Cumulative Step Length Adaptation for Evolution Strategies Using Negative Recombination Weights

Dirk V. Arnold and D.C. Scott Van Wart

Faculty of Computer Science, Dalhousie University
Halifax, Nova Scotia, Canada B3H 1W5
{dirk,wart}@cs.dal.ca

Abstract. Cumulative step length adaptation is a mutation strength control mechanism commonly employed with evolution strategies. When using weighted recombination with negative weights it can be observed to be prone to failure, often leading to divergent behaviour in low-dimensional search spaces. This paper traces the reasons for this breakdown of step length control. It then proposes a novel variant of the algorithm that reliably results in convergent behaviour for the test functions considered. The influence of the dimensionality as well as of the degree of ill-conditioning on optimisation performance are evaluated in computer experiments. Implications for the use of weighted recombination with negative weights are discussed.

1 Introduction

Evolution strategies [6, 13] are nature inspired, iterative algorithms most commonly used for solving numerical optimisation problems. Such problems typically require that the mutation strength, which controls the step length of the strategies, be adapted in the course of the search. Several adaptation mechanisms have been proposed. Among them are the 1/5th success rule [11], mutative self-adaptation [11, 13], and cumulative step length adaptation [10]. The latter algorithm is particularly significant as it is the step length adaptation mechanism used in covariance matrix adaptation evolution strategies (CMA-ES) [7, 9].

In the basic form of evolution strategies, all of the selected candidate solutions are weighted equally. Weighted recombination, i.e., having candidate solutions enter recombination with different weights that are based on their ranks in the set of all offspring, has been known for some time to be an effective means for speeding up local convergence. Rudolph [12] discovered that assigning negative weights to especially unfavourable candidate solutions can result in a further speed-up. Underlying the use of negative weights is the assumption that the opposite of an unfavourable step is likely to be favourable. More recently, it has been seen that for the idealised environments of the quadratic sphere and the parabolic ridge, if the search space dimensionality is high, weighted recombination with optimally chosen weights (half of which are negative) can speed up convergence by a factor of up to 2.5 compared to unweighted recombination [2, 3, 4]. Moreover, optimal weights agree for both of those cases.

However, cumulative step length adaptation is not without problems when used in connection with negative weights. It can be observed that if the dimensionality of the

M. Giacobini et al. (Eds.): EvoWorkshops 2008, LNCS 4974, pp. 545–554, 2008.

optimisation problem at hand is low, cumulative step length adaptation may fail to generate useful step lengths and result in divergent behaviour for objective functions as simple as the sphere model. Moreover, for some problems the use of negative weights in combination with cumulative step length adaptation results in significantly reduced performance even if the search space dimensionality is high. Presumably, these problems are the reason that CMA-ES refrain from using negative weights altogether [7, 9].

In this paper, we propose a modification to the cumulative step length adaptation mechanism that enables evolution strategies employing negative weights to converge reliably in low-dimensional search spaces. Its remainder is organised as follows. Section 2 describes multirecombination evolution strategies with cumulative step length adaptation. In Section 3, the reason for the breakdown of cumulative step length control when using negative recombination weights in low-dimensional search spaces is discussed, and a modification to the mechanism is proposed. In Section 4, we evaluate the performance of the novel algorithm using several test functions. Section 5 concludes with a brief discussion.

2 Strategy

For an N-dimensional optimisation problem with objective function $f : \mathbb{R}^N \to \mathbb{R}$, the state of a weighted multirecombination evolution strategy with cumulative step length adaptation as proposed in [10] is described by search point $\mathbf{x} \in \mathbb{R}^N$, search path $\mathbf{s} \in \mathbb{R}^N$ (initialised to the zero vector), and mutation strength $\sigma \in \mathbb{R}$. The strategy repeatedly updates those quantities using the following five steps (where $\leftarrow$ denotes the assignment operator):

