Improved Step Size Adaptation for the MO-CMA-ES
Thomas Voß, Nikolaus Hansen, Christian Igel

To cite this version:

HAL Id: hal-00503251
https://hal.archives-ouvertes.fr/hal-00503251
Submitted on 18 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
ABSTRACT
The multi-objective covariance matrix adaptation evolution strategy (MO-CMA-ES) is an evolutionary algorithm for continuous vector-valued optimization. It combines indicator-based selection based on the contributing hypervolume with the efficient strategy parameter adaptation of the elitist covariance matrix adaptation evolution strategy (CMA-ES). Step sizes (i.e., mutation strengths) are adapted on individual-level using an improved implementation of the 1/5-th success rule. In the original MO-CMA-ES, a mutation is regarded as successful if the offspring ranks better than its parent in the elitist, rank-based selection procedure. In contrast, we propose to regard a mutation as successful if the offspring is selected into the next parental population. This criterion is easier to implement and reduces the computational complexity of the MO-CMA-ES, in particular of its steady-state variant. The new step size adaptation improves the performance of the MO-CMA-ES as shown empirically using a large set of benchmark functions. The new update scheme in general leads to larger step sizes and thereby counters premature convergence. The experiments comprise the first evaluation of the MO-CMA-ES for problems with more than two objectives.

Categories and Subject Descriptors
G.1.6 [Optimization]: Global Optimization; I.2.8 [Problem Solving, Control Methods, and Search]: Heuristic methods

General Terms
Algorithms, Performance

Keywords
multi-objective optimization, step size adaptation, covariance matrix adaptation, evolution strategy, MO-CMA-ES

ABSTRACT

In the MO-CMA-ES, step sizes (i.e., mutation strengths) are adapted on individual-level. The step size update procedure originates in the well-known 1/5-th rule originally presented by [18] and extended by [17]. If the success rate, that is, the fraction of successful mutations, is high, the step size is increased, otherwise it is decreased. In the original MO-CMA-ES, a mutation is regarded as successful if the resulting offspring is better than its parent. In this study, we propose to replace this criterion and to consider a mutation as being successful if the offspring becomes a member of the next parent population. We argue that this notion of success is easier to implement, computationally less demanding, and improves the performance of the MO-CMA-ES.

In the next section, we briefly review the MO-CMA-ES. In Sec. 3, we discuss our new notion of success for the step size adaptation. Then, we empirically evaluate the resulting algorithms. In this evaluation, the MO-CMA-ES is for the first time benchmarked on functions with more than two objectives. As a baseline, we consider a new variant of the NSGA-II, in which the crowding distance is replaced by the contributing hypervolume for sorting individuals at the same level of non-dominance.

2. THE MO-CMA-ES
In the following, we briefly outline the MO-CMA-ES according to [14, 16, 19], see Algorithm 1. For a detailed description and a performance evaluation on bi-objective benchmark functions we refer to [14, 21]. We consider objective functions $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $x \mapsto (f_1(x), \ldots, f_m(x))^T$.

In the MO-CMA-ES, a candidate solution $a_i^{(g)}$ in generation $g$ is a tuple $[x_i^{(g)}, \tilde{p}_{\text{succ},i}^{(g)}, \sigma_i^{(g)}, p_i^{(g)}(a_i^{(g)}), C_i^{(g)}]$, where $x_i^{(g)} \in \mathbb{R}^n$ is the current search point, $\tilde{p}_{\text{succ},i}^{(g)} \in [0, 1]$ is the smoothed success probability, $\sigma_i^{(g)} \in \mathbb{R}_+^n$ is the global step size, $p_i^{(g)}(a_i^{(g)}) \in \mathbb{R}$ is the cumulative evolution path, $C_i^{(g)} \in \mathbb{R}^{n \times n}$ is the covariance matrix of the search distribution. For an individual $a$ encoding search point $x$, we write $f(a)$ for $f(x)$ with a slight abuse of notation.
We first describe the general ranking procedure and summarize the other parts of the MO-CMA-ES. The MO-CMA-ES relies on the non-dominated sorting selection scheme [6]. As in the SMS-EMOA [2], the hypervolume-indicator serves as second-level sorting criterion to rank individuals at the same level of non-dominance. Let \( A \) be a population, and let \( a, a' \) two individuals in \( A \). Let the non-dominated solutions in \( A \) be denoted by \( \text{ndom}(A) = \{ a \mid 3a' \in A : a' < a \} \), where \( < \) denotes the Pareto-dominance relation. The elements in \( \text{ndom}(A) \) are assigned a level of non-dominance of 1. The other ranks of non-dominance are defined recursively by considering the set \( A \) without the solutions with lower ranks [6]. Formally, let \( \text{dom}_1(A) = A, \text{dom}_i(A) = \text{dom}_{i-1}(A) \setminus \text{ndom}(A) \), and \( \text{ndom}(A) = \text{dom}(\text{dom}_{i-1}(A)) \) for \( i \geq 1 \). For \( a \in A \) we define the level of non-dominance rank \((a, A)\) to be \( i \) if \( a \in \text{dom}_i(A) \).

