Improved step size adaptation for the MO-CMA-ES

3 authors, including:

Thomas Voß

Ruhr-Universität Bochum

9 PUBLICATIONS   **143** CITATIONS

SEE PROFILE

Nikolaus Hansen

National Institute for Research in Computer Science and Control

180 PUBLICATIONS   **15,888** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Music Classification on Personal Mobile Devices View project

Project   New adaptive methods for constrained multi objective CMA-ES View project

Improved Step Size Adaptation for the MO-CMA-ES

Thomas Voß
Institut für Neuroinformatik
Ruhr-Universität Bochum
44780 Bochum, Germany
thomas.voss@ini.rub.de

Nikolaus Hansen
Université de Paris-Sud
Centre de recherche INRIA
Saclay – Île-de-France
F-91405 Orsay Cedex, France
hansen@lri.fr

Christian Igel
Institut für Neuroinformatik
Ruhr-Universität Bochum
44780 Bochum, Germany
christian.igel@ini.rub.de

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 counteracts 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

1. INTRODUCTION

The multi-objective covariance matrix adaptation evolution strategy (MO-CMA-ES, [14, 16, 19]) is an extension of the CMA-ES [12, 11] for real-valued multi-objective optimization. It combines the mutation and strategy adaptation of the (1+1)-CMA-ES [14, 15, 19] with a multi-objective selection procedure based on non-dominated sorting [6] and the contributing hypervolume [2] acting on a population of individuals.

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 $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m, \boldsymbol{x} \mapsto (f_1(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x}))^{\mathrm{T}}$. In the MO-CMA-ES, a candidate solution $a_i^{(g)}$ in generation g is a tuple $\left[\boldsymbol{x}_i^{(g)}, \bar{p}_{\mathrm{succ},i}^{(g)}, \sigma_i^{(g)}, \boldsymbol{p}_{i,c}^{(g)}, \boldsymbol{C}_i^{(g)} \right]$, where $\boldsymbol{x}_i^{(g)} \in \mathbb{R}^n$ is the current search point, $\bar{p}_{\mathrm{succ},i}^{(g)} \in [0, 1]$ is the smoothed success probability, $\sigma_i^{(g)} \in \mathbb{R}_0^+$ is the global step size, $\boldsymbol{p}_{i,c}^{(g)} \in \mathbb{R}^n$ is the cumulative evolution path, $\boldsymbol{C}_i^{(g)} \in \mathbb{R}^{n \times n}$ is the covariance matrix of the search distribution. For an individual a encoding search point $\boldsymbol{x}$, we write $\boldsymbol{f}(a)$ for $\boldsymbol{f}(\boldsymbol{x})$ 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' be two individuals in A. Let the non-dominated solutions in A be denoted by $\mathrm{ndom}(A) = \{a \in A \mid \nexists a' \in A : a' \prec a\}$, where $\prec$ denotes the Pareto-dominance relation. The elements in $\mathrm{ndom}(A)$ 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 $\mathrm{dom}_0(A) = A, \mathrm{dom}_l(A) = \mathrm{dom}_{l-1}(A) \setminus \mathrm{ndom}_l(A)$, and $\mathrm{ndom}_l(A) = \mathrm{ndom}(\mathrm{dom}_{l-1}(A))$ for $l \geq 1$. For $a \in A$ we define the level of non-dominance $\mathrm{rank}(a, A)$ to be i iff $a \in \mathrm{ndom}_i(A)$.

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

$$\mathcal{S}_{\boldsymbol{f}^{\mathrm{ref}}}(A) = \Lambda \left(\bigcup_{a \in A} \left[f_1(a), f_1^{\mathrm{ref}} \right] \times \cdots \times \left[f_m(a), f_m^{\mathrm{ref}} \right] \right) , \tag{1}$$

with $\boldsymbol{f}^{\mathrm{ref}} \in \mathbb{R}^m$ referring to an appropriately chosen reference point and $\Lambda(\cdot)$ being the Lebesgue measure. The contributing hypervolume of a point $a \in A' = \mathrm{ndom}(A)$ is given by

$$\Delta_{\mathcal{S}}(a, A') = \mathcal{S}_{\boldsymbol{f}^{\mathrm{ref}}}(A') - \mathcal{S}_{\boldsymbol{f}^{\mathrm{ref}}}(A' \setminus \{a\}) . \tag{2}$$

Now we define the *contribution rank* $\mathrm{cont}(a, A')$ 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 $c_0(A') = \mathrm{argmin}_{a \in A'} \Delta_{\mathcal{S}}(a, A')$ and

$$c_i(A') = c_0 \left(A' \setminus \bigcup_{j=0}^{i-1} \left\{ c_j(A') \right\} \right) \tag{3}$$

for $i > 0$. For $a \in A'$ we define the contribution rank $\mathrm{cont}(a, A')$ to be $i + 1$ iff $a = c_i(A')$. In the ranking procedure ties are broken at random. We refer to the points $\{a \mid a = \mathrm{argmin}_{a \in A} f_i(a), i = 1, \ldots, m\}$ as boundary elements of A. The reference point $\boldsymbol{f}^{\mathrm{ref}}$ is chosen in each iteration such that an individual with fitness $\boldsymbol{f}^{\mathrm{ref}}$ 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:

$$a \prec_A a' \Leftrightarrow \mathrm{rank}(a, A) < \mathrm{rank}(a', A) \vee$$
$$\left[\mathrm{rank}(a, A) = \mathrm{rank}(a', A) \wedge \right.$$
$$\left. \mathrm{cont}(a, \mathrm{ndom}(A)) > \mathrm{cont}(a', \mathrm{ndom}(A)) \right] \tag{4}$$

