Qi et al. [10] give two broad classes of applications for optimization on manifolds: "equality-constrained optimization problems where the constraints specify a submanifold of $\mathbb{R}^N$; and problems where the objective function has continuous invariance properties that we want to eliminate for various reasons". Our own interest in optimization on manifolds is motivated by the need to optimize quaternion variables, which commonly arises in 3D registration tasks where quaternion variables are used to encode orientation.

Several applications of evolutionary algorithms to optimization on Riemannian manifolds other than Euclidean spaces can be found in the literature. An early instance is work by Kissinger et al. [9], who propose a variant of evolutionary programming for optimization involving quaternion variables. Arguably the most sophisticated evolutionary approach to optimization on general Riemannian manifolds is that by Colutto et al. [5], who propose a variant of covariance matrix adaptation evolution strategies (CMA-ES) [7] for optimization on Riemannian manifolds. Their algorithm uses parallel transport, a tool for transporting geometrical data along smooth curves, to transform search paths and covariance matrices between iterations. They find that their approach more effectively solves a two-dimensional multimodal optimization problem than restart

T. Bartz-Beielstein et al. (Eds.): PPSN XIII 2014, LNCS 8672, pp. 882–891, 2014.

variants of gradient based and Newton-Armijo methods proposed by Yang [12] for optimization on manifolds. However, they also observe that their strategy can (infrequently) be observed to fail to locate the optimal solution to a simple unimodal optimization problem on spherical manifolds.

The aim of this paper is to develop an analytically based understanding of the behaviour of a simplified variant of the algorithm of Colutto et al. [5] that adapts its global step size using cumulative step size adaptation, but does not adapt the full covariance matrix. We derive results characterizing the behaviour of the algorithm when applied to a class of unimodal optimization problems on spherical manifolds, both providing an explanation for the observed inability to converge to the optimal solution in some instances and suggesting how to avoid this situation.

2 Problem and Algorithm

Let M be a Riemannian manifold with tangent space T_pM at point $p \in M$. We consider the case that $M = S^{N-1}$ (i.e., the unit $(N-1)$-sphere defined by $S^{N-1} = \{\mathbf{x} \in \mathbb{R}^N \mid \|\mathbf{x}\| = 1\}$). Note that the universal cover of the 3D rotation group $SO(3)$ is diffeomorphic to S^3, and that problems on spherical manifolds thus naturally arise in combination with the optimization of orientations in 3D. We consider the class of optimization problems $f : S^{N-1} \to \mathbb{R}$ that possess a unique optimal solution and where objective function values depend only on the distance from that solution and vary strictly monotonically with it. By choosing a coordinate system such that the optimal solution is located at $\mathbf{x} = (1, 0, \ldots, 0)^{\mathrm{T}}$ and taking into account that evolution strategies perform selection based only on comparisons of objective function values, we can without loss of generality consider objective function

$$f(\mathbf{x}) = x_1 \tag{1}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_N)^{\mathrm{T}} \in S^{N-1}$ and the task is maximization. Effectively the same problem has been used by Colutto et al. [5] in the experimental evaluation of their algorithm. Arguably, Eq. (1) constitutes the analogue of the "sphere model" introduced by Rechenberg [11] for the study of the behaviour of evolution strategies in Euclidean spaces.

The CMA-ES for optimization on Riemannian manifolds proposed by Colutto et al. [5] generates mutation vectors in the tangent space at the current population centroid, and it uses the Riemannian exponential map to map them onto the manifold. Parallel transport is used as a means for mapping search paths and the covariance matrix of the mutation distribution from the tangent space at the current population centroid to that at the next. The algorithm considered here is in essence the same, with the single major difference that in order to admit an analytically based investigation, rather than adapting the entire covariance matrix, only the global step size is adapted. A single iteration of the algorithm is given in Fig. 1. The population size parameters μ and λ are positive integers with $\mu < \lambda$. Cumulation parameter $c \in (0, 1]$ and damping parameter $D \in \mathbb{R}^+$

