Covariance matrix adaptation revisited - The CMSA Evolution Strategy

Hans-Georg Beyer, Bernhard Sendhoff

2008

Preprint:

This is an accepted article published in Parallel Problem Solving From Nature X. The final authenticated version is available online at: https://doi.org/[DOI not available]

Covariance Matrix Adaptation Revisited – the CMSA Evolution Strategy –

Hans-Georg Beyer[1] and Bernhard Sendhoff[2]

[1] Vorarlberg University of Applied Sciences
6850 Dornbirn, Austria
hans-georg.beyer@fh-vorarlberg.ac.at
[2] Honda Research Institute Europe GmbH
Carl-Legien-Strasse 30, 63073 Offenbach/Main, Germany
bs@honda-ri.de

Abstract. The covariance matrix adaptation evolution strategy (CMA-ES) rates among the most successful evolutionary algorithms for continuous parameter optimization. Nevertheless, it is plagued with some drawbacks like the complexity of the adaptation process and the reliance on a number of sophisticatedly constructed strategy parameter formulae for which no or little theoretical substantiation is available. Furthermore, the CMA-ES does not work well for large population sizes. In this paper, we propose an alternative – simpler – adaptation step of the covariance matrix which is closer to the "traditional" mutative self-adaptation. We compare the newly proposed algorithm, which we term the CMSA-ES, with the CMA-ES on a number of different test functions and are able to demonstrate its superiority in particular for large population sizes.

1 Introduction

State-of-the-art Evolutionary Algorithms (EA) in real-valued search domains use non-isotropic mutation distributions in order to explore the search space. The Covariance Matrix Adaptation Evolution Strategy (CMA-ES), proposed by Hansen, Ostermeier, and Gawelczyk [1] and further developed in [2, 3], is currently the most widely used, and in its restart version [4] arguably the best performing EA for continuous optimization on a (sub-)set of test functions [5].[3]

At the same time, the CMA-ES is also plagued with a couple of drawbacks which we want to address in this paper by proposing an alternative adaptation of the covariance matrix. As we will see in the next section, the adaptation process in the CMA-ES is rather complex and involves a number of free parameters which have to be set with no or little theoretical guidance. Although thorough empirical investigations have been performed to identify suitable parameter settings [2, 3], still the application of the algorithm requires extensive experience.

Secondly, the performance of the CMA-ES does not scale well with increasing population size. This problem has been alleviated by the introduction of the hybrid version

[3] According to "Tutorial: Covariance Matrix Adaptation (CMA) Evolution Strategy", presented by N. Hansen at PPSN Conference, Sep. 8, 2006, Reykjavik.

of the CMA-ES [3] with direct covariance matrix estimation, which will be our starting point in the next section and which will be used for comparison with our suggested algorithm.

Additionally, the CMA-ES due to the cumulative step size adaptation experiences problems when the fitness information is disturbed by heavy noise (noisy objective functions) [6, 7] and instabilities can occur when very large populations are needed [8].

Extensions of the CMA-ES and alternative approaches to covariance matrix adaptation have been proposed in the literature. Auger et al. [9] proposed an alternative method to calculate the covariance matrix by locally estimating the Hessian (Taylor expansion) matrix, however, at the expense of a large computational overhead of $\mathcal{O}(N^6)$. A first multi-objective $(1 + \lambda)$-CMA-ES has been described in [10] that uses the "traditional" 1/5-rule for controlling the global step size.

In this paper, we will proceed in a different direction and revisit the mutative self-adaptation process in the context of covariance matrix adaptation. In the next section, we will briefly recall the CMA-ES and propose our new algorithm in Section 3. The empirical comparison between both algorithms will be described in Section 4 followed by the conclusion in the last section.

2 The $(\mu/\mu_W, \lambda)$-CMA-ES

In Figure 1 the basic $(\mu/\mu_W, \lambda)$-CMA-ES is presented. This is done at a level that assumes that the reader is already acquainted with the (hybrid) CMA-ES as described in [3].

