Objective Reduction in Many-objective Optimization: Linear and Nonlinear Algorithms

Dhish Kumar Saxena, João A. Duro, Ashutosh Tiwari, Kalyanmoy Deb and Qingfu Zhang
KanGAL Report No. 2010008

***Abstract*—The difficulties faced by existing Multi-objective Evolutionary Algorithms (MOEAs) in handling many-objective problems relate to the inefficiency of selection operators, high computational cost and difficulty in visualization of objective space. While many approaches aim to counter these difficulties by increasing the fidelity of the standard selection operators, the objective reduction approach attempts to eliminate objectives that are not essential to describe the Pareto-optimal Front (POF). If the number of essential objectives are found to be two or three, the problem could be solved by the existing MOEAs. It implies that objective reduction could make an otherwise unsolvable (many-objective) problem solvable. Even when the essential objectives are four or more, the reduced representation of the problem will have favorable impact on the search efficiency, computational cost and decision-making. Hence, development of generic and robust objective reduction approaches becomes important. This paper presents a Principal Component Analysis and Maximum Variance Unfolding based framework for linear and nonlinear objective reduction algorithms, respectively. The major contribution of this paper includes: (a) the enhancements in the core components of the framework for higher robustness in terms of–applicability to a range of problems with disparate degree of redundancy; mechanisms to handle an input data that is mis-representative of the true POF; and dependence on fewer parameters to minimize the variability in performance, (b) proposition of an error measure to assess the quality of results, (c) sensitivity analysis of the proposed algorithms for the parameters involved and on the quality of the input data, and (d) study of the performance of the proposed algorithms vis-$\grave{a}$-vis dominance relation preservation based algorithms, on a wide range of test problems (scaled up to 50 objectives) and two real-world problems.**

Keywords: Evolutionary Multi-objective Optimization, Many-objective Optimization, Principal Component Analysis, Maximum Variance Unfolding and Kernels.

I. Introduction

Optimization problems involving four or more conflicting objectives are typically referred to as *many-objective* problems. With the multiple conflicting demands faced by the industry today for superior quality, low cost and higher safety etc., competitive edge could only be established by designing the products and processes that account for as many performance objectives as possible. It implies that many objectives need to be simultaneously dealt with, in an optimization problem. While this recognition is growing, it is alarming that the search ability of some of the most well known MOEAs severely deteriorates when more than three objectives are involved [1]–[3]. This makes evolutionary many-objective optimization one of the most challenging research areas in the field of evolutionary optimization and explains the growing research emphasis in this direction. The main difficulties associated with many-objective problems relate to:

1) High computational cost: If a continuous multi-objective optimization problem (with M objectives) meets the regularity property, the dimensionality of its Pareto-optimal front (POF) can be $M-1$ [4]. Therefore, the number of points needed for approximating the whole POF increase exponentially with M. The same phenomenon can also be observed in discrete problems.

2) Poor scalability of most existing MOEAs: With an increase in M, almost the entire population acquires the same-rank of non-domination. This makes the Pareto-dominance based primary selection ineffective and the role of secondary selection based on diversity preservation becomes crucial. The density based MOEAs (such as NSGA-II [5], SPEA2 [6] and PESA [7]) worsen the situation by favoring the remote and boundary solutions, implying that the best diversity gets associated with poorly converged solutions. This explains the performance deterioration reported in [1]–[3]. The more recent MOEA/D [8], [9] and indicator based MOEAs [10], [11] are found to cope better with an increase in M. For example, in SMS-EMOA [11], the problems faced by the density based MOEAs are countered by the use of $\mathcal{S}$-metric [12] based secondary selection. However, the utility of such MOEAs is limited by the fact that their running times increase exponentially with M [13]–[15].

3) Difficulty in visualization of a POF for problems with $M \geq 4$, which in turn makes the task of decision making more difficult. Although techniques such as decision maps [16] and self-organizing maps [17] exist to aid in visualization, they require a large number of solutions.

In the context of many-objective problems, the above difficulties are often referred to as the *curse of dimensionality*. Here, it is important to discriminate between the curse of dimensionality caused by: (i) the difficulties inherent in the problems, e.g., problems with a high-dimensional Pareto-set with complicated shapes [18], and (ii) the poor scalability of most of the existing MOEAs for $M \geq 4$, even when the corresponding Pareto-set has simple shapes. For instance, a many-objective problem whose Pareto-set is 1- or 2-dimensional with a linear or a quadratic shape, represents a case where the

D. K. Saxena, J. A. Duro and A. Tiwari are with the Manufacturing Department, Cranfield University, Cranfield, U.K (email:{d.saxena, j.a.duro, a.tiwari}@cranfield.ac.uk).

K. Deb is with the Department of Mechanical Engineering, Indian Institute of Technology, Kanpur, India (email: deb@iitk.ac.in)

Q. Zhang is with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, U.K (email: qzhang@essex.ac.uk).

problem in itself does not have the curse of dimensionality, yet most of the existing MOEAs may suffer from it.

The approaches for dealing with many-objective problems can be broadly categorized as:

1) *Preference-ordering* approaches: These approaches assume that no objective in the given problem is redundant and aim to counter the low selection pressure for convergence by inducing a preference ordering over the non-dominated solutions. Some of these approaches include: reducing the number of non-dominated points [19]; assigning different ranks to non-dominated points [20]–[23]; using scalarizing functions for fitness evaluation [2], [24]; using indicator functions [10], [11], [25]; or using decision maker's preference information [26]–[29]. While more details can be found in [30], it may be noted that most of these approaches either aim for only a part of the POF, report an improvement in convergence with the loss in diversity, or their computational time increases exponentially with the number of objectives.

2) *Objective reduction* approaches: These approaches [31]–[35] assume the existence of redundant objectives in a given M-objective problem. Operating on the objective vectors of the non-dominated solutions obtained from an MOEA, these approaches aim to identify a smallest set of m ($m \leq M$) conflicting objectives which: (i) generates the same POF as the original problem, or (ii) alternatively, preserves the dominance relations of the original problem. Such m objectives are termed as *essential* and the remaining ones as *redundant*. If $m \leq 3$, an otherwise unsolvable problem will become solvable. Even if $4 \leq m \leq M$, objective reduction will contribute to higher search efficiency, lower computational cost and ease in visualization and decision-making.

This paper builds upon [34], [35] and presents a framework for both linear and nonlinear objective reduction algorithms, namely, L-PCA and NL-MVU-PCA. The distinctive contribution of this paper relates to:

1) The four goals that the framework pursues:
   - Generality: The scope of [34], [35] was limited to highly redundant problems, where $m \ll M$. The scope of the proposed framework is broadened to include problems with moderate ($m < M$) and negligible ($m \approx M$) redundancy. This is significant because the core components of the framework need to adapt to conflicting goals while handling the above cases with different degree of redundancy.
   - De-noising of the input data: Unlike [34], [35], the proposed framework accounts for the fact that the input data, namely, the non-dominated solutions obtained from an MOEA is mis-representative of the true POF (noisy). Towards it, an eigenvalue based dynamic interpretation of the strength of correlation between objective vectors of the non-dominated solutions is proposed, which serves to negate the effect of noise in the data.
   - Parameter reduction: The number of parameters involved in the proposed framework have been minimized as much as possible, so that the variability in the results corresponding to different parameter settings is minimized. This is a significant shift in approach from [34], [35], where a lot of parameters were involved, making the algorithms less robust.
   - Proposition of an error measure: An error measure has been proposed (not in [34], [35]) that allows an assessment of the quality of the obtained results.
2) Extensive simulations and results: The results presented in this paper are new and are based on 12800 simulations[1] performed on 30 versions of 7 test problems and two real-world problems. In this, the sensitivity of the proposed algorithms on: (i) the randomness, (ii) the parameters involved, and (iii) the size and the quality of the underlying non-dominated sets, is discussed.
3) Performance comparison: Here, the performance of the proposed algorithms vis-à-vis dominance relation preservation [31] based algorithms is studied, with reference to their general strengths and limitations. It is established that while the scope of the latter is limited to *linear* objective reduction, the proposed NL-MVU-PCA efficiently performs *nonlinear* objective reduction. Besides this, the misinterpretation by peer researchers of the previous versions of this work, in terms of the computational complexity, is clarified.

This paper is organized as follows: The fundamental issues in objective reduction and the existing approaches are discussed in Section II. Section III presents the rationale for viewing of objective reduction as a machine learning problem, while Section IV presents the proposed framework. The test-suite and the experimental settings are defined in Section V. The working of the proposed algorithms is demonstrated in Section VI, while Section VII presents the experimental results for the considered test-suite. The proposed and the alternative objective reduction algorithms are compared in Section VIII, while the real-world problems are covered in Section IX. The paper concludes with Section X.