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

In each generation, λ offspring individuals are sampled (lines 3–7). If λ does not equal μ (e.g., in case of the steady-

---

Algorithm 1: $(\mu + \lambda)$-MO-CMA-ES

1 $g \leftarrow 0$, initialize parent population $Q^{(0)}$;
2 repeat
3   **for** $k = 1, \ldots, \lambda$ **do**
4a $i_k \leftarrow \mathcal{U} \left(1, | \mathrm{ndom} \left(Q^{(g)} \right) | \right)$;
4b $i_k \leftarrow k$;
5 $a_k'^{(g+1)} \leftarrow a_{i_k}^{(g)}$;
6 $\boldsymbol{x}_k'^{(g+1)} \sim \boldsymbol{x}_{i_k}^{(g)} + \sigma_{i_k}^{(g)} \mathcal{N} \left(\boldsymbol{0}, \boldsymbol{C}_{i_k}^{(g)} \right)$;
7 $Q^{(g)} \leftarrow Q^{(g)} \cup \left\{ a_k'^{(g+1)} \right\}$;
8   **for** $k = 1, \ldots, \lambda$ **do**
9 $\bar{p}_{\mathrm{succ},k}'^{(g+1)} \leftarrow$
     $(1 - c_p) \bar{p}_{\mathrm{succ},k}'^{(g+1)} + c_p \mathrm{succ}_{Q^{(g)}} \left(a_{i_k}^{(g)}, a_k'^{(g+1)} \right)$;
10 $\sigma_k'^{(g+1)} \leftarrow \sigma_k'^{(g+1)} \exp \left(\frac{1}{d} \frac{\bar{p}_{\mathrm{succ},k}'^{(g+1)} - p_{\mathrm{succ}}^{\mathrm{target}}}{1 - p_{\mathrm{succ}}^{\mathrm{target}}} \right)$;
11   **if** $\bar{p}_{\mathrm{succ},k}'^{(g+1)} < p_{\mathrm{thresh}}$ **then**
12 $\boldsymbol{p}_{c,k}'^{(g+1)} \leftarrow$
       $(1 - c_c) \boldsymbol{p}_{c,k}'^{(g+1)} + \sqrt{c_c(2 - c_c)} \frac{\boldsymbol{x}_k'^{(g+1)} - \boldsymbol{x}_{i_k}^{(g)}}{\sigma_{i_k}^{(g)}}$;
13 $\boldsymbol{C}_k'^{(g+1)} \leftarrow$
       $(1 - c_{\mathrm{cov}}) \boldsymbol{C}_k'^{(g+1)} + c_{\mathrm{cov}} \boldsymbol{p}_{c,k}'^{(g+1)} \boldsymbol{p}_{c,k}'^{(g+1)\mathrm{T}}$;
14   **else**
15 $\boldsymbol{p}_{c,k}'^{(g+1)} \leftarrow (1 - c_c) \boldsymbol{p}_{c,k}'^{(g+1)}$;
16 $\boldsymbol{C}_k'^{(g+1)} \leftarrow (1 - c_{\mathrm{cov}}) \boldsymbol{C}_k'^{(g+1)} +$
       $c_{\mathrm{cov}} \left(\boldsymbol{p}_{c,k}'^{(g+1)} \boldsymbol{p}_{c,k}'^{(g+1)\mathrm{T}} + c_c \left(2 - c_c \right) \boldsymbol{C}_k'^{(g+1)} \right)$;
17 $\bar{p}_{i_k}^{(g)} \leftarrow (1 - c_p) \bar{p}_{i_k}^{(g)} + c_p \mathrm{succ}_{Q^{(g)}} \left(a_{i_k}^{(g)}, a_k'^{(g+1)} \right)$;
18 $\sigma_{i_k}^{(g)} \leftarrow \sigma_{i_k}^{(g)} \exp \left(\frac{1}{d} \frac{\bar{p}_{\mathrm{succ},i_k}^{(g)} - p_{\mathrm{succ}}^{\mathrm{target}}}{1 - p_{\mathrm{succ}}^{\mathrm{target}}} \right)$;
19 $g \leftarrow g + 1$;
20 $Q^{(g)} \leftarrow \left\{ Q_{\prec:i}^{(g-1)} \big| 1 \leq i \leq \mu \right\}$;
   until stopping criterion is met ;

---

state MO-CMA-ES), the parent is chosen uniformly at random from the set of non-dominated individuals $\mathrm{ndom} \left(Q^{(g)} \right)$ (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 $a_{i_k}^{(g)}$ are adapted (line 17 and 18). 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, $Q_{\prec:i}^{(g)}$ denotes the ith 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-

mark problem $\boldsymbol{f}$ with box constraints, we consider a penalized fitness function

$$\boldsymbol{f}^{\text{penalty}}(\boldsymbol{x}) = \boldsymbol{f}(\text{feasible}(\boldsymbol{x})) + \alpha\|\boldsymbol{x} - \text{feasible}(\boldsymbol{x})\|_2^2 \ , \quad (5)$$

where feasible$(\boldsymbol{x})$ returns the closest feasible point to $\boldsymbol{x}$ w.r.t. the L_1-norm.

