Benchmarking Multi- and Many-objective Evolutionary Algorithms under Two Optimization Scenarios

Ryoji Tanabe, *Member, IEEE,* and Hisao Ishibuchi, *Fellow, IEEE* and Akira Oyama, *Member, IEEE,*

***Abstract*—Recently, a large number of multi-objective evolutionary algorithms (MOEAs) for many-objective optimization problems (MaOPs) have been proposed in the evolutionary computation community. However, an exhaustive benchmarking study has never been performed. As a result, the performance of MOEAs has not been well understood, yet. Moreover, in almost all previous studies, the performance of MOEAs was evaluated based on nondominated solutions in the final population at the end of the search. Such traditional benchmarking methodology has several critical issues. In this paper, we exhaustively investigate the anytime performance of 21 MOEAs using an Unbounded External Archive (UEA), which stores all nondominated solutions found during the search process. Each MOEA is evaluated under two optimization scenarios called UEA and reduced UEA in addition to the standard final population scenario. These two scenarios are more practical in real-world applications than the final population scenario. Experimental results obtained under the two scenarios are significantly different from the previously reported results under the final population scenario. For example, results on the WFG test problems with up to six objectives indicate that some recently proposed MOEAs are outperformed by some classical MOEAs. We also analyze the reason why some classical MOEAs work well under the UEA and reduced UEA scenarios.**

***Index Terms*—Multi-objective optimization, evolutionary multi-objective optimization, benchmarking study**

I. Introduction

A multi-objective optimization problem (MOP) is the problem of minimizing M potentially conflicting objective functions $f_1, ..., f_M$ simultaneously. MOPs frequently appear in engineering problems, such as aerodynamic wing design problems [1], ship parametric design problems [2], oil well problems [3], and unit commitment problems [4]. More formally, an unconstrained (bound-constrained) MOP can be formulated as follows:

$$\begin{aligned} &\text{minimize} \quad \boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), ..., f_M(\boldsymbol{x}))^{\mathrm{T}} \\ &\text{subject to} \quad \boldsymbol{x} \in \mathbb{S} \subseteq \mathbb{R}^D \end{aligned} \tag{1}$$

where, $\boldsymbol{f} : \mathbb{S} \rightarrow \mathbb{R}^M$ is an objective function vector that consists of M objective functions, and $\mathbb{R}^M$ is the objective function space. Here, $\boldsymbol{x} = (x_1, ..., x_D)^{\mathrm{T}}$ is a D-dimensional solution vector, and $\mathbb{S} = \Pi_{j=1}^{D}[x_j^{\min}, x_j^{\max}]$ is the bound-constrained search space where $x_j^{\min} \leq x_j \leq x_j^{\max}$ for each variable index $j \in \{1, ..., D\}$.

We say that $\boldsymbol{x}^1$ dominates $\boldsymbol{x}^2$ and denote $\boldsymbol{x}^1 \prec \boldsymbol{x}^2$ if and only if $f_i(\boldsymbol{x}^1) \leq f_i(\boldsymbol{x}^2)$ for all $i \in \{1, ..., M\}$ and $f_i(\boldsymbol{x}^1) < f_i(\boldsymbol{x}^2)$ for at least one index i. Here, $\boldsymbol{x}^*$ is a Pareto-optimal solution if there exists no $\boldsymbol{x} \in \mathbb{S}$ such that $\boldsymbol{x} \prec \boldsymbol{x}^*$. In this case, $\boldsymbol{f}(\boldsymbol{x}^*)$ is a Pareto-optimal objective function vector. The set of all $\boldsymbol{x}^*$ in $\mathbb{S}$ is the Pareto-optimal solution set (PS), and the set of all $\boldsymbol{f}(\boldsymbol{x}^*)$ is the Pareto front (PF). Usually, one optimal solution that minimizes all objective functions does not exist for MOPs. Thus, the goal of MOPs is to find a set of nondominated solutions that are well-distributed and close to the PF in the objective function space.

A multi-objective evolutionary algorithm (MOEA) is an efficient approach for solving MOPs [5]. Since MOEAs use a set of individuals (solutions of a given problem) for the search, they can find good nondominated solutions in a single run. NSGA-II [6] and SPEA2 [7], proposed in the early 2000s, are representative MOEA algorithms. While classical MOEAs such as NSGA-II and SPEA2 perform relatively well on MOPs with $M \leq 3$, their performance significantly degrades for MOPs with $M \geq 4$ [8], [9]. An MOP with $M \geq 4$ is referred to as a many-objective optimization problem (MaOP). Since MaOPs frequently appear in real-world applications, researchers in the evolutionary computation community have attempted to design novel MOEAs that can handle a large number of objectives in the past few years [5]. Table I shows 21 MOEAs, including recently proposed methods for MaOPs (e.g., NSGA-III [10]) and classical methods which were originally designed only for MOPs with $M \leq 3$ (e.g., NSGA-II [6]).

In order to evaluate the performance of MOEAs, the following three optimization scenarios can be considered:

- **Final population scenario**: All nondominated solutions in the final population are used for the performance assessment, where we assume that the population size is finite (not infinite such as [27]).
- **UEA scenario**: All nondominated solutions in an unbounded external archive (UEA) are used for the performance assessment, where the UEA stores all nondominated solutions found during the search process. It should be noted that the UEA can be introduced into all MOEAs with no changes in the original algorithmic framework [28]–[33].
- **Reduced UEA scenario**: A pre-specified number of selected nondominated solutions from the UEA are used for the perfor-

R. Tanabe and H. Ishibuchi are with Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China. e-mail: (rt.ryoji.tanabe@gmail.com, hisao@sustc.edu.cn).

A. Oyama is with Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagamihara, Japan. e-mail: (oyama@flab.isas.jaxa.jp).

TABLE I: Properties of 21 MOEAs benchmarked in this paper: C1 Pareto-dominance based MOEAs, C2 relaxed-dominance based MOEAs, C3 decomposition based MOEAs, C4 indicator based MOEAs, and C5 reference-vector based MOEAs. We also show $\mathrm{FEvals}^{\max}$ used in each paper and its publication year. MOEA/D-07 and MOEA/D-09 correspond to the original names "MOEA/D" and "MOEA/D-DE", respectively.

MOEAs	C1	C2	C3	C4	C5	Year	$\mathrm{FEvals}^{\max}$
NSGA-II [6]	✓					2002	2.5×10^4
NSGA-III [10]	✓				✓	2014	5.5×10^5
SPEA2 [7]	✓					2001	1.0×10^6
SPEA2+SDE [11]	✓					2014	1.0×10^5
GrEA [12]		✓				2013	1.0×10^5
VaEA [13]	✓				✓	2016	5.5×10^5
θ-DEA [14]	✓		✓		✓	2016	5.5×10^5
MOEA/D-07 [15]			✓			2007	7.5×10^4
MOEA/D-09 [16]			✓			2009	3.0×10^5
MOEA/D-DRA [17]			✓			2009	3.0×10^5
MOEA/D-STM [18]			✓			2014	3.0×10^5
MOEA/DD [19]	✓		✓			2015	5.5×10^5
MOEA/D-DU [20]			✓			2016	2.6×10^5
I-DBEA [21]	✓		✓			2015	1.4×10^6
EFR-RR [20]			✓			2016	2.6×10^5
RVEA [22]			✓		✓	2016	2.8×10^5
$\mathrm{IBEA}_{\epsilon+}$ [23]				✓		2004	2.0×10^4
IBEA_{HD} [23]				✓		2004	2.0×10^4
HypE [24]	✓			✓		2011	1.0×10^4
BiGE [25]	✓			✓		2015	1.0×10^5
MOMBI-II [26]				✓		2015	4.4×10^5

mance assessment. This scenario is to compare MOEAs using the obtained solution sets of the same size (since the size of the UEA obtained by each MOEA is different).

The final population scenario is the most widely used optimization scenario. This scenario is also used in the experimental studies of all the papers listed in Table I. The UEA scenario was not so popular in the evolutionary computation community and used in only some previous studies [28], [29]. However, the UEA scenario comes to be used for the performance evaluation in recent work [30]–[33]. The UEA scenario is also adopted in the recently proposed BBOB-biobj benchmark suite [34] in the COCO framework[1]. As far as we know, the reduced UEA scenario is not so frequently found in previous work but is used in some recent work (e.g., [17], [35], [36]).

A large number of nondominated solutions can be presented for decision makers under the UEA scenario. However, in practice, human users of an MOEA often want to know only a small number of representative, well-distributed nondominated solutions (instead of a large number of nondominated solutions in the UEA). They do not want to examine a large number of nondominated solutions [15]. We do not intend to claim that the UEA scenario is impractical. A large number of nondominated solutions are helpful for visualizing the shape of the entire PF [37] and understanding the properties of an application task [38]. Thus, which optimization scenario is more practical is totally application dependent.

While the final population scenario is frequently used in the evolutionary computation community, it is not always a practically desirable optimization scenario [35]. Under the final population scenario, when the number of nondominated solutions found in the search exceeds the predefined population size, they are removed from the population to keep the population size constant. This operation is undesirable when a user wants to carefully examine the entire PF using a large number of nondominated solutions. Moreover, a good potential solution found during the search process is likely to be discarded from the population [35]. Such a problem is easily addressed by using the UEA. There is no particular reason not to incorporate the UEA into MOEAs (especially when the computational cost for archive maintenance is very small in comparison with the objective function evaluation of each solution as in many real-world application problems). Also, as pointed out in [39], when the hypervolume (HV) [40] and Inverted Generational Distance (IGD) [41] indicators are used for the performance assessment under the final population scenario, a large population size is always more beneficial than a small one. This indicates that the fair comparison of MOEAs with different population sizes is not accomplished under the final population scenario.

In addition to the above-mentioned issue related to the population size, the final population scenario has another issue: timing of evaluation. In most studies in Table I, the performance of MOEAs was evaluated based on the end-of-the-run results. More specifically, suppose the performance comparison between n MOEAs $\{\mathrm{MOEA}_1, ..., \mathrm{MOEA}_n\}$ on a specific problem instance I. In the traditional benchmarking methodology (i.e., the final population scenario), the performance rank is determined as follows: For each $\mathrm{MOEA}_i, i \in \{1, ..., n\}$, an independent run of MOEA_i on I is performed until reaching the maximum number of function evaluations ($\mathrm{FEvals}^{\max}$). At the end of the search, the final population $\boldsymbol{P}^{\mathrm{final},i}$ is recorded as a set of obtained nondominated solutions. By repeating this procedure T times, $\boldsymbol{P}^{\mathrm{final},i,1}, ..., \boldsymbol{P}^{\mathrm{final},i,T}$ are obtained. Then, quality of nondominated solutions in each final population is calculated using arbitrary indicators such as HV and IGD. Finally, $\{\mathrm{MOEA}_1, ..., \mathrm{MOEA}_n\}$ are ranked based on a representative value (e.g., mean and median) of the obtained T indicator values. For instance, if the mean HV value of MOEA_1 is higher than that of MOEA_2, then MOEA_1 is evaluated as being better than MOEA_2.

Note that the performance rank based on the above traditional benchmarking procedure strongly depends on the setting of $\mathrm{FEvals}^{\max}$. For example, the conclusion "MOEA_1 performs better than MOEA_2" obtained by an experiment with $\mathrm{FEvals}^{\max} = 7 \times 10^5$ cannot be applied when $\mathrm{FEvals}^{\max}$ is set to 2×10^4. Since compared results only at $\mathrm{FEvals}^{\max}$ are usually reported, it is impossible to discuss the anytime performance of MOEAs. While a well-performing MOEA

[1] http://coco.gforge.inria.fr/

should be able to return a set of good nondominated solutions to a user at any time, such anytime performance cannot be measured using the end-of-the-run results. Thus, as pointed out in [32], the comparison based on the end-of-the-run results does not provide sufficient information about the performance of MOEAs.

As far as we know, there is no common setting of $\mathrm{FEvals}^{\max}$ for the final population scenario in the evolutionary computation community, except for some competitions (e.g., [42], [43]). Thus, researchers in the community set $\mathrm{FEvals}^{\max}$ to an arbitrary number[2]. Table I shows the $\mathrm{FEvals}^{\max}$ used in each study. $\mathrm{FEvals}^{\max}$ values differ significantly from each other. For example, $\mathrm{FEvals}^{\max}$ was set to 1×10^4 in the HypE paper [24], whereas it was set to 5.5×10^5 in the recent studies [10], [13], [14], [19]. Some previous studies (e.g., [11], [12], [14], [18], [19]) have reported the poor performance of HypE. However, since HypE was originally designed for optimization with $\mathrm{FEvals}^{\max} = 1 \times 10^4$, performance comparison with a larger $\mathrm{FEvals}^{\max}$ may be unfair.

As discussed above, the traditional methodology for benchmarking MOEAs (i.e., the final population scenario) has several issues. Taking into account the issues of the final population scenario, in this paper, we exhaustively investigate the anytime performance of the 21 MOEAs in Table I on MaOPs with up to six objectives using the the UEA and reduced UEA scenarios. As far as we know, this study is the first attempt to benchmark a number of MOEAs for MaOPs using the the UEA and reduced UEA scenarios. While a large number of MOEAs for MaOPs have been proposed in 2014 – 2016 as shown in Table I, an exhaustive benchmarking study has never been performed. Thus, the best MOEA is unknown. In practice, users want to apply a high-performance MOEA to their real-world problems to obtain well-approximated nondominated solutions. Algorithm designers develop novel, efficient MOEAs by analyzing and improving well-performing MOEAs. Therefore, the situation that the best MOEA is unknown is undesirable for both users and designers of MOEAs.

There are some previous studies on benchmarking MOEAs for MaOPs. Wagner et al. evaluated the performance of eight MOEAs on the DTLZ1 and DTLZ2 functions with up to six objectives [8]. Eight MOEAs were also compared for various problems, including multi-objective TSPs, with up to ten objectives in [46]. Very recently, the scalability of nine MOEAs with respect to the number of objectives and variables was investigated in [47]. Also, Bezerra et al. presented a large scale benchmarking study of nine MOEAs with tuned control parameter settings on the DTLZ and WFG problems with up to 10 objectives [48]. However, in the above-mentioned studies, the performance of MOEAs was measured mainly based on the final population scenario. Thus, the anytime performance of MOEAs was not much discussed in the previous studies. In addition, considered MOEAs did not include recently proposed state-of-the-art methods, and the number of algorithms for comparison was limited up to nine.

While the final population scenario has been widely used in the literature, some recent work (e.g., [30]–[33], [35], [36], [49]) tried to benchmark MOEAs under the UEA and reduced UEA scenarios. However, the previous studies [33], [36], [49] reported only the end-of-the-run results and did not evaluate the anytime performance of MOEAs. Although some previous work (e.g., [44], [50], [51]) evaluated the anytime performance of MOEAs, these studies were based on the population at each iteration.

In [32], Brockhoff et al. established the data profile-based methodology for visualizing the anytime performance of MOEAs. Their work was extended to the recently designed BBOB-biobj functions [34] in the COCO platform. At the GECCO Workshop on Real-Parameter Black-Box Optimization Benchmarking (BBOB 2016), the performance of 16 optimization algorithms were evaluated on the BBOB-biobj functions. Experimental data can be downloaded from the workshop website[3]. Such a well-established and systematic methodology allows researchers to design and benchmark new algorithms constructively. Unfortunately, COCO *currently* provides only two-objective MOPs, and thus the performance of MOEAs for MOPs with $M \geq 3$ cannot be evaluated using the current COCO platform. Bringmann et al. evaluated the anytime performance of four MOEAs under the reduced UEA scenario [35]. However, the performance of MOEAs was investigated only on two-objective MOPs in their study.

This paper is an extended version of an earlier conference paper [52]. The differences between the previous conference paper and this journal paper are as follows: (1) While the anytime performance of MOEAs was measured only under the UEA scenario in the conference paper, we evaluate it under both the UEA and reduced UEA scenarios in this paper. By comparing the results obtained under both scenarios, more important information about the performance of MOEAs is provided. (2) Analysis of the performance of MOEAs is significantly improved in this paper. For example, while only the HV indicator was used in the conference paper, various indicators are used to analyze the performance of MOEAs in this paper. (3) Classical and recently proposed MOEAs are reviewed in this paper to understand the properties of the examined MOEAs in this paper.

The remainder of this paper is organized as follows. Section II reviews classical and advanced MOEAs. We introduce experimental settings for our benchmarking study in Section III. Section IV describes experimental results. We discuss the experimental results and some previous work in Section V. Finally, we conclude this paper in Section VI.

II. Review of classical and advanced MOEAs

MOEAs can be roughly classified into the following three categories: (1) dominance based MOEAs, (2) decomposition based MOEAs, and (3) indicator based MOEAs. In this section, we review classical and recently proposed MOEAs in each category in Subsection II-A, II-B, and II-C, respectively.

[2]We do not have an intention to claim that $\mathrm{FEvals}^{\max}$ can be set to an arbitrary number for *all possible problems*. Usually, $\mathrm{FEvals}^{\max}$ cannot be set to an arbitrary number in practice. Since some real-world problems require the execution of a simulation that takes a long time to evaluate the solution, $\mathrm{FEvals}^{\max}$ is dependent on user's available time [44], [45].