II. Objective reduction: Definition, Usefulness, Salient issues and existing Approaches

Objective reduction refers to finding an *essential* objective set for a given problem. In that:

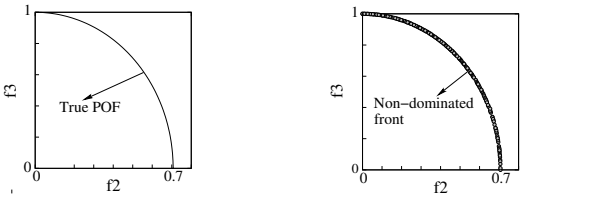
- An *essential* objective set is defined as the smallest set of conflicting objectives ($\mathcal{F}_\mathcal{T}$, $|\mathcal{F}_\mathcal{T}| = m$) which can generate the same POF as that by the original problem, given by $\mathcal{F}_0 = \{f_1, f_2, \ldots, f_M\}$.
- The *dimensionality* of the problem refers to the number of essential objectives[2] (m, $m \leq M$).
- The *redundant* objective set associated with an $\mathcal{F}_\mathcal{T}$ ($\mathcal{F}_{redn} = \mathcal{F}_0 \setminus \mathcal{F}_\mathcal{T}$) refers to the set of objectives, elimination of which does not affect the POF of the given problem. Notably, an objective could be redundant if it is non-conflicting (or correlated) with some other objectives.

[1]Each of the proposed L-PCA and NL-MVU-PCA has been applied for 32 test cases. For every test case, the populations obtained from 20 runs each of NSGA-II and ϵ-MOEA are used, resulting in 2560 simulations. These simulations are repeated for four different algorithm-parameter settings and also for dominance relation preservation based objective reduction approach.

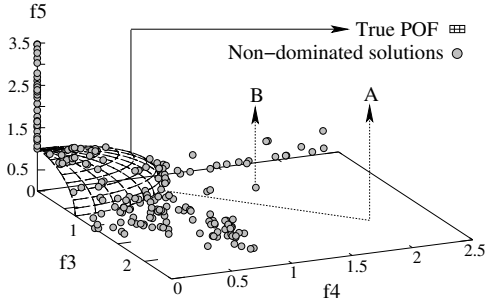
[2]Under regularity condition [4], for m conflicting objectives, the dimensionality of the POF is $m - 1$. However, in the current context, it is meant to refer to the number of essential objectives.

The usefulness of objective reduction is illustrated below in the context of the non-dominated solutions obtained from NSGA-II (denoted by, $\mathcal{N}_{\mathcal{NS}}$) for two problems, namely, DTLZ5$(3,5)$ and DTLZ5$(2,3)$. While these problems are instances of the DTLZ5(I,M) problem introduced later, here it is sufficient to note that:

- DTLZ5$(2,3)$ is a 3-objective problem, where (i) f_1–f_2 are positively correlated, and (ii) the essential objective set: $\mathcal{F}_{\mathcal{T}} = \{f_2, f_3\}$ with $m = 2$.
- DTLZ5$(3,5)$ is a 5-objective problem, where (i) f_1–f_2–f_3 are positively correlated, and (ii) the essential objective set: $\mathcal{F}_{\mathcal{T}} = \{f_3, f_4, f_5\}$ with $m = 3$.



(a) DTLZ5$(2,3)$: True POF (b) DTLZ5$(2,3)$: $\mathcal{N}_{\mathcal{NS}}$



(c) DTLZ5$(3,5)$: Obtained $\mathcal{N}_{\mathcal{NS}}$ vis-à-vis the true POF

Fig. 1. Illustrating that the presence of redundant objectives can hinder the search efficiency of an MOEA. The $\mathcal{N}_{\mathcal{NS}}$ corresponds to a population size of 200 and 2000 generations (one run).

NSGA-II is known to perform well for problems with two or three conflicting objectives. This is affirmed in the case of DTLZ5$(2,3)$, where the obtained $\mathcal{N}_{\mathcal{NS}}$ (Figure 1a) can be seen to conform with the true POF (Figure 1b). Ironically, in the case of DTLZ5$(3,5)$ which has three conflicting objectives, Figure 1c shows that NSGA-II could not converge to the true POF. This can be attributed to the presence of redundant objectives, as explained below. In the $\mathcal{F}_{\mathcal{T}}$ subspace shown in Figure 1c, solution A dominates B. However, B qualifies as a non-dominated solution by being better than A in one of the redundant objectives, namely, f_1 or f_2. This illustrates how the presence of redundant objectives hindered the search efficiency of the NSGA-II, resulting in its poor convergence to the true POF. If NSGA-II were allowed to run for infinitely many generations, the spurious solutions like B would die out and the population would converge to the true POF. However, with a reasonable number of generations, the spurious solutions will prevail. This example illustrates the importance of objective reduction. In that:

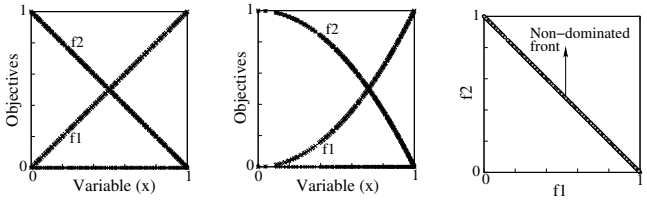
- If the problem could be reduced to $m \leq 3$: an otherwise unsolvable problem could be solved using any of the existing MOEAs. For example, in the above problem, if $\mathcal{F}_{\mathcal{T}}$ were known, NSGA-II would have converged to the true POF.
- Even if $4 \leq m \leq M$: the benefits of objective reduction could be realized in the form of relatively higher search efficiency, lower computational cost and ease in visualization and decision-making.
- Objective reduction could play a complimentary role for the preference-ordering approaches. If the latter were to be applied to the reduced representation of the original problem, higher gains can be expected.
- Besides the above, the benefits of the gain in knowledge about the problem itself, can not be discounted in real-world problems. This is significant because quite often, there is a tendency among practitioners to account for all the performance indicators as objectives, while it may not be intuitive if some of them are correlated.

A. Modes and scope of objective reduction approaches

It is important to highlight some salient issues around objective reduction, before introducing the existing approaches. Firstly, that all these approaches operate on the objective vectors of the non-dominated solutions obtained from an MOEA. Secondly, that objective reduction can be performed during an MOEA run (*online* reduction) to simplify the search or it can be performed post MOEA run (*offline* reduction) to assist in the decision making. Thirdly, that an *essential* objective set may be expressed as a subset of the original problem or its elements may be expressed as a linear combination of the original objectives. While the former approach is referred to as feature *selection*, the latter, as feature *extraction*. It may be noted that all the existing objective reduction approaches pursue feature *selection* towards the ease in subsequent decision making.

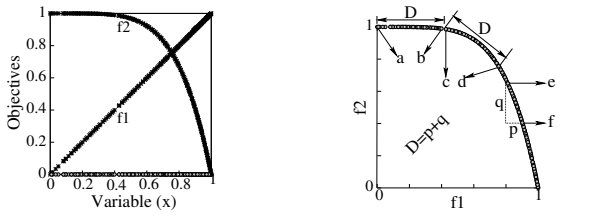
Finally, objective reduction approaches can also be distinguished as *linear* and *nonlinear* reduction approaches. These two refer to identification of an *essential* objective set when the underlying non-dominated front is linear and *nonlinear*, respectively. Here, it is important to discriminate between the linearity and nonlinearity of the non-dominated front and that of the objective functions. While the latter depends on how an objective varies across the search space, the former depends on how the given objectives covary. In that:
1) While linear objectives (Figures 2a) ensure a linear non-dominated front (Figure 2c), even two nonlinear objectives (Figure 2b) can lead to a linear non-dominated front (Figure 2c) *provided* the degree of nonlinearity of both the objectives is the same (identically nonlinear).
2) If the objectives have different degree of nonlinearity (for example, if one objective is linear and the other is nonlinear), then the resulting non-dominated front can only be nonlinear. This is evident from Figures 3a and 3b.

It is important to recognize that an objective reduction approach would need to rely on some form of a distance measure while evaluating the non-dominated solutions. Figure 3b highlights that a distance measure based on either L_1 or L_2 norm will be inadequate to account for nonlinearity, as both of these, do not account for the distribution of solutions



(a) $f_1 = x$ and $f_2 = 1 - x$ (b) $f_1 = x^2$ and $f_2 = 1 - x^2$ (c) Non-dominated front for (a) and (b)

Fig. 2. Illustration: (a) Two linear objectives, and (b) Two identically nonlinear objectives, resulting in (c) a *linear* non-dominated front. $x \in$ [0,1].



(a) $f_1 = x$ and $f_2 = 1 - x^5$ (b) Non-dominated front

Fig. 3. Illustration: (a) Two objectives with different degree of nonlinearity resulting in (b) a *nonlinear* non-dominated front, where use of L_1 or L_2 norm will be inadequate to account for nonlinearity. $x \in$ [0,1].

between any two solutions of interest. For example, since $L_1(a, b) = L_1(e, f) = D$, the distribution of intermediate solutions will be assumed to be identical for both the solution pairs. The same will hold for the solution pairs (a, b) and (c, d), in the case of L_2 norm, since $L_2(a, b) = L_2(c, d) = D$. To summarize, the scope of an objective reduction approach relying either on L_1 or L_2 norm will be limited to *linear* objective reduction (unless used in clusters–where the variability of solutions in each cluster is linear; or used after unfolding the data–as discussed later).

B. The existing objective reduction approaches

With the background established above, the existing objective reduction approaches are presented below.

