Scalarization versus Indicator-based Selection in Multi-Objective CMA Evolution Strategies

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

Nicola Beume
Lehrstuhl für Algorithm Engineering
Technische Universität Dortmund
44221 Dortmund, Germany
nicola.beume@tu-dortmund.de

Günter Rudolph
Lehrstuhl für Algorithm Engineering
Technische Universität Dortmund
44221 Dortmund, Germany
guenter.rudolph@tu-dortmund.de

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

Abstract—**While scalarization approaches to multi-criteria optimization become infeasible in the case of many objectives, for few objectives the benefits of a population compared to a set of independent single-objective optimization trials on scalarized functions are not obvious.**

The multi-objective covariance matrix adaptation evolution strategy (MO-CMA-ES) is a powerful algorithm for real-valued multi-criteria optimization. This population-based approach combines mutation and strategy adaptation from the elitist CMA-ES with multi-objective selection.

We empirically compare the steady-state MO-CMA-ES with different scalarization algorithms, in which the elitist CMA-ES is used as single-objective optimizer. Although only bicriteria benchmark problems are considered, the MO-CMA-ES performs best in the overall comparison. However, if the scalarized problems have a structure that can easily exploited by the CMA-ES and that is less apparent in the vector-valued fitness function, the CMA-ES with scalarization outperforms the population-based approach.

I. Introduction

The goal of multi-objective optimization (MOO) can be defined as efficiently finding a set of solutions that diversely and accurately approximates the set of Pareto-optimal solutions. Evolutionary algorithms (EAs) are particularly well-suited for MOO. This may have several reasons, but one of the most intriguing arguments is that the search in MOO for a set of solutions perfectly matches the concept of populations in EAs. In this study, we empirically investigate the role of the population dynamics in a recently proposed evolutionary MOO algorithm.

Many established deterministic approaches to MOO consider the scalarization of the vector-valued problem. To find a diverse approximation of the set of Pareto-optimal solutions, the vector-valued problem is reduced to a number of different scalar-valued optimization problems, which favor different trade-offs between objectives. Prominent examples of scalarization methods are the basic weighted-sum approach and the Tchebycheff-method [1].

In contrast, most state-of-the-art evolutionary methods for solving MOO problems more directly rely on the notion of Pareto-dominance for finding good approximations of the Pareto-optimal set. This does not explicitly need information regarding the importance of single criteria. However, selection in general requires a ranking of the individuals in the current population including the comparison of solutions that are not related in terms of Pareto-dominance. This ranking may be viewed as some kind of scalarization, which is, however, dynamic and usually for each solution dependent on the other solutions represented in the current population. In the population-based algorithms considered here, this ranking is induced by the hypervolume in the objective space covered by the solutions in the population: Selection aims at preserving those solutions that maximize the covered hypervolume.

Scalarization approaches as outlined above are only feasible if there are only very few objectives, as the number of combinations of weighting coefficients expressing relative preferences for objectives grows exponentially with the number of objectives and so scales the number of single-objective optimization trials in simple scalarization methods. However, for two or three objectives the benefits of directly evolving a set of solutions compared to independent (parallel) single-objective optimization trials on scalarized functions are not obvious. Note that there exists scalarization methods (e.g., based on the Tchebychef metric) that do not suffer from the often mentioned problems of weighted-sum approaches with non-convex MOO problems.

In this study, we consider the covariance matrix adaptation evolution strategy (CMA-ES [2], [3]), which is an elaborate algorithm for real-valued optimization performing well in real-world applications [4] as well as in benchmark scenarios [5], and its multi-objective variant, the MO-CMA-ES [6], [7]. The latter features a population-based approach combining the methods for mutation and strategy adaptation from the elitist CMA-ES with multi-objective selection. This makes the algorithm ideal for studying the influence of the population dynamics in the case of few objectives, because we can directly compare the MO-CMA-ES with single-objective optimizations using the elitist CMA-ES. To this end, we formulate "hybrid" algorithms combining established scalarization methods with the CMA-ES.

The remainder of this paper is organized as follows. First, the single- and multi-objective elitist CMA-ESs are introduced. Then, two basic methods for scalarization are presented and combined with the CMA-ES. After that, the experimental setup is described and results of the performance assessment are presented. The paper closes with conclusions and suggestions for future work.

II. Scalarization of Multi-Objective Optimization Problems

This section introduces the weighted-sum method and the method of weighted metrics for the scalarization of vector-valued optimization problems. The ability of these methods to find Pareto-optimal solutions are briefly discussed.

Real-valued, m-objective optimization problems

$$\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m, \boldsymbol{x} \mapsto (f_1(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x}))$$

to be minimized in each objective are considered.

A. The Weighted-Sum Method

The weighted-sum method (e.g., [8], [9]) associates each objective function with a weighting coefficient and minimizes the weighted sum of the objectives. Thus, the vector-valued optimization problem is transformed into a scalar optimization problem of the following form:

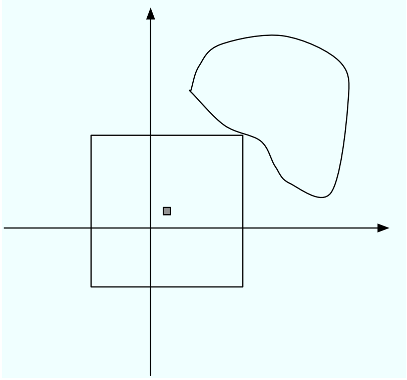
$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad \omega^{(\boldsymbol{f},\boldsymbol{w})}(\boldsymbol{x}) = \sum_{i=1}^{m} w_i f_i(\boldsymbol{x}) \ ,$$

where $\boldsymbol{w}$ is a weighting-vector with $w_i \geq 0$ for all $i = 1, \ldots, m$ and $\sum_{i=1}^{m} w_i = 1$.

B. The Method of Weighted Metrics

In this method, the distance between some reference point and the feasible objective region is minimized. Most often, the utopian point $\boldsymbol{u}^*$ is used as a reference and L_p-metrics are used to measure distance. Moreover, each single objective function is associated with a weighting coefficient. Thus, the multi-objective optimization problem $\boldsymbol{f}$ is replaced with the following single-objective optimization problem, called the *weighted L_p-problem*:

Fig. 1. Illustration of the Tchebycheff-method for the scalarization of a multi-objective optimization problem. The figure emphasizes the ability of the method to find solutions within concave regions of the Pareto-front.



