

Objective Reduction in Many-Objective Optimization: Evolutionary Multiobjective Approach and Critical 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 trade-off 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 the decision support to the user. Moreover, we conduct a critical analysis of two major categories of objective reduction approaches based on several theorems, aiming to disclose their general strengths and limitations. Lastly, the performance of the proposed multiobjective algorithms is studied extensively on various benchmark problems and two real-world problems. The numerical results and comparisons not only show the effectiveness and superiority of the proposed multiobjective algorithms but also agree with our critical analysis very well.

Index Terms—Objective reduction, many-objective optimization, multiobjective optimization, multiobjective evolutionary algorithms.

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, 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 nearly all the 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 last few years, some contributions have been made to counter the limitations of existing MOEAs in many-objective optimization, mainly including the alternative dominance relations [15]–[17] and the new diversity promotion mechanisms [18]–[21] in Pareto dominance-based MOEAs, the approximation of hypervolume values [22], [23] and more

computationally efficient performance indicators [24], [25] in indicator-based MOEAs, and the novel updating strategies [26], [27] and the generation of more uniformly distributed weight vectors [28], [29] in decomposition-based MOEAs. Nevertheless, these newly proposed many-objective evolutionary algorithms may not be strong enough to adequately handle MaOPs with more than around 15 objectives [17], [20], and their effectiveness is still needed to be further examined on more real-world problems. Moreover, the improvements on MOEAs cannot alleviate the difficulty of visualizing the high-dimensional Pareto front and making a decision to choose preferred solutions in many-objective optimization [20].

Instead of improving the scalability of existing MOEAs, *objective reduction* considers to decrease the problem difficulty by reducing the number of objectives in the decision making stage or during the search process, which is another important technique to deal with MaOPs [2], [8], [20], [30]. The motivation behind it is that, for many problems with m objectives in practice, there exists a smallest set of k ($k < m$) conflicting objectives that can generate the same Pareto front as the original problem. Such k objectives are often called as *essential*, while the rest as *redundant* [2]. The potential benefits of objective reduction have been clearly pointed out in several studies [1], [2], [31], which can be illustrated from two major aspects. On one hand, for a MaOP at hand, if the number of objectives can be reduced to no more than three, any state-of-the-art MOEA can be expected to solve it sufficiently; on the other hand, if the remaining number of objectives is larger than three, objective reduction is still helpful for making the subsequent search of multi- or many-objective optimizers more effective and efficient, and also ease the visualization of the Pareto front and the decision making process.

Note that, following the previous research [1], [2], [8], [30], objective reduction in this paper refers in particular to the selection of a subset of given objectives describing the original MOP best, which is somewhat similar to *feature selection*. However, it is worth pointing out that there is another related research direction aiming to determine a small set of arbitrary objectives for the original MOP, which is analogous to *feature transformation* and appears commonly in visualization of many-objective solution sets [32], [33]. In this respect, a prior work similar to ours is from Köppen and Yoshida [34], where the authors attempted to map a set of nondominated points in higher-dimensional space to a two-dimensional Euclidean space for visualization, whilst preserving both the distance relations and the Pareto dominance relations as good as

possible. They used NSGA-II to find a trade-off between the preservation of the two relations.

This paper focuses on the objective reduction related with feature selection. Feature selection has been extensively studied in machine learning and statistics, and plenty of relevant methods [35] are 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], [36], [37] or the conflict relation between objectives [31] needs to be taken into account, but their basic ideas in data analysis are still quite useful for developing effective approaches to objective reduction, which have already inspired several significant contributions [2], [31], [38] to this topic. Evolutionary computation (EC) approaches have gained much attention in feature selection [39] due to their global search potential in the large-scale search space. However, to our knowledge, there is no research applying EC to objective reduction in the sense of feature selection, despite the fact that the δ -MOSS and k -EMOSS problems [1] resulting from objective reduction have been both proved to be NP-hard and the exact algorithm is not applicable for larger instance of practical size [1]. Moreover, most existing objective reduction algorithms [2], [37], [40], [41] only return a unique reduced objective set as the outcome in a single simulation run, without considering various requirements in different situations, such as the error tolerance and the desired number of objectives, so that they generally fail to offer the decision support [42] to the user. Given the above, this paper presents the first study on evolutionary multiobjective approaches to objective reduction, exploiting both the global search ability of EC and the decision support characteristic of multiobjective optimization. More specifically, we propose three different multiobjective formulations of the objective reduction problem using three different kinds of errors (δ , η and γ) respectively, where the first two are based on the Pareto dominance structure and the last one is based on the correlation structure; and 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 critical analysis of existing objective reduction algorithms, which can be roughly categorized into *dominance structure-based approaches* and *correlation-based approaches* [1], [2], [37]. The motivation comes from the fact that, although the two types of approaches certainly have their own advantages and disadvantages, the past studies on objective reduction usually fail to analyze them carefully, or there exist some limitations in the analysis conducted. For example, Brockhoff and Zitzler [1] pointed out that the correlation-based approaches cannot guarantee that the Pareto dominance relation is preserved, but they did not further indicate in what scenarios the correlation-based approaches would fail due to this reason. And also, they did not discuss the potential drawbacks of dominance structure-based approaches and the rationality of correlation-based approaches. Saxena *et al.* [2] implied that the dominance structure-based approaches are extremely sensitive to misdirection, but the inference is only based on the performance of the algorithms proposed in

[1] with the δ error [1] fixed at 0. Additionally, their analysis is purely according to the experimental results, thereby making it difficult to disclose the inherent strength of dominance structure-based approaches and the inherent limitation of correlation-based approaches in general. In this paper, we take a more objective and comprehensive view toward both the two categories of approaches, and intend to clearly indicate their general strongness and weakness by combining the theoretical foundations of objective reduction, so that the user of these algorithms can have a better understanding of why, how and when they work. Moreover, the analysis is also expected to provide deeper insights to design more effective objective reduction algorithms in the future.

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

Our work mainly takes inspiration from Brockhoff and Zitzler [1], Saxena *et al.* [2], Jaimes *et al.* [31], and Singh *et al.* [37]. Based on these previous studies, our contributions to the topic of objective reduction are summarized as follows:

1) Inspired by [1], we propose to formulate objective reduction as a MOP and use NSGA-II to derive a trade-off 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 [37], we introduce a new error based on the dominance structure change, referred to as η , which is easier to comprehend than δ error in the context of decision support.

3) Inspired by the algorithms introduced in [31], we present a new error, referred to as γ , to measure the change of correlation structure. The computing of γ is characterized by the using 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], [31], [37], 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], [2], we make a further analysis on their strengths and limitations with both theorems and experiments. 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 multiob-

jective approaches are described in detail. Section IV conducts a critical 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 further applied to two real-world problems. In Section VII, the benefits of the proposed approaches are illustrated in optimization, visualization and decision making. Lastly, the concluding remarks and future work directions are given in Section VIII.

II. PRELIMINARIES AND BACKGROUND

A. Multi and Many-Objective Optimization

A general MOP can be mathematically formulated as

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

$\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ is a n -dimensional decision vector in the decision space Ω ; $\mathbf{f} : \Omega \rightarrow \Theta \subseteq \mathbb{R}^m$, 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, $m > 3$. $g_i(\mathbf{x}) \leq 0$ and $h_j(\mathbf{x}) = 0$ represent inequality and equality constraints respectively.