1) Dominance Relation Preservation (DRP): The objective reduction approach proposed in [31] is based on preserving the dominance relations in the given non-dominated solutions. For a given objective set $\mathcal{F}$, if the dominance relations among the objective vectors remains unchanged when an objective $f \in \mathcal{F}$ is removed, then f is considered to be non-conflicting with the other objectives in $\mathcal{F}$. Based on this criterion, an exact and a greedy algorithm is proposed to address δ-MOSS (finding the minimum objective subset corresponding to a given error δ) and k-EMOSS (finding an objective subset of size k with minimum possible error) problems. However, due to the underlying assumptions [36], this approach is limited to *linear* objective reduction and equally distributed solutions in the objective space (discussed in Section VIII).

2) Unsupervised Feature Selection: This approach [32] utilizes the correlation between objectives and treats the more distant objectives as more conflicting ones. In this: (i) the objective set is divided into neighborhoods of size q around each objective (where, q is reduced during the search), and (ii) the most compact neighborhood is selected, whose center is retained and the neighbors are eliminated. This process is repeated until a pre-specified criterion is achieved. Based on this, two algorithms have been proposed to deal with δ-MOSS and k-EMOSS problems, respectively. However, the role of the parameter q on the performance of the technique and the issue of how to adaptively reduce q needs to be investigated.

3) Pareto Corner Search: A Pareto corner search evolutionary algorithm (PCSEA) has been proposed in [33], which instead of aiming for the complete POF, searches for only the corners of the POF. The solutions so obtained are assumed to appropriately capture the dependency of the POF on different objectives. Then objective reduction is based on the premise that omitting a *redundant* and an *essential* objective will have negligible and substantial effect, respectively, on the number of non-dominated solutions in the population.

4) Machine Learning based objective reduction: This approach [34], [35] utilizes the machine learning techniques like Principal Component Analysis (PCA) and Maximum Variance Unfolding (MVU) to remove the second- and higher-order dependencies in the non-dominated solutions. As this approach lies at the heart of the framework proposed in this paper, the basic concepts are discussed below.

III. Machine Learning based objective reduction

This approach is based on the premise that the *intrinsic structure* of a *garbled* high-dimensional data can be revealed by transforming it such that the effect of *noise* and *redundancy* (dependencies) is minimized. PCA [37] achieves this goal by projecting the given data X on the eigenvectors of the correlation matrix of X.

Notably, PCA is based on removing the *second order dependencies* in the given data. Hence, PCA is likely to be ineffective in capturing the structure of data sets with multi-modal Gaussian or non-Gaussian distributions [37], as in Figure 4(a). Several nonlinear dimensionality reduction methods, such as Kernel PCA [38] and Graph-based methods [39], exist for removing the *higher order dependencies*. In that, the former nonlinearly transforms the data by using a standard kernel function and then applies PCA in the transformed/kernel space. However, its success depends on the *apriori* chosen kernel. The latter does away with this limitation by deriving "data-dependent" kernels.

Maximum Variance Unfolding (MVU) [40] is a graph-based method that computes the low-dimensional representation by explicitly attempting to 'unfold' the high-dimensional data manifold, as in Figure 4. The unfolding is achieved by maximizing the Euclidean distances between the data points while locally preserving the distances and angles between *nearby* points. Mathematically, this can be posed as a semidefinite programming (SDP) problem [40], the output of which is the kernel matrix (say, K) representing the kernel space to which PCA can be applied.

The discussion below presents the rationale for viewing of the objective reduction in many-objective optimization as a machine learning problem:

- The *intrinsic structure* of the POF refers to its intrinsic dimensionality (m) and the essential components ($\mathcal{F}_\mathcal{T}$).

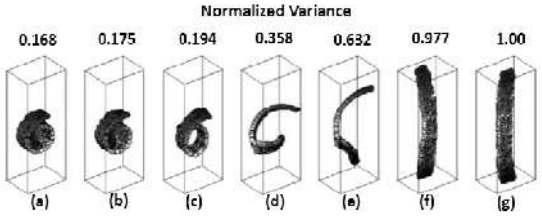


Fig. 4. Maximum Variance Unfolding (taken from [40] and edited). The use of L_2 norm based PCA will be erroneous in (a) but appropriate in (g).

- The *garbled* high-dimensional data set refers to the non-dominated solutions (denoted by, $\mathcal{N}$) obtained from an MOEA. It has previously been mentioned that the objective reduction approaches operate on the objective vectors of $\mathcal{N}$ to identify an *essential* objective set. Hence, an important pre-requisite for accurate objective reduction is that the given $\mathcal{N}$ should be representative of the true POF. However, as has been discussed in Sections I and II, the $\mathcal{N}$ obtained from most of the existing MOEAs fail on this requirement owing to poor convergence and good diversity in areas of poor convergence. This explains the usage of the term garbled for $\mathcal{N}$.
- *Unnoised signal* refers to those non-dominated solutions which are exact optimal solutions. For example, the $\mathcal{N}_{\mathcal{NS}}$ for DTLZ5$(2,3)$ (Figure 1b) represents *unnoised signal*. Similarly, the fraction of $\mathcal{N}_{\mathcal{NS}}$ for DTLZ5$(3,5)$ which conformed with the true POF (solutions like A in Figure 1c) represents the *unnoised signal*.
- *Noised signal* refers to those non-dominated solutions which are not optimal. For example, the fraction of $\mathcal{N}_{\mathcal{NS}}$ for DTLZ5$(3,5)$ which did not conform with the true POF (solutions like B in Figure 1c) represents the *noised signal*. *Noise* refers to the departure in the characteristics of the *noised* signal and those of the *unnoised* signal, for example, the difference in the dimensionality (m) of *unnoised* signal and that of the *noised* signal (which could be greater than m).
- *Redundancy* refers to the presence of objectives which are non-conflicting (or correlated) with some other objectives. Based on discussions in Section II, it can be inferred that *redundancy* may contribute to *garbled data*.

Such a conformance in the definitions and terminology, justifies the viewing of objective reduction as a machine learning problem. It also needs to be recognized that MOEAs are stochastic methods, hence, they can be considered as sampling methods. Depending on the characteristics of different MOEAs and the different properties of the problems, the solutions sampled in objective space may have a uniform, Gaussian (more solutions close to the midrange than the extremes) or non-Gaussian (say, more solutions close to the some/all extremes than midrange) distributions.

IV. Proposed Framework for Linear and Nonlinear Objective Reduction

In the wake of the above background, this section proposes the framework for both linear and nonlinear objective reduction algorithms, namely, L-PCA and NL-MVU-PCA, respectively. Presented as Framework 1, it aims to find an *essential* objective set for a given problem, *offline*. To achieve this aim, the framework assumes that the non-dominated solution set obtained from an MOEA is representative of the true POF, and adopts the following steps: (i) treats the objective vectors of the non-dominated solutions as the input data, (ii) determines the directions of significant variance in the data, (iii) identifies the conflicting objectives along these significant directions and composes a set of such objectives, and (iv) finally eliminates the identically correlated objectives in the above set.

Framework 1: The proposed framework for linear and nonlinear objective reduction algorithms

Input:
$t = 0$ and $\mathcal{F}_t = \{f_1, f_2, \ldots, f_M\}$.

begin
- Obtain a set of non-dominated solutions by running an MOEA corresponding to $\mathcal{F}_t$, for N_g generations with a population size of N.
- Compute a positive semi-definite matrix: R (Equation 1) or K (Equation 2), for L-PCA and NL-MVU-PCA, respectively.
- Compute the eigenvalues and eigenvectors of R or K as the case may be (Section IV-B).
- Perform the Eigenvalue Analysis (Section IV-C) to identify the set of important objectives $\mathcal{F}_e \subseteq \mathcal{F}_t$.
- Perform the Reduced Correlation Matrix Analysis (Section IV-D) to identify the identically correlated subsets ($\mathcal{S}$) in $\mathcal{F}_e$. If there is no such subset, $\mathcal{F}_s = \mathcal{F}_e$.
- Apply the selection scheme (Section IV-E) to identify the most significant objective in each $\mathcal{S}$, to arrive at $\mathcal{F}_s$, such that $\mathcal{F}_s \subseteq \mathcal{F}_e \subseteq \mathcal{F}_t$.
- Compute and store $\mathcal{E}_t$ (Equation 6).
- **if** $\mathcal{F}_s = \mathcal{F}_t$ **then**
  - Stop and declare $\mathcal{F}_t$ as the essential objective set;
  - Set $T = t$ and compute the total error $\mathcal{E}_T$ (Equation 7).
- **end**
- **else**
  - set $t = t + 1$, $\mathcal{F}_t = \mathcal{F}_s$, and go to Step 2.
- **end**

end

While the preliminary proof-of-concept of this approach can be found in [34], [35], this paper is novel in multiple ways, as summarized in Section I. In that, besides the extensive range of results and their analysis, the proposed framework distinctively pursues the four goals of *generality*, *de-noising* of input data, *parameter reduction* and *proposition of an error measure*. These goals are achieved through conceptual enhancements which are highlighted in the description of the framework's steps below, and also summarized in Section IV-G.

As can be seen in Framework 1, it is designed to perform objective reduction iteratively, until the objective sets deduced as essential–in two successive iterations, remain the same. As

the framework's steps are identical for every iteration, they are described below with reference to the first iteration.