$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad L_p^{(\boldsymbol{f},\boldsymbol{w},\boldsymbol{u}^*)}(\boldsymbol{x}) = \left(\sum_{i=1}^{m} w_i |f_i(\boldsymbol{x}) - u_i^*|^p\right)^{1/p}$$

with $1 \leq p < \infty$, $w_i \geq 0$ for all $i = 1, \ldots, m$ and $\sum_{i=1}^{m} w_i = 1$. In the case of $p = \infty$, the metric is also called the *Tchebycheff metric* and the *weighted L_∞*- or the *Tchebycheff problem* is of the form:

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad \tau^{(\boldsymbol{f},\boldsymbol{w},\boldsymbol{u}^*)}(\boldsymbol{x}) = \max_{i=1,\ldots,m} w_i |f_i(\boldsymbol{x}) - u_i^*|$$

The utopian point $\boldsymbol{u}^* = \boldsymbol{z}^* - \boldsymbol{\varepsilon}$ can be set via the *ideal point* $\boldsymbol{z}^*$ with $z_i^* = \min\{f_i(\boldsymbol{x}) : x \in \mathbb{R}^n\}$ and $\boldsymbol{\varepsilon} > 0$.

C. Properties of Scalarization Methods

Now we review some properties of the methods introduced before regarding their ability to find every Pareto-optimal solution of a vector-valued optimization problem

Recall that a function $g : \mathbb{R}^n \to \mathbb{R}$ is convex if for all $\boldsymbol{x}^1, \boldsymbol{x}^2 \in \mathbb{R}^n$ and $0 \leq \beta \leq 1$ the following equation holds

$$g(\beta\boldsymbol{x}^2 + (1-\beta)\boldsymbol{x}^2) \leq \beta g(\boldsymbol{x}^1) + (1-\beta) g(\boldsymbol{x}^2) \ .$$

A set $S \subseteq \mathbb{R}^n$ is convex if $\boldsymbol{x}^1, \boldsymbol{x}^2 \in S$ implies that $\beta\boldsymbol{x}^1 + (1-\beta)\boldsymbol{x}^2 \in S$ with $0 \leq \beta \leq 1$.

A multi-objective optimization problem is convex, if all the objective functions and the feasible region are convex.

Let $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ be a multi-objective optimization problem with the feasible region $S \subset \mathbb{R}^n$. As stated in [1], if $\boldsymbol{f}$ is convex and $\boldsymbol{x}^* \in S$ is Pareto-optimal, then there exists a weighting-vector $\boldsymbol{w} \in \mathbb{R}^m$ such that $\boldsymbol{x}^*$ is a solution of $\omega^{(\boldsymbol{f},\boldsymbol{w})}$.

The Tchebycheff-method is able to find every Pareto-optimal solution regardless of the vector-valued optimization problem $\boldsymbol{f}$ being convex. That is, let $\boldsymbol{x}^* \in S$ be Pareto-optimal. Then there exists a weighting vector $\boldsymbol{w} \in \mathbb{R}^m$ such that $\boldsymbol{x}^*$ is a solution of $\tau^{f,\boldsymbol{w},\boldsymbol{u}^*}$.

III. Covariance Matrix Adaptation Evolution Strategies in Multi-objective Optimization

In this section we describe the covariance matrix adaptation evolution strategy (CMA-ES). First we review the elitist single-objective CMA-ES, and then we outline the steady-state multi-objective CMA-ES.

The key idea of CMA-ES algorithms is to alter the mutation distribution such that the probability to reproduce steps in the search space that led to the actual population (i.e., produced offspring that were selected) is increased. This enables the algorithm to detect correlations between object variables and to become invariant under affine transformations of the search space. The algorithms implement several important concepts for adapting search distributions. The first one is known as *derandomization* meaning that the mutation distribution is adapted in a deterministic way. The second principle is *cumulation*, which refers to taking the search path of the population over the past generations into account, where the influence of previous steps decays exponentially.

A. Nomenclature

In the elitist CMA-ES, an individual, a, in generation g is a 5-tuple $a^{(g)} = [\boldsymbol{x}^{(g)}, \bar{p}_{\text{succ}}^{(g)}, \sigma^{(g)}, \boldsymbol{p}_c^{(g)}, \boldsymbol{C}^{(g)}]$ comprising its candidate solution vector $\boldsymbol{x}^{(g)} \in \mathbb{R}^n$, an averaged success rate $\bar{p}_{\text{succ}}^{(g)} \in [0,1]$, the global step size $\sigma^{(g)} \in \mathbb{R}_+$, an evolution path $\boldsymbol{p}_c^{(g)} \in \mathbb{R}^n$, and the covariance matrix $\boldsymbol{C}^{(g)} \in \mathbb{R}^{n\times n}$. For the generational multi-objective CMA-ES, an individual $a_k^{(g)}$ denotes the k-th individual in generation g. Additionally, the following nomenclature is used:

- $f : \mathbb{R}^n \to \mathbb{R}, \boldsymbol{x} \mapsto f(\boldsymbol{x})$ is the scalar fitness function to be minimized. For the multi-objective CMA-ES, $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m, \boldsymbol{x} \mapsto \boldsymbol{f}(\boldsymbol{x})$ is the vector-valued fitness function.
- $\mathcal{N}(\boldsymbol{m}, \boldsymbol{C})$ is a multi-variate normal distribution with mean vector $\boldsymbol{m}$ and covariance matrix $\boldsymbol{C}$. The notation $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{C})$ denotes that random variable $\boldsymbol{x}$ is distributed according to the distribution $\mathcal{N}(\boldsymbol{m}, \boldsymbol{C})$.
- $\boldsymbol{x}_{1:\lambda}^{(g)} \in \mathbb{R}^n$ is the best point from $\left\{\boldsymbol{x}_1^{(g)}, \ldots, \boldsymbol{x}_\lambda^{(g)}\right\}$, that is, $f(\boldsymbol{x}_{1:\lambda}^{(g)}) \leq f(\boldsymbol{x}_i^{(g)})$ for all $i = 1, \ldots, \lambda$.

B. Elitist Single-Objective CMA Evolution Strategy

The (1+1)-CMA-ES combines success-based step size adaptation and elitist selection [10], [11], [12] with the covariance matrix adaptation presented in [2].