Input: population centroid $\mathbf{x} \in M$, mutation strength $\sigma \in \mathbb{R}^+$, search path $\mathbf{s} \in T_{\mathbf{x}}M$

1. Generate offspring candidate solutions

$$\mathbf{y}^{(i)} = \exp_{\mathbf{x}}\left(\sigma \mathbf{z}^{(i)}\right) \qquad\qquad i = 1, \ldots, \lambda$$

   where the mutation vectors $\mathbf{z}^{(i)}$ are standard normally distributed in $T_{\mathbf{x}}M$ and $\exp_{\mathbf{x}}(\cdot)$ denotes the Riemannian exponential map.

2. Compute $f(\mathbf{y}^{(i)})$ for $i = 1, \ldots, \lambda$. Let $(k; \lambda)$ denote the index of the offspring candidate solution with the kth largest objective function value and

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

3. Update the search path according to

$$\mathbf{s} \leftarrow (1 - c)\mathbf{s} + \sqrt{\mu c(2 - c)}\mathbf{z}^{(\text{avg})}.$$

4. Update the population centroid according to

$$\mathbf{x} \leftarrow \exp_{\mathbf{x}}\left(\sigma \mathbf{z}^{(\text{avg})}\right)$$

   and use parallel transport to transform the search path from the old population centroid to the new one.

5. Update the mutation strength according to

$$\sigma \leftarrow \sigma \exp\left(\frac{\|\mathbf{s}\|^2 - N}{2DN}\right).$$

Fig. 1. Single iteration of the strategy for optimization on Riemannian manifolds

are constants. After initialization (to be discussed below), the algorithm in Fig. 1 is iterated until a stopping condition is met.

Regarding the implementation of the algorithm for the case that $M = S^{N-1}$, sampling standard normally distributed mutation vectors in $T_{\mathbf{x}}S^{N-1}$ can be accomplished by sampling standard normally distributed mutation vectors $\mathbf{w}$ in $\mathbb{R}^N$ and projecting them onto the tangent space $T_{\mathbf{x}}S^{N-1}$ according to

$$\mathbf{z} = \mathbf{w} - \langle \mathbf{x}, \mathbf{w} \rangle \mathbf{x} \tag{2}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product. The Riemannian exponential map $\exp_{\mathbf{x}}(\cdot) : T_{\mathbf{x}}M \to M$ for the case that $M = S^{N-1}$ is described by

$$\exp_{\mathbf{x}}(\sigma \mathbf{z}) = \mathbf{x} \cos(\sigma \|\mathbf{z}\|) + \mathbf{z} \, \frac{\sin(\sigma \|\mathbf{z}\|)}{\|\mathbf{z}\|} \, . \tag{3}$$

Finally, Huckemann et al. [8] show that parallel transport on spherical manifolds is accomplished for $\mathbf{x} \neq \pm\mathbf{x}'$ by

$$\mathbf{w}' = \mathbf{w} - \langle \mathbf{w}, \mathbf{v} \rangle \left[(1 - \langle \mathbf{x}, \mathbf{x}' \rangle) \, \mathbf{v} + \langle \mathbf{v}, \mathbf{x}' \rangle \mathbf{x} \right] \tag{4}$$

where

$$\mathbf{v} = \frac{\mathbf{x}' - \langle \mathbf{x}, \mathbf{x}' \rangle \mathbf{x}}{\| \mathbf{x}' - \langle \mathbf{x}, \mathbf{x}' \rangle \mathbf{x} \|} \tag{5}$$

and $\mathbf{w}'$ is the parallel transplant of $\mathbf{w} \in T_{\mathbf{x}} S^{N-1}$ to $T_{\mathbf{x}'} S^{N-1}$.

3 Analysis

In order to analyze the behaviour of the algorithm thus described when applied to the class of problems defined by Eq. (1) we first consider single iterations and determine the expected step. In analogy to related work in Euclidean spaces [11, 2] we then obtain simpler expressions by making the assumption that the steps that the strategy takes are small. It will be seen that that assumption is a valid one to make only if the strategy has reached the vicinity of the optimal solution. Finally, we consider the multi-iteration behaviour of the algorithm.