A. Construction of a Positive Semi-definite Matrix

This step describes the construction of the correlation (R) and the kernel (K) matrix, to be used for linear and nonlinear objective reduction, respectively. First, the non-dominated solutions are obtained by running an MOEA with the initial objective set $\mathcal{F}_0 = \{f_1, \ldots, f_M\}$, corresponding to a population size of N. For each objective vector in the non-dominated set, $f_i \in \mathbb{R}^N$, let the mean and standard deviation be given by μ_{f_i} and σ_{f_i}, respectively, and let $\ddot{f}_i = (f_i - \mu_{f_i})/\sigma_{f_i}$. Then, the input data X, an $M \times N$ matrix, can be composed as $X = [\ddot{f}_1 \ \ddot{f}_2 \ldots \ddot{f}_M]^T$.

1) Construction of correlation matrix for linear objective reduction (L-PCA): For a given X, the correlation matrix R is defined by Equation 1. It may be noted that L-PCA will be based on de-correlation of R, i.e., removal of second-order dependencies in X. Hence, the scope of L-PCA will largely be limited to *linear* objective reduction.

$$R = \frac{1}{M} X X^T \tag{1}$$

2) Construction of kernel matrix for nonlinear objective reduction (NL-MVU-PCA): It may be noted that NL-MVU-PCA will be based on de-correlation of the kernel matrix K, i.e., removal of higher-order dependencies in X (due to unfolding by MVU). This will allow for nonlinear objective reduction. K can be *learnt* from the SDP formulation[3] presented in Equation 2.

$$\begin{array}{l}\text{Maximize } trace(K) = \sum_{ij} \frac{(K_{ii} - 2K_{ij} + K_{jj})}{2M} \\ \text{subject to the following constraints :} \\ (a) K_{ii} - 2K_{ij} + K_{jj} = R_{ii} - 2R_{ij} + R_{jj}, \forall\ \eta_{ij} = 1 \\ (b) \sum_{ij} K_{ij} = 0 \\ (c) K \text{ is positive-semidefinite,} \\ \text{where : } R_{ij} \text{ is the } (i,j)^{th} \text{ element of the correlation matrix } R\end{array} \tag{2}$$

The neighborhood relation η_{ij} in Equation 2 is governed by the parameter q. For each input feature ($f_i \in \mathbb{R}^N$), q represents the number of neighbors with respect to which the local *isometry*[4] is to be retained during the unfolding, as in Figure 4. While a high value of q ensures better retention of isometry, it delays the unfolding process (say, incremental unfolding from (a) to (b), in Figure 4). In contrast, a low value of q offers fast unfolding (say, direct unfolding from (a) to (g), in Figure 4) but at the risk of distorting the isometry. Given this trade-off, proper selection of q is crucial. While $q = 4$ is mostly used in literature [40], experiments were conducted in [35] to identify the q as a function of M and $q = \lceil \sqrt{M} \rceil$ was found to offer accurate results.

This paper aims to propose a robust framework for objective reduction which relies on as few parameters, as possible. Towards it, the most constrained case of $q = M - 1$ is recommended and used[5]. The rationale for this choice of q is the following: (i) the number of unknown elements in K equals to $M(M+1)/2$ (accounting for symmetry), (ii) Equations 2(b) and (c) jointly represent two constraints, and (iii) for $q = M - 1$, Equations 2(a) leads to $M(M-1)/2$ constraints. Hence, even with $q = M - 1$, the matrix K is not fully specified by the constraints and sufficient degree of freedom ($M(M+1)/2 - M(M-1)/2 - 2 = M - 2$) is available for the unfolding process, while ensuring that the local isometry is retained.

It may be noted that: (i) this SDP is convex for which polynomial time off-the-shelf solvers, such as SeDuMi [41] and CSDP [42] toolbox in MATLAB, are available in public domain, and (ii) computational complexity for solving sparse SDPs [42] is $O(M^3 + c^3)$ where $c = Mq$, represents the number of constraints. In this paper, Sedumi has been used.

B. Computation of Eigenvalues and Eigenvectors

Once R or K are computed (each, an $M \times M$ matrix), their eigenvalues and eigenvectors are determined. Let these be given by: $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_M$ and $V_1, V_2, \ldots V_M$, respectively. To lay the background for discussions ahead, it is important to note the following:

(a) If $e_j = \lambda_j / \sum_{j=1}^{M} \lambda_i$, then $\sum_{j=1}^{M} e_j = 1$.
(b) The i^{th} component of j^{th} principal component, say f_{ij}, reflects on the contribution of f_i towards $V_j \in \mathbb{R}^M$.
(c) From orthonormality of eigenvectors, it follows that: $|V_j| = \sum_{i=1}^{M} f_{ij}^2 = 1 \ \forall j = 1, \ldots, M$.
(d) The contribution of f_i for all V_j's can be given by $c_i^M = \sum_{j=1}^{M} e_j f_{ij}^2$, where $\sum_{i=1}^{M} c_i^M = 1$.

C. Eigenvalue Analysis

This step aims to identify the conflicting objectives (in $\mathcal{F}_0$) along the *significant* principal components (V_js), as follows:

1) The number (N_v) of *significant* V_js are determined using $\sum_{j=1}^{N_v} e_j \geq \theta$, where the variance threshold $\theta \in [0, 1]$ is an algorithm parameter. In this study, $\theta = 0.997$ is used[6].
2) Each *significant* V_j is interpreted as follows[7]:
   (i) Let $\mathcal{F}^+ = \{f_i | f_{ij} \geq 0\}$ and $\mathcal{F}^- = \{f_i | f_{ij} < 0\}$. Also, let f_p and f_n be the objectives with the highest magnitudes in $\mathcal{F}^+$ and $\mathcal{F}^-$, respectively.
   - if $|f_p| \geq |f_n|$, then f_p and all the objectives in $\mathcal{F}^-$ are picked as conflicting objectives.
   - if $|f_p| < |f_n|$, then f_n and all the objectives in $\mathcal{F}^+$ are picked as conflicting objectives.

[3]This is the novel implementation of the original MVU, as proposed in [35]

[4]Two data sets $\{x_i\}_{i=1}^M$ and $\{y_i\}_{i=1}^M$ that are in one-to-one correspondence are said to be q-locally *isometric* if: for every point ϕ_i, there exists a rotation and translation that maps ϕ_i and its q nearest neighbors $\{x_{i1}, \ldots x_{iq}\}$ precisely onto ψ_i and $\{y_{i1}, \ldots y_{iq}\}$. In other words, the distances and angles between *nearby* inputs are preserved.

[5]Experiments are also performed with $q = \lceil \sqrt{M} \rceil$ and the results are presented in the supplementary file provided with this paper

[6]This is an enhancement over [34], [35], to make the interpretation scheme more robust in terms of its ability to handle problems with moderate or negligible redundancy wherein more principal components may be required to identify the conflicting objectives. This value is simply chosen in analogy with Gaussian distributions where it accounts for $\pm 3\sigma$. It does not imply that the scope of the framework is limited to Gaussian distributions.

[7]This is an enhancement over [34], [35] wherein a maximum of two objectives per principal component were picked as conflicting, depending on several parameters. The interpretation scheme proposed here is parameterless and also allows for more objectives per principal component to be picked as conflicting. These two factors contribute to the goal(i) and (iii) of framework.

(ii) If $\mathcal{F}^+ = \emptyset$ or $\mathcal{F}^- = \emptyset$, then the objectives with the highest and the second highest magnitudes are picked.

Let M_E denote the number of objectives deduced as important after the above eigenvalue analysis and let the corresponding set be referred as $\mathcal{F}_e$. Clearly, $\mathcal{F}_e \subseteq \mathcal{F}_0$.

D. Reduced Correlation Matrix (RCM) Analysis

This step aims to identify the subsets of *identically* correlated objectives within $\mathcal{F}_e$. If such subsets exist, then the most significant objective in each subset can be retained while the rest could be discarded, allowing for further reduction of $\mathcal{F}_e$. Towards it, the following steps are proposed:

1) Construction of a Reduced Correlation Matrix (RCM): RCM is the same as R except that the columns corresponding to the objectives in $\mathcal{F}_0 \setminus \mathcal{F}_e$, are not included[8].
2) Identification of *identically* correlated subset for each objective in $\mathcal{F}_e$: For each $f_i \in \mathcal{F}_e$, all $f_j \in \mathcal{F}_e$ which satisfy the conditions of *identical* correlation given by Equation 3, constitute an *identically* correlated subset $\mathcal{S}$.

$$\begin{array}{l}(i)\ \text{sign}(R_{ik}) = \ \text{sign}(R_{jk}),\ k = 1, 2, \ldots, M \\ (ii)\ R_{ij} \geq \text{the correlation threshold } (T_{cor})\end{array} \tag{3}$$

Equation 3(i) ensures that correlation between objectives is interpreted as a set-based property. Equation 3(ii) accounts for the fact the non-dominated set based on which R is computed is *garbled* (Section III) and hence the correlations evident from R may not represent the true correlations on the POF. In a sense, Equation 3(ii) serves to have a *de-noising* effect, while interpreting a matrix based on *gabled* data, to minimize the chances of inaccurate identification of $\mathcal{S}$.