The algorithm is described within three routines. In the main routine, `(1+1)-CMA-ES`, the new candidate solution is sampled and the parent solution $a^{(g)}$ is updated depending on whether the new solution $a^{(g+1)}$ is better than $a^{(g)}$.

```
Procedure (1+1)-CMA-ES ( $\boldsymbol{x}^{(0)}$ )
$g = 0$, initialize $a^{(g)}$ with $\boldsymbol{x}^{(0)}$
repeat
    $a^{(g+1)} \leftarrow a^{(g)}$
    $\boldsymbol{x}'^{(g+1)} \sim \mathcal{N}\left(\boldsymbol{x}^{(g)}, {\sigma^{(g)}}^2 \boldsymbol{C}^{(g)}\right)$
    $\sigma$-update $\left(a^{(g+1)}, \mathbb{I}\left(f\left(\boldsymbol{x}'^{(g+1)}\right) \leq f\left(\boldsymbol{x}^{(g)}\right)\right)\right)$
    if $f\left(\boldsymbol{x}'^{(g+1)}\right) \leq f\left(\boldsymbol{x}^{(g)}\right)$ then
        $\boldsymbol{x}^{(g+1)} \leftarrow \boldsymbol{x}'^{(g+1)}$
        $\boldsymbol{C}$-update $\left(a^{(g+1)}, \frac{\boldsymbol{x}^{(g+1)} - \boldsymbol{x}^{(g)}}{\sigma^{(g)}}\right)$
    $g \leftarrow g + 1$
until stopping criterion is met
```

Here the indicator function $\mathbb{I}(\cdot)$ is one if its argument is true and zero otherwise. Thus, $\mathbb{I}\left(f\left(\boldsymbol{x}'^{(g+1)}\right) \leq f\left(\boldsymbol{x}^{(g)}\right)\right)$ is one if the last mutation has been successful and zero otherwise. After sampling the new candidate solution, the step size is updated based on the success with a learning rate c_p ($0 < c_p \leq 1$) using a target success rate $p_{\text{succ}}^{\text{target}}$.

```
Procedure $\sigma$-update ( $a = [\boldsymbol{x}, \bar{p}_{\text{succ}}, \sigma, \boldsymbol{p}_c, \boldsymbol{C}], p_{\text{succ}}$ )
$\bar{p}_{\text{succ}} \leftarrow (1 - c_p)\,\bar{p}_{\text{succ}} + c_p p_{\text{succ}}$
$\sigma \leftarrow \sigma \cdot \exp\left(\frac{1}{d}\left(\bar{p}_{\text{succ}} - \frac{p_{\text{succ}}^{\text{target}}}{1 - p_{\text{succ}}^{\text{target}}}\,(1 - \bar{p}_{\text{succ}})\right)\right)$
```

This update rule is rooted in the $1/5$-success-rule proposed by [10] and is an extension from the rule proposed by [13].

If the best new candidate solution was better than the parent individual (see main routine), the covariance matrix is updated as in the original CMA-ES (see [2]).

```
Procedure $\boldsymbol{C}$-update ( $a = [\boldsymbol{x}, \bar{p}_{\text{succ}}, \sigma, \boldsymbol{p}_c, \boldsymbol{C}], \boldsymbol{x}_{\text{step}}$ )
if $\bar{p}_{\text{succ}} < p_{\text{thresh}}$ then
    $\boldsymbol{p}_c \leftarrow (1 - c_c)\boldsymbol{p}_c + \sqrt{c_c(2 - c_c)}\,\boldsymbol{x}_{\text{step}}$
    $\boldsymbol{C} \leftarrow (1 - c_{\text{cov}})\boldsymbol{C} + c_{\text{cov}} \cdot \boldsymbol{p}_c {\boldsymbol{p}_c}^{\mathrm{T}}$
else
    $\boldsymbol{p}_c \leftarrow (1 - c_c)\boldsymbol{p}_c$
    $\boldsymbol{C} \leftarrow (1 - c_{\text{cov}})\boldsymbol{C} + c_{\text{cov}} \cdot \left(\boldsymbol{p}_c {\boldsymbol{p}_c}^{\mathrm{T}} + c_c(2 - c_c)\boldsymbol{C}\right)$
```

The update of the evolution path $\boldsymbol{p}_c$ depends on the value of $\bar{p}_{\text{succ}}$. If the smoothed success rate $\bar{p}_{\text{succ}}$ is high, that is, above $p_{\text{thresh}} < 0.5$, the update of the evolution path $\boldsymbol{p}_c$ is stalled. This prevents a too fast increase of axes of $\boldsymbol{C}$ when the step size is far too small, for example, in a linear surrounding. The constants c_c and c_{cov} ($0 \leq c_{\text{cov}} < c_c \leq 1$) are learning rates for the evolution path and the covariance matrix, respectively. The factor $\sqrt{c_c(2 - c_c)}$ normalizes the variance of $\boldsymbol{p}_c$ viewed as a random variable (see [2]). The evolution path $\boldsymbol{p}_c$ is used to update the covariance matrix. The new covariance matrix is a weighted mean of the old covariance matrix and the outer product of $\boldsymbol{p}_c$. In the second case (line 5), the second summand in the update of $\boldsymbol{p}_c$ is missing and the length of $\boldsymbol{p}_c$ shrinks. Although of minor relevance, the term $c_{\mathrm{c}}(2-c_{\mathrm{c}})\boldsymbol{C}$ (line 6) compensates for this shrinking in $\boldsymbol{C}$.

The (external) strategy parameters are the population size, target success probability $p_{\mathrm{succ}}^{\mathrm{target}}$, 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 [6] and used in this paper, are: $d = 1+n/2$, $p_{\mathrm{succ}}^{\mathrm{target}} = (5+\sqrt{1/2})^{-1}$, $c_p = p_{\mathrm{succ}}^{\mathrm{target}}/(2+p_{\mathrm{succ}}^{\mathrm{target}})$, $c_{\mathrm{c}} = 2/(n+2)$, $c_{\mathrm{cov}} = 2/(n^2+6)$, and $p_{\mathrm{thresh}} = 0.44$.

The elements of the initial individual $a^{(0)}$ are set to $\overline{p}_{\mathrm{succ}} = p_{\mathrm{succ}}^{\mathrm{target}}$, $\boldsymbol{p}_c = \boldsymbol{0}$, and $\boldsymbol{C} = \mathbf{I}$. The initial candidate solution $\boldsymbol{x}^{(0)} \in \mathbb{R}^n$ and the initial $\sigma \in \mathbb{R}^+$ must be chosen problem dependent. The optimum should presumably be within the cube $\bigotimes_{i=1}^{n}[\,x_i^{(0)}-2\,\sigma, x_i^{(0)}+2\,\sigma\,]$.

C. Multi-Objective CMA Evolution Strategy

In the following, we briefly outline the multi-objective covariance matrix adaptation evolution strategy (MO-CMA-ES) according to [7], for a detailed description and a performance evaluation on benchmark functions (in particular a comparison with real-valued NSGA-II [14]) we refer to [6].

The MO-CMA-ES relies on the non-dominated sorting selection scheme suggested by [14]. As in the SMS-EMOA [15], the hypervolume-indicator serves as second-level sorting criterion to rank individuals on 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 $\mathrm{ndom}(A) = \{a \in A | \nexists a' \in A : a' \prec a\}$ 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 without the solutions with lower ranks ([14]). 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 $a \in \mathrm{ndom}_i(A)$. Let $\Delta_{\mathcal{S}}(a, A)$ be the contributing hypervolume of a with respect to $\{a' \in A|\,\mathrm{rank}(a', A) = \mathrm{rank}(a, A)\}$ [15]. Moreover, let $\mathrm{cont}(a, A)$ be the contribution rank of a with respect to its contributing hypervolume. Then the following relation is defined.

$$\begin{aligned} a \prec_A a' \Leftrightarrow \quad & \mathrm{rank}(a, A) < \mathrm{rank}(a', A) \vee \\ & (\mathrm{rank}(a, A) = \mathrm{rank}(a', A) \wedge \\ & \mathrm{cont}(a, A) > \mathrm{cont}(a', A)) \end{aligned}$$

The standard version of the steady-state MO-CMA-ES reads:

Algorithm 4: Steady-State MO-CMA-ES

```
1  $g = 0$, initialize $Q^{(g)} \leftarrow \left\{a_k^{(g)} | 1 \leq k \leq \mu\right\}$
2  repeat
3      $i \leftarrow \mathcal{U}(1, |\,\mathrm{ndom}\left(Q^{(g)}\right)|)$
4      $a^{(g+1)} \leftarrow Q_{\prec:i}^{(g)}$
5      $a'^{(g+1)} \leftarrow a^{(g+1)}$
6      $\boldsymbol{x}'^{(g+1)} \sim \mathcal{N}\left(\boldsymbol{x}^{(g+1)}, {\sigma^{(g+1)}}^2 \boldsymbol{C}^{(g+1)}\right)$
7      $Q^{(g)} \leftarrow Q^{(g)} \cup \left\{a'^{(g+1)}\right\}$
8      $\boldsymbol{C}$-update $\left(a'^{(g+1)}, \frac{\boldsymbol{x}'^{(g+1)} - \boldsymbol{x}^{(g+1)}}{\sigma^{(g+1)}}\right)$
9      $\sigma$-update $\left(a^{(g+1)}, \mathbb{I}\left(a'^{(g+1)} \prec_{Q^{(g)}} a^{(g+1)}\right)\right)$
10     $\sigma$-update $\left(a'^{(g+1)}, \mathbb{I}\left(a'^{(g+1)} \prec_{Q^{(g)}} a^{(g+1)}\right)\right)$
11     $Q^{(g+1)} \leftarrow Q^{(g)} \setminus Q_{\prec:\mu+1}^{(g)}$
12     $g \leftarrow g + 1$
13 until stopping criterion is met
```

The parent of the only offspring is selected uniformly at random among the non-dominated solutions in the current population and an offspring is generated (line 4–7). The step size of the parent and its offspring are updated depending on whether the mutation was successful, that is, whether the offspring is better than the parent according to the relation $\prec_{Q^{(g)}}$.

Both step size and covariance matrix update are the same as in the single-objective (1+1)-CMA-ES described above.

The best μ individuals in $Q^{(g)}$ sorted by $\prec_{Q^{(g)}}$ form the next parent generation (line 11, where $Q_{\prec:i}^{(g)}$ is the i-th best offspring in $Q^{(g)}$ w.r.t. $\prec_{Q^{(g)}}$).

D. Hybrid MO-CMA-ES

In this section we introduce the canonical version of an algorithm for solving vector-valued optimization problems by using the scalarization methods introduced previously and the single-objective, elitist (1+1)-CMA-ES. The new algorithm is denoted by (1+1)-CMA-ES$^{\tau,\omega}$, where the superscript indicates the scalarization method: τ denotes the Tchebycheff and ω the weighted sum approach.

We restrict ourselves to bicriteria optimization problems $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^2$ that allow for a simple and efficient scheme to systematically select the weighting-vectors. That is, given a number of weight steps n_{weights}, $0 < n_{\text{weights}}$, a set of weighting-vectors $W = \{(\alpha, 1-\alpha) | \alpha = i/\left(n_{\text{weights}}-1\right), 0 \leq i \leq n_{\text{weights}}-1\} \subset \mathbb{R}^2$ is populated. For each $\boldsymbol{w} \in W$, $w_i \geq 0, i = 1, 2$ and $\sum_{i_1}^{2} w_i = 1$, the (1+1)-CMA-ES$^{\tau(\boldsymbol{f},\boldsymbol{w},\boldsymbol{u}^*)}$ and the (1+1)-CMA-ES$^{\omega(\boldsymbol{f},\boldsymbol{w})}$ (line 6), respectively, are run for some number of generations, producing a final solution $\boldsymbol{x}^*$ and the corresponding objective vector $\boldsymbol{f}(\boldsymbol{x}^*)$.

Given a weighting-vector $\boldsymbol{w}$ and the corresponding solution $\boldsymbol{x}^{\boldsymbol{w}}$, Rudolph et al. suggested in [16]

Procedure (1+1) -CMA-ES$^{\tau,\omega}(n_{\text{weights}})$

```
$\mathcal{F} \leftarrow \emptyset$
$W \leftarrow \{(\alpha, 1-\alpha) | \alpha = i/\left(n_{\text{weights}}-1\right), 0 \leq i \leq n_{\text{weights}}-1\}$
foreach $\boldsymbol{w} \in W$ do
    Initialize $\boldsymbol{x}^{(0)}$
    $(\boldsymbol{x}^*, \boldsymbol{f}(\boldsymbol{x}^*)) \leftarrow$ (1+1)-CMA-ES $^{\{\tau,\omega\}}(\boldsymbol{x}^{(0)})$
    $\mathcal{F} \leftarrow \mathcal{F} \cup \{\boldsymbol{f}(\boldsymbol{x}^*)\}$
```

to use $\boldsymbol{x}^{\boldsymbol{w}}$ as initial search point in solving the single-objective problem corresponding to $\boldsymbol{w}' = \boldsymbol{w} + (1/\left(n_{\text{weights}}-1\right), 1/\left(n_{\text{weights}}-1\right))$. Apart from the initial search point, all parameters of the (1+1)-CMA-ES are reinitialized, in particular the global step-size σ. The modified algorithm is denoted with (1+1)-CMA-ES$^{*,\tau,\omega}$ and is outlined in algorithm 6.