The hypervolume measure or \( \mathcal{A} \)-metric was introduced in the domain of evolutionary multi-objective optimization (MOO) in [26]. It is defined as

\[
S_{\text{ref}}(A) = \mathcal{A} \left( \bigcup_{a \in A} f_1(a), f_2(a), \ldots, f_m(a) \right),
\]

with \( f_{ref} \in \mathbb{R}^m \) referring to an appropriately chosen reference point and \( \mathcal{A} \) the Lebesgue measure. The contributing hypervolume of a point \( a \in A' = \text{ndom}(A) \) is given by

\[
\Delta_{\mathcal{A}}(a, A') = S_{\text{ref}}(A') - S_{\text{ref}}(A' \setminus \{ a \}).
\]

Now we define the contribution rank \( \text{cont}(a, A') \) of \( a \). This is again done recursively. The element, say \( a \), with the smallest contributing hypervolume is assigned contribution rank 1. The next rank is assigned by considering \( A' \setminus \{ a \} \) etc. More precisely, let \( c_l(A') = \arg\min_{a \in A'} \Delta_{\mathcal{A}}(a, A') \) and

\[
c_i(A') = c_0 \left( A' \setminus \bigcup_{j=0}^{i-1} \{ c_j(A') \} \right)
\]

for \( i > 0 \). For \( a \in A' \) we define the contribution rank \( \text{cont}(a, A') \) to be \( i + 1 \) if \( a = c_i(A') \). In the ranking procedure ties are broken at random. We refer to the points \( \{ a \mid a = \arg\min_{a \in A} f_i(a), i = 1, \ldots, m \} \) as boundary elements of \( A \). The reference point \( f_{ref} \) is chosen in each iteration such that an individual with fitness \( f_{ref} \) would be dominated by all individuals in the current population and such that all boundary elements get the highest contribution ranks (i.e., the boundary elements are always selected). Such a reference point always exists, is easy to find and its exact choice does not matter for the MO-CMA-ES as long as the boundary elements get the highest ranks.

Finally, the following relation between individuals \( a, a' \in A \) is defined:

\[
a < a' \Leftrightarrow \text{rank}(a, A) < \text{rank}(a', A) \land
\text{cont}(a, \text{ndom}(A)) > \text{cont}(a', \text{ndom}(A))
\]

The success indicator \( \text{succ}_{Q^{(g)}}(a^{(g)}, a^{(g+1)}) \) in Algorithm 1 evaluates to one if the mutation that has created \( a^{(g+1)} \) is considered to be successful and to zero otherwise, see Sec. 3.

In each generation, \( \lambda \) offspring individuals are sampled (lines 3–7). If \( \lambda \) does not equal \( \mu \) (e.g., in case of the steady-state MO-CMA-ES), the parent is chosen uniformly at random from the set of non-dominated individuals \( \text{ndom}(Q^{(g)}) \) (line 4a). Otherwise, one offspring individual is created from every parent individual (line 4b). Thereafter, the strategy parameters of parent and offspring individuals are adapted (lines 9–18). The decision whether a new candidate solution is better than its parent is made in the context of the population \( Q^{(g)} \) of parent and offspring individuals subject to the indicator-based selection strategy implemented in the algorithm. The step sizes and the covariance matrix of the offspring individuals are updated (lines 9–16). Subsequently, the step sizes \( \sigma_{i,k}^{(g)} \) of the parent individuals \( a_{i,k} \) are adapted (line 17). Finally, the new parent population is selected from the set of parent and offspring individuals according to the indicator-based selection scheme (line 20).

Here, \( Q^{(g)} \) denotes the \( g \)-th best individual in \( Q^{(g)} \) ranked by non-dominated sorting and the contributing hypervolume according to (Eq. 4).