The (external) strategy parameters are the population size, initial global step size, target success probability $p_{\text{succ}}^{\text{target}}$, step size damping d, success rate averaging parameter c_p, cumulation time horizon parameter c_c, and covariance matrix learning rate c_{cov}. Default values as given in [14] and used in this paper are $d = 1 + n/2$, $p_{\text{succ}}^{\text{target}} = (5 + \sqrt{1/2})^{-1}$, $c_p = p_{\text{succ}}^{\text{target}}/(2 + p_{\text{succ}}^{\text{target}})$, $c_c = 2/(n + 2)$, $c_{\text{cov}} = 2/(n^2 + 6)$, 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 $x_i^u - x_i^l = x_j^u - x_j^l$ for $1 \leq i, j \leq n$ to $0.6 \cdot (x_1^u - x_1^l)$).

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)

$$\bar{p}_{\text{succ}} = (1 - c_p)\,\bar{p}_{\text{succ}} + c_p\, \text{succ}_Q\left(a, a'\right) \ , \quad (6)$$

with a learning rate c_p ($0 < c_p \leq 1$) according to

$$\sigma = \sigma \cdot \exp\left(\frac{1}{d}\frac{\bar{p}_{\text{succ}} - p_{\text{succ}}^{\text{target}}}{1 - p_{\text{succ}}^{\text{target}}}\right) \ . \quad (7)$$

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 $\bar{p}_{\text{succ}} > p_{\text{succ}}^{\text{target}}$ the argument is greater than zero and the step size increases; for $\bar{p}_{\text{succ}} < p_{\text{succ}}^{\text{target}}$ the argument is smaller than zero and the step size decreases; for $\bar{p}_{\text{succ}} = p_{\text{succ}}^{\text{target}}$ the argument becomes zero and no change of σ takes place.

The argument to the exponential function is always smaller than $1/d$ and larger than $-1/d$ if $p_{\text{succ}}^{\text{target}} < 0.5$ (a necessary assumption). Therefore, the damping parameter d controls the rate of the step size adaptation. Using $\bar{p}_{\text{succ}}$ instead of $\text{succ}_{Q^{(g)}}\left(a_i^{(g)}, a_i^{\prime(g+1)}\right)$ 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 $\#\mathbf{P}$-hard (see [5]), where $\#\mathbf{P}$ is the analog of $\mathbf{NP}$ 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 λ worst. In addition, only if there is a need to pick $m \leq \lambda$ individuals from the same level of non-dominance, say from $\text{ndom}_l(A)$ with $k = |\text{ndom}_l(A)|$ 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 $\sum_{i=k-m}^{k} i = \frac{2mk - m^2 + 2k - m}{2}$ 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 σ_i as well as the update of the covariance matrix $\boldsymbol{C}_i$.

In general, a conservative notion of success results in a low success rate and thereby in a decrease of the step sizes σ_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 $a_i^{\prime(g+1)}$ with its parent $a_i^{(g)}$ w.r.t. the level of non-dominance and the contribution rank. Thus, we have

$$\text{succ}_{Q^{(g)}}^I\left(a_i^{(g)}, a_i^{\prime(g+1)}\right) = \begin{cases} 1 & \text{if } a_i^{\prime(g+1)} \prec a_i^{(g)} \\ 0 & \text{otherwise} \end{cases} . \quad (8)$$

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.* If 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 $a_i^{\prime(g+1)}$ is considered successful if it is selected for the next parent population $P^{(g+1)} = \left\{Q_{\prec:i}^{(g)}|1 \leq i \leq \mu\right\}$:

$$\text{succ}_{Q^{(g)}}^P\left(a_i^{(g)}, a_i^{\prime(g+1)}\right) = \begin{cases} 1 & \text{if } a_i^{\prime(g+1)} \in Q^{(g+1)} \\ 0 & \text{otherwise} \end{cases} . \quad (9)$$

This criterion is strictly more optimistic than the parent-based notion of success in the sense that

$$\mathrm{succ}^{I}_{Q^{(g)}}\left(a_i^{(g)}, a_i'^{(g+1)}\right) = 1 \Rightarrow \mathrm{succ}^{P}_{Q^{(g)}}\left(a_i^{(g)}, a_i'^{(g+1)}\right) = 1 \; ,$$
(10)

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[I] 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[P] and the $(\mu+1)$-MO-CMA-ES[P] to the results of the original $(\mu+\lambda)$-MO-CMA-ES[I] and the $(\mu+1)$-MO-CMA-ES[I]. 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 rotated functions ELLI1, ELLI2, CIGTAB1 and CIGTAB2 (see [14]), with the distance of the optima of the single objectives set to the default value two. In the case of three objectives, the seven constrained functions DTLZ1–7 (see [7]) have been chosen.

We defined a new class of test functions based on ELLI1 (see [16]), which is scalable to an arbitrary number of objectives $m \leq n$, for this study. Let $\boldsymbol{O} \in \mathbb{R}^{n \times n}$ be an orthogonal matrix, $\boldsymbol{D} \in \mathbb{R}^{n \times n}$ be a diagonal matrix. Each candidate solution $\boldsymbol{x} \in \mathbb{R}^n$ is transformed by $\boldsymbol{v} = \boldsymbol{DOx}$. Moreover, a matrix $\boldsymbol{M} \in \mathbb{R}^{m \times n}$ defining the centers of the m objectives is needed. The objective functions are then given by $f_m(\boldsymbol{v}) = \frac{1}{\alpha^2 \cdot n} \cdot \sum_{i=1}^n (v_i - M_{mi})^2$. Varying $\boldsymbol{M}$ produces different benchmark functions. Here, $\boldsymbol{M} \in \mathbb{R}^{m \times n}$ is set to $\boldsymbol{M}_{ij} = 0$ if $j > m$, $\boldsymbol{M}_{ij} = \sqrt{\frac{d-1}{d}}$ if $i = j$, and $\boldsymbol{M}_{ij} = \frac{-1}{\sqrt{d \cdot (d-1)}}$ otherwise. Here, $d \in \mathbb{R}^{>0}$ and $\boldsymbol{D} \in \mathbb{R}^{n \times n}$ is set to $\boldsymbol{D}_{ij} = \alpha^{\frac{i-1}{n-1}}$ if $i = j$ and to 0 otherwise. For this study, the parameter α was set to 1000 and d was chosen as 2. The m optima of the single objective functions are placed on the unit sphere centered at the origin of Z such that they have maximum distance (i.e., they form a hyper-

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.

4.2 Statistical Evaluation

We consider the unary hypervolume-indicator as performance measure [26, 27]. We want to compare $k = 5$ algorithms on a particular optimization problem $\boldsymbol{f}$ 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 $\mathcal{R}$. The upper bounds of the reference set of the respective fitness function translated by $\boldsymbol{1}$ 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 García 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[P] and the $(\mu+1)$-MO-CMA-ES[P] performed significantly better than the $(\mu+\lambda)$-MO-CMA-ES[I], $(\mu+1)$-MO-CMA-ES[I] 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[P] and the new $(\mu+1)$-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
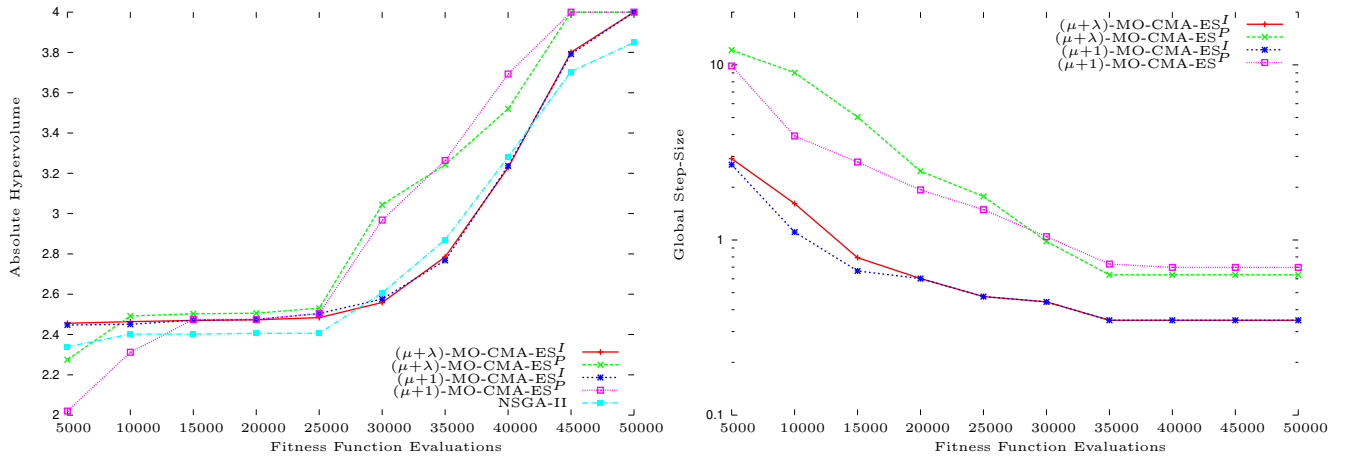
Figure 1: Evolution of the absolute hypervolume (left) and of the corresponding global step size (right) for the fitness function ELLI2 over the number of objective function evaluations. All plots refer to the medians over 25 trials.

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 $(\mu+\lambda)$-MO-CMA-ES^P as well as its steady-state variant $(\mu+1)$-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].

6. REFERENCES

[1] N. Beume. S-metric calculation by considering dominated hypervolume as Klee's measure problem. *Evolutionary Computation*, 17(4):477–492, 2009.

[2] N. Beume, B. Naujoks, and M. Emmerich. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181(3):1653–1669, 2007.

[3] N. Beume and G. Rudolph. Faster S-metric calculation by considering dominated hypervolume as Klee's measure problem. In *IASTED International Conference on Computational Intelligence*, pages 231–236. ACTA Press, 2006.

[4] S. Bleuler, M. Laumanns, L. Thiele, and E. Zitzler. PISA – A platform and programming language independent interface for search algorithms. In C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, editors, *Evolutionary Multi-Criterion Optimization (EMO 2003)*, volume 2632 of *LNCS*, pages 494 – 508. Springer-Verlag, 2003.

[5] K. Bringmann and T. Friedrich. Approximating the least hypervolume contributor: NP-Hard in General, But Fast in Practice. In M. Ehrgott, C. M. Fonseca, X. Gandibleux, J.-K. Hao, and M. Sevaux, editors, *EMO*, volume 5467 of *Lecture Notes in Computer Science*, pages 6–20. Springer, 2009.

[6] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6:182–197, 2002.

[7] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. Scalable multi-objective optimization test problems. In *Congress on Evolutionary Computation (CEC)*, pages 825–830. IEEE Press, 2002.

[8] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.

[9] S. Garcıa and F. Herrera. An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons. *Journal of Machine Learning Research*, 9:2677–2694, 2008.
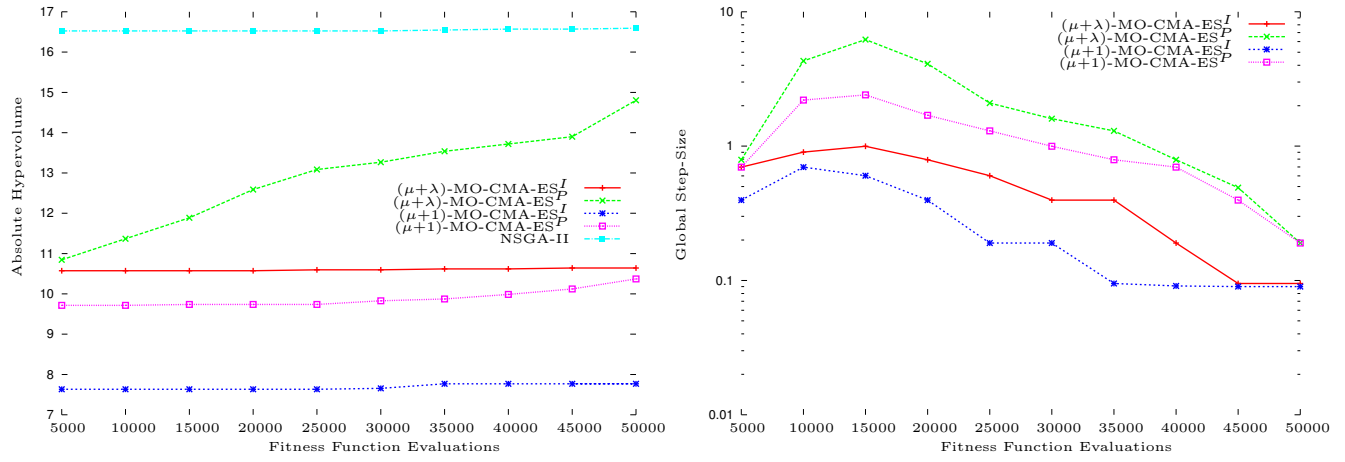
Figure 2: Evolution of the absolute hypervolume (left) and of the global step size (right) for the fitness function IHR4.
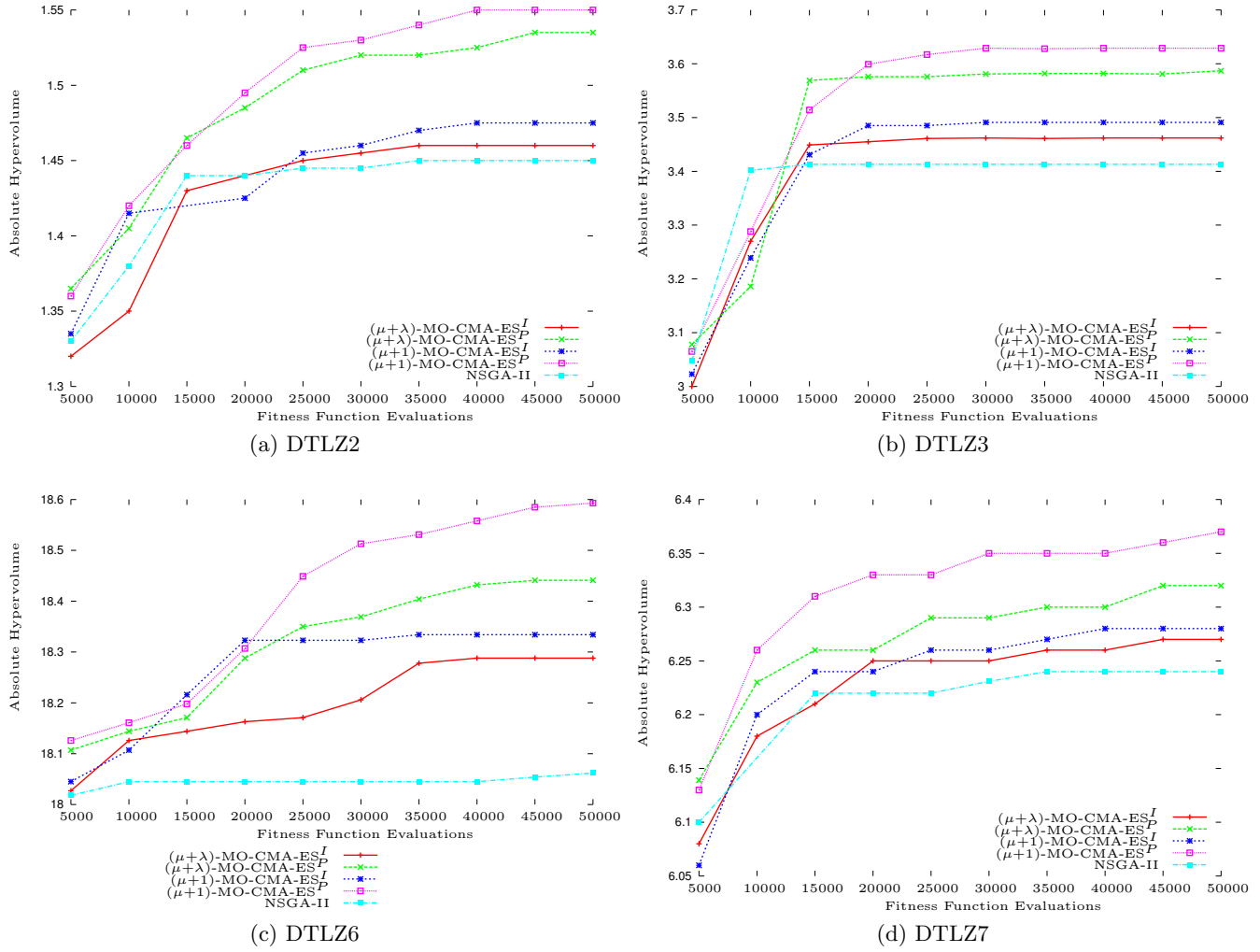

(a) DTLZ2

(b) DTLZ3

(c) DTLZ6

(d) DTLZ7

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

	$(\mu+\lambda)$-MO-CMA-ES^I	$(\mu+\lambda)$-MO-CMA-ES^P	$(\mu+1)$-MO-CMA-ES^I	$(\mu+1)$-MO-CMA-ES^P	NSGA-II
Two-objective functions					
ZDT1	6.04312^{V}	$6.08654^{\text{I,III,V}}$	$6.06543^{\text{I,V}}$	$\mathbf{6.08921}^{\text{I,II,III,V}}$	6.04301
ZDT2	4.33030^{V}	$4.33181^{\text{I,III,V}}$	$4.33092^{\text{I,V}}$	$\mathbf{4.33201}^{\text{I,II,III,V}}$	4.31906
ZDT3	3.89724^{V}	$3.89845^{\text{I,III,V}}$	$3.89755^{\text{I,V}}$	$\mathbf{4.89991}^{\text{I,II,III,V}}$	3.89696
ZDT4	5.7355682^{III}	$6.9222784^{\text{I,III,IV}}$	5.0322781	5.33372^{III}	$\mathbf{7.92227}^{\text{I,II,III,IV}}$
ZDT6	9.14021^{V}	$9.14039^{\text{I,III,V}}$	$9.14029^{\text{I,V}}$	$\mathbf{9.14052}^{\text{I,II,III,V}}$	8.99023
IHR1	10.00087^{V}	$10.00951^{\text{I,III,V}}$	$10.00100^{\text{I,V}}$	$\mathbf{10.00999}^{\text{I,II,III,V}}$	9.35870
IHR2	21.35308^{V}	$22.46712^{\text{I,III,V}}$	$22.00009^{\text{I,V}}$	$\mathbf{22.55552}^{\text{I,II,III,V}}$	19.26734
IHR3	11.01029^{V}	$11.94161^{\text{I,III,V}}$	$11.50321^{\text{I,V}}$	$\mathbf{11.99028}^{\text{I,II,III,V}}$	10.90063
IHR4	$9.63114^{\text{III,IV}}$	$12.00649^{\text{I,III,IV}}$	7.40903	8.93219^{III}	$\mathbf{16.29601}^{\text{I,II,III,IV}}$
IHR6	7.49222^{V}	$8.30291^{\text{I,III,V}}$	$7.51206^{\text{I,V}}$	$\mathbf{8.44444}^{\text{I,II,III,V}}$	7.04526
ELLI1	2.28712^{V}	$2.29654^{\text{I,III,V}}$	$2.28990^{\text{I,V}}$	$\mathbf{2.30657}^{\text{I,II,III,V}}$	2.23001
ELLI2	2.90507^{V}	$2.94182^{\text{I,III,V}}$	2.91999^{V}	$\mathbf{2.96430}^{\text{I,II,III,V}}$	2.84432
CIGTAB1	3.92059^{V}	$3.94653^{\text{I,III,V}}$	$3.93485^{\text{I,V}}$	$\mathbf{4.13942}^{\text{I,II,III,V}}$	3.91111
CIGTAB2	7.09623^{V}	$7.09914^{\text{I,III,V}}$	$7.09745^{\text{I,V}}$	$\mathbf{7.10005}^{\text{I,II,III,V}}$	6.82424
Three-objective functions					
DTLZ1	$210.09418^{\text{III,IV}}$	$231.77529^{\text{I,III,IV}}$	190.11114	198.77843^{III}	$\mathbf{362.99741}^{\text{I,II,III,IV}}$
DTLZ2	1.45493^{V}	$1.51007^{\text{I,III,V}}$	$1.45590^{\text{I,V}}$	$\mathbf{1.52587}^{\text{I,II,III,V}}$	1.44501
DTLZ3	3.46292^{V}	$3.57601^{\text{I,III,V}}$	$3.48666^{\text{I,V}}$	$\mathbf{3.61829}^{\text{I,II,III,V}}$	3.41438
DTLZ4	3.75548^{V}	$3.87132^{\text{I,III,V}}$	$3.80002^{\text{I,V}}$	$\mathbf{3.91733}^{\text{I,II,III,V}}$	3.71349
DTLZ5	1.74993^{V}	$1.81004^{\text{I,III,V}}$	$1.78524^{\text{I,V}}$	$\mathbf{1.86991}^{\text{I,II,III,V}}$	1.73421
DTLZ6	18.17184^{V}	$18.35190^{\text{I,III,V}}$	$18.32435^{\text{I,V}}$	$\mathbf{18.45000}^{\text{I,II,III,V}}$	18.04588
DTLZ7	6.25983^{V}	$6.29887^{\text{I,III,V}}$	$6.26779^{\text{I,V}}$	$\mathbf{6.33241}^{\text{I,II,III,V}}$	6.22765
GELLI^3	5.34226^{V}	$5.82398^{\text{I,III,V}}$	$5.42099^{\text{I,V}}$	$\mathbf{6.02909}^{\text{I,II,III,V}}$	5.04991

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 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^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.

[10] S. Garcıa, D. Molina, M. Lozano, and F. Herrera. A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: A case study on the CEC'2005 special session on real parameter optimization. *Journal of Heuristics*, 15:617–644, 2009.

[11] 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.

[12] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.

[13] C. Igel, T. Glasmachers, and V. Heidrich-Meisner. Shark. *Journal of Machine Learning Research*, 9:993–996, 2008.

[14] C. Igel, N. Hansen, and S. Roth. Covariance matrix adaptation for multi-objective optimization. *Evolutionary Computation*, 15(1):1–28, 2006.

[15] C. Igel, T. Suttorp, and N. Hansen. A computational efficient covariance matrix update and a (1+1)-CMA for evolution strategies. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2006)*, pages 453–460. ACM Press, 2006.

[16] C. Igel, T. Suttorp, and N. Hansen. Steady-state selection and efficient covariance matrix update in the multi-objective CMA-ES. In *Fourth International Conference on Evolutionary Multi-Criterion Optimization (EMO 2007)*, volume 4403 of *LNCS*. Springer-Verlag, 2007.

[17] S. Kern, S. D. Müller, N. Hansen, D. Büche, J. Ocenasek, and P. Koumoutsakos. Learning probability distributions in continuous evolutionary algorithms – a comparative review. *Natural Computing*, 3:77–112, 2004.

[18] I. Rechenberg. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution.* Frommann-Holzboog, 1973.

[19] T. Suttorp, N. Hansen, and C. Igel. Efficient covariance matrix update for variable metric evolution strategies. *Machine Learning*, 75(2):167–197, 2009.

[20] L. G. Valiant. The complexity of computing 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	NSGA-II
	Two-objective functions				
ZDT1	6.69909^{V}	$6.71001^{\text{I,III,V}}$	$6.69899^{\text{I,V}}$	$\mathbf{6.73491}^{\text{I,II,III,V}}$	6.69625
ZDT2	4.33332^{V}	$4.33482^{\text{I,III,V}}$	$4.433402^{\text{I,V}}$	$\mathbf{4.433502}^{\text{I,II,III,V}}$	4.32001
ZDT3	3.99924^{V}	$3.99989^{\text{I,III,V}}$	$3.99929^{\text{I,V}}$	$\mathbf{4.02111}^{\text{I,II,III,V}}$	3.99901
ZDT4	5.96642^{III}	$7.45576^{\text{I,III,IV}}$	5.10009	5.76452^{IV}	$\mathbf{8.13997}^{\text{I,II,III,IV}}$
ZDT6	11.10091^{V}	$11.63232^{\text{I,III,V}}$	$11.29834^{\text{I,V}}$	$\mathbf{11.65779}^{\text{I,II,III,V}}$	11.09981
IHR1	11.47301^{V}	$11.59002^{\text{I,III,V}}$	$11.49309^{\text{I,V}}$	$\mathbf{11.71329}^{\text{I,II,III,V}}$	11.46899
IHR2	21.49398^{V}	$22.51000^{\text{I,III,V}}$	$22.01000^{\text{I,V}}$	$\mathbf{22.72727}^{\text{I,II,III,V}}$	20.27754
IHR3	12.00024^{V}	$12.00091^{\text{I,III,V}}$	$12.00029^{\text{I,V}}$	$\mathbf{12.00093}^{\text{I,II,III,V}}$	12.00019
IHR4	$9.66304^{\text{III,IV}}$	$13.97065^{\text{I,III,IV}}$	7.50005	9.44219^{III}	$\mathbf{16.39299}^{\text{I,II,III,IV}}$
IHR6	8.99222^{V}	$9.37221^{\text{I,III,V}}$	$9.11209^{\text{I,V}}$	$\mathbf{9.39911}^{\text{I,II,III,V}}$	8.10921
ELLI1	2.37769^{V}	$2.37782^{\text{I,III,V}}$	$2.37779^{\text{I,V}}$	$\mathbf{2.37791}^{\text{I,II,III,V}}$	2.37664
ELLI2	4.00016^{V}	$4.00261^{\text{I,III,V}}$	4.00101^{III}	$\mathbf{4.00411}^{\text{I,II,III,V}}$	3.90004
CIGTAB1	4.32756^{V}	$4.32865^{\text{I,III,V}}$	$4.32765^{\text{I,V}}$	$\mathbf{4.33065}^{\text{I,II,III,V}}$	4.10921
CIGTAB2	7.57643^{V}	$7.90036^{\text{I,III,V}}$	$7.87643^{\text{I,V}}$	$\mathbf{8.10911}^{\text{I,II,III,V}}$	7.20192
	Three-objective functions				
DTLZ1	$214.29212^{\text{III,IV}}$	$233.72501^{\text{I,III,IV}}$	191.90762	198.77843^{III}	$\mathbf{366.21213}^{\text{I,II,III,IV}}$
DTLZ2	1.46973^{V}	$1.53591^{\text{I,III,V}}$	$1.47532^{\text{I,V}}$	$\mathbf{1.55009}^{\text{I,II,III,V}}$	1.45592
DTLZ3	3.46299^{V}	$3.58811^{\text{I,III,V}}$	$3.49291^{\text{I,V}}$	$\mathbf{3.66377}^{\text{I,II,III,V}}$	3.41494
DTLZ4	3.77945^{V}	$3.89059^{\text{I,III,V}}$	$3.82637^{\text{I,V}}$	$\mathbf{3.96828}^{\text{I,II,III,V}}$	3.73446
DTLZ5	1.75773^{V}	$1.83225^{\text{I,III,V}}$	$1.78524^{\text{I,V}}$	$\mathbf{1.86923}^{\text{I,II,III,V}}$	1.73587
DTLZ6	18.28844^{V}	$18.44184^{\text{I,III,V}}$	$18.33333^{\text{I,V}}$	$\mathbf{18.59409}^{\text{I,II,III,V}}$	18.06321
DTLZ7	6.27901^{V}	$6.32499^{\text{I,III,V}}$	$6.28969^{\text{I,V}}$	$\mathbf{6.37965}^{\text{I,II,III,V}}$	6.24907
GELLI^3	5.38871^{V}	$6.00012^{\text{I,III,V}}$	$5.46952^{\text{I,V}}$	$\mathbf{6.22111}^{\text{I,II,III,V}}$	5.33869

Table 2: 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 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^P, $(\mu+1)$-MO-CMA-ES^I, $(\mu+1)$-MO-CMA-ES^P, and NSGA-II, respectively. The best value in each row is marked bold.

permanent. *Theoretical Computer Science*, 8:189–201, 1979.

[21] T. Voß, N. Beume, G. Rudolph, and C. Igel. Scalarization versus indicator-based selection in multi-objective CMA evolution strategies. In *IEEE Congress on Evolutionary Computation 2008 (CEC 2008)*, pages 3041–3048. IEEE Press, 2008.

[22] T. Voß, N. Hansen, and C. Igel. Recombination for learning strategy parameters in the MO-CMA-ES. In M. Ehrgott, C. Fonseca, X. Gandibleux, J.-K. Hao, and M. Sevaux, editors, *Fifth International Conference on Evolutionary Multi-Criterion Optimization (EMO 2009)*, volume 5467 of *LNCS*. Springer-Verlag, 2009.

[23] L. While. A new analysis of the LebMeasure algorithm for calculating hypervolume. In C. A. Coello Coello, E. Zitzler, and A. Hernandez Aguirre, editors, *Third International Conference on Evolutionary Multi-Criterion Optimization (EMO 2005)*, volume 3410 of *LNCS*, pages 326–340. Springer-Verlag, 2005.

[24] E. Zitzler. Hypervolume metric calculation. Technical report, Computer Engineering and Networks

Laboratory, (TIK), Swiss Federal Institute of Technology (ETH) Zurich, 2001.

[25] E. Zitzler, K. Deb, and L. Thiele. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation*, 8(2):173–195, 2000.

[26] E. Zitzler and L. Thiele. Multiobjective optimization using evolutionary algorithms — a comparative case study. In A. E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel, editors, *Fifth International Conference on Parallel Problem Solving from Nature (PPSN-V)*, volume 1498 of *LNCS*, pages 292–301. Springer-Verlag, 1998.

[27] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. Grunert da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132, 2003.