Not all parents are equal for MO-CMA-ES

Ilya Loshchilov[1,2], Marc Schoenauer[1,2], and Michèle Sebag[2,1]

[1] TAO Project-team, INRIA Saclay - Île-de-France**
[2] Laboratoire de Recherche en Informatique (UMR CNRS 8623)
Université Paris-Sud, 91128 Orsay Cedex, France
FirstName.LastName@inria.fr

Abstract. The Steady State variants of the Multi-Objective Covariance Matrix Adaptation Evolution Strategy (SS-MO-CMA-ES) generate one offspring from a uniformly selected parent. Some other parental selection operators for SS-MO-CMA-ES are investigated in this paper. These operators involve the definition of multi-objective rewards, estimating the expectation of the offspring survival and its Hypervolume contribution. Two selection modes, respectively using tournament, and inspired from the Multi-Armed Bandit framework, are used on top of these rewards. Extensive experimental validation comparatively demonstrates the merits of these new selection operators on unimodal MO problems.

1 Introduction

The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [9, 8] is considered today as the state-of-the art method for continuous optimization at large, at least for small to medium-sized search space (up to dimension 100) [7]. Its efficiency mostly derives from its invariance properties; not only is it invariant with respect to monotonous transformations of the objective function, like all comparison-based optimization algorithms; it is also invariant with respect to orthogonal transformations of the coordinate system, thanks to the on-line adaptation of the covariance matrix of the Gaussian mutation. The multi-objective version of CMA-ES proposed by Igel et al. [10], called MO-CMA-ES, benefits from these invariance properties (though the hypervolume indicator is not invariant) and performs very well on non-separable problems like the IHR family. MO-CMA-ES proceeds as a $(\mu + \mu)$ algorithm, where μ parents give birth to μ offspring, and the best μ individuals in the sense of Pareto dominance (out of parents plus offspring) become the parents of the next generation. As shown by [4] however, Evolutionary Multi-Objective Optimization can benefit from steady-state strategies. Accordingly, two steady state variants, hereafter called SS-MO-CMA-ES, have been proposed by Igel et al. [12], implementing some $(\mu + 1)$ selection strategy: a parent is selected uniformly (either from the whole parent population, or among the non-dominated parents), and is used to generate a single offspring, which is inserted back into the population at each time

** Work partially funded by FUI of System@tic Paris-Region ICT cluster through contract DGT 117 407 *Complex Systems Design Lab* (CSDL).

step. Significant improvements over the generational version were reported on unimodal benchmark problems.

The present paper investigates some alternative choices of the fertile parent in SS-MO-CMA-ES, based on the conjecture that not all (non-dominated) parents are equal. Several indicators, estimating the expectation of offspring survival or its hypervolume contribution, are considered. These indicators are exploited within a simple tournament selection, or borrowing Multi-Armed Bandits principles [2] to deal with the Exploration vs Exploitation dilemma.

This paper is organized as follows. Section 2 recalls the basics of MO-CMA-ES; the generational and the steady state variants are described within a generic scheme. Section 3 details the proposed parent selection operators and how they fit in the generic scheme. These operators involve a rewarding procedure estimating the goodness of parents, and a selection procedure. In Section 4 the resulting algorithms are experimentally assessed on some well-known benchmark functions, comparatively to the previous versions of MO-CMA-ES, and the paper concludes with some perspectives for further research in Section 5.

2 State of the art

This section briefly recalls the formal background of multi-objective optimization, and the basics of MO-CMA-ES and SS-MO-CMA-ES, referring the reader to [10] and [12] for a more comprehensive description.

Let $\mathcal{D} \subseteq \mathbb{R}^d$ be the decision space, and let $f_1, \ldots, f_m$ denote m objectives defined on the decision space ($f_i : \mathcal{D} \mapsto \mathbb{R}$). The objective space is given by $\mathbb{R}^m$ and the image of x in the objective space is defined as $o_x = (f_1(x), \ldots f_m(x))$. Given a pair of points $(x, y) \in \mathcal{D}$, it is said that x *dominates* y (denoted $x \prec y$) iff x is not worse than y over all objectives, and x is strictly better than y on at least one objective. It is said that $o_x \prec\ o_y$ iff $x \prec y$.

2.1 MO-CMA-ES

Originally, MO-CMA-ES involves a set of μ $(1+1)$-CMA-ES, each of which performs step-size and covariance matrix updates based on its own evolution path, and a Pareto-based survival selection mechanism that selects μ individuals from the population of size 2μ built from all parents and offspring.

Regarding the $(1+1)$-ES, the general rules used for the adaptation of the step-size and the covariance matrix in CMA-ES [9, 8] cannot be used within the $(1+1)$ setting. Specific rules have hence been proposed [11], based on the success rate of the previous evolution steps, *à la* 1/5th rule [16]. The detailed description of those rules fall outside the scope of this paper, though their formal description is given in lines 8-16 in Algorithm 1 for the sake of reproducibility.

Regarding the survival selection mechanism, it is inspired by the Non-dominated Sorting procedure first proposed within the NSGA-II algorithm [6]. Two hierarchical criteria are used in turn: the Pareto rank, and the hypervolume contribution [4], that replaces the original crowding distance. Let $A = \{a_1, \ldots, a_p\}$ denote a set of p points of the objective space.