When in this study the MO-CMA-ES is applied to a bench-

\[
\text{Algorithm 1: } (\mu + \lambda)-\text{MO-CMA-ES}
\]

1. \( g \leftarrow 0 \), initialize parent population \( Q^{(0)} \);  
2. repeat  
3. \( \text{for } k = 1, \ldots, \lambda \) do  
4a. \( i_k \leftarrow \mathcal{U}(1,|\text{ndom}(Q^{(g)})|) \);  
4b. \( i_k \leftarrow k \);  
5. \( a_{i,k}^{(g+1)} \leftarrow a_{i_k}^{(g)} \);  
6. \( x_{i,k}^{(g+1)} \sim x_{i_k}^{(g)} + \sigma_{i_k}^{(g)} N(0, \sigma_{i_k}^{(g)}); \)  
7. \( Q^{(g)} \leftarrow Q^{(g)} \cup \{ a_{i,k}^{(g+1)} \} \);  
4. \( \text{for } k = 1, \ldots, \lambda \) do  
9. \( \tilde{p}_{i,k}^{(g+1)} \leftarrow (1-c_p)p_{i,k}^{(g+1)} + c_p \text{succ}_{Q^{(g)}}(a_{i,k}^{(g)}, a_{i,k}^{(g+1)}); \)  
10. \( \sigma_{i,k}^{(g+1)} \leftarrow \sigma_{i,k}^{(g)} \exp \left( \frac{p_{i,k}^{(g+1)} - \text{target}}{1 - \text{target}} \right); \)  
11. if \( \tilde{p}_{i,k}^{(g+1)} < p_{\text{break}} \) then  
12. \( \sigma_{i,k}^{(g+1)} = (1-c_c)\sigma_{i,k}^{(g+1)} + \sqrt{c_c(2-cc)} \frac{\sigma_{i,k}^{(g+1)} - \sigma_{i,k}^{(g)}}{\sigma_{i,k}^{(g)}}; \)  
13. \( C_{i,k}^{(g+1)} = (1-c_c)C_{i,k}^{(g+1)} + c_c(2-cc)C_{i,k}^{(g)}(g); \)  
14. else \( \tilde{p}_{i,k}^{(g+1)} = (1-c_c)\tilde{p}_{i,k}^{(g+1)}; \)  
15. \( \sigma_{i,k}^{(g+1)} = \exp \left( \frac{\tilde{p}_{i,k}^{(g+1)} - \text{target}}{1 - \text{target}} \right); \)  
16. \( \sigma_{i,k}^{(g+1)} = \exp \left( \frac{\tilde{p}_{i,k}^{(g+1)} - \text{target}}{1 - \text{target}} \right); \)  
17. \( g \leftarrow g + 1 \);  
18. \( Q^{(g)} \leftarrow \{ Q_{i}^{(g)} \mid 1 \leq i \leq \mu \}; \)  
20. until stopping criterion is met ;
mark problem $f$ with box constraints, we consider a penalized fitness function
\[
\tilde{f}(\mathbf{x}) = f(\text{feasible}(\mathbf{x})) + \alpha \| \mathbf{x} - \text{feasible}(\mathbf{x}) \|_2^2,
\] (5)
where feasible($\mathbf{x}$) returns the closest feasible point to $\mathbf{x}$ w.r.t. the $L_1$-norm.

The (external) strategy parameters are the population size, initial global step size, target success probability $p_{\text{target}}$, step size damping $d$, learning rate averaging parameter $c_p$, cumulation time horizon parameter $c_t$, and covariance matrix learning rate $c_{\text{cov}}$. Default values as given in [14] and used in this paper are $d = 1 + n/2$, $p_{\text{target}} = (5 + \sqrt{17/2})^{-1}$, $c_p = p_{\text{target}} / (2 + p_{\text{target}})$, $c_t = 2/(n + 2)$, $c_{\text{cov}} = 2/(n^2 + 6)$, and $p_{\text{thresh}} = 0.44$. In the constraint handling we set $\alpha = 10^{-6}$.

The initial global step sizes $\sigma_i^{(0)}$ are set dependent on the problem (e.g., in the case of box constraints, see below, with $x_i^* - x_i = x_i^* - x_j$ for $1 \leq i, j \leq n$ to $0.6 \cdot (x_i^* - x_i)$).

### 2.1 Step Size Update Procedure

The focus of this study is the step size adaptation, which is described in more detail in the following. After sampling the new candidate solutions, the step size of parent $a$ is updated based on the smoothed success rate (different notions of success are discussed in Sec. 3)
\[
p_{\text{succ}} = (1 - cp) p_{\text{succ}} + cp \text{ succ}_{\text{Q}}(a, a'),
\] (6)
with a learning rate $cp$ ($0 < cp \leq 1$) according to
\[
\sigma = \sigma \cdot \exp \left( \frac{1}{d} \frac{p_{\text{succ}} - p_{\text{target}}}{1 - p_{\text{succ}}} \right). 
\] (7)

The update rule is rooted in the 1/5-success-rule proposed by [18] and is an extension from the rule proposed by [17]. It implements the well-known heuristic that the step size should be increased if the success rate (i.e., the fraction of offspring better than the parent) is high, and the step size should be decreased if the success rate is low. The rule is reflected in the argument of the exponential function. For $p_{\text{succ}} > p_{\text{target}}$ the argument is greater than zero and the step size increases; for $p_{\text{succ}} < p_{\text{target}}$ the argument is smaller than zero and the step size decreases; for $p_{\text{succ}} = p_{\text{target}}$ the argument becomes zero and no change of $\sigma$ takes place.