1. Generate λ offspring candidate solutions $\mathbf{y}^{(i)} = \mathbf{x} + \sigma\mathbf{z}^{(i)}$, $i = 1, \ldots, \lambda$. The $\mathbf{z}^{(i)}$ are vectors consisting of N independent, standard normally distributed components and are referred to as mutation vectors. The nonnegative mutation strength σ determines the step length of the strategy.
2. Determine the objective function values $f(\mathbf{y}^{(i)})$ of the offspring candidate solutions and order the $\mathbf{y}^{(i)}$ according to those values.
3. Compute the progress vector

$$\mathbf{z}^{(\text{avg})} = \sum_{k=1}^{\lambda} w_{k,\lambda}\mathbf{z}^{(k;\lambda)} \tag{1}$$

   where index $k;\lambda$ refers to the kth best of the λ offspring. The $w_{k,\lambda}$ are used to weight the mutation vectors and depend on the rank of the corresponding candidate solution in the set of all offspring.
4. Update the search point and search path according to

$$\mathbf{x} \leftarrow \mathbf{x} + \sigma\mathbf{z}^{(\text{avg})} \tag{2}$$

$$\mathbf{s} \leftarrow (1-c)\mathbf{s} + \sqrt{\mu_{\text{eff}}\, c(2-c)}\,\mathbf{z}^{(\text{avg})} \tag{3}$$

   where $c \in (0,1)$ is a cumulation parameter and where $\mu_{\text{eff}} = 1/\sum_{k=1}^{\lambda} w_{k,\lambda}^2$ denotes the "variance effective selection mass" [7].

5. Update the mutation strength according to

$$\sigma \leftarrow \sigma \exp\left(c\,\frac{\|\mathbf{s}\| - \chi_N}{\chi_N d}\right) \tag{4}$$

where $d > 0$ is a damping constant and where $\chi_N = \sqrt{2}\,\Gamma((N+1)/2)/\Gamma(N/2)$ denotes the expected value of a χ_N-distributed random variable.

Using mutative self-adaptation as a starting point, Hansen and Ostermeier [9] describe the development of cumulative step length adaptation as a sequence of steps aimed at derandomising the former mechanism. While mutative self-adaptation adapts the mutation strength based solely on differences in the length of mutation steps, cumulative step length adaptation attempts to measure correlations between the directions of successive progress vectors. The search path $\mathbf{s}$ as updated in Eq. (3) implements an exponentially fading record of steps taken by the strategy. The cumulation parameter c determines how quickly the memory of the strategy expires, with larger values resulting in a more rapid decay of the information present in $\mathbf{s}$. Higher-dimensional problems require a longer memory, and c is often chosen to be asymptotically inversely proportional to N. The coefficient in Eq. (3) which the progress vector is multiplied with is chosen such that under random selection (i.e., on a flat fitness landscape), after initialisation effects have faded the search path has independent, normally distributed components with zero mean and unit variance.

Cumulative step length adaptation relies on the assertion that ideally, consecutive steps of the strategy should be uncorrelated. Uncorrelated random steps lead to the search path having expected length χ_N. According to Eq. (4), a search path of that length results in the mutation strength remaining unchanged. If the most recently taken steps are positively correlated, they tend to point in similar directions and efficiency could be gained by making fewer but longer steps. In that case, the length of the search path exceeds χ_N and Eq. (4) acts to increase the mutation strength. Similarly, negative correlations between successive steps (i.e., the strategy stepping back and forth) lead to the mutation strength being reduced. Larger values of the damping constant d in Eq. (4) act to moderate the magnitude of the changes in mutation strength. Lower-dimensional problems in particular may require significant damping. On the other hand, too large a damping constant can negatively impact the performance of the strategy as it prevents rapid adaptation of the mutation strength and therefore rapid convergence for "simple" problems. Typically, d is chosen to be asymptotically independent of N.

Several settings for the weights $w_{k,\lambda}$ occurring in Eq. (1) can be found in the literature:

– The most common choice is to use unweighted recombination in connection with truncation selection. The corresponding weights are

$$w_{k,\lambda} = \begin{cases} 1/\mu & \text{if } k \leq \mu \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

for some $\mu \in \{1, \ldots, \lambda - 1\}$. Typically, $\mu \approx \lambda/4$. The resulting strategy is referred to as $(\mu/\mu, \lambda)$-ES [6].