3.1 Large-Step Behaviour

Considering a single iteration of the algorithm described in Fig. 1, by choosing the coordinate system appropriately we can without loss of generality assume that $\mathbf{x} = (x_1, \sqrt{1 - x_1^2}, 0, \ldots, 0)^{\mathrm{T}}$. From Eq. (1) with Eq. (3), the objective function value of offspring candidate solution $\mathbf{y} = \exp_{\mathbf{x}}(\sigma \mathbf{z})$ is

$$f(\mathbf{y}) = x_1 \cos(\sigma \|\mathbf{z}\|) + z_1 \frac{\sin(\sigma \|\mathbf{z}\|)}{\|\mathbf{z}\|} \ . \tag{6}$$

The lengths $\|\mathbf{z}\|$ of mutation vectors are χ_{N-1}-distributed with mean $l_{\mathbf{z}} = \sqrt{2}\Gamma(N/2)/\Gamma((N-1)/2)$, which for large N is well approximated by $\sqrt{N}$, and a coefficient of variation that goes to zero as N increases. The impact of variations in $\|\mathbf{z}\|$ on offspring objective function values thus decreases with increasing dimension N. The ordering of the offspring by objective function values in Step 2 of the algorithm in Fig. 1 will thus increasingly be an ordering by values of z_1. From Eq. (2), $z_1 = (1 - x_1^2) w_1 - \sqrt{1 - x_1^2} x_1 w_2$, where w_1 and w_2 are independently standard normally distributed. Mutation vector component z_1 is thus normally distributed with mean zero and variance $1 - x_1^2$. The selected z_1-components are the μ largest order statistics of a sample of normally distributed random variates, and their expected average according to [2] equals

$$\mathrm{E}\left[z_1^{(\mathrm{avg})} \right] = \mathrm{sgn}\left(\sin(\sigma \|\mathbf{z}\|) \right) \sqrt{1 - x_1^2} c_{\mu/\mu,\lambda} \tag{7}$$

where $c_{\mu/\mu,\lambda}$ denotes the progress coefficient. Defining the progress rate

$$\varphi = \mathrm{E}\left[f\left(\exp_{\mathbf{x}}\left(\sigma \mathbf{z}^{(\mathrm{avg})} \right) \right) \right] - f(\mathbf{x}) \tag{8}$$

as the expected improvement in objective function value in a single iteration, it follows from Eq. (7) that for large N

$$\varphi = \operatorname{sgn}\left(\sin(\sigma\|\mathbf{z}\|)\right)\frac{\sin(\sigma\|\mathbf{z}^{(\mathrm{avg})}\|)}{\|\mathbf{z}^{(\mathrm{avg})}\|}\sqrt{1-x_1^2}c_{\mu/\mu,\lambda} - \left[1-\cos(\sigma\|\mathbf{z}^{(\mathrm{avg})}\|)\right]x_1 \quad (9)$$

approximately holds. As $\mathbf{z}^{(\mathrm{avg})}$ is the average of μ vectors $N-2$ of the $N-1$ components of which are random and uncorrelated, for large N the expected length of $\mathbf{z}^{(\mathrm{avg})}$ can be approximated by $l_{\mathbf{z}}/\sqrt{\mu}$ [2].