The argument to the exponential function is always smaller than $1/d$ and larger than $-1/d$ if $p_{\text{target}} < 0.5$ (a necessary assumption). Therefore, the damping parameter $d$ controls the rate of the step size adaptation. Using $p_{\text{succ}}$ instead of $\text{succ}_{\text{Q}}(a_i^{(g)}, a_i^{(g+1)})$ primarily smooths the single step size changes.

### 2.2 Hypervolume Computations

Computing the contributing hypervolume (see Eq. 2) is computationally demanding [24, 23, 3, 1]. In fact, calculating the contributing hypervolume is $\#P$ hard (see [5]), where $\#P$ is the analog of NP for counting problems (see [20]).

Thus, in an efficient implementation of the MO-CMA-ES this should be done as rarely as possible. For selection, it is not necessary to rank all $\mu + \lambda$ individuals. It is sufficient to determine the $\lambda$ worst. In addition, only if there is a need to pick $m \leq \lambda$ individuals from the same level of non-dominance, say from $\text{ndom}(A)$ with $k = |\text{ndom}(A)|$ and $m < k$, we need to determine $k - m$ times the individual with the least hypervolume contribution. Because in each of these rounds the cardinality of the set we have to consider is reduced by one, the contributing hypervolume (Eq. 2) needs to be computed $\sum_{i=k-m}^{k} (2n^2 + m^2) + 2k - m$ times. For the special case of the steady-state MO-CMA-ES with $\lambda = 1$, we therefore need to compute a contributing hypervolume for at most $\mu + 1$ points, because we just have to determine a single individual to discard. However, in the original MO-CMA-ES additional contributing hypervolume computations are required as discussed in the following section.

### 3. NEW NOTION OF SUCCESS FOR STEP SIZE ADAPTATION

For the success rule based adaptation of the step sizes, we need a notion of what is considered to be a successful mutation. The choice of the success criterion is crucial for the step size update procedure (see Eq. 7). It directly affects (the smoothed) success rate associated with the individual that in turn influences the update of the global step-size $\sigma_i$ as well as the update of the covariance matrix $C_i$.

In general, a conservative notion of success results in a low success rate and thereby in a decrease of the step sizes $\sigma_i$. If the criterion is too conservative, the convergence rate of the MO-CMA-ES slows down. In contrast, an optimistic notion of success results in a higher success rate and larger steps.

In the (1+1)-CMA-ES defining the notion of success is unambiguous. A mutation has been successful if the parent is replaced by the offspring. This can be expressed in two ways. A mutation has been successful if (i) the offspring is better than the parent, (ii) the offspring is selected. While these criteria are equivalent in the (1+1)-CMA-ES, they lead to different success indicators succ_{Q} in the MO-CMA-ES.

### 3.1 Parent-Based Notion of Success

In the original MO-CMA-ES, a mutation was considered as being successful if the offspring is better than the parent. This requires a direct comparison of an offspring individual $a_i^{(g+1)}$ with its parent $a_i^{(g)}$ w.r.t. the level of non-dominance and the contribution rank. Thus, we have
\[
succ_{Q}(a_i^{(g)}, a_i^{(g+1)}) = \begin{cases} 1 \text{ if } a_i^{(g+1)} \prec a_i^{(g)} \\ 0 \text{ otherwise} \end{cases}. 
\] (8)

This direct comparison of parent and offspring may require more contributing hypervolume computations than the environmental selection procedure (see Sec. 2.2) if we use exactly the same ranking method as described in Sec. 2. To make parent and offspring are selected and have the same level of non-dominance (which frequently happens in real-valued multi-objective optimization after the first generations), additional hypervolume computations are required to determine whether the parent or the offspring rank higher.

### 3.2 Population-Based Notion of Success

We propose a simpler, at least as intuitive notion of success. An offspring individual $a_i^{(g+1)}$ is considered successful if it is selected for the next parent population $P^{(g+1)} = \{Q_i^{(g+1)} | 1 \leq i \leq \mu\}$:
\[
succ_{Q}^{P}(a_i^{(g)}, a_i^{(g+1)}) = \begin{cases} 1 \text{ if } a_i^{(g+1)} \in Q_i^{(g+1)} \\ 0 \text{ otherwise} \end{cases}. 
\] (9)
This criterion is strictly more optimistic than the parent-based notion of success in the sense that
\[
\text{succ}_{Q(s)}^{f,g} \left( a_i^{(g)}, a_i^{(g+1)} \right) = 1 \Rightarrow \text{succ}_{Q(s)}^{g,n} \left( a_i^{(g)}, a_i^{(n+1)} \right) = 1,
\]
for all selected individuals requiring a step size update.