– Hansen and Kern [7] set $\mu = \lfloor \lambda/2 \rfloor$ and use weighted recombination with weights

$$w_{k,\lambda} = \begin{cases} \frac{\ln(\mu+1)-\ln(k)}{\mu\ln(\mu+1)-\sum_{i=1}^{\mu}\ln(i)} & \text{if } k \leq \mu \\ 0 & \text{otherwise} \end{cases}. \tag{6}$$

They note that strategies employing this choice of weights "only slightly outperform [$(\mu/\mu, \lambda)$-style recombination with] $\mu \approx \lambda/4$" when used in the context of the CMA-ES.

– In [2], setting

$$w_{k,\lambda} = \frac{E_{k,\lambda}}{\kappa} \tag{7}$$

for some $\kappa > 0$ is proposed, where

$$E_{k,\lambda} = \frac{1}{\sqrt{2\pi}} \frac{\lambda!}{(\lambda-k)!(k-1)!} \int_{-\infty}^{\infty} x\mathrm{e}^{-\frac{1}{2}x^2} [\Phi(x)]^{\lambda-k} [1-\Phi(x)]^{k-1} \mathrm{d}x$$

denotes the expected value of the $(\lambda + 1 - k)$th order statistic of a sample of λ independent, standard normally distributed random variables and Φ denotes the cumulative distribution function of the standard normal distribution. In [2, 3, 4], this setting has been seen to be the optimal choice of weights for both the quadratic sphere and parabolic ridge models in the limit of very high search space dimensionality. The resulting strategy is referred to as $(\lambda)_{\text{opt}}$-ES and in both of those environments outperforms the $(\mu/\mu, \lambda)$-ES with optimally chosen μ by a factor of up to 2.5 in the limit $N \to \infty$. The scalar quantity κ in Eq. (7) is referred to as the rescaling factor and can be used to control the length of the progress vectors relative to that of the mutation vectors.

3 Direction Based Cumulative Step Length Adaptation

While for sufficiently high search space dimensionality the $(\lambda)_{\text{opt}}$-ES often converges faster than evolution strategies that use nonnegative recombination weights, it may fail to converge altogether even for simple test functions if N is small. Figure 1 contrasts the performance of the $(\lambda)_{\text{opt}}$-ES with $\lambda = 10$ with that of a positively weighted strategy when optimising the quadratic sphere model (see Table 1 for a definition). For that model, evolution strategies converge linearly provided that the mutation strength is adapted successfully. The normalised quality gain Δ^* as defined for example in [1, Chapter 6] is a measure for the speed of convergence[1]. The positively weighted strategy employs weights $w_{k,\lambda} = \max(0, E_{k,\lambda}/\kappa)$ (i.e., the weights are as prescribed by Eq. (7) for ranks $k \leq \lambda/2$, and they are zero for the remaining candidate solutions). The rescaling factor κ is set to 3.68 as for $\lambda = 10$, that setting leads to approximately the same value of μ_{eff} as Eq. (6) does. The resulting choice of weights is nearly identical to that proposed by Hansen and Kern [9], and the strategy is referred to as $(\lambda)_{\text{pos}}$-ES. The $(\lambda)_{\text{opt}}$-ES differs from the $(\lambda)_{\text{pos}}$-ES only in that it uses negative weights as prescribed by Eq. (7) for candidate solutions with ranks $k > \lambda/2$. The solid lines in Fig. 1 show

[1] For the sphere model, this definition agrees with the definition of the log-progress rate in [5].