The CMA-ES uses weighted recombination which is indicated by the subscript "W" in the strategy parentheses. The correlated mutations are generated in a two-step process where at first a vector $\mathbf{N}_l(\mathbf{0}, \mathbf{I})$ of i.i.d. standard normal random components is transformed by the matrix $\sqrt{\mathbf{C}}$ in step (L1). The resulting random vectors $\mathbf{z} = \sqrt{\mathbf{C}}\,\mathbf{N}(\mathbf{0}, \mathbf{I})$ are $\mathbf{N}(\mathbf{0}, \mathbf{C})$ distributed. The matrix $\sqrt{\mathbf{C}}$ may be interpreted as the "square root" of the covariance matrix $\mathbf{C}$. The standard way in CMA-ES [2, 3] to obtain $\sqrt{\mathbf{C}}$ is based on eigenvalue decomposition solving the eigenvalue problem. After producing the correlated Gaussian vector $\mathbf{s}$, it is scaled in length in (L2), thus, representing the mutation $\sigma\mathbf{s}$ which is finally added to the old parental state producing the offspring in (L2). The offspring's fitness is evaluated in (L3). The new parental state is calculated in (L4) by recombination of the μ best offspring realized by weighted averaging. The adaptation of $\mathbf{C}$ is performed in (L6) using a cumulated $\mathbf{p}$ vector and the generational cross momentum matrix estimate $\langle \mathbf{s}\mathbf{s}^{\mathrm{T}} \rangle_{\mathrm{w}}$ weighted by the μ_{eff}^{-1} factor. (L6) performs an exponential smoothing (averaging) where the $\mathbf{C}$ "memory" decays with $(1 - \tau_{\mathrm{c}}^{-1})^g$ (g - generation counter). The quantity τ_{c} can be interpreted as a decay time constant determining the number of generations g needed to "forget" the initial $\mathbf{C}$ matrix. It is quite clear that τ_{c} must be a function of the endogenous strategy parameters and the problem dimensionality N. In (L5) exponential smoothing is used to update the $\mathbf{p}$ vector with the direction $\langle \mathbf{s} \rangle_{\mathrm{w}}$ of the actually taken step from parent $\mathbf{y}$ at generation g to $g + 1$ which has taken place in (L4). Therefore, $\mathbf{p}$ may be regarded as the average search step. The update of the covariance matrix $\mathbf{C}$ via the $\mathbf{p}$ vector is done in such a way that *selected* steps from the past on average are also preferred in future. This resembles the *momentum term*

$(\mu/\mu_W, \lambda)$-CMA-ES (one generation cycle)

For $l = 1$ **To** λ

$$\mathbf{s}_l \leftarrow \sqrt{\mathbf{C}}\,\mathbf{N}_l(\mathbf{0}, \mathbf{I}) \quad \text{(L1)}$$

$$\mathbf{y}_l \leftarrow \mathbf{y} + \sigma\mathbf{s}_l \quad \text{(L2)}$$

$$f_l \leftarrow f(\mathbf{y}_l) \quad \text{(L3)}$$

End

$$\mathbf{y} \leftarrow \mathbf{y} + \sigma\langle\mathbf{s}\rangle_{\mathrm{w}} \quad \text{(L4)}$$

$$\mathbf{p} \leftarrow \left(1 - \frac{1}{\tau_{\mathrm{p}}}\right)\mathbf{p} + \sqrt{\frac{1}{\tau_{\mathrm{p}}}\left(2 - \frac{1}{\tau_{\mathrm{p}}}\right)}\,\sqrt{\mu_{\mathrm{eff}}}\,\langle\mathbf{s}\rangle_{\mathrm{w}} \quad \text{(L5)}$$

$$\mathbf{C} \leftarrow \left(1 - \frac{1}{\tau_{\mathrm{c}}}\right)\mathbf{C} + \frac{1}{\tau_{\mathrm{c}}}\left[\frac{1}{\mu_{\mathrm{eff}}}\mathbf{p}\mathbf{p}^T + \left(1 - \frac{1}{\mu_{\mathrm{eff}}}\right)\langle\mathbf{s}\mathbf{s}^T\rangle_{\mathrm{w}}\right] \quad \text{(L6)}$$

$$\mathbf{p}_\sigma \leftarrow \left(1 - \frac{1}{\tau_\sigma}\right)\mathbf{p}_\sigma + \sqrt{\frac{1}{\tau_\sigma}\left(2 - \frac{1}{\tau_\sigma}\right)}\,\sqrt{\mu_{\mathrm{eff}}}\,\langle\mathbf{N}(\mathbf{0}, \mathbf{I})\rangle_{\mathrm{w}} \quad \text{(L7)}$$

$$\sigma \leftarrow \sigma \exp\left[\frac{\|\mathbf{p}_\sigma\| - \overline{\chi_N}}{d\,\overline{\chi_N}}\right] \quad \text{(L8)}$$