No additional (contributing) hypervolume computations are needed in addition to those needed for selection as described in Sec. 2.2. Especially for the steady state MO-CMA-ES, this considerably decreases the computation time for the strategy parameter update.

For the remainder of this work, the terms MO-CMA-ES$^f$ and MO-CMA-ES$^p$ refer to the MO-CMA-ES relying on the individual-based and population-based notion of success, respectively.

4. **EMPIRICAL EVALUATION**

This section presents a performance evaluation of the revised step size adaptation on a broad range of two- and three-objective benchmark functions. We compare the ($\mu$+$\lambda$)-MO-CMA-ES$^S$ and the ($\mu$+$\lambda$)-MO-CMA-ES$^P$ to the results of the original ($\mu$+$\lambda$)-MO-CMA-ES$^S$ and the ($\mu$+$\lambda$)-MO-CMA-ES$^P$. The results of both algorithms are compared to results of a "new" variant of the well-known NSGA-II [6]. Because it is well-known that the MO-CMA-ES in general outperforms the standard NSGA-II (e.g., see [16]) for real-valued optimization, we replaced the second-level sorting criterion in the NSGA-II and use the contributing hypervolume (as in the MO-CMA-ES) instead of the crowding distance.

The resulting algorithm is a non-steady-state version of the SMS-EMOA [2]. All experiments have been conducted using the Shark machine learning library [13].

4.1 **Experimental Setup**

We compare the algorithms on several classes of benchmark functions. The bi-criteria constrained benchmark functions ZDT1–4 and ZDT6 (see [25]) and their rotated variants IHR1–4 and IHR6 (see [14]) have been chosen for the performance evaluation. Additionally, the set of bi-objective test problems has been augmented by the unconstrained and bi-criteria constrained benchmark functions. The bi-criteria constrained benchmark functions ZDT1–4 and ZDT6 (see [25]) and their rotated variants are needed in addition to those needed for selection as described in Sec. 2.2. Especially for the steady state MO-CMA-ES, this considerably decreases the computation time for the strategy parameter update.

We defined a new class of test functions based on ELLI1 (see [16]), which is scalable to an arbitrary number of objectives. In the case of three objectives, the seven constrained functions DTLZ1–7 (see [8]) have been chosen.

We consider the unary hypervolume-indicator as performance measure [26, 27]. We want to compare $k = 5$ algorithms on a particular optimization problem $f$ after $g$ fitness evaluations and we assume that we have conducted $t$ trials. We consider the non-dominated individuals of the union of all $k \cdot t$ populations after $g$ evaluations. These individuals make up the reference set $R$. The upper bounds of the reference set of the respective fitness function translated by $\tau$ serve as the reference point for the calculation of the unary hypervolume-indicator.

Several ways to compare multiple direct search heuristics on multiple objective functions have been proposed in the literature [4, 9]. Such a statistical comparison is not straightforward, because one has to account for multiple testing. For the overall evaluation of the algorithms, we follow the recommendation in [9] and use non-parametric statistical tests in a step-wise procedure. For each fitness function, we rank the algorithms and then compute the average ranks. Then, we apply a (Friedman) test to check whether the ranks are different from the mean rank. If so, we determine ad hoc whether two algorithms differ by pairwise comparison (using Bergmann-Hommel’s dynamic procedure). We fixed a significance level of $p = 0.001$. For a detailed description of the test procedure we refer to the literature [8, 9, 10]. We rely on the open source software supplied by Garcia and Herrera [9] in our evaluation. In addition, we compare the results on the individual benchmark functions using a standard two-sided Wilcoxon rank sum test ($p < 0.001$). All results highlighted in the following sections are statistically significant, if not stated otherwise.

4.3 **Results**

The results of the performance evaluation after 25,000 and 50,000 fitness function evaluations are presented in Tables 1 and 2. In the overall comparison, the ($\mu$+$\lambda$)-MO-CMA-ES$^S$ and the ($\mu$+$\lambda$)-MO-CMA-ES$^P$ performed significantly better than the ($\mu$+$\lambda$)-MO-CMA-ES$^S$, ($\mu$+$\lambda$)-MO-CMA-ES$^P$ and the NSGA-II across the set of benchmark functions. The steady-state MO-CMA-ES with the new step-size update is the best of the five algorithms.