The following description is on the determination of T_{cor}. There exist a few guidelines on interpreting the strength of correlation, one of which is the Cohen scale [43]. It interprets a correlation strength of 0.1 to 0.3 as *weak*, 0.3 to 0.5 as *moderate* and 0.5 to 1.0 as *strong*. However, in the current context, a rigid interpretation of the correlation strength will not be helpful because T_{cor} needs to adapt to conflicting goals depending on the problem being solved. In that, a low and a high T_{cor} is desired while solving problems with high and negligible redundancy, respectively. Towards it, this paper proposes a dynamic computation of T_{cor} based on the spectrum of the eigenvalues for a specific problem. This is based on the understanding that with an increase in the redundancy of a problem: (i) the first eigenvalue e_1 becomes predominantly higher, and (ii) fewer principal components are required to account for a certain variance threshold. For a given M-objective problem, let $M_{2\sigma}$ denote the number of principal components required to account for 95.4% variance[9], then the proposed T_{cor} is given by Equation 4.

$$T_{cor} \quad = 1.0 - e_1(1.0 - M_{2\sigma}/M) \tag{4}$$

[8]This is an enhancement over [34], [35], where both the rows and columns corresponding to the objectives $\mathcal{F}_0 \setminus \mathcal{F}_e$ were eliminated from R. It acknowledges that correlation is a set based property, hence any inference on whether or not two objectives in $\mathcal{F}_e$ are correlated, should be based on the entire set of objectives for which R is computed, namely, $\mathcal{F}_0$.

[9]This is simply chosen in analogy with Gaussian distributions where it accounts for $\pm 2\sigma$ but here it is recommended for all purposes.

It can be realized that for a highly redundant problem where e_1 will be very high, and $M_{2\sigma}$ will be small compared to M, T_{cor} will have a small value. In contrast, for a problem with low redundancy, e_1 will be low, and $M_{2\sigma}$ will be comparable to M, T_{cor} will have a very high value.

E. Selection Scheme for Final Reduction based on RCM Analysis

Once the subsets of identically correlated objectives in $\mathcal{F}_e$ are identified by the RCM analysis, the goal is to identify and retain the most significant objective in each subset ($\mathcal{S}$) and eliminate the remaining ones. This is achieved by: (i) attributing a selection score[10] for each objective, as given by Equation 5, and (ii) retaining the objective with highest c_i in each $\mathcal{S}$ and eliminating the rest.

$$c_i \quad = \sum_{j=1}^{N_v} e_j |f_{ij}| \tag{5}$$

Doing so, will reduce $\mathcal{F}_e$ to $\mathcal{F}_s$ which becomes an *essential* objective set, after one iteration of the proposed framework.

F. Computation of Error

This section proposes a measure for the error incurred in one iteration of the proposed framework. The error measure proposed here aims to compute the variance that is left unaccounted when objectives constituting $\mathcal{F}_{redn} = \mathcal{F}_0 \setminus \mathcal{F}_s$ are discarded, as given below:

$$\left.\begin{array}{ll}\mathcal{E}_t & = \sum\limits_{i \in \mathcal{F}_{redn}} c_i^M (1.0 - \max\limits_{j \in \mathcal{F}_s} \{\delta_{i\,j}.R_{ij}\}) \\ \text{where:} & \\ c_i^M & = \sum_{k=1}^{M} e_k f_{ik}^2 \\ \delta_{i\,j} & = \begin{cases} 1, & \text{if } f_i \text{ and } f_j \text{ are identically correlated} \\ 0, & \text{otherwise} \end{cases} \\ R_{ij} & = \text{strength of correlation between} f_i \text{ and} f_j \end{array}\right\} \tag{6}$$

The rationale for the proposed measure can be realized from the following: (i) when $f_i \in \mathcal{F}_{redn}$ is not identically correlated with any $f_j \in \mathcal{F}_s$: $\delta_{i\,j} = 0$ and the variance left unaccounted by elimination of f_i is c_i^M, and (ii) when $f_i \in \mathcal{F}_{redn}$ is identically correlated with some $f_j \in \mathcal{F}_s$: the variance left unaccounted by elimination of f_i reduces by a factor of R_{ij} as this is already accounted for, by f_j. Equation 6 generalizes this argument.

Suppose, for a given problem, the framework terminates after T iterations. While the error $\mathcal{E}_t$ (Equation 6) represents the error incurred in each iteration, the total error incurred by the framework in T iterations can be given by Equation 7, the rationale for which is the following. For the first iteration, the error $\mathcal{E}_0$ is the variance that will be left unaccounted if the objectives in $\mathcal{F}_{redn}$ were to be eliminated. Hence, the error $\mathcal{E}_1$ in the second iteration will only be with respect to $1 - \mathcal{E}_0$ variance of the original problem. This argument is generalized in Equation 7.

$$\mathcal{E}_T \quad = \mathcal{E}_0 + \sum_{t=1}^{T} \mathcal{E}_t (1 - \mathcal{E}_{t-1}) \tag{7}$$

[10]This is based on capturing an objective's contribution to the *significant* principal components, given by $c_i = \sum_{j=1}^{N_v} e_j f_{ij}^2$ (Section IV-B). However, e_j and f_{ij} being individually less than one, may lead to indiscriminatingly small values for c_i. A remedial form is proposed in Equation 5.

TABLE I
SUMMARY OF THE ENHANCEMENTS IN THE PROPOSED FRAMEWORK, OVER [34], [35]

Proposed Steps	Status in [34], [35]	Classification of proposed enhancements				Conflicting settings for problems with:	
		Adaptation	Correction	Addition	Parameter reduction	$m \ll M$	$m \approx M$
A). q	A parameter				✓		
C). θ	Suitable for $m \ll M$	✓				Low	High
$\mathcal{F}_e$	Parameter dependent, suitable for $m \ll M$	✓			✓	Few objectives per principal component so that: $\lvert\mathcal{F}_e\rvert \ll \lvert\mathcal{F}_0\rvert$	More objectives per principal component so that: $\lvert\mathcal{F}_e\rvert \approx \lvert\mathcal{F}_0\rvert$
D). RCM	Erroneous[a]		✓				
Eq. 3(i)			✓				
Eq. 3(ii)	Absent			✓		Low	High
E). Eq. 5	Absent[b]			✓			
F). Eq. 6	Absent			✓			

[a] The manner is which the reduced correlation matrix (RCM) was constructed, did not allow for correlation to be treated as a set based property.
[b] It was pursued on an ad hoc basis.

G. On the generality and efficiency of the framework

This section summarizes the enhancements over [34], [35] and reflects on the broader issues of generality and efficiency of the proposed framework.

Table I shows, that while the basic steps of the proposed framework overlap with those in the previous versions, these steps are implemented differently in the wake of the four goals that the framework pursues. In that, the steps in [34], [35] have either been *adapted* to cater to the requirements of problems with disparate degree of redundancy; *corrected* for their anomalies; incorporated with *additional* features to denoise the input data or compute the error; or made more robust through *parameter reduction* which helps in minimizing the variability in performance.

Table I also details the facts that the requirements of the proposed framework for handling problems with contrasting degree of redundancy, are opposite in nature. It could be inferred from this table, that the algorithms (resulting from the proposed framework) could be customized to offer higher efficiency for different problems. For example, for problems where $m \ll M$ (as in [34], [35]): (i) a low θ, (ii) composition of $\mathcal{F}$ by picking only a few objectives as conflicting, per principal component, and (iii) a low T_{cor}, may be used. Similarly, for problems where $m \approx M$ or $m = M$, the step of eigenvalue analysis may be skipped and RCM analysis may directly be performed on $\mathcal{F}_0$ with a high value of T_{cor}. It is important to recognize that while doing so will improve the efficiency of the algorithms on the problems they are customized for, it will come at the loss of generality. In that, an algorithm customized for the case of $m \ll M$ will perform poorly for the case of $m \approx M$ or $m = M$, and vice-versa[11].

[11]In relative terms, the efficiency of an algorithm customized for $m \ll M$, when applied to $m \approx M$ or $m = M$ will be lower than the reverse case. This because, in the former case, the requirement of the latter that $|\mathcal{F}_e| \approx |\mathcal{F}_0|$ is likely to be violated, leaving no room for accurate deductions. However, in the reverse case, $|\mathcal{F}_e|$ would just be over-sized than desired for $m \ll M$ but it could still be reduced to $\mathcal{F}_\mathcal{T}$ based on T_{cor}. This explains why the right tail of the left curve in Figure 5 is lower than the left tail of the right curve.

For the sake of generality, the proposed framework assumes that no *a priori* information about the nature of the problem is available. In that, it adopts a high value of θ and a uniform approach of composing $\mathcal{F}_e$, but dynamically assigns T_{cor} based on the problem information revealed by the eigenvalues. Figure 5 summarizes the above discussions.

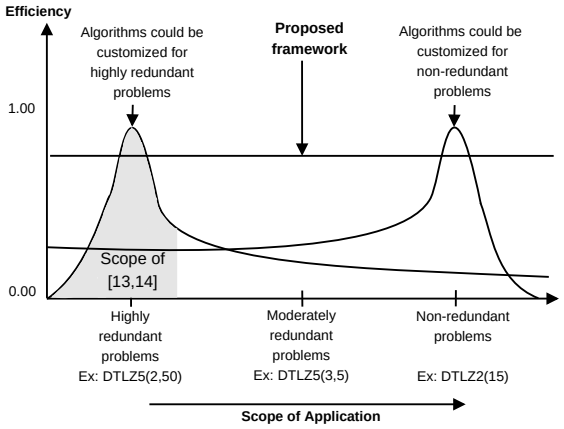


Fig. 5. Highlighting the scope and efficiency of the proposed framework.

V. TEST SUITE AND EXPERIMENTAL SETTINGS

The test suite used in this paper consists of both redundant and non-redundant problems, described below.