Pareto Ranking The Pareto ranks w.r.t. A of the points in A are iteratively determined. All non-dominated points in A (denoted $ndom_1(A)$ or simply $ndom(A)$), are given rank 1. The set $ndom(A)$ is then removed from A; from this reduced set, the non-dominated points (denoted $ndom_2(A)$) are given rank 2; the process continues until all points of A have received a Pareto rank. The Pareto rank of point $a \in A$ is denoted $PR(a, A)$.

Hypervolume contribution The *Hypervolume* of a set of points A is sometimes also called "S-Metric" [18]. Let a_{ref} denote a reference point, dominated by all points in A. The hypervolume of A is then the volume of the union of the hypercubes defined by one point of the set and a_{ref}. Formally,

$$H(A) = Volume(\bigcup_{i=1}^{i=p} Rect(a_i, a_{ref}))$$

where $Rect(a, b)$ is the hyper-rectangle whose diagonal is the segment $[ab]$. It is clear that only the non-dominated points in A (i.e. the points in $ndom(A)$) contribute to the hypervolume. The *Hypervolume contribution* of some non-dominated point a is defined as the difference between the hypervolume of the whole set A and that of the set from which a has been removed.

$$\varDelta H(a, A) = H(A) - H(A \backslash \{a\})$$

For dominated points, the hypervolume contribution can also be defined by considering only the points that have the same rank. More precisely, if $PR(a) = k$, i.e. $a \in ndom_k(A)$, then

$$\varDelta H(a, A) = H(ndom_k(A)) - H(ndom_k(A) \backslash \{a\})$$

Survival Selection in MO-CMA-ES All above definitions are extended to points in the decision space as follows. Given a set $X = \{x_1, \dots x_p\}$ in the decision space, given the set $A = \{o_{x1}, \dots, o_{xp}\}$ of their image in the objective space, the Pareto rank (resp. hypervolume contribution) of any point x in X is set to the Pareto rank (resp. hypervolume contribution) of o_x in A.

Using Pareto ranking as first criterion, and the hypervolume contribution as secondary criterion (rather than the crowding distance proposed with the original NSGA-II, as advocated in [4]), a total preorder relation $\prec_X$ is defined on any finite subset X of the decision space, as follows:

$$\begin{aligned} x \prec_X y \Leftrightarrow PR(x, X) < PR(y, X) & \qquad \textit{// lower Pareto rank} \\ \text{or} & \qquad \textit{// same Pareto rank and higher HC} \\ PR(x, X) = PR(y, X) \text{ and } \varDelta H(x, X) > \varDelta H(y, X) & \end{aligned} \tag{1}$$

Ties on hypervolume contributions are broken at random.

Specific care must be taken with the extreme points of the set, i.e. the points for which the hypervolume contribution depends on the choice of the reference

Algorithm 1 (μ+λ)-MO-CMA-ES   Generic MO-CMA-ES scheme

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

point. By convention, they are associated an infinite hypervolume contribution, and they thus dominate in the sense of Eq. (1) all points with same Pareto rank.

The survival selection in MO-CMA-ES finally proceeds as the standard deterministic selection of the $(\mu + \mu)$-ES algorithm: at generation g, after each of the μ parents has generated one offspring, let $Q^{(g)}$ denote the union of the μ parents and the μ offspring. Then the best μ individuals according to $\prec_{Q^{(g)}}$ become the parents of generation $g + 1$.

2.2 Generational and Steady State MO-CMA-ES Algorithms

Algorithm 1 is a generic description of all MO-CMA-ES algorithms, where μ parents generate λ offspring. Borrowing Igel et al's notations [12], $a_i^{(g)}$ denotes some structure containing the i^{th} point x_i of the population at generation g together with its parameters related to the mutation (step-size, covariance matrix, average success rate, . . .).

Lines 3-7 is the loop of offspring generation: at line 5 the parent is copied onto the offspring together with its parameters. It is mutated at line 6 (and evaluated). Then, depending on the success of the offspring (its performance compared to its parents'), the parameters of both individuals are updated between lines 8 and 16 (update of the covariance matrix in the case of successful offspring[3], and update of success rate and mutation step in any case).

The original MO-CMA-ES thus instantiates the generic Alg. 1 by taking $\lambda = \mu$ and having the parent selection on line 4 simply return its second argument ($i_k = k$, i.e., all parents generate exactly one offspring in turn).

In SS-MO-CMA-ES, $\lambda = 1$: only one offspring is generated at each generation. The parent selection (line 4 of Algorithm 1) proceeds by uniformly selecting either one from the μ parents (variant $(\mu+1)$-MO-CMA in the following); or one from the non-dominated parents (variant $(\mu_{\prec}+1)$-MO-CMA in the following).

The survival selection (line 17) then amounts to replacing the worst parent x_w with the offspring if the latter improves on x_w according to $\prec_{Q^{(g)}}$, or discarding the offspring otherwise.

3 New Parent Selections for Steady-State MO-CMA-ES

After [12], the more greedy variant $(\mu_{\prec}+1)$-MO-CMA outperforms all other variants on all unimodal problems. In contrast, on multi-modal problems such as ZDT4 and IHR4, $(\mu+1)$-MO-CMA performs better than $(\mu_{\prec}+1)$-MO-CMA [12], but it does not perform too well, and neither does the generational version of MO-CMA-ES, comparatively to other MOEAs.

These remarks naturally lead to propose more greedy parent selection operators within SS-MO-CMA-ES (line 4 of Alg. 1), in order to further improve its performances on unimodal problems, leaving aside at the moment the multimodality issue. A parent selection operator is based on i) a selection mechanism; and ii) a rewarding procedure (line 18). A family of such operators is presented in this section; the selection procedure either is based on a standard tournament selection (section 3.1), or inspired from the Multi-Armed Bandit paradigm (section 3.2). The rewarding procedures are described in section 3.3.