Procedure (1+1) -CMA-ES$^{*,\tau,\omega}(n_{\text{weights}})$

```
$\mathcal{F} \leftarrow \emptyset$, initialize $\boldsymbol{x}^{(0)}$
$W \leftarrow \{(\alpha, 1-\alpha) | \alpha = i/\left(n_{\text{weights}}-1\right), 0 \leq i \leq n_{\text{weights}}-1\}$
foreach $\boldsymbol{w} \in W$ do
    $(\boldsymbol{x}^*, \boldsymbol{f}(\boldsymbol{x}^*)) \leftarrow$(1+1)-CMA-ES$^{\{\tau,\omega\}}(\boldsymbol{x}^{(0)})$
    $\mathcal{F} \leftarrow \mathcal{F} \cup \{\boldsymbol{f}(\boldsymbol{x}^*)\}$
    $\boldsymbol{x}^{(0)} \leftarrow \boldsymbol{x}^*$
```

a) Calculating the utopian point: The Tchebycheff-method requires the utopian point $\boldsymbol{u}^*$ for its operation. Unfortunately, this point is not known in advance given a multi-objective optimization problem. Thus, we rely on the (1+1)-CMA-ES to calculate an estimate of the ideal point $\boldsymbol{z}^*$ and set $\boldsymbol{u}^* = \boldsymbol{z}^* - \varepsilon$ for some $\varepsilon > 0$. Let $\boldsymbol{f}$ be a bicritera optimization problem, and let $f_1\left(\boldsymbol{x}^{f_1}\right)$,$f_2\left(\boldsymbol{x}^{f_2}\right)$ be the objective values of the final points found by the (1+1)-CMA-ES running on f_1 and f_2 respectively. Then we choose $z_1^* = f_1\left(\boldsymbol{x}^{f_1}\right)$ and $z_2^* = f_2\left(\boldsymbol{x}^{f_2}\right)$.

IV. Experiments

The experimental setup for the performance evaluation has been chosen similar to the one suggested in [7].

The results of both variants of the (1+1)-CMA-ES$^{\{\omega,\tau\}}$ were compared to results of the steady-state MO-CMA-ES using the S-Measure as second-level sorting criterion, termed with $\mathcal{S}$-MO-CMA-ES. We rely on the evaluation module of PISA [17] presented in [18] for the performance assessment of the different algorithms. We briefly outline the evaluation process below and refer to [19], [20] for a detailed description of the methods.

For k algorithms $A_1, \ldots, A_k$ and a benchmark function $\boldsymbol{f}$ a reference Pareto-front approximation is calculated from the normalized approximations generated by $A_1, \ldots, A_k$ (see below). Finally, the results of $A_1, \ldots, A_k$ are compared to the reference set of objective vectors by means of unary quality indicators. Here, the ϵ- and the hypervolume-indicator are measured.

Before the performance indicators are computed, the data are normalized. We want to compare k 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 $\mathcal{R}$. Their objective vectors are normalized such that for every objective the smallest and largest objective function values are mapped to 1 and 2, respectively, by an affine transformation.

The closer the indicator value for an algorithm is to the corresponding indicator value for $\mathcal{R}$ the better the performance. The difference between these values are reported in the results section (i.e., lower values are preferable) [18], [7].

A. Benchmark Functions

Several benchmark functions have been chosen from the literature. The functions ZDT1, ZDT2, ZDT3, and ZDT6 [21] and rotated variants of the ZDT-suite termed IHR1, IHR3 and IHR6 [6]. Additionally, the unconstrained and rotated functions ELLI$_1$, ELLI$_2$, CIGTAB$_1$, and CIGTAB$_2$ proposed in [6] were used. The benchmark function WFG1 introduced in [22] has been chosen to illustrate the ability of a scalarization method to find solutions in concave regions of the Pareto-optimal front depending on the scalarization method. In table I the definitions of the benchmark functions are presented.

B. Parameter Setup

For each of the algorithms, 50 independent trials with 100000 fitness function evaluations were carried out. For the hybrid algorithms a number of weight steps $n_{\text{weights}} = 100$ with the single-objective CMA-ES running for 1000 generations was chosen. For the $\mathcal{S}$-MO-CMA-ES, a population size of $\mu = 100$ and an offspring population size of $\lambda = 1$ was chosen. Accordingly, the number of generations was set to 100000. For all other parameters standard values were used.

C. Results

Table II shows the results of the performance evaluation after 50 trials with 100000 fitness evaluations. Further experiments with 50000 fitness evaluations were carried out but showed similar results. The $\mathcal{S}$-MO-CMA-ES performs significantly better than the (1+1)-

TABLE I
Benchmark problems to be minimized, $\boldsymbol{y} = \boldsymbol{O}\boldsymbol{x}$, where $\boldsymbol{O} \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, and $y_{\text{MAX}} = 1/\max_j(|o_{1j}|)$. For the definition of h, h_f, and h_g we refer to [21]. For the rotated benchmark functions, with $a = 1000$, $b = 100$, $\boldsymbol{y} = \boldsymbol{O}_1\boldsymbol{x}$, and $\boldsymbol{z} = \boldsymbol{O}_2\boldsymbol{x}$, where $\boldsymbol{O}_1$ and $\boldsymbol{O}_2$ are orthogonal matrices. For the benchmark function WFG1, the input vector $\boldsymbol{x}$ is transformed multiple times resulting in the vector $\boldsymbol{t}^4$, with $k = 2$. We refer to [22] for the definition of the transformations.

Problem	n	Variable bounds	Objective function	Optimal solution
IHR1	10	$[-1, 1]$	$f_1(\boldsymbol{x}) = \|y_1\|$ $f_2(\boldsymbol{x}) = g(\boldsymbol{y})\, h_f\Big(1 - \sqrt{h(y_1)/g(\boldsymbol{y})}\Big)$ $g(\boldsymbol{y}) = 1 + 9\left(\sum_{i=2}^{n} h_g(y_i)\right) / (n-1)$	$y_1 \in [0, y_{\text{max}}]$ $y_i = 0$ $i = 2, \ldots n$
IHR3	10	$[-1, 1]$	$f_1(\boldsymbol{x}) = \|y_1\|$ $f_2(\boldsymbol{x}) = g(\boldsymbol{y})\, h_f\Big(1 - \sqrt{h(y_1)/g(\boldsymbol{y})} - \frac{h(y_1)}{g(\boldsymbol{y})} \sin(10\pi y_1)\Big)$ $g(\boldsymbol{y}) = 1 + 9\left(\sum_{i=2}^{n} h_g(y_i)\right) / (n-1)$	$y_1 \in [0, y_{\text{max}}]$ $y_i = 0$ $i = 2, \ldots n$
IHR6	10	$[-1, 1]$	$f_1(\boldsymbol{x}) = 1 - \exp(-4\,\|y_1\|) \sin^6(6\pi y_1)$ $f_2(\boldsymbol{x}) = g(\boldsymbol{y})\, h_f\Big(1 - (f_1(\boldsymbol{x})/g(\boldsymbol{y}))^2\Big)$ $g(\boldsymbol{y}) = 1 + 9\left[\left(\sum_{i=2}^{n} h_g(y_i)\right) / (n-1)\right]^{0.25}$	$y_1 \in [-y_{\text{max}}, y_{\text{max}}]$ $y_i = 0$ $i = 2, \ldots n$
WFG1	10	$[0, 1]$	$f_1(\boldsymbol{t}^4) = 1 - \cos(0.5 t_1^4 \pi)$ $f_2(\boldsymbol{t}^4) = 1 - t_1^4 - (\cos(2 \cdot 5 \cdot \pi t_1^4 + \pi/2))/(2 \cdot 5 \cdot \pi)$	$x_i = 2i + 0.35$ $i = k+1, \ldots, n$
Unconstrained Benchmark Functions				
ELLI$_1$	10	$[-10, 10]$	$f_1(\boldsymbol{y}) = \frac{1}{a^2 n} \sum_{i=1}^{n} a^{2\frac{i-1}{n-1}} y_i^2$ $f_2(\boldsymbol{y}) = \frac{1}{a^2 n} \sum_{i=1}^{n} a^{2\frac{i-1}{n-1}} (y_i - 2)^2$	$y_1 = \cdots = y_n$ $y_1 \in [0, 2]$
ELLI$_2$	10	$[-10, 10]$	$f_1(\boldsymbol{y}) = \frac{1}{a^2 n} \sum_{i=1}^{n} a^{2\frac{i-1}{n-1}} y_i^2$ $f_2(\boldsymbol{z}) = \frac{1}{a^2 n} \sum_{i=1}^{n} a^{2\frac{i-1}{n-1}} (z_i - 2)^2$	
CIGTAB$_1$	10	$[-10, 10]$	$f_1(\boldsymbol{y}) = \frac{1}{a^2 n} \left[y_1^2 + \sum_{i=2}^{n-1} a y_i^2 + a^2 y_n^2\right]$ $f_2(\boldsymbol{y}) = \frac{1}{a^2 n} \left[(y_1 - 2)^2 + \sum_{i=2}^{n-1} a\,(y_i - 2)^2 + a^2 (y_n - 2)^2\right]$	$y_1 = \cdots = y_n$ $y_1 \in [0, 2]$
CIGTAB$_2$	10	$[-10, 10]$	$f_1(\boldsymbol{y}) = \frac{1}{a^2 n} \left[y_1^2 + \sum_{i=2}^{n-1} a y_i^2 + a^2 y_n^2\right]$ $f_2(\boldsymbol{z}) = \frac{1}{a^2 n} \left[(z_1 - 2)^2 + \sum_{i=2}^{n-1} a\,(z_i - 2)^2 + a^2 (z_n - 2)^2\right]$	