A. Redundant Problems

In the DTLZ test suite [44], DTLZ5 and DTLZ6 were designed to degenerate to $m = 2$ for any given M. However, it has been reported in [45] that when $M \geq 4$, these problems fail to degenerate to $m = 2$. For DTLZ5, this issue has been addressed in the form of DTLZ5(I, M). While its formulation is presented in Table II, it may be noted that:

- For any M, the dimension of the POF (characterized by parameter $g = 0$) is $I \leq M$,

- All objectives within $\{f_1, \ldots, f_{M-I+1}\}$ are positively correlated, while each objective in $\{f_{M-I+2}, \ldots, f_M\}$ is conflicting with every other objective in the problem,
- $\mathcal{F}_\mathcal{T} = \{f_k, f_{M-I+2}, \ldots, f_M\}$ defines the true POF, where $k \in \{1, \ldots, M-I+1\}$. Notably, among the possible indices for k, the variance of $f_{k=M-I+1}$ is the highest. Given that the PCA based dimensionality reduction is based on the assumption that large variances have important dynamics:

$$\underset{\text{DTLZ5}(I,M)}{\mathcal{F}_\mathcal{T}} = \{f_{M-I+1}, \ldots, f_M\} \quad \text{(For L-PCA and NL-MVU-PCA)} \tag{8}$$

TABLE II
DESCRIPTION OF DTLZ5(I, M). M DENOTES THE NUMBER OF OBJECTIVES AND I DENOTES THE DIMENSIONALITY OF THE POF. ALL OBJECTIVES ARE TO BE MINIMIZED. ALL $x_i \in [0, 1]$. AS AN EXAMPLE, THE COMPOSITION OF $\mathcal{F}_\mathcal{T}$ IS DERIVED FOR DTLZ5$(4, 5)$[a]

f_1	$= (1+g)\ 0.5\ \Pi_{i=1}^{M-1} cos(\theta_i)$
$f_{m=2:M-1}$	$= (1+g)\ \Pi_{i=1}^{M-m} cos(\theta_i)\ sin(\theta_{M-m+1})$
f_M	$= (1+g)\ sin(\theta_1)$
g	$= \Sigma_{i=M}^{M+k-1}(x_i - 0.5)^2$
$\theta_{i=1:I-1}$	$= \frac{\pi}{2} x_i$
$\theta_{i=I:M-1}$	$= \frac{\pi}{4(1+g)}(1 + 2gx_i)$
subject to $M-I+1$ constraints: $\Sigma_{j=0}^{I-2} f_{M-j}^2 + 2^{p_i} f_i^2 \geqslant 1$	
where $i = 1, \ldots, M-I+1$, and:	
$p_{i=1}$	$= M - I$
$p_{i=2:M-I+1}$	$= (M-I+2) - i$

[a] Substituting $M = 5$, $I = 4$ and $g = 0$, leads to following Equations:
i. $\cos(\theta_1)\cos(\theta_2)\cos(\theta_3) = \sqrt{2} f_1 = \sqrt{2} f_2$ *iii.* $\cos(\theta_1)\sin(\theta_2) = f_4$
ii. $\cos(\theta_1)\cos(\theta_2)\sin(\theta_3) = f_3$ *iv.* $\sin(\theta_1) = f_5$

Squaring and adding Equations i. to iv. helps define the POF as:
$2f_2^2 + f_3^2 + f_4^2 + f_5^2 = 1$ or $2f_1^2 + f_3^2 + f_4^2 + f_5^2 = 1$.
Hence, $\mathcal{F}_\mathcal{T} = \{f_k, f_3, f_4, f_5\}$; $k \in \{1, 2\}$.

Another redundant problem considered in this paper is WFG3 [45], in which the POF degenerates into a linear hyperplane, such that (i) $\Sigma_{m=1}^{M} f_m = 1$, and (ii) all incomparable solutions evolve out of the conflict between only two objectives, f_M and any one of the rest. Hence, $\mathcal{F}_\mathcal{T} = \{f_k, f_M\}$, where, $k \in \{1, \ldots, M-1\}$. Notably, among all possible indices for k, the variance of $f_{k=M-1}$ is the highest, hence, based on the above argument:

$$\underset{\text{WFG3(M)}}{\mathcal{F}_\mathcal{T}} = \{f_{M-1}, f_M\} \quad \text{(For L-PCA and NL-MVU-PCA)} \tag{9}$$

B. Non-redundant Problems

Among the non-redundant problems in the scalable DTLZ suite: the DTLZ1 to DTLZ4, and the DTLZ7 problems are considered. This is because in each of these, the degree of convergence is explicitly quantified by a problem parameter. This will enable the study of the impact of the quality (in terms of convergence and diversity) of the non-dominated solution set on the performance of the objective reduction algorithms.

C. Variable-settings and Properties for the Test Problems

The authors in [45] have categorized the variables as either a 'distance' variable (say, κ: changing which results in comparable solutions) or a 'position' variable (say, ρ: changing which results in incomparable solutions). Hence, the total number of variables in a given problem is equal to the sum of κ and ρ. For an M-objective version of the problems considered in this paper, while $|\rho| = M - 1$ is used, the values used for κ are reported in Table III.

TABLE III
TEST PROBLEMS: VARIABLE-SETTINGS AND QUALITY INDICATORS

Name	Obj	$\lvert\kappa\rvert$	Quality Indicators for POF[a]	
			$g(X)$	$\mathcal{D}_\mathcal{T}^2$
DTLZ1(M)	$f_{1:M}$	5	0	0.25M
DTLZ2(M)	$f_{1:M}$	10	0	M
DTLZ3(M)	$f_{1:M}$	10	0	M
DTLZ4(M)	$f_{1:M}$	10	0	M
DTLZ5(I, M)	$f_{1:M}$	10	0	Note[b]
DTLZ7(M)	$f_{1:M-1}$	–	1	$M - 1 + t_{max}^2$, Note[c]
	f_M	20		
WFG3(M)	$f_{1:M}$	20	NA	NA

[a] While $g(X)$ values are picked from [44], the derivations for $\mathcal{D}_\mathcal{T}^2$ can be found in the supplementary file provided with this paper.
[b] $\left(\frac{1}{2}\right)^{M-I} + \sum_{i=2}^{M-I+1} \left(\frac{1}{2}\right)^{M-I+2-i} + I - 1$
[c] $t = \sum_{i=1}^{M-1} x_i(1 + \sin(3\pi x_i))$ and $t_{max} \approx 0.8594009(M-1)$.

D. Experimental Settings for the MOEAs used

To enable a study on the effect of the underlying $\mathcal{N}$ on the objective reduction algorithms, two MOEAs, namely, NSGA-II and ϵ-MOEA [46] have been used. The obtained non-dominated sets are referred as $\mathcal{N}_{\mathcal{NS}}$, and $\mathcal{N}_\epsilon$, respectively. The choice of these algorithms is driven by the fact that while the quality of $\mathcal{N}_{\mathcal{NS}}$ is known to deteriorate with an increase in M, $\mathcal{N}_\epsilon$ is likely to be better because the underlying notion of ϵ-dominance [47] has been shown to simultaneously preserve the convergence and diversity properties.

For both these algorithms, the following settings are used for experiments: the probability of crossover and mutation as 0.9 and 0.1, respectively; the distribution index for crossover and mutation as 5 and 20, respectively; and, the population size (N) and the number of generations (N_g) as 200 and 2000, respectively. For each test problem, experiments are performed for 20 different runs (evenly spaced seeds in the range between $[0, 1]$). In case of ϵ-MOEA, different values such as $\epsilon = 0.1, 0.2$ and 0.3 were experimented with. However, for the test problems used in this paper, the quality of $\mathcal{N}_\epsilon$ did not significantly change with ϵ. Hence, $\epsilon = 0.3$ was arbitrarily chosen for use in this paper.

E. Quality Indicators for the Non-dominated Solution Sets

The quality of $\mathcal{N}_{\mathcal{NS}}$ and $\mathcal{N}_\epsilon$ for each for the DTLZ problems including DTLZ5(I, M) will be assessed in terms of (i) convergence: measured by the $g(X)$, and (ii) diversity: measured by the normalized maximum spread indicator denoted by I_s [48]. As the name suggests, I_s normalizes the actual dispersal of solutions in the obtained non-dominated set (say, $\mathcal{D}_\mathcal{A}$) by the dispersal of solutions on the true POF (say, $\mathcal{D}_\mathcal{T}$); i.e., $I_s = \mathcal{D}_\mathcal{A}/\mathcal{D}_\mathcal{T}$. The $g(X)$ and $\mathcal{D}_\mathcal{T}$ for the test problems considered (wherever applicable) are reported in Table III.