The left hand side of Fig. 2 compares predictions from Eq. (9) with $\|\mathbf{z}\|$ and $\|\mathbf{z}^{(\mathrm{avg})}\|$ replaced with their expected values with measurements made in one-iteration experiments of the algorithm for several values of N. Values of $x_1 \in \{-0.5, 0.2, 0.9\}$ have been chosen as representatives of situations where the search is at a great distance, an intermediate distance, and in relative proximity to the optimal solution and thus of different stages in the optimization process. Clearly, the accuracy of the predictions improves with increasing N and decreases with increasing σ. The primary reason for the inaccuracies are variations in the length of mutation vectors. In low dimensions and for large mutation strengths, selection is increasingly on the basis of the length of the mutation vectors. Nonetheless, it is clear from the figure that the dependence of the progress rate on the mutation strength differs markedly from that for the Euclidean sphere model, where the progress rate after an initial increase monotonically decreases to negative infinity [11, 2]. On the spherical manifold, multiple modes can be observed that result from the Riemannian exponential map tracing out geodesic paths. Additionally, the curves exhibit discontinuities that stem from the lengths of the average of the selected mutation vectors being reduced compared to those of the mutation vectors themselves. Mutation vectors of a length resulting in the sine function in Eq. (6) being positive may result in negative values of the sine function in the numerator in Eq. (9) and vice versa.

3.2 Small-Step Behaviour

All of the curves in the graphs on the left hand side of Fig. 1 have in common that for small mutation strengths they initially increase. More often than never, the first mode encountered yields optimal or near optimal performance. To investigate the behaviour of the strategy for small mutation strengths, we replace $\|\mathbf{z}\|$ and $\|\mathbf{z}^{(\mathrm{avg})}\|$ in Eq. (9) with $\sqrt{N}$ and $\sqrt{N/\mu}$, respectively, expand the sine and cosine functions into Taylor series at zero and abort after the linear and quadratic terms, respectively, and for $x_1 \neq \pm 1$ introduce normalized mutation strength $\sigma^* = N|x_1|\sigma/\sqrt{1-x_1^2}$ and normalized progress rate $\varphi^* = N|x_1|\varphi/(1-x_1^2)$, resulting in

$$\varphi^* = \sigma^* c_{\mu/\mu,\lambda} - \operatorname{sgn}(x_1)\frac{\sigma^{*2}}{2\mu} \ . \quad (10)$$

Notice that for $x_1 > 0$ Eq. (10) has the same form as the progress rate law for the Euclidean sphere model [2].