Fig. 1. The algorithmic "essence" of the CMA-ES. Endowed with initialization an outer generation loop and an appropriated termination condition, an approximation of the optimizer is given by the final result of the parent $\mathbf{y}$. In general weighted recombination, denoted by "$\langle\cdot\rangle_{\mathrm{w}}$", is used. Note, the individuals' $\mathbf{N}$ vectors used in (L7) are from the selected individuals that have been generated in (L1). $\overline{\chi_N} = \mathrm{E}[\chi_N]$ is the expected value of the χ distribution with N degrees of freedom being the search space dimensionality. Initially, $\mathbf{C} = \sqrt{\mathbf{C}} = \mathbf{I}$, $\mathbf{p}_\sigma = \mathbf{0}$, and $\mathbf{p} = \mathbf{0}$. Basically, the following parameters have to be chosen d, τ_σ, τ_{c}, τ_{p}, and μ_{eff}.

approach in nonlinear programming. Therefore, the original form of the CMA-ES [2] can also be regarded as a *randomized momentum term* strategy.

3 The $(\mu/\mu_I, \lambda)$-CMA-σ-SA-ES

There are two main ingredients to build an efficient ES that works well on arbitrarily rotated ellipsoidal success domains:

1. a covariance matrix adaptation algorithm which is able to learn the shape of the success domain sufficiently exact and fast,
2. a routine that adapts a global step size σ

As we mentioned already in the introduction, the disadvantage of the different versions of the CMA-ES presented in the literature is the large number of exogenous strategy parameters needed. There are five main parameters (d, τ_σ, τ_{c}, τ_{p}, and μ_{eff}^{-1}) interacting

with each other dynamically. While the effect of d and τ_σ has been analyzed on the sphere model [11], the interaction with the other time constants remains unclear. Furthermore, the CMA-ES does not always behave well in robust optimization scenarios [8, 12] when the number of offspring λ is significantly larger than the parameter space dimension.

3.1 The $(\mu/\mu_I, \lambda)$-CMA-σ-SA-ES Algorithm

In the following, the CMSA-ES will be proposed based on a radical simplification of the covariance learning rule and a revival of the well-known σ-self-adaptation (σSA) approach. Figure 2 shows the contents of the generation loop. As customary in self-

$(\mu/\mu_I, \lambda)$-CMA-σ-SA-ES (one generation cycle)

For $l = 1$ **To** λ

$$\sigma_l \leftarrow \langle\sigma\rangle \mathrm{e}^{\tau \mathcal{N}_l(0,1)} \quad \text{(R1)}$$

$$\mathbf{s}_l \leftarrow \sqrt{\mathbf{C}}\,\mathbf{N}_l(\mathbf{0}, \mathbf{I}) \quad \text{(R2)}$$

$$\mathbf{z}_l \leftarrow \sigma_l \mathbf{s}_l \quad \text{(R3)}$$

$$\mathbf{y}_l \leftarrow \mathbf{y} + \mathbf{z}_l \quad \text{(R4)}$$

$$f_l \leftarrow f(\mathbf{y}_l) \quad \text{(R5)}$$

End

$$\mathbf{y} \leftarrow \mathbf{y} + \langle \mathbf{z} \rangle \quad \text{(R6)}$$

$$\mathbf{C} \leftarrow \left(1 - \frac{1}{\tau_\mathrm{c}}\right) \mathbf{C} + \frac{1}{\tau_\mathrm{c}} \langle \mathbf{s}\mathbf{s}^T \rangle \quad \text{(R7)}$$