When looking at the single benchmark functions, the new ($\mu$+$\lambda$)-MO-CMA-ES$^S$ and the new ($\mu$+$\lambda$)-MO-CMA-ES$^P$ performed always better than their counterparts with the original step size adaptation (in one single case the difference is not significant at the level $p < 0.001$). Further, the MO-CMA-ES outperformed NSGA-II across the set of benchmark functions, except for ZDT4, IHR4, and DTLZ1. The latter three functions are multi-modal and it is well-known that the elitist variant of the (MO-)CMA-ES suffers from tetrahedron). The function is called GELLI $m$, where the superscript $m$ indicates the number of objectives.

The default search space dimension for constrained and non-rotated benchmark functions has been chosen to be 30. In case of rotated benchmark functions, the search space dimensions has been chosen to be 10. The number of parent and offspring individuals has been set to $\mu = \lambda = 100$. We conducted 25 independent trials with 100000 fitness function evaluations each. We sampled the performance of the algorithms after every 500th fitness function evaluation and carried out the statistical evaluation after 25,000 and 50,000 fitness function evaluations.
convergence into suboptimal local optima on multi-modal fitness landscapes (see [14]). The new step size adaptation attenuates this problem. The enhanced performance can be attributed to the larger global step sizes realized by the new variant (see Fig. 1 and Fig. 2) which in turn prevents both algorithms from getting stuck in local optima early. Nevertheless, the NSGA-II still outperforms all MO-CMA-ES variants on these fitness functions.

The good performance of both the \((\mu+\lambda)\text{-MO-CMA-ES}^P\) as well as its steady-state variant \((\mu+1)\text{-MO-CMA-ES}^P\) show that the new notion of success clearly improves the convergence properties of the algorithm.

5. CONCLUSIONS

We presented a new step size adaptation procedure for the MO-CMA-ES that improves the convergence speed and at the same time reduces the risk of convergence into suboptimal local optima. Additionally, the new update scheme lowers the computational complexity of both the generational MO-CMA-ES as well as its steady-state variant considerably by reducing the number of required computations of the contributing hypervolume. Our experiments showed the significantly improved performance of the new approach.

As baseline for the empirical evaluation, we considered a new variant of the NSGA-II relying on the hypervolume indicator as second-level sorting criterion. This comparison demonstrated that the superior performance of the MO-CMA-ES is not only due to the selection procedure but that the powerful strategy parameter adaptation in the CMA-ES plays a major role. For the first time, it was shown that the MO-CMA-ES outperforms the NSGA-II also in the case of more than two objectives (see Fig. 3). In summary, we strongly recommend the new update rule for the MO-CMA-ES.

In future work, additional notions of success will be considered. For instance, one could check whether the offspring individual is at least at the same level of non-dominance as the parent individual. Further, we will study the impact of the new step size update procedure on the performance of the MO-CMA-ES with recombination of strategy parameters presented in [22].

Acknowledgments

CI acknowledges support from the German Federal Ministry of Education and Research within the National Network Computational Neuroscience under grant number 01GQ0951.

6. REFERENCES

Figure 2: Evolution of the absolute hypervolume (left) and of the global step size (right) for the fitness function IHR4.