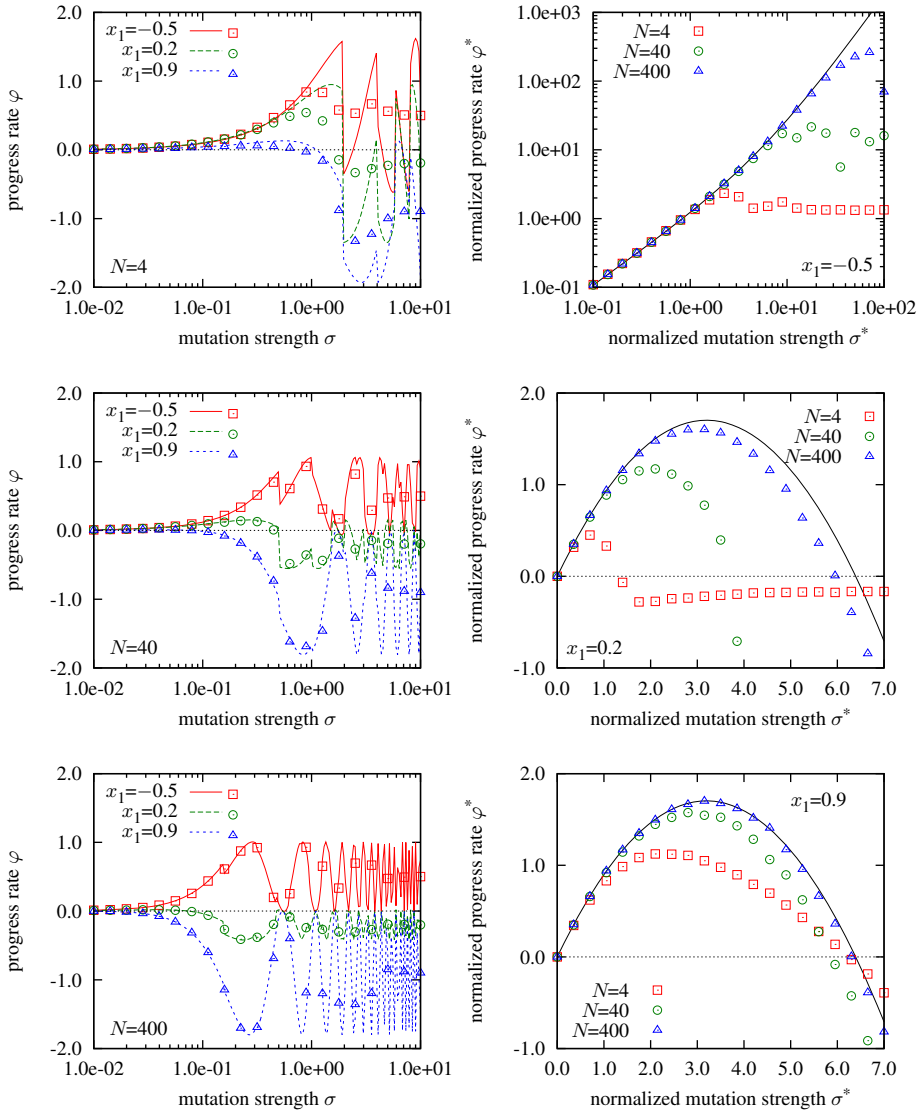

Fig. 2. Left: Progress rate φ plotted against mutation strength σ for $\mu = 3$ and $\lambda = 10$, parent locations $x_1 \in \{-0.5, 0.2, 0.9\}$, and, from top to bottom, search space dimensions $N = 4$, 40, and 400. Right: Normalized progress rate φ^* plotted against normalized mutation strength σ^* for search space dimensions $N \in \{4, 40, 400\}$ and, from top to bottom, parent locations $x_1 = -0.5$, 0.2, and 0.9. The lines represent predictions from Eqs. (9) and (10), respectively. The points mark values obtained by averaging over 20,000 one-iteration experiments for each data point shown.

The right hand side of Fig. 2 compares predictions from Eq. (10) with measurements made in one-iteration experiments of the algorithm. For $x_1 = -0.5$ Eq. (10) predicts that the normalized progress rate increases indefinitely with increasing normalized mutation strength. This is of course impossible for finite N, and the solid curve fails to correctly predict the experimental data if the mutation strength is too large for the truncated Taylor series to well represent the trigonometric functions. However, as increasing σ results in tracing out a geodesic path, the hemisphere with $x_1 > 0$ can always be reached in a single step. For $x_1 = 0.2$ the accuracy of the predictions increases significantly with increasing N. The truncated Taylor series become poor approximations to the trigonometric functions for $\sigma\|\mathbf{z}\| \approx \pi/2$ and thus for $\sigma^* \approx \pi|x_1|\sqrt{N}/(2\sqrt{1-x_1^2})$, and the maxima of the experimental data for $N = 4$ and $N = 40$ quite closely correspond to those values. Finally, for $x_1 = 0.9$ the qualitative behaviour of the algorithm is captured quite well for N as small as 4. If the mutation strength of the strategy is controlled properly, values of x_1 in excess of 0.9 can be reached in a relatively small number of iterations, and much of the computational effort will be incurred where the predictions from Eq. (10) are quite accurate.

3.3 Step Size Adaptation

To analyze the performance of cumulative step size adaptation on spherical manifolds, we employ the same approach as in Euclidean spaces [1]. The state of the strategy is determined by the population centroid $\mathbf{x}$, the mutation strength σ, and the search path $\mathbf{s}$. The parallel transport in Step 4 of the algorithm Fig. 1 uses vector $\mathbf{v}$, which can be computed from Eq. (5) as $\mathbf{z}^{(\mathrm{avg})}/\|\mathbf{z}^{(\mathrm{avg})}\|$. Using primes to indicate values of a quantity after an iteration of the algorithm, it follows that $\langle \mathbf{x}, \mathbf{x}' \rangle = \cos(\sigma\|\mathbf{z}^{(\mathrm{avg})}\|)$ and $\langle \mathbf{v}, \mathbf{x}' \rangle = \sin(\sigma\|\mathbf{z}^{(\mathrm{avg})}\|)$. Omitting terms that disappear in the limit $N \to \infty$ and using the small-step approximation from Sect. 3.2, the update of the search path in Steps 3 and 4 of the algorithm in Fig. 1 is thus described by