CMA-ES$^{\tau(\boldsymbol{f},\boldsymbol{w},\boldsymbol{u}^*)}$ and the (1+1)-CMA-ES$^{*,\tau(\boldsymbol{f},\boldsymbol{w},\boldsymbol{u}^*)}$ on all benchmark functions except for ELLI$_1$ and ELLI$_2$.

For the function ELLI$_2$ the values of the difference of the indicator values exceed the numerical precision of the result table. For this reason, box plots of the values are provided in figure 2.

The function WFG1 has a non-convex Paeto front, and thus it does not come as a surprise that the weighted-sum approach performs badly on WFG1.

The good results of the (1+1)-CMA-ES$^{\omega(\boldsymbol{f},\boldsymbol{w})}$ algorithms on the functions CIGTAB$_1$, CIGTAB$_2$, ELLI$_1$ and ELLI$_2$ are striking. But these are ideal test functions for this method. First, as all objective functions are convex, the weighted-sum approach is in principle able to find all Pareto-optimal points. Second, for fixed weighting coefficients, these functions reduce to quadratic optimization problems (i.e., for these two functions $w_1 f_1(\boldsymbol{x}) + w_2 f_2(\boldsymbol{x})$ are elliptic paraboloids), and exactly these problems are ideal for the CMA-ES. That is, on CIGTAB$_1$, CIGTAB$_2$, ELLI$_1$, and ELLI$_2$, the algorithms (1+1)-CMA-ES$^{\omega(\boldsymbol{f},\boldsymbol{w})}$ and (1+1)-CMA-ES$^{*,\omega(\boldsymbol{f},\boldsymbol{w})}$ decompose the multi-objective optimization problem into subproblems perfectly suited for the CMA-ES, and this results in superior performance on these benchmark functions.

V. Conclusions

Scalarization approaches as considered in this paper suffer from the "curse of dimensionality" and are not suited for many-objective optimization (MOO). However, for bicriteria optimization single-objective evolutionary algorithms applied to a set of scalarized fitness functions may be competitive to population-based approaches, in which the population of a single trial is considered as the solution to the MOO problem.

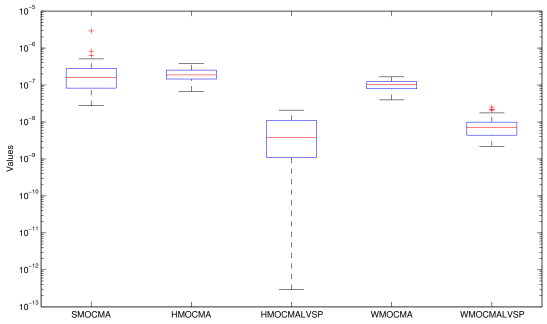
Here we compared the multi-objective CMA evolution strategy (MO-CMA-ES), which is based on the elitist CMA-ES, with application of the elitist CMA-ES to the optimization of scalarized fitness functions. To this end, we introduced new "hybrid" algorithms combining established scalarization methods with the CMA-ES. These algorithms have the same invariance properties as the MO-CMA-ES.

On most standard benchmark functions, the MO-CMA-ES outperformed the elitist CMA-ES combined with scalarization. But when the scalarization produced

TABLE II

RESULTS ON BENCHMARK PROBLEMS. THE UPPER AND LOWER PART OF EACH TABLE SHOWS THE MEDIAN OF 50 TRIALS AFTER 100000 FITNESS EVALUATIONS OF THE HYPERVOLUME-INDICATOR AND THE ϵ-INDICATOR RESPECTIVELY. FOR THE HYPERVOLUME-INDICATOR, THE MEDIAN OF THE ACTUAL HYPERVOLUME COVERED BY THE RESPECTIVE PARETO-FRONT APPROXIMATION W.R.T. THE REFERENCE POINT $(2.1, 2.1)$ IS PUT IN PARENTHESIS. THE SMALLEST VALUE IN EACH ROW IS PRINTED IN BOLD, THE LARGEST IN ITALICS. THE SUPERSCRIPTS I, II, III, IV AND V INDICATE WHETHER AN ALGORITHM IS STATISTICALLY SIGNIFICANTLY BETTER THAN THE $\mathcal{S}$-MO-CMA-ES, (1+1)-CMA-ES$^{\tau(\boldsymbol{f},\boldsymbol{w},\boldsymbol{u}^*)}$, (1+1)-CMA-ES$^{*,\tau(\boldsymbol{f},\boldsymbol{w},\boldsymbol{u}^*)}$, (1+1)-CMA-ES$^{\omega(\boldsymbol{f},\boldsymbol{w})}$ AND (1+1)-CMA-ES$^{*,\omega(\boldsymbol{f},\boldsymbol{w})}$, RESPECTIVELY (TWO-SIDED WILCOXON RANK SUM TEST, $p < 0.001$, SLANTED SUPERSCRIPTS REFER TO A SIGNIFICANCE LEVEL OF $p < 0.01$).