3.1 Tournament Selection

Standard tournament selection is parameterized from a tournament size $t \in \mathbb{N}$. Given a set X, t-tournament selection proceeds by uniformly selecting t individuals (with or without replacement) from X and returning the best one according to criterion at hand (here, the $\prec_{Q^{(g)}}$ criterion, see Eq. (1)). The parent selection procedure (line 4 of Alg. 1) thus becomes `TournamentSelection(`$Q^{(g)}$`)`.

The rewarding procedure (line 18 of Alg. 1) only computes for each parent its Pareto rank and Hypervolume contribution[4].

The Steady-State MO-CMA-ES using t-size Tournament Selection is denoted $(\mu +_t 1)$-MO-CMA in the following, or $(\mu +_t 1)$ for short. Parameter t thus

[3] The formulation of Algorithm 1 was chosen for its genericity. In practice however, only the surviving offspring will actually adapt their parameters; the update phase thus takes place after the survival selection (line 17).

[4] It is thus redundant with the Survival Selection (line 17), and can be omitted.

controls the selection greediness; the larger t, the more often points with high Hypervolume contribution will be selected on average.

3.2 Multi-Armed Bandit-inspired Selection

Another parent selection procedure (line 4 of Alg. 1) inspired from the Multi-Armed Bandit (MAB) paradigm is described here. How to define the underlying rewards (line 18) will be detailed in next subsection.

The standard MAB setting considers several options, also called arms, each one with an (unknown but fixed) reward distribution [2]. The MAB problem is to find a selection strategy, selecting an arm $i(t)$ in each time step t and getting an instance of the corresponding reward distribution, such that this strategy optimizes the cumulative reward.

An algorithm yielding an optimal result has been proposed by Auer et al. [2]; this algorithm proceeds by selecting the arm which maximizes the sum of an exploitation term (the empirical quality, or average of rewards the arm has ever actually received) and an exploration term (enforcing that non-optimal arms be selected sufficiently often to enforce the identification of the truly optimal arm).

Considering that our setting is a dynamic one (as evolution proceeds toward the Pareto front), no algorithm with theoretical guarantees is available, and some heuristic adaptation of the above MAB algorithm is used:

1. The average reward of an arm (a parent) is replaced by its average reward along a time window of size w;
2. The exploration is enforced by selecting once every arm which i) occurs only once in the time window and ii) is about to disappear from the time window (it was selected w time steps ago);
3. In all other cases, the selection is on the exploitation side, and the arm with best average reward along the last w time steps is selected.

In summary, the MAB-like selection (line 4 of Alg. 1) always selects the parent with best average reward in the last w time steps, except for case 2 (a current parent is about to disappear from the time window). Parameter w thus controls the exploration strength of the selection. Experimentally however, the sensitivity of the algorithm w.r.t. w seems to be rather low, and w was set to 500 in all experiments (section 4).

3.3 Defining Rewards

This subsection describes the rewards underlying the MAB-like selection mechanism (line 18 of Alg. 1). A key related issue is how to share the reward between parents and offspring. On the one hand, if an offspring survives, it is better that some old parents and might thus be a good starting point for further advances toward the Pareto front. The offspring must thus inherit a sufficient fraction of its parent reward, to enable its exploitation. On the other hand, the reward of a parent should be high when it yields good-performing offspring, and in particular no reward should be awarded to the parent if the newborn offspring does not survive. Several reward indicators have been considered.

($\boldsymbol{\mu + 1_{succ}}$) A first possibility is to consider boolean rewards. If an offspring makes it to the next generation, both the offspring and the parent receives reward 1. Formally:

$$r^{(g)} = 1 \text{ if } a_1'^{(g+1)} \in Q^{(g+1)}$$

Such boolean rewards entail a very greedy behavior. The newborn offspring, receiving 1 as instant reward, gets 1 as average reward over the time window; it will thus very rapidly (if not immediately) be selected at next parent. Likewise, its parent which already had a top average reward (it was selected), will improve its average reward and tend to be selected again.

($\boldsymbol{\mu + 1_{rank}}$) A smoother reward is defined by taking into account the rank of the newly inserted offspring:

$$r^{(g)} = 1 - \frac{rank(a_1'^{(g+1)})}{\mu} \text{ if } a_1'^{(g+1)} \in Q^{(g+1)}$$

where $rank(a_1'^{(g+1)})$ is the rank of the newly inserted offspring in population $Q^{(g+1)}$ (using comparison operator $\prec_{Q^{(g)}}$ defined by Eq. (1); the top individual gets rank 0). Along this line, the reward ranges linearly from 1 (for a non-dominated individual with best hypervolume contribution) to 0. A newborn offspring will here be selected rapidly only if it makes it into the top-ranked individuals of the current population. The average reward of the parent can decrease if its offspring gets a poor rank, even if the offspring survives.