Fig. 2. Contents of the generation loop of the self-adaptive CMA-ES. Recombination, expressed by the "$\langle\cdot\rangle$" notation, is done (in the simplest case) by mean value calculation. The covariance matrix is initially chosen to be the identity matrix, i.e. $\mathbf{C} = \sqrt{\mathbf{C}} = \mathbf{I}$. For the choice of the strategy parameters τ and τ_c, see the text.

adaptation ES, each of the λ offspring individuals has its own mutation strength σ_l which is generated by the log-normal rule in line (R1). The generation of the object parameter $\mathbf{y}_l$ is done consecutively in line (R2 – R4). First, correlated random direction $\mathbf{s}_l$ is generated in (R2). This random direction is scaled in length by the individual's mutation strength σ_l in (R3) and finally added to the parental state $\mathbf{y}$ in line (R4) producing the offspring's object parameter vector $\mathbf{y}$. Its fitness f_l is evaluated in (R5).

In line (R6), recombination of the μ best offspring is performed. In the experiments done so far, $w_m = 1/\mu$ appeared as a reasonable choice, i.e., the angular bracket operation $\langle\cdot\rangle$ is simply an averaging over the μ best offspring individuals.

The covariance matrix adaptation takes place in (R7). Comparing with the rules used in the hybrid CMA-ES in lines (L5) and (L6), Fig. 1, one sees how simple this new rule

is. Actually, it could be recovered from (L6) for $\mu_{\text{eff}} \to \infty$. As will be shown in the experimental Section 4, this CMA rule together with σ-self-adaptation yields comparable and even better results. As in the case of the object parameter recombination, recombining the generational cross momentum matrices $\mathrm{s}_m \mathrm{s}_m^{\mathrm{T}}$ is done with uniform weights (i.e., simple averaging over the contribution of the μ best individuals).

Due to the simplicity of the newly proposed self-adaptive CMSA-ES, there is a certain chance to put the choice of the (only) two endogenous strategy parameters, the learning rate τ and the covariance cumulation time constant τ_{c}, on a theoretically motivated basis.

3.2 Parameter Settings for the CMSA-ES

The Learning Parameter $\boldsymbol{\tau}$. This parameter basically influences the time needed to learn the global step size σ and its accuracy. Assuming a locally ellipsoidal fitness landscape and provided that the covariance is adapted correctly, the $\sigma \mathbf{N}_l(\mathbf{0}, \mathbf{I})$ vectors in the CMSA-ES of Fig. 2 "experience" conditions similar to a spherical landscape. That is, under steady state conditions, one can use the τ which maximizes the steady state progress rate on the sphere model. As can be shown (due to space restrictions the derivation is beyond the scope of this paper) for sufficiently large μ, λ, and N this is the case for

$$\tau_{\text{opt}} = \frac{1}{\sqrt{2N}}. \tag{1}$$

This value has been used in the simulations of the CMSA-ES presented below. Note, this choice is *not* the optimal one for the initial phase of covariance adaptation. If one wants to increase the speed by which the $\mathbf{C}$ matrix is adapted, smaller values (e.g. $\tau = \tau_{\text{opt}}/2$) should be used. A strategy that provides a "second order" adaptation of τ could be envisioned, but has not been tested yet.

The $\boldsymbol{\tau_{\mathrm{c}}}$ Time Constant. The covariance learning rule (R7) contains the covariance learning time constant τ_{c}, the choice of which can be derived by information theoretical means. There are two aspects that must be considered: (1) the information dynamics of the covariance update; and (2) the minimum information needed to determine a covariance matrix. Again we must defer the derivation steps to an upcoming paper. The final result of the derivation is

$$\tau_{\mathrm{c}} = 1 + \frac{N(N+1)}{2\mu}. \tag{2}$$

This formula will be used in the simulations of the CMSA-ES in Section 4.