	$\mathcal{S}$-MO-CMA-ES	(1+1)-CMA-ES$^{\tau(\boldsymbol{f},\boldsymbol{w},\boldsymbol{u}^*)}$	(1+1)-CMA-ES$^{*,\tau(\boldsymbol{f},\boldsymbol{w},\boldsymbol{u}^*)}$	(1+1)-CMA-ES$^{\omega(\boldsymbol{f},\boldsymbol{w})}$	(1+1)-CMA-ES$^{*,\omega(\boldsymbol{f},\boldsymbol{w})}$
			Hypervolume-Indicator		
CIGTAB$_1$	**0.001536(1.124140)**$^{\mathrm{II,III,IV,V}}$	*0.053042(1.072630)*	0.010117(1.115560)$^{\mathrm{II}}$	0.003112(1.122560)$^{\mathrm{II,III}}$	0.001841(1.123830)$^{\mathrm{II,III,IV}}$
CIGTAB$_2$	**0.000000(1.210000)**$^{\mathrm{II},\mathit{III},\mathrm{IV,V}}$	0.000002(1.210000)$^{\mathrm{IV},\mathit{V}}$	0.000000(1.210000)$^{\mathrm{II,IV,V}}$	*0.000007(1.209990)*	0.000002(1.210000)$^{\mathrm{IV}}$
ELLI$_1$	*0.066303(1.040650)*	0.002458(1.104500)$^{\mathrm{I}}$	0.002516(1.104440)$^{\mathrm{I}}$	0.001379(1.105580)$^{\mathrm{I,II,III}}$	**0.000931(1.106030)**$^{\mathrm{I,II,III,IV}}$
ELLI$_2$	0.000000(1.210000)	*0.000000(1.210000)*	**0.000000(1.210000)**$^{\mathrm{I,II,IV,V}}$	0.000000(1.210000)$^{\mathrm{I,II}}$	0.000000(1.210000)$^{\mathrm{I,II,IV}}$
IHR1	**0.000424(1.207610)**	0.002778(1.205260)$^{\mathrm{IV}}$	0.002658(1.205380)$^{\mathrm{II,IV,V}}$	*0.002843(1.205200)*	0.002659(1.205380)$^{\mathrm{II,IV}}$
IHR3	**0.000004(1.209800)**$^{\mathrm{II}}$	*0.025611(1.184200)*	0.000270(1.209540)$^{\mathrm{II,IV}}$	0.012111(1.197700)$^{\mathrm{II}}$	0.000216(1.209590)$^{\mathrm{II,III,IV}}$
IHR6	**0.014747(1.159390)**$^{\mathrm{II,III,IV,V}}$	0.177218(0.996915)$^{\mathrm{IV}}$	0.169071(1.005060)$^{\mathrm{II,IV}}$	*0.200065(0.974068)*	0.146458(1.027680)$^{\mathrm{II,III,IV}}$
WFG1	**0.002983(0.829102)**$^{\mathrm{II,III,IV,V}}$	0.014696(0.817389)$^{\mathrm{IV,V}}$	0.012367(0.819718)$^{\mathit{II},\mathrm{IV,V}}$	0.059198(0.772888)$^{\mathrm{V}}$	*0.064734(0.767351)*
ZDT1	**0.000630(1.161750)**$^{\mathrm{II,III,IV,V}}$	*0.050794(1.111590)*	0.000814(1.161570)$^{\mathrm{II,IV,V}}$	0.021237(1.141140)$^{\mathrm{II}}$	0.012361(1.150020)$^{\mathrm{II,IV}}$
ZDT2	**0.006558(1.114180)**$^{\mathrm{II,III,IV,V}}$	*0.038717(1.082020)*	0.034567(1.086170)$^{\mathrm{II,IV,V}}$	0.038245(1.082490)	0.034744(1.085990)$^{\mathrm{II,IV}}$
ZDT3	**0.000254(1.115980)**$^{\mathrm{II,III,IV,V}}$	*0.104633(1.011600)*	0.019863(1.096370)$^{\mathrm{II,IV},\mathit{V}}$	0.056590(1.059640)$^{\mathrm{II}}$	0.026171(1.090060)$^{\mathrm{II,IV}}$
ZDT6	**0.003013(1.150740)**$^{\mathrm{II,III,IV,V}}$	0.032967(1.120790)$^{\mathrm{IV,V}}$	0.032950(1.120800)$^{\mathrm{II,IV,V}}$	*0.033002(1.120750)*	0.032976(1.120780)
			ϵ-Indicator		
CIGTAB$_1$	**0.002869**$^{\mathrm{II,III,IV,V}}$	*0.090382*	0.018137$^{\mathrm{II}}$	0.008138$^{\mathrm{II,III}}$	0.003577$^{\mathrm{II,III,IV}}$
CIGTAB$_2$	**0.000009**$^{\mathrm{II,III,IV}}$	*0.000026*	0.000011$^{\mathrm{II,IV}}$	0.000016$^{\mathrm{II}}$	0.000010$^{\mathrm{II,III,IV}}$
ELLI$_1$	*0.057773*	0.005842$^{\mathrm{I}}$	0.006129$^{\mathrm{I}}$	0.004103$^{\mathrm{I,II,III}}$	**0.003783**$^{\mathrm{I,II,III,IV}}$
ELLI$_2$	*0.000004*	0.000000$^{\mathrm{I}}$	0.000000$^{\mathrm{I,II}}$	0.000000$^{\mathrm{I,II,III}}$	**0.000000**$^{\mathrm{I,II,III,IV}}$
IHR1	**0.001281**	0.004291$^{\mathrm{IV}}$	0.004200$^{\mathrm{II,IV,V}}$	*0.004367*	0.004201$^{\mathrm{II,IV}}$
IHR3	**0.000119**$^{\mathrm{II},\mathit{IV}}$	*0.017937*	0.003412$^{\mathrm{II,IV,V}}$	0.014500$^{\mathrm{II}}$	0.004128$^{\mathrm{II,IV}}$
IHR6	**0.039137**$^{\mathrm{II,III,IV,V}}$	0.164918$^{\mathrm{IV}}$	0.130065$^{\mathrm{II,IV,V}}$	*0.192874*	0.161322$^{\mathrm{II,IV}}$
WFG1	**0.004463**$^{\mathrm{II,III,IV,V}}$	0.034220$^{\mathit{III},\mathrm{IV,V}}$	0.036850$^{\mathrm{IV,V}}$	*0.098827*	0.098597$^{\mathit{IV}}$
ZDT1	**0.001806**$^{\mathrm{II,III,IV,V}}$	*0.084596*	0.007526$^{\mathrm{II,IV,V}}$	0.043881$^{\mathrm{II}}$	0.039923$^{\mathrm{II},\mathit{IV}}$
ZDT2	**0.017523**$^{\mathrm{II,III,IV,V}}$	*0.110446*	0.107069$^{\mathrm{II,IV,V}}$	0.107188$^{\mathrm{II}}$	0.107075$^{\mathrm{II,IV}}$
ZDT3	**0.001299**$^{\mathrm{II,III,IV,V}}$	*0.141574*	0.047396$^{\mathrm{II,IV,V}}$	0.099733$^{\mathrm{II}}$	0.059055$^{\mathrm{II,IV}}$
ZDT6	**0.013306**$^{\mathrm{II,III,IV,V}}$	0.080600$^{\mathrm{IV,V}}$	0.080600$^{\mathrm{II,IV,V}}$	0.080605$^{\mathrm{V}}$	*0.080612*