($\boldsymbol{\mu + 1_{\Delta H_1}}$) Another way of getting smooth rewards is based on the hypervolume contribution of the offspring. Let us set the reward to 0 for dominated offspring (noting that most individuals are non-dominated in the end of evolution); for non-dominated offspring, one sets the reward to the increase of the total Hypervolume contribution from generation g to $g+1$:

$$r^{(g)} = \sum_{a \in Q^{(g+1)}} \Delta H(a, Q^{(g+1)}) - \sum_{a \in Q^{(g)}} \Delta H(a, Q^{(g)})$$

($\boldsymbol{\mu + 1_{\Delta H_i}}$) In the early stages of evolution, many offspring are dominated and the above Hypervolume-based reward thus gives little information. A relaxation of the above reward, involving a rank-based penalization is thus defined. Formally, if k denote the Pareto rank of the current offspring, the reward is:

$$r^{(g)} = \frac{1}{2^{k-1}} \left(\sum_{ndom_k(Q^{(g+1)})} \Delta H(a, ndom_k(Q^{(g+1)})) - \sum_{ndom_k(Q^{(g)})} \Delta H(a, ndom_k(Q^{(g)})) \right)$$

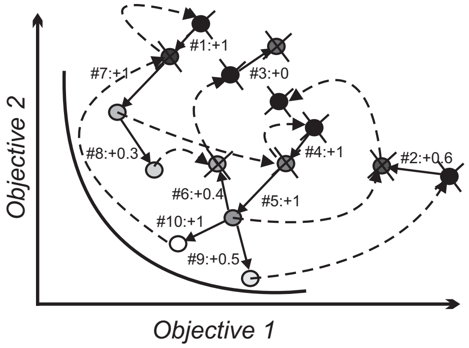

Fig. 1. Reward-based multi-objective optimization with bounded population. "#6:+0.4" means reward 0.4 on 6th iteration.

3.4 Discussion

The difficulty of associating a reward to a pair (parent, offspring) in Multi-Objective optimization is twofold. On the one hand, defining absolute indicators (e.g. reflecting some aggregation of the objective values) goes against the very spirit of MO. On the other hand, relative indicators such as above-defined must be taken with care: they give a snapshot of the current situation, which evolves along the population progress toward the Pareto front. The well-founded Multi-Armed Bandit setting, and its trade-off between Exploration and Exploitation, must thus be modified to account for non-stationarity.

Another difficulty is related to the finiteness of the population: while new arms appear, some old arms must disappear. The parent selection, e.g. based on the standard deterministic selection (Eq. (1)) is biased toward exploitation as it does not offer any way of "cooling down" the process. Such a bias is illustrated in Fig. 1. Let the population size of steady-state EMOA be 5, and consider a sequence of 10 evolution steps, generating 10 new points (oldest, resp. newest points are black resp. white). At each iteration the parent with best reward generates an offspring, then 6 points are compared using Eq. (1), and the worst point (crossed out) is eliminated. The instant parent reward reflects the quality of the offspring. Along evolution, some prospective points/arms are thus eliminated because they progress more slowly than others, although they do progress, due to the fixed population size. Expectedly, this bias toward exploitation adversely affects the discovery of multi-modal and/or disconnected Pareto front. We shall return to this issue in section 5.

4 Experimental Validation

This section reports on the validation of the proposed schemes, comparatively to the baseline MO-CMA-ES algorithms, detailing the experimental setting in section 4.1 before discussing the experimental evidence in section 4.2.

4.1 Experimental Setting

Algorithms. The experimentation involves:

- The steady-state MO-CMA-ES with tournament-based parent selection, where the tournament size t is set to 2 and 10 (respectively noted $(\mu +_2 1)$ and $(\mu +_{10} 1)$);
- The steady-state MO-CMA-ES with MAB-based parent selection, considering the four rewards described in section 3.2 (respectively noted $(\mu + 1_{succ})$, $(\mu + 1_{rank})$, $(\mu + 1_{\Delta H_1})$ and $(\mu + 1_{\Delta H_i})$);
- The baseline algorithms include the generational $(\mu + \mu)$-MO-CMA [10], and its steady-state variants $(\mu + 1)$-MO-CMA and $(\mu_{\prec} + 1)$-MO-CMA [12] (section 2.2).

All parameters of MO-CMA-ES are set to their default values [10] (in particular, $\mu = 100$); all algorithms only differ by their parent selection procedure. All reported results are based on 31 independent runs with at most 200,000 fitness evaluations, and median results are reported when the target precision was reached.

Problems. The well-known bi-criteria ZDT1:3-6 problems [17] and their rotated variants IHR1:3-6 [10] have been considered. Note however that the true Pareto front of all ZDT problems lies on the boundary of the decision space, which might make it easier to discover it. For the sake of an unbiased assessment, the true Pareto front is thus shifted in decision space: $x'_i \leftarrow |x_i - 0.5|$ for $2 \leq i \leq n$, where n is the problem dimension. The shifted ZDT problems are denoted sZDT. The set of recently proposed benchmark problems LZ09-1:5 [14] have also been used for their complicated Pareto front in decision space (Fig. 4).

Performance Measures. Following [13], the algorithmic EMO performance is measured from the hypervolume indicator I_H. Let P be a μ-size approximation of Pareto front and let P^* be the approximate μ-optimal distribution of optimal Pareto points [3]. The approximation error of the Pareto front is defined by $\Delta H(P^*, P) = I_H(P^*) - I_H(P)$.

Furthermore, to support an easy comparison of different algorithms across different problems, all results will be presented 'the horizontal way', i.e., reporting the number of function evaluations needed to reach a given precision. This procedure rigorously supports claims[5] such as *algorithm A is 2 times faster than algorithm B*.

4.2 Result Analysis

All empirical results are displayed in Table 1. These results show that the proposed algorithms generally outperform the baseline MO-CMA-ES approaches,

[5] In opposition, a claim such as *Algorithm A can reach a precision 10 times smaller than algorithm B* is hard to assess when considering different problems.