Figure 3: Evolution of the absolute hypervolume for the fitness functions DTLZ2, DTLZ3, DTLZ6, and DTLZ7 having three objectives.
Two-objective functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Value</th>
<th>Value</th>
<th>Value</th>
<th>Value</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZDT1</td>
<td>6.04312 (^{V})</td>
<td>6.08654 (^{III,IV})</td>
<td>6.06543 (^{V})</td>
<td>(6.08921^{III,III,IV})</td>
<td>6.04301</td>
</tr>
<tr>
<td>ZDT2</td>
<td>4.33039 (^{V})</td>
<td>4.33181 (^{III,IV})</td>
<td>4.33092 (^{V})</td>
<td>(4.33201^{III,III,IV})</td>
<td>4.31906</td>
</tr>
<tr>
<td>ZDT3</td>
<td>3.89724 (^{V})</td>
<td>3.89854 (^{III,IV})</td>
<td>3.89755 (^{V})</td>
<td>(4.89991^{III,III,IV})</td>
<td>3.89696</td>
</tr>
<tr>
<td>ZDT4</td>
<td>5.7355682 (^{III})</td>
<td>6.9222784 (^{III,IV})</td>
<td>5.0322781</td>
<td>(5.33372^{III})</td>
<td>7.92227 (^{III,III,IV})</td>
</tr>
<tr>
<td>ZDT6</td>
<td>9.14021 (^{V})</td>
<td>9.14039 (^{III,IV})</td>
<td>9.14029 (^{V})</td>
<td>(9.14052^{III,III,IV})</td>
<td>8.99023</td>
</tr>
<tr>
<td>IHR1</td>
<td>10.00087 (^{V})</td>
<td>10.00951 (^{III,IV})</td>
<td>10.00106 (^{V})</td>
<td>(10.00099^{III,III,IV})</td>
<td>9.35870</td>
</tr>
<tr>
<td>IHR2</td>
<td>21.35308 (^{V})</td>
<td>22.46712 (^{III,IV})</td>
<td>22.00089 (^{V})</td>
<td>(22.55552^{III,III,IV})</td>
<td>19.26374</td>
</tr>
<tr>
<td>IHR3</td>
<td>11.01029 (^{V})</td>
<td>11.91461 (^{III,IV})</td>
<td>11.50321 (^{V})</td>
<td>(11.99028^{III,III,IV})</td>
<td>10.90063</td>
</tr>
<tr>
<td>IHR4</td>
<td>9.63114 (^{III,IV})</td>
<td>12.00649 (^{III,IV})</td>
<td>7.40903</td>
<td>(8.93219^{III})</td>
<td>16.29601 (^{III,III,IV})</td>
</tr>
<tr>
<td>IHR6</td>
<td>7.49222 (^{V})</td>
<td>8.30291 (^{III,IV})</td>
<td>7.51206 (^{V})</td>
<td>(8.44444^{III,III,IV})</td>
<td>7.04526</td>
</tr>
<tr>
<td>ELLI1</td>
<td>2.28712 (^{V})</td>
<td>2.29654 (^{III,IV})</td>
<td>2.28990 (^{V})</td>
<td>(2.30657^{III,III,IV})</td>
<td>2.23001</td>
</tr>
<tr>
<td>ELLI2</td>
<td>2.90507 (^{V})</td>
<td>2.94182 (^{III,IV})</td>
<td>2.91999 (^{V})</td>
<td>(2.96430^{III,III,IV})</td>
<td>2.84432</td>
</tr>
<tr>
<td>CIGTAB1</td>
<td>3.92059 (^{V})</td>
<td>3.94653 (^{III,IV})</td>
<td>3.93485 (^{V})</td>
<td>(4.13942^{III,III,IV})</td>
<td>3.91111</td>
</tr>
<tr>
<td>CIGTAB2</td>
<td>7.09623 (^{V})</td>
<td>7.09914 (^{III,IV})</td>
<td>7.09745 (^{V})</td>
<td>(7.10005^{III,III,IV})</td>
<td>6.82424</td>
</tr>
</tbody>
</table>

Three-objective functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Value</th>
<th>Value</th>
<th>Value</th>
<th>Value</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTLZ1</td>
<td>210.09418 (^{III,IV})</td>
<td>231.77529 (^{III,IV})</td>
<td>190.11114</td>
<td>(198.77843^{III})</td>
<td>362.99741 (^{III,III,IV})</td>
</tr>
<tr>
<td>DTLZ2</td>
<td>1.45493 (^{V})</td>
<td>1.51007 (^{III,IV})</td>
<td>1.45590 (^{V})</td>
<td>(1.52587^{III,III,IV})</td>
<td>1.44501</td>
</tr>
<tr>
<td>DTLZ3</td>
<td>3.46292 (^{V})</td>
<td>3.57601 (^{III,IV})</td>
<td>3.48666 (^{V})</td>
<td>(3.61829^{III,III,IV})</td>
<td>3.41438</td>
</tr>
<tr>
<td>DTLZ4</td>
<td>3.75548 (^{V})</td>
<td>3.87328 (^{III,IV})</td>
<td>3.80002 (^{V})</td>
<td>(3.91733^{III,III,IV})</td>
<td>3.71349</td>
</tr>
<tr>
<td>DTLZ5</td>
<td>1.74993 (^{V})</td>
<td>1.81004 (^{III,IV})</td>
<td>1.78524 (^{V})</td>
<td>(1.86991^{III,III,IV})</td>
<td>1.73421</td>
</tr>
<tr>
<td>DTLZ6</td>
<td>18.17184 (^{V})</td>
<td>18.35190 (^{III,IV})</td>
<td>18.32435 (^{V})</td>
<td>(18.45000^{III,III,IV})</td>
<td>18.04588</td>
</tr>
<tr>
<td>DTLZ7</td>
<td>6.25983 (^{V})</td>
<td>6.29887 (^{III,IV})</td>
<td>6.26779 (^{V})</td>
<td>(6.33241^{III,III,IV})</td>
<td>6.22765</td>
</tr>
<tr>
<td>GELLI</td>
<td>5.34226 (^{V})</td>
<td>5.82398 (^{III,IV})</td>
<td>5.42099 (^{V})</td>
<td>(6.02909^{III,III,IV})</td>
<td>5.04991</td>
</tr>
</tbody>
</table>