In a 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 trade-offs between objectives. Several basic concepts about MOPs are provided below.

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

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

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

Definition 4 (Weak ϵ -Dominance [43]): A vector $\mathbf{u} = (u_1, u_2, \dots, u_k)^T$ is said to *weakly ϵ -dominate* another vector $\mathbf{v} = (v_1, v_2, \dots, v_k)^T$, denoted by $\mathbf{u} \preceq^\epsilon \mathbf{v}$, iff $\forall i \in \{1, 2, \dots, k\} : u_i - \epsilon \leq v_i$.

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 distribute 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 Eq. (1); the original (or universe) objective set is denoted as $\mathcal{F}_0 = \{f_1, f_2, \dots, f_m\}$; PF_0 refers to the Pareto front of the original

MOP. For the 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 $\mathbf{u} = (f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x}))^T$, $\mathcal{F} := \{f_1, f_3\}$, then $\mathbf{u}^{(\mathcal{F})} = (f_1(\mathbf{x}), f_3(\mathbf{x}))^T$.

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 $\mathcal{F}' := \mathcal{F}_0 \setminus \mathcal{F}$ is $PF' := \{\mathbf{u}^{(\mathcal{F}')} \mid \mathbf{u} \in PF_0\}$

Definition 6: An objective set \mathcal{F} is said to be an *essential objective set*, iff $\mathcal{F}_0 \setminus \mathcal{F}$ 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 3-objective WFG3 problem [44], [45], which is a straight line from $(0, 0, 6)^T$ to $(1, 2, 0)^T$. From Fig. 1, 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.

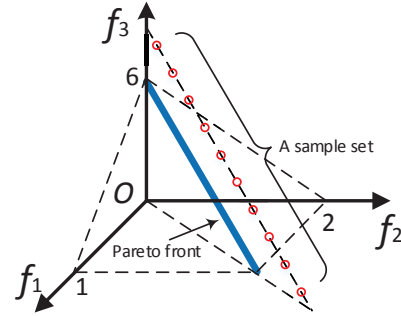


Fig. 1. The Pareto front and a sample set of the 3-objective WFG3 problem.

“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 $\forall \mathbf{u}, \mathbf{v} \in PF_0 : \mathbf{u}^{(\{f_i\})} \leq \mathbf{v}^{(\{f_i\})} \Leftrightarrow \mathbf{u}^{(\{f_j\})} \leq \mathbf{v}^{(\{f_j\})}$.

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 proved in Section IV-A. 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 a 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 conforming well with the dominance structure or the correlation structure of the Pareto front. The difference between the “approximation” and the “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 3-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 very easily and reliably. But note that, this set is indeed not a reasonably good Pareto front-approximation since it is a bit far away from the 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], [42] in it. The misdirection¹ here refers to the difference in the dominance structure or the correlation structure, between the Pareto front and the sample set [42]. To further explain this, Fig. 2 shows a sample set of 3-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 can be interpreted as *misdirection*, while the rest can be interpreted as *signal*. Similarly, from Fig. 2(b), the correlation structure 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].

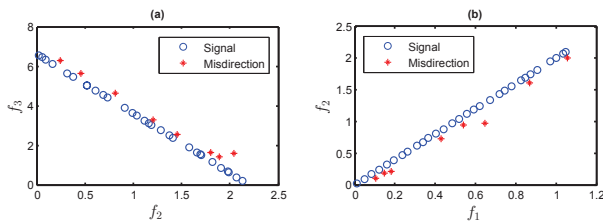


Fig. 2. The illustration of misdirection using a sample set of 3-objective WFG3.

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: dominance structure- and 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. Suppose there is an 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 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 $|\mathcal{F}| \leq k_0$.

This study by Brockhoff and Zitzler [1] proposed an exact algorithm for both the δ -MOSS and k -EMOSS problems. Given 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 (PCA) and maximum variance unfolding (MVU) [2]: Based on two well-known dimensionality reduction techniques, i.e., PCA and MVU, this study proposed two algorithms, namely L-PCA and NL-MVU-PCA, for linear 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 context of identifying the essential objective set.

Besides the above two works, there are a few other contributions in objective reduction. Jaimes *et al.* [31] 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.* [37] 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.* [40] 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 combining mutual information and correlation coefficient. Duro *et al.* [42] extended the framework proposed in [2] to include the analysis analogous to δ -MOSS and k -EMOSS, which could serve as a decision support for the decision makers. De Freitas *et al.* [46] introduced the concept of aggregation trees for the visualization of the results of

¹The references [2] and [42] 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.

MaOPs, and also used it as a means of performing objective reduction. Guo *et al.* [47] claimed that the corners used in PCSEA are not sufficient to reflect the entire Pareto front, and thereby developed an algorithm using the representative nondominated solutions instead. Wang and Yao [41] adopted the nonlinear correlation information entropy to measure both the 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 [48]–[52] have focussed 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 only *offline* objective reduction is considered. In Table I, we summarize the reviewed studies on objective reduction in each category.

TABLE I
SUMMARY OF THE STUDIES REVIEWED IN THIS PAPER

Algorithm Category	References
Dominance structure-based approaches	[1], [37], [46] [47], [48]
Correlation-based approaches	[2], [31], [40], [41] [42], [49]–[52]

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 e.g., [1], [31], devoted to the first issue and did not explicitly considered the second one, while the others, e.g., [2], [37], only concerned the second issue. The algorithms for the first issue can provide the decision support for the user by investigating δ -MOSS problems with different δ s and k -EMOSS problems with different k s; and 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 properly small δ value [2]. In this paper, the proposed multiobjective approaches adequately address the first issue that is usually more practical. But it appears difficult to have an experimental evaluation and comparison of the effect of objective reduction in this respect. So, 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 a 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 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 the number of selected objectives k ($k = |\mathcal{F}|$) and the δ error proposed in [1]. As depicted in Section II-D, δ is mathematically expressed as follows

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

δ can be seen as a sort of criterion that measures 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 an extreme case is that the maximum possible k value, i.e., $k = |\mathcal{F}_0|$, definitely leads to the minimum possible δ value, i.e., $\delta = 0$. With δ -OR formulation, the goal of objective reduction is to explore the Pareto front of objective subsets.

However, when using δ error, it should be assumed that all objective values 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 [42]. Here, inspired from the definition of the parameter R in [37], we propose an alternative error to measure the dominance structure change, denoted as η , which is free from 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 $\mathcal{N}_{\mathcal{NS}} = \{\mathbf{u} \in \mathcal{N} \mid \nexists \mathbf{v} \in \mathcal{N} : \mathbf{v}^{(\mathcal{F})} \prec \mathbf{u}^{(\mathcal{F})}\}$ and $\mathcal{N}_{\mathcal{DS}} := \mathcal{N} \setminus \mathcal{N}_{\mathcal{NS}}$. Then η is expressed as

$$\eta = |\mathcal{N}_{\mathcal{DS}}|/|\mathcal{N}| \quad (3)$$

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 , and the maximum $k = |\mathcal{F}_0|$ also makes η reach 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 subsection 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 [31].

The computing of γ needs to exploit the correlation between every two objectives in \mathcal{F}_0 . Instead of using Pearson's correlation coefficient as in the works [2] and [31], Kendall's rank correlation coefficient [53] is adopted here. This is because Pearson's correlation coefficient is sensitive only to linear relationships between two variables, while Kendall's rank correlation coefficient measures the extent to which, as one variable increases, the other variable tends to increase, no matter 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 on one objective would deteriorate/improve the other objective [36].

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.

Now, the procedures to compute γ for an objective subset $\mathcal{F} := \{f_{i_1}, f_{i_2}, \dots, f_{i_k}\} \subseteq \mathcal{F}_0$ is as follows. First, split \mathcal{F} into k different clusters C_1, C_2, \dots, 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 $f_l \in \mathcal{F}_0 \setminus \mathcal{F}$ 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

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

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.

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

Input: an objective subset $\mathcal{F} := \{f_{i_1}, f_{i_2}, \dots, f_{i_k}\} \subseteq \mathcal{F}_0$; the original objective set $\mathcal{F}_0 = \{f_1, f_2, \dots, f_m\}$

Output: k objective set clusters C_1, C_2, \dots, C_k

for $j = 1$ **to** k **do**

$C_j \leftarrow \{f_{i_j}\};$

for each objective f_l **in** $\mathcal{F}_0 \setminus \mathcal{F}$ **do**

$\min \leftarrow d(l, i_1);$

$c \leftarrow 1;$

for $j = 2$ **to** k **do**

if $d(l, i_j) < \min$ **then**

$\min \leftarrow d(l, i_j);$

$c \leftarrow j;$

$C_c \leftarrow C_c \cup f_l;$

The aim of γ -OR is to minimize both the number of selected objectives k and the γ value, where $\gamma \in [0, 1]$ since $d(i, j) \in [0, 1]$. Similar to that in δ -OR and η -OR, the two objectives of γ -OR are conflicting with each other to some extent. In addition, as indicated by Eq. (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 of “1” or “0” means that the corresponding objective is selected or excluded, respectively.

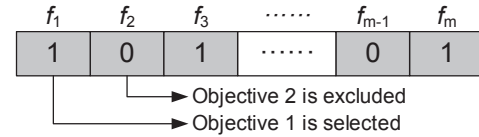


Fig. 3. Representation of chromosome as a binary string.

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 loops until the maximum generations is satisfied. At each generation g , the binary tournament selection [9], single-point crossover [54], and bit flip mutation [54] 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 in the elite 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 the 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 modelled as a 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 a deeper insight into the objective reduction problem and making 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 $PF_\delta = \{(k_j, \delta_j)^T \mid j = 1 : \kappa\}$ 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 $(k_\mu, \delta_\mu)^T$, where

$$\mu = \operatorname{argmax}_j \{\delta_j \mid \delta_j \leq \delta_0, j = 1 : \kappa\} \quad (5)$$

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

$$\nu = \operatorname{argmax}_j \{k_j \mid k_j \leq k_0, j = 1 : \kappa\} \quad (6)$$

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 the 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 if 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 the 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 $PF' \subseteq \{\mathbf{u}^{(\mathcal{F}')} \mid \mathbf{u} \in PF_0\}$.

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 $\nexists \mathbf{u}, \mathbf{v} \in PF_0 : \mathbf{u}^{(\mathcal{F}')} \prec \mathbf{v}^{(\mathcal{F}')} , \text{ where } \mathcal{F}' := \mathcal{F}_0 \setminus \mathcal{F}$.

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 “redundant”, 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 $\forall \mathbf{u}, \mathbf{v} \in PF_0 : \mathbf{u}^{(\mathcal{F}')} \preceq \mathbf{v}^{(\mathcal{F}')} \Rightarrow \mathbf{u}^{(\mathcal{F})} \preceq \mathbf{v}^{(\mathcal{F})}$.

Theorem 3: Given two nonempty objective subsets $\mathcal{F}, \mathcal{F}' \subset \mathcal{F}_0$ and $\mathcal{F} \cap \mathcal{F}' = \emptyset$, if $\forall \mathbf{u}, \mathbf{v} \in PF_0 : \mathbf{u}^{(\mathcal{F}')} \preceq \mathbf{v}^{(\mathcal{F}')} \Rightarrow \mathbf{u}^{(\mathcal{F})} \preceq \mathbf{v}^{(\mathcal{F})}$, then \mathcal{F} is redundant.

Based on Theorem 3, we have the following corollary, which establishes the relationship between the concepts “non-conflicting” 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 $\mathcal{N} := PF_0$ and $|PF_0| < +\infty$, PF_η denotes the Pareto front of η -OR problem. Then $\exists k^* \in \mathbb{N} : (k^*, 0)^T \in PF_\eta$ and any objective subset corresponding to $(k^*, 0)^T$ 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 with sufficient computation time. Take η -OR for instance, Theorem 4 indicates that, under the most ideal situation, i.e., $\mathcal{N} := PF_0$, it is ensured that PF_η includes the vector $(k^*, 0)^T$ that corresponds to an essential objective set.

However, it is often expecting too much that \mathcal{N} is a perfect 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's have a closer look at the criteria δ and η . For the criterion δ , suppose the following $\mathcal{N} = \{\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{2n-1}, \mathbf{u}_{2n}\}$, where

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

In this case, \mathbf{u}_{2n} could resemble misdirection since it is an outlier. And 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 $\{f_1, f_2\}$ unable to become an essential objective set. As for the criterion η , we take 3-objective WFG3 as an example again. Fig. 4 shows the projections of a sample set \mathcal{N} (including \mathcal{N}_{NS} and \mathcal{N}_{DS}) on f_2 - f_3 objective space. In this case, $\eta = 0.52$ for $\{f_2, f_3\}$, which is a significant proportion. It means that it is not safe to interpret the solutions in \mathcal{N}_{DS} as misdirection according to the large η error and $\{f_2, f_3\}$ would not be identified as the essential objective set, which does not conform well to the intuition for Fig. 4. The problem lies in that, although \mathcal{N}_{DS} accounts for a significant proportion of \mathcal{N} , most of its objective vectors are very close to those in \mathcal{N}_{NS} and can be indeed interpreted as the misdirection with high possibility. In summary, a general major limitation of dominance structure-based approaches is that they may have difficulty in handling varying degree of misdirection in the sample set effectively.

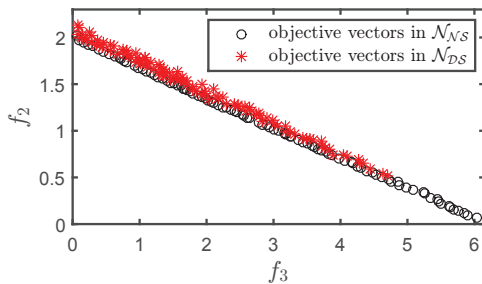


Fig. 4. The projection of \mathcal{N} (including \mathcal{N}_{NS} and \mathcal{N}_{DS}) on f_2 - f_3 objective space for 3-objective WFG3.

Another limitation of 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 \mathcal{N} is not so large and diverse that its distribution restricts to only a part of PF_0 , resulting that the solutions in \mathcal{N} are also mutually nondominated in a lower dimensional (e.g., 10-dimensional) objective subspace that 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 since $|\mathcal{F}| < 15$, 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., $|\mathcal{N}| \gg m$, 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 with dominance structure-based approaches. This is mainly because correlation only pays attention to 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

$$r_\tau = \begin{pmatrix} 1 & 0.901 & -0.943 \\ 0.901 & 1 & -0.952 \\ -0.943 & -0.952 & 1 \end{pmatrix} \quad (8)$$

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 3-objective WFG3. Recall that it is very difficult to identify an essential objective set in this case by using η criterion, 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 also can find out what objectives are conflicting and what are nonconflicting, which would be helpful 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 very 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,

one objective is redundant can be due to its relationship with an objective set more than just a single objective, whereas conflicting 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 2-objective DTLZ2 [55] 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

$$r_\tau = \begin{pmatrix} 1 & -1 & -0.016 \\ -1 & 1 & 0.016 \\ -0.016 & 0.016 & 1 \end{pmatrix} \quad (9)$$

From Eq. (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.

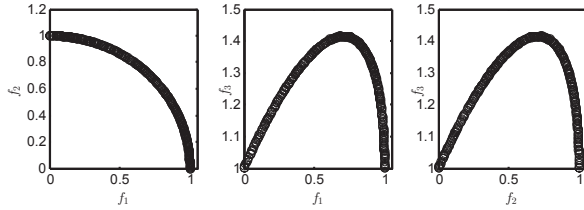


Fig. 5. The projections of a sample set for the modified DTLZ2 problem on the objective subspaces f_1 - f_2 , f_1 - f_3 and f_2 - f_3 .

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 the accurate result can still be obtained, but meanwhile the strength in handling the misdirection would be sacrificed to some extent.

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

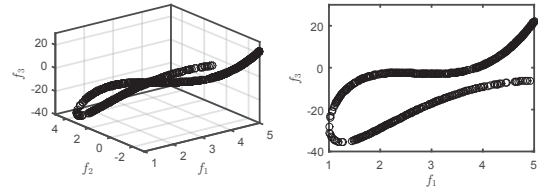


Fig. 6. A high-quality sample set obtained for the spline problem: $f_1 = x^2 + 1$, $f_2 = -x^2 + x + 3$ and $f_3 = -(f_1 + f_2^3)$, where $x \in [-2, 2]$, and its projection on the f_1 - f_3 objective subspace.

V. BENCHMARK EXPERIMENTS

In this section, we first describe several experimental design issues 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], [38], WFG3(m) [44], [45], and DTLZ2(m) [55], 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) [55], 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 $\mathcal{F}_T = \{f_l, f_{m-I+2}, \dots, f_m\}$, where $l \in \{1, 2, \dots, m - I + 1\}$.

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 $l \in \{1, 2, \dots, m - 1\}$.

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

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 totally has $2m$ objectives. 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 $\mathcal{F}_T = \{f_{l_1}, f_{l_2}, \dots, f_{l_m}\}$, 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 $\{f_1, f_2, \dots, f_m\}$ 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 $\mathcal{F}_T = \{f_1, f_2, \dots, f_m\}$.

Except for WFG3(m), the number of decision variables for the other problems depends on a parameter p , i.e., $n = m + p - 1$. We set p to 10 for these problems as recommended in [55]. 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 a SPEA2 variant referred as SPEA2-SDE [19] with the same number of generations for 30 independent repetitions. \mathcal{N}_2 requires more computational effort (within 2,000 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 of g to 0 and the closer of I_s to 1 mean 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 made nearly one magnitude less than that of \mathcal{N}_1 , so that they have clear difference in quality.

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 Java² and run on an Intel 3.20 GHz Xeon processor with 16.0 GB of RAM.

ExactAlg is not suited for large-scale instances, so some of its results are not available 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, the exhaustive search would be enough for small instances, e.g., $m \leq 10$, but

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

for the uniformity, we just solve all the instances using the evolutionary multiobjective search.

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	$[2.5m]$
Crossover probability	0.9
Mutation probability	$1/m$

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].

D. Investigation on the Behavior of Multiobjective Approaches

The multiobjective approaches return a number of non-dominated 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 trade-off 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:

²The source code of the proposed objective reduction algorithms is available on line: <https://github.com/yyxhdy/MOOR/archive/master.zip>.

TABLE IV
THE 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	$\{f_9, f_{13}, f_{16}, f_{17}, f_{19}, f_{20}\}$

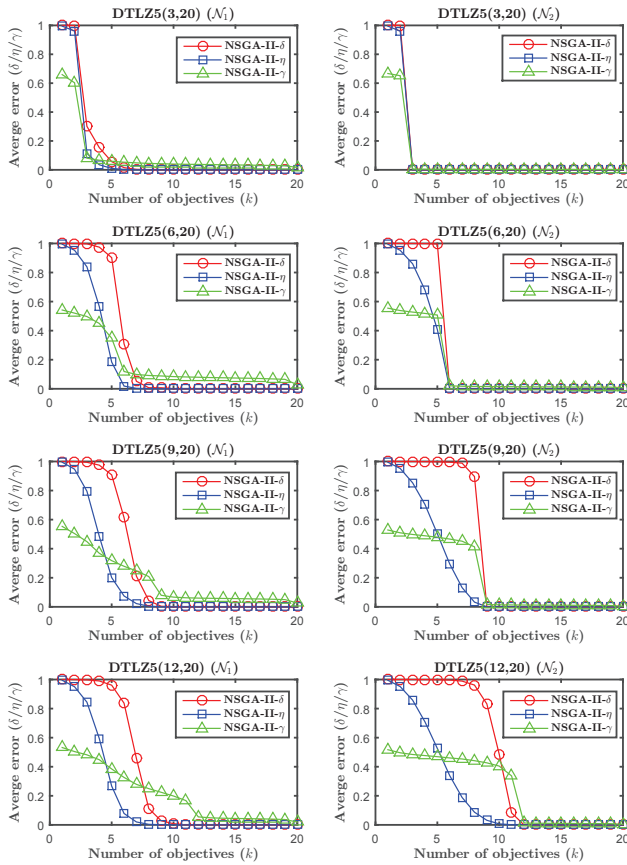


Fig. 7. Results of NSGA-II- δ , NSGA-II- η and NSGA-II- γ on DTLZ5(3,20), DTLZ5(6,20), DTLZ5(9,20) and DTLZ5(12,20), corresponding to \mathcal{N}_1 and \mathcal{N}_2 . The error for each point is averaged over 30 sample sets in \mathcal{N}_1 or \mathcal{N}_2 .

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.

TABLE V
THE 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 , 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 subsection, 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

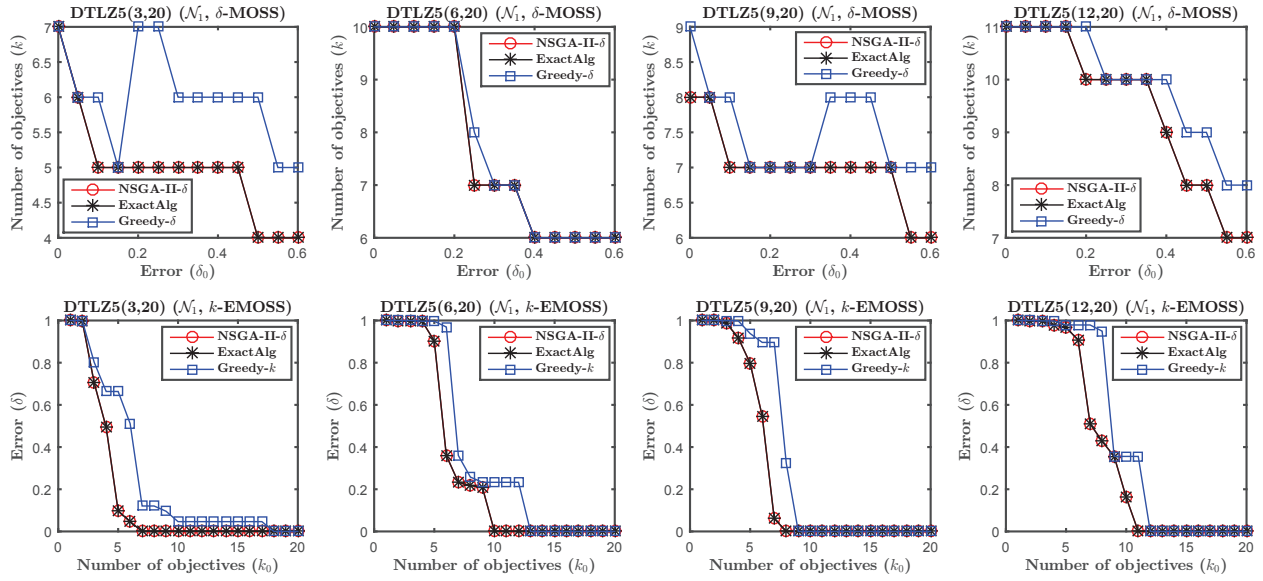


Fig. 8. Comparison of NSGA-II- δ , ExactAlg, and Greedy- δ (Greedy- k) for the δ -MOSS or k -EMOSS problem.

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 from the side.

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 spends 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 hours on the sample set of DTLZ5(12,20). Compared with Greedy- k , NSGA-II- δ also uses relatively short 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 the 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.

F. Comparison in Identifying the Essential Objective Set

In this subsection, 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 multi-objective approaches can return a number of nondominated objective subsets which contributes to the 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 possible one. But for the 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 exponential 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 certain 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 of 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 non-conflicting 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

TABLE VII
THE FREQUENCY OF SUCCESS IN IDENTIFYING THE ESSENTIAL OBJECTIVE SET \mathcal{F}_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	NSGA	NSGA	Exact	Greedy	L	NL-MVU	NSGA	NSGA	NSGA	Exact	Greedy	L	NL-MVU
	-II- δ	-II- η	-II- γ	Alg	- δ	-PCA	-PCA	-II- δ	-II- η	-II- γ	Alg	- δ	-PCA	-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.

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 ap-

proaches, 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; 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 $|\mathcal{F}_T|$ is relatively low ($|\mathcal{F}_T| \leq 6$), 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

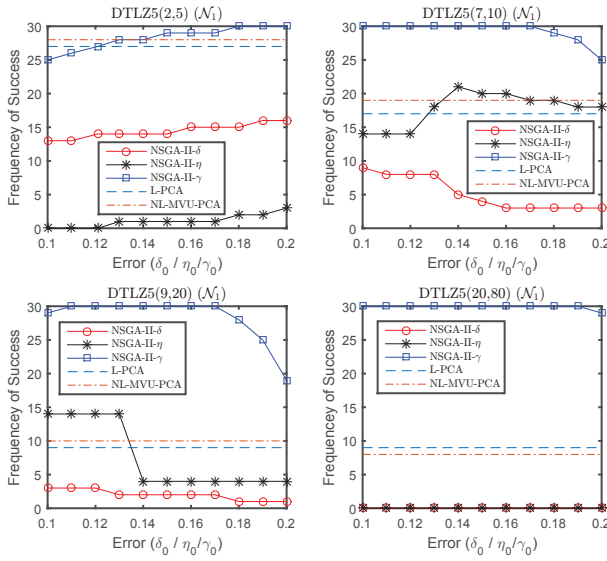


Fig. 9. Illustration of the influence of $\delta_0/\eta_0/\gamma_0$ on the performance of the proposed multiobjective approaches in identifying the essential objective set. The performance of L-PCA and NL-MVU-PCA is provided as a reference.

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 $|\mathcal{F}_T|$ objectives on the average, which can be attributed to their general limitation in handling the 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.

TABLE VIII

THE AVERAGE NUMBER OF OBJECTIVES IDENTIFIED PER RUN BY NSGA-II- δ , NSGA-II- η AND GREEDY- δ ON THE TEST INSTANCES WITH THE DIMENSIONALITY $|\mathcal{F}_T| \leq 6$, 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

2) When the dimensionality $|\mathcal{F}_T|$ is moderately high ($7 \leq |\mathcal{F}_T| \leq 10$), they would confront both the 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 difficul-

ties. And as observed from Table IX, the number of objectives can be either larger or smaller than $|\mathcal{F}_T|$, 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 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 alleviate both the two difficulties with these algorithms effectively.

TABLE IX

THE AVERAGE NUMBER OF OBJECTIVES IDENTIFIED PER RUN BY NSGA-II- δ , NSGA-II- η AND GREEDY- δ ON THE TEST INSTANCES WITH THE DIMENSIONALITY $7 \leq |\mathcal{F}_T| \leq 10$, 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

3) As the dimensionality $|\mathcal{F}_T|$ gets higher ($|\mathcal{F}_T| \geq 12$), 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., $|\mathcal{F}_T| \geq 12$) 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 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

TABLE X
THE AVERAGE NUMBER OF OBJECTIVES IDENTIFIED PER RUN
BY NSGA-II- δ , NSGA-II- η AND GREEDY- δ ON THE TEST
INSTANCES WITH THE DIMENSIONALITY $|\mathcal{F}_T| \geq 12$,
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.

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 $k_1 < |\mathcal{F}_T|$ 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 $|\mathcal{F}_T|$ 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 overall better than that of Greedy- δ , which is particularly clear on the SUM-DTLZ2(m) instances.

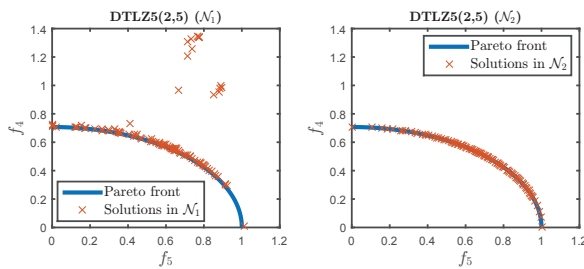


Fig. 10. DTLZ5(2,5): a randomly selected sample set in \mathcal{N}_1 and \mathcal{N}_2 respectively, corresponding to $\mathcal{F}_T = \{f_4, f_5\}$.

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 $i, j \in [1, 15]$.

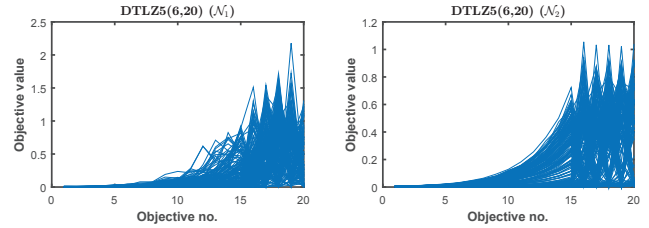


Fig. 11. DTLZ5(6,20): the parallel coordinate plots for a randomly selected sample set in \mathcal{N}_1 and \mathcal{N}_2 respectively.

2) L-PCA and NL-MVU-PCA perform quite well on DTLZ5(I, m) with a relatively low dimensionality and all WFG3(m) instances, e.g., DTLZ5(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 such ability deteriorates seriously for the redundant problems with a 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 the 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 performs overall 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 generally 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-MVU-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 on 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 concerned SUM-DTLZ2(m) instances perfectly.

2) Except on SUM-DTLZ2(m), NSGA-II- γ performs better than, or as good as four dominance structure-based algorithms corresponding to both \mathcal{N}_1 and \mathcal{N}_2 , owing to its great power in handling the 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 the problem with the 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, the dominance structure-based approaches would be 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 [56] and car side-impact problem [50], [57].

A. Water Resource Problem

The water resource problem consists of 5 objectives and 7 constraints, which relates to optimal planning for a storm drainage system [56]. As a basis for the analysis, a sample set is first produced by running SPEA2-SDE with a population size of 200 and 2,000 generations.

In Table XI, we show the detail 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

TABLE XI
THE 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	$\{f_1, f_2, f_3, f_4, f_5\}$
	5	0	$\{f_1, f_2, f_3, f_4, f_5\}$

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

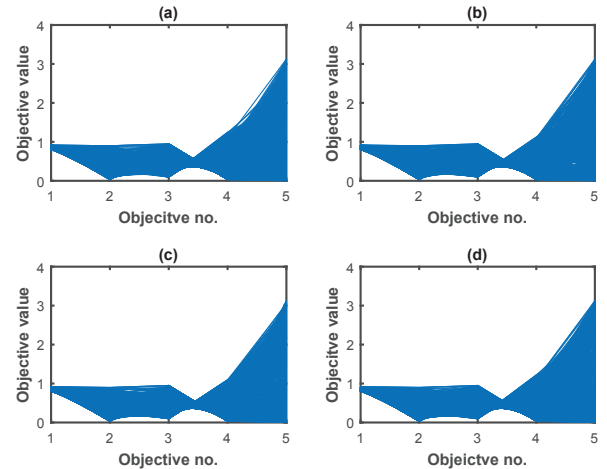


Fig. 12. Water resource problem: parallel coordinate plots for the nondominated solutions obtained by running SPEA2-SDE. (a) On the original objective set. (b) On the reduced objective set $\{f_2, f_3, f_5\}$. (c) On the reduced objective set $\{f_1, f_2, f_4, f_5\}$. (d) On the reduced objective set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

B. Car Side-Impact Problem

The car side-impact problem is transformed into a many-objective problem with 11 objectives and 10 constraints [50]. The sample set is still generated by SPEA2-SDE with a population size of 200 and 2,000 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 trade-off for NSGA-II- δ is obviously achieved at $k = 6$, whereas the best trade-off 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 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 non-dominated 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.

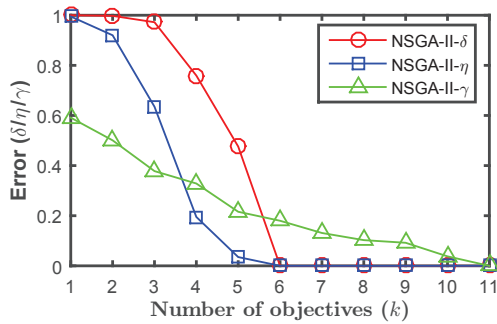


Fig. 13. The results of NSGA-II- δ , NSGA-II- η and NSGA-II- γ for the car side-impact problem.

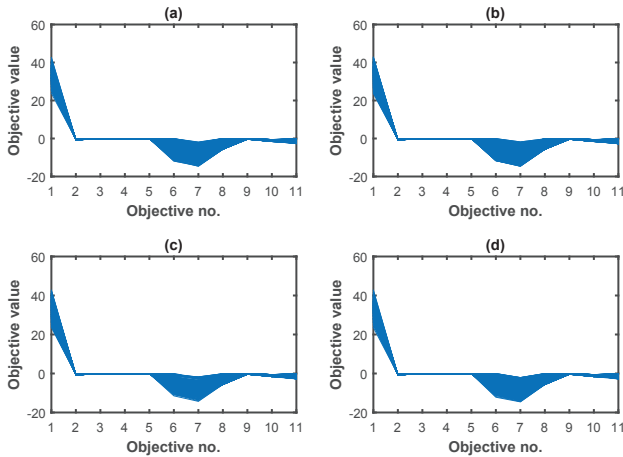


Fig. 14. Car side-impact problem: parallel coordinate plots for the non-dominated solutions obtained by running SPEA2-SDE. (a) On the original objective set. (b) On the reduced objective set $\{f_1, f_4, f_5, f_9, f_{10}, f_{11}\}$. (c) On the reduced objective set $\{f_1, f_4, f_5, f_8, f_9, f_{11}\}$.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

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 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. The 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.

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 $\{f_1, f_2, f_4, f_5\}$
[37]	$\{f_1, f_2, f_5\}$ or $\{f_2, f_3, f_5\}$
[47]	$\{f_3, f_4\}$
[40], [50]	$\{f_2, f_3, f_4, f_5\}$

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], [37], [47]. 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 good 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. Fig. 12 and Fig. 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 illustrating examples.

A. On Objective Reduction Assisted Optimization

Generally, the objective reduction approaches can assist optimization in two manners. 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, a MOEA continues to run on the reduced set for a number of generations and returns the nondominated solutions obtained. The second one is more commonly used, which is so called online objective reduction [48], [49], i.e., the objective reduction algorithm is embedded into the iterations of the MOEA. Now, we use both of the two manners 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 SDE_{online} . To show the effectiveness, SDE_{so} and SDE_{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 non-uniformly distributed Pareto fronts. For a fair comparison, SDE_{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.

TABLE XIII

THE 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}	SDE_{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

First, we use a benchmark problem, i.e., 15-objective WFG3, for illustration purposes. The inverted generational distance (IGD) [59] is used as the performance metric, where the reference set is produced by uniformly sampling 1,000 points on the true Pareto front. A-NSGA-III initially uses two-layered reference vectors with 2 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 SDE_{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 SDE_{online} promotes the convergence of the many-objective optimizer.

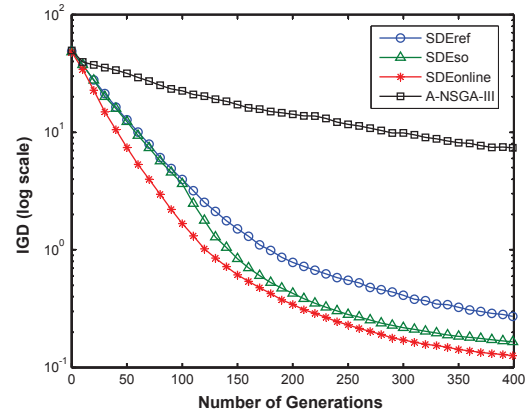


Fig. 15. Convergence curves of the average IGD (over 30 runs) obtained with the number of generations for SDE_{ref} , SDE_{so} , SDE_{online} and A-NSGA-III on 15-objective WFG3.

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

TABLE XIV

THE 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}	SDE_{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

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 a hypervolume-based MOEA or the objective computations of the MaOP are expensive. The former scenario can be referred

to the study by Brockhoff and Zitzler [48], while the latter can be seen in a recent work by Carreras *et al.* [60].

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 (e.g., [38], [51], [61]), 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], [62]–[64] 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 visualisation methods in the literature, e.g., parallel coordinate plots, heatmaps, bubble chart. The interested reader is referred to [32] or [33] 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 an 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 an 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 know that the true dimensionality of DTLZ5(6,20) is much lower than 20 by NSGA-II- γ , he/she will 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). Obviously, Fig. 16(b) is much easier than Fig. 16(a) for the user to comprehend and reveal good trade-offs.

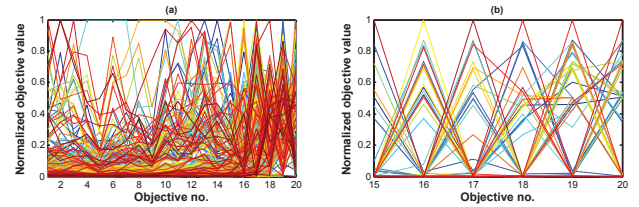


Fig. 16. Comparison of the parallel coordinate plots obtained before and after objective reduction.

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 AND FUTURE WORK

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: one is the number of objectives selected (k) and

the other is the error (δ , η or γ) incurred by removing the unselected objectives, and 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 trade-off among the two objectives and return a set of nondominated objective subsets that can offer the 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 disclose the characteristics of the compared algorithms which match well with our critical analysis. The possible directions for the future research are suggested as follows:

1) Since dominance structure- and correlation-based approaches have complementary advantages, it would be possible to develop the hybrid approaches that can combine the merits of both and meanwhile overcome their disadvantages to a certain degree.

2) It is necessary to further investigate on-line objective reduction based on the proposed algorithms. On one hand, more advanced integration schemes need to be developed. On the other hand, the benefits of on-line objective reduction are worthy to be further examined by comparing with those of decomposition-based algorithms with reference vector adaptation [21], [62]–[64] more extensively.

3) There are only limited types of benchmark problems available in the literature for objective reduction. It is very desirable to design more benchmark problems with new characteristics so that the objective reduction algorithms can be evaluated more comprehensively.

REFERENCES

- [1] D. Brockhoff and E. Zitzler, "Objective reduction in evolutionary multiobjective optimization: theory and applications," *Evolutionary Computation*, 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 Transactions on Evolutionary Computation*, vol. 17, no. 1, pp. 77–99, 2013.
- [3] P. J. Fleming, R. C. Purshouse, and R. J. Lygoe, "Many-objective optimization: An engineering design perspective," in *Proceedings of Evolutionary Multi-Criterion Optimization*, 2005, pp. 14–32.
- [4] E. J. Hughes, "Radar waveform optimisation as a many-objective application benchmark," in *Proceedings of Evolutionary Multi-Criterion Optimization*, 2007, pp. 700–714.
- [5] M. Asafuddoula, H. K. Singh, and T. Ray, "Six-sigma robust design optimization using a many-objective decomposition-based evolutionary algorithm," *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 4, pp. 490–507, 2015.
- [6] W. Mkaouer, M. Kessentini, A. Shaout, P. Koligheue, S. Bechikh, K. Deb, and A. Ouni, "Many-objective software modularization using NSGA-III," *ACM Transactions on Software Engineering and Methodology*, vol. 24, no. 3, p. 17, 2015.
- [7] T. Wagner, N. Beume, and B. Naujoks, "Pareto-, aggregation-, and indicator-based methods in many-objective optimization," in *Proceedings of Evolutionary Multi-Criterion Optimization*, 2007, pp. 742–756.
- [8] H. Ishibuchi, N. Tsukamoto, and Y. Nojima, "Evolutionary many-objective optimization: A short review," in *Proceedings of IEEE Congress on Evolutionary Computation*, 2008, pp. 2419–2426.
- [9] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.
- [10] E. Zitzler, M. Laumanns, L. Thiele, E. Zitzler, E. Zitzler, L. Thiele, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm," *Proceedings of Evolutionary Methods for Design, Optimisation and Control*, 2002, pp. 95–100.
- [11] E. Zitzler and S. Künzli, "Indicator-based selection in multiobjective search," in *Proceedings of International Conference on Parallel Problem Solving from Nature*, 2004, pp. 832–842.
- [12] N. Beume, B. Naujoks, and M. Emmerich, "SMS-EMOA: Multiobjective selection based on dominated hypervolume," *European Journal of Operational Research*, vol. 181, no. 3, pp. 1653–1669, 2007.
- [13] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712–731, 2007.
- [14] E. J. Hughes, "Multiple single objective Pareto sampling," in *Proceedings of IEEE Congress on Evolutionary Computation*, 2003, pp. 2678–2684.
- [15] S. Yang, M. Li, X. Liu, and J. Zheng, "A grid-based evolutionary algorithm for many-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 5, pp. 721–736, 2013.
- [16] Z. He, G. G. Yen, and J. Zhang, "Fuzzy-based Pareto optimality for many-objective evolutionary algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 2, pp. 269–285, 2014.
- [17] Y. Yuan, H. Xu, B. Wang, and X. Yao, "A new dominance relation based evolutionary algorithm for many-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 1, pp. 16–37, 2015.
- [18] S. F. Adra and P. J. Fleming, "Diversity management in evolutionary many-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 2, pp. 183–195, 2011.
- [19] M. Li, S. Yang, and X. Liu, "Shift-based density estimation for pareto-based algorithms in many-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 3, pp. 348–365, 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 Transactions on Evolutionary Computation*, vol. 18, no. 4, pp. 577–601, 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 Transactions on Evolutionary Computation*, vol. 18, no. 4, pp. 602–622, 2014.
- [22] H. Ishibuchi, N. Tsukamoto, Y. Sakane, and Y. Nojima, "Indicator-based evolutionary algorithm with hypervolume approximation by achievement scalarizing functions," in *Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation*, 2010, pp. 527–534.
- [23] J. Bader and E. Zitzler, "HypE: An algorithm for fast hypervolume-based many-objective optimization," *Evolutionary computation*, vol. 19, no. 1, pp. 45–76, 2011.
- [24] C. A. Rodríguez Villalobos and C. A. Coello Coello, "A new multi-objective evolutionary algorithm based on a performance assessment indicator," in *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation*, 2012, pp. 505–512.
- [25] R. Hernandez Gomez and C. Coello Coello, "MOMBI: A new meta-heuristic for many-objective optimization based on the R2 indicator," in *Proceedings of IEEE Congress on Evolutionary Computation*, 2013, pp. 2488–2495.
- [26] M. Asafuddoula, T. Ray, and R. Sarker, "A decomposition-based evolutionary algorithm for many objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 3, pp. 445–460, 2015.
- [27] Y. Yuan, H. Xu, B. Wang, B. Zhang, and X. Yao, "Balancing convergence and diversity in decomposition-based many-objective optimizers," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 2, pp. 180–198, 2016.
- [28] Y. Tan, Y. Jiao, H. Li, and X. Wang, "MOEA/D+ uniform design: A new version of MOEA/D for optimization problems with many objectives," *Computers & Operations Research*, vol. 40, no. 6, pp. 1648–1660, 2013.
- [29] I. Giagkiozis, R. C. Purshouse, and P. J. Fleming, "Generalized decomposition and cross entropy methods for many-objective optimization," *Information Sciences*, vol. 282, pp. 363–387, 2014.
- [30] A. López Jaimes and C. A. Coello Coello, "Some techniques to deal with many-objective problems," in *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation*, 2009, pp. 2693–2696.
- [31] A. López Jaimes, C. A. Coello Coello, and D. Chakraborty, "Objective reduction using a feature selection technique," in *Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation*, 2008, pp. 673–680.

- [32] D. J. Walker, R. Everson, and J. E. Fieldsend, "Visualizing mutually nondominating solution sets in many-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 2, pp. 165–184, 2013.
- [33] T. Tušar and B. Filipič, "Visualization of Pareto front approximations in evolutionary multiobjective optimization: A critical review and the projection method," *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 2, pp. 225–245, 2015.
- [34] M. Köppen and K. Yoshida, "Visualization of Pareto-sets in evolutionary multi-objective optimization," in *Proceedings of 7th International Conference on Hybrid Intelligent Systems*, 2007, pp. 156–161.
- [35] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.
- [36] R. C. Purshouse and P. J. Fleming, "Conflict, harmony, and independence: Relationships in evolutionary multi-criterion optimisation," in *Proceedings of Evolutionary Multi-Criterion Optimization*, 2003, pp. 16–30.
- [37] H. K. Singh, A. Isaacs, and T. Ray, "A Pareto corner search evolutionary algorithm and dimensionality reduction in many-objective optimization problems," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 4, pp. 539–556, 2011.
- [38] K. Deb and D. Saxena, "Searching for Pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems," in *Proceedings of IEEE Congress on Evolutionary Computation*, 2006, pp. 3352–3360.
- [39] B. Xue, M. Zhang, W. Browne, and X. Yao, "A survey on evolutionary computation approaches to feature selection," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 4, pp. 606–626, 2016.
- [40] X. Guo, Y. Wang, and X. Wang, "Using objective clustering for solving many-objective optimization problems," *Mathematical Problems in Engineering*, vol. 2013, 2013.
- [41] H. Wang and X. Yao, "Objective reduction based on nonlinear correlation information entropy," *Soft Computing*, vol. 20, no. 6, pp. 2393–2407, 2016.
- [42] 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, 2014.
- [43] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler, "Combining convergence and diversity in evolutionary multiobjective optimization," *Evolutionary Computation*, vol. 10, no. 3, pp. 263–282, 2002.
- [44] S. Huband, P. Hingston, L. Barone, and L. While, "A review of multiobjective test problems and a scalable test problem toolkit," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 5, pp. 477–506, 2006.
- [45] H. Ishibuchi, H. Masuda, and Y. Nojima, "Pareto fronts of many-objective degenerate test problems," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 5, pp. 807–813, 2016.
- [46] A. R. de Freitas, P. J. Fleming, and F. G. Guimarães, "Aggregation trees for visualization and dimension reduction in many-objective optimization," *Information Sciences*, vol. 298, pp. 288–314, 2015.
- [47] X. Guo, Y. Wang, and X. Wang, "An objective reduction algorithm using representative Pareto solution search for many-objective optimization problems," *Soft Computing*, vol. 20, no. 12, pp. 4881–4895, 2016.
- [48] D. Brockhoff and E. Zitzler, "Improving hypervolume-based multiobjective evolutionary algorithms by using objective reduction methods," in *Proceedings of IEEE Congress on Evolutionary Computation*, 2007, pp. 2086–2093.
- [49] A. L. Jaimes, C. A. C. Coello, and J. E. U. Barrientos, "Online objective reduction to deal with many-objective problems," in *Proceedings of Evolutionary Multi-Criterion Optimization*, 2009, pp. 423–437.
- [50] A. Sinha, D. K. Saxena, K. Deb, and A. Tiwari, "Using objective reduction and interactive procedure to handle many-objective optimization problems," *Applied Soft Computing*, vol. 13, no. 1, pp. 415–427, 2013.
- [51] S. Bandyopadhyay and A. Mukherjee, "An algorithm for many-objective optimization with reduced objective computations: A study in differential evolution," *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 3, pp. 400–413, 2015.
- [52] 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," *Applied Soft Computing*, vol. 35, pp. 75–112, 2015.
- [53] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1/2, pp. 81–93, 1938.
- [54] D. E. Goldberg, "Genetic algorithms in search, optimization and machine learning," 1989.
- [55] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable multi-objective optimization test problems," in *Proceedings of IEEE Congress on Evolutionary Computation*, 2002, pp. 825–830.
- [56] K. Musselman and J. Talavage, "A tradeoff cut approach to multiple objective optimization," *Operations Research*, vol. 28, no. 6, pp. 1424–1435, 1980.
- [57] L. Gu, R. Yang, C. Cho, M. Makowskit, M. Faruquet, and Y. Li, "Optimisation and robustness for crashworthiness of side impact," *International Journal of Vehicle Design*, vol. 26, no. 4, pp. 348–360, 2001.
- [58] T. Ray, K. Tai, and K. C. Seow, "Multiobjective design optimization by an evolutionary algorithm," *Engineering Optimization*, vol. 33, no. 4, pp. 399–424, 2001.
- [59] D. A. Van Veldhuizen and G. B. Lamont, "Multiobjective evolutionary algorithm research: A history and analysis," Citeseer, Tech. Rep., 1998.
- [60] J. Carreras, C. Pozo, D. Boer, G. Guillén-Gosálbez, J. A. Caballero, R. Ruiz-Femenia, and L. Jiménez, "Systematic approach for the life cycle multi-objective optimization of buildings combining objective reduction and surrogate modeling," *Energy and Buildings*, vol. 130, pp. 506–518, 2016.
- [61] 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," *Computers & Chemical Engineering*, vol. 87, pp. 36–48, 2016.
- [62] S. Jiang, Z. Cai, J. Zhang and Y. S. Ong, "Multiobjective optimization by decomposition with Pareto-adaptive weight vectors," in *Proceedings of 7th International Conference on Natural Computation*, 2011, pp. 1260–1264.
- [63] Y. Qi, X. Ma, F. Liu, L. Jiao, J. Sun, and J. Wu, "MOEA/D with adaptive weight adjustment," *Evolutionary Computation*, vol. 22, no. 2, pp. 231–264, 2014.
- [64] R. Cheng, Y. Jin, M. Olhofer, and B. Sendhoff, "A reference vector guided evolutionary algorithm for many-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 5, pp. 773–791, 2016.