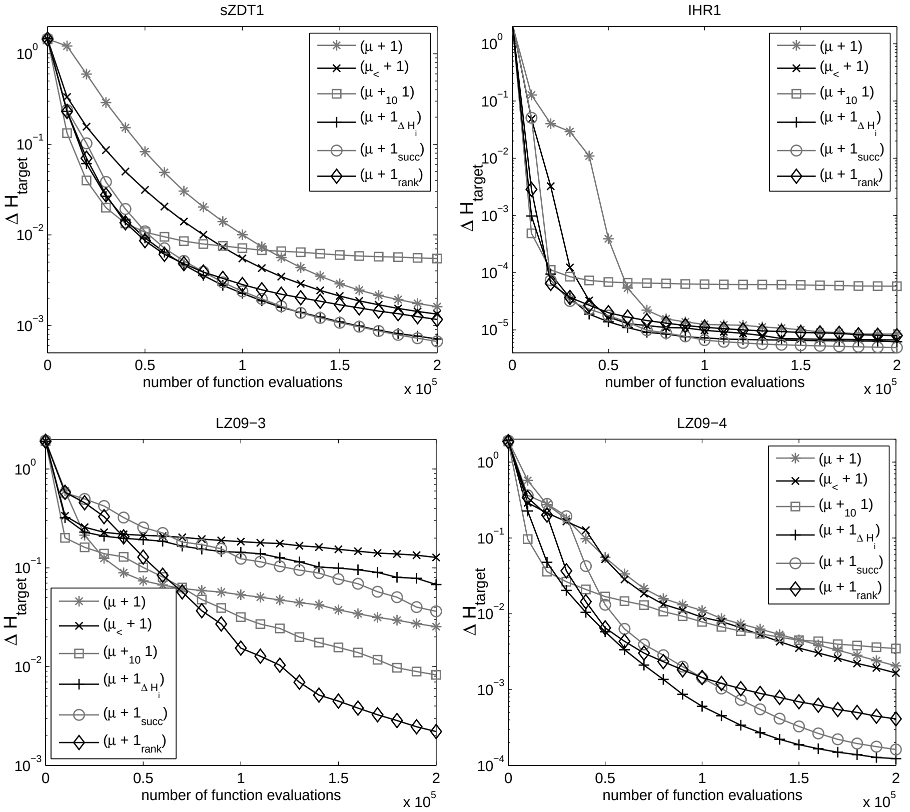

Fig. 2. On-line performances of baseline and proposed variants of steady-state MO-CMA-ES on sZDT1, IHR1, LZ09-3 and LZ09-4 problems (median out of 31 runs).

with the exception of problems sZDT3, IHR6 and LZ09-5. A first general remark is that the steady-state variants of MO-CMA-ES outperform the generational one on unimodal benchmark problems; as already noted in [12], the greedier $(\mu_{\prec}+1)$-MO-CMA is usually faster than the original steady-state on sZDT and IHR problems; in counterpart, it is too greedy on LZ09 problems (see below). Another general remark is that $(\mu+1_{\Delta H_i})$-MO-CMA is usually more robust and faster than $(\mu+1_{\Delta H_1})$-MO-CMA; this fact is explained as the former exploit a better informed hypervolume contribution based reward, considering also the contribution of dominated points.

The on-line performance of most considered algorithms on sZDT1, IHR1, LZ09-3 and LZ09-4 shows the general robustness of $(\mu+1_{rank})$-MO-CMA (Fig.2, displaying $\Delta H(P^*, P)$ versus the number of function evaluations). The comparatively disappointing results of $(\mu+1)$-MO-CMA on IHR1 are explained from the structure of the Pareto front, which includes an easy-to-find segment. This

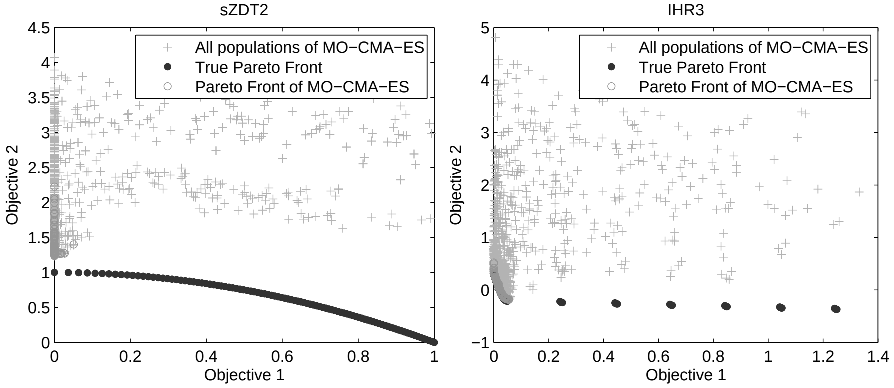

Fig. 3. Typical behavior of $(\mu + 1_{succ})$-MO-CMA on sZDT2 (left) and IHR3 (right) problems: premature convergence after 5,000 fitness function evaluations.

segment can be discovered by selecting the extreme parent (in objective space), thus with probability $1/\mu$ within a uniform selection scheme. Quite the contrary, reward-based selection schemes quickly catch the fruitful directions of search.