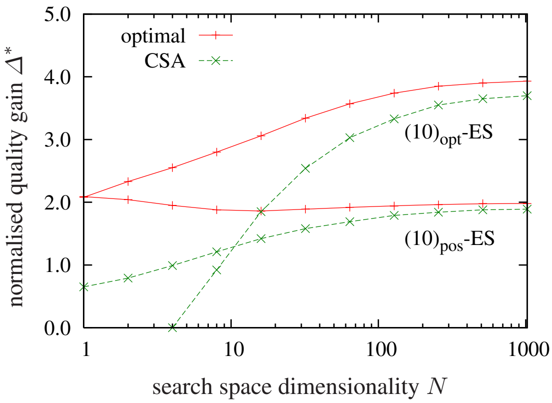

Fig. 1. Normalised quality gain Δ^* for the sphere model of $(10)_{\text{pos}}$-ES and $(10)_{\text{opt}}$-ES plotted against search space dimensionality N. Shown are both the quality gain achieved with optimal step lengths and the quality gain achieved with cumulative step length adaptation.

the maximal quality gain that can be achieved if at every step the mutation strength is set optimally. The data points have been obtained by numerically optimising the normalised mutation strength of the strategy (which, of course, does not constitute a viable step length control algorithm in general as it assumes knowledge of the distance of the search point from the location of the optimum). It can be seen that the $(\lambda)_{\text{opt}}$-ES is capable of nearly twice the normalised quality gain of the $(\lambda)_{\text{pos}}$-ES if the search space dimensionality is high. For smaller values of N the potential advantage of the negatively weighted strategy is smaller, but it is present for N as small as two. The dashed lines in Fig. 1 show the normalised quality gain of the strategies when using cumulative step length adaptation. In accordance with [9], the cumulation parameter c that appears in Eq. (3) is set to $4/(N+4)$. The damping constant d has been optimised numerically. It can be seen that for large N the performance of the adaptive strategies appears to approach optimal behaviour for both $(\lambda)_{\text{pos}}$-ES and $(\lambda)_{\text{opt}}$-ES. However, while for low search space dimensionalities cumulative step length adaptation results in convergence for the $(\lambda)_{\text{pos}}$-ES, the $(\lambda)_{\text{opt}}$-ES fails to generate positive quality gain.

It is not immediately clear why cumulative step length adaptation enables the $(\lambda)_{\text{pos}}$-ES to converge for the sphere model while the $(\lambda)_{\text{opt}}$-ES does not when N is small. It is certainly true that the opposite of a bad mutation is not always a good mutation, and the use of negative weights may contribute to bad steps being made. However, Fig. 1 clearly illustrates that the negatively weighted strategy is capable of convergence even for small N, and that the observed failure to converge is due to imperfect step length adaptation rather than to the use of negative weights per se. Interestingly, not shown, the $(\lambda)_{\text{pos}}$-ES employing cumulative step length adaptation is capable of convergence on the one-dimensional sphere even if the cumulation parameter c is set to 1. As for $c = 1$ no cumulation takes place, it cannot be correlations between consecutive steps that enable the $(\lambda)_{\text{pos}}$-ES to converge. Instead, the strategy is able to use information with regard to the length of the selected mutation vectors. If successful mutation vectors are short, then so is the search path and Eq. (4) acts to reduce the mutation strength. If it is the longer than average steps that are successful, then the mutation strength will be increased. The $(\lambda)_{\text{opt}}$-ES on the other hand utilises not only good mutation vectors but also bad ones. Unsuccessful mutation vectors enter the averaging in Eq. (1) with a negative sign, eliminating the correlation between the length of successful mutation vectors and the length of the search path and thus preventing successful adaptation of the mutation strength.