[3]http://coco.gforge.inria.fr/doku.php?id=algorithms-biobj

A. Dominance based MOEAs

NSGA-II [6] and SPEA2 [7] are well-known and frequently-used dominance based MOEAs. Both MOEA algorithms commonly use (1) the Pareto dominance relationship and (2) the distance-based density estimation in the objective function space for the mating and environmental selections. The Pareto dominance based selection (1) enhances the selection pressure towards the PF and is used as a primary criterion. The density based selection (2) promotes the diversity of individuals in the population and is used as a secondary criterion when compared individuals have the same rank based on the first criterion (1).

NSGA-II and SPEA2 perform poorly on MOPs with more than four objectives (i.e., MaOPs) [8], [9]. The main reason for their poor performance on MaOPs is a weak selection pressure towards the PF. Since almost all individuals in the population are nondominated with each other for MaOPs, the Pareto dominance based selection (1) does not work well [8], [9]. As a result, the selection is performed mainly based on the density based selection (2), which prevents the convergence of the population to the PF. This phenomenon is called Active Diversity Promotion (ADP) and the main reason for the poor performance of classical dominance based MOEAs, such as NSGA-II and SPEA2 [13], [53].

In order to address this ADP problem, some improved density based selection techniques have been proposed. Wagner et al. proposed an alternative NSGA-II algorithm that assigns the zero crowding distance value to extreme individuals in the population to encourage the convergence toward the center of the PF [8]. Köppen and Yoshida proposed a new crowding distance assignment strategy that assigns a high distance value to an individual near the nondominated individuals in the objective function space [54]. In the proposed method in [55], when the spread indicator [41] value of individuals in the population exceeds a value of one, the procedure of the crowding distance-based selection is not done in this iteration. Li et al. proposed Shift-based Density Estimation (SDE) strategy [11], where large crowding distance values are assigned to poorly converged individuals by virtually shifting them to the crowding region. Cheng et al. proposed an improved version of mating and environmental selection methods based on a scalarizing function for NSGA-II [56]. Knee point-based mating and environmental selection strategies are also proposed in [57].

Recently, some studies have proposed improved variants of dominance based MOEAs for MaOPs that use reference vectors-based niching selection, instead of the crowding distance-based selection. After the generation of the new solutions, all individuals in a union of the population and children are associated with each reference vector. Then, the environmental selection is performed by each niche. Such typical examples are NSGA-III [10], Vector angle based EA (VaEA) [13], and SPEA/R [58].

Various types of new relaxed-dominance relationships $\prec_{\text{relaxed}}$ have been designed to enhance the convergence ability of dominance based MOEAs toward the PF. The relaxed-dominance relationship $\prec_{\text{relaxed}}$ is used for the mating and/or environmental selections, instead of the exact Pareto-dominance relationship $\prec$. The representative relaxed-dominance relationships are α-dominance [59], ϵ-dominance [60], ranking-dominance [61], CDAS-dominance [62], grid-dominance [12], and θ-dominance [14]. In particular, recent studies show that MOEAs with the grid-dominance and the θ-dominance perform well [12], [14].

In the grid-dominance [12], the objective function space is divided into div^M hyperboxes based on individuals in the population, where the number of divisions div is the control parameter. Then, each individual is assigned to one hyperbox. A dominance relationship between individuals is decided based on the position of the hyperbox that the individual is assigned. Grid-based EA (GrEA) [12] use three criteria (grid ranking, grid crowding distance, and grid coordinate point distance) based on the grid-dominance described here for assigning fitness values to the individuals. In the θ-dominance [14], first, individuals in the population are grouped according to the reference-vector based niching procedure. Then, individuals in each group are ranked based on the values of the PBI function of MOEA/D [15] (see Section II-B).

B. Decomposition based MOEAs

A decomposition based MOEA decomposes a given MOP with M objectives into N single-objective sub-problems[4] and tries to find good solutions for all the subproblems. Representative decomposition based MOEAs are Multi-Objective Genetic Local Search (MOGLS) [64], Cellular Multi-Objective GA (C-MOGA) [65], Multiple Single-Objective Pareto Sampling (MSOPS) [66], [67], and MOEA based on Decomposition (MOEA/D) [15]. Since recent studies have reported the promising performance of MOEA/D-type algorithms [68], we here mainly review them.

MOEA/D [15] decomposes an MOP with M objectives into μ single-objective sub-problems $g(\boldsymbol{x}|\boldsymbol{w}_1), ..., g(\boldsymbol{x}|\boldsymbol{w}_\mu)$ using a set of weight vectors $\boldsymbol{W} = \{\boldsymbol{w}^1, ..., \boldsymbol{w}^\mu\}$ and a scalarizing function $g(\boldsymbol{x}|\boldsymbol{w})$, where $\boldsymbol{w} = (w_1, ..., w_M)^{\mathrm{T}}, \sum_{j=1}^M w_j = 1$. Each individual $\boldsymbol{x}^i, i \in \{1, ..., \mu\}$ in the population $\boldsymbol{P} = \{\boldsymbol{x}^1, ..., \boldsymbol{x}^\mu\}$ has own weight vector $\boldsymbol{w}^i$. MOEA/D tries to find the optimal solution of all subproblems simultaneously. While the parent individuals are selected from the whole population in general MOEAs, they are selected from a set of neighborhood individuals of each weight vector $\boldsymbol{w}^i$ in MOEA/D.

Below, improved variants of MOEA/D are described. We call the original MOEA/D proposed in [15] MOEA/D-07 for clarity. Most studies that attempt to improve the performance of MOEA/D focus on (1) how to select an individual to be evolved, (2) how to select parents to generate a new individual, (3) how to select individuals to be compared with a newly generated one, and (4) how to specify/select/adjust a scalarizing function g.

For (2) and (3), Li and Zhang proposed MOEA/D-DE [16], which is a modified version of MOEA/D-07. The differences between MOEA/D-DE and MOEA/D-07 are as follows: (i) the parent individuals are selected from the whole population

[4]Some decomposition based MOEAs (e.g., MOEA/D-M2M [63]) decompose a given MOP into a set of simple MOPs.

with some probability, (ii) the maximum number of replaced individuals by a child is introduced, and (iii) variation operators of Differential Evolution (DE) [69] are used. It is worth mentioning that the two algorithmic components (i) and (ii) of MOEA/D-DE are introduced into most of improved MOEA/D algorithms described below. In this paper, we call an MOEA/D-DE algorithm extracted (iii) MOEA/D-09.

For (1), an evolved individual for each iteration is selected according to the lexicographic order in MOEA/D. For this reason, unnecessary computational resource may be allocated to a well-optimized subproblem. Zhang et al. proposed MOEA/D with Dynamical Resource Allocation (MOEA/D-DRA) [17], which adaptively selects a subproblem to be evolved based on improvement rates of scalarizing function values. Its improved version, called MOEA/D with Generalized Resource Allocation (MOEA/D-GRA) [70], was also proposed.

For (3), while replaced individuals are selected from only the neighborhood individuals in the most of MOEA/D algorithms, this restriction is removed in MOEA/D with Stable Matching model (MOEA/D-STM) [18]. In the environmental selection of MOEA/D-STM, while each subproblem prefers an individual with a good scalarizing function value, each individual prefers a subproblem whose weight vector $\boldsymbol{w}$ is closed to $\boldsymbol{f}(\boldsymbol{x})$ in the objective function space. In MOEA/D with Adaptive Global Replacement (MOEA/D-AGR) [71], a child $\boldsymbol{u}$ is assigned to a subproblem j that minimizes the scalarizing function value $g(\boldsymbol{u}|\boldsymbol{w}_j)$. Then, an individual that has already been assigned to the subproblem j and its neighborhood individuals are replaced by the child $\boldsymbol{u}$ if their scalarizing function values are worse than that of $\boldsymbol{u}$. In MOEA/D-DU [20], once the child $\boldsymbol{u}$ is generated, the perpendicular distance between $\boldsymbol{f}(\boldsymbol{u})$ and each weight vector in $\boldsymbol{W}$ is calculated. Then, the replacement occurs only for K closed subproblems, where K is the control parameter of MOEA/D-DU. In addition to the MOEA/D algorithms mentioned above, some approaches for (3) have been proposed (e.g., [72]–[74]).

In addition to the MOEA/D variants mentioned above, various improved MOEA/D algorithms have been proposed, such as MOEA/DD [19] and Improved Decomposition-based EA (I-DBEA) [21], which use the Pareto-dominance principle. An adaptive neighborhood size control strategy [75] and a special neighborhood representation based on the distance in the solution space [76] were also proposed. MOEA/D uses a scalarizing function g to decompose an MOP into multiple sub problems. Thus, the performance of MOEA/D algorithms significantly depends on the type of scalarizing functions [77]. Typical functions are the weighted sum, Tchebycheff, Penalty-based Boundary Intersection (PBI) functions [15]. Some recent studies proposed improved variants of scalarizing functions (e.g., [18], [77]–[80]).

In addition to MOEA/D variants described in this section, other decomposition based MOEAs have also been studied. Among them are Ensemble Fitness Ranking (EFR) [81], EFR with Ranking Restriction (EFR-RR) [20], and Reference Vector guided Evolutionary Algorithm (RVEA) [22].

C. Indicator based MOEAs

Quality indicators are used to quantitatively evaluate a set of nondominated solutions found in the search of an MOEA. An indicator based MOEA, which was proposed by Zitzler and Künzli in 2004, is a framework of MOEAs to use such a quality indicator for multi-objective search [23]. In indicator based MOEAs, individuals in the population are evaluated or ranked based on a predefined indicator. Then, the obtained fitness values of the individuals are used for the mating and environmental selections.

The first indicator based MOEA is Indicator-Based Evolutionary Algorithm (IBEA) [23]. At the beginning of the environmental selection of IBEA, a fitness value $F(\boldsymbol{x})$ of each individual $\boldsymbol{x}$ in a union of the population $\boldsymbol{P}$ and children $\boldsymbol{Q}$ is assigned by a paired comparison of the other individuals in $\boldsymbol{P} \cup \boldsymbol{Q} \backslash \{\boldsymbol{x}\}$ based on quality indicator values. The fitness values also used for the parent selection. Zitzler and Künzli proposed two types of IBEA algorithms, IBEA$_{\epsilon+}$ and IBEA$_{HD}$. IBEA$_{\epsilon+}$ uses the binary additive ϵ-indicator $I_{\epsilon+}$ [41], and IBEA$_{HD}$ uses the hypervolume difference indicator I_{HD}. IBEA is a general framework that any quality indicators can be incorporated into. Various type of IBEAs with different indicators have been proposed such as R2-IBEA [82], Many-Objective Metaheuristic Based on the R2 Indicator (MOMBI) [26], [83], and DDE [84].

One of the critical problems of IBEA$_{\epsilon+}$ is that the distribution of individuals in the population is biased to specific regions of the objective function space [8], [26], [85]. In order to address this issue, Li et al. proposed SRA [85], which simultaneously uses multiple indicators based on the stochastic ranking [86]. In [85], the SRA with $I_{\epsilon+}$ and I_{SDE} performs significantly better than compared MOEAs, where I_{SDE} is an indicator derived from SPEA2+SDE [11]. Bi-Criterion Evolution (BCE) was also proposed to address the same issue of IBEA$_{\epsilon+}$ and MOEA/D [87]. BCE is a general framework to improve the performance of non-dominance based MOEAs by incorporating a finite-sized external archive that maintains evenly distributed nondominated solutions in the objective function space. Two_Arch2 [88], which is an improved version of Two_Arch [89], utilizes a Convergence Archive (CA) and a Diversity Archive (DA) to maintain the diversity of individuals in the population. After the generation of children, the environmental selection is performed by CA and DA, separately. The $I_{\epsilon+}$ indicator and a L_p norm based crowding distance metric are applied to the individuals in CA and DA, respectively.

In general, the goal of MOPs is to obtain well-converged and well-distributed nondominated solutions in the objective function space [90]. Bi-Goal Evolution (BiGE) [25] deals with the two sub-goals of MOPs as two-objective MOPs. The sum of the objective function value $\sum_{i=1}^{M} f_i(\boldsymbol{x})$ and a sharing function-based diversity metric are used for the two objective functions of BiGE. In other words, BiGE translates a given M-objective MOP into a two-objective MOP and tries to find good nondominated solutions in the translated problem.

S-Metric Selection-based Evolutionary Multi-objective Optimization Algorithm (SMS-EMOA) [91] is a steady-state hybrid algorithm using both the Pareto dominance and indicator based selection. In the environmental selection of SMS-EMOA, individuals in a union of the population and the child are ranked based on the nondominated sort of NSGA-II. When the number of individuals in the final front is more than two, their contribution for the HV value is calculated. Then, an individual with the lowest HV contribution is removed from the population. The experimental results in [91] show the promising performance of SMS-EMOA on two- and three-objective MOPs. However, it is difficult to apply SMS-EMOA to MaOPs due to the expensive computational cost for the HV calculation. Hypervolume Estimation Algorithm (HypE) [24] was proposed to address this issue. In HypE, the HV contribution of individuals is estimated by Monte Carlo simulation, instead of the exact HV calculation. The good performance of HypE on MaOPs with up to 50 objectives was reported in [24].

III. Experimental settings

This section describes our experimental settings. Experimental results are reported in Section IV. As mentioned in Section I, we evaluate the anytime performance of the 21 MOEAs in Table I under the two scenarios: UEA (Unbounded External Archive), and reduced UEA. A selection method of a pre-specified number of nondominated solutions from the UEA is necessary for the reduced UEA scenario. First, we introduce the distance based selection method in Subsection III-A. Then, we describe test problems and a performance indicator in Subsection III-B. Finally, the parameter settings for MOEAs are described in Subsection III-C.

A. Solution selection from UEA

Let us assume that we compare different UEAs by selecting the same number of nondominated solutions (say b solutions, $b \geq M$) from each UEA. In the following we explain how to select b solutions from an UEA $\boldsymbol{A}^{\mathrm{UEA}}$ of size a (i.e., with a solutions), where $b < a$ (when $b \geq a$, simply $\boldsymbol{B} \leftarrow \boldsymbol{A}^{\mathrm{UEA}}$). This task can be considered as being a hypervolume subset selection problem (HSSP) [24], [92]. The HSSP is the problem of finding b solutions that maximize the HV indicator value from a set of nondominated solutions with the size a. When the number of objectives M of a given MOP is small, some efficient subset selection methods such as [92]–[94] are available. However, since we deal MaOPs with a large M in this study, such selection methods are unavailable. Since most of existing subset selection methods require iterative HV calculations to find a well-approximated subset, their computational cost becomes impractical for large UEAs of MaOPs with a large M. Fortunately, some computationally cheap selection methods of representative nondominated solutions have been proposed (e.g., [17]). In this study, we used a slightly modified version of the distance based selection method proposed in [17].

Algorithm 1 shows our selection method used in this experimental study. The function $\mathrm{distance}_{\mathrm{E}}(\boldsymbol{x}^1, \boldsymbol{x}^2)$ in Algorithm 1 returns the Euclidean distance between two vectors $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$. First, M extreme solutions that have the minimum objective function value $f_i(\boldsymbol{x})$ for f_i, $i \in \{1, ..., M\}$ are selected (Algorithm 1, line 4–6) and added to $\boldsymbol{B}$. After selecting the M extreme solutions, a solution that has the maximum distance to solutions in $\boldsymbol{B}$ is selected for $\boldsymbol{B}$ (Algorithm 1, line 10–12). By repeatedly adding an isolated solution to $\boldsymbol{B}$ until $|\boldsymbol{B}| = b$, it is expected that a set of uniformly distributed nondominated solutions in the objective function space are obtained.

Algorithm 1: Our selection method of a pre-specified number of nondominated solutions from the UEA

input : A set of a nondominated solutions in the UEA $\boldsymbol{A}^{\mathrm{UEA}}$
output: A set of b uniformly distributed nondominated solutions $\boldsymbol{B}$

`// Initialization`
1 $\boldsymbol{B} \leftarrow \emptyset$;
2 **for** $l \in \{1, ..., a\}$ **do**
3     $V_l \leftarrow$ false, $D_l \leftarrow \infty$;
`// Select` M `extreme solutions`
4 **for** $i \in \{1, ..., M\}$ **do**
5     $j \leftarrow \underset{k \in \{1,...,a\} | V_k = \mathrm{false}}{\mathrm{argmin}} f_i(\boldsymbol{x}^k)$;
6     $\boldsymbol{B} \leftarrow \boldsymbol{B} \cup \{\boldsymbol{x}^j\}$, $V_j \leftarrow$ true;
7   **for** $l \in \{1, ..., a\}$ **do**
8   **if** $V_l =$ false **then**
9             $D_l \leftarrow \min(\mathrm{distance}_{\mathrm{E}}(\boldsymbol{f}(\boldsymbol{x}^l), \boldsymbol{f}(\boldsymbol{x}^j)), D_l)$;
`// Select an isolated solution to` $\boldsymbol{B}$
10 **while** $|\boldsymbol{B}| < b$ **do**
11     $j \leftarrow \underset{k \in \{1,...,a\} | V_k = \mathrm{false}}{\mathrm{argmax}} D_k$;
12     $\boldsymbol{B} \leftarrow \boldsymbol{B} \cup \{\boldsymbol{x}^j\}$, $V_j \leftarrow$ true;
13   **for** $l \in \{1, ..., a\}$ **do**
14   **if** $V_l =$ false **then**
15             $D_l \leftarrow \min(\mathrm{distance}_{\mathrm{E}}(\boldsymbol{f}(\boldsymbol{x}^l), \boldsymbol{f}(\boldsymbol{x}^j)), D_l)$;

B. Test problems and performance indicator

In our benchmarking study, we used the nine WFG test problems [95] with $M \in \{2, 3, 4, 5, 6\}$. Table II shows the properties of the WFG test problems. The shapes of the PFs of the WFG1, WFG2, and WFG3 test problems are complicated, discontinuous, and partially degenerate respectively. The WFG4, ..., WFG9 problems have nonconvex PFs. Note that a part of the PF of the WFG3 problem is non-degenerate [96]. The WFG4 and WFG9 test problems are multimodal. The objective functions in the WFG2, WFG6, WFG8, and WFG9 problems are also nonseparable. The WFG5 and WFG9 problems are deceptive problems. As suggested in [95], the position parameter k was set to $k = 2\,(M-1)$, and the distance parameter l was set to $l = 20$, where the number of variables is $k + l$.