$$\mathbf{s}' = (1-c)\mathbf{s} + \sqrt{\mu c(2-c)}\left[\mathbf{z}^{(\mathrm{avg})}\cos(\sigma\|\mathbf{z}^{(\mathrm{avg})}\|) - \mathbf{x}\|\mathbf{z}^{(\mathrm{avg})}\|\sin(\sigma\|\mathbf{z}^{(\mathrm{avg})}\|)\right] \tag{11}$$

where it is assumed that $c = 1/\sqrt{N}$ and $D = 1/c$.[1]

Due to the symmetry inherent in the problem at hand, the location of the population centroid is adequately described by x_1, and the search path is characterized by its components s_1 and $s_\odot = \sum_{i=2}^N x_i s_i / \sqrt{\sum_{i=2}^N x_i^2}$, along with its squared length $\|\mathbf{s}\|^2$. Iterating the algorithm in Fig. 1 generates a Markov process in a five-dimensional state space with variables x_1, σ, s_1, $s_\odot$, and $\|\mathbf{s}\|^2$. We compute an approximation to the average values characterizing the search path

[1] Detailed calculations cannot be reproduced here due to space limitations, but can be found at `http://www.cs.dal.ca/~dirk/PPSN2014addendum.pdf` instead.

by requiring that an iteration of the algorithm results in no change in expectation. That is, we require that $\mathrm{E}[s'_1] = s_1$, $\mathrm{E}[s'_\odot] = s_\odot$, and $\mathrm{E}[\|\mathbf{s}'\|^2] = \|\mathbf{s}\|^2$. Dropping terms that disappear for large N and solving for $\|\mathbf{s}\|^2$ yields

$$\|\mathbf{s}\|^2 = N + 2\,\frac{\mu c_{\mu/\mu,\lambda}}{c}\left[c_{\mu/\mu,\lambda} - \operatorname{sgn}(x_1)\frac{\sigma^*}{\mu}\right] \tag{12}$$

for the squared length of the search path.

As in [1] we refer to the mutation strength for which no change in step size is expected as the target mutation strength of the strategy. For $x_1 > 0$, from the update of the mutation strength in Step 5 of the algorithm in Fig. 1 with Eq. (12), the target mutation strength is $\sigma^*_{\text{target}} = \mu c_{\mu/\mu,\lambda}$, which is optimal according to Eq. (10). However, the normalized mutation strength actually attained by the strategy differs from the target mutation strength as the distance from the optimal solution decreases simultaneously with the step size and adaptation is not instantaneous. Calculations equivalent to those in [1] yield

$$\sigma^* = \mu c_{\mu/\mu,\lambda}\left[\left(1 - x_1^2\right)\operatorname{sgn}(x_1) + \sqrt{1 + x_1^4}\right] \tag{13}$$

for the mutation strength attained by the strategy. That is, normalized mutation strengths generated by cumulative step size adaptation for $x_1 \lessapprox 1$ exceed optimal ones by a factor of $\sqrt{2}$, resulting in a 17% loss of performance (compare [3]). For $x_1 \gtrapprox 0$ Eq. (13) suggests that mutation strengths generated by cumulative step size adaptation are nearly twice as large as optimal, resulting in near zero progress and thus stagnation of the strategy. However, as seen above, the small-step predictions are highly inaccurate for $x_1 \approx 0$ and the validity of the findings needs to be confirmed experimentally.

Figure 3 shows partial traces from typical runs of the evolution strategy for different search space dimensions. For each combination of parameter settings, 99 runs were conducted until either a solution with an objective function value within 10^{-6} of optimal was generated or 20,000 iterations were reached. The runs shown in the figure are those with the median number of iterations until termination, where ties were broken arbitrarily. Each run was initialized with $x_1 = 0$ and an initial mutation strength $\sigma_0 \in \{0.1, 1.0, 10.0\}$. It can be seen that the behaviour of the algorithm depends qualitatively on the initial mutation strength. Too large a value of σ_0 (where what is "too large" depends on N) results in the strategy operating past the first mode observed in Fig. 2. Cumulative step size adaptation in that situation either does not decrease the step size or decreases it only very slowly. The graphs on the right hand side of Fig. 3 show that the strategy in this situation jumps about apparently symmetrically about $x_1 = 0$, without reaching a point in the vicinity of the optimal solution. It can also be seen that once a point in the vicinity of the optimal solution is reached, cumulative step size adaptation controls the mutation strength as expected, and Eq. (13) quite closely predicts the further behaviour of the strategy. The observed metastable states characterized by relatively large mutation strengths and expected x_1 values of zero are more easily broken out of for small values of N, where the variance of the observed x_1 values is larger.