As seen in Fig. 1, the use of negative weights has significant potential benefits. It is thus desirable to have a step length control mechanism that is capable of generating useful mutation strengths in low-dimensional search spaces even if length information about successful mutation vectors cannot be utilised. Such a mechanism can be devised by making mutation strength updates more explicitly dependent on directions of recent progress vectors, using another level of indirection in order to be able to cope with the noisy signal. Specifically, we propose to introduce a mutation strength modifier $h \in \mathbb{R}$, and to replace steps 4 and 5 of the algorithm described in Section 2 with:

4'. Update the search point, mutation strength modifier, and search path according to

$$\begin{aligned} \mathbf{x} &\leftarrow \mathbf{x} + \sigma \mathbf{z}^{(\text{avg})} \\ h &\leftarrow (1 - c_h)h + c_h \mathbf{s} \cdot \mathbf{z}^{(\text{avg})} \\ \mathbf{s} &\leftarrow (1 - c)\mathbf{s} + \sqrt{\mu_{\text{eff}}\, c(2 - c)}\mathbf{z}^{(\text{avg})} \end{aligned} \tag{8}$$

where $c \in (0, 1)$ and $c_h \in (0, 1)$ are cumulation parameters.

5'. Update the mutation strength according to

$$\sigma \leftarrow \sigma \exp\left(\frac{h}{Nd}\right) \tag{9}$$

where $d > 0$ is a damping constant.

The inner product $\mathbf{s} \cdot \mathbf{z}^{(\text{avg})}$ in Eq. (8) is greater than zero if the progress vector points in the predominant direction of the previous steps accumulated in the search path. It is negative if the direction of $\mathbf{z}^{(\text{avg})}$ is opposite to that of $\mathbf{s}$. Rather than immediately using the inner product to update the mutation strength, $\mathbf{s} \cdot \mathbf{z}^{(\text{avg})}$ contributes to the mutation strength modifier h. This additional level of indirection is reminiscent of the use of a momentum term and acts as a low pass filter. It is useful as it reduces fluctuations of the mutation strength in low-dimensional search spaces, and it is without relevance if N is large. We have found setting $c_h = 0.1$ useful in all of our experiments.

4 Experimental Evaluation

In order to evaluate the usefulness of the step length control mechanism proposed in Section 3, we conduct computer experiments involving a subset of the test functions employed in [9]. The functions used are listed in Table 1. The first four are convex quadratic and have a unique optimum at $(0, 0, \ldots, 0)$. For each of them, the search point of the evolution strategies is initialised to $\mathbf{x} = (1, 1, \ldots, 1)$. While the eigenvectors of the Hessians of f_{cigar}, f_{discus}, and $f_{\text{ellipsoid}}$ are aligned with the coordinate

Table 1. Test functions

sphere	$f_{\text{sphere}}(\mathbf{x}) = \sum_{i=1}^{N} x_i^2$
cigar	$f_{\text{cigar}}(\mathbf{x}) = x_1^2 + \sum_{i=2}^{N} (ax_i)^2$
discus	$f_{\text{discus}}(\mathbf{x}) = (ax_1)^2 + \sum_{i=2}^{N} x_i^2$
ellipsoid	$f_{\text{ellipsoid}}(\mathbf{x}) = \sum_{i=1}^{N} \left(a^{\frac{i-1}{N-1}} x_i\right)^2$
Rosenbrock	$f_{\text{Rosen}}(\mathbf{x}) = \sum_{i=1}^{N-1} \left(100\left(x_i^2 - x_{i+1}\right)^2 + (x_i - 1)^2\right)$

axes, none of the strategies considered here make use of the separability of those functions, and the same results would be observed if an arbitrary rotation were applied to the coordinate system. The parameter a controls the degree of ill-conditioning of the functions. For $a = 1$ all of them are identical to the sphere function. Rosenbrock's function is characterised by a long, bent valley that needs to be followed in order to arrive at the global optimum at location $(1, 1, \ldots, 1)$. For f_{Rosen}, the initial search point is $\mathbf{x} = (0, 0, \ldots, 0)$. For all of the functions, the optimal objective function value is zero. The initial mutation strength is $\sigma = 1$ for all functions but Rosenbrock's, for which it is $\sigma = 0.1$ as this setting effectively prevents convergence to the local optimum the existence of which is noted in [9]. Optimisation proceeds until an objective function value $f(\mathbf{x}) \leq 10^{-10}$ is reached.