The price to pay for this is depicted on Fig. 3, showing $(\mu+1_{succ})$-MO-CMA on sZDT2 and IHR3 problems. On these problems, a premature convergence toward a small segment of the Pareto front is observed after circa 5,000 function evaluations. The interpretation provided for this premature convergence goes as follows. As one part of the Pareto front is easier to find than others, points aiming at this part quickly reach their goal; due to non-dominated sorting (and to the fixed population size), these eliminate other points, resulting in a very poor diversity (in objective space) of the population. This remark suggests that some additional diversity preserving technique should be used together with MO-CMA-ES; note that, even in the original MO-CMA-ES, a premature convergence is observed on IHR3.

LZ09 problems have also been considered because of their non-linear Pareto front in decision space (Fig. 4), contrasting with the linear Pareto front of all sZDT and IHR problems. The results depicted in Fig. 4 show that $(\mu + 1_{rank})$-MO-CMA better approximates the Pareto front than $(\mu + 1)$ and $(\mu +_{10} 1)$-MO-CMA, for all problems except LZ09-5. It is interesting to note that the results of $(\mu+1_{rank})$-MO-CMA after 100,000 fitness evaluations match those of MOEA/D-DE after 150,000 fitness evaluations [14].

Overall (see also Table 1), $(\mu + 1_{rank})$-MO-CMA and $(\mu +_{10} 1)$-MO-CMA perform best on most problems, while $(\mu + 1_{\Delta H_i})$-MO-CMA is slightly more robust. Most generally, all greedy versions of MO-CMA-ES get better results on problems with a convex Pareto front; on problems with a concave or disconnected Pareto front, they suffer from premature convergence, entailed by a loss of diversity, due to non-dominated sorting and bounded population size.

5 Conclusion and Perspectives

The goal and main contribution of the paper is to speed up MO-CMA-ES using new parent selection schemes, based on tournament and reward-based approaches inspired from the Multi-Armed Bandit framework, in order to quickly identify the most fruitful directions of search. Experiments on several bi-objective problems have shown a significantly speed-up of MO-CMA-ES on unimodal problems (for both generational and previous steady-state variants). However, the proposed approach results in a poor convergence on multi-modal multi-objective problems, or problems where some parts of the Pareto front are much easier to reach than others, such as IHR3 (Fig. 3, and discussion in sections 3.4 and 4.2).

These remarks open some perspectives for further research, aimed at preserving the benefits of parent selection schemes while addressing the premature convergence on multi-modal landscapes. A first perspective is to maintain the points located at the border of the already visited region, and to give them some chance to produce offspring as well although they are dominated. The question thus becomes to handle yet another exploitation vs exploration dilemma, and distribute the fitness evaluations between the current population and the borderline points; it also remains to extend the reward definition for the borderline points. Such an approach is similar in spirit to the so-called BIPOP-CMA-ES designed to overcome premature convergence within single-objective evolutionary optimization [7]; BIPOP-CMA-ES proceeds by maintaining one large population for exploration purposes, and a small one for fast and accurate convergence.

A second perspective is to design a more integrated multi-objective CMA-ES based algorithm, by linking the reward mechanism used in the parent selection and the internal update rules of CMA-ES. Indeed, the success rate used to control the $(1 + 1)$-ES evolution and the empirical success expectation used in $(\mu + 1_{succ})$-MO-CMA are strongly related. Further work will consider how to use the success rate *in lieu* of reward for parental selection, expectedly resulting in a more consistent evolution progress. Meanwhile, the CMA update rules might want to consider the discarded offspring (possibly weighting their contribution depending on their hypervolume contribution), since they might contain useful information even though they are discarded. Again, similar ideas have been investigated in the single objective case: the Active Covariance Matrix Adaptation [1] does use unsuccessful trials to update the distribution of mutation parameters. Some other recent proposals [5] might also help accelerating even further the MO-CMA-ES on separable functions: mirrored sampling systematically evaluates two symmetric points w.r.t. the mean of the Gaussian distribution, and sequential selection stops generating offspring after the first improvement over the parent.

Last but not least, the MO-CMA-ES and the proposed parent selection schemes must be analysed and compared with other state-of-the art MOEAs, specifically SMS-EMOA [4], the first algorithm to advocate the use of steady state within EMO to our best knowledge; it also proposed separable variation operators, resulting in excellent results comparatively to MO-CMA-ES on separable problems. How to extend these variation operators in the non-separable case, borrowing approximation ideas from [15] will be investigated.