TABLE IV
DTLZ5$(3,5)$: THE R AND THE K MATRIX WITH THEIR CORRESPONDING EIGENVALUES AND EIGENVECTORS, FOR $\mathcal{N}_{\mathcal{NS}}$ (ONE RUN)

(a) Correlation matrix (R)

	f_1	f_2	f_3	f_4	f_5
f_1	1.0000	0.7391	0.8291	-0.3985	-0.3653
f_2	0.7391	1.0000	0.8761	-0.4357	-0.3226
f_3	0.8291	0.8761	1.0000	-0.3410	-0.2537
f_4	-0.3985	-0.4357	-0.3410	1.0000	-0.4598
f_5	-0.3653	-0.3226	-0.2537	-0.4598	1.0000

(b) Kernel matrix (K)

	f_1	f_2	f_3	f_4	f_5
f_1	3.2335	1.1500	3.2519	-3.8127	-3.8228
f_2	1.1500	2.5984	3.3141	-3.9315	-3.1310
f_3	3.2519	3.3141	16.3678	-12.3313	-10.6025
f_4	-3.8127	-3.9315	-12.3313	23.3217	-3.2461
f_5	-3.8228	-3.1310	-10.6025	-3.2461	20.8026

(c) Eigenvalues and eigenvectors of R

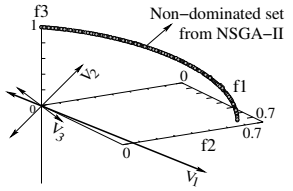
e_1=0.5893	e_2=0.2913	e_3=0.0604	e_4=0.0501	e_5=0.0087
V_1	V_2	V_3	V_4	V_5
0.535	-0.054	0.366	-0.652	-0.390
0.545	-0.008	-0.260	0.609	-0.514
0.548	-0.021	-0.540	-0.212	0.602
-0.281	-0.672	-0.531	-0.269	-0.339
-0.196	0.738	-0.475	-0.293	-0.326

(d) Eigenvalues and eigenvectors of K

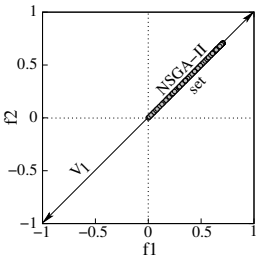
e_1=0.5492	e_2=0.3773	e_3=0.0481	e_4=0.0252	e_5=0.0000
V_1	V_2	V_3	V_4	V_5
0.184	-0.045	0.665	-0.567	0.447
0.176	-0.018	0.339	0.809	0.447
0.640	-0.100	-0.601	-0.137	0.447
-0.613	-0.616	-0.208	-0.041	0.447
-0.387	0.780	-0.195	-0.064	0.447

e_1=0.961	e_2=0.038	$e_3 \approx 0$
V_1	V_2	V_3
0.583	0.400	0.707
0.583	0.400	-0.707
-0.565	0.825	0.000

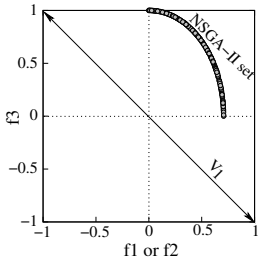
(a) The principal components



(b) The non-dominated set and the principal components



(c) Positive correlation between f_1 and f_2 captured along V_1



(d) Conflict between f_3 and f_1 or f_2 captured along V_1

Fig. 6. Use of PCA for feature selection: Demonstrated on DTLZ5$(2,3)$.

VI. WORKING OF THE PROPOSED ALGORITHMS

This section first demonstrates the basis on which the proposed algorithms pursue the feature *selection* approach for objective reduction. Then, the working demonstration of the L-PCA and NL-MVU-PCA on the DTLZ5$(3,5)$ problem, based on both $\mathcal{N}_{\mathcal{NS}}$ and $\mathcal{N}_{\epsilon}$, is presented. The DTLZ5$(3,5)$ with moderate degree of redundancy is chosen for demonstration, to highlight the absence of any bias in the proposed algorithms for a particular class of problems (discussed in Section IV-G).

A. Feature selection by the proposed algorithms

The proposed framework infers the objectives as conflicting or non-conflicting, depending on whether their contributions along any principal components are with the same-signs or opposite-signs. This is demonstrated on DTLZ5$(2,3)$ problem for which the principal components and the corresponding eigenvalues obtained from R are presented in Figure 6a, and also plotted in Figure 6b vis-à-vis the underlying $\mathcal{N}_{\mathcal{NS}}$.

Notably, the V_1 accounts for 96.1% variance and while the contribution of both f_1 and f_2 towards it is positive in sign ($f_{11} = f_{21} = 0.583$), the contribution of f_3 is negative ($f_{31} = -0.565$). These facts are reflected in: (i) Figure 6c where both f_1 and f_2 can be seen to simultaneously increase or decrease, from any point on V_1 to another, and (ii) Figure 6d where an increase/decrease in $f1$ and $f2$ corresponds to a decrease/increase in f_3. These observations imply that either f_1 or f_2 is redundant and also demonstrate the basis for PCA based feature selection in many-objective context.

In the above problem: (i) the $\mathcal{N}_{\mathcal{NS}}$ fully conformed with the true POF (no *noise*, Section III), and (ii) visualization of the 3-dimensional objective space was possible. Hence, the redundancy of either f_1 or f_2 was evident directly from Figure 6b, without requiring any analysis of the principal components. However, this may not be possible in the case of many-objective problems where not only the visualization is difficult but the underlying data is predominantly characterized by *noise*. This is evident for the DTLZ5$(3,5)$ problem in Figure 1c, where the redundancy of f_1 and f_2 cannot be deduced visually. This justifies the need for an analytical approach, which the proposed algorithms provide.

B. Linear (L-PCA) and Nonlinear (NL-MVU-PCA) Objective Reduction for DTLZ5$(3,5)$ based on $\mathcal{N}_{\mathcal{NS}}$

The correlation matrix R and the kernel matrix K based on $\mathcal{N}_{\mathcal{NS}}$, and also the corresponding eigenvalues and eigenvectors, are presented in Table IV.

1) L-PCA based on $\mathcal{N}_{\mathcal{NS}}$: L-PCA is based on the eigenvalues/eigenvectors of R. Table Va shows that all the principal components together, meet $\theta = 0.997$, to give $\mathcal{F}_e = \{f_1, f_2, f_3, f_4, f_5\}$. Towards the RCM analysis (Table Vb), R in Table IVa shows that the signs of correlation between f_1 and all the remaining objectives are the same as those of f_2 or f_3. In other words, $\{f_1, f_2, f_3\}$ satisfies the Equation 3(i).

TABLE V
DTLZ5-$(3,5)$: ITERATION 1 OF L-PCA WITH $\mathcal{N}_{\mathcal{NS}}$

(a) Eigenvalue Analysis

PCA	Variance	Cumulative	Objectives Selected				
(N^v)	(%)	(%)	f_1	f_2	f_3	f_4	f_5
1	58.93	58.93			f_3	f_4	f_5
2	29.13	88.06	f_1	f_2	f_3	f_4	f_5
3	06.04	94.10	f_1		f_3		
4	05.01	99.11	f_1	f_2			
5	00.87	99.98	f_1	f_2	f_3	f_4	f_5

(b) RCM Analysis

1) Corr. over whole set	Eq 3(i)	$\mathcal{S}_{possible} = \{f_1, f_2, f_3\}$
2) Corr. threshold: T_{cor}	Eq 4	1.0-0.5893(1.0-4/5) = 0.8821
3) Corr. meeting T_{cor}	Eq 3(ii)	$\mathcal{S} = \emptyset$

TABLE VI
DTLZ5-$(3,5)$: ITERATION 1 OF NL-MVU-PCA WITH $\mathcal{N}_{\mathcal{NS}}$

(a) Eigenvalue Analysis

PCA	Variance	Cumulative	Objectives Selected				
(N^v)	(%)	(%)	f_1	f_2	f_3	f_4	f_5
1	54.92	54.92			f_3	f_4	f_5
2	37.73	92.65	f_1	f_2	f_3	f_4	f_5
3	04.81	97.46	f_1		f_3	f_4	f_5
4	02.52	99.98	f_1	f_2	f_3	f_4	f_5

(b) RCM Analysis

1) Corr. over whole set	Eq 3(i)	$\mathcal{S}_{possible} = \{f_1, f_2, f_3\}$
2) Corr. threshold: T_{cor}	Eq 4	1.0-0.5492(1.0-3/5) = 0.7803
3) Corr. meeting T_{cor}	Eq 3(ii)	$\mathcal{S}_1 = \{f_1, f_3\}$, $\mathcal{S}_2 = \{f_2, f_3\}$

(c) Selection scheme

	e_1=0.5492	e_2=0.3773	e_3=0.0481	e_4=0.0252	c_i
	V_1	V_2	V_3	V_4	
f_1	0.184	-0.045	0.665	-0.567	0.1642
f_2	0.176	-0.018	0.339	0.809	0.1401
f_3	0.640	-0.100	-0.601	-0.137	0.4219

However, the condition imposed by Equation 3(ii) is violated as $R_{12} = 0.7391$, $R_{13} = 0.8291$ and $R_{23} = 0.8761$, while $T_{cor} = 0.8821$. It implies that $\{f_1, f_2, f_3\}$ can not be considered as an identically correlated set. This eliminates the need for the last step of the algorithm (Section IV-E) and $\mathcal{F}_s = \mathcal{F}_e$, which does not coincide with $\mathcal{F}_\mathcal{T}$. Clearly, as $\mathcal{F}_{redn} = \emptyset$, $\mathcal{E}_t = 0$ based on Equation 6.