An Alternative Approach to $\sqrt{\mathbf{C}}$. Calculating $\sqrt{\mathbf{C}}$ via spectral decomposition requires the solution of the eigenvalue problem. While that approach provides additional information w.r.t. the sensitivity of the fitness landscape in the vicinity of the optimizer state, it is computationally demanding and not always required. Dropping the symmetry

of the $\sqrt{\mathbf{C}}$ matrix, the Cholesky decomposition offers a much simpler alternative which does not need the eigenvalue decomposition. Standard Cholesky decomposition yields a upper triangular matrix in $\mathcal{O}(N^3)$ floating point operations the outcome of which can directly be used as $\sqrt{\mathbf{C}}^{\mathrm{T}}$. That is, the $\mathbf{s}$ vectors are obtained by matrix multiplication of the transposed outcome of the Cholesky algorithm with the standard normal vector $\mathbf{N}(\mathbf{0}, \mathbf{I})$.

4 Comparison between CMSA-ES and CMA-ES

In order to demonstrate the effectiveness of the $\mathbf{C}$ adaptation rule (R7) in Fig. 2 and the choice of the parameters, empirical investigations are necessary to evaluate the behavior of the CMSA-ES and to compare it with the state-of-the-art $(\mu/\mu_W, \lambda)$-CMA-ES [13].

The CMSA-ES is a straightforward implementation of the algorithm in Fig. 2 using (2) and (1) for τ_{c} and τ, respectively. A truncation ratio of $\mu/\lambda = 1/4$ has been used throughout the simulations. This may be regarded as a compromise w.r.t. the progress rate under non-noisy conditions and final fitness error under additive symmetric fitness noise with constant strength (e.g. constant standard deviation) [8]. Furthermore, this choice is consonant with Hansen's recommendation to use "variance effective selection mass" $\mu_{\mathrm{eff}} = \lambda/4$ in the hybrid CMA-ES which transfers to $\mu = \lambda/4$ in the case of intermediate (uniformly weighted) recombination.

4.1 Test Functions

Tests have been performed on 12 test functions belonging to different problem classes. We will report results for four of them displayed in Tab. 1 each representing one class. The results of the other eight functions are qualitatively similar to these classes. We chose the sphere function as a kind of baseline for all continuous optimization tasks, the Schwefel ellipsoid because of the required adaptation of the covariance matrix and its special spectrum, the Rosenbrock function because it requires continuous change of the covariance matrix and the Rastrigin function because of its multi-modality.

Name	Function	$\mathbf{y}_{\mathrm{init}}$	σ_{init}	f_{stop}
Sphere	$f_{\mathrm{Sp}}(\mathbf{y}) := \sum_{i=1}^{N} y_i^2$	$(1, \ldots, 1)$	1	10^{-10}
Schwefel Ellipsoid	$f_{\mathrm{Sch}}(\mathbf{y}) := \sum_{i=1}^{N} \left(\sum_{j=1}^{i} y_i \right)^2$	$(1, \ldots, 1)$	1	10^{-10}
Rosenbrock	$f_{\mathrm{Ros}}(\mathbf{y}) := \sum_{i=1}^{N-1} \left(100(y_i^2 - y_{i+1})^2 + (y_i - 1)^2 \right)$	$(0, \ldots, 0)$	0.1	10^{-10}
Rastrigin	$f_{\mathrm{Ras}}(\mathbf{y}) := 10N + \sum_{i=1}^{N} \left(y_i^2 - 10\cos(2\pi y_i) \right)$	$\|\mathbf{y}\| = 10$	5	10^{-10}

Table 1. Test functions, initialization, and stop criterion for the evaluation of the CMA-ES.

Note, all test functions except Rastrigin's use a deterministic initialization for the object parameter vector $\mathbf{y}$. In the case of Rastrigin's function, the initial vector is randomly initialized on a hypersphere with given radius $\|\mathbf{y}\|$.

4.2 Simulation Settings

The simulation settings are directly taken from [13]. Both algorithms are compared for search space dimensionalities $N = 2, 3, 5, 10, 20, 40, 80$, and 160 considering offspring populations sizes $\lambda = 8$, $\lambda = 4N$, and $\lambda = 4N^2$. For the latter population sizes, the maximum dimensionality of $N = 80$ has been chosen in order to keep the simulation time within reasonable limits. For each N-λ-combination, 20 independent runs have been used to obtain the average number of generations to reach f_{stop} (given in Tab. 1). These average generation numbers together with the corresponding standard deviation (displayed as error bars) vs. search space dimensionality N are displayed in the plots.