Throughout this section, the strategies with step length adaptation as described in Section 2 are referred to as $(\lambda)_{\text{opt}}$-CSA-ES and $(\lambda)_{\text{pos}}$-CSA-ES, depending on whether negative weights are used or not. The negatively weighted strategy with direction based cumulative step length adaptation as described in Section 3 is referred to as $(\lambda)_{\text{opt}}$-dCSA-ES. In all runs, $\lambda = 10$ and $\kappa = 3.68$. Setting the cumulation parameter to $c = 4/(N + 4)$ has proven useful for all strategy variants. The damping constant is set to $d = 1 + c$ for the $(\lambda)_{\text{pos}}$-CSA-ES according to a recommendation in [9]. For the $(\lambda)_{\text{opt}}$-CSA-ES and the $(\lambda)_{\text{opt}}$-dCSA-ES, parameter settings of $d = 0.25 + 4/N$ and $d = 1 + 7/N$, respectively, have been employed after some numerical experimentation.

We first conduct a series of experiments with the goal of comparing the performance of the step length adaptation mechanisms for mildly ill-conditioned problems. This is motivated by the CMA-ES striving to learn covariance matrices that locally transform arbitrary objective functions into the sphere model [9]. If covariance matrix adaptation is successful, it can be hoped that objective functions can locally be approximated by convex quadratic functions with low condition numbers of their Hessians. Figure 2 compares the optimisation performance of the three strategy variants for the four convex quadratic test functions, where $a = 4$ for the cigar, discus, and ellipsoid functions. It can be seen that the strategies employing negative weights are significantly superior to the $(\lambda)_{\text{pos}}$-CSA-ES if the search space dimensionality is high, outperforming the latter by a factor between 1.9 and 2.8. Moreover, while the $(\lambda)_{\text{opt}}$-CSA-ES fails to converge for small values of N for three of the four test functions, direction based cumulative step length adaptation is successful in that the $(\lambda)_{\text{opt}}$-dCSA-ES consistently either matches or outperforms the $(\lambda)_{\text{pos}}$-CSA-ES even in low-dimensional search spaces.

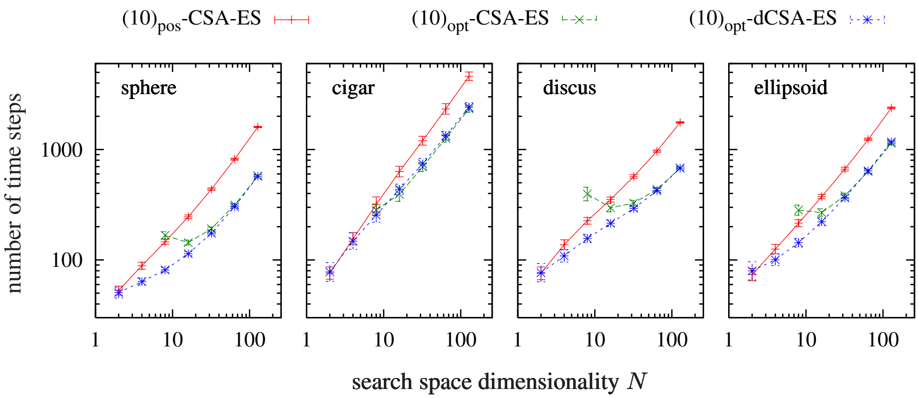

Fig. 2. Number of time steps required to reach an objective function value of $f(\mathbf{x}) = 10^{-10}$ plotted against the dimensionality N of the search space for $a = 4$. Results are averaged over 100 independent runs, with error bars indicating the standard deviation of the measurements.

Next, we investigate the effect of the degree of ill-conditioning on optimisation performance. Figure 3 contrasts the performance of $(\lambda)_{\text{pos}}$-CSA-ES and $(\lambda)_{\text{opt}}$-dCSA-ES for the anisotropic convex quadratic objective functions. It can be seen that the advantage afforded by the use of negative weights that is present for small values of a generally decreases or even turns into a disadvantage as the degree of ill-conditioning increases. While the use of negative weights generally seems to do no harm for the cigar function, refraining from weighting mutation vectors negatively results in significantly better performance for the discus and ellipsoid functions unless a is sufficiently small or N is sufficiently large.