References

1. D. V. Arnold and N. Hansen. Active covariance matrix adaptation for the (1+1)-CMA-ES. In J. Branke et al., editor, *ACM-GECCO*, pages 385–392, 2010.
2. P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multi-armed bandit problem. *Machine Learning*, 47(2-3):235–256, 2002.
3. A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Theory of the hypervolume indicator: Optimal μ-distributions and the choice of the reference point. In *FOGA*, pages 87–102. ACM, 2009.
4. 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.
5. D. Brockhoff, A. Auger, N. Hansen, D. V. Arnold, and T. Hohm. Mirrored sampling and sequential selection for evolution strategies. In R. S. et al., editor, *Parallel Problem Solving from Nature (PPSN XI)*, volume 6238 of *LNCS*, pages 11–20. Springer, 2010.
6. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A Fast Elitist Multi-Objective Genetic Algorithm: NSGA-II. *IEEE TEC*, 6:182–197, 2000.
7. N. Hansen, A. Auger, R. Ros, S. Finck, and P. Posík. Comparing results of 31 algorithms from the black-box optimization benchmarking BBOB-2009. In J. Branke et al., editor, *GECCO (Companion)*, pages 1689–1696. ACM, July 2010.
8. N. Hansen, S. Müller, and P. Koumoutsakos. Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES). *Evolution Computation*, 11(1), 2003.
9. N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.
10. C. Igel, N. Hansen, and S. Roth. Covariance Matrix Adaptation for Multi-objective Optimization. *Evolutionary Computation*, 15(1):1–28, 2007.
11. C. Igel, T. Suttorp, and N. Hansen. A computational efficient covariance matrix update and a (1+1)-CMA for evolution strategies. In M. Keijzer et al., editor, *GECCO 2006*, pages 453–460. ACM Press, 2006.
12. C. Igel, T. Suttorp, and N. Hansen. Steady-state selection and efficient covariance matrix update in the multi-objective CMA-ES. In S. Obayashi et al., editor, *EMO'07*, volume 4403 of *LNCS*, pages 171–185. Springer-Verlag, 2007.
13. J. Knowles, L. Thiele, and E. Zitzler. A tutorial on the performance assessment of stochastic multiobjective optimizers. Technical Report TIK 214, ETH Zürich, 2006.
14. H. Li and Q. Zhang. Multiobjective Optimization Problems With Complicated Pareto Sets, MOEA/D and NSGA-II. *IEEE Trans. Evolutionary Computation*, 13(2):284–302, 2009.
15. R. Ros and N. Hansen. A Simple Modification in CMA-ES Achieving Linear Time and Space Complexity. In G. Rudolph et al., editor, *Parallel Problem Solving from Nature (PPSN X)*, volume 5199 of *LNCS*, pages 296–305, 2008.
16. M. Schumer and K. Steiglitz. Adaptive step size random search. *Automatic Control, IEEE Transactions on*, 13:270–276, 1968.
17. E. Zitzler, K. Deb, and L. Thiele. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation*, 8:173–195, 2000.
18. E. Zitzler and L. Thiele. Multiobjective optimization using evolutionary algorithms – a comparative case study. In A.E. Eiben et al., editor, *PPSN V*, volume 1498 of *LNCS*, pages 292–301. Springer Verlag, 1998.

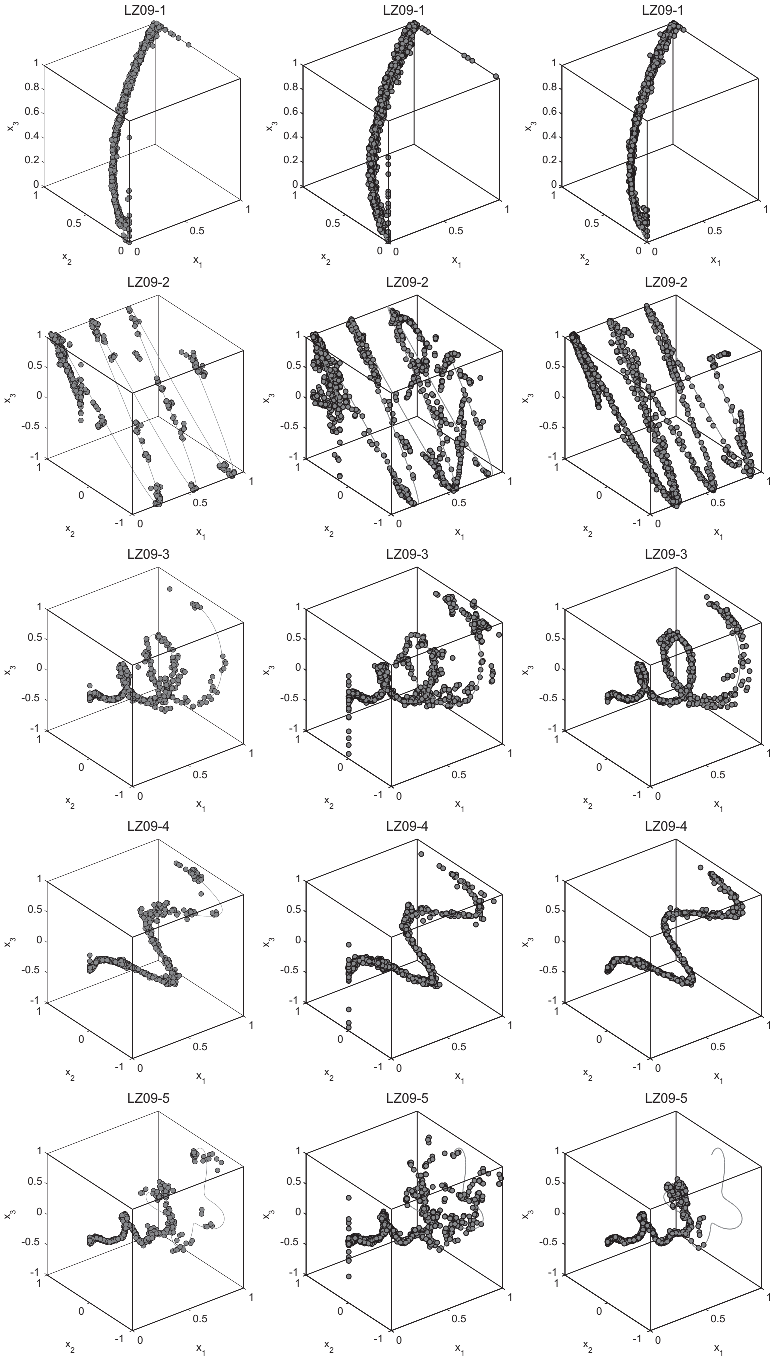

Fig. 4. Plots of all 10 populations found by $(\mu + 1)$-MO-CMA (left), $(\mu +_{10} 1)$-MO-CMA (center) and $(\mu + 1_{rank})$-MO-CMA (right) in the $x_1 - x_2 - x_3$ space on LZ09-1:5 after 100,000 function evaluations.