On all the WFG test problems, the scale of each objective function is different. In general, an MOEA with an efficient normalization strategy for handling differently scaled objective function values performs well on such scaled MOPs [10]. Since objective function values of many real-world problems are differently scaled, normalization strategies are mandatory in practice. Therefore, MOEAs with efficient normalization strategies have been proposed, such as NSGA-III [10],

TABLE II: Properties of the WFG test problems.

Problem	Shape of PF	Multimodality	Separability	Others
WFG1	Mixed		✓	Biased
WFG2	Discontinuous	✓		
WFG3	Partially Degenerate			
WFG4	Nonconvex	✓	✓	
WFG5	Nonconvex		✓	Deceptive
WFG6	Nonconvex			
WFG7	Nonconvex		✓	Biased
WFG8	Nonconvex			Biased
WFG9	Nonconvex			Deceptive, Biased

TABLE III: Websites that we downloaded the source codes of the corresponding MOEAs.

MOEA	Download site
SPEA2, NSGA-II, IBEA$_{HD}$, MOEA/D-09, and MOEA/D-DRA	jMetal 4.5[5]
RVEA and I-DBEA	MOEA Framework 2.11[6]
MOEA/D-STM and MOEA/DD	Ke Li's website [7]
NSGA-III, θ-DEA, MOEA/D-DU, and EFR-RR	Yao's website[8]
HypE	PISA[9]
GrEA, SPEA2+SDE, and BiGE	Miqing Li's website [10]
MOMBI-II	Gómez's website [11]
VaEA	Yi's website[12]

MOMBI-II [26], and, RVEA [22]. However, in the 21 MOEAs, while some MOEAs adopt sophisticated normalization strategies, the others do not use any normalization strategies. Since we want to evaluate the performance of MOEAs as fair as possible, we used the normalized WFG test problems to remove the effect of the normalization strategies as [20], [97]. For the normalized WFG test problems, each objective function value $f_i(\boldsymbol{x}), i \in \{1, ..., M\}$ was normalized using the ideal point $(0, ..., 0)^{\mathrm{T}}$ and the nadir point $(2, ..., 2M)^{\mathrm{T}}$ as in [20], [97]. As suggested in [20], [97], the reference point for calculating the HV value was set to $(1.1, ..., 1.1)^{\mathrm{T}}$. Note that the nadir point of the normalized WFG test problems is $(1, ..., 1)^{\mathrm{T}}$. In this setting, the HV range for all of the WFG test problems is $[0, 1.1^M]$. We further normalized HV values $\in [0, 1.1^M]$ to the range $[0, 1]$ by divided by 1.1^M.

A normalization strategy of an MOEA is usually a "removable", "attachable", and "exchangeable" algorithmic component. For example, the normalization strategy of NSGA-III can be treated as an independent algorithmic component and added to other MOEAs (e.g., MOEA/D). Benchmarking on normalized test problems like this experimental study can be considered to be the performance evaluation of MOEAs with the perfect normalization strategy. A benchmarking study of various normalization strategies should be our future work.

The computational cost of the HV calculation exponentially increases both with the number of objectives M and the number of nondominated solutions in the UEA $|\boldsymbol{A}^{\mathrm{UEA}}|$. In the COCO software with the two-objective BBOB-biobj functions [34], when a newly generated solution enters the UEA $\boldsymbol{A}^{\mathrm{UEA}}$, the HV value of the nondominated solutions in $\boldsymbol{A}^{\mathrm{UEA}}$ is recalculated immediately. However, for MaOPs with $M \geq 4$, since the computational cost of the HV calculation is very high, the on-the-fly HV calculation like the COCO software with the two-objective BBOB-biobj functions is almost impossible. In order to address this issue, in this paper the HV value of the nondominated solutions in $\boldsymbol{A}^{\mathrm{UEA}}$ was calculated only in every $1\,000$ function evaluations, i.e., when FEvals $\in \{1\,000, 2\,000, ..., 49\,000, 50\,000\}$.

C. Parameter settings for MOEAs

We used the SBX crossover and the polynomial mutation for all 21 MOEAs. As suggested in [10], we set the control parameters of the variation operators as follows: $p_c = 1.0$, $\eta_c = 30$, $p_m = 1/D$, and $\eta_m = 20$. For most of the 21 MOEAs, we used the source codes downloaded from websites described in Table III. We replaced the DE operator with the SBX crossover in MOEA/D-DE (MOEA/D-09), MOEA/D-DRA, and MOEA/D-STM to remove the effect of variation operators (i.e., the three MOEAs use the SBX crossover and the polynomial mutation). A slightly modified source code of MOEA/D-09 was used for MOEA/D-07. While buggy codes had been used for the performance evaluation of IBEA$_{HD}$ so far [98], the bug-fixed jMetal source code of IBEA$_{HD}$ was used in this study. We implemented IBEA$_{\epsilon+}$ by modifying the source code of IBEA$_{HD}$ in reference to the PISA code.

For each $M \in \{2, 3, 4, 5, 6\}$, we set the population size μ to 100, 91, 220, 210, and 182 respectively for all MOEAs, except for GrEA, BiGE, SPEA2+SDE, and MOMBI-II. Since μ in the source code of the four exceptional MOEAs must be set as a multiple of four, μ was set to 100, 92, 220, 212, and 184 for each M, respectively. The μ values used in this paper have been widely used in previous studies such as [10], [19]–[21]. For each $M \in \{2, 3, 4, 5, 6\}$, the number of solutions b (see Section III-A) to be selected from the UEA for the HV calculation was set to $100, 91, 220, 210$, and 182 respectively.

For the decomposition and reference vector based MOEAs, weight/reference vectors were generated using the simplex-lattice design (only for $M = 6$, its two-layered version [10] was used). For GrEA, for each M, we set the div value as suggested in [20], [22]. For HypE, the upper bound defining the reference point and the number of samples were set to 2 and 10^4, respectively. The maximum number of

[5] http://jmetal.sourceforge.net/
[6] http://moeaframework.org/index.html
[7] http://www.cs.bham.ac.uk/~likw/publications.html
[8] http://www.cs.bham.ac.uk/~xin/journal_papers.html
[9] http://www.tik.ee.ethz.ch/sop/download/supplementary/hype/
[10] http://www.cs.bham.ac.uk/~limx/
[11] https://www.cs.cinvestav.mx/~EVOCINV/software/MOMBI-II/MOMBI-II.html
[12] https://www.researchgate.net/profile/Xiang_Yi9/publications

function evaluations (FEvals$^{\max}$) was set to 5×10^4 for all test problems, and 21 independent runs were performed.

The performance of MOEAs depends on control parameter settings and can be improved by parameter tuning [31]. For example, the parameter study in [99] reveals that a simple parameter tuning significantly improves the performance of NSGA-III. The parameter study of the population size in [36] also shows that the suitable population size is dependent on the type of MOEAs. For more fair comparison, we should run MOEAs with individually tuned control parameters. However, tuning control parameter settings of all the 21 MOEAs for MaOPs is not a trivial task and beyond the scope of this paper. Its investigation is our future direction.

IV. Experimental results

In this section, we report our experimental results by the 21 MOEAs in Table I. Here, we explain overall performance of each MOEA over all WFG test problems. Our results on each WFG test problem are reported in Appendix. We discuss the performance of the 21 MOEAs in Section V.

A. Overall performance

We describe the overall performance of the 21 MOEAs on all nine WFG problems with each M. For this comparison, we used the average performance score (APS) [24]. Suppose that n algorithms $\{A_1, ..., A_n\}$ are compared for a given problem instance using the HV values obtained in multiple runs. For each $i \in \{1, ..., n\}$ and $j \in \{1, ..., n\} \backslash \{i\}$, let $\delta_{i,j} = 1$, if A_j significantly outperforms A_i using the Wilcoxon rank-sum test with $p < 0.05$, otherwise $\delta_{i,j} = 0$. Then, the performance score $P(A_i)$ is defined as follows: $P(A_i) = \sum_{j \in \{1,...,n\} \backslash \{i\}}^{n} \delta_{i,j}$. The score $P(A_i)$ represents the number of algorithms outperforming A_i. The APS is the average of the $P(A_i)$ values for all problem instances. In other words, the APS value of A_i represents how good (relatively) the performance of A_i is among the n algorithms for all problem instances. A small APS value $P(A_i)$ indicates that the performance of A_i is better than other algorithms.

Figure 1 and 2 show the results for the 21 MOEAs for all the nine WFG problems with $M \in \{2, 3, 4, 5, 6\}$ under the UEA and reduced UEA scenarios, respectively. Below, we discuss the results for each algorithm category:

• **Seven Pareto dominance-based MOEAs**: Compared to the results for the other types of MOEAs described below, the performance rank of the seven dominance-based MOEAs is not significantly affected by the choice of an optimization scenario. Therefore, we describe the performance of the seven dominance-based MOEAs based on only the results for the UEA scenario.

Interestingly, in this category, the worst and best performers are SPEA2 and SPEA2+SDE respectively. SPEA2+SDE is an improved version of SPEA2 that incorporates the SDE strategy [11]. Based on these results, the SDE strategy appears to contribute significantly to the outstanding performance of SPEA2+SDE. For almost all M, the convergence speed of NSGA-III is lower than that of NSGA-II. That is, NSGA-II performs better than NSGA-III for smaller FEvals, even when $M = 6$. NSGA-III is an improved version of NSGA-II for MaOPs by replacing the crowding distance-based selection with the reference vectors-based niching selection. The newly introduced niching selection appears to make the convergence speed of NSGA-III slow. The two relaxed-dominance based MOEAs (GrEA and θ-DEA) perform similarly to each other for the WFG problems with 2–6 objectives.

• **Nine decomposition-based MOEAs**: First, we describe the results for the UEA scenario. For $M = 2$, MOEA/D-STM clearly performs better than the other decomposition-based MOEAs. However, within FEvals $= 2 \times 10^3$, the classical MOEA/D-07 outperforms its improved variants. For $M \in \{4, 5, 6\}$, RVEA performs best for much smaller FEvals, followed by MOEA/D-07. For $M = 4$, in an early stage of evolution, MOEA/D-DRA performs very well, but MOEA/D-DU outperforms MOEA/D-DRA after FEvals $= 4.3 \times 10^4$. For $M = 6$, the best performer is also MOEA/D-DRA, followed by MOEA/D-STM and MOEA/D-09. I-DBEA and MOEA/DD clearly perform worse than the other decomposition-based MOEAs including MOEA/D-07. In this category, only I-DBEA and MOEA/DD use the Pareto-dominance for the selection, and it might cause their poor performance. However, with increasing M, the APS value of I-DBEA decreases gradually. Thus, it is possible that I-DBEA performs well for MaOPs with a large number of objectives.

Next, we describe the results for the reduced UEA scenario. Interestingly, the APS values of some MOEA/D variants (MOEA/D-07, MOEA/D-09, and MOEA/D-DRA) are improved, while those of MOEA/D-DU and EFR-RR are deteriorated. In particular, the APS values of MOEA/D-DRA is significantly improved, and MOEA/D-DRA is the best performer among the 21 MOEAs after FEvals $= 2.3 \times 10^4$ for $M = 5$.

• **Five indicator-based MOEAs**:

First, the results for the UEA scenario are described. In this category, MOMBI-II performs poorly, but it outperforms some MOEAs in the other categories. For $M \in \{2, 3, 4, 5\}$, the best performer is clearly IBEA$_{\epsilon+}$, followed by IBEA$_{HD}$. For $M = 6$, HypE performs very well within FEvals $= 1 \times 10^4$. However, as the search progresses, the APS value of HypE gradually deteriorates for all M. In contrast to the results of HypE, the APS value of BiGE is gradually improved as the search progresses. For FEvals $> 1 \times 10^4$, IBEA$_{\epsilon+}$ outperforms the four remaining indicator-based MOEAs.

Next, we describe the results for the reduced UEA scenario. The degradation speed of the APS values of HypE is slower than that under the UEA scenario for $M \in \{4, 5, 6\}$. Although IBEA$_{\epsilon}$ catches up with HypE for FEvals $= 3 \times 10^4$, HypE performs very well for $M = 6$. The APS values of BiGE is also improved for a larger number of objectives compared to those under the UEA scenario. Taking into account the results, when M is increasing, HypE and BiGE might perform well for a smaller and larger FEvals, respectively.

• **All 21 MOEAs**:

For the UEA scenario, the best performer is IBEA$_{\epsilon+}$, followed by IBEA$_{HD}$ for $M \in \{2, 3\}$. For $M \in \{4, 5, 6\}$, RVEA performs best for much smaller FEvals. For $M = 4$, IBEA$_{\epsilon+}$ still outperforms the other MOEAs. For $M \in \{5, 6\}$,

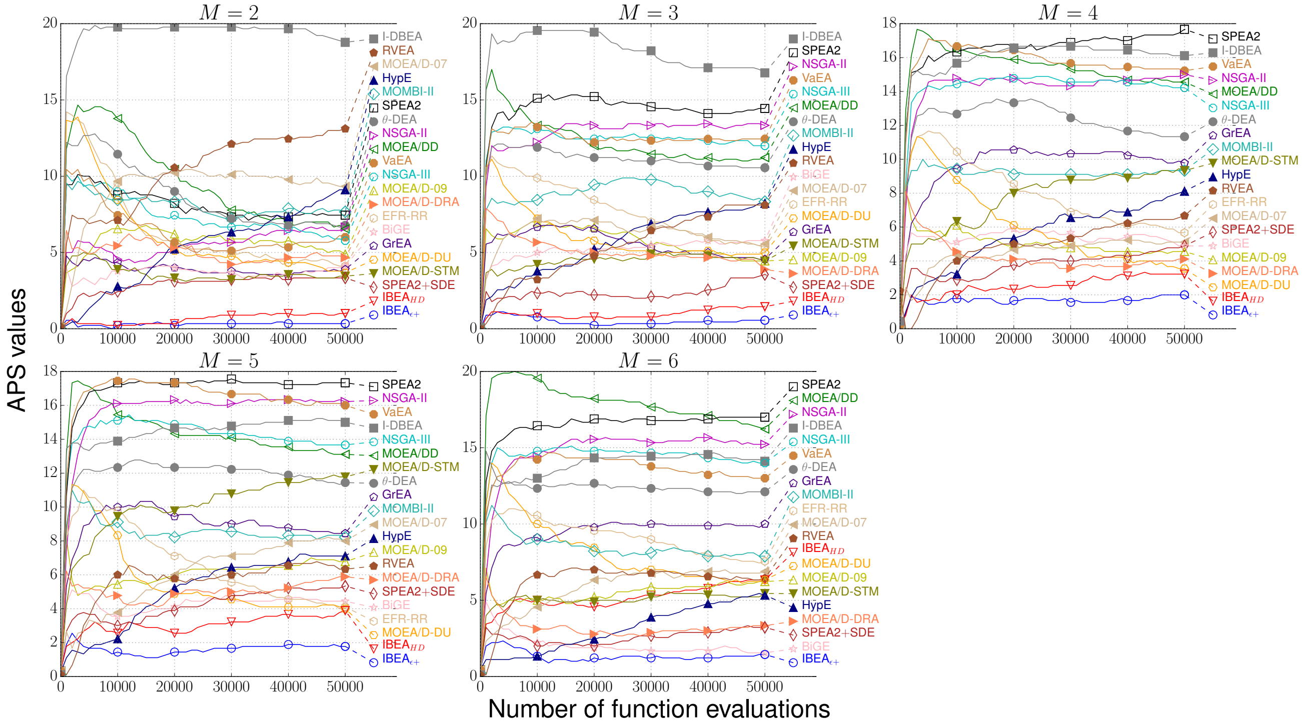


Fig. 1: UEA scenario: APS values based on the HV values of all nondominated solutions in the UEA of the 21 MOEAs for all WFG problems with $M \in \{2, 3, 4, 5, 6\}$ (lower is better).

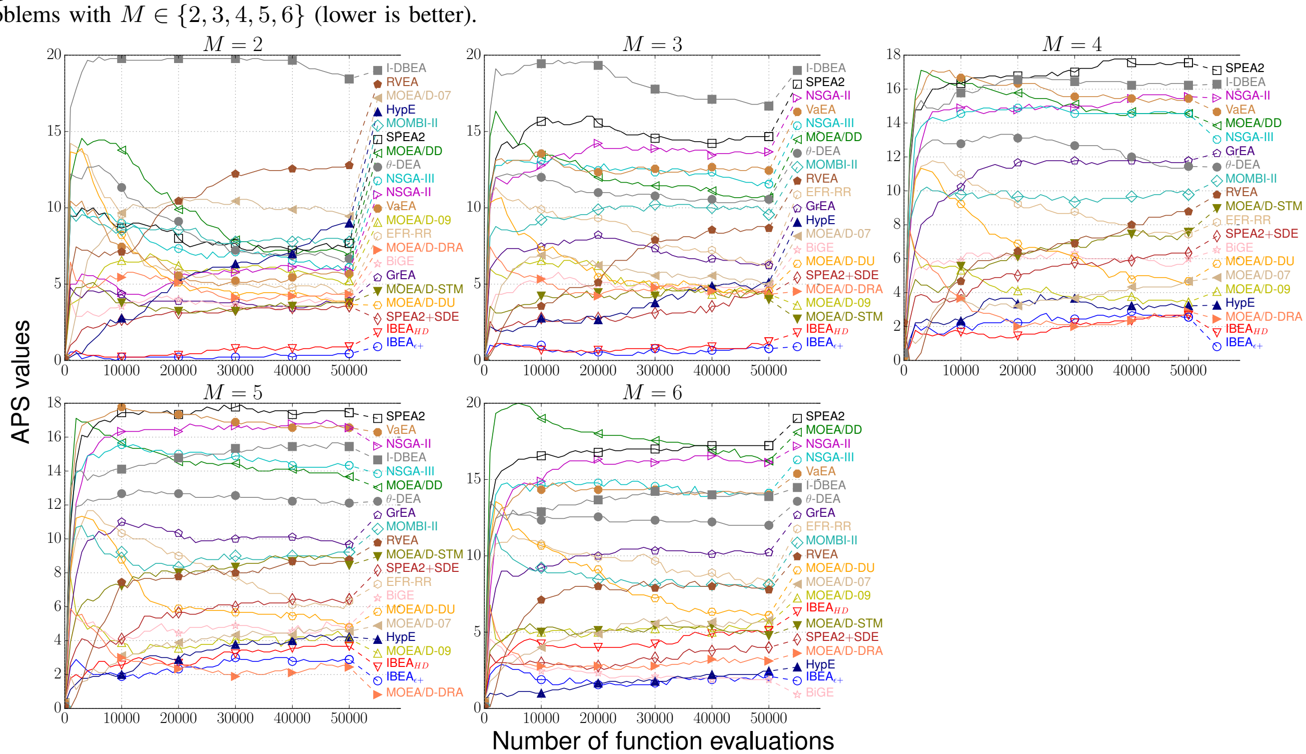


Fig. 2: Reduced UEA scenario: APS values based on the HV values of selected nondominated solutions in the UEA of the 21 MOEAs for all WFG problems with $M \in \{2, 3, 4, 5, 6\}$ (lower is better).

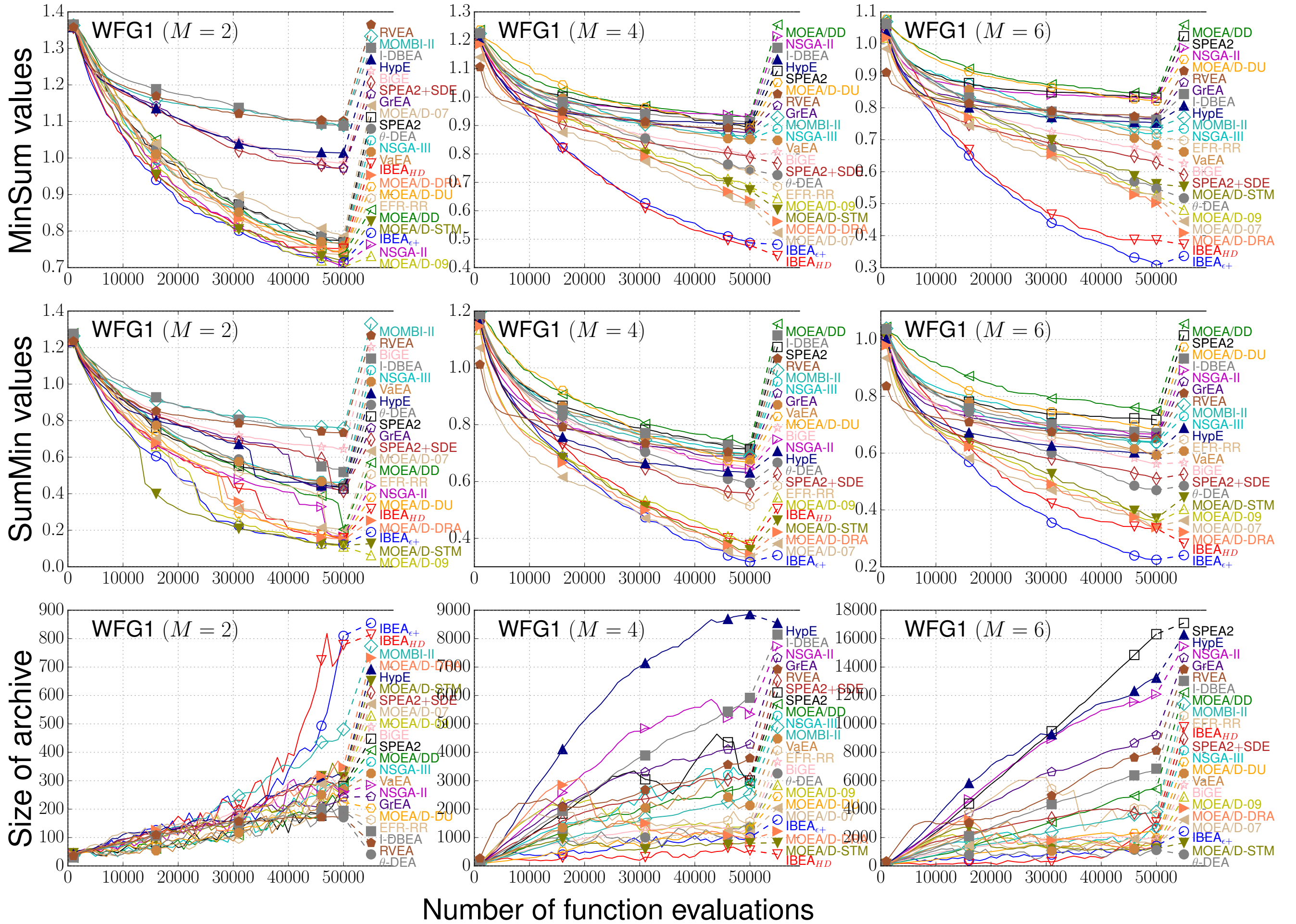


Fig. 3: Convergence behavior of nondominated solutions in the UEAs of the 21 MOEAs. The figures show the values of the MinSum, SumMin, and $|\boldsymbol{A}^{\mathrm{UEA}}|$ over the number of problem evaluations on the WFG1 problem with $M \in \{2, 4, 6\}$. The median of each indicator value across all runs are shown.

among the 21 MOEAs, the best MOEA is HypE for the smaller FEvals, followed by IBEA$_{\epsilon+}$.

For the reduced UEA scenario, the results for $M \in \{2, 3, 4\}$ are almost same as the results for the UEA scenario, but the APS values of MOEA/D-DRA is improved for $M = 4$. For $M = 5$, MOEA/D-DRA is the best performer after FEvals $= 2 \times 10^4$, while MOEA/D-DRA does not perform well for the UEA scenario. For $M = 6$, the APS values of MOEA/D-DRA, HypE, and BiGE are improved compared to the results for the UEA scenario, and BiGE outperforms IBEA$_{\epsilon+}$ at FEvals$=$ 5×10^4. The excellent performance of RVEA for much smaller FEvals can also be found under the reduced UEA scenario.

V. Discussion

In this section, we discuss the experimental results described in Section IV and Appendix. First, we investigate the properties of nondominated solutions in the UEAs of the 21 MOEAs using various performance indicators in Subsection V-A. Next, the distribution of nondominated solutions under each optimization scenario is investigated in Subsection V-B. In Subsection V-C, we analyze the dependency of the performance rank on optimization scenarios. Finally, an overall discussion of our experimental results is presented in Subsection V-D.

A. *Nondominated solutions in UEA of each MOEA*

In order to evaluate the quality of a set of nondominated solutions obtained by MOEAs, we used the HV indicator in Section IV. The HV indicator measures both the convergence toward the PF and diversity of nondominated solutions. In other words, the convergence and diversity of MOEAs cannot be individually evaluated using the HV indicator. In this section, we further investigate the properties of the 21 MOEAs using another type of indicators.

We used two convergence metrics (the MinSum and SumMin indicators [9]) and the number of nondominated solutions stored in the UEA $\boldsymbol{A}^{\mathrm{UEA}}$. Since the GD and IGD indicators [41] are unsuitable for comparing nondominated solution sets of different size, we substituted the MinSum and SumMin indicators for the GD and IGD metrics. While a number of indicators for assessing the diversity of the solutions in the *bounded* archive have been proposed (e.g., [41], [100], [101]), a good diversity indicator for the *unbounded* archive does not exist. Therefore, we used the size of the UEA $|\boldsymbol{A}^{\mathrm{UEA}}|$ to roughly measure the diversity. We do not intend to claim that $|\boldsymbol{A}^{\mathrm{UEA}}|$ is a good indicator to evaluate the diversity of nondominated solutions.

The MinSum indicator evaluates only the convergence quality of a set of nondominated solutions $\boldsymbol{A}$ to the center of the PF. The MinSum value of $\boldsymbol{A}$ is calculated as follows:

$$\mathrm{MinSum}(\boldsymbol{A}) = \min_{\boldsymbol{x} \in \boldsymbol{A}} \Big\{ \sum_{j=1}^{M} f_j(\boldsymbol{x}) \Big\} \tag{2}$$

A small MinSum value indicates that $\boldsymbol{A}$ is closed to the center of the PF in the objective function space.

In contrast to the MinSum indicator defined in Equation (2), the SumMin indicator evaluates only the convergence performance of $\boldsymbol{A}$ toward the PF around its M edges:

$$\mathrm{SumMin}(\boldsymbol{A}) = \sum_{j=1}^{M} \min_{\boldsymbol{x} \in \boldsymbol{A}} \{ f_j(\boldsymbol{x}) \} \tag{3}$$

A small SumMin value indicates that an MOEA achieves well-converged solutions on each objective function.

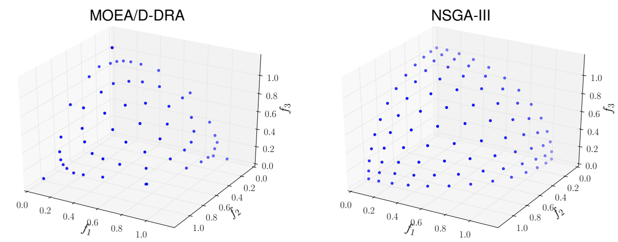
Figure 3 shows the MinSum, SumMin, and $|\boldsymbol{A}^{\mathrm{UEA}}|$ values for the WFG1 problem with $M \in \{2, 4, 6\}$. For $M = 2$, IBEA$_{\epsilon+}$ achieves the lowest MinSum value until FEvals $= 4.5 \times 10^4$, but after that, MOEA/D-09 and NSGA-II show the good MinSum values. As seen from the SumMin indicator values, the convergence performance of MOEA/D-STM toward the PF around its M edges is better than that of the other MOEAs. MOEA/D-09 and IBEA$_{\epsilon+}$ catch up with MOEA/D-STM for FEvals $= 4.0 \times 10^4$. According to the size of the UEA $|\boldsymbol{A}^{\mathrm{UEA}}|$, IBEA$_{\epsilon+}$ and IBEA$_{HD}$ obtain a large number of nondominated solutions after FEvals $= 4.0 \times 10^4$.

For $M \in \{4, 6\}$, RVEA and MOEA/D-07 achieve the best MinSum and SumMin values for much smaller FEvals. This behavior is consistent with the outstanding performance of RVEA and MOEA/D-07 at the beginning of the search as shown in Section IV. RVEA also found the largest number of nondominated solutions until HypE catches up with RVEA at FEvals $= 9 \times 10^3$. HypE finds a large number of nondominated solutions for $M = 4$. After several function evaluations, IBEA$_{\epsilon+}$ and IBEA$_{HD}$ achieve low MinSum values, which indicate their good convergence to the center of the PF. For $M = 6$, in particular, $|\boldsymbol{A}|$ of SPEA2 approximately linearly increases and the size of $\boldsymbol{A}$ becomes 3×10^4 at FEvals $= 5 \times 10^4$. This result shows that approximately 60 percent of individuals generated by SPEA2 are nondominated to each other, and the search of SPEA2 is clearly stagnated due to the ADP phenomenon [53] described in Section II-A.

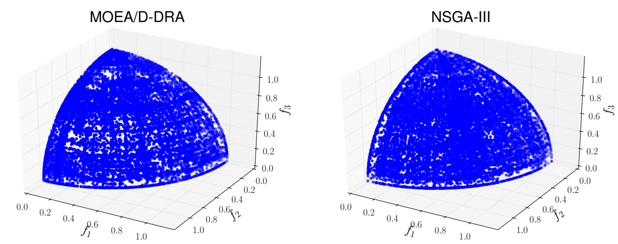
Although we here analyzed the properties of the 21 MOEAs on the WFG1 problem, the results on the remaining WFG problems are not similar to the results on the WFG1 problem. A further investigation should be our future work.

B. Distribution of nondominated solutions under each scenario

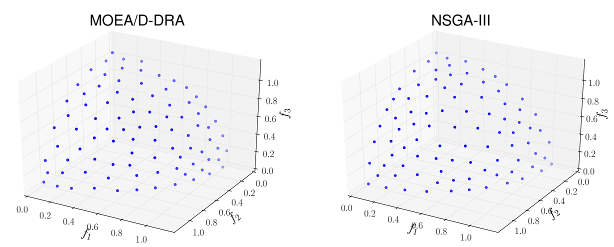
Here, we discuss how the distribution of nondominated solutions is affected by the type of optimization scenarios. Figure 4 shows the distribution of nondominated solutions obtained by MOEA/D-DRA and NSGA-III for the three-objective WFG4 problem under (a) the final population scenario, (b) the UEA scenario, and (c) the reduced UEA scenario.



(a) Nondominated solutions under the final population scenario



(b) Nondominated solutions under the UEA scenario



(c) Nondominated solutions under the reduced UEA scenario

Fig. 4: Distribution of nondominated solutions under (a) the final population scenario, (b) the UEA scenario, and (c) the reduced UEA scenario at FEvals $= 5 \times 10^4$ for the three-objective WFG4 problem. Data for a single run with a median HV value are shown (MOEA/D-DRA and NSGA-III).

As shown in Figure 4(a) and (b), while NSGA-III maintains the uniformly distributed solutions in the population, the solutions obtained by MOEA/D-DRA are not evenly distributed. The biased distribution of the MOEA/D algorithm with the Tchebycheff decomposition approach is also reported in previous studies (e.g., [10], [15], [87]). In contrast to the results for the final population scenario, as shown in Figure 4(b), both MOEAs successfully maintain the densely distributed solutions covering the entire PF in the UEA. Figure 4(c) also shows that even under the reduced UEA scenario, the selected nondominated solutions obtained by MOEA/D-DRA are uniformly distributed on the whole PF.

It is likely that MOEA/D-DRA (the MOEA/D with the Tchebycheff approach) can generate well-distributed solutions but cannot maintain them in the population. This issue of MOEA/D-DRA is addressed by incorporating the UEA, and its performance for approximating the PF can be improved. This result is consistent with the previous study [87]. In fact, Li et al. have reported the promising performance of the

MOEA/D with the Tchebycheff approach with a (bounded) external archive for maintaining nondominated solutions found during the search process [87].

C. Dependency of performance rank on optimization scenarios

In this section, we investigate the dependency of the performance rank. Here, we compare the performance ranks of the 21 MOEAs under the three scenarios: final population, UEA, and reduced UEA. Recall that all nondominated solutions in the *fixed size population* are used for the performance assessment for the final population scenario. For the final population and reduced UEA scenarios, it is assumed that a user of an MOEA wants to know only a small number of well-distributed nondominated solutions of a given MOP.

Table IV shows the comparison of the performance rank for the final population, UEA, and reduced UEA scenarios. The 21 MOEAs are ranked according to their APS values for all the nine WFG problems at FEvals= 5×10^4. It should be emphasized that Table IV *does not* summarize all the experimental results in our benchmarking study. The performance rank described in Table IV is based on only the end-of-the-run results, which do not provide sufficient information about the performance of MOEAs [32], [102]. Our interest here is an investigation how the performance rank differs depending on optimization scenarios.

As shown in Table IV, the performance ranks of particular MOEAs are significantly affected by the type of optimization scenarios. For example, for $M = 2$, while the rank of IBEA$_{\epsilon+}$ is 10th out of the 21 methods for the final population scenario, that is the first for the UEA and reduced UEA scenarios. For $M = 5$, while the rank of MOEA/D-DRA is 14th for the final population scenario, that is the first for the reduced UEA scenario. The ranks of MOEA/D variants (MOEA/D-07, MOEA/D-09, MOEA/D-DRA, MOEA/D-STM, MOEA/DD) and some indicator based MOEAs (HypE and BiGE) are also significantly changed depending on the type of optimization scenarios. In particular, the ranks of MOEA/D-07, MOEA/D-09, MOEA/D-DRA, and MOEA/D-STM are notably improved for the UEA and reduced UEA scenarios. In contrast to the four MOEA/D algorithms, the performance ranks of three latest MOEA/D variants (MOEA/D-DU, EFR-RR, and I-DBEA) for the UEA and reduced UEA scenarios become worse than those for the final population scenario (except for the results of MOEA/D-DU on $M = 4$ under the UEA scenario). MOEA/D-DU, EFR-RR, and I-DBEA were originally developed for the final population scenario [20], [21] and are capable of maintaining evenly-distributed nondominated solutions in their population. However, such an ability that keeps well-distributed solutions in the population is not necessary for the UEA and reduced UEA scenarios. Thus, it is likely that the incorporation of the UEA is not so beneficial for MOEA/D-DU, EFR-RR, and I-DBEA compared to MOEA/D-07, MOEA/D-09, MOEA/D-DRA, and MOEA/D-STM.

In contrast to the indicator and decomposition based MOEAs, the ranks of the seven dominance based methods degrade in many cases under the reduced UEA scenario. This reason seems to be the same with the case of MOEA/D-DU, EFR-RR, and I-DBEA. As pointed out in [87], some non-Pareto dominance based MOEAs have a characteristic that is difficult to maintain well-distributed nondominated solutions in their population. However, as shown in Section V-B, this undesirable characteristic of decomposition and indicator based MOEAs can be easily addressed by incorporating the UEA that maintain all nondominated solutions found during the search process. This is the reason why the ranks of some non-Pareto dominance based MOEAs are improved for the reduced UEA scenario.

Taking into account of above discussion, we conclude that the type of optimization scenarios significantly affects the performance rank of MOEAs. Although the final population scenario has been widely used in the evolutionary computation community, other types of optimization scenarios should be considered to evaluate the performance of MOEAs for more fair comparison.

D. Overall discussion on the performance of the 21 MOEAs

Unsurprisingly, the results described in Section IV and Appendix show that the performance of the 21 MOEAs significantly depends on the problem type, the number of objectives M, and the number of function evaluations (FEvals). For example, for the six-objective WFG8 problem, HypE performs best within FEvals $= 1 \times 10^4$, followed by BiGE. Thus, determining the best MOEA among the 21 MOEAs is difficult, even though for the same problem instance. Keeping in mind this fact, in the following, we roughly discuss the performance of the 21 MOEAs based on their APS values (Figure 1 and 2).

IBEA$_{\epsilon+}$ is the first indicator based MOEA proposed in 2004 and was not specially designed for MaOPs. The poor performance of IBEA$_{\epsilon+}$ for MaOPs as well as MOPs with $M < 4$ has been reported in numerous articles (e.g., [8], [23], [26], [85], [103]). However, in our benchmarking study, the classical IBEA$_{\epsilon+}$ shows high performance among the 21 MOEAs, including recently proposed state-of-the-art MOEAs, on the WFG problems with two to six objectives under the UEA scenario. IBEA$_{\epsilon+}$ is also competitive with the other MOEAs under the reduced UEA scenario.

For $M = 6$, HypE is the best MOEA up to FEvals $= 1 \times 10^4$ under both the UEA and reduced UEA scenarios, while RVEA performs best immediately after the start of the search. HypE, proposed in 2011, was specially designed for MaOPs with up to 50 objectives but is a relatively classical algorithm among the 21 MOEAs. HypE is considered as a standard, benchmark algorithm and has frequently been compared to newly proposed MOEAs (e.g., [11], [12], [14], [19]). In other words, it is widely believed that recently proposed MOEAs perform better than HypE. However, such conventional wisdom is clearly contradicted by the experimental results in this paper.

Although MOEA/D-DRA, proposed in 2009, was the winner of the IEEE CEC2009 competition on multi-objective optimization[13], MOEA/D-DRA is a relatively classical algorithm among the nine decomposition based MOEAs. In fact,

[13]http://dces.essex.ac.uk/staff/zhang/moeacompetition09.htm

TABLE IV: The comparison of the performance ranks under the final population, UEA, and reduced UEA scenarios (from left to right). For example, the performance ranks of NSGA-II (15/14/12) on the WFG problems with $M = 2$ under the final population, UEA, and reduced UEA scenarios are 15, 14, and 12, respectively. The 21 MOEAs are ranked according to their APS values for all the nine WFG problems {WFG1, ..., WFG9} at FEvals= 5×10^4. For each M, data that the rank of an MOEA for the UEA and reduced UEA scenarios is better than that for the final population scenario are shaded.

MOEAs	$M = 2$	$M = 3$	$M = 4$	$M = 5$	$M = 6$
NSGA-II	15/14/12	20/19/19	20/18/19	16/20/19	19/19/19
NSGA-III	09/11/13	10/17/17	12/16/16	12/17/17	15/17/18
SPEA2	12/16/16	16/20/20	19/21/21	18/21/21	20/21/21
SPEA2+SDE	05/03/03	03/03/05	03/06/09	05/06/10	03/03/05
GrEA	20/06/06	04/07/11	10/14/15	10/13/14	08/14/14
VaEA	13/12/11	12/18/18	16/19/18	14/19/19	16/16/17
θ-DEA	10/15/14	08/15/15	11/15/14	11/14/15	11/15/15
MOEA/D-07	14/19/19	18/10/09	18/08/06	19/11/06	18/11/09
MOEA/D-09	07/10/10	17/05/05	17/05/05	19/09/04	17/07/08
MOEA/D-DRA	06/09/08	15/04/05	13/04/03	14/07/01	12/04/04
MOEA/D-STM	01/04/05	14/05/03	14/12/11	17/15/11	12/06/06
MOEA/DD	08/13/15	21/16/16	21/17/16	21/16/16	21/20/20
MOEA/D-DU	03/05/04	07/08/07	04/03/06	03/03/07	05/08/10
I-DBEA	21/21/21	18/21/21	14/20/20	13/18/18	14/18/16
EFR-RR	02/07/09	05/09/12	05/09/10	02/04/09	06/12/13
RVEA	19/20/20	09/12/13	07/10/12	07/08/12	10/10/11
IBEA$_{\epsilon+}$	10/01/01	01/01/01	01/01/01	01/01/02	01/01/02
IBEA$_{HD}$	03/02/02	02/02/02	02/02/02	04/02/03	07/08/07
HypE	17/18/18	11/13/10	09/11/04	08/10/05	04/05/03
BiGE	18/07/07	13/11/08	06/06/08	06/05/08	02/02/01
MOMBI-II	16/17/17	06/14/14	08/12/13	09/12/13	09/13/12

a number of recently proposed decomposition based MOEAs outperform MOEA/D-DRA. However, MOEA/D-DRA shows high performance for $M \geq 4$ on the reduced UEA scenario and performs best among the 21 MOEAs for $M = 5$ after FEvals $= 2 \times 10^4$. MOEA/D-07, which is the first MOEA/D algorithm, also performs better than some MOEA/D variants. Compared to MOEA/D-09, MOEA/D-07 performs well for smaller FEvals.

In summary, our benchmarking results presented in Section IV and Appendix are significantly different from the results reported in previous studies. One reason may be the difference of optimization scenarios used to evaluate the performance of MOEAs. We evaluated the anytime performance of the 21 MOEAs on the UEA and reduced UEA scenarios, whereas previous studies evaluated MOEAs on the final population scenario. In general, the population maintains the good solutions obtained during the search process, but the population size is limited. Therefore, when the number of good solutions found in the search exceeds the population size, they must be removed from the population to keep the population size constant. Thus, a carefully designed environmental selection method is required for MOEAs so that the nondominated solutions in the population are uniformly distributed. Recently proposed MOEAs for MaOPs (e.g., NSGA-III, MOEA/DD, and VaEA) were designed according to this policy. However, MOEAs do not require such well-designed environmental selection method under the UEA and reduced UEA scenarios because all nondominated solutions found in the search are automatically maintained in the UEA independent from the population (see Subsection V-B and V-C).

Another reason for the poor performance of recently proposed MOEAs is likely to be the setting of the maximum number of function evaluations (FEvals$^{\max}$). As shown in Table I, FEvals$^{\max}$ used for comparative studies is increasing yearly. For example, while FEvals$^{\max}$ used in the HypE paper [24], published in 2011, was only 1×10^4, FEvals$^{\max}$ used in recent papers (e.g., NSGA-III [10], MOEA/DD [19], θ-DEA [14], and VaEA [13]) was 5.5×10^5. The latter is 55 times larger than the former. Some real-world problems require the execution of a computationally expensive simulation in order to evaluate the solution [44], [45]. Thus, in practice, users of MOEAs cannot always set FEvals$^{\max}$ to a large number (e.g., 5.5×10^5). As mentioned above, many articles report the poor performance of HypE [11], [12], [14], [19], but we believe that this is because of the different settings of FEvals$^{\max}$. HypE was designed for optimization with FEvals$^{\max}$ $\leq 1 \times 10^4$, but the settings of FEvals$^{\max}$ used in other previous studies (e.g., [13], [14], [19]) were clearly much larger. In fact, as shown in Figure 1 and 2, HypE performs very well within FEvals $= 1 \times 10^4$ for $M = 6$. Such excellent performance of HypE for small FEvals cannot be revealed under the final population scenario, which is frequently used in the evolutionary computation community.

VI. Conclusion

In this paper, we exhaustively investigated the anytime performance of the 21 MOEAs in Table I on the WFG test problems with $M \in \{2, 3, 4, 5, 6\}$. We used the UEA and reduced UEA scenarios to fairly evaluate MOEAs. Our experimental results are significantly different from the results reported in previous studies under the final population scenario. The performance of some recently proposed MOEAs is not as good as some classical MOEAs, some classical MOEAs perform very well (e.g., the two IBEA variants [23], MOEA/D-DRA [17] and HypE [24]) in our experimental study. We also analyzed the reason why the performance rank of the 21 MOEAs is significantly different from that in previous studies.

Our exhaustive benchmarking study will be useful for both users and algorithm designers of MOEAs. The experimental results of this study are helpful for a user who wants to use an MOEA as an optimization tool. For example, if the number of objective functions in a given real-world problem is five, and available computational budget FEvals$^{\max}$ is within 3×10^3, RVEA is likely to be the most suitable algorithm for both the UEA and reduced UEA scenarios. Moreover, the experimental data presented herein are available on the website[14]. Thus, researchers can easily compare their newly designed algorithms to the 21 MOEAs. We hope that our benchmarking study presented here boosts researchers in the evolutionary computation community to a constructive race of algorithm developments.

In this benchmarking study, the (normalized) WFG functions [95] were used for evaluating the performance of the 21 MOEAs. However, the WFG functions have some unnatural, exploitable problem features [32], [97]. For example, MOEAs with uniformly generated reference and weight vectors (e.g., MOEA/D variants) can successfully find eventually distributed nondominated solutions on the WFG functions because the shape of the distribution of the reference and weight vectors is the same as the shape of the PF [97]. The comparison of MOEAs on other benchmark functions is an interesting future direction.

Appendix
Results on each WFG test problem

Figures 5 – 13 show the results of the 21 MOEAs on each WFG problem instance with $M \in \{2, 4, 6\}$. The left (a) and right (b) side of Figures 5 – 13 show the anytime performance on the UEA and reduced UEA scenarios, respectively. Note that all of the HV values under the reduced UEA scenario is always lower than those under the UEA scenario, because the number of nondominated solutions for the HV calculation is significantly different in both optimization scenarios. A monotonic increase of the HV value over time is not ensured in the case of the reduced UEA scenario [30]. Therefore, the performance deterioration can be found under the reduced UEA scenario (i.e., the performance of an MOEA at the iteration t becomes worse than that at the iteration $t-1$). For example, the performance curve of SPEA2 on the six-objective WFG3 problem shown in Figure 7(b) is zigzag. In the following, we describe the results on each WFG problem:

- **WFG1**:

No significant difference between the results for the UEA and reduced UEA scenarios is found on the WFG1 problem. IBEA$_{\epsilon+}$ and MOEA/D variants perform well in $M \in \{2, 4, 6\}$. For $M = 2$, MOEA/D-STM achieves the highest HV value until FEvals $= 3 \times 10^4$, but MOEA/D-09 and IBEA$_{\epsilon+}$ catch up with MOEA/D-STM for FEvals $> 3 \times 10^4$. For $M = 4$ and 6, the best performer is MOEA/D-07 for FEvals $< 3 \times 10^4$ and $< 1.3 \times 10^4$ respectively. After that, IBEA$_{\epsilon+}$ outperforms the other MOEAs.

- **WFG2**:

As in the results on the WFG1 problem, the results for both scenarios are quite similar on the WFG2 problem. For $M = 2$, EFR-RR outperforms the other MOEAs after FEvals $= 2 \times 10^4$. For $M = 4$ and $M = 6$, HypE and MOEA/D-DU achieve high HV values, whereas MOEA/D-07 and MOEA/D-09 perform poorly. In particular, the performance of HypE shows the good anytime behavior.

- **WFG3**: In contrast to the results on the WFG1 and WFG2 problems, the results for the UEA and reduced UEA scenarios are significantly different from each other on the WFG3 problem with $M \in \{4, 6\}$. The same tendency can be found in the results on the other test problems (WFG4, ..., WFG9). Therefore, in the following, we describe the results under each scenario separately for each test problem.

◦ **UEA scenario**: For $M = 2$, IBEA$_{HD}$ performs relatively well after FEvals $= 1 \times 10^4$. While the performance of IBEA$_{HD}$ is better than that of other MOEAs for $M = 2$, the performance of IBEA$_{HD}$ deteriorates gradually with increasing M. For $M = 4$ and $M = 6$, while MOEA/D-09, MOEA/D-DRA, IBEA$_{\epsilon+}$, and BiGE perform well for the larger FEvals, the best MOEA for the smaller FEvals is HypE.

◦ **Reduced UEA scenario**: While after FEvals $\in \{1.3 \times 10^4, ..., 3 \times 10^4\}$ HypE is inferior to some algorithms on the UEA scenario, HypE clearly outperforms the other MOEAs for $M \in \{4, 6\}$ at any time (in particular, $M = 6$).

- **WFG4**:

◦ **UEA scenario**: For $M = 2$ and $M = 4$, IBEA$_{HD}$ exhibits good performance at all times, while for $M = 6$, its performance becomes inferior to the other MOEAs. On the other hand, for $M = 6$, BiGE and IBEA$_{\epsilon+}$ achieve high HV values, and MOEA/D-DU outperforms the remaining MOEAs at approximately FEvals $= 5 \times 10^4$.

◦ **Reduced UEA scenario**: For $M \in \{4, 6\}$, the performance of NSGA-II and SPEA2 significantly deteriorates compared to the results for the UEA scenario. The rank of HypE for the reduced UEA scenario is slightly better than that for the UEA scenario. This phenomenon also can be found in the results of WFG5, WFG6, WFG7, WFG8, and WFG9 problems.

[14] https://sites.google.com/site/benchmarkingmoeas

• **WFG5**:

◦ **UEA scenario**: For $M = 2$ and $= 4$, the best MOEA is RVEA within FEvals $= 8 \times 10^3$ and $= 2.1 \times 10^4$ respectively. IBEA$_{HD}$ exhibits the best performance on the six-objective WFG5 problem.

◦ **Reduced UEA scenario**: The difference of the HV values of IBEA$_{HD}$ and other MOEAs is slightly emphasized compared to the results on UEA scenario. The same tendency can be found in the results of the WFG6 problem.

• **WFG6**:

◦ **UEA scenario**: For $M = 2$, IBEA$_{HD}$ and IBEA$_{\epsilon+}$ perform well for the smaller FEvals. The two IBEA variants also outperform the remaining MOEAs for $M = 4$. IBEA$_{HD}$ is also the best performer for $M = 6$.

◦ **Reduced UEA scenario**: For $M = 6$, the relative performance of BiGE becomes better than that for the UEA scenario. BiGE is still outperformed by IBEA$_{HD}$, but the anytime performance of BiGE is better than that of IBEA$_{\epsilon+}$.

• **WFG7**:

◦ **UEA scenario**: The two IBEA variants show the best anytime performance for $M = 2$. For $M = 4$, IBEA$_{HD}$ and IBEA$_{\epsilon+}$ also perform well, and MOEA/D-DU and EFR-RR catch up them for larger FEvals. For $M = 6$, IBEA$_{\epsilon+}$ still performs well while the performance of IBEA$_{HD}$ is degraded. BiGE and SPEA2+SDE are competitive with IBEA$_{\epsilon+}$.

◦ **Reduced UEA scenario**: The poor performance of NSGA-II and SPEA2 is more emphasized than that for the UEA scenario. For $M = 6$, BiGE has the best performance after FEvals $= 1.6 \times 10^4$.

• **WFG8**:

◦ **UEA scenario**: MOEA/D-STM outperforms the other MOEAs for $M = 2$ at all times. For $M = 4$, the best MOEA is IBEA$_{HD}$ within FEvals $= 2 \times 10^4$, and MOEA/D-DU outperforms the other MOEAs. Similarly, for $M = 6$, HypE achieves the highest HV value until FEvals $= 1 \times 10^4$, and BiGE performs best after that.

◦ **Reduced UEA scenario**: For $M = 6$, the good performance of HypE at the beginning of the search is emphasized compared to the results for the UEA scenario. BiGE, MOEA/D-DU, IBEA$_{\epsilon+}$, and SPEA2+SDE achieve the high HV values after FEvals $= 4.6 \times 10^4$.

• **WFG9**:

◦ **UEA scenario**: For $M = 2$, SPEA2+SDE performs relatively well while the evolution of the five MOEAs (GrEA, MOEA/D-STM, EFR-RR, NSGA-II, and I-DBEA) clearly stagnates just after the beginning of the search. For $M = 4$, RVEA exhibits the good performance at all times and is clearly the best MOEA. For $M = 6$, IBEA$_{HD}$ performs well, but SPEA2+SDE and IBEA$_{\epsilon+}$ outperform IBEA$_{HD}$ for larger FEvals.

◦ **Reduced UEA scenario**: RVEA still has the best performance for $M = 4$, IBEA$_{HD}$ is the best performer for $M = 6$. The slow convergence of NSGA-II and MOEA/DD seems to be more emphasized for $M = 6$.

Acknowledgment

This research was supported by MEXT as "Priority Issue on Post-K computer" (Development of Innovative Design and Production Processes that Lead the Way for the Manufacturing Industry in the Near Future).

References

[1] Y. S. Ong, P. B. Nair, and A. J. Keane, "Evolutionary optimization of computationally expensive problems via surrogate modeling," *AIAA journal*, vol. 41, no. 4, pp. 687–696, 2003.
[2] M. G. Parsons and R. L. Scott, "Formulation of Multicriterion Design Optimization Problems for Solution With Scalar Numerical Optimization Methods," *J. Ship Research*, vol. 48, no. 1, pp. 61–76, 2004.
[3] M. N. Le, Y. Ong, S. Menzel, Y. Jin, and B. Sendhoff, "Evolution by adapting surrogates," *Evol. Comput.*, vol. 21, no. 2, pp. 313–340, 2013.
[4] A. Trivedi, D. Srinivasan, K. Pal, C. Saha, and T. Reindl, "Enhanced Multiobjective Evolutionary Algorithm Based on Decomposition for Solving the Unit Commitment Problem," *IEEE Trans. Industrial Informatics*, vol. 11, no. 6, pp. 1346–1357, 2015.
[5] B. Li, J. Li, K. Tang, and X. Yao, "Many-objective evolutionary algorithms: A survey," *ACM C. S.*, vol. 48, no. 1, pp. 13:1–13:35, 2015.
[6] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE TEVC*, vol. 6, no. 2, pp. 182–197, 2002.
[7] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the Strength Pareto Evolutionary Algorithm," ETHZ, Tech. Rep., 2001.
[8] T. Wagner, N. Beume, and B. Naujoks, "Pareto-, Aggregation-, and Indicator-Based Methods in Many-Objective Optimization," in *EMO*, 2007, pp. 742–756.
[9] H. Ishibuchi, N. Tsukamoto, and Y. Nojima, "Evolutionary many-objective optimization: A short review," in *IEEE CEC*, 2008, pp. 2419–2426.
[10] K. Deb and H. Jain, "An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints," *IEEE TEVC*, vol. 18, no. 4, pp. 577–601, 2014.
[11] M. Li, S. Yang, and X. Liu, "Shift-Based Density Estimation for Pareto-Based Algorithms in Many-Objective Optimization," *IEEE TEVC*, vol. 18, no. 3, pp. 348–365, 2014.
[12] S. Yang, M. Li, X. Liu, and J. Zheng, "A Grid-Based Evolutionary Algorithm for Many-Objective Optimization," *IEEE TEVC*, vol. 17, no. 5, pp. 721–736, 2013.
[13] Y. Xiang, Y. Zhou, M. Li, and Z. Chen, "A Vector Angle-Based Evolutionary Algorithm for Unconstrained Many-Objective Optimization," *IEEE TEVC*, vol. 21, no. 1, pp. 131–152, 2017.
[14] Y. Yuan, H. Xu, B. Wang, and X. Yao, "A New Dominance Relation-Based Evolutionary Algorithm for Many-Objective Optimization," *IEEE TEVC*, vol. 20, no. 1, pp. 16–37, 2016.
[15] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE TEVC*, vol. 11, no. 6, pp. 712–731, 2007.
[16] H. Li and Q. Zhang, "Multiobjective Optimization Problems With Complicated Pareto Sets, MOEA/D and NSGA-II," *IEEE TEVC*, vol. 13, no. 2, pp. 284–302, 2009.
[17] Q. Zhang, W. Liu, and H. Li, "The performance of a new version of MOEA/D on CEC09 unconstrained MOP test instances," in *IEEE CEC*, 2009, pp. 203–208.
[18] K. Li, Q. Zhang, S. Kwong, M. Li, and R. Wang, "Stable Matching-Based Selection in Evolutionary Multiobjective Optimization," *IEEE TEVC*, vol. 18, no. 6, pp. 909–923, 2014.
[19] K. Li, K. Deb, Q. Zhang, and S. Kwong, "An Evolutionary Many-Objective Optimization Algorithm Based on Dominance and Decomposition," *IEEE TEVC*, vol. 19, no. 5, pp. 694–716, 2015.
[20] Y. Yuan, H. Xu, B. Wang, B. Zhang, and X. Yao, "Balancing Convergence and Diversity in Decomposition-Based Many-Objective Optimizers," *IEEE TEVC*, vol. 20, no. 2, pp. 180–198, 2016.
[21] M. Asafuddoula, T. Ray, and R. A. Sarker, "A Decomposition-Based Evolutionary Algorithm for Many Objective Optimization," *IEEE TEVC*, vol. 19, no. 3, pp. 445–460, 2015.
[22] R. Cheng, Y. Jin, M. Olhofer, and B. Sendhoff, "A Reference Vector Guided Evolutionary Algorithm for Many-Objective Optimization," *IEEE TEVC*, vol. 20, no. 5, pp. 773–791, 2016.
[23] E. Zitzler and S. Künzli, "Indicator-based selection in multiobjective search," in *PPSN*, 2004, pp. 832–842.
[24] J. Bader and E. Zitzler, "HypE: An Algorithm for Fast Hypervolume-Based Many-Objective Optimization," *Evol. Comput.*, vol. 19, no. 1, pp. 45–76, 2011.

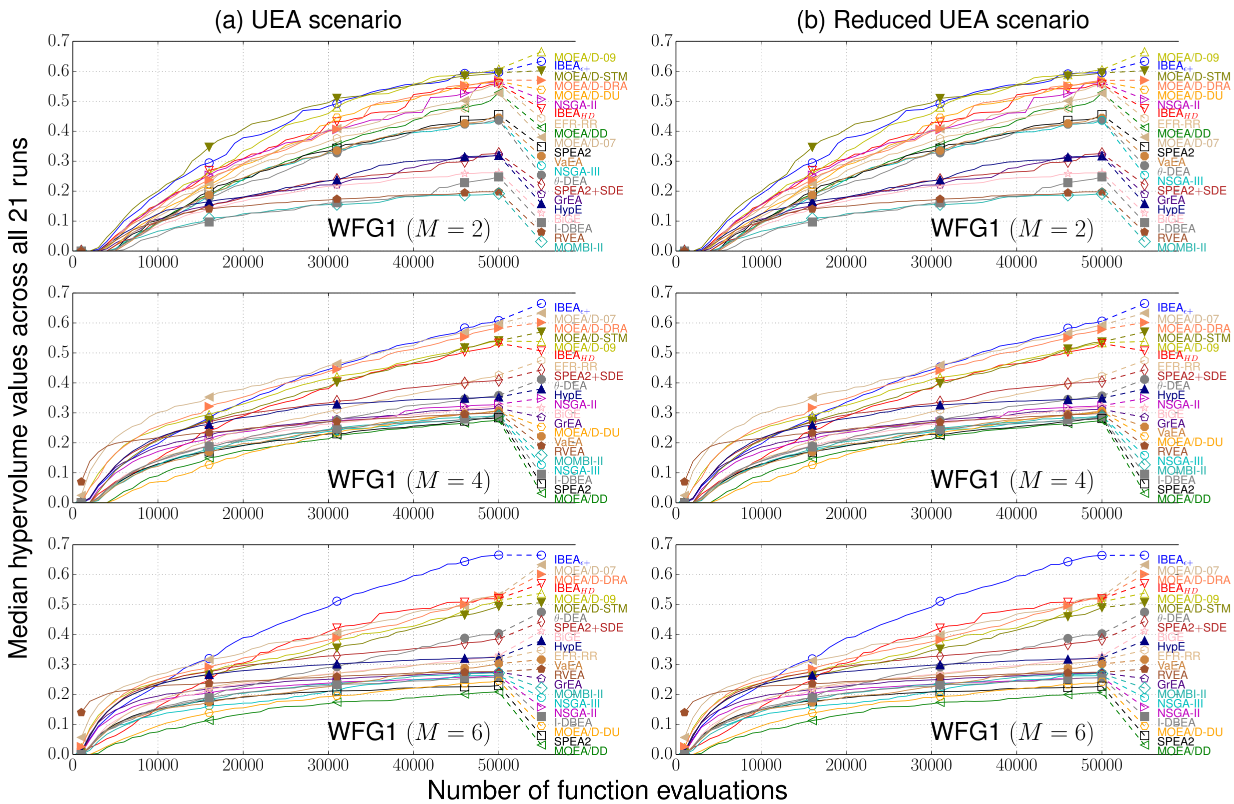


Fig. 5: Anytime performance of the 21 MOEAs on the normalized WFG1 problem ($M \in \{2, 4, 6\}$).

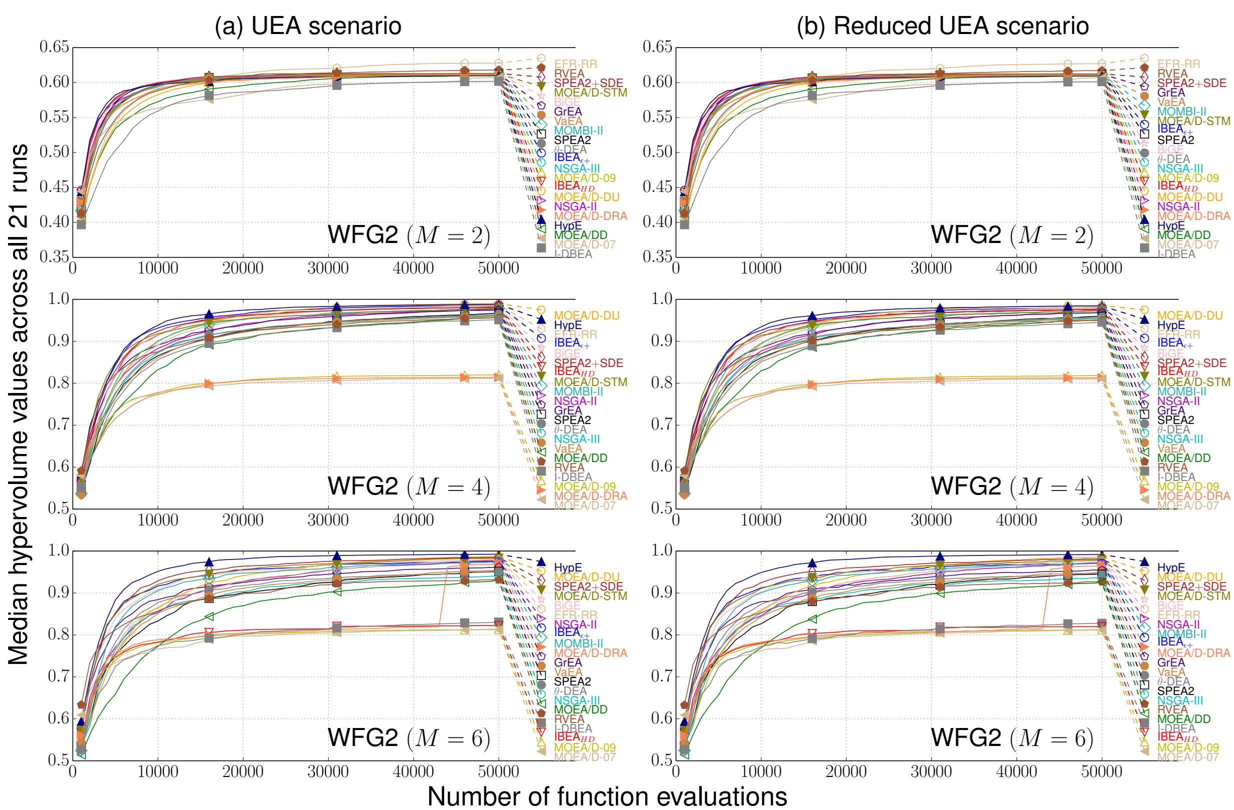


Fig. 6: Anytime performance of the 21 MOEAs on the normalized WFG2 problem ($M \in \{2, 4, 6\}$).

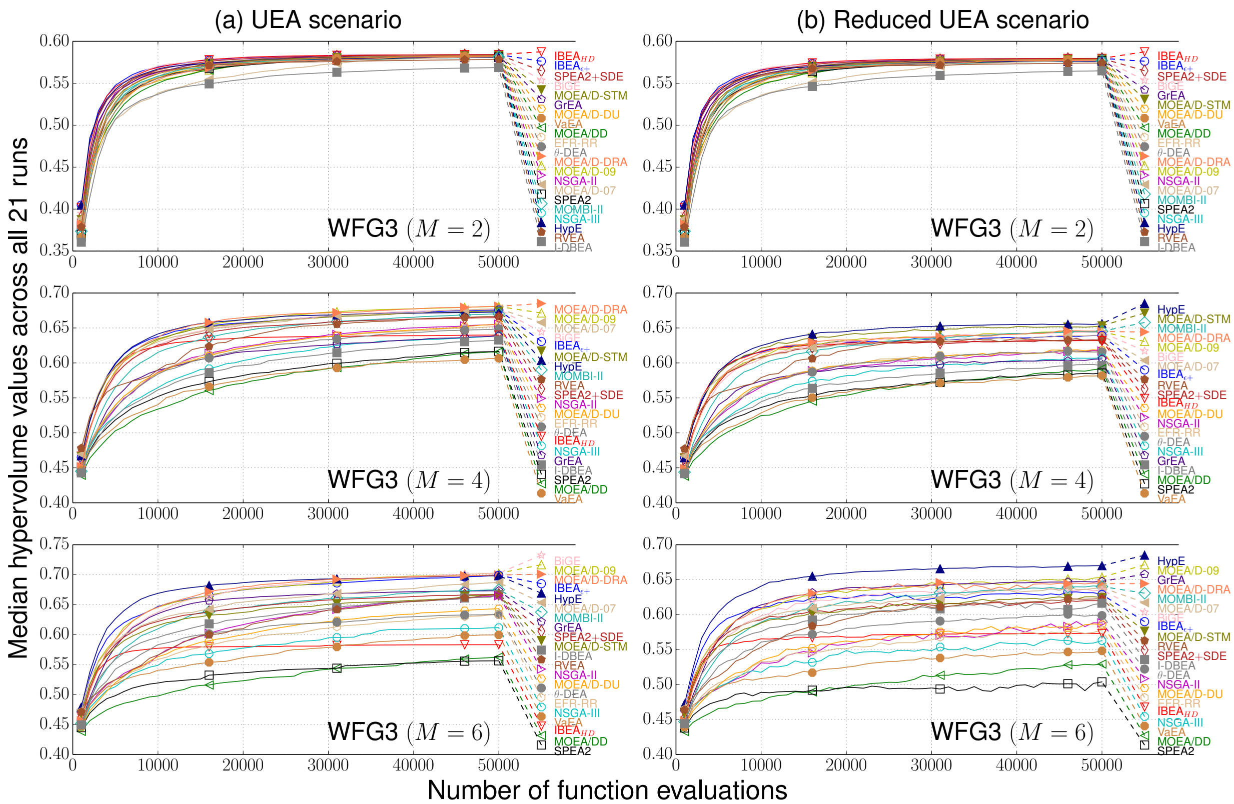


Fig. 7: Anytime performance of the 21 MOEAs on the normalized WFG3 problem ($M \in \{2, 4, 6\}$).

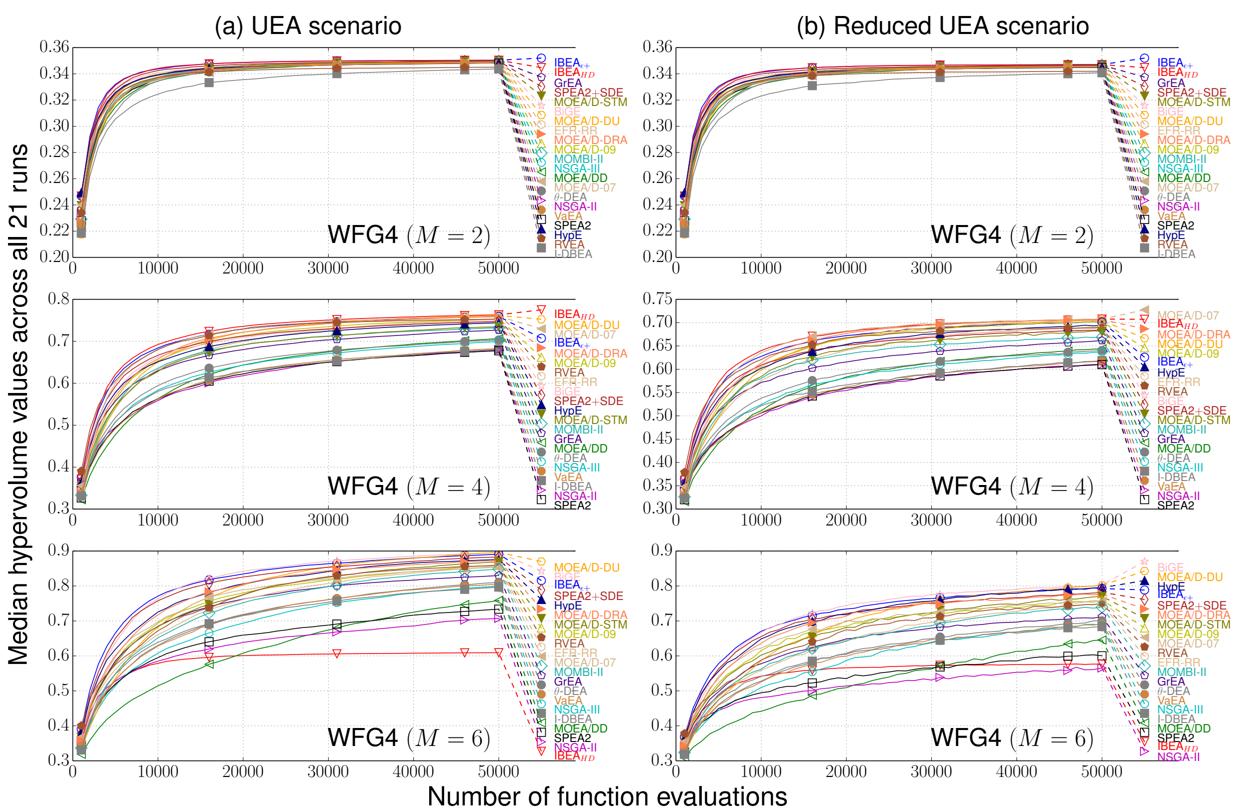


Fig. 8: Anytime performance of the 21 MOEAs on the normalized WFG4 problem ($M \in \{2, 4, 6\}$).

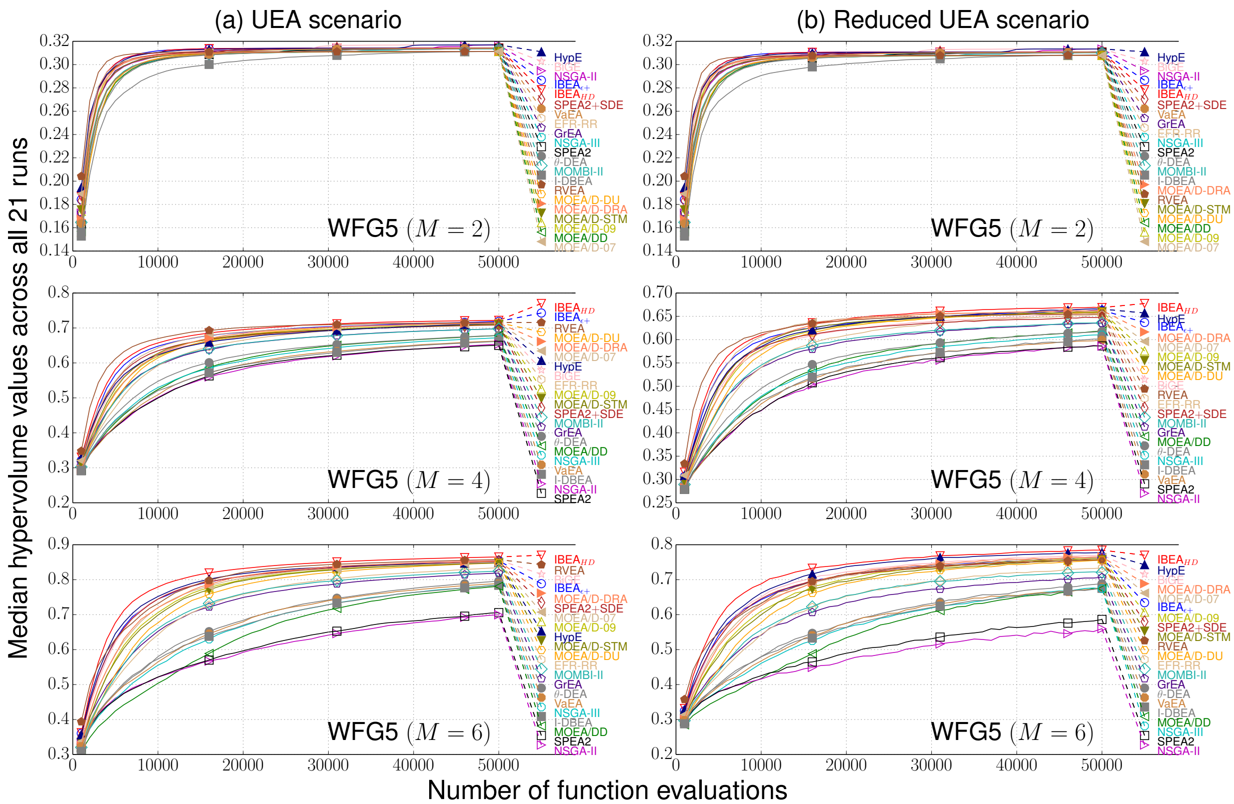


Fig. 9: Anytime performance of the 21 MOEAs on the normalized WFG5 problem ($M \in \{2, 4, 6\}$).

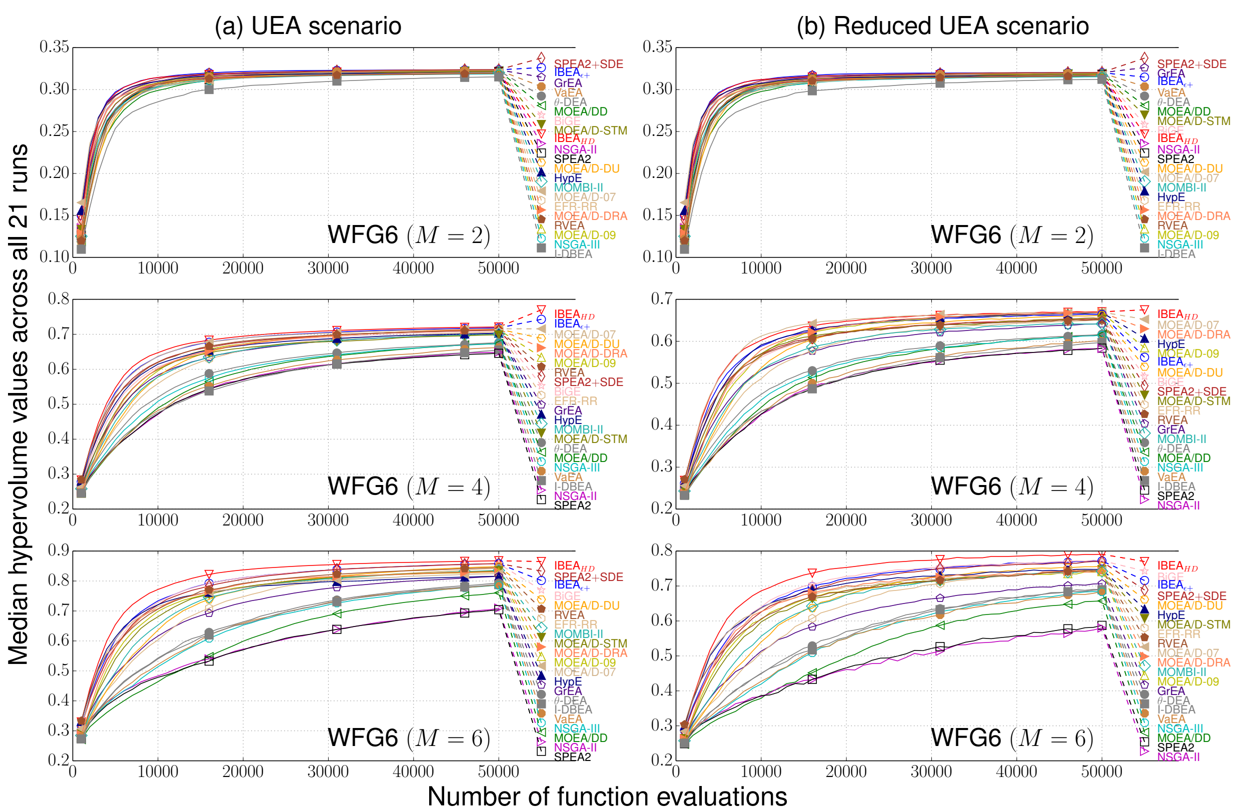


Fig. 10: Anytime performance of the 21 MOEAs on the normalized WFG6 problem ($M \in \{2, 4, 6\}$).

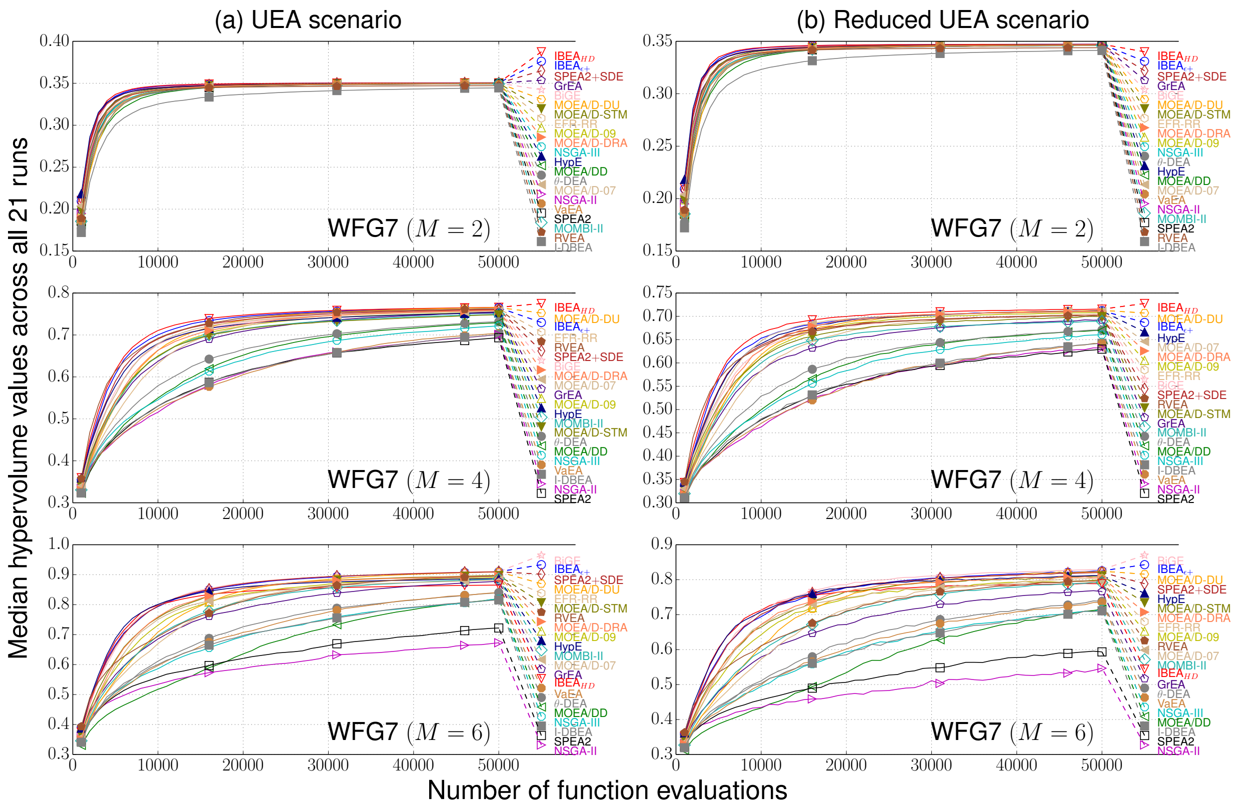


Fig. 11: Anytime performance of the 21 MOEAs on the normalized WFG7 problem ($M \in \{2, 4, 6\}$).

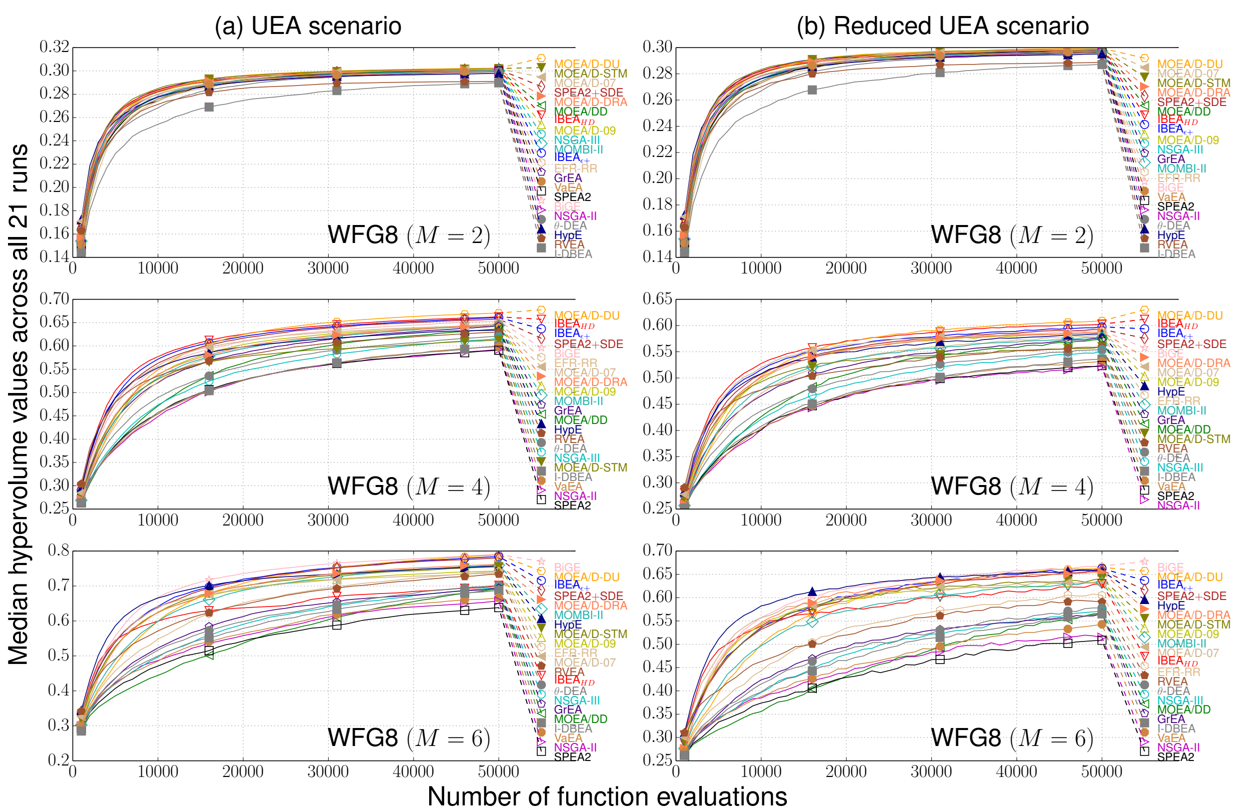


Fig. 12: Anytime performance of the 21 MOEAs on the normalized WFG8 problem ($M \in \{2, 4, 6\}$).

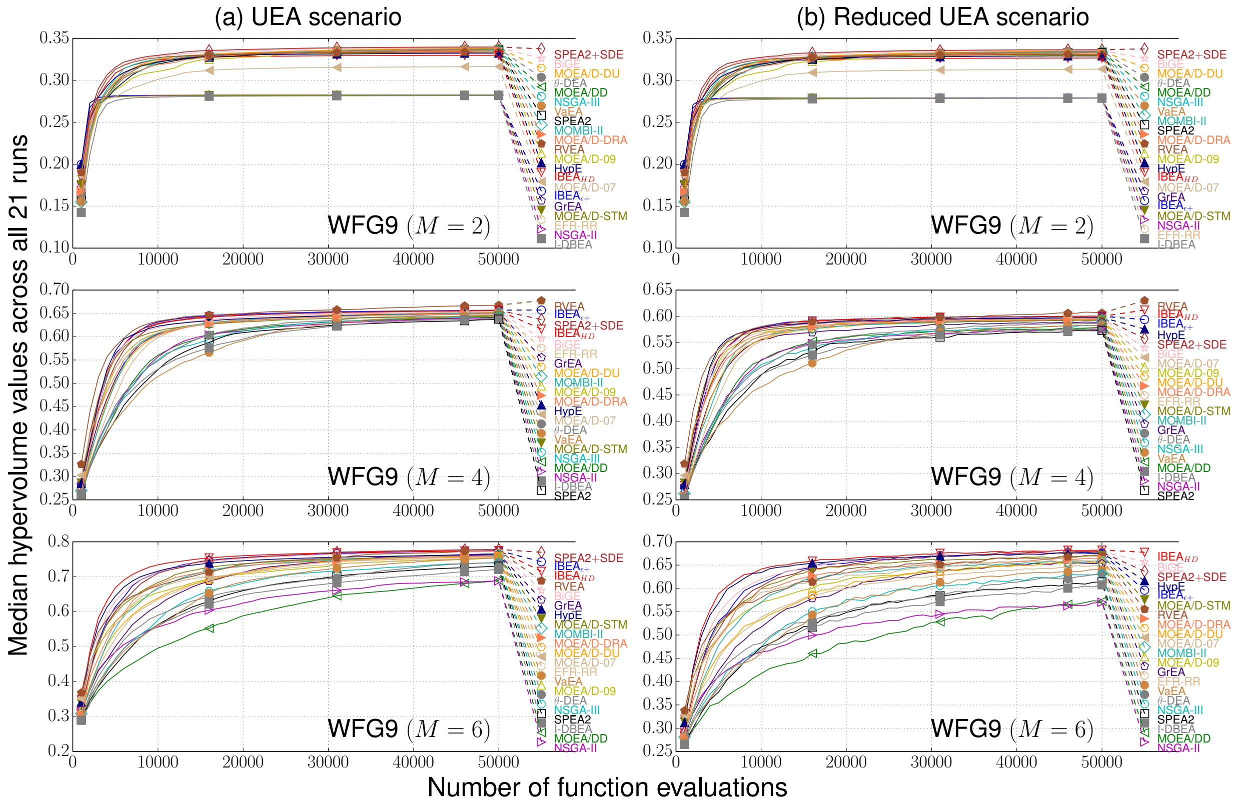


Fig. 13: Anytime performance of the 21 MOEAs on the normalized WFG9 problem ($M \in \{2, 4, 6\}$).

[25] M. Li, S. Yang, and X. Liu, “Bi-goal evolution for many-objective optimization problems,” *Artif. Intell.*, vol. 228, pp. 45–65, 2015.

[26] R. H. Gómez and C. A. C. Coello, “Improved Metaheuristic Based on the R2 Indicator for Many-Objective Optimization,” in *GECCO*, 2015, pp. 679–686.

[27] O. Krause, T. Glasmachers, N. Hansen, and C. Igel, “Unbounded Population MO-CMA-ES for the Bi-Objective BBOB Test Suite,” in *GECCO (Companion)*, 2016, pp. 1177–1184.

[28] T. Hanne, “On the convergence of multiobjective evolutionary algorithms,” *EJOR*, vol. 117, no. 3, pp. 553–564, 1999.

[29] J. E. Fieldsend, R. M. Everson, and S. Singh, “Using unconstrained elite archives for multiobjective optimization,” *IEEE TEVC*, vol. 7, no. 3, pp. 305–323, 2003.

[30] M. López-Ibáñez, J. D. Knowles, and M. Laumanns, “On Sequential Online Archiving of Objective Vectors,” in *EMO*, 2011, pp. 46–60.

[31] A. Radulescu, M. López-Ibáñez, and T. Stützle, “Automatically Improving the Anytime Behaviour of Multiobjective Evolutionary Algorithms,” in *EMO*, 2013, pp. 825–840.

[32] D. Brockhoff, T. Tran, and N. Hansen, “Benchmarking Numerical Multiobjective Optimizers Revisited,” in *GECCO*, 2015, pp. 639–646.

[33] R. Wang, Z. Zhou, H. Ishibuchi, T. Liao, and T. Zhang, “Localized weighted sum method for many-objective optimization,” *IEEE TEVC*, 2017 (in press).

[34] T. Tusar, D. Brockhoff, N. Hansen, and A. Auger, “COCO: The Bi-objective Black Box Optimization Benchmarking (bbob-biobj) Test Suite,” *CoRR*, 2016.

[35] K. Bringmann, T. Friedrich, and P. Klitzke, “Generic Postprocessing via Subset Selection for Hypervolume and Epsilon-Indicator,” in *PPSN*, 2014, pp. 518–527.

[36] H. Ishibuchi, Y. Setoguchi, H. Masuda, and Y. Nojima, “How to compare many-objective algorithms under different settings of population and archive sizes,” in *IEEE CEC*, 2016, pp. 1149–1156.

[37] T. Tusar and B. Filipic, “Visualization of Pareto Front Approximations in Evolutionary Multiobjective Optimization: A Critical Review and the Prosection Method,” *IEEE TEVC*, vol. 19, no. 2, pp. 225–245, 2015.

[38] S. Bandaru, A. H. C. Ng, and K. Deb, “Data mining methods for knowledge discovery in multi-objective optimization: Part A – Survey,” *Expert Syst. Appl.*, vol. 70, pp. 139–159, 2017.

[39] H. Ishibuchi, H. Masuda, and Y. Nojima, “Comparing solution sets of different size in evolutionary many-objective optimization,” in *IEEE CEC*, 2015, pp. 2859–2866.

[40] E. Zitzler and L. Thiele, “Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach,” *IEEE TEVC*, vol. 3, no. 4, pp. 257–271, 1999.

[41] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. da Fonseca, “Performance assessment of multiobjective optimizers: an analysis and review,” *IEEE TEVC*, vol. 7, no. 2, pp. 117–132, 2003.

[42] Q. Zhang, A. Zhou, S. Zhao, P. N. Suganthan, W. Liu, and S. Tiwari, “Multiobjective optimization test instances for the CEC 2009 special session and competition,” Univ. of Essex, Tech. Rep., 2008.

[43] R. Cheng, M. Li, Y. Tian, X. Zhang, S. Yang, Y. Jin, and X. Yao, “Benchmark Functions for the CEC’2017 Competition on Many-Objective Optimization,” Univ. of Birmingham, Tech. Rep., 2017.

[44] A. J. Nebro, J. J. Durillo, C. A. C. Coello, F. Luna, and E. Alba, “A Study of Convergence Speed in Multi-objective Metaheuristics,” in *PPSN*, 2008, pp. 763–772.

[45] Y. Jin, “Surrogate-assisted evolutionary computation: Recent advances and future challenges,” *Swarm and Evol. Comput.*, vol. 1, no. 2, pp. 61–70, 2011.

[46] M. Li, S. Yang, X. Liu, and R. Shen, “A Comparative Study on Evolutionary Algorithms for Many-Objective Optimization,” in *EMO*, 2013, pp. 261–275.

[47] J. Maltese, B. M. Ombuki-Berman, and A. P. Engelbrecht, “A Scalability Study of Many-Objective Optimization Algorithms,” *IEEE TEVC*, 2017 (in press).

[48] L. C. T. Bezerra, M. López-Ibáñez, and T. Stützle, “A Large-Scale Experimental Evaluation of High-Performing Multi-and Many-Objective Evolutionary Algorithms,” IRIDIA, Tech. Rep. 2017-005, 2017.

[49] R. Wang, R. C. Purshouse, and P. J. Fleming, “Preference-Inspired Coevolutionary Algorithms for Many-Objective Optimization,” *IEEE TEVC*, vol. 17, no. 4, pp. 474–494, 2013.

[50] K. Deb, M. Mohan, and S. Mishra, “Evaluating the epsilon-Domination Based Multi-Objective Evolutionary Algorithm for a Quick Computation of Pareto-Optimal Solutions,” *Evol. Comput.*, vol. 13, no. 4, pp. 501–525, 2005.

[51] D. Hadka and P. M. Reed, "Diagnostic Assessment of Search Controls and Failure Modes in Many-Objective Evolutionary Optimization," *Evol. Comput.*, vol. 20, no. 3, pp. 423–452, 2012.
[52] R. Tanabe and A. Oyama, "Benchmarking MOEAs for multi- and many-objective optimization using an unbounded external archive," in *GECCO*, 2017, pp. 633–640. [Online]. Available: https://ryojitanabe.github.io/pdf/to-gecco2017.pdf
[53] R. C. Purshouse and P. J. Fleming, "On the Evolutionary Optimization of Many Conflicting Objectives," *IEEE TEVC*, vol. 11, no. 6, pp. 770–784, 2007.
[54] M. Köppen and K. Yoshida, "Substitute Distance Assignments in NSGA-II for Handling Many-objective Optimization Problems," in *EMO*, 2007, pp. 727–741.
[55] S. F. Adra and P. J. Fleming, "Diversity management in evolutionary many-objective optimization," *IEEE TEVC*, vol. 15, no. 2, pp. 183–195, 2011.
[56] J. Cheng, G. G. Yen, and G. Zhang, "A Many-Objective Evolutionary Algorithm With Enhanced Mating and Environmental Selections," *IEEE TEVC*, vol. 19, no. 4, pp. 592–605, 2015.
[57] X. Zhang, Y. Tian, and Y. Jin, "A Knee Point-Driven Evolutionary Algorithm for Many-Objective Optimization," *IEEE TEVC*, vol. 19, no. 6, pp. 761–776, 2015.
[58] S. Jiang and S. Yang, "A Strength Pareto Evolutionary Algorithm Based on Reference Direction for Multi-objective and Many-objective Optimization," *IEEE TEVC*, 2017 (in press).
[59] K. Ikeda, H. Kita, and S. Kobayashi, "Failure of Pareto-based MOEAs: does non-dominated really mean near to optimal?" in *IEEE CEC*, 2001, pp. 957–962.
[60] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler, "Combining Convergence and Diversity in Evolutionary Multiobjective Optimization," *Evol. Comput.*, vol. 10, no. 3, pp. 263–282, 2002.
[61] S. Kukkonen and J. Lampinen, "Ranking-Dominance and Many-Objective Optimization," in *IEEE CEC*, 2007, pp. 3983–3990.
[62] H. Sato, H. E. Aguirre, and K. Tanaka, "Controlling Dominance Area of Solutions and Its Impact on the Performance of MOEAs," in *EMO*, 2007, pp. 5–20.
[63] H. Liu, F. Gu, and Q. Zhang, "Decomposition of a Multiobjective Optimization Problem Into a Number of Simple Multiobjective Subproblems," *IEEE TEVC*, vol. 18, no. 3, pp. 450–455, 2014.
[64] H. Ishibuchi and T. Murata, "A multi-objective genetic local search algorithm and its application to flowshop scheduling," *IEEE Trans. SMC, Part C*, vol. 28, no. 3, pp. 392–403, 1998.
[65] T. Murata, H. Ishibuchi, and M. Gen, "Specification of Genetic Search Directions in Cellular Multi-objective Genetic Algorithms," in *EMO*, 2001, pp. 82–95.
[66] E. J. Hughes, "Evolutionary many-objective optimisation: many once or one many?" in *IEEE CEC*, 2005, pp. 222–227.
[67] ——, "MSOPS-II: A general-purpose many-objective optimiser," in *IEEE CEC*, 2007, pp. 3944–3951.
[68] A. Trivedi, D. Srinivasan, K. Sanyal, and A. Ghosh, "A Survey of Multiobjective Evolutionary Algorithms Based on Decomposition," *IEEE TEVC*, vol. 21, no. 3, pp. 440–462, 2017.
[69] R. Storn and K. Price, "Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces," *J. Global Optimiz.*, vol. 11, no. 4, pp. 341–359, 1997.
[70] A. Zhou and Q. Zhang, "Are All the Subproblems Equally Important? Resource Allocation in Decomposition-Based Multiobjective Evolutionary Algorithms," *IEEE TEVC*, vol. 20, no. 1, pp. 52–64, 2016.
[71] Z. Wang, Q. Zhang, A. Zhou, M. Gong, and L. Jiao, "Adaptive Replacement Strategies for MOEA/D," *IEEE Trans. Cyber.*, vol. 46, no. 2, pp. 474–486, 2016.
[72] G. Marquet, B. Derbel, A. Liefooghe, and E. Talbi, "Shake Them All! - Rethinking Selection and Replacement in MOEA/D," in *PPSN*, 2014, pp. 641–651.
[73] K. Li, S. Kwong, Q. Zhang, and K. Deb, "Interrelationship-Based Selection for Decomposition Multiobjective Optimization," *IEEE Trans. Cyber.*, vol. 45, no. 10, pp. 2076–2088, 2015.
[74] H. Sato, "Chain-reaction solution update in MOEA/D and its effects on multi- and many-objective optimization," *Soft Comput.*, vol. 20, no. 10, pp. 3803–3820, 2016.
[75] S. Zhao, P. N. Suganthan, and Q. Zhang, "Decomposition-Based Multiobjective Evolutionary Algorithm With an Ensemble of Neighborhood Sizes," *IEEE TEVC*, vol. 16, no. 3, pp. 442–446, 2012.
[76] T. Chiang and Y. Lai, "MOEA/D-AMS: Improving MOEA/D by an adaptive mating selection mechanism," in *IEEE CEC*, 2011, pp. 1473–1480.
[77] M. Pescador-Rojas, R. H. Gómez, E. Montero, N. Rojas-Morales, M. C. Riff, and C. A. C. Coello, "An Overview of Weighted and Unconstrained Scalarizing Functions," in *EMO*, 2017, pp. 499–513.
[78] H. Ishibuchi, Y. Sakane, N. Tsukamoto, and Y. Nojima, "Simultaneous use of different scalarizing functions in MOEA/D," in *GECCO*, 2010, pp. 519–526.
[79] Y. Qi, X. Ma, F. Liu, L. Jiao, J. Sun, and J. Wu, "MOEA/D with Adaptive Weight Adjustment," *Evol. Comput.*, vol. 22, no. 2, pp. 231–264, 2014.
[80] H. Sato, "Analysis of inverted PBI and comparison with other scalarizing functions in decomposition based MOEAs," *J. Heuristics*, vol. 21, no. 6, pp. 819–849, 2015.
[81] Y. Yuan, H. Xu, and B. Wang, "Evolutionary many-objective optimization using ensemble fitness ranking," in *GECCO*, 2014, pp. 669–676.
[82] D. H. Phan and J. Suzuki, "R2-IBEA: R2 indicator based evolutionary algorithm for multiobjective optimization," in *IEEE CEC*, 2013, pp. 1836–1845.
[83] R. H. Gómez and C. A. C. Coello, "MOMBI: A new metaheuristic for many-objective optimization based on the R2 indicator," in *IEEE CEC*, 2013, pp. 2488–2495.
[84] C. A. R. Villalobos and C. A. C. Coello, "A new multi-objective evolutionary algorithm based on a performance assessment indicator," in *GECCO*, 2012, pp. 505–512.
[85] B. Li, K. Tang, J. Li, and X. Yao, "Stochastic Ranking Algorithm for Many-Objective Optimization Based on Multiple Indicators," *IEEE TEVC*, vol. 20, no. 6, pp. 924–938, 2016.
[86] T. P. Runarsson and X. Yao, "Stochastic ranking for constrained evolutionary optimization," *IEEE TEVC*, vol. 4, no. 3, pp. 284–294, 2000.
[87] M. Li, S. Yang, and X. Liu, "Pareto or Non-Pareto: Bi-Criterion Evolution in Multiobjective Optimization," *IEEE TEVC*, vol. 20, no. 5, pp. 645–665, 2016.
[88] H. Wang, L. Jiao, and X. Yao, "Two_Arch2: An Improved Two-Archive Algorithm for Many-Objective Optimization," *IEEE TEVC*, vol. 19, no. 4, pp. 524–541, 2015.
[89] K. Praditwong and X. Yao, "A New Multi-objective Evolutionary Optimisation Algorithm: The Two-Archive Algorithm," in *CIS*, 2006, pp. 95–104.
[90] A. Toffolo and E. Benini, "Genetic Diversity as an Objective in Multi-Objective Evolutionary Algorithms," *Evol. Comput.*, vol. 11, no. 2, pp. 151–167, 2003.
[91] N. Beume, B. Naujoks, and M. T. M. Emmerich, "SMS-EMOA: multiobjective selection based on dominated hypervolume," *EJOR*, vol. 181, no. 3, pp. 1653–1669, 2007.
[92] A. P. Guerreiro, C. M. Fonseca, and L. Paquete, "Greedy Hypervolume Subset Selection in Low Dimensions," *Evol. Comput.*, vol. 24, no. 3, pp. 521–544, 2016.
[93] H. Ishibuchi, Y. Sakane, N. Tsukamoto, and Y. Nojima, "Selecting a small number of representative non-dominated solutions by a hypervolume-based solution selection approach," in *FUZZ-IEEE*, 2009, pp. 1609–1614.
[94] M. Basseur, B. Derbel, A. Goëffon, and A. Liefooghe, "Experiments on Greedy and Local Search Heuristics for *d*dimensional Hypervolume Subset Selection," in *GECCO*, 2016, pp. 541–548.
[95] S. Huband, P. Hingston, L. Barone, and R. L. While, "A review of multiobjective test problems and a scalable test problem toolkit," *IEEE TEVC*, vol. 10, no. 5, pp. 477–506, 2006.
[96] H. Ishibuchi, H. Masuda, and Y. Nojima, "Pareto Fronts of Many-Objective Degenerate Test Problems," *IEEE TEVC*, vol. 20, no. 5, pp. 807–813, 2016.
[97] H. Ishibuchi, Y. Setoguchi, H. Masuda, and Y. Nojima, "Performance of Decomposition-Based Many-Objective Algorithms Strongly Depends on Pareto Front Shapes," *IEEE TEVC*, vol. 21, no. 2, pp. 169–190, 2017.
[98] D. Brockhoff, "A Bug in the Multiobjective Optimizer IBEA: Salutary Lessons for Code Release and a Performance Re-Assessment," in *EMO*, 2015, pp. 187–201.
[99] R. Tanabe and A. Oyama, "The Impact of Population Size, Number of Children, and Number of Reference Points on the Performance of NSGA-III," in *EMO*, 2017, pp. 606–621.
[100] M. Li, S. Yang, and X. Liu, "Diversity Comparison of Pareto Front Approximations in Many-Objective Optimization," *IEEE Trans. Cyber.*, vol. 44, no. 12, pp. 2568–2584, 2014.
[101] H. Wang, Y. Jin, and X. Yao, "Diversity Assessment in Many-Objective Optimization," *IEEE Trans. Cyber.*, vol. 47, no. 6, pp. 1510–1522, 2017.

[102] T. Weise, R. Chiong, J. Lässig, K. Tang, S. Tsutsui, W. Chen, Z. Michalewicz, and X. Yao, "Benchmarking Optimization Algorithms: An Open Source Framework for the Traveling Salesman Problem," *IEEE CIM*, vol. 9, no. 3, pp. 40–52, 2014.

[103] H. Ishibuchi, N. Akedo, and Y. Nojima, "Behavior of Multiobjective Evolutionary Algorithms on Many-Objective Knapsack Problems," *IEEE TEVC*, vol. 19, no. 2, pp. 264–283, 2015.