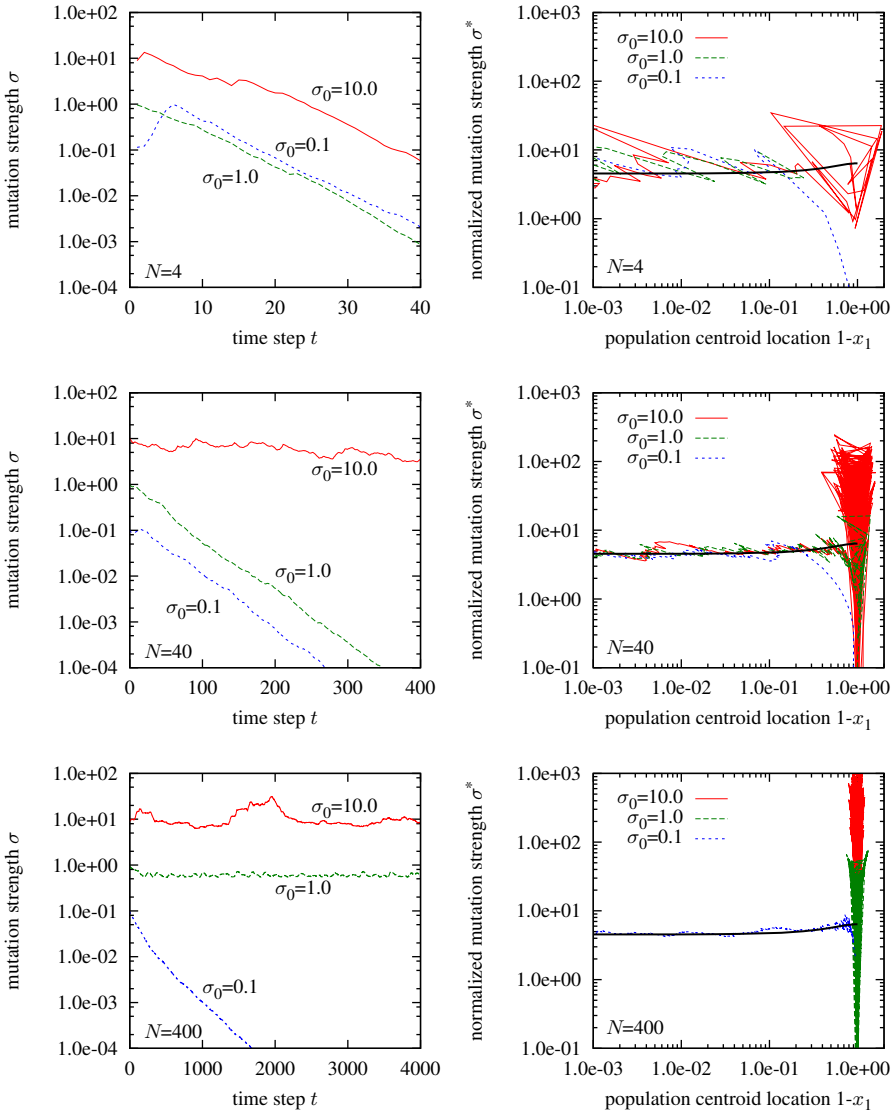

Fig. 3. Left: Mutation strength σ generated using cumulative step-size adaptation plotted against iteration number t for typical runs with $\mu = 3$ and $\lambda = 10$, initial step sizes in $\{0.1, 1.0, 10.0\}$, and, from top to bottom, search space dimensions $N = 4$, 40, and 400. Right: Normalized mutation strength σ^* from the same runs as shown on the left plotted against the transformed location $1 - x_1$ of the population centroid. The bold solid lines on the right hand side represent predictions from Eq. (13).

4 Discussion

We have presented a small-step approximation to the behaviour of a simplified variant of the algorithm of Colutto et al. [5] applied to a class of unimodal optimization problems on spherical manifolds. The approximation quite accurately describes the behaviour of the strategy in the vicinity of the optimal solution, but it is insufficient as a model in greater distance from that solution. In the latter case, large mutation strengths can result in the strategy operating in a metastable state rather than converging to the optimal solution. The analysis also suggests an approach for avoiding such metastable states: limiting the step size to at most $\sigma \approx \pi/(2\sqrt{N})$ ensures that the strategy does not operate significantly past the first mode in Fig. 2 and effectively prevents the long periods of stagnation observed in Fig. 3. In future work, we will attempt to derive an equivalent cap in the case of general Riemannian manifolds and compare the performance of the resulting algorithm with that of the approaches in *Manopt* [4].

References

[1] Arnold, D.V., Beyer, H.-G.: Performance analysis of evolutionary optimization with cumulative step length adaptation. IEEE Transactions on Automatic Control 49(4), 617–622 (2004)
[2] Beyer, H.-G.: The Theory of Evolution Strategies. Springer (2001)
[3] Beyer, H.-G., Arnold, D.V.: Qualms regarding the optimality of cumulative path length control in CSA/CMA-evolution strategies. Evolutionary Computation 11(1), 19–28 (2003)
[4] Boumal, N., Mishra, B., Absil, P.-A., Sepulchre, R.: Manopt, a Matlab toolbox for optimization on manifolds. Journal of Machine Learning Research 15, 1455–1459 (2014)
[5] Colutto, S., Frühauf, F., Fuchs, M., Scherzer, O.: The CMA-ES on Riemannian manifolds to reconstruct shapes in 3-D voxel images. IEEE Transactions on Evolutionary Computation 14(2), 227–245 (2010)
[6] do Carmo, M.P.: Riemannian Geometry. Birkhäuser (1992)
[7] Hansen, N., Ostermeier, A.: Completely derandomized self-adaptation in evolution strategies. Evolutionary Computation 9(2), 159–195 (2001)
[8] Huckemann, S., Hotz, T., Munk, A.: Intrinsic MANOVA for Riemannian manifolds with an application to Kendall's space of planar shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(4), 593–603 (2010)
[9] Kissinger, C.R., Gehlhaar, D.K., Fogel, D.B.: Rapid automated molecular replacement by evolutionary search. Acta Crystallographica D55, 484–491 (1999)
[10] Qi, C., Gallivan, K.A., Absil, P.-A.: Riemannian BFGS algorithm with applications. In: Diehl, M., et al. (eds.) Recent Advances in Optimization and its Applications in Engineering, pp. 183–192. Springer (2010)
[11] Rechenberg, I.: Evolutionsstrategie — Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Friedrich Frommann Verlag (1973)
[12] Yang, Y.: Globally convergent optimization algorithms on Riemannian manifolds: Uniform framework for unconstrained and constrained optimization. Journal of Optimization Theory and Applications 132(2), 245–265 (2007)