Table 1. Comparative results of two baseline EMOAs, namely generational and steady-state MO-CMA-ES and several versions of steady-state MO-CMA-ES with different parent selection schemes. Median number of function evaluations (out of 31 independent runs) to reach ΔHtarget values, normalized by Best: a value of 1 indicates the best result, a value $X > 1$ indicates that the corresponding algorithm needed X times more evaluations than the best to reach the same precision.

ΔHtarget	1	0.1	0.01	1	0.1	0.01	1	0.1	0.01	1	0.1	0.01
	sZDT1			sZDT2			sZDT3			sZDT6		
Best	2500	12000	47000	2500	15000	59000	3000	18500	70000	4500	141200	.
$(\mu + \mu)$	7.3	4.3	2.2	8.6	4.3	2.1	5.5	3	1.5	8.2	**1**	.
$(\mu + 1)$	6.5	3.9	2.1	7.6	3.9	2	5.1	2.8	1.4	7.2	1.1	.
$(\mu_{\prec} + 1)$	1.5	2.2	1.7	**1**	2	1.5	1.3	1.8	1.2	**1**	.	.
$(\mu +_2 1)$	3.7	2.4	1.5	4.4	2.4	1.4	3.1	1.7	**1**	4.3	.	.
$(\mu +_{10} 1)$	1.2	**1**	1.1	1.4	**1**	1.3	**1**	**1**	.	1.3	.	.
$(\mu + 1_{\Delta H_1})$	3.5	1.5	**1**	3.4	1.6	**1**	2.5	.	.	1.7	.	.
$(\mu + 1_{\Delta H_i})$	1.7	1.3	**1**	1.8	1.3	**1**	1.1	.	.	1.5	.	.
$(\mu + 1_{succ})$	1.2	1.7	1.1	1.6	.	.	**1**	.	.	2.1	.	.
$(\mu + 1_{rank})$	**1**	1.4	**1**	1.4	.	.	**1**	.	.	1.7	.	.
	IHR1			IHR2			IHR3			IHR6		
Best	500	1500	6000	1500	4000	8500	1000	.	.	6000	.	.
$(\mu + \mu)$	8.4	8.8	6.9	6.4	4.8	3.3	8.2	.	.	5.6	.	.
$(\mu + 1)$	7	7.3	6.7	5.6	4.1	2.9	7	.	.	5	.	.
$(\mu_{\prec} + 1)$	**1**	**1**	3	**1**	1.6	1.7	**1**	.	.	**1**	.	.
$(\mu +_2 1)$	4	4.3	4	3.3	2.5	1.9	4	.	.	3	.	.
$(\mu +_{10} 1)$	2	1.6	1.1	**1**	**1**	**1**	**1**	.	.	**1**	.	.
$(\mu + 1_{\Delta H_1})$	2	1.6	**1**	2	1.5	1.2	2.5	.	.	1.4	.	.
$(\mu + 1_{\Delta H_i})$	2	2.3	**1**	1.3	1.3	1.1	1.5	.	.	1.2	.	.
$(\mu + 1_{succ})$	2	2.3	2	5.3	2.7	1.7	1.5	.	.	1.9	.	.
$(\mu + 1_{rank})$	2	2	1.5	1.6	1.7	1.3	1.5	.	.	1.6	.	.
	LZ09-1			LZ09-2			LZ09-3			LZ09-4		
Best	500	6000	17000	3500	144000	.	1500	35000	120500	1000	10000	40500
$(\mu + \mu)$	11.4	5.1	3.2	3.6	.	.	4.1	1.2	.	5.7	3.2	2.4
$(\mu + 1)$	9	4.7	3	3.2	.	.	3.6	**1**	.	5	3.9	2.5
$(\mu_{\prec} + 1)$	2	2.5	2.2	**1**	.	.	**1**	.	.	**1**	4.3	2.3
$(\mu +_2 1)$	6	2.8	1.9	2.2	.	.	2.3	1.7	.	3.5	2.7	1.9
$(\mu +_{10} 1)$	2	**1**	**1**	**1**	.	.	**1**	1.4	1.4	1.5	**1**	2
$(\mu + 1_{\Delta H_1})$	9	2.1	1.5	2	.	.	1.6	5.6	.	2	1.8	**1**
$(\mu + 1_{\Delta H_i})$	2	1.5	1.3	2.1	**1**	.	2	4.2	.	2.5	1.5	**1**
$(\mu + 1_{succ})$	**1**	2.1	1.4	3.5	.	.	1.3	3.6	.	2	3.5	1.3
$(\mu + 1_{rank})$	**1**	1.9	1.3	5.8	1.1	.	1.3	1.6	**1**	1.5	2.4	**1**
	LZ09-5											
Best	1500	19000	.									
$(\mu + \mu)$	3.4	1.6	.									
$(\mu + 1)$	3.3	1.4	.									
$(\mu_{\prec} + 1)$	**1**	.	.									
$(\mu +_2 1)$	2	**1**	.									
$(\mu +_{10} 1)$	**1**	1.7	.									
$(\mu + 1_{\Delta H_1})$	1.3	.	.									
$(\mu + 1_{\Delta H_i})$	1.6	1.9	.									
$(\mu + 1_{succ})$	1.3	.	.									
$(\mu + 1_{rank})$	**1**	.	.									