2) NL-MVU-PCA based on $\mathcal{N}_{\mathcal{NS}}$: NL-MVU-PCA is based the eigenvalues/eigenvectors of K (Tables IVb and IVd). It can be seen that Table VIa deduces $\mathcal{F}_e = \{f_1, f_2, f_3, f_4, f_5\}$. Further, based on R in Table IVa, $\{f_1, f_2, f_3\}$ satisfies Equation 3(i) and is the potential set of identically correlated objectives. However, as $R_{12} = 0.7391$, $R_{13} = 0.8291$, $R_{23} = 0.8761$ and $T_{cor} = 0.7803$, Equation 3(ii) is violated by the pair of f_1 and f_2. Hence, $\mathcal{S}_1 = \{f_1, f_3\}$ and $\mathcal{S}_2 = \{f_2, f_3\}$ become the identically correlated sets. Finally, the selection scheme shown in Table VIc identifies f_3 as the more significant objective in both $\mathcal{S}_1$ and $\mathcal{S}_2$. To summarize, NL-MVU-PCA identifies $\mathcal{F}_s = \{f_3, f_4, f_5\}$, which coincides with $\mathcal{F}_\mathcal{T}$. Given that $\mathcal{F}_{redn} = \{f_1, f_2\}$, the error computation based on Equation 6 is as follows:

$$\begin{aligned}\mathcal{E}_t &= \sum_{i\in\{1,2\}} c_i^M (1.0 - \max_{j\in\{3,4,5\}} \{\delta_{i\,j}.R_{ij}\}) \\ &= c_1^M(1.0 - R_{13}) + c_2^M(1.0 - R_{23}) \\ &= 0.0131804 \text{ or } 1.3\%\end{aligned}$$

since:

$$\begin{aligned}c_1^M &= \textstyle\sum_{k=1}^M e_k f_{1k}^2 = 0.048754 \text{ and } R_{13} = 0.8291 \\ c_2^M &= \textstyle\sum_{k=1}^M e_k f_{2k}^2 = 0.039178 \text{ and } R_{23} = 0.8761\end{aligned}$$

TABLE VII
DTLZ5$(3,5)$: THE R AND THE K MATRIX WITH THEIR CORRESPONDING EIGENVALUES AND EIGENVECTORS, FOR $\mathcal{N}_\epsilon$ (ONE RUN)

(a) Correlation matrix (R)

	f_1	f_2	f_3	f_4	f_5
f_1	1.0000	0.9875	0.9736	-0.4310	-0.3384
f_2	0.9875	1.0000	0.9644	-0.4496	-0.3040
f_3	0.9736	0.9644	1.0000	-0.4689	-0.3229
f_4	-0.4310	-0.4496	-0.4689	1.0000	-0.5136
f_5	-0.3384	-0.3040	-0.3229	-0.5136	1.0000

(b) Kernel matrix (K)

	f_1	f_2	f_3	f_4	f_5
f_1	2.4039	2.4928	4.1692	-4.4405	-4.6255
f_2	2.4928	2.7527	4.4084	-5.0861	-4.5678
f_3	4.1692	4.4084	8.1727	-9.0340	-7.7163
f_4	-4.4405	-5.0861	-9.0340	27.0187	-8.4579
f_5	-4.6255	-4.5678	-7.7163	-8.4579	25.3677

(c) Eigenvalues and eigenvectors of R

e_1=0.6567	e_2=0.3007	e_3=0.0333	e_4=0.0069	e_5=0.0021
V_1	V_2	V_3	V_4	V_5
0.546	-0.056	-0.219	0.225	-0.775
0.544	-0.030	-0.274	0.505	0.610
0.545	-0.030	-0.059	-0.820	0.165
-0.285	-0.660	-0.686	-0.114	0.007
-0.166	0.748	-0.635	-0.097	-0.020

(d) Eigenvalues and eigenvectors of K

e_1=0.5322	e_2=0.4608	e_3=0.0057	e_4=0.0011	e_5=0.0000
V_1	V_2	V_3	V_4	V_5
0.059	0.269	-0.376	-0.764	-0.447
0.076	0.282	-0.553	0.639	-0.447
0.142	0.490	0.730	0.086	-0.447
-0.821	-0.337	0.108	0.027	-0.447
0.545	-0.704	0.091	0.011	-0.447

C. Linear (L-PCA) and Nonlinear (NL-MVU-PCA) Objective Reduction for DTLZ5$(3,5)$ based on $\mathcal{N}_\epsilon$

The correlation matrix R and the kernel matrix K based on $\mathcal{N}_\epsilon$, and also the corresponding eigenvalues and eigenvectors, are presented in Table VII.

1) L-PCA based on $\mathcal{N}_\epsilon$: From Table VIII, it follows that $\mathcal{F}_s = \{f_1, f_4, f_5\}$. Hence, the corresponding error is:
$\mathcal{E}_t = c_2^M(1.0 - R_{21}) + c_3^M(1.0 - R_{31}) = 0.007766$ or 0.77%.

2) NL-MVU-PCA based on $\mathcal{N}_\epsilon$: From Table IX, it follows that $\mathcal{F}_s = \{f_3, f_4, f_5\}$. Hence, the corresponding error is: $\mathcal{E}_t = c_1^M(1.0 - R_{13}) + c_2^M(1.0 - R_{23}) = 0.002455$ or 0.24%

TABLE VIII
DTLZ5-$(3, 5)$: ITERATION 1 OF L-PCA WITH $\mathcal{N}_\epsilon$

(a) Eigenvalue Analysis

PCA	Variance	Cumulative	Objectives Selected				
(N^v)	(%)	(%)	f_1	f_2	f_3	f_4	f_5
1	65.67	65.67	f_1			f_4	f_5
2	30.07	95.74	f_1	f_2	f_3	f_4	f_5
3	03.33	99.07				f_4	f_5
4	00.69	99.76	f_1	f_2	f_3		

(b) RCM Analysis

1) Corr. over whole set	Eq 3(i)	$\mathcal{S}_{possible} = \{f_1, f_2, f_3\}$
2) Corr. threshold: T_{cor}	Eq 4	1.0-0.6567(1.0-2/5) = 0.6059
3) Corr. meeting T_{cor}	Eq 3(ii)	$\mathcal{S} = \{f_1, f_2, f_3\}$

(c) Selection scheme

	e_1=0.6567	e_2=0.3007	e_3=0.0333	e_4=0.0069	c_i
	V_1	V_2	V_3	V_4	
f_1	0.546	-0.056	-0.219	0.225	0.3859
f_2	0.544	-0.030	-0.274	0.505	0.3805
f_3	0.545	-0.030	-0.059	-0.820	0.3749

TABLE IX
DTLZ5-$(3, 5)$: ITERATION 1 OF NL-MVU-PCA WITH $\mathcal{N}_\epsilon$

(a) Eigenvalue Analysis

PCA	Variance	Cumulative	Objectives Selected				
(N^v)	(%)	(%)	f_1	f_2	f_3	f_4	f_5
1	53.22	53.22	f_1	f_2	f_3	f_4	f_5
2	46.08	99.30	f_1	f_2	f_3		f_5
2	00.57	99.87	f_1	f_2	f_3		

(b) RCM Analysis

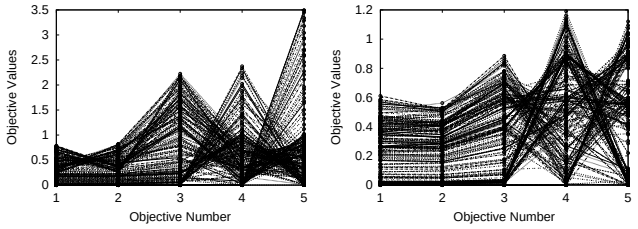
1) Corr. over whole set	Eq 3(i)	$\mathcal{S}_{possible} = \{f_1, f_2, f_3\}$
2) Corr. threshold: T_{cor}	Eq 4	1.0-0.5322(1.0-2/5) = 0.6806
3) Corr. meeting T_{cor}	Eq 3(ii)	$\mathcal{S} = \{f_1, f_2, f_3\}$

(c) Selection scheme

	e_1=0.5322	e_2=0.4608	e_3=0.0057	c_i
	V_1	V_2	V_3	
f_1	0.059	0.269	-0.376	0.1580
f_2	0.076	0.282	-0.553	0.1743
f_3	0.142	0.490	0.730	0.3054

D. Discussion: Effect of the quality of the non-dominated solutions on the performance of L-PCA and NL-MVU-PCA

This section revisits the performance of the proposed algorithms vis-à-vis the underlying data set, to highlight some key issues. For the DTLZ5$(3, 5)$ problem:



(a) $\mathcal{N}_{\mathcal{NS}}$: Parallel coordinate plot (b) $\mathcal{N}_\epsilon$: Parallel coordinate plot

Fig. 7. DTLZ5-$(3, 5)$: Parallel coordinate plots for $\mathcal{N}_{\mathcal{NS}}$ and $\mathcal{N}_\epsilon$ (one run).

TABLE X
CONVERGENCE AND DIVERSITY: FOR THE REDUNDANT DTLZ5(I,M) AND NON-REDUNDANT DTLZ PROBLEMS. THE MEAN (μ) AND STANDARD DEVIATION (σ) ARE AVERAGED OVER 20 RUNS

Problems	Convergence (g)		Diversity (I_s)	
DTLZ5 I M	$\mathcal{N}_\epsilon$ $(\mu \pm \sigma)$	$\mathcal{N}_{\mathcal{NS}}$ $(\mu \pm \sigma)$	$\mathcal{N}_\epsilon$ $(\mu \pm \sigma)$	$\mathcