Objective Reduction in Many-Objective Optimization: Evolutionary Multiobjective Approaches and Comprehensive Analysis

Yuan Yuan, Yew-Soon Ong, Abhishek Gupta, and Hua Xu

***Abstract*—Many-objective optimization problems bring great difficulties to the existing multiobjective evolutionary algorithms, in terms of selection operators, computational cost, visualization of the high-dimensional tradeoff front, and so on. Objective reduction can alleviate such difficulties by removing the redundant objectives in the original objective set, which has become one of the most important techniques in many-objective optimization. In this paper, we suggest to view objective reduction as a multiobjective search problem and introduce three multiobjective formulations of the problem, where the first two formulations are both based on preservation of the dominance structure and the third one utilizes the correlation between objectives. For each multiobjective formulation, a multiobjective objective reduction algorithm is proposed by employing the nondominated sorting genetic algorithm II to generate a Pareto front of nondominated objective subsets that can offer decision support to the user. Moreover, we conduct a comprehensive analysis of two major categories of objective reduction approaches based on several theorems, with the aim of revealing their strengths and limitations. Lastly, the performance of the proposed multiobjective algorithms is studied extensively on various benchmark problems and two real-world problems. Numerical results and comparisons are then shown to highlight the effectiveness and superiority of the proposed multiobjective algorithms over existing state-of-the-art approaches in the related field.**

***Index Terms*—Many-objective optimization, multiobjective evolutionary algorithms (MOEAs), multiobjective optimization, objective reduction.**

I. INTRODUCTION

IN PRACTICE, it is desirable with most applications to consider as many objectives as possible in order to better satisfy various performance demands [1], [2]. Resultant problem formulations lead to the existence of multiobjective optimization problems (MOPs) with more than three objectives, which are typically referred to as many-objective optimization problems (MaOPs) in many real-world scenarios [3]–[6]. Recently, MaOPs have stirred growing interest in the evolutionary multiobjective optimization (EMO) community, since they have posed great challenges to most existing categories of multiobjective evolutionary algorithms (MOEAs) [7], [8], including Pareto dominance- [9], [10], indicator- [11], [12], and decomposition-based [13], [14] MOEAs.

Over the past decade, some contributions have been made to counter the limitations of existing MOEAs in many objective optimization. For instance, alternative dominance relations [15]–[17] and new diversity promotion mechanisms [18]–[22] were proposed for Pareto dominance-based MOEAs. Fast approximation of hypervolume (HV) values [23], [24] and other computationally efficient performance indicators [25], [26] were studied for indicator-based MOEAs. Further, novel updating strategies [27], [28] and the improved generation of well distributed weight vectors [29]–[31] were presented for decomposition-based MOEAs. Nevertheless, the recently proposed many-objective algorithms may not adequately handle MaOPs with more than around 15 objectives [17], [20], with their effectiveness needing further examination on real-world problems. Moreover, the improvements on MOEAs cannot alleviate the difficulty of visualizing the high-dimensional Pareto front which presents a steep challenge in choosing preferred solutions in many-objective optimization [20].

Instead of improving the scalability of existing MOEAs, *objective reduction* attempts to decrease the problem difficulty by reducing the number of objectives in the decision making stage and/or during the search process [2], [7], [20], [32]. The motivation behind the approach is that for many problems with m objectives, there exists a minimum cardinality subset of k ($k < m$) conflicting objectives that can generate the same Pareto front as the original problem. Such k objectives are often deemed as being *essential*, while all the other objectives are termed as *redundant* [2]. The potential benefits of objective reduction have been clearly highlighted in several studies [1], [2], [33], which can be understood from two main perspectives. On one hand, for a MaOP of interest, if the number of objectives can be reduced to no more than three, any state-of-the-art MOEA can be expected to solve

Manuscript received April 23, 2016; revised September 13, 2016 and January 1, 2017; accepted January 22, 2017. Date of publication February 22, 2017; date of current version March 28, 2018. This work was supported in part by the A*Star-TSRP, in part by the Singapore Institute of Manufacturing Technology—Nanyang Technological University (NTU) Joint Laboratory and Collaborative Research Programme on Complex Systems, and in part by the Computational Intelligence Graduate Laboratory at NTU. The work of H. Xu was supported by the National Natural Science Foundation of China under Grant 61673235.

Y. Yuan, Y.-S. Ong, and A. Gupta are with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: yyxhdy@gmail.com; asysong@ntu.edu.sg; abhishekg@ntu.edu.sg).

H. Xu is with the State Key Laboratory of Intelligent Technology and Systems, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: xuhua@tsinghua.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TEVC.2017.2672668

it efficiently; on the other hand, if the remaining number of objectives is larger than three, objective reduction continues to be helpful in making the subsequent search of multiobjective or many-objective optimizers more effective and efficient, and also eases the visualization of the Pareto front and the decision making process.

Note that, the following previous research [1], [2], [7], [32], objective reduction in this paper refers in particular to the selection of a subset of given objectives which best describe the original MOP. This approach is somewhat similar to *feature selection*. However, it is worth noting there exists another related research direction aiming to determine a small set of arbitrary objectives for the original MOP in the spirit of *feature transformation*, as is commonly used in visualization of many-objective solution sets [34], [35]. In this regard, a prior work similar to ours is by Köppen and Yoshida [36], where the authors attempted to map a set of nondominated points in higher-dimensional space to a 2-D Euclidian space for visualization, whilst preserving both distance relations and Pareto dominance relations as much as possible. They used nondominated sorting genetic algorithm (NSGA)-II to find a tradeoff between the preservation of the two relations.

This paper focuses on objective reduction related to feature selection. Feature selection has been extensively studied in machine learning and statistics, and plenty of relevant methods [37] are now available in the literature. Unfortunately, these standard feature selection techniques cannot be applied directly to objective reduction in many-objective optimization since the Pareto dominance structure [1], [38], [39] or the conflict relation between objectives [33] must be taken into account. However, the basic ideas in data analysis are still quite useful for developing effective approaches to objective reduction, and have already inspired several significant contributions [2], [33], [40]. Evolutionary computation (EC) approaches have gained much attention in feature selection [41] due to their global search potential in the large-scale search spaces. However, to the best of our knowledge, no prior research has been made on using EC for objective reduction in the sense of feature selection. This is despite the fact that the δ-MOSS and k-EMOSS problems [1] resulting from objective reduction have been both proven to be NP-hard and the exact algorithm is not applicable for larger instances of practical size [1]. Moreover, most existing objective reduction algorithms [2], [39], [42], [43] only return a unique reduced objective set as the outcome in a single simulation run, without considering conflicting requirements, such as the error tolerance and the desired number of objectives. Thus, they fail to offer necessary flexibility for decision support [44] to the user. Taking this cue, this paper presents the first study on evolutionary multiobjective approaches to objective reduction, exploiting both global search ability of EC and decision support characteristic of multiobjective optimization. More specifically, we propose three different multiobjective formulations of the objective reduction problem using three different kinds of error measures (δ, η, and γ), where the first two measures consider the Pareto dominance structure while the last focuses on the correlation structure. Then, we employ NSGA-II as an effective solver for the three formulated MOPs, leading to three evolutionary multiobjective approaches (NSGA-II-δ, NSGA-II-η, and NSGA-II-γ) to objective reduction.

Another highlight of this paper lies in the comprehensive analysis of existing objective reduction algorithms, which can be roughly categorized into *dominance structure-based approaches* and *correlation-based approaches* [1], [2], [39]. The motivation comes from the fact that, although the two types of approaches have their own advantages and disadvantages, past studies on objective reduction fail to analyze their behavior appropriately. For example, Brockhoff and Zitzler [1] pointed out that the correlation-based approaches cannot guarantee preservation of Pareto dominance relation, but they did not further indicate in what scenarios the correlation-based approaches would fail due to this reason. Also, they did not discuss the potential drawbacks of dominance structure-based approaches and the rationale of correlation-based approaches. Saxena *et al.* [2] implied that the dominance structure-based approaches are extremely sensitive to misdirection, but their inference is only based on the performance of the algorithms proposed in [1] with the δ error fixed at 0. Additionally, their conclusions were drawn mainly based on numerical results, thereby making it difficult to satisfactorily uncover the true strengths of dominance structure-based approaches and the inherent limitations of correlation-based approaches in general. In this paper, we undertake a comprehensive study of the two categories of approaches through a theoretical treatment of objective reduction which clearly highlights their general strengths and weaknesses. Our goal is to enable users of these algorithms to have a better understanding of why, how and when they work. Moreover, our analysis is expected to provide deeper insights for designing more effective objective reduction algorithms in the future.

In order to verify the performance of the proposed multiobjective methods and analyze the two categories of approaches comprehensively, we conduct a series of experiments with benchmark problems with detailed comparisons to current state-of-the-art algorithms. Experimental results not only show the benefits of our proposed algorithms but also agree well with the analysis. Furthermore, we apply the proposed multiobjective approaches to two real-world applications and demonstrate their efficacies in optimization, visualization and decision making.

This paper takes inspiration from Brockhoff and Zitzler [1], Saxena *et al.* [2], Jaimes *et al.* [33], and Singh *et al.* [39]. Based on these previous studies, our contributions to the topic of objective reduction are summarized as follows.

1) Inspired by Brockhoff and Zitzler [1], we propose to formulate objective reduction as an MOP and use NSGA-II to derive a tradeoff between the δ error and the size of the objective subset. Compared to the greedy algorithms in [1] that can only solve a specific δ-MOSS or k-EMOSS problem at a time, the proposed approach indeed solves all δ-MOSS and k-EMOSS problems at once, thereby providing more comprehensive analysis and better decision support in objective reduction. Moreover, the evolutionary multiobjective search is verified to be superior to the greedy and exact search used in [1].

2) Inspired by the parameter R in [39], we introduce a new error based on the dominance structure change, referred to as η, which is easier to comprehend than δ error from the point of view of decision support.
3) Inspired by the algorithms introduced in [33], we present a new error, referred to as γ, to measure the change of correlation structure. The deduction of γ is characterized by the use of Kendall's rank correlation coefficient instead of Pearson's, and the clustering procedure that takes the selected objectives as cluster centers.
4) Compared to [2], [33], and [39], the proposed approaches differ not only in the decision support characteristic derived from the multiobjective formulation, but also in the search mechanism.
5) Inspired by the comments on the dominance structure- and correlation-based approaches in [1] and [2], we carry out further analysis of their strengths and limitations based on theoretical and experimental results. Some new insights are provided for the two categories of objective reduction approaches.

The remainder of this paper is organized as follows. Section II presents the preliminaries and background on objective reduction. In Section III, the proposed evolutionary multiobjective approaches are described in detail. Section IV conducts a rigorous analysis of dominance structure- and correlation-based approaches based on several theorems in order to show their general strengths and limitations. Section V presents experiments with some objective reduction algorithms for benchmark problems. In Section VI, the proposed multiobjective approaches are applied to two real-world problems. In Section VII, the benefits of the proposed approaches are illustrated in optimization, visualization and decision making. Lastly, concluding remarks and directions for future work are given in Section VIII.

II. Preliminaries and Background

A. Multiobjective and Many-Objective Optimization

A general MOP can be mathematically formulated as

$$\begin{aligned} \min\ & \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}))^{\mathrm{T}} \\ \text{subject to}\ & g_i(\mathbf{x}) \le 0,\ \ i = 1, 2, \ldots, u \\ & h_j(\mathbf{x}) = 0,\ \ j = 1, 2, \ldots, v \\ & \mathbf{x} \in \Omega. \end{aligned} \tag{1}$$

$\mathbf{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ is a n-dimensional decision vector in the decision space Ω; $\mathbf{f} : \Omega \to \Theta \subseteq \mathbb{R}^m$, is an objective vector consisting of m objective functions, which maps n-dimensional decision space Ω to m-dimensional attainable objective space Θ. For MaOPs, $m > 3$. $g_i(\mathbf{x}) \le 0$ and $h_j(\mathbf{x}) = 0$ represent inequality and equality constraints, respectively.

In an MOP, there typically exists no solution that minimizes all objective functions simultaneously. Therefore, the attention is paid to approximating the *Pareto front* that represents optimal tradeoffs between objectives. Several basic concepts about MOPs are provided below.

Definition 1 (Weak Pareto Dominance): A vector $\mathbf{u} = (u_1, u_2, \ldots, u_k)^{\mathrm{T}}$ is said to *weakly dominate* another vector $\mathbf{v} = (v_1, v_2, \ldots, v_k)^{\mathrm{T}}$, denoted by $\mathbf{u} \preceq \mathbf{v}$, iff $\forall i \in \{1, 2, \ldots, k\} : u_i \le v_i$.

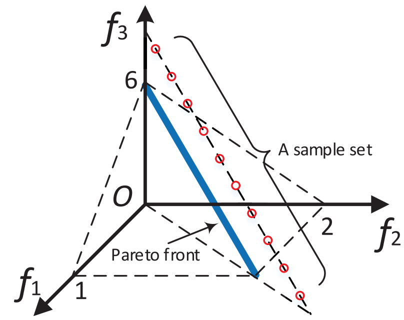


Fig. 1. Pareto front and a sample set of the three-objective WFG3 problem.

Definition 2 (Pareto Dominance): A vector $\mathbf{u} = (u_1, u_2, \ldots, u_k)^{\mathrm{T}}$ is said to *dominate* another vector $\mathbf{v} = (v_1, v_2, \ldots, v_k)^{\mathrm{T}}$, denoted by $\mathbf{u} \prec \mathbf{v}$, iff $\forall i \in \{1, 2, \ldots, k\} : u_i \le v_i$ and $\exists j \in \{1, 2, \ldots, k\} : u_j < v_j$.

Definition 3 (Pareto Front): The Pareto front of an MOP is defined as $\mathrm{PF} := \{\mathbf{f}(\mathbf{x}^*) \in \Theta | \nexists \mathbf{x} \in \Omega, \mathbf{f}(\mathbf{x}) \prec \mathbf{f}(\mathbf{x}^*)\}$.

Definition 4 (Weak ϵ-Dominance [45]): A vector $\mathbf{u} = (u_1, u_2, \ldots, u_k)^{\mathrm{T}}$ is said to *weakly ϵ-dominate* another vector $\mathbf{v} = (v_1, v_2, \ldots, v_k)^{\mathrm{T}}$, denoted by $\mathbf{u} \preceq^{\epsilon} \mathbf{v}$, iff $\forall i \in \{1, 2, \ldots, k\} : u_i - \epsilon \le v_i$.

Roughly speaking, the goal of MOEAs is to find the best possible Pareto front-*approximation*, i.e., the obtained nondominated objective vectors should be close to the Pareto front (convergence) and also distributed well along the Pareto front (diversity).

B. Basic Concepts in Objective Reduction

Unless otherwise specified, the given MOP for objective reduction always has the formulation defined in (1); the original (or universe) objective set is denoted as $\mathcal{F}_0 = \{f_1, f_2, \ldots, f_m\}$; PF_0 refers to the Pareto front of the original MOP. For convenience, the notation $\mathbf{u}^{(\mathcal{F})}$ is used to denote the subvector of $\mathbf{u}$ given by a nonempty objective subset $\mathcal{F}$. For example, if $\mathbf{u} = (f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x}))^{\mathrm{T}}$, $\mathcal{F} := \{f_1, f_3\}$, then $\mathbf{u}^{(\mathcal{F})} = (f_1(\mathbf{x}), f_3(\mathbf{x}))^{\mathrm{T}}$.

One of the basic aims of objective reduction is to find an *essential objective set* [2] for a given MOP, whose definition and related concepts are given as follows.

Definition 5: An objective subset $\mathcal{F} \subset \mathcal{F}_0$ is said to be *redundant*, iff the Pareto front corresponding to $\mathcal{F}' := \mathcal{F}_0 \backslash \mathcal{F}$ is $\mathrm{PF}' = \{\mathbf{u}^{(\mathcal{F}')} | \mathbf{u} \in \mathrm{PF}_0\}$

Definition 6: An objective set $\mathcal{F}$ is said to be an *essential objective set*, iff $\mathcal{F}_0 \backslash \mathcal{F}$ is a redundant objective subset with the largest cardinality.

Definition 7: The *dimensionality* of a given MOP or PF_0 refers to the cardinality of an essential objective set.

To illustrate the above definitions, Fig. 1 shows the Pareto front of the three-objective WFG3 problem [46], [47], which is a straight line from $(0, 0, 6)^{\mathrm{T}}$ to $(1, 2, 0)^{\mathrm{T}}$.[1] From Fig. 1,

[1] Ishibuchi *et al.* [47] showed that the Pareto front of the original WFG3 [46] has a nondegenerate part. In this paper, we study the transformed version of WFG3 [47], where the nondegenerate part is removed by the introduction of constraints.

the objectives sets $\{f_1\}$ and $\{f_2\}$ are both redundant; there exist two essential objective sets $\{f_1, f_3\}$ and $\{f_2, f_3\}$; and hence the dimensionality of this problem is 2. As illustrated, there may exist different essential objective sets for a given MOP; the concept "redundant" is with respect to $\mathcal{F}_0$ and two redundant objective sets cannot always be removed simultaneously.