Table 1: Performance comparison of the (\(\mu+\lambda\))-MO-CMA-ES \(^{i}\), (\(\mu+\lambda\))-MO-CMA-ES \(^{p}\), (\(\mu+1\))-MO-CMA-ES \(^{i}\), (\(\mu+1\))-MO-CMA-ES \(^{p}\), and NSGA-II using the hypervolume indicator as second-level sorting criterion. The table shows the median of 25 trials after 25,000 generations of the hypervolume-indicator (the higher the better). The superscripts III,III,IV and V indicate whether the respective algorithm performs significantly better than the (\(\mu+1\))-MO-CMA-ES \(^{i}\), (\(\mu+1\))-MO-CMA-ES \(^{p}\), (\(\mu+1\))-MO-CMA-ES \(^{i}\), (\(\mu+1\))-MO-CMA-ES \(^{p}\), and NSGA-II, respectively (two-sided Wilcoxon rank sum test, \(p < 0.001\)). The best value in each row is marked bold.

[20] L. G. Valiant. The complexity of computing the
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ZDT2</td>
<td>$4.33332^V$</td>
<td>$4.33482^{III,IV}$</td>
<td>$4.433402^{I,V}$</td>
<td>$4.435502^{I,I,IV}$</td>
<td>$4.32001$</td>
</tr>
<tr>
<td>ZDT3</td>
<td>$3.99924^V$</td>
<td>$3.9989^{III,IV}$</td>
<td>$3.9993^{I,IV}$</td>
<td>$4.0211^{I,II,IV}$</td>
<td>$3.9991$</td>
</tr>
<tr>
<td>ZDT4</td>
<td>$5.96662^{III}$</td>
<td>$7.45762^{III,IV}$</td>
<td>$5.1009$</td>
<td>$5.7645^{IV}$</td>
<td>$8.1399^V$</td>
</tr>
<tr>
<td>ZDT6</td>
<td>$11.10091^{IV}$</td>
<td>$11.63232^{III,IV}$</td>
<td>$11.2983^{I,IV}$</td>
<td>$11.65779^{III}$</td>
<td>$11.09981$</td>
</tr>
<tr>
<td>IHR1</td>
<td>$11.4730^I$</td>
<td>$11.5002^{I,III,IV}$</td>
<td>$11.4930^{I,IV}$</td>
<td>$11.71329^{I,II,IV}$</td>
<td>$11.46899$</td>
</tr>
<tr>
<td>IHR3</td>
<td>$12.0002^V$</td>
<td>$12.0009^{I,IV}$</td>
<td>$12.0029^{I,IV}$</td>
<td>$12.0003^{I,II,IV}$</td>
<td>$12.00019$</td>
</tr>
<tr>
<td>IHR6</td>
<td>$8.9922^{V}$</td>
<td>$9.37221^{III,IV}$</td>
<td>$9.1120^{I,IV}$</td>
<td>$9.3991^{I,II,IV}$</td>
<td>$8.10921$</td>
</tr>
<tr>
<td>ELL1</td>
<td>$2.37769^V$</td>
<td>$2.37772^{III,IV}$</td>
<td>$2.3777^{V}$</td>
<td>$2.3779^{I,II,IV}$</td>
<td>$2.37664$</td>
</tr>
<tr>
<td>ELL2</td>
<td>$4.00016^V$</td>
<td>$4.00261^{III,IV}$</td>
<td>$4.00101^{V}$</td>
<td>$4.0041^{I,IV}$</td>
<td>$3.90004$</td>
</tr>
<tr>
<td>CIGTA1</td>
<td>$4.32756^{V}$</td>
<td>$4.3286^{III,IV}$</td>
<td>$4.3276^{I,IV}$</td>
<td>$4.3306^{IV}$</td>
<td>$4.10921$</td>
</tr>
<tr>
<td>CIGTA2</td>
<td>$7.5764^{V}$</td>
<td>$7.90036^{III,IV}$</td>
<td>$7.8764^{IV}$</td>
<td>$8.1091^{I,II,IV}$</td>
<td>$7.20192$</td>
</tr>
</tbody>
</table>

Table 2: Performance comparison of the ($\mu$+$\lambda$)-MO-CMA-ES$^I$, ($\mu$+$\lambda$)-MO-CMA-ES$^V$, ($\mu$+$\lambda$)-MO-CMA-ES$^F$, and NSGA-II using the hypervolume indicator as second-level sorting criterion. The table shows the median of 25 trials after 50,000 generations of the hypervolume-indicator (the higher the better). The superscripts I,II,III,IV and V indicate whether the respective algorithm performs significantly better than the ($\mu$+$\lambda$)-MO-CMA-ES$^I$, ($\mu$+$\lambda$)-MO-CMA-ES$^F$, ($\mu$+$\lambda$)-MO-CMA-ES$^V$, and NSGA-II, respectively. The best value in each row is marked bold.