Finally, we are interested in the potential of the use of negative recombination weights for "less artificial" objective functions. While the degree of ill-conditioning of

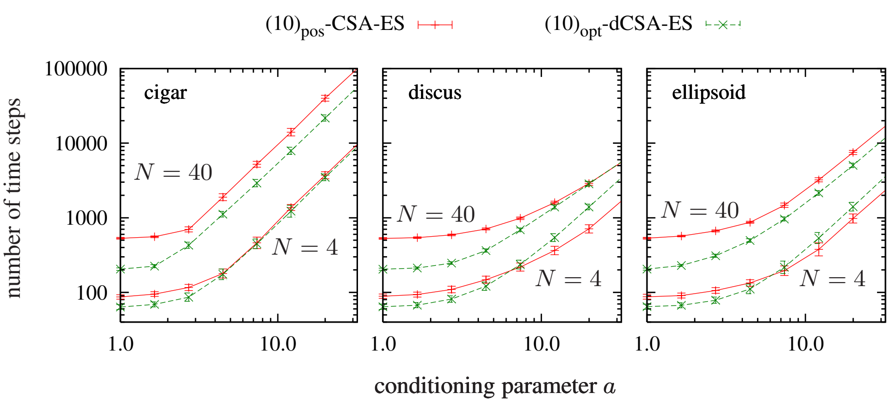

Fig. 3. Number of time steps required to reach an objective function value of $f(\mathbf{x}) = 10^{-10}$ plotted against the conditioning parameter a for $N = 4$ and $N = 40$. Results are averaged over 100 independent runs, with error bars indicating the standard deviation of the measurements.

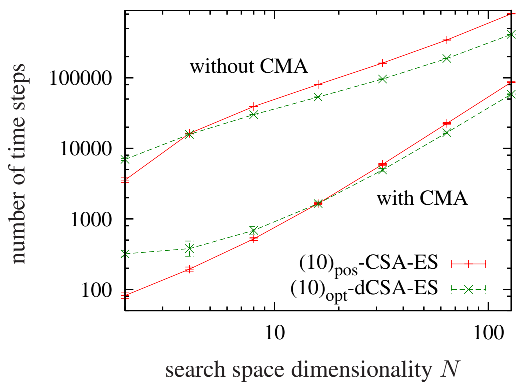

Fig. 4. Number of time steps required to reach an objective function value of $f(\mathbf{x}) = 10^{-10}$ plotted against the search space dimensionality N for Rosenbrock's function. Results are averaged over 100 independent runs, with error bars indicating the standard deviation of the measurements.

f_{cigar}, f_{discus}, and $f_{\text{ellipsoid}}$ can be increased indefinitely, it is unclear how relevant such large condition numbers are for real-world optimisation. As noted above, a covariance matrix adaptation algorithm may lead to the strategies "seeing" only moderately ill-conditioned problems, which the experiments above suggest negative weights may be beneficial for. We have thus equipped the evolution strategies with the covariance matrix adaptation algorithm described in [9]. The settings of the cumulation parameters of the $(\lambda)_{\text{pos}}$-ES are as described in that reference. For the $(\lambda)_{\text{opt}}$-ES, the parameter c_{cov} that governs the rate at which the covariance matrix is updated is reduced from $2/(N+\sqrt{2})^2$ to $2/(N+8)^2$ in order to achieve reliable convergence for small values of N.

Figure 4 contrasts the performance of the $(\lambda)_{\text{opt}}$-dCSA-ES with that of the $(\lambda)_{\text{pos}}$-CSA-ES, both with and without covariance matrix adaptation, for Rosenbrock's function. It can be seen that the use of negative weights when optimising f_{Rosen} is generally beneficial if the search space dimensionality is high. Without covariance matrix adaptation, the $(\lambda)_{\text{opt}}$-dCSA-ES outperforms the $(\lambda)_{\text{pos}}$-CSA-ES for $N > 4$. With covariance matrix adaptation, the slower adaptation of that matrix in the $(\lambda)_{\text{opt}}$-dCSA-ES that results from the different setting of c_{cov} leads to the $(\lambda)_{\text{pos}}$-CSA-ES enjoying a performance advantage up to $N = 16$.