"Nonconflicting" is another common concept in objective reduction, whose definition is as follows.

Definition 8: Two objectives $f_i, f_j \in \mathcal{F}_0$ are said to be *nonconflicting*, if $\forall \mathbf{u}, \mathbf{v} \in \mathrm{PF}_0 : \mathbf{u}^{(\{f_i\})} \leq \mathbf{v}^{(\{f_i\})} \Leftrightarrow \mathbf{u}^{(\{f_j\})} \leq \mathbf{v}^{(\{f_j\})}$.

Its relationship with the concept redundant is that, if f_i and f_j are nonconflicting, then both $\{f_i\}$ and $\{f_j\}$ are redundant (e.g., f_1 and f_2 in Fig. 1), which will be proven in the supplementary material. The degree of conflict can be estimated by the correlation between two objectives. In general, a more negative correlation between two objectives indicates they are more conflicting, whereas a more positive correlation means they tend more to be nonconflicting.

C. Pareto Front-Representation and Misdirection

Objective reduction approaches generally operate on a set of nondominated objective vectors obtained from an MOEA, hereafter called the *sample set*. In objective reduction, the sample set is not necessarily required to be a good Pareto front-*approximation*, but a good Pareto front-*representation* that refers to a solution set that conforms well with the dominance structure or the correlation structure of the Pareto front. The difference between "approximation" and "representation" has been explained in detail by Saxena *et al.* [2].

In Fig. 1, the presented sample set provides a perfect Pareto front-representation of the three-objective WFG3 problem, because its dominance/correlation structure is completely consistent with that of the Pareto front, i.e., f_1 and f_2 are nonconflicting with each other and either of them is entirely conflicting with f_3. Given this sample set, an objective reduction algorithm based on dominance or correlation can identify the essential objectives reliably. However, note that, this set is indeed not a good Pareto front-approximation since it is far away from the true Pareto front.

In practice, the sample set is generally not a perfect Pareto front-representation, and there is almost always a certain degree of *misdirection* [2], [44] in it. The misdirection here refers to the difference in the dominance structure or the correlation structure, between the Pareto front and the sample set [44].[2] To further explain this, Fig. 2 shows a sample set of three-objective WFG3 which is projected on the $f_2 - f_3$ and $f_1 - f_2$ objective subspaces, respectively. As seen from Fig. 2(a), the dominance structure in this sample set is a little different from that in the Pareto front, because some solutions become dominated with respect to the $f_2 - f_3$ objective subspace; these solutions contribute to the difference that can be interpreted as misdirection, while the rest can be interpreted as *signal*. Similarly, from Fig. 2(b), the correlation structure

[2]References [2] and [44] used the term "noise" to refer to this concept. In this paper, we use the term "misdirection" instead to avoid confusion stemming from the general meaning of noise.

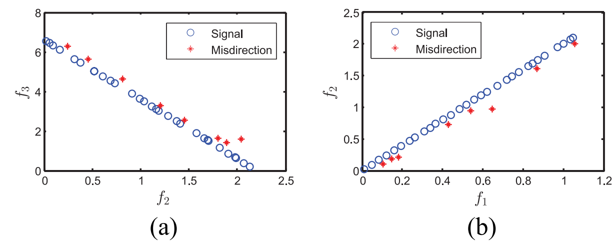


Fig. 2. Illustration of misdirection using a sample set of three-objective WFG3. Projection onto the (a) f_2-f_3 plane and (b) f_1-f_2 plane.

in the Pareto front is also violated slightly by this sample set, where f_1 and f_2 are not perfectly positively correlated. A well-designed objective reduction algorithm is expected to work well with a varying degree of misdirection [2].

D. Existing Approaches to Objective Reduction

In recent years, some objective reduction algorithms have been proposed in the literature. As mentioned before, these algorithms can be broadly classified into two categories in essence: 1) dominance structure- and 2) correlation-based approaches. Dominance structure-based approaches aim to preserve the dominance structure as much as possible after removing objectives, while correlation-based approaches exploit the correlation within objective pairs and consider to keep the most conflicting objectives and eliminate the objectives that are nonconflicting with others. In the following, two typical studies are highlighted, which belong to the dominance structure- and correlation-based approaches, respectively.

1) Dominance Relation Preservation [1]: This objective reduction approach is based on preserving the weak Pareto dominance relations. For a particular objective subset $\mathcal{F} \subseteq \mathcal{F}_0$, this approach considers each solution pair $\mathbf{u}, \mathbf{v} \in \mathcal{N}$ satisfying $\mathbf{u}^{(\mathcal{F})} \preceq \mathbf{v}^{(\mathcal{F})}$, and computes the minimum nonnegative ϵ that ensures $\mathbf{u} \preceq^{\epsilon} \mathbf{v}$ for each pair. After examining all possible solution pairs in $\mathcal{N}$, the error equals to the maximum ϵ recorded. This error is denoted as δ, which can be viewed as a criterion to measure the degree of conflict between $\mathcal{F}$ and $\mathcal{F}_0$. Based on this criterion, two problems regarding objective reduction are formalized as follows.

Definition 9 (δ-MOSS Problem): Given a $\delta_0 \geq 0$ and a sample set $\mathcal{N}$, the problem is to compute a smallest subset $\mathcal{F} \subseteq \mathcal{F}_0$ satisfying the condition that its associated δ value is no greater than δ_0.

Definition 10 (k-EMOSS Problem): Given a $k_0 \in \mathbb{N}^+$ and a sample set $\mathcal{N}$, the problem is to compute an objective subset $\mathcal{F} \subseteq \mathcal{F}_0$ with the minimum δ value in the premise $|\mathcal{F}| \leq k_0$.

In the study by Brockhoff and Zitzler [1], an exact algorithm for both δ-MOSS and k-EMOSS problems was proposed. Given that the two problems are both NP-hard, three greedy algorithms were also developed for large instances of δ-MOSS and k-EMOSS, respectively.

2) Principal Component Analysis and Maximum Variance Unfolding [2]: Based on two well-known dimensionality reduction techniques, i.e., principal component analysis (PCA) and maximum variance unfolding (MVU), this paper proposed two algorithms, namely L-PCA and NL-MVU-PCA, for linear

TABLE I
SUMMARY OF THE STUDIES REVIEWED IN THIS PAPER

Algorithm Category	References
Dominance structure-based approaches	[1], [39], [48] [49], [50]
Correlation-based approaches	[2], [33], [42], [43] [44], [51]–[54]

and nonlinear objective reduction, respectively. The core idea is to find the smallest set of conflicting objectives preserving the correlation structure in the given sample set, which is achieved by removing objectives that are nonconflicting along the significant eigenvectors of the correlation matrix (for L-PCA) or the kernel matrix (for NL-MVU-PCA). In the experiments, L-PCA and NL-MVU-PCA were compared with the exact and greedy algorithms in [1] in the context of identifying the essential objective set.

Besides the aforementioned works, there are a few other noteworthy contributions in objective reduction. Jaimes *et al.* [33] presented an approach based on an unsupervised feature selection technique, where the objective set is divided into neighborhoods of fixed size around each objective according to the correlation strength. Singh *et al.* [39] proposed a Pareto corner search evolutionary algorithm (PCSEA) that only searches for the corners of the Pareto front instead of the complete Pareto front. The population obtained by PCSEA is then used for the objective reduction, which is based on the premise that the number of nondominated solutions will be affected substantially by omitting an essential objective. Guo *et al.* [42] proposed to employ a k-medoids clustering algorithm to identify potentially redundant objectives by merging the more correlated objectives into the same cluster, where the clustering is based on a metric combing mutual information and correlation coefficient. Duro *et al.* [44] extended the framework proposed in [2] to include the analysis analogous to δ-MOSS and k-EMOSS, which could serve as decision support for decision makers. De Freitas *et al.* [48] introduced the concept of aggregation trees for the visualization of the results of MaOPs, and also used it as a means of performing objective reduction. Guo *et al.* [49] claimed that the corners used in PCSEA are insufficient to reflect the entire Pareto front, and thereby developed an algorithm using the representative nondominated solutions instead. Wang and Yao [43] adopted the nonlinear correlation information entropy to measure both linear and nonlinear correlation between objectives and then used a simple method to select the most conflicting objectives.

Moreover, it is worth mentioning that several studies [50]–[54] have focused on integrating the objective reduction algorithms into MOEAs in order to simplify the search. This paradigm is often referred to as *online* objective reduction. However, in this paper, our attention is concentrated on the objective reduction algorithms themselves, so *offline* objective reduction is mainly considered. In Table I, we summarize the reviewed studies on objective reduction in each category.

At this juncture, it is important to distinguish two different issues related to the objective reduction algorithms.

1) δ-MOSS and k-EMOSS analysis.
2) Identification of an essential objective set.

Some existing studies (see [1], [33]) are devoted to the first issue and do not explicitly consider the second one, while the others (see [2], [39]) only considered the second issue. The algorithms for the first issue can provide the decision support for the user by investigating δ-MOSS problems with different δs, and k-EMOSS problems with different ks; the selected objective subset may not be an essential objective set, but a desirable one consistent with the user's preference in a specific application scenario. As for the second issue, it only depends on the MaOP at hand and has nothing to do with the user's preference. Note that an algorithm for the first issue can be easily used for the second one by solving a δ-MOSS problem with a sufficiently small δ value [2]. In this paper, the proposed multiobjective approaches adequately address the first issue that is contended to be more practical. However, it appears difficult to have an experimental evaluation and comparison of the effect of objective reduction in this respect. Therefore, the proposed multiobjective approaches are applied to the second issue when analyzing their strengths and limitations in objective reduction.

III. PROPOSED MULTIOBJECTIVE APPROACHES

This section introduces the proposed multiobjective approaches in detail. First, we describe how to formulate the objective reduction problem as an MOP. Three multiobjective formulations are provided: the first two (introduced in Section III-A) are based on the dominance structure while the third (introduced in Section III-B) is based on the correlation between objectives. Then, Section III-C describes how to use MOEAs to solve the MOPs formulated in Sections III-A and III-B. Lastly, Section III-D illustrates that the proposed multiobjective approaches are well suited to objective reduction.

A. Dominance Structure-Based Multiobjective Formulation

The objective reduction problem takes the original objective set $\mathcal{F}_0$ and the sample set $\mathcal{N}$ as input, and its candidate solution is in the form of an objective subset $\mathcal{F}$.

The first multiobjective formulation of objective reduction, namely δ-OR, has two objectives, which are minimizing both number of selected objectives k ($k = |\mathcal{F}|$) and δ error proposed in [1]. As depicted in Section II-D, δ is mathematically expressed as follows:

$$\delta = \max \left\{ \left\{ \min_{\substack{\epsilon \geq 0 \\ \mathbf{u} \preceq^{\epsilon} \mathbf{v}}} \epsilon | \mathbf{u}, \mathbf{v} \in \mathcal{N}, \mathbf{u}^{(\mathcal{F})} \preceq \mathbf{v}^{(\mathcal{F})} \right\} \bigcup \{0\} \right\}. \quad (2)$$

δ can be seen as a criterion measuring the degree of dominance structure change between $\mathcal{F}$ and $\mathcal{F}_0$. k and δ are conflicting with each other to some extent. Usually, δ would become smaller when k gets larger, and in the extreme case of maximum possible k value, i.e., $k = |\mathcal{F}_0|$, the minimum possible delta value is achieved, i.e., $\delta = 0$. With δ-OR formulation, the goal of objective reduction is to explore the Pareto front of objective subsets.

Note that, when using δ error, all objective values must have the same scale such that ϵ values are comparable among objectives [1]. Moreover, δ error might be misleading if the objective functions have different degree of nonlinearity [44]. Here, inspired by the definition of the parameter R in [39], we propose an alternative error to measure the dominance structure change, denoted as η, which is free from the potential disadvantages of δ mentioned above.

To compute the η error for an objective subset $\mathcal{F} \subseteq \mathcal{F}_0$, the sample set $\mathcal{N}$ is divided into two disjoint sets $\mathcal{N}_{\mathcal{NS}}$ and $\mathcal{N}_{\mathcal{DS}}$, where $\mathcal{N}_{\mathcal{NS}} = \{\mathbf{u} \in \mathcal{N} | \nexists \mathbf{v} \in \mathcal{N} : \mathbf{v}^{(\mathcal{F})} \prec \mathbf{u}^{(\mathcal{F})}\}$ and $\mathcal{N}_{\mathcal{DS}} := \mathcal{N} \backslash \mathcal{N}_{\mathcal{NS}}$. Then η is expressed as

$$\eta = |\mathcal{N}_{\mathcal{DS}}|/|\mathcal{N}|. \tag{3}$$

Note that, unlike δ, η naturally lies in the range $[0, 1)$, which is problem independent and indeed represents the proportion of dominated (with respect to $\mathcal{F}$) objective vectors among the sample set $\mathcal{N}$.

By replacing δ error in δ-OR with η error, we have the second multiobjective formulation of objective reduction, namely η-OR. The conflict relation between k and η is similar to that between k and δ, with the maximum $k = |\mathcal{F}_0|$ also ensuring η reaches a minimum value 0, since the objective vectors in $\mathcal{N}$ are mutually nondominated with respect to $\mathcal{F}_0$.

B. Correlation-Based Multiobjective Formulation

This section introduces another multiobjective formulation of objective reduction, namely γ-OR. Unlike δ-OR and η-OR, γ-OR is based on the correlation rather than dominance structure. Although there exist several correlation-based objective reduction algorithms as reviewed in Section II-D, there is no off-the-shelf criterion that evaluates any given objective subset based on the correlation analysis. Now, we provide such a criterion, denoted as γ, which is inspired by the objective reduction procedures described in [33].

The computing of γ needs to exploit the correlation between every pair of objectives in $\mathcal{F}_0$. Instead of using Pearson's correlation coefficient as in [2] and [33], Kendall's [55] rank correlation coefficient is adopted here. This is because Pearson's correlation coefficient is sensitive only to linear relationships between two variables. On the other hand, Kendall's rank correlation coefficient measures the extent to which, as one variable increases, the other variable tends to increase, without regard to whether the increase is represented by a linear or nonlinear relationship, which matches well with the purpose of correlation analysis in objective reduction, i.e., identifying whether the improvement of one objective would deteriorate/improve the other objective [38].

Let $r_\tau(i, j)$ denote the Kendall's rank correlation coefficient between two objectives $f_i, f_j \in \mathcal{F}_0$, which can be calculated by the given sample set $\mathcal{N}$. The closer $r_\tau(i, j)$ is to 1 (-1), the stronger the monotonic increasing (decreasing) relationship is. Define $d(i, j) = (1 - r_\tau(i, j))/2$ as the distance between f_i and f_j. Thus $d(i, j) \in [0, 1]$, 0 (1) indicates f_i and f_j are completely positively (negatively) correlated. Note that $d(i, j) = d(j, i)$ since $r_\tau(i, j)$ is symmetric.

Algorithm 1: Clustering($\mathcal{F}, \mathcal{F}_0$)

Input: an objective subset $\mathcal{F} := \{f_{i_1}, f_{i_2}, \ldots, f_{i_k}\} \subseteq \mathcal{F}_0$; the original objective set $\mathcal{F}_0 = \{f_1, f_2, \ldots, f_m\}$
Output: k objective set clusters $C_1, C_2, \ldots C_k$

```
for j = 1 to k do
    C_j ← {f_{i_j}};
for each objective f_l in F_0\F do
    min ← d(l, i_1);
    c ← 1;
    for j = 2 to k do
        if d(l, i_j) < min then
            min ← d(l, i_j);
            c ← j;
    C_c ← C_c ∪ f_l;
```

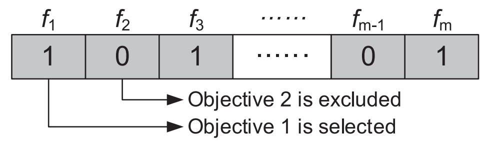


Fig. 3. Representation of chromosome as a binary string.

Now, the procedures to compute γ for an objective subset $\mathcal{F} := \{f_{i_1}, f_{i_2}, \ldots, f_{i_k}\} \subseteq \mathcal{F}_0$ is as follows. First, split $\mathcal{F}$ into k different clusters $C_1, C_2, \ldots, C_k$, where f_{i_j} falls into C_j and is set as the center of C_j, $j = 1 : k$. Then, every objective $f_l \in \mathcal{F}_0 \backslash \mathcal{F}$ is associated to the nearest cluster in terms of its distance to cluster centers. The clustering process is described in detail in Algorithm 1. Once clustering is completed, every objective in $\mathcal{F}_0$ belongs exclusively to one of the k clusters, and the criterion γ is computed from

$$\gamma = \max_{j=1:k} \max_{f_l \in C_j} d\big(i_j, l\big). \tag{4}$$

γ can be seen as a measurement reflecting the degree of correlation structure change between $\mathcal{F}$ and $\mathcal{F}_0$, and if an essential objective is not selected by $\mathcal{F}$, then there usually exist two relatively conflicting objectives in the same cluster, leading to a poor γ value.

The aim of γ-OR is to minimize both number of selected objectives k and γ value, where $\gamma \in [0, 1]$ since $d(i, j) \in [0, 1]$. Similar to δ-OR and η-OR, the two objectives of γ-OR are conflicting with each other to some extent. In addition, as indicated by (4), the maximum k value, i.e., $k = |\mathcal{F}_0|$, also corresponds to a minimum γ value, i.e., $\gamma = 0$.

C. Using Multiobjective Evolutionary Algorithms

To solve the formulated MOP (δ-OR, η-OR, or γ-OR) using MOEAs, a representation for the candidate solution must be chosen and encoded as a chromosome. In this paper, a chromosome is a binary string with m binary bits as shown in Fig. 3, where m is the size of the original objective set, i.e., $|\mathcal{F}_0|$. Each binary bit encodes a single objective, and a bit value of "1" or "0" means that the corresponding objective is selected or excluded, respectively.

Generally, based on this representation, any MOEA can serve the purpose of evolving the Pareto optimal objective subsets for each of the three formulated MOPs. Here, a popular MOEA, i.e., NSGA-II [9], is employed. The basic procedures of NSGA-II-based objective reduction is summarized as follows. Given the original objective set $\mathcal{F}_0$, the sample set $\mathcal{N}$ and the formulated MOP (δ-OR, η-OR, or γ-OR), the algorithm first randomly generates an initial population with N chromosomes. Then the algorithm goes into iterations until the stopping criterion specifying the maximum number of generations is satisfied. At each generation g, binary tournament selection [9], single-point crossover [56], and bit flip mutation [56] are performed on the current population P_g to produce the offspring population Q_g. Then, the best N chromosomes are selected as the next population P_{g+1} from the union population U_g using fast nondominated sorting and crowding distance. Moreover, it should be noted that $k = 0$ is meaningless for objective reduction, so once $k = 0$ for a chromosome, its two objective values are both immediately set to $+\infty$ in the proposed multiobjective approaches, making such chromosomes disappear easily during elitist selection.

For convenience, the resulting NSGA-II-based objective reduction algorithms are referred to as NSGA-II-δ, NSGA-II-η, and NSGA-II-γ, respectively, depending on which multiobjective formulation is used.

D. Why Multiobjective Approaches

The benefit of using multiobjective approaches for objective reduction mainly comes from two aspects. On one hand, objective reduction is inherently a multiobjective task. When performing objective reduction, we generally hope to keep the reduced objective set as small as possible. However, since a given sample set is only an approximation of the Pareto front, the smaller reduced objective set usually means that we would take a higher risk of losing problem information. Therefore, the user needs to make a compromise between the two conflicting aspects when finally determining a reduced objective set. This application scenario can be naturally modeled as an MOP. In each of our multiobjective formulations, the first objective (k) corresponds to the size of the objective subset, while the second objective (δ, η, or γ) measures the degree of risk in a particular way.

On the other hand, the multiobjective approaches are able to obtain a set of estimated Pareto optimal objective subsets, from which the user can gain deeper insights into the objective reduction problem. This enables the user to make a better decision when choosing the final reduced objective set. To further illustrate this, we reveal the relationship between NSGA-II-δ and the work in [1] by the following theorem.

Theorem 1: Given a sample set $\mathcal{N}$, let $\text{PF}_\delta = \{(k_j, \delta_j)^\text{T} | j = 1 : \kappa\}$ be the Pareto front of δ-OR problem. Then:

1) the solution to the δ-MOSS problem given by δ_0 is any objective subset corresponding to $(k_\mu, \delta_\mu)^\text{T}$, where

$$\mu = \underset{j}{\text{argmax}}\left\{\delta_j | \delta_j \leq \delta_0, j = 1 : \kappa\right\} \tag{5}$$

2) the solution to the k-EMOSS problem given by k_0 is any objective subset corresponding to $(k_\nu, \delta_\nu)^\text{T}$, where

$$\nu = \underset{j}{\text{argmax}}\left\{k_j | k_j \leq k_0, j = 1 : \kappa\right\}. \tag{6}$$

Theorem 1 implies that NSGA-II-δ can obtain solutions to all possible δ-MOSS and k-EMOSS problems related to $\mathcal{N}$ in a single simulation run, i.e., it can provide the necessary information for decision support once for all. Whereas in [1], a simulation run of the algorithm only concerns one specific δ-MOSS (given by δ_0) or k-EMOSS (given by k_0) problem, and while using heuristic algorithms, the two kinds of problems have to be addressed by different greedy algorithms.

Note that, Theorem 1 along with the definitions of δ-MOSS and k-EMOSS can be generalized to other types of errors, e.g., η and γ. But for simplicity, we still use the terms "δ-MOSS" and "k-EMOSS" no matter what error is used.

IV. Analysis of Dominance Structure- and Correlation-Based Approaches

In this section, we first put forward several theorems on objective reduction. Based on them, we analyze the strengths and limitations of dominance structure- and correlation-based approaches in terms of identifying the essential objective set, respectively. Note that the analysis in this section also applies to the proposed multiobjective approaches, since they are based essentially on the dominance structure or the correlation.

A. Theoretical Foundations

In objective reduction, it is important to understand the original Pareto front and the Pareto front corresponding to an objective subset. The following theorem gives the relationship between the two.

Lemma 1: Let PF' be the Pareto front corresponding to an objective subset $\mathcal{F}' \subseteq \mathcal{F}_0$, then $\text{PF}' \subseteq \{\mathbf{u}^{(\mathcal{F}')} | \mathbf{u} \in \text{PF}_0\}$.

Based on Definition 5 and Lemma 1, the following theorem provides a necessary and sufficient condition to decide whether an objective subset is redundant or not, which can be viewed as the principle of dominance structure-based objective reduction approaches.

Theorem 2: An objective subset $\mathcal{F} \subset \mathcal{F}_0$ is redundant, iff $\nexists \mathbf{u}, \mathbf{v} \in \text{PF}_0 : \mathbf{u}^{(\mathcal{F}')} \prec \mathbf{v}^{(\mathcal{F}')}$, where $\mathcal{F}' := \mathcal{F}_0 \backslash \mathcal{F}$.

Further, Theorem 3 shows a way to judge whether an objective subset is redundant by exploiting its relationship with another one. It needs to be stressed that the condition given in Theorem 3 is sufficient yet not necessary for "redundancy," which is different from that in Theorem 2. In other words, if $\mathcal{F}$ is redundant, there may exist no such objective subset $\mathcal{F}'$ satisfying $\forall \mathbf{u}, \mathbf{v} \in \text{PF}_0 : \mathbf{u}^{(\mathcal{F}')} \preceq \mathbf{v}^{(\mathcal{F}')} \Rightarrow \mathbf{u}^{(\mathcal{F})} \preceq \mathbf{v}^{(\mathcal{F})}$.

Theorem 3: Given two nonempty objective subsets $\mathcal{F}, \mathcal{F}' \subset \mathcal{F}_0$ and $\mathcal{F} \cap \mathcal{F}' = \emptyset$, if $\forall \mathbf{u}, \mathbf{v} \in \text{PF}_0 : \mathbf{u}^{(\mathcal{F}')} \preceq \mathbf{v}^{(\mathcal{F}')} \Rightarrow \mathbf{u}^{(\mathcal{F})} \preceq \mathbf{v}^{(\mathcal{F})}$, then $\mathcal{F}$ is redundant.

Based on Theorem 3, we have the following corollary, which establishes the relationship between the concepts nonconflicting and redundant.

Corollary 1: Given two objectives $f_i, f_j \in \mathcal{F}_0$, if f_i and f_j are nonconflicting, then $\{f_i\}$ and $\{f_j\}$ are both redundant.

Theorem 3 together with Corollary 1 will be used later in Section IV-C to illustrate the principle and the limitation of correlation-based approaches.

Lastly, we provide a theorem about the multiobjective formulation η-OR, which will be used later to demonstrate the strength of dominance structure-based approaches.

Theorem 4: Given a sample set $\mathcal{N} := \text{PF}_0$ and $|\text{PF}_0| < +\infty$, PF_η denotes the Pareto front of η-OR problem. Then $\exists k^* \in \mathbb{N} : (k^*, 0)^{\text{T}} \in \text{PF}_\eta$ and any objective subset corresponding to $(k^*, 0)^{\text{T}}$ is an essential objective set for the given MOP.

B. Strengths and Limitations of Dominance Structure-Based Approaches

It can be inferred from Theorem 2 that the revelation of an essential objective set can be transformed to the examination of the dominance relation in the sample set with respect to every objective subset. The dominance structure-based approaches generally exploit this characteristic to guide the algorithm design, with major difference in how to measure the degree of violation of the condition in Theorem 2. Thus their algorithm mechanisms can fit the nature of objective reduction very well. In theory, as long as the sample set $\mathcal{N}$ provides a perfect Pareto front-representation, a basic dominance structure-based algorithm can identify the essential objective set accurately on all the problems given sufficient computation time. Take η-OR for instance, Theorem 4 indicates that, under the most ideal situation, i.e., $\mathcal{N} := \text{PF}_0$, it is ensured that PF_η includes the vector $(k^*, 0)^{\text{T}}$ that corresponds to an essential objective set.

However, it is often expecting too much that $\mathcal{N}$ is a prefect Pareto front-representation. The misdirection in $\mathcal{N}$ is very likely to make dominance structure-based approaches fail to capture the true dimensionality of PF_0. This is mainly because it is sometimes hard to reasonably determine whether some solutions can be interpreted as misdirection by only exploiting the mutual dominance relationship. To explain this further, let us have a closer look at the criteria δ and η. For the criterion δ, suppose the following $\mathcal{N} = \{\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{2n-1}, \mathbf{u}_{2n}\}$, where:

$$\mathbf{u}_i = \begin{cases} \left(10^{-6}, 2n, 0\right)^{\text{T}}, & i = 0 \\ (i, 2n-i, i)^{\text{T}}, & i = 1 : n \\ (i, 2n-i, 3n-i)^{\text{T}}, & i = n+1 : 2n-1 \\ (0, 2n, +\infty)^{\text{T}} & i = 2n. \end{cases} \tag{7}$$

In this case, $\mathbf{u}_{2n}$ could resemble misdirection since it is an outlier. If $\mathbf{u}_{2n}$ is ignored, $\{f_1, f_2\}$ would be identified as a unique essential objective set, because the dominance structure is not changed only with respect to the $f_1 - f_2$ objective subspace. However, the computing of δ cannot automatically ignore $\mathbf{u}_{2n}$ and indeed involves all solutions in $\mathcal{N}$, thus the δ error of $\{f_1, f_2\}$ would tend to be positive infinity by considering the dominance relation between $\mathbf{u}_0$ and $\mathbf{u}_{2n}$, making $\{f_1, f_2\}$ disqualify in becoming an essential objective set. As for criterion η, we take three-objective WFG3 as an example again. Fig. 4 shows the projections of a sample set $\mathcal{N}$ (including $\mathcal{N}_{\mathcal{NS}}$ and $\mathcal{N}_{\mathcal{DS}}$) on $f_2 - f_3$ objective space. In this case, $\eta = 0.52$ for $\{f_2, f_3\}$ is a significant proportion. This implies that it is not safe to interpret the solutions in $\mathcal{N}_{\mathcal{DS}}$ as misdirection according to the large η error and consequently $\{f_2, f_3\}$ would not be identified as an essential objective set, which does not conform well with the intuition for Fig. 4. The problem lies in that, although $\mathcal{N}_{\mathcal{DS}}$ accounts for a significant proportion of $\mathcal{N}$, most of its objective vectors are very close to those in $\mathcal{N}_{\mathcal{NS}}$ and can indeed be interpreted as misdirection. In summary, a major limitation of dominance structure-based approaches is that they may have difficulty in handling varying degree of misdirection in the sample set effectively.

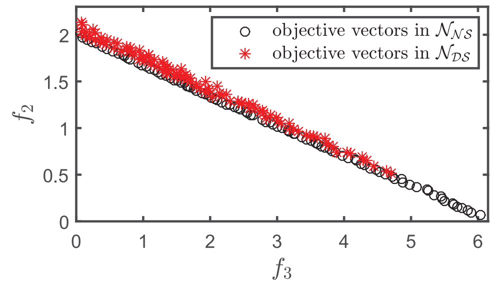


Fig. 4. Projection of $\mathcal{N}$ (including $\mathcal{N}_{\mathcal{NS}}$ and $\mathcal{N}_{\mathcal{DS}}$) on $f_2 - f_3$ objective space for three-objective WFG3.

Another limitation of the dominance structure-based approaches is that, they may over-reduce the objectives if $\mathcal{N}$ fails to have a good coverage of PF_0. This phenomenon is more likely to occur when the dimensionality of PF_0 is higher, since exponentially more solutions are needed to represent PF_0 well. For example, suppose the dimensionality of PF_0 is 15, and the limited size and diversity of $\mathcal{N}$ restricts its distribution to only a part of PF_0. Due to the inadequate representation of PF_0, the solutions in $\mathcal{N}$ may also be mutually nondominated in a lower dimensional (e.g., 10-dimensional) objective subspace which corresponds to an objective subset $\mathcal{F}$. So the associated error (e.g., δ or η) of $\mathcal{F}$ already reaches the minimal value, i.e., 0; and because $|\mathcal{F}| < 15$, the dominance structure-based approaches would prefer to select $\mathcal{F}$ as the essential objective set, leading to an unsuccessful identification.

Moreover, the computational complexity of dominance structure-based approaches is dominated by analyzing the dominance relation between every two solutions, whereas that of correlation-based approaches is dominated by exploiting the correlation relation between objective pairs. Considering the number of solutions is normally much larger than the number of objectives, i.e., $|\mathcal{N}| >> m$, dominance structure-based approaches are generally much more computationally expensive than correlation-based approaches.

C. Strengths and Limitations of Correlation-Based Approaches

Correlation-based approaches can generally cope with misdirection in the sample set more naturally and effectively when compared to dominance structure-based approaches. This is mainly because correlation only emphasizes on the overall increasing or decreasing trend between objectives, without focusing on the relationship between particular solutions like the dominance structure-based approaches. Hence, correlation-based approaches can usually negate the effect of misdirection easily. For example, the Kendall's rank correlation matrix based on $\mathcal{N}$ shown in Fig. 4 is computed as follows:

$$r_\tau = \begin{pmatrix} 1 & 0.901 & -0.943 \\ 0.901 & 1 & -0.952 \\ -0.943 & -0.952 & 1 \end{pmatrix}. \tag{8}$$

It can be observed from this matrix that while f_1 and f_2 are strongly positively correlated (nonconflicting) with each other, each of them is strongly negatively correlated (conflicting) with f_3. Thus, it can be concluded that $\{f_1\}$ and $\{f_2\}$ are both redundant, and $\{f_2, f_3\}$ and $\{f_1, f_3\}$ are both essential objective sets, which is consistent with the true Pareto front of three-objective WFG3. Recall that it is very difficult to identify an essential objective set in this case by using η criterion, thereby highlighting the strength of correlation-based approaches in handling the misdirection. Considering this strength along with their higher computational efficiency as mentioned before, they appear to be a better choice than dominance structure-based approaches when embedded into the iteration of MOEAs to perform online objective reduction. In addition, different from dominance structure-based approaches, correlation-based approaches do not only provide the essential objective set, but can also identify what objectives are conflicting and what are nonconflicting, which would be helpful information for decision makers.

The major limitation of correlation-based approaches is that reducing objectives by correlation analysis is not completely consistent with the original intention of objective reduction. Recall that correlation-based approaches use correlation coefficient to measure the degree of conflict between each pair of objectives, aiming to keep the highly conflicting objectives and remove the objectives that are nonconflicting with others. Corollary 1 justifies the rationality of correlation-based approaches in this respect. However, as seen from Theorem 3, the redundancy of an objective can be caused due to its relationship with an objective subset more than just a single objective. However, conflict or correlation is restricted to capturing the relationship between only two objectives. To illustrate this, we add a third objective $f_3 = f_1 + f_2$ to the well known two-objective DTLZ2 [57] problem, and obtain a nearly perfect sample set for this modified problem by NSGA-II. Fig. 5 shows the projections of the sample set on the objective subspaces $f_1 - f_2$, $f_1 - f_3$, and $f_2 - f_3$, respectively. Based on this sample set, the underlying Kendall's rank correlation matrix is computed as follows:

$$r_\tau = \begin{pmatrix} 1 & -1 & -0.016 \\ -1 & 1 & 0.016 \\ -0.016 & 0.016 & 1 \end{pmatrix}. \tag{9}$$

From (9), it can be concluded that f_1 and f_2 are strongly conflicting, but each of them appears neither conflicting nor nonconflicting with f_3 since $r_\tau(1, 3)$ and $r_\tau(2, 3)$ are both very close to zero. Thus, f_3 cannot be safely removed by correlation analysis. However, according to $f_3 = f_1 + f_2$ and Theorem 3, it can be determined that $\{f_3\}$ is redundant. Indeed, by Fig. 5 and Theorem 2, it can be further inferred that $\{f_3\}$ is the unique redundant objective set and $\{f_1, f_2\}$ is the unique essential objective set for this problem.