4.3 Results

The somewhat surprising results for the sphere function are presented in Fig. 3. Usually it is expected that the CMA-ES works better than self-adaptive ES on the sphere model due to the use of *cumulative step length* adaptation (CSA) in the CMA-ES [6]. Since

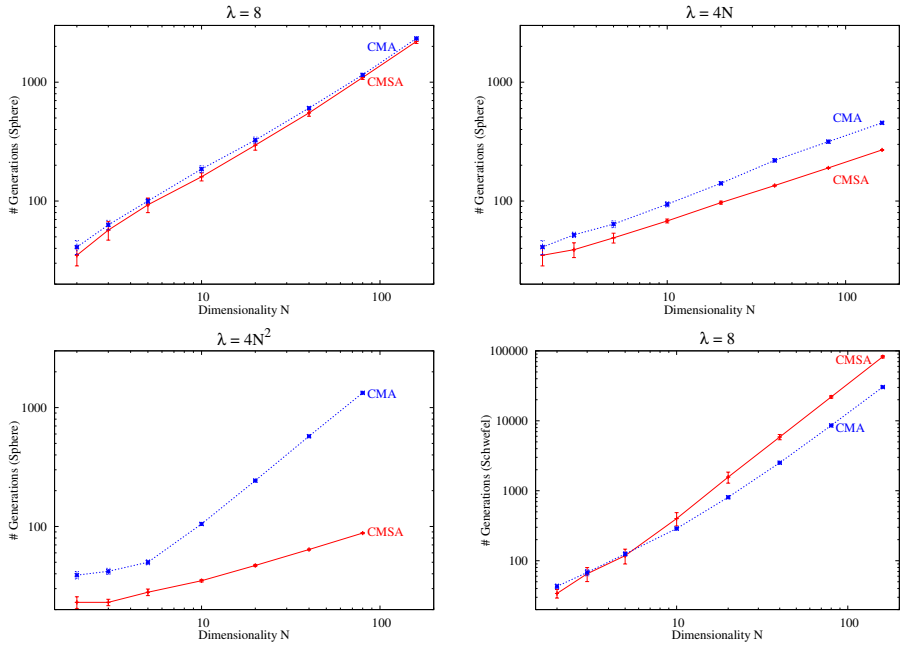

Fig. 3. Top row and bottom left: performance comparison on the sphere test function. Bottom right: performance comparison on Schwefel's Ellipsoid test function for constant $\lambda = 8$.

both CMA and CMSA start with an initial covariance matrix $\mathbf{C} = \mathbf{I}$, i.e., with isotropic mutations, the superiority of CSA must be questioned. This is consonant with observations that the CMA-ES does not work well with population sizes $\lambda \gg N$. However, even more remarkable is the observation that the new *self-adaptive* CMSA-ES works comparably well in the small population and small search space dimensionality regime.

Originally, the CMA-ES and its recent hybrid versions were designed to adapt to arbitrary quadratic test functions. Therefore, the comparison of the performance on the ellipsoidal test function class provides a good basis to evaluate the different strategies. "Schwefel's Ellipsoid" is a rotated ellipsoid with moderately increasing eigenvalue spectrum (w.r.t. N), but an isolated largest eigenvalue. As can be seen in the left graph of Fig.4, the performance of the CMSA changes to the worse (compared to CMA) if

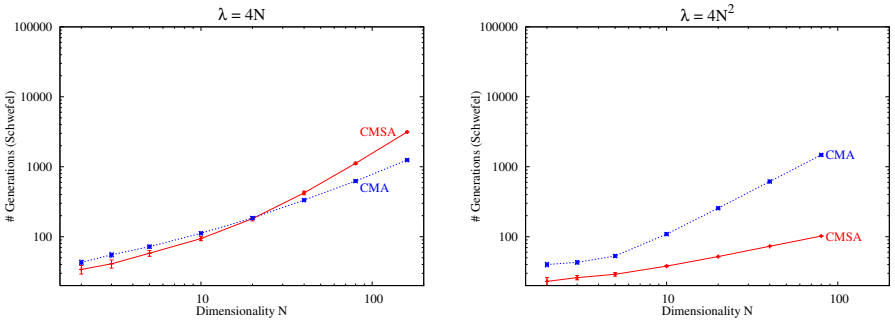

Fig. 4. Detailed performance comparison on Schwefel's Ellipsoid test function.

N gets larger. This is due to the increasing condition number of the mixing matrix in the ellipsoid function when N gets larger. However, as to large population sizes (right graph in Fig.4), CMSA performs better.

The Rosenbrock function seems to be somewhat harder for the CMSA-ES as can be seen in Fig. 5 in the case of constant and linear population sizing. In the case of quadratic population sizing both strategies perform nearly equally well. It seems that the path cumulation with decay rates proportional to $1/N$ (or larger) is a necessary ingredient in CMA-ES to effectively change the covariance matrix. This cannot be accomplished by the simple update rule (R7) used in our CMSA-ES when using population sizes of $\mathcal{O}(N)$.

For Rastrigin's function, only the quadratic population sizing has been used because the constant $\lambda = 8$ and the linear population sizing $\lambda = 4N$ does not ensure convergence to the global optimizer. It is to be mentioned that the quadratic population sizing $\lambda = 4N^2$ is *not* the optimal population sizing for this problem class. Actually, the optimal population sizing is weakly sublinear so that $\lambda \propto N$ would be the better choice. However, the proportionality factor is rather large. That is why, one observes convergence to local optima in runs with $\lambda = 4N^2$ for small N. This is also reflected in the larger standard deviations of the generation numbers in Fig. 5 (bottom right). As to the performance, one sees that the CMSA-ES clearly beats the CMA-ES. Similar behav-

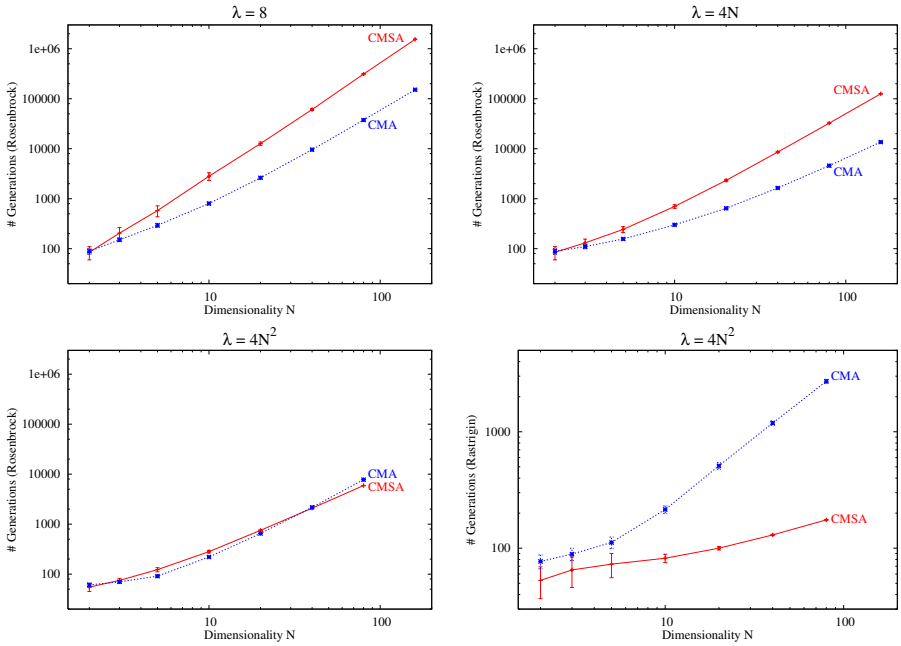

Fig. 5. Detailed performance comparison on Rosenbrock's test function and on Rastrigin test function (bottom right figure).

ior can be expected for other multi-modal test functions where the global optimizer is surrounded by a huge number of local optima.

5 Summary and Conclusion

In this paper, we have outlined the new $(\mu/\mu_I, \lambda)$-CMA-σ-SA-ES algorithm that uses mutative self-adaptation instead of cumulative step length adaptation to adjust the global step size σ during search. Compared to the standard CMA-ES which has (at least) *four* exogenous strategy parameters to be fixed, our new strategy contains only two, the time constants τ and τ_c. While the choice of some of those strategy parameters in CMA-ES is based on extensive empirical investigations, the new CMSA-ES time constants rely on information theoretical considerations.