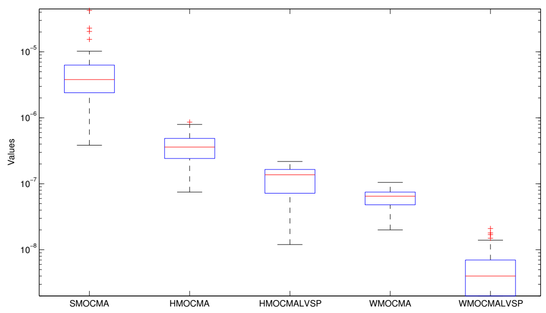


Fig. 2. Box plots of the results of the performance assessment for the fitness function ELLI$_2$. The left plot shows the hypervolume-indicator, the right plot shows the ϵ-indicator. In both plots, from left to right, results of the $\mathcal{S}$-MO-CMA-ES, (1+1)-CMA-ES$^{\tau(\boldsymbol{f},\boldsymbol{w},\boldsymbol{u}^*)}$, (1+1)-CMA-ES$^{*,\tau(\boldsymbol{f},\boldsymbol{w},\boldsymbol{u}^*)}$, (1+1)-CMA-ES$^{\omega(\boldsymbol{f},\boldsymbol{w})}$, (1+1)-CMA-ES$^{*,\omega(\boldsymbol{f},\boldsymbol{w})}$ are shown.

perfect quadratic fitness functions, the powerful strategy adaptation of the CMA-ES could exploit this ideal structure and the CMA-ES combined with weighted-sum scalarization performed on par with the truely population-based MO-CMA-ES. However, across the test problem scalarization using the Tchebycheff metric performed better than weighted aggregation, in particular because of the known problems of the latter approach with non-convex MOO problems.

Future work will include the evaluation of the proposed algorithms on more than two objectives with the hypothesis that the MO-CMA-ES will outperform the hybrid methods even more clearly.

References

[1] K. Miettinen, *Nonlinear Multiobjective Optimization*, ser. Kluwer's International Series in Operations Research & Management Science. Kluwer Academic Publishers, 1999, vol. 12.

[2] N. Hansen and A. Ostermeier, "Completely derandomized self-adaptation in evolution strategies," *Evolutionary Computation*, vol. 9, no. 2, pp. 159–195, 2001.

[3] N. Hansen, "The CMA evolution strategy: A comparing review," in *Towards a new evolutionary computation. Advances on estimation of distribution algorithms.*, I. I. J.A. Lozano, P. Larraãga and E. Bengoetxea, Eds. Springer-Verlag, 2005.

[4] ——, http://www.bionik.tu-berlin.de/user/niko/cmaapplications.pdf, 2006.

[5] A. Auger and N. Hansen, "A restart CMA evolution strategy with increasing population size," in *Proceedings of the IEEE Congress on Evolutionary Computation (CEC 2005)*. IEEE Press, 2005, pp. 1769–1776.

[6] C. Igel, N. Hansen, and S. Roth, "Covariance matrix adaptation for multi-objective optimization," *Evolutionary Computation*, vol. 15, no. 1, 2006.

[7] 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)*, ser. LNCS, vol. 4403. Springer-Verlag, 2007.

[8] S. Gass and T. Saaty, "The Computational Algorithm for the Parametric Objective Function," *Naval Research Logistics Quarterly*, vol. 2, no. 2, pp. 39–45, 1955.

[9] L. Zadeh, "Optimality and Non-Scalar-Valued Performance Criteria," *IEEE Transactions on Automatic Control*, vol. 8, no. 8, pp. 59–60, 1963.

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

[11] H.-P. Schwefel, *Evolution and Optimum Seeking*, ser. Sixth-Generation Computer Technology Series. John Wiley & Sons, 1995.

[12] H.-G. Beyer and H.-P. Schwefel, "Evolution strategies: A comprehensive introduction," *Natural Computing*, vol. 1, no. 1, pp. 3–52, 2002.

[13] 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*, vol. 3, pp. 77–112, 2004.

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

[15] N. Beume, B. Naujoks, and M. Emmerich, "SMS-EMOA: Multiobjective selection based on dominated hypervolume," *European Journal of Operational Research*, 2007.

[16] G. Rudolph, B. Naujoks, and M. Preuss, "Capabilities of MOEA to Detect and Preserve Equivalent Pareto Subsets," in *Proceedings of the 4th International Conference on Evolutionary Multi-Criterion Optimization (EMO 2007)*, S. Obayashi, Ed. Springer-Verlag, 2007, pp. 36–50.

[17] C. Fonseca, J. Knowles, L. Thiele, and E. Zitzler, "A Platform and Programming Language Independent Interface for Search Algorithms (PISA)," 2005. [Online]. Available: http://www.tik.ee.ethz.ch/pisa/

[18] J. Knowles, L. Thiele, and E. Zitzler, "A tutorial on the performance assessment of stochastic multiobjective optimizers," Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Switzerland, 214, Feb. 2006, revised version.

[19] E. Zitzler and L. Thiele, "Multiobjective optimization using evolutionary algorithms — a comparative case study," in *Fifth International Conference on Parallel Problem Solving from Nature (PPSN-V)*, A. E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel, Eds. Berlin, Germany: Springer-Verlag, 1998, pp. 292–301.

[20] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. Grunert da Fonseca, "Performance assesment of multiobjective optimizers: An analysis and review," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 2, pp. 117–132, 2003.

[21] E. Zitzler, K. Deb, and L. Thiele, "Comparison of multiobjective evolutionary algorithms: Empirical results," *Evolutionary Computation*, vol. 8, no. 2, pp. 173–195, 2000.

[22] S. Huband, P. Hingston, L. Barone, and L. While, "A review of multi-objective test problems and a scalable test problem toolkit," in *IEEE Transactions on Evolutionary Computation*, vol. 10, October 2007, pp. 477–506.