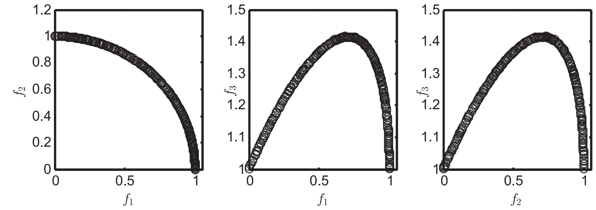


Fig. 5. Projections of a sample set for the modified DTLZ2 problem on the objective subspaces $f_1 - f_2$, $f_1 - f_3$, and $f_2 - f_3$.

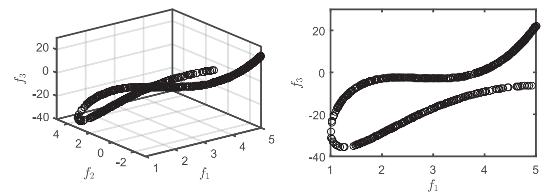


Fig. 6. High-quality sample set obtained for the spline problem: $f_1 = x^2 + 1$, $f_2 = -x^2 + x + 3$, and $f_3 = -(f_1 + f_2^3)$, where $x \in [-2, 2]$, and its projection on the $f_1 - f_3$ objective subspace.

Note that Saxena *et al.* [2] indicated that the correlation-based approaches may lead to inaccurate results in some cases where the misdirection interpreted is in fact the signal. To demonstrate this point, they constructed an artificial problem shown in Fig. 6, in which f_1 and f_3 are globally correlated but locally conflicting and the latter is not significant enough. However, any algorithm with certain ability in handling the misdirection would suffer from this potential side-effect inevitably, so it may not be regarded as an inherent limitation of correlation-based approaches. To ensure a lower error tolerance may relieve this issue. For example, $r_\tau(1, 3) = 0.652$ for the objectives f_1 and f_3 in Fig. 6, if it is assumed that two objectives are seen as nonconflicting only when the correlation coefficient is larger than 0.7, then accurate results can still be obtained, but at the cost of sacrificing the strength in handling misdirection to some extent.

In summary, the correlation-based approaches cannot find the essential objective set exactly on all problems, but their strengths mentioned above make them popular for objective reduction. Indeed, as shown in Table I, the majority of existing studies focus on correlation-based approaches.

V. Benchmark Experiments

In this section, we first describe several experimental design conditions including the benchmark problems used, the generation of sample sets, and the algorithms used for comparison purpose. Then, the extensive experiments and comparisons are conducted on various benchmark problems from three different aspects.

A. Benchmark Problems

To study the performance of objective reduction algorithms, we employ five different test problems with varying number of objectives and dimensionality in the experiments. These problems include three well-known benchmark problems, i.e., DTLZ5(I, m) [2], [40], WFG3(m) [46], [47], and DTLZ2(m) [57], and two new problems constructed based on

DTLZ2(m), namely POW-DTLZ2(p, m) and SUM-DTLZ2(m), all of which are described below.

DTLZ5(I, m) is a redundant problem derived from the scalable DTLZ5(m) [57], where $I \leq m$, I denotes the dimensionality of the Pareto front and m denotes the original number of objectives for the problem. The first $m-I+1$ objectives over the Pareto front are mutually nonconflicting, and an essential objective set is given by $\mathcal{F_T} = \{f_l, f_{m-I+2}, \ldots, f_m\}$, where $l \in \{1, 2, \ldots, m-I+1\}$.

WFG3(m) is an m-objective optimization problem whose dimensionality of Pareto front is 2. Except f_m, the other objectives over the Pareto front are nonconflicting with each other. Thus, an essential objective set is given by $\mathcal{F_T} = \{f_l, f_m\}$, where $l \in \{1, 2, \ldots, m-1\}$.

DTLZ2(m) is a nonredundant problem with m objectives, and its essential objective set consists of all m objectives, i.e., $\mathcal{F_T} = \{f_1, f_2, \ldots, f_m\}$.

POW-DTLZ2(p, m) is used to test the ability of the objective reduction algorithms to handle strong nonlinearity. The problem is constructed by adding m new objectives into DTLZ2(m), so it has $2m$ objectives in total. The first m objectives are the original objectives of DTLZ2(m), while the remaining m objectives are formulated as $f_{m+j} = f_j^p$, $j = 1{:}m$. An essential objective set is given by $\mathcal{F_T} = \{f_{l_1}, f_{l_2}, \ldots, f_{l_m}\}$, where $l_j \in \{j, m+j\}$ for $j = 1 : m$.

SUM-DTLZ2(m) is designed to disclose the possible limitation of correlation-based approaches, meanwhile show the strength of dominance structure-based approaches in certain aspect. Like POW-DTLZ2(p, m), the first m objectives of this problem come from DTLZ2(m). The sums of every two different objectives in $\{f_1, f_2, \ldots, f_m\}$ form the remaining $(m(m-1))/2$ objectives. Thus SUM-DTLZ2(m) has $(m(m+1))/2$ objectives in total. The essential objective set is given by $\mathcal{F_T} = \{f_1, f_2, \ldots, f_m\}$.

Except for WFG3(m), the number of decision variables for the other problems depends on a parameter p, i.e., $n = m + q - 1$. We set q to 10 for these problems as recommended in [57]. As for WFG3(m), the number of decision variables is set to 28 and the position related parameter is set to $m - 1$.

B. Generation of Sample Sets

The quality of the sample set would certainly influence the behavior of these algorithms. However, few studies on objective reduction emphasized this issue except the work by Saxena *et al.* [2], where four types of sample sets with different qualities were used to study the performance of the presented algorithms. In our experiments, two types of sample sets, i.e., $\mathcal{N}_1$ and $\mathcal{N}_2$, are used for each test instance, and each of $\mathcal{N}_1$ and $\mathcal{N}_2$ has 30 different sets, which are obtained by an SPEA2 variant referred as strength Pareto evolutionary algorithm 2-shift-based density estimation (SPEA2-SDE) [19] with the same number of generations for 30 independent repetitions. $\mathcal{N}_2$ requires more computational effort (within 2000 generations) than $\mathcal{N}_1$ (within 200 generations) and achieves a high-quality Pareto front-representation, whereas $\mathcal{N}_1$ suffers from much more misdirection than $\mathcal{N}_2$ and thus poses greater challenge to the algorithms in coping with the misdirection. Note that, different from [2], we do not consider the randomly generated sample set and the sample set distributed evenly on the true Pareto front, since the conclusions are obvious in the two scenarios.

Following the practice in [2], we assess the quality of $\mathcal{N}_1$ and $\mathcal{N}_2$ using the problem parameter g (convergence, not for WFG3) and the normalized maximum spread indicator I_s (diversity) [18]. The closer g is to 0 and the closer I_s is to 1 implies better convergence and diversity to some extent, respectively. For $\mathcal{N}_1$ and $\mathcal{N}_2$, the average g and I_s are as follows.

1) $\mathcal{N}_1 : g = 0.213$ and $I_s = 1.636$.
2) $\mathcal{N}_2 : g = 0.019$ and $I_s = 1.044$.

It can be seen that the average g of $\mathcal{N}_2$ is nearly one order of magnitude less than that of $\mathcal{N}_1$, so that they have clear difference in quality.

TABLE II
LIST OF OTHER ALGORITHMS IN COMPARISON

Name	Category	Description
ExactAlg [1]	C1	an exact algorithm for both δ-MOSS and k-EMOSS
Greedy-δ [1]	C1	a greedy algorithm for δ-MOSS
Greedy-k [1]	C1	a greedy algorithm for k-EMOSS
L-PCA [2]	C2	an algorithm based on principal component analysis
NL-MVU-PCA [2]	C2	an algorithm based on maximum variance unfolding

C1: Dominance structure-based approaches
C2: Correlation-based approaches

C. Algorithms in Comparison

Several other objective reduction algorithms are involved for comparison with the proposed multiobjective approaches. Table II provides their short names used in this paper, categories and brief descriptions. All these algorithms are implemented in Java and run on an Intel 3.20-GHz Xeon processor with 16.0 GB of RAM.[3]

ExactAlg is not suited for large-scale instances, so some of its results are unavailable in the experiments. There are two greedy algorithms for k-EMOSS problem in [1], here Greedy-k refers to the one based on omission of objectives. The proposed multiobjective approaches adopt the parameters given in Table III, where m is the number of objectives in the problem to be reduced. Note that, exhaustive search can be considered for the small instances, e.g., $m \leq 10$, but for uniformity, we just solve all the instances using the evolutionary multiobjective search.

Before performing ExactAlg, Greedy-δ, Greedy-k, and NSGA-II-δ, the objective values of the sample set are first normalized to a range of $[0, 1]$ in each dimension, because one assumption of δ error is that all objective values have the same scale [1].

[3]The source code of the proposed objective reduction algorithms is available online: https://github.com/yyxhdy/MOOR/archive/master.zip.

TABLE III
PARAMETERS IN THE PROPOSED MULTIOBJECTIVE APPROACHES (NSGA-II-δ, NSGA-II-η, AND NSGA-II-γ)

Parameter	Value
Population size	$5m$
Maximum generations	$\lfloor 2.5m \rfloor$
Crossover probability	0.9
Mutation probability	$1/m$

TABLE IV
RESULT OF NSGA-II-δ FOR A SAMPLE SET OF DTLZ5(3, 20) CORRESPONDING $\mathcal{N}_1$

k	δ	Objective Set
1	1	$\{f_i\}, i = 6, 13, 15, 16, 19, 20$
2	0.997	$\{f_{19}, f_{20}\}$
3	0.255	$\{f_{13}, f_{19}, f_{20}\}$
4	0.133	$\{f_{11}, f_{16}, f_{19}, f_{20}\}$
5	0.053	$\{f_{11}, f_{15}, f_{16}, f_{19}, f_{20}\}$
6	0	$\{f_9, f_{13}, f_{15}, f_{16}, f_{19}, f_{20}\}$
6	0	$\{f_{11}, f_{15}, f_{16}, f_{18}, f_{19}, f_{20}\}$
6	0	$\{f_9, f_{13}, f_{16}, f_{17}, f_{19}, f_{20}\}$

D. Investigation on the Behavior of Multiobjective Approaches

The multiobjective approaches return a number of nondominated objective subsets along with the corresponding errors for an objective reduction problem, which can provide decision support to the user. As an illustration, Table IV shows the result of NSGA-II-δ for a sample set of DTLZ5(3, 20) corresponding to $\mathcal{N}_1$. It is clear from Table IV that an objective set with smaller size would lead to a larger error. So, the user need to make a tradeoff when choosing a preferred one. For example, when selecting $\{f_{13}, f_{19}, f_{20}\}$ as the reduced objective set, the user should consider whether the error $\delta = 0.255$ is tolerable or not for the current application.

Fig. 7 provides a clearer picture of how the error decreases with the number of objectives increasing in the results of NSGA-II-δ, NSGA-II-η, and NSGA-II-γ. Four DTLZ5(I, 20) instances are used for illustration here, and the results are shown for both $\mathcal{N}_1$ and $\mathcal{N}_2$. In each chart of this figure, the horizontal axis shows the number of objectives selected (k), and the vertical axis shows the error (δ, η, or γ) for the corresponding k-EMOSS problem averaged over 30 sample sets in $\mathcal{N}_1$ or $\mathcal{N}_2$, which can be obtained directly from the results of multiobjective approaches as indicated in Theorem 1. Based on Fig. 7, the following observations are helpful for better understanding the behavior of the multiobjective approaches.

1) The variation trends of the errors are different in NSGA-II-δ, NSGA-II-η, and NSGA-II-γ. It is very reasonable since the three types of errors have quite different implications.
2) All three multiobjective approaches achieve a relatively small error at $k = |\mathcal{F_T}|$ in most cases, which validates the rationality of the three kinds of errors to some extent.
3) For $\mathcal{N}_2$, the errors in the multiobjective approaches usually have a sharp drop at $k = |\mathcal{F_T}| - 1$ to a very small value at $k = |\mathcal{F_T}|$, so it would be easy to pick out the essential objective set by exploiting this characteristic. However, for $\mathcal{N}_1$, the phenomena is not so obvious, implying that $\mathcal{N}_1$ would lead to greater difficulty in identifying the essential objective set.
4) The error η in NSGA-II-η almost decreases exponentially to 0 as k increases except on DTLZ5(3, 20) and it usually has already reached a sufficiently small value before k is increased to $|\mathcal{F_T}|$, which is quite different from the other two types of errors. The possible reason is that, with the dimension increasing, the number of nondominated solutions in the sample set grows exponentially, and the cardinality of $\mathcal{N_{DS}}$ would become more inadequate to reflect the real degree of the dominance structure change in the Pareto front. In view of this, we suggest that the tolerable η error corresponding to a larger k should become smaller to relieve the curse of dimensionality.

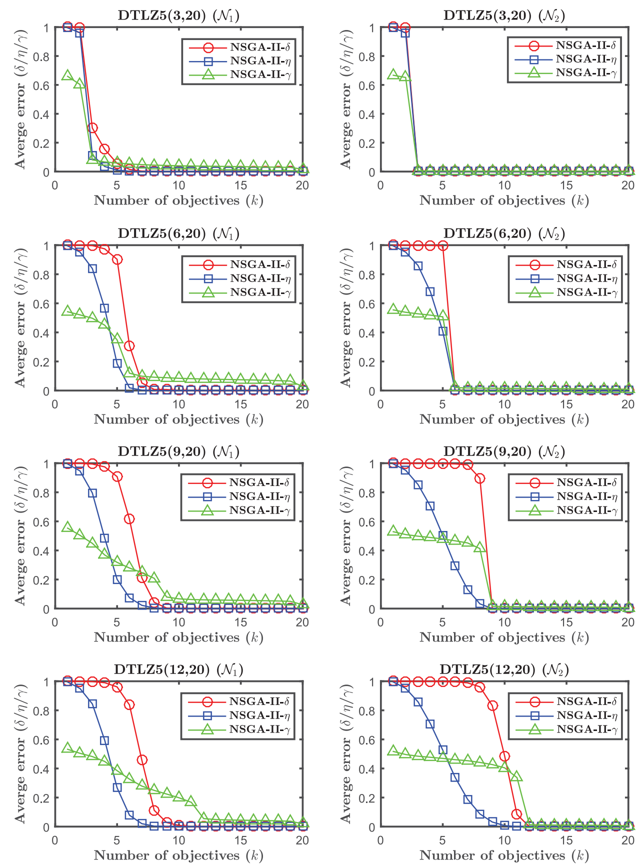


Fig. 7. Results of NSGA-II-δ, NSGA-II-η, and NSGA-II-γ on DTLZ5(3, 20), DTLZ5(6, 20), DTLZ5(9, 20), and DTLZ5(12, 20), corresponding to $\mathcal{N}_1$ and $\mathcal{N}_2$. The error for each point is averaged over 30 sample sets in $\mathcal{N}_1$ or $\mathcal{N}_2$.

TABLE V
AVERAGE COMPUTATION TIME (IN SECONDS) FOR EACH RUN ON FOUR CONSIDERED DTLZ5(I, 20) INSTANCES

NSGA-II-δ	NSGA-II-η	NSGA-II-γ
4.45	4.59	0.12

In Table V, we show the average computation time required by each run of NSGA-II-δ, NSGA-II-η, and NSGA-II-γ,

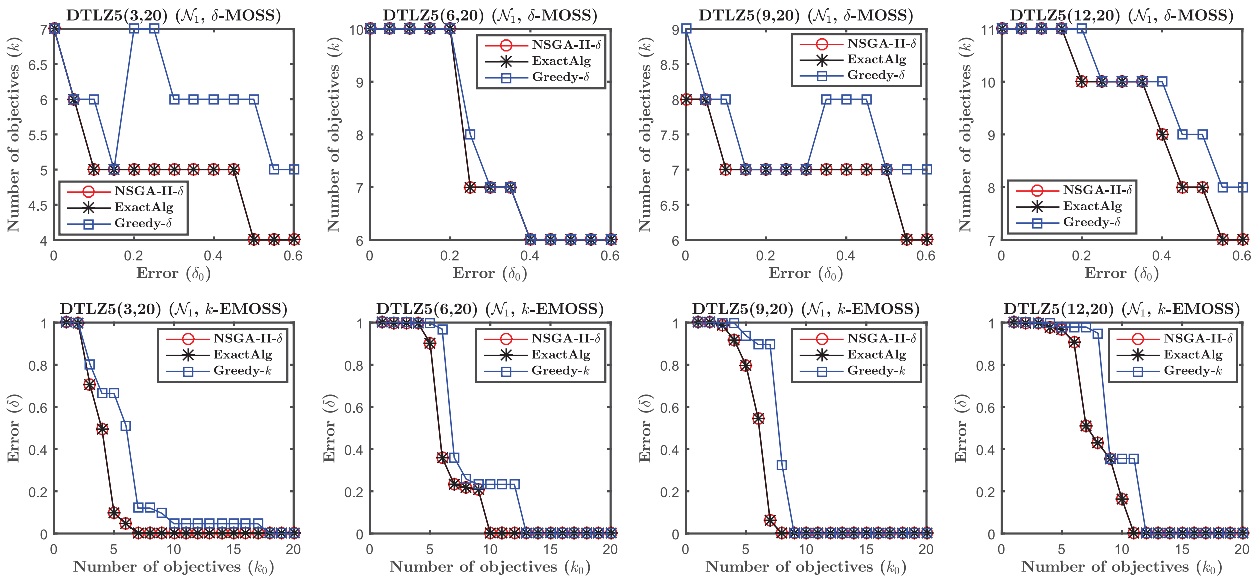


Fig. 8. Comparison of NSGA-II-δ, ExactAlg, and Greedy-δ (Greedy-k) for the δ-MOSS or k-EMOSS problem.

respectively. As seen from Table V, the dominance structure-based algorithms NSGA-II-δ and NSGA-II-η need much more computational effort than the correlation-based algorithm NSGA-II-γ, which is accordant with the analysis in Section IV-B.