The comparison of the CMSA-ES with the current state-of-the-art Evolution Strategy for real-valued parameter optimization, revealed a general pattern. While the CMA-ES performed slightly better for small population sizes, the newly proposed CMSA-ES achieved considerably better results for large population sizes. Surprisingly, for the sphere function both algorithms worked equally well even for small population sizes. Generally, we believe that due to its improved clarity and simplicity, the newly proposed algorithm is a serious competitor to the established CMA-ES. In case of large populations, we clearly recommend to employ the CMSA-ES. Large populations are required

in particular in the context of robust optimization [8, 12]. Even for practical applications large populations can be feasible, e.g. in the context of massive parallelization or rapid serialization of experiments like in quantum control [14]. Therefore, the increased performance for larger population sizes of the proposed CMSA-ES has potential practical implications.

References

[1] N. Hansen, A. Ostermeier, and A. Gawelczyk. On the Adaptation of Arbitrary Normal Mutation Distributions in Evolution Strategies: The Generating Set Adaptation. In L. J. Eshelman, editor, *Proc. 6th Int'l Conf. on Genetic Algorithms*, pages 57–64, San Francisco, CA, 1995. Morgan Kaufmann Publishers, Inc.

[2] N. Hansen and A. Ostermeier. Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation*, 9(2):159–195, 2001.

[3] N. Hansen, S.D. Müller, and P. Koumoutsakos. Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES). *Evolutionary Computation*, 11(1):1–18, 2003.

[4] A. Auger and N. Hansen. A Restart CMA Evolution Strategy with Increasing Population Size. In *Congress on Evolutionary Computation*, volume 2, pages 1769–1776. IEEE, 2005.

[5] P.N. Suganthan, N. Hansen, J.J. Liang, K. Deb, Y.-P. Chen, A. Auger, and S. Tiwari. Problem Definitions and Evaluation Criteria for the CEC 2005 Special Session on Real-Parameter Optimization. Technical Report, Nanyang Tech. University, Singapore, 2005.

[6] H.-G. Beyer and D.V Arnold. Qualms Regarding the Optimality of Cumulative Path Length Control in CSA/CMA-Evolution Strategies. *Evolutionary Computation*, 11(1):19–28, 2003.

[7] H.-G. Beyer, M. Olhofer, and B. Sendhoff. On the Behavior of $(\mu/\mu_I, \lambda)$-ES Optimizing Functions Disturbed by Generalized Noise. In K. De Jong, R. Poli, and J. Rowe, editors, *Foundations of Genetic Algorithms, 7*, pages 307–328, San Francisco, CA, 2003. Morgan Kaufmann.

[8] H.-G. Beyer and B. Sendhoff. Evolution Strategies for Robust Optimization. In *Congress on Evolutionary Computation (CEC)*, pages 1346–1353. IEEE Press, 2006.

[9] A. Auger, M. Schoenauer, and N. Vanhaecke. LS-CMA-ES: A Second-Order Algorithm for Covariance Matrix Adaptation. In X. Yao at al., editors, *Parallel Problem Solving from Nature 8*, pages 182–191, Berlin, 2004. Springer.

[10] C. Igel, N. Hansen, and S. Roth. Covariance Matrix Adaptation for Multi-Objective Optimization. *Evolutionary Computation*, 15(1):1–28, 2007.

[11] 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.

[12] H.-G. Beyer and B. Sendhoff. Robust Optimization - A Comprehensive Survey. *Computer Methods in Applied Mechanics and Engineering*, 196(33–34):3190–3218, 2007.

[13] N. Hansen and S. Kern. Evaluating the CMA Evolution Strategy on Multimodal Test Functions. In X. Yao at al., editors, *Parallel Problem Solving from Nature 8*, pages 282–291, Berlin, 2004. Springer.

[14] B. Amstrup, G.J. Tóth, G. Szabó, H. Rabitz, and A. Lörincz. Genetic Algorithm with Migration on Topology Conserving Maps for Optimal Control of Quantum Systems. *J. Phys. Chem.*, 99:5206–5213, 1995.