5 Conclusions

To summarise, in low-dimensional search spaces the existing approach to cumulative step length adaptation implicitly combines information with regard to the length of selected mutation vectors with directional information about consecutive steps. For evolution strategies using negative recombination weights, the progress vector does not contain useful length information and cumulative step length adaptation is prone to failure. This paper has introduced a novel variant of cumulative step length adaptation that more explicitly relies on directional information, and that introduces an additional level of indirection in order to improve robustness in low-dimensional search spaces.

Computer experiments have confirmed that the novel algorithm reliably results in convergent behaviour and outperforms strategies that do not use negative weights for moderately ill-conditioned problems. For higher degrees of ill-conditioning, the use of negative weights can deteriorate performance unless the search space dimensionality is high.

In future work we will study the effect of the number of offspring on the findings made here. Using the improved covariance matrix update for large populations proposed in [8] may affect the rate at which adaptation can occur for negatively weighted strategies in low-dimensional search spaces. A further goal will be to devise a mechanism that can be used to adaptively scale the negative recombination weights with the goal of maximising the convergence rate for various degrees of ill-conditioning.

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

References

[1] Arnold, D.V.: Noisy Optimization with Evolution Strategies. Kluwer Academic Publishers, Dordrecht (2002)
[2] Arnold, D.V.: Optimal weighted recombination. In: Wright, A.H., et al. (eds.) Foundations of Genetic Algorithms 8, pp. 215–237. Springer, Heidelberg (2005)
[3] Arnold, D.V.: Weighted multirecombination evolution strategies. Theoretical Computer Science 361(1), 291–308 (2006)
[4] Arnold, D.V., MacDonald, D.: Weighted recombination evolution strategies on the parabolic ridge. In: Proc. of the 2006 IEEE Congress on Evolutionary Computation, pp. 411–418. IEEE Press, Los Alamitos (2006)
[5] Auger, A., Hansen, N.: Reconsidering the progress rate theory for evolution strategies in finite dimensions. In: Proc. of the 8th Annual Conference on Genetic and Evolutionary Computation, pp. 445–452. ACM Press, New York (2006)
[6] Beyer, H.-G., Schwefel, H.-P.: Evolution strategies — A comprehensive introduction. Natural Computing 1(1), 3–52 (2002)
[7] Hansen, N., Kern, S.: Evaluating the CMA evolution strategy on multimodal test functions. In: Yao, X., Burke, E.K., Lozano, J.A., Smith, J., Merelo-Guervós, J.J., Bullinaria, J.A., Rowe, J.E., Tiňo, P., Kabán, A., Schwefel, H.-P. (eds.) PPSN 2004. LNCS, vol. 3242, pp. 282–291. Springer, Heidelberg (2004)
[8] Hansen, N., Müller, S.D., Koumoutsakos, P.: Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). Evolutionary Computation 11(1), 1–18 (2003)
[9] Hansen, N., Ostermeier, A.: Completely derandomized self-adaptation in evolution strategies. Evolutionary Computation 9(2), 159–195 (2001)
[10] Ostermeier, A., Gawelczyk, A., Hansen, N.: Step-size adaptation based on non-local use of selection information. In: Davidor, Y., Männer, R., Schwefel, H.-P. (eds.) PPSN 1994. LNCS, vol. 866, pp. 189–198. Springer, Heidelberg (1994)
[11] Rechenberg, I.: Evolutionsstrategie — Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Friedrich Frommann Verlag (1973)
[12] Rudolph, G.: Convergence Properties of Evolutionary Algorithms. Verlag Dr. Kovač (1997)
[13] Schwefel, H.-P.: Evolution and Optimum Seeking. John Wiley & Sons, Chichester (1995)