E. Effectiveness of Evolutionary Multiobjective Search

NSGA-II-δ, ExactAlg, and Greedy-δ (Greedy-k) all utilize δ error to measure the change of the dominance structure, and the difference between them mainly lies in the search mechanism. In this section, we compare NSGA-II-δ with ExactAlg and Greedy-δ (Greedy-k) in solving δ-MOSS or k-EMOSS problems, in order to verify the effectiveness of the evolutionary multiobjective search.

Here, as an illustration, we still use the four test instances adopted in Section V-D. For each test instance, we choose a single sample set from 30 ones in $\mathcal{N}_1$, which is associated with the result closest to the average convergence metric. And for each sample set, we vary δ_0 from 0 to 0.6 with an interval of 0.05 and k_0 from 1 to 20 with an interval of 1, leading to 13 δ-MOSS problems and all related k-EMOSS problems.

Fig. 8 shows the comparison results of the considered algorithms on the δ-MOSS problems and k-EMOSS problems, respectively. As seen from Fig. 8, all the results of NSGA-II-δ match exactly with those of ExactAlg, which means that NSGA-II-δ has achieved the optimal solutions to all considered δ-MOSS and k-EMOSS problems. The performance of greedy algorithms including Greedy-δ and Greedy-k cannot compare with that of NSGA-II-δ. Indeed, they can only obtain the optimal solutions sometimes and always perform worse than NSGA-II-δ in the other situations. For example, for the sample set of DTLZ5(3, 20), Greedy-k yields worse δ values than NSGA-II-δ on 15 out of total 20 k-EMOSS problems, and achieves the same results only on the remaining 5 ones. Moreover, when solving the δ-MOSS problem, it is natural that the obtained number of objectives should be nonincreasing with the increase of δ_0, but it is not always the case for Greedy-δ as shown in Fig. 8. This is because the greedy search in Greedy-δ may only reach a local optimal solution that depends on the given δ_0, which reflects the necessity of the global search in NSGA-II-δ.

TABLE VI
COMPARISON OF THE COMPUTATIONAL TIME (IN SECONDS) OF NSGA-II-δ, EXACTALG, GREEDY-k, AND GREEDY-δ

Test Instance	NSGA-II-δ[a]	ExactAlg[a]	Greedy-k[a]	Greedy-δ[b]
DTLZ5(3,20)	3.19	5280.79	7.16	0.17
DTLZ5(6,20)	3.80	1533.35	7.55	0.21
DTLZ5(9,20)	4.71	11588.49	7.59	0.20
DTLZ5(12,20)	5.15	14491.03	7.12	0.26

[a] The computational time for all related k-EMOSS problems.
[b] The average computational time for a specific δ-MOSS problem.

Table VI reports the computational time of NSGA-II-δ, ExactAlg, and Greedy-k for solving all related k-EMOSS problems, and the average computational time of Greedy-δ for solving a specific δ-MOSS problem. From Table VI, it can be seen that NSGA-II-δ incurs much less time than ExactAlg on all four sample sets of considered test instances. Indeed, the time required by ExactAlg may not be acceptable in practice, which even reaches about 4 h on the sample set of DTLZ5(12, 20). Compared with Greedy-k, NSGA-II-δ also uses relatively shorter time, usually a few seconds less, in all test cases. Although Greedy-δ requires less time than NSGA-II-δ, Greedy-δ only solves a specific δ-MOSS problem here whereas NSGA-II-δ indeed provides solutions to all δ-MOSS problems (the number is infinite) in a single run.

From the above, it can be concluded that, the evolutionary multiobjective search can achieve comparable results with the global exact search but shows an overwhelming advantage in the computational efficiency, and it is also superior to the greedy search in terms of both effectiveness and efficiency.

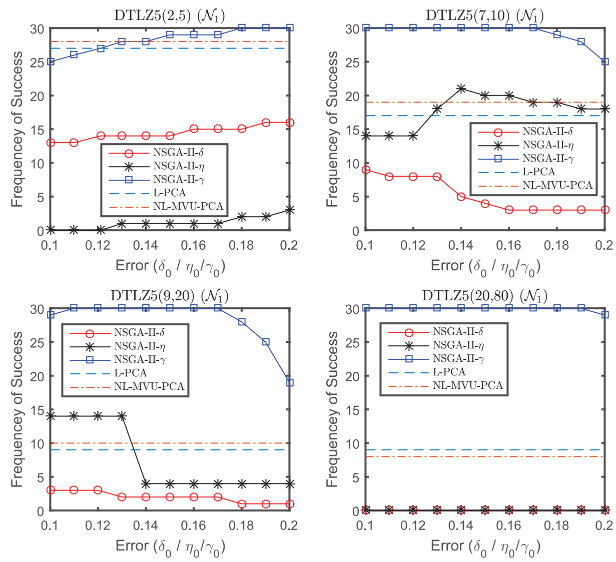


Fig. 9. Illustration of the influence of $\delta_0/\eta_0/\gamma_0$ on the performance of the proposed multiobjective approaches in identifying the essential objective set. The performance of L-PCA and NL-MVU-PCA is provided as a reference.

F. Comparison in Identifying the Essential Objective Set

In this section, we compare the proposed multiobjective approaches with ExactAlg, Greedy-δ, L-PCA, and NL-MVU-PCA in identifying the essential objective set, in order to demonstrate the effectiveness of the proposed algorithms and confirm the analysis conducted in Section IV. The multiobjective approaches can return a number of nondominated objective subsets which contributes to decision support, but a single one should be chosen as the essential objective set here for the comparison. In practice, we can visually see the obtained Pareto front of objective subsets for a specific problem and find the most likely one. But for simplicity, we just select the solution to a given δ-MOSS problem, where the threshold of error (δ_0, η_0, or γ_0) is specified beforehand and kept the same for all the test instances. And as suggested in Section V-D, η_0 is decreased by an empirical factor of $\sqrt{3}$ from $k = 2$ to counteract the impact of the exponentially growing number of nondominated solutions in the higher-dimensional objective space.

How to determine the threshold value is an important issue. If it is set too small, the algorithm will not have the necessary ability of handling the misdirection. An extreme case can be seen in [2] where δ of ExactAlg and Greedy-δ is fixed at 0, leading to very poor results by them on the sample sets with misdirection as reported by Saxena *et al*. [2]. However, if it is set too large, we will usually omit essential objectives by mistake in many scenarios. For example, if we set γ_0 to 0.3, then it would mean that we can regard the two objectives as nonconflicting once the correlation coefficient between them reaches 0.4, which is obviously unreliable. Based on our experience with many problems and the sample sets with different quality, we suggest to use the threshold values within the range [0.1, 0.2], which can usually make the algorithm not only handle the moderate degree of misdirection but also do a reliable identification in various scenarios. In Fig. 9, we investigate the influence of $\delta_0/\eta_0/\gamma_0 \in [0.1, 0.2]$ on the performance of multiobjective approaches, in terms of the frequency of success in identifying the essential objective set out of 30 runs. From Fig. 9, the performance of each algorithm is overall stable, and more importantly, the relative superior or inferior relationship between them remains similar no matter what the threshold value is set in the considered range. Note that, we only show four cases in Fig. 9 due to space limitation, but the situation is similar in the other case, and also we find that their performance will become much less sensitive to the threshold values on $\mathcal{N}_2$. In the following experiments, we just set the same threshold value, i.e., 0.15, for the three multiobjective algorithms, ExactAlg and Greedy-δ, to ensure a fair comparison. As for L-PCA and NL-MVU-PCA, a variance threshold θ is set to 0.997 and the correlation threshold T_{cor} is determined by an empirical formula as recommended in [2].

Table VII shows the results of all the algorithms in terms of the frequency of success out of 30 runs, where each run is associated with a unique sample set in $\mathcal{N}_1$ ($\mathcal{N}_2$). Based on Table VII, we first consider the dominance structure-based approaches, i.e., NSGA-II-δ, NSGA-II-η, ExactAlg, and Greedy-δ, in comparison. As analyzed in Section IV-B, there mainly exist two potential difficulties that make these algorithms fail to identify the essential objective set: 1) the misdirection in the sample set and 2) the inadequate coverage of the high dimensional Pareto front by the sample set. The dominance structure-based approaches tend to select an excessive number of objectives in the presence of the first difficulty, while they would over-reduce the objectives due to the second difficulty. Keeping the two difficulties in mind, we can provide a sound explanation for the results of these algorithms, which will be illustrated from the following three situations.

1) When the dimensionality $|\mathcal{F}_\mathcal{T}|$ is relatively low ($|\mathcal{F}_\mathcal{T}| \leq 6$), they mainly suffer from the first difficulty for $\mathcal{N}_1$. Table VIII shows some of these instances separately and also reports the average number of objectives obtained by NSGA-II-δ, NSGA-II-η, and Greedy-δ corresponding to $\mathcal{N}_1$. The results of ExactAlg (if available) are omitted in Table VIII since they are always the same with those of NSGA-II-δ. From Tables VII and VIII, the considered algorithms normally fail to obtain a satisfying performance on such instances for $\mathcal{N}_1$ and would choose more than $|\mathcal{F}_\mathcal{T}|$ objectives on the average, which can be attributed to their limitation in handling misdirection. However, when it comes to $\mathcal{N}_2$, they perform obviously better and achieve the optimal or near optimal results on all these instances. It can be explained that, there exists much less misdirection in $\mathcal{N}_2$ than in $\mathcal{N}_1$, leading to a significant alleviation of the first difficulty with these algorithms.
2) When the dimensionality $|\mathcal{F}_\mathcal{T}|$ is moderately high ($7 \leq |\mathcal{F}_\mathcal{T}| \leq 10$), they would confront both two difficulties for $\mathcal{N}_1$. In Table IX, we further show the average number of objectives identified by NSGA-II-δ, NSGA-II-η, and Greedy-δ on such instances for $\mathcal{N}_1$. As seen from Table VII, they generally perform poorly on these instances corresponding to $\mathcal{N}_1$, which can be due to the synthetic effect of the two difficulties. And as observed

TABLE VII
FREQUENCY OF SUCCESS IN IDENTIFYING THE ESSENTIAL OBJECTIVE SET $\mathcal{F}_{\mathcal{T}}$ OUT OF 30 RUNS, CORRESPONDING TO $\mathcal{N}_1$ AND $\mathcal{N}_2$. THE BEST FREQUENCIES ARE SHOWN IN BOLD

Test Instance	$\mathcal{N}_1$							$\mathcal{N}_2$						
	NSGA -II-δ	NSGA -II-η	NSGA -II-γ	Exact Alg	Greedy -δ	L -PCA	NL-MVU -PCA	NSGA -II-δ	NSGA -II-η	NSGA -II-γ	Exact Alg	Greedy -δ	L -PCA	NL-MVU -PCA
D5(2,3)	23	0	**30**	23	23	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**
D5(2,5)	14	1	**29**	14	14	27	28	**30**	**30**	**30**	**30**	**30**	**30**	**30**
D5(3,5)	7	6	**30**	7	3	29	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**
D5(3,10)	10	19	**29**	10	5	28	28	29	**30**	**30**	29	28	**30**	**30**
D5(5,10)	0	4	28	0	0	28	**29**	**30**	**30**	**30**	**30**	29	**30**	28
D5(7,10)	4	20	**30**	4	6	17	19	**30**	**30**	**30**	**30**	27	**30**	22
D5(3,20)	10	16	25	10	2	**26**	**26**	**30**	**30**	**30**	**30**	28	**30**	**30**
D5(6,20)	3	13	**29**	3	0	4	5	**30**	**30**	**30**	**30**	27	**30**	29
D5(9,20)	2	4	**30**	2	4	9	10	29	29	**30**	29	29	20	16
D5(12,20)	0	0	**25**	–	0	3	6	5	5	**30**	–	7	19	14
D5(5,50)	0	9	**30**	–	0	28	28	**30**	**30**	**30**	–	**30**	**30**	**30**
D5(8,50)	14	4	**30**	–	6	1	0	**30**	**30**	**30**	–	29	27	26
D5(10,50)	0	0	**30**	–	3	0	2	29	29	**30**	–	29	28	28
D5(15,50)	0	0	**30**	–	1	13	13	12	12	**30**	–	16	26	23
D5(25,50)	0	0	**29**	–	0	9	9	0	0	**30**	–	0	23	22
D5(5,80)	2	25	**30**	–	0	29	29	**30**	**30**	**30**	–	25	**30**	29
D5(10,80)	1	1	**28**	–	4	0	0	**30**	29	**30**	–	29	24	23
D5(15,80)	0	0	**30**	–	0	2	3	11	11	**30**	–	16	25	22
D5(20,80)	0	0	**30**	–	0	9	8	2	2	**30**	–	5	25	22
D5(30,80)	0	0	**30**	–	0	16	16	0	0	**30**	–	0	27	27
W3(5)	0	0	29	0	0	**30**	**30**	**30**	**30**	**30**	**30**	22	**30**	**30**
W3(15)	0	0	**30**	–	0	**30**	**30**	**30**	**30**	**30**	–	23	**30**	**30**
W3(25)	0	0	**30**	–	0	**30**	**30**	**30**	**30**	**30**	–	**30**	**30**	**30**
D2(5)	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**	**30**
D2(15)	0	0	**30**	–	0	**30**	28	4	4	**30**	–	6	**30**	28
D2(25)	0	0	**30**	–	0	**30**	29	0	0	**30**	–	0	**30**	28
P2(3,5)	**30**	**30**	**30**	**30**	**30**	3	29	**30**	**30**	**30**	**30**	**30**	0	**30**
P2(3,10)	10	10	**30**	–	13	0	0	23	23	**30**	–	23	0	21
P2(8,5)	**30**	**30**	**30**	**30**	**30**	0	0	**30**	**30**	**30**	**30**	**30**	0	1
P2(8,10)	6	6	**30**	–	9	0	0	27	27	**30**	–	28	0	1
S2(3)	**30**	**30**	0	**30**	0	0	0	**30**	**30**	0	**30**	0	0	0
S2(5)	**30**	**30**	0	**30**	0	0	0	**30**	**30**	0	**30**	0	0	0

For brevity, DTLZ5(I, m), WFG3(m), DTLZ2(m), POW-DTLZ2(p, m), SUM-DTLZ2(m) are abbreviated as D5(I, m), W3(m), D2(m), P2(p, m), S2(m), respectively; "–" means the corresponding data is not available.

from Table IX, the number of objectives can be either larger or smaller than $|\mathcal{F}_{\mathcal{T}}|$, which may depend on which difficulty has a greater impact on the algorithm. When concerning $\mathcal{N}_2$, they yield the excellent results on all these instances, most of which reach the best frequency of 30. It can be inferred that, within this range of dimensionality, the much higher quality of $\mathcal{N}_2$ can simultaneously provide the significant reduction in misdirection and the adequate improvement in the coverage of the Pareto front, thus alleviating both two difficulties with these algorithms effectively.

3) As the dimensionality $|\mathcal{F}_{\mathcal{T}}|$ gets higher ($|\mathcal{F}_{\mathcal{T}}| \geq 12$), the second difficulty would gradually become a dominating factor that influences the behavior of these algorithms corresponding to $\mathcal{N}_1$, which may lead to the severe over-reduction of objectives. It is interesting to note that, even when it comes to $\mathcal{N}_2$, they still perform poorly on such instances, although there generally exists certain performance improvement. To have a further investigation, in Table X, we list these instances separately and show the average number of objectives obtained per run corresponding to $\mathcal{N}_1$ and $\mathcal{N}_2$. As seen from Table X, the phenomena of the over-reduction of objectives occurs on each concerned instance for both $\mathcal{N}_1$ and $\mathcal{N}_2$. Thus, it can be inferred that, although the first difficulty could be almost eliminated on $\mathcal{N}_2$, the second difficulty is not reduced substantially here and still imposes the negative influence to these algorithms because it is generally very hard for a sample set to achieve an adequate coverage of the Pareto front in such a high-dimensional (i.e., $|\mathcal{F}_{\mathcal{T}}| \geq 12$) objective space.

Besides the above common features, we can obtain the following observations from Table VII by making a comparison among the four dominance structure-based algorithms.

1) For $\mathcal{N}_1$, NSGA-II-η and NSGA-II-δ yield the same results on 21 out of 32 instances. As for $\mathcal{N}_2$, the

TABLE VIII
AVERAGE NUMBER OF OBJECTIVES IDENTIFIED PER RUN BY NSGA-II-δ, NSGA-II-η, AND GREEDY-δ ON THE TEST INSTANCES WITH THE DIMENSIONALITY $|\mathcal{F}_\mathcal{T}| \leq 6$, CORRESPONDING TO $\mathcal{N}_1$

Test Instance	NSGA-II-δ	NSGA-II-η	Greedy-δ
DTLZ5(2,3)	2.23	3.00	2.23
DTLZ5(2,5)	2.70	3.37	2.70
DTLZ5(3,5)	3.83	3.83	4.03
DTLZ5(3,10)	3.93	3.53	4.43
DTLZ5(5,10)	6.80	6.10	6.97
DTLZ5(3,20)	4.23	3.93	4.77
DTLZ5(6,20)	7.07	6.73	7.53
DTLZ5(5,50)	6.50	5.70	7.40
DTLZ5(5,80)	6.13	5.17	6.63
WFG3(5)	3.63	3.57	3.90
WFG3(15)	7.10	14.57	7.90
WFG3(25)	10.07	18.73	10.90

TABLE IX
AVERAGE NUMBER OF OBJECTIVES IDENTIFIED PER RUN BY NSGA-II-δ, NSGA-II-η, AND GREEDY-δ ON THE TEST INSTANCES WITH THE DIMENSIONALITY $7 \leq |\mathcal{F}_\mathcal{T}| \leq 10$, CORRESPONDING TO $\mathcal{N}_1$

Test Instance	NSGA-II-δ	NSGA-II-η	Greedy-δ
DTLZ5(7,10)	6.10	7.27	6.30
DTLZ5(9,20)	7.70	8.20	8.00
DTLZ5(8,50)	7.77	9.73	8.10
DTLZ5(10,50)	8.73	8.50	9.07
DTLZ5(10,80)	8.90	8.60	9.40
POW-DTLZ2(3,10)	9.30	9.23	9.43
POW-DTLZ2(8,10)	9.20	9.03	9.30

TABLE X
AVERAGE NUMBER OF OBJECTIVES IDENTIFIED PER RUN BY NSGA-II-δ, NSGA-II-η, AND GREEDY-δ ON THE TEST INSTANCES WITH THE DIMENSIONALITY $|\mathcal{F}_\mathcal{T}| \geq 12$, CORRESPONDING TO $\mathcal{N}_1$ AND $\mathcal{N}_2$

Test Instance	$\mathcal{N}_1$			$\mathcal{N}_2$		
	NSGA-II-δ	NSGA-II-η	Greedy-δ	NSGA-II-δ	NSGA-II-η	Greedy-δ
D5(12,20)	8.37	8.30	8.77	10.77	10.77	11.00
D5(15,50)	12.67	12.60	12.97	14.23	14.23	14.53
D5(25,50)	14.03	14.03	14.93	15.63	15.67	16.23
D5(15,80)	12.50	12.43	13.00	14.23	14.23	14.43
D5(20,80)	13.23	13.27	14.07	17.90	17.90	18.47
D5(30,80)	15.67	15.50	16.57	15.70	15.67	16.50
D2(15)	11.50	11.50	12.13	13.33	13.33	13.73
D2(25)	11.87	11.83	12.73	13.80	13.80	14.77

For brevity, DTLZ5(I, m) and DTLZ2(m) are abbreviated as D5(I, m) and D2(m) respectively.

performance of NSGA-II-η is identical to that of NSGA-II-δ on all the instances except DTLZ5(3, 5), where there only exists a slight difference. Taken together, we conclude that NSGA-II-η can generally provide comparable performance to NSGA-II-δ.

2) NSGA-II-δ always obtains the same results as that of ExactAlg (if available) for both $\mathcal{N}_1$ and $\mathcal{N}_2$. This is because NSGA-II-δ has the strong ability to achieve the optimal solutions to the δ-MOSS problems, which has been verified in detail previously.

3) It is interesting to find that NSGA-II-δ and ExactAlg do not show the absolute superiority over Greedy-δ in identifying the essential objective set, although they generally obtain better solutions to δ-MOSS problems as demonstrated before. By further observation, Greedy-δ could be slightly better than NSGA-II-δ and ExactAlg on the instances with relatively high dimensionality. This can also be attributed to the inadequate coverage of Pareto front by the given sample set. In such situation, NSGA-II-δ and ExactAlg tend to over-reduce the objectives as illustrated before, in other words, they would usually return an objective subset with size $k_1 < |\mathcal{F}_\mathcal{T}|$ for a given δ_0. But due to the weaker search ability, Greedy-δ may yield an objective subset with size $k_2 > k_1$ by solving the same δ-MOSS problem and thus may hit $|\mathcal{F}_\mathcal{T}|$ with a higher probability. Nevertheless, it seems unreasonable to regard the unintentional success as an advantage of Greedy-δ. Indeed, the performance of NSGA-II-δ and ExactAlg is still in overall better than that of Greedy-δ, which is particularly clear on the SUM-DTLZ2(m) instances.

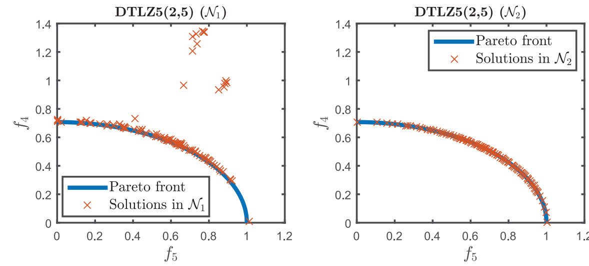


Fig. 10. DTLZ5(2, 5): a randomly selected sample set in $\mathcal{N}_1$ and $\mathcal{N}_2$, respectively, corresponding to $\mathcal{F}_\mathcal{T} = \{f_4, f_5\}$.

Based on Table VII, we now consider the correlation-based approaches, i.e., NSGA-II-γ, L-PCA, and NL-MVU-PCA, in comparison. On the whole, we can obtain the following findings about them.

1) NSGA-II-γ achieves remarkable performance on all the test instances except SUM-DTLZ2(m), corresponding to both $\mathcal{N}_1$ and $\mathcal{N}_2$. And it even obtains the optimal results on all such instances for $\mathcal{N}_2$. These results clearly indicate that NSGA-II-γ has great and stable strength in dealing with varying degree of misdirection in the sample set. To visually reflect this strength of NSGA-II-γ, we show a sample set of DTLZ5(2,5) corresponding to $\mathcal{N}_1$ and $\mathcal{N}_2$, respectively, in Fig. 10; and we use the parallel coordinate plot to show a sample set of an instance with higher dimensionality, i.e., DTLZ5(6, 20), for $\mathcal{N}_1$ and $\mathcal{N}_2$, respectively, in Fig. 11. From the two figures, $\mathcal{N}_1$ shows features obviously affected by misdirection in contrast to $\mathcal{N}_2$: for DTLZ5(2, 5), a fraction of solutions in $\mathcal{N}_1$ are far away from the true Pareto front; as for DTLZ5(6, 20), some solutions in $\mathcal{N}_1$ show conflict between f_i and f_j, where $i, j \in [1, 15]$.

2) L-PCA and NL-MVU-PCA perform quite well on DTLZ5(I, m) with a relatively low dimensionality and all

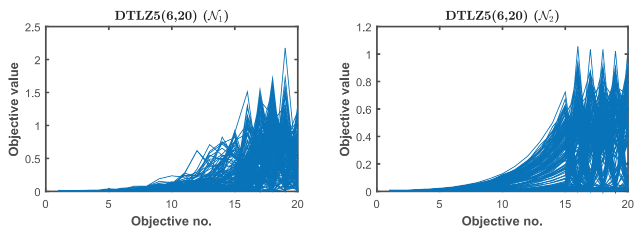


Fig. 11. DTLZ5(6, 20): the parallel coordinate plots for a randomly selected sample set in $\mathcal{N}_1$ and $\mathcal{N}_2$, respectively.

WFG3(m) instances, e.g., DTLT5(5,50) and WFG3(25), for both $\mathcal{N}_1$ and $\mathcal{N}_2$. This implies that they have a certain capability to handle the misdirection effectively. However, it appears that their ability deteriorates seriously on the redundant problems at higher dimensionality, e.g., DTLZ5(10,50) and DTLZ5(20,80). This inference is made according to the fact that L-PCA and NL-MVU-PCA yield the inferior results on such instances for $\mathcal{N}_1$ but still show decent performance on them for $\mathcal{N}_2$. Moreover, it is worth noting that both two algorithms can adequately address the nonredundant problem, i.e., DTLZ2(m), in all the considered cases.

3) The performance of NSGA-II-γ is generally better than or at least equivalent to that of L-PCA and NL-MVU-PCA. Notably, NSGA-II-γ shows much advantage over both L-PCA and NL-MVU-PCA on the instances with very strong nonlinearity between redundant objectives, i.e., POW-DTLZ2(8,5) and POW-DTLZ2(8,10). Thus, it can be safely concluded that, compared with L-PCA and NL-MVU-PCA, NSGA-II-γ has stronger ability not only in coping with misdirection but also in handling the nonlinearity on these problems.
4) NL-MVU-PCA in overall performs better than L-PCA corresponding to $\mathcal{N}_1$, implying that NL-MVU-PCA may be better at coping with the misdirection than L-PCA. However, L-PCA in general slightly outperforms NL-MVU-PCA on DTLZ5(I, m) instances corresponding to $\mathcal{N}_2$, which makes sense because there is only linear relationship between redundant objectives over the Pareto front of DTLZ5(I, m) and there is not so much misdirection in $\mathcal{N}_2$. Moreover, it is worth noting that NL-MUV-PCA can address POW-DTLZ2(3, 5) and POW-DTLZ2(3, 10) corresponding to $\mathcal{N}_2$ very well, whereas L-PCA fails to identify the essential objective set in all trials here. So, it can be inferred that NL-MVU-PCA is clearly better than L-PCA in terms of handling the nonlinearity, although it cannot compare with NSGA-II-γ in this respect.

Still based on Table VII, we gain the following insights through the comparison between the considered dominance structure-based approaches and correlation-based approaches.

1) All three correlation-based algorithms do not work well for SUM-DTLZ2(m) in all the cases due to their inherent limitation as indicated in Section IV-C, whereas NSGA-II-δ, NSGA-II-η, and ExactAlg can address the two SUM-DTLZ2(m) instances of interest perfectly.
2) Except on SUM-DTLZ2(m), NSGA-II-γ performs better than, or competitively to all four dominance structure-based algorithms corresponding to both $\mathcal{N}_1$ and $\mathcal{N}_2$, due to its strength in handling misdirection.
3) The situation is much more complex when having a comparison between L-PCA (NL-MVU-PCA) and dominance structure-based algorithms, which is mainly because the higher dimensionality would decrease the ability of L-PCA (NL-MVU-PCA) in handling the misdirection as mentioned before. But by and large, without considering SUM-DTLZ2(m), for the problem with very low or very high dimensionality, L-PCA and NL-MVU-PCA usually perform better; whereas for problems with medium dimensionality, the four dominance structure-based algorithms would exhibit certain advantage.
4) In summary, when the user has a high-quality sample set at hand and know that the true dimensionality may not be very high, dominance structure-based approach would serve as a better choice since they generally apply to all types of problems under such condition; otherwise the correlation-based approaches are suggested for objective reduction.

Lastly, it is worth pointing out that, although NSGA-II-δ and NSGA-II-η have shown comparable performance in our experiments, NSGA-II-δ needs to be executed on a normalized sample set. Since normalization is a nontrivial task in the presence of outliers, NSGA-II-η appears to be a better choice than NSGA-II-δ from this perspective.

VI. Applications to Real-World Problems

In this section, the performance of the proposed multiobjective approaches are further investigated on two real-world problems: water resource problem [58] and car side-impact problem [52], [59].

A. Water Resource Problem

The water resource problem consists of five objectives and seven constraints, which relates to optimal planning for a storm drainage system [58]. As a basis for the analysis, a sample set is first produced by running SPEA2-SDE with a population size of 200 and 2000 generations.

In Table XI, we show the detailed results of the proposed multiobjective approaches. From this table, it is reasonable to identify the reduced objective set as $\{f_2, f_3, f_5\}$ for both NSGA-II-δ and NSGA-II-η. As for NSGA-II-γ, two objective sets $\{f_1, f_2, f_4, f_5\}$ and $\{f_2, f_3, f_4, f_5\}$ lead to the same error 0.109, and appear to be the best choices for the reduced objective set. To validate the results, Fig. 12 uses the parallel coordinate plots to visualize the nondominated solutions obtained by running SPEA2-SDE on the original objective set and the three reduced objective sets, respectively, where the objective values are scaled by constant factors suggested in [60] to aid visualization. As seen from Fig. 12, the parallel coordinate plot corresponding to each of the three reduced objective set closely matches with that obtained using the original objective set, which confirms that either of the three

TABLE XI
RESULTS OF NSGA-II-δ, NSGA-II-η, AND NSGA-II-γ FOR THE WATER RESOURCE PROBLEM

Algorithm	k	$\delta/\eta/\gamma$	Objective Set
NSGA-II-δ	1	1	$\{f_i\}, i = 1, 2, 3, 4, 5$
	2	0.789	$\{f_3, f_4\}$
	3	0.078	$\{f_2, f_3, f_5\}$
	4	0	$\{f_2, f_3, f_4, f_5\}$
NSGA-II-η	1	0.995	$\{f_i\}, i = 1, 2, 3, 4, 5$
	2	0.51	$\{f_3, f_4\}$
	3	0.03	$\{f_2, f_3, f_5\}$
	4	0	$\{f_2, f_3, f_4, f_5\}$
NSGA-II-γ	1	0.7	$\{f_i\}, i = 2, 5$
	2	0.388	$\{f_i, f_4\}, i = 1, 3$
	3	0.313	$\{f_2, f_3, f_4\}$
	3	0.313	$\{f_1, f_2, f_4\}$
	3	0.313	$\{f_2, f_3, f_5\}$
	3	0.313	$\{f_1, f_2, f_5\}$
	4	0.109	$\{f_1, f_2, f_4, f_5\}$
	4	0.109	$\{f_2, f_3, f_4, f_5\}$
	5	0	$\{f_1, f_2, f_3, f_4, f_5\}$

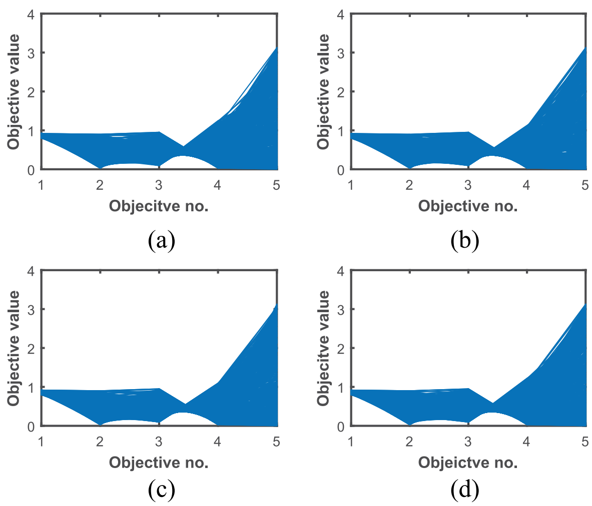


Fig. 12. Water resource problem: parallel coordinate plots for the nondominated solutions obtained by running SPEA2-SDE. (a) On the original objective set. On the reduced objective set (b) $\{f_2, f_3, f_5\}$, (c) $\{f_1, f_2, f_4, f_5\}$, and (d) $\{f_2, f_3, f_4, f_5\}$.

reduced objective sets is sufficient to generate a good estimate of the Pareto front of this problem.

B. Car Side-Impact Problem

The car side-impact problem is transformed into a many-objective problem with 11 objectives and 10 constraints [52]. The sample set is still generated by SPEA2-SDE with a population size of 200 and 2000 generations.

Fig. 13 shows how error decreases with the number of objectives increasing in the results of three multiobjective approaches for the car side-impact problem. From Fig. 13, the best tradeoff for NSGA-II-δ is obviously achieved at $k = 6$, whereas the best tradeoff for NSGA-II-η or NSGA-II-γ is not so apparent, which depends on the user's preference. Here, we set the same threshold as in Section V-F for η and γ

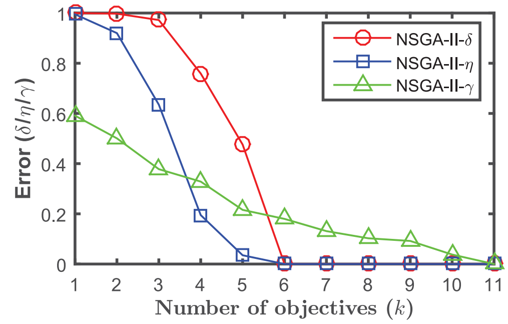


Fig. 13. Results of NSGA-II-δ, NSGA-II-η, and NSGA-II-γ for the car side-impact problem.

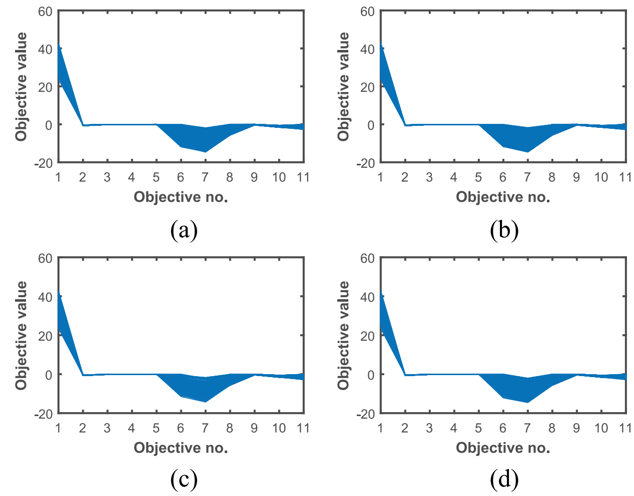


Fig. 14. Car side-impact problem: parallel coordinate plots for the nondominated solutions obtained by running SPEA2-SDE. (a) On the original objective set. On the reduced objective set (b) $\{f_1, f_4, f_5, f_9, f_{10}, f_{11}\}$, (c) $\{f_1, f_4, f_5, f_8, f_9, f_{11}\}$, and (d) $\{f_1, f_2, f_5, f_6, f_8, f_9, f_{11}\}$.

errors so as to choose a single objective set. Eventually, both NSGA-II-δ and NSGA-II-η return $\{f_1, f_4, f_5, f_9, f_{10}, f_{11}\}$ or $\{f_1, f_4, f_5, f_8, f_9, f_{11}\}$ as the reduced objective set, while NSGA-II-γ selects $\{f_1, f_2, f_5, f_6, f_8, f_9, f_{11}\}$. In Fig. 14, we show the parallel coordinate plots corresponding to the nondominated solutions obtained by running SPEA2-SDE on the original objective set and three reduced objective sets, respectively. As seen from Fig. 14, the four parallel coordinate plots are quite similar, verifying that all the three reduced objective sets are enough to obtain a good estimate to the Pareto front of this problem.

C. Discussion

From the above results, all three multiobjective algorithms are effective in reducing the number of objectives for the two real-world problems. The reduced objective sets yielded by NSGA-II-δ and NSGA-II-η are the same, but are different from those by NSGA-II-γ. The true Pareto front is unknown for the two real-world problems, so it seems difficult to have a definite conclusion which result is better. But because the objective sets obtained by NSGA-II-δ (NSGA-II-η) have a smaller size and are verified to be sufficient to represent the original objective set, they are more likely to be the true

TABLE XII
ESSENTIAL OBJECTIVE SETS OBTAINED FOR THE WATER RESOURCE PROBLEM IN DIFFERENT STUDIES

Reference	Essential Objective Set
[2]	$\{f_2, f_3, f_4, f_5\}$ or $\{f_1, f_2, f_4, f_5\}$
[39]	$\{f_1, f_2, f_5\}$ or $\{f_2, f_3, f_5\}$
[49]	$\{f_3, f_4\}$
[42], [52]	$\{f_2, f_3, f_4, f_5\}$

TABLE XIII
AVERAGE IGD (INCLUDING THE STANDARD DEVIATION IN THE BRACKET) AND THE AVERAGE COMPUTATION TIME (IN SECONDS) OVER 30 RUNS ON 15-OBJECTIVE WFG3

	SDE_{ref}	SDE_{so}	$\text{SDE}_{\text{online}}$	A-NSGA-III
IGD	0.2820 (0.1044)	0.1466 (0.0662)	0.1098 (0.0475)	7.3570 (2.6868)
Time	52.32	27.50	22.99	44.39

essential objective set. One possible reason leading to the difference is that the correlation analysis in the correlation-based approaches, e.g., NSGA-II-γ, is not enough to capture the complex relationship between objectives in the two real-world problems, although it generally works well on the existing benchmark problems. Further explanation about this can be found in Section IV-C.

Moreover, it is interesting to find that the identified essential objective sets for the two real-world problems usually vary in different studies. Table XII shows the results on the water resource problem taken from five existing studies. Considering that these algorithms can usually identify the essential objective set exactly for numerous benchmark problems, it can be inferred that the two real-world problems considered here may pose greater challenge to the existing objective reduction algorithms, especially the correlation-based approaches.

Lastly, it is worth noting that the qualitative method used to validate the obtained reduced objective set in this section follows the practice in [2], [39], and [49]. This validation method is reasonable considering the Pareto fronts of the real-world problems are generally unknown. However, its limitation is that it relies on an evolutionary many-objective optimizer which is a stochastic algorithm and may not have enough optimization ability. To alleviate the limitation, we indeed run a high performing algorithm (i.e., SPEA2-SDE) for a large number of generations until convergence to obtain a nondominated solution set for the parallel coordinate plot. Figs. 12 and 14 only show the results of one simulation run, but many repetitions are made in practice and almost unchanged parallel coordinate plots are observed.

VII. BENEFITS OF THE PROPOSED APPROACHES

This section aims to discuss the benefits of the proposed objective reduction algorithms in optimization, visualization, and decision making, with some illustrative examples.

A. On Objective Reduction Assisted Optimization

Generally, the objective reduction approaches can assist optimization in two ways. The first one, referred to as subsequent optimization in this paper, is straightforward: after a reduced objective set is obtained by an objective reduction algorithm, an MOEA continues to run on the reduced set for a number of generations and returns the nondominated solutions obtained. The second, more commonly used approach, is the so-called online objective reduction [50], [51], i.e., the objective reduction algorithm is embedded into the iterations of the MOEA. Next, we use both techniques based on NSGA-II-γ in order to demonstrate the benefits of the proposed objective reduction approaches in optimization.

For the subsequent optimization, we first run a many-objective algorithm, i.e., SPEA2-SDE, for 100 generations, then the set of nondominated solutions obtained is input into NSGA-II-γ for objective reduction, lastly we further run SPEA2-SDE on the reduced objective set for 300 generations. For convenience, we denote this optimization mode as SDE_{so}. As for online objective reduction, we still employ SPEA2-SDE as the underlying optimizer, and NSGA-II-γ is incorporated into SPEA2-SDE using the integration scheme in [1]. The resultant online objective reduction algorithm is denoted by $\text{SDE}_{\text{online}}$. To show the effectiveness, SDE_{so} and $\text{SDE}_{\text{online}}$ are compared with SPEA2-SDE without any objective reduction, denoted by SDE_{ref} for short. We also include A-NSGA-III [21] for comparison, which is a decomposition-based algorithm with reference vector adaptation and aims to better address MaOPs with nonuniformly distributed Pareto fronts. For a fair comparison, $\text{SDE}_{\text{online}}$, SDE_{ref}, and A-NSGA-III use 400 generations for optimization and the same population size is set for all the algorithms. The other parameters for SPEA2-SDE and A-NSGA-III follow the original studies [19], [21]. NSGA-II-γ uses the same parameter setting as in Section V-F.

First, we use a benchmark problem, i.e., 15-objective WFG3, for illustration purposes. The inverted generational distance (IGD) [61] is used as the performance metric, where the reference set is produced by uniformly sampling 1000 points on the true Pareto front. A-NSGA-III initially uses two-layered reference vectors with two divisions in both boundary and inside layers, leading to a population size of 240. Table XIII reports the average IGD (including the standard deviation) and the average computational time over 30 runs by four algorithms in comparison. From Table XIII, SDE_{so} and $\text{SDE}_{\text{online}}$ clearly outperform SDE_{ref} in terms of both effectiveness and efficiency, which demonstrates the usefulness of NSGA-II-γ in facilitating the effect of many-objective optimization. A-NSGA-III performs quite poorly here, and the large IGD indicates that it fails to approach the Pareto front. The observation of A-NSGA-III also implies that reference vector adaption may not be strong enough to locate a lower-dimensional front from a high-dimensional objective space. Fig. 15 further compares the convergence curves of average IGD with the number of generations. It can be seen that the proposed objective reduction algorithm in SDE_{so} and $\text{SDE}_{\text{online}}$ promotes the convergence of the many-objective optimizer.

Next, we consider a real-world problem, i.e., the water resource problem. Since its Pareto front is unknown, the

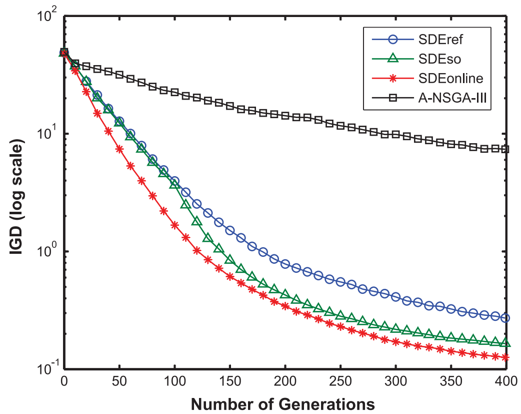


Fig. 15. Convergence curves of the average IGD (over 30 runs) obtained with the number of generations for SDE_{ref}, SDE_{so}, $\text{SDE}_{\text{online}}$, and A-NSGA-III on 15-objective WFG3.

TABLE XIV
AVERAGE HV (INCLUDING THE STANDARD DEVIATION IN THE BRACKET) AND THE AVERAGE COMPUTATION TIME (IN SECONDS) OVER 30 RUNS ON THE WATER RESOURCE PROBLEM

	SDE_{ref}	SDE_{so}	$\text{SDE}_{\text{online}}$	A-NSGA-III
HV	0.4200 (0.0376)	1.0030 (0.0435)	1.0377 (0.0157)	1.0537 (0.0036)
Time	35.41	31.25	31.87	30.64

HV [12] is used as the performance metric and is computed in the same way as that in [17]. Table XIV shows the average HV along with the standard deviation and the average computational time. From this table, the two objective reduction assisted optimization algorithms, i.e., SDE_{so} and $\text{SDE}_{\text{online}}$, achieve considerable performance improvement over SDE_{ref} in terms of HV. Unlike on 15-objective WFG3, A-NSGA-III performs well on this problem and is even a little better than SDE_{so} and $\text{SDE}_{\text{online}}$. Given that the water resource problem may only have one or two redundant objectives as depicted in Section VI-A, the possible reason is that the reference vector adaption in A-NSGA-III can effectively handle a slight degree of degeneration of Pareto front here.

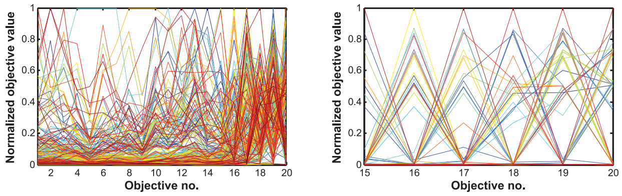
According to Tables XIII and XIV, the proposed objective reduction algorithm can reduce the overall time of optimization although it needs additional computational costs. This benefit will become clearer when the underlying optimizer is an HV-based MOEA or the objective computations of the MaOP are expensive. The former scenario can be referred to the study by Brockhoff and Zitzler [50], while the latter can be seen in a recent work by Carreras *et al.* [62].

Although the effectiveness of objective reduction assisted optimization has been shown by the experiments, we should also realize its limitation. That is, if the objective reduction algorithm fails to work properly, the removal of essential objectives during optimization may bias the search process. The manner of subsequent optimization is more prone to this risk, since the removed objectives have no chance to be considered again. Nevertheless, based on our experimental results and those in the literature (see [40], [53], [63]), combining objective reduction with MOEAs is still a good avenue in practical many-objective optimization, particularly for those MaOPs with redundant objectives. The decomposition-based algorithms with reference vector adaption [21], [64]–[66] may be a competitive alternative for redundant MaOPs, but as far as our experimental study on A-NSGA-III is concerned, this kind of techniques do not show the clear advantage over objective reduction assisted optimization algorithms. In the future, more experimental comparisons between the two classes of techniques are still needed to further understand their strengths and weakness.

B. On Visualization and Decision Making

It is important to visualize Pareto front approximations in many-objective optimization, which can assist the user to better understand the nondominated solutions obtained and then make a final selection from them. There exist a number of many-objective visualization methods in the literature, e.g., parallel coordinate plots, heatmaps, bubble chart. The interested reader is referred to [34] or [35] for more details. These methods can often provide a useful visualization, but the representation of most of them can become very confusing when the number of objectives or nondominated solutions increases. The objective reduction approaches can remove some redundant objectives and obtain a lower-dimensional problem in some cases. If the number of objectives can be reduced to two or three, e.g., for WFG3, the visualization will be very intuitive. Otherwise, with the help of objective reduction, the visualization method used can still produce an easier visualization for the user by considering fewer objectives and a lower number of nondominated solutions.

Here, for illustration, we use the proposed NSGA-II-γ along with a popular many-objective visualization method, i.e., parallel coordinate plots, on the DTLZ5(6, 20) problem. Given the high number of objectives, the user usually uses a relatively large population size, e.g., 500, to cover the high-dimensional Pareto front with certain diversity. After running a optimizer, e.g., SPEA2-SDE, the user will obtain a nondominated set that is shown by the parallel coordinate plots in Fig. 16(a). However, Fig. 16(a) is indeed very unclear for the user to do visual analysis, and due to the misdirection, the nonconflicting relationships between objectives are not reflected obviously in this figure. After running NSGA-II-γ on the nondominated set shown in Fig. 16(a), an essential objective set $\{f_{15}, f_{16}, f_{17}, f_{18}, f_{19}, f_{20}\}$ is identified with the γ error 0.11, and the other objectives are regarded to behave in a nonconflicting manner. So, the user can infer that the true dimensionality of DTLZ5(6, 20) is much lower than 20 by NSGA-II-γ, and he/she will tend to use a nondominated set with a much smaller size, e.g., 50, for the appropriate coverage. To realize this, the user can first use the environmental selection in SPEA2-SDE to pick out 50 diverse distributed ones among the nondominated solutions in Fig. 16(a), and then refine them by optimizing the essential objectives identified by NSGA-II-γ. The final nondominated set obtained is shown in Fig. 16(b). Clearly, Fig. 16(b) is much easier than Fig. 16(a) for the user to comprehend and reveal good tradeoffs.



(a) (b)

Fig. 16. Comparison of the parallel coordinate plots obtained (a) before and (b) after objective reduction.

Due to the same reason as visualization, the objective reduction approaches can also ease the decision making of the user to select the final solution for redundant MaOPs. Note that, the proposed NSGA-II-η and NSGA-II-γ do not require that each objective of the nondominated solution set is normalized to the same range. But this is strongly recommended for decision making after the essential objectives are identified using NSGA-II-η or NSGA-II-γ, which can avoid the inconvenience in comparing differently scaled objective values. Moreover, when using the dominance structure-based objective reduction algorithms, a possible scenario in the decision making is that although two solutions are nondominated to each other in the reduced objective space, they have big differences in some removed objective dimensions. At this time, the user may feel uncomfortable with just dropping these objectives, because he/she would think that such differences are of concern to his/her decision. We suggest a possible solution for this scenario. That is, instead of completely neglecting the removed objectives in the decision making, we aggregate them into a single objective using a preferred aggregation function, and this objective together with the identified essential objectives will be presented to the user for consideration. This method can be expected to make a compromise between the effect of objective reduction and the user's concern about big differences in removed objectives.

VIII. Conclusion

In this paper, we have conducted a study on evolutionary multiobjective approaches to objective reduction. Specifically, we propose to view objective reduction as a multiobjective search problem and introduce three different multiobjective formulations of this problem, aiming to preserve either the dominance structure or the correlation structure of the given sample set. For each multiobjective formulation, two conflicting objectives are considered to be minimized simultaneously: the first is the number of objectives selected (k) and the other is the error (δ, η, or γ) incurred by removing the unselected objectives, Then, a multiobjective objective reduction algorithm (NSGA-II-δ, NSGA-II-η, or NSGA-II-γ) is formed by employing NSGA-II to achieve a good tradeoff among the two objectives and return a set of nondominated objective subsets that can offer decision support to the user. Furthermore, we have provided a detailed analysis of dominance structure- and correlation-based approaches based on several theorems, which is intended to clearly reveal the general strengths and limitations of the two major kinds of objective reduction techniques. Extensive experimental results and comparisons on a wide range of problems demonstrate the effectiveness of the proposed multiobjective approaches, and also unveil the characteristics of the algorithms investigated, which agree well with our theoretical analysis. The possible directions for future research are suggested as follows.

1) Since dominance structure- and correlation-based approaches have complementary advantages, it is possible to develop hybrid approaches that can combine the merits of both while simultaneously overcoming their disadvantages to a certain degree.
2) It is necessary to further investigate online objective reduction based on the proposed algorithms. On one hand, more advanced integration schemes need to be developed. A possible way is to perform objective reduction and multiobjective optimization simultaneously using the evolutionary multitasking paradigm [67]–[69]. On the other hand, the benefits of online objective reduction are worthy of further examination by extensive comparisons against decomposition-based algorithms with reference vector adaption [21], [64]–[66].
3) There are only a few types of many-objective degenerate benchmark problems available in the literature for objective reduction. It is very desirable to design new benchmark problems with characteristics suitable for comprehensive evaluation of objective reduction algorithms.

References

[1] D. Brockhoff and E. Zitzler, "Objective reduction in evolutionary multiobjective optimization: Theory and applications," *Evol. Comput.*, vol. 17, no. 2, pp. 135–166, 2009.

[2] D. K. Saxena, J. A. Duro, A. Tiwari, K. Deb, and Q. Zhang, "Objective reduction in many-objective optimization: Linear and nonlinear algorithms," *IEEE Trans. Evol. Comput.*, vol. 17, no. 1, pp. 77–99, Feb. 2013.

[3] P. J. Fleming, R. C. Purshouse, and R. J. Lygoe, "Many-objective optimization: An engineering design perspective," in *Proc. Evol. Multi-Criterion Optim.*, Guanajuato, Mexico, 2005, pp. 14–32.

[4] M. Asafuddoula, H. K. Singh, and T. Ray, "Six-sigma robust design optimization using a many-objective decomposition-based evolutionary algorithm," *IEEE Trans. Evol. Comput.*, vol. 19, no. 4, pp. 490–507, Aug. 2015.

[5] W. Mkaouer *et al.*, "Many-objective software remodularization using NSGA-III," *ACM Trans. Softw. Eng. Methodol.*, vol. 24, no. 3, p. 17, 2015.

[6] Y. Yuan and H. Xu, "Multiobjective flexible job shop scheduling using memetic algorithms," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 1, pp. 336–353, Jan. 2015.

[7] H. Ishibuchi, N. Tsukamoto, and Y. Nojima, "Evolutionary many-objective optimization: A short review," in *Proc. IEEE Congr. Evol. Comput.*, Hong Kong, 2008, pp. 2419–2426.

[8] B. Li, J. Li, K. Tang, and X. Yao, "Many-objective evolutionary algorithms: A survey," *ACM Comput. Surveys*, vol. 48, no. 1, p. 13, 2015.

[9] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[10] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm," in *Proc. Evol. Methods Design Optim. Control*, Athens, Greece, 2002, pp. 95–100.

[11] E. Zitzler and S. Künzli, "Indicator-based selection in multiobjective search," in *Proc. Int. Conf. Parallel Problem Solving Nat.*, Birmingham, U.K., 2004, pp. 832–842.

[12] N. Beume, B. Naujoks, and M. Emmerich, "SMS-EMOA: Multiobjective selection based on dominated hypervolume," *Eur. J. Oper. Res.*, vol. 181, no. 3, pp. 1653–1669, 2007.
[13] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Trans. Evol. Comput.*, vol. 11, no. 6, pp. 712–731, Dec. 2007.
[14] E. J. Hughes, "Multiple single objective Pareto sampling," in *Proc. IEEE Congr. Evol. Comput.*, Canberra, ACT, Australia, 2003, pp. 2678–2684.
[15] X. Zou, Y. Chen, M. Liu, and L. Kang, "A new evolutionary algorithm for solving many-objective optimization problems," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 5, pp. 1402–1412, Oct. 2008.
[16] S. Yang, M. Li, X. Liu, and J. Zheng, "A grid-based evolutionary algorithm for many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 17, no. 5, pp. 721–736, Oct. 2013.
[17] Y. Yuan, H. Xu, B. Wang, and X. Yao, "A new dominance relation-based evolutionary algorithm for many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 20, no. 1, pp. 16–37, Feb. 2016.
[18] S. F. Adra and P. J. Fleming, "Diversity management in evolutionary many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 15, no. 2, pp. 183–195, Apr. 2011.
[19] M. Li, S. Yang, and X. Liu, "Shift-based density estimation for Pareto-based algorithms in many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 18, no. 3, pp. 348–365, Jun. 2014.
[20] 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 Trans. Evol. Comput.*, vol. 18, no. 4, pp. 577–601, Aug. 2014.
[21] H. Jain and K. Deb, "An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part II: Handling constraints and extending to an adaptive approach," *IEEE Trans. Evol. Comput.*, vol. 18, no. 4, pp. 602–622, Aug. 2014.
[22] H. Wang, L. Jiao, and X. Yao, "Two_arch2: An improved two-archive algorithm for many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 19, no. 4, pp. 524–541, Aug. 2015.
[23] H. Ishibuchi, N. Tsukamoto, Y. Sakane, and Y. Nojima, "Indicator-based evolutionary algorithm with hypervolume approximation by achievement scalarizing functions," in *Proc. 12th Annu. Conf. Genet. Evol. Comput.*, Portland, OR, USA, 2010, pp. 527–534.
[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.
[25] C. A. R. Villalobos and C. A. C. Coello, "A new multi-objective evolutionary algorithm based on a performance assessment indicator," in *Proc. 14th Annu. Conf. Genet. Evol. Comput.*, Philadelphia, PA, USA, 2012, pp. 505–512.
[26] R. H. Gómez and C. A. C. Coello, "MOMBI: A new metaheuristic for many-objective optimization based on the R2 indicator," in *Proc. IEEE Congr. Evol. Comput.*, Cancún, Mexico, 2013, pp. 2488–2495.
[27] M. Asafuddoula, T. Ray, and R. Sarker, "A decomposition-based evolutionary algorithm for many objective optimization," *IEEE Trans. Evol. Comput.*, vol. 19, no. 3, pp. 445–460, Jun. 2015.
[28] Y. Yuan, H. Xu, B. Wang, B. Zhang, and X. Yao, "Balancing convergence and diversity in decomposition-based many-objective optimizers," *IEEE Trans. Evol. Comput.*, vol. 20, no. 2, pp. 180–198, Apr. 2016.
[29] Y.-Y. Tan, Y.-C. Jiao, H. Li, and X.-K. Wang, "MOEA/D+ uniform design: A new version of MOEA/D for optimization problems with many objectives," *Comput. Oper. Res.*, vol. 40, no. 6, pp. 1648–1660, 2013.
[30] I. Giagkiozis, R. C. Purshouse, and P. J. Fleming, "Generalized decomposition and cross entropy methods for many-objective optimization," *Inf. Sci.*, vol. 282, pp. 363–387, Oct. 2014.
[31] R. Wang, R. C. Purshouse, and P. J. Fleming, "Preference-inspired co-evolutionary algorithms using weight vectors," *Eur. J. Oper. Res.*, vol. 243, no. 2, pp. 423–441, 2015.
[32] A. L. Jaimes and C. A. C. Coello, "Some techniques to deal with many-objective problems," in *Proc. 11th Annu. Conf. Companion Genet. Evol. Comput.*, Montreal, QC, Canada, 2009, pp. 2693–2696.
[33] A. L. Jaimes, C. A. C. Coello, and D. Chakraborty, "Objective reduction using a feature selection technique," in *Proc. 10th Annu. Conf. Genet. Evol. Comput.*, Atlanta, GA, USA, 2008, pp. 673–680.
[34] D. J. Walker, R. Everson, and J. E. Fieldsend, "Visualizing mutually nondominating solution sets in many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 17, no. 2, pp. 165–184, Apr. 2013.
[35] T. Tušar and B. Filipič, "Visualization of Pareto front approximations in evolutionary multiobjective optimization: A critical review and the prosection method," *IEEE Trans. Evol. Comput.*, vol. 19, no. 2, pp. 225–245, Apr. 2015.
[36] M. Köppen and K. Yoshida, "Visualization of Pareto-sets in evolutionary multi-objective optimization," in *Proc. 7th Int. Conf. Hybrid Intell. Syst.*, Kaiserslautern, Germany, 2007, pp. 156–161.
[37] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Elect. Eng.*, vol. 40, no. 1, pp. 16–28, 2014.
[38] R. C. Purshouse and P. J. Fleming, "Conflict, harmony, and independence: Relationships in evolutionary multi-criterion optimisation," in *Proc. Evol. Multi-Criterion Optim.*, Faro, Portugal, 2003, pp. 16–30.
[39] H. K. Singh, A. Isaacs, and T. Ray, "A Pareto corner search evolutionary algorithm and dimensionality reduction in many-objective optimization problems," *IEEE Trans. Evol. Comput.*, vol. 15, no. 4, pp. 539–556, Aug. 2011.
[40] K. Deb and D. Saxena, "Searching for Pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems," in *Proc. IEEE Congr. Evol. Comput.*, Vancouver, BC, Canada, 2006, pp. 3353–3360.
[41] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A survey on evolutionary computation approaches to feature selection," *IEEE Trans. Evol. Comput.*, vol. 20, no. 4, pp. 606–626, Aug. 2016.
[42] X. Guo, Y. Wang, and X. Wang, "Using objective clustering for solving many-objective optimization problems," *Math. Problems Eng.*, vol. 2013, Apr. 2013, Art. no. 584909.
[43] H. Wang and X. Yao, "Objective reduction based on nonlinear correlation information entropy," *Soft Comput.*, vol. 20, no. 6, pp. 2393–2407, 2016.
[44] J. A. Duro, D. K. Saxena, K. Deb, and Q. Zhang, "Machine learning based decision support for many-objective optimization problems," *Neurocomputing*, vol. 146, pp. 30–47, Dec. 2014.
[45] 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.
[46] S. Huband, P. Hingston, L. Barone, and L. While, "A review of multiobjective test problems and a scalable test problem toolkit," *IEEE Trans. Evol. Comput.*, vol. 10, no. 5, pp. 477–506, Oct. 2006.
[47] H. Ishibuchi, H. Masuda, and Y. Nojima, "Pareto fronts of many-objective degenerate test problems," *IEEE Trans. Evol. Comput.*, vol. 20, no. 5, pp. 807–813, Oct. 2016.
[48] A. R. de Freitas, P. J. Fleming, and F. G. Guimarães, "Aggregation trees for visualization and dimension reduction in many-objective optimization," *Inf. Sci.*, vol. 298, pp. 288–314, Mar. 2015.
[49] X. Guo, Y. Wang, and X. Wang, "An objective reduction algorithm using representative Pareto solution search for many-objective optimization problems," *Soft Comput.*, vol. 20, no. 12, pp. 4881–4895, 2016.
[50] D. Brockhoff and E. Zitzler, "Improving hypervolume-based multiobjective evolutionary algorithms by using objective reduction methods," in *Proc. IEEE Congr. Evol. Comput.*, Singapore, 2007, pp. 2086–2093.
[51] A. L. Jaimes, C. A. C. Coello, and J. E. U. Barrientos, "Online objective reduction to deal with many-objective problems," in *Proc. Evol. Multi-Criterion Optim.*, Nantes, France, 2009, pp. 423–437.
[52] A. Sinha, D. K. Saxena, K. Deb, and A. Tiwari, "Using objective reduction and interactive procedure to handle many-objective optimization problems," *Appl. Soft Comput.*, vol. 13, no. 1, pp. 415–427, 2013.
[53] S. Bandyopadhyay and A. Mukherjee, "An algorithm for many-objective optimization with reduced objective computations: A study in differential evolution," *IEEE Trans. Evol. Comput.*, vol. 19, no. 3, pp. 400–413, Jun. 2015.
[54] B. Min, C. Park, I. Jang, J. M. Kang, and S. Chung, "Development of Pareto-based evolutionary model integrated with dynamic goal programming and successive linear objective reduction," *Appl. Soft Comput.*, vol. 35, pp. 75–112, Oct. 2015.
[55] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, nos. 1–2, pp. 81–93, 1938.
[56] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Boston, MA, USA: Addison-Wesley Longman, 1989.
[57] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable multi-objective optimization test problems," in *Proc. IEEE Congr. Evol. Comput.*, Honolulu, HI, USA, 2002, pp. 825–830.
[58] K. Musselman and J. Talavage, "A tradeoff cut approach to multiple objective optimization," *Oper. Res.*, vol. 28, no. 6, pp. 1424–1435, 1980.
[59] L. Gu *et al.*, "Optimisation and robustness for crashworthiness of side impact," *Int. J. Veh. Design*, vol. 26, no. 4, pp. 348–360, 2001.
[60] T. Ray, K. Tai, and K. C. Seow, "Multiobjective design optimization by an evolutionary algorithm," *Eng. Optim.*, vol. 33, no. 4, pp. 399–424, 2001.

[61] D. A. Van Veldhuizen and G. B. Lamont, "Multiobjective evolutionary algorithm research: A history and analysis," Dept. Elect. Comput. Eng., Air Force Inst. Technol., Wright-Patterson, OH, USA, Tech. Rep. 98-03, 1998.

[62] J. Carreras *et al.*, "Systematic approach for the life cycle multi-objective optimization of buildings combining objective reduction and surrogate modeling," *Energy Build.*, vol. 130, pp. 506–518, Oct. 2016.

[63] P. J. Copado-Méndez, C. Pozo, G. Guillén-Gosálbez, and L. Jiménez, "Enhancing the ϵ-constraint method through the use of objective reduction and random sequences: Application to environmental problems," *Comput. Chem. Eng.*, vol. 87, pp. 36–48, Apr. 2016.

[64] S. Jiang, Z. Cai, J. Zhang, and Y.-S. Ong, "Multiobjective optimization by decomposition with Pareto-adaptive weight vectors," in *Proc. 7th Int. Conf. Nat. Comput.*, Shanghai, China, 2011, pp. 1260–1264.

[65] Y. Qi *et al.*, "MOEA/D with adaptive weight adjustment," *Evol. Comput.*, vol. 22, no. 2, pp. 231–264, 2014.

[66] R. Cheng, Y. Jin, M. Olhofer, and B. Sendhoff, "A reference vector guided evolutionary algorithm for many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 20, no. 5, pp. 773–791, Oct. 2016.

[67] A. Gupta, Y.-S. Ong, and L. Feng, "Multifactorial evolution: Toward evolutionary multitasking," *IEEE Trans. Evol. Comput.*, vol. 20, no. 3, pp. 343–357, Jun. 2016.

[68] A. Gupta, Y.-S. Ong, L. Feng, and K. C. Tan, "Multiobjective multifactorial optimization in evolutionary multitasking," *IEEE Trans. Cybern.*, to be published.

[69] Y.-S. Ong and A. Gupta, "Evolutionary multitasking: A computer science view of cognitive multitasking," *Cogn. Comput.*, vol. 8, no. 2, pp. 125–142, 2016.

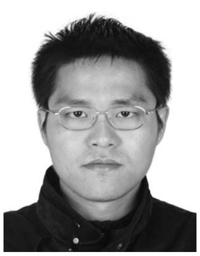
Yew-Soon Ong received the Ph.D. degree in artificial intelligence in complex design from the Computational Engineering and Design Center, University of Southampton, Southampton, U.K., in 2003.

He is a Professor and the Chair of the School of Computer Science and Engineering, Nanyang Technological University, Singapore, where he is the Director of the Data Science and Artificial Intelligence Research Center and the A*Star SIMTech-NTU Joint Laboratory on Complex Systems, and the Principal Investigator of the Data Analytics and Complex System Programme, Rolls-Royce@NTU Corporate Laboratory. His current research interests include computational intelligence span across memetic computation, complex design optimization, intelligent agents, and big data analytics.

Dr. Ong was a recipient of the 2015 *IEEE Computational Intelligence Magazine* Outstanding Paper Award and the 2012 IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION Outstanding Paper Award for his work pertaining to memetic computation. He is the Founding Editor-in-Chief of the IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE, an Associate Editor of the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, the IEEE TRANSACTIONS ON NEURAL NETWORK AND LEARNING SYSTEMS, and the IEEE TRANSACTIONS ON CYBERNETICS.

Abhishek Gupta received the B.Tech. degree from the National Institute of Technology Rourkela, Rourkela, India, and the Ph.D. degree in engineering science from the University of Auckland, Auckland, New Zealand, in 2014.

He is currently a Research Fellow with the Rolls-Royce@NTU Corporate Laboratory, Nanyang Technological University, Singapore. His current research interests include computational science, spanning the fields of continuum mechanics as well as computational intelligence, with recent emphasis on memetic computation and its application in evolutionary design.



Yuan Yuan received the Ph.D. degree from the Department of Computer Science and Technology, Tsinghua University, Beijing, China, in 2015.

He is currently a Post-Doctoral Research Fellow with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, USA. From 2014 to 2015, he was a visiting Ph.D. student with the Centre of Excellence for Research in Computational Intelligence and Applications, University of Birmingham, Birmingham, U.K. He was a Research Fellow with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, from 2015 to 2016. His current research interests include evolutionary computation, machine learning, and their applications.

Hua Xu received the B.S. degree from Xi'an Jiaotong University, Xi'an, China, in 1998, and the M.S. and Ph.D. degrees from Tsinghua University, Beijing, China, in 2000 and 2003, respectively.

He is an Associate Professor with the Department of Computer Science and Technology, Tsinghua University. He has published over 50 academic papers and holds ten invention patents of advanced controller. He is also the copyright owner of six software systems. His current research interests include data mining, intelligent information processing, and advanced process controllers for IC manufacturing equipments.

Dr. Xu was a recipient of the Second Prize from the National Science and Technology Progress of China, the First Prize from Beijing Science and Technology, and the Third Prize from Chongqing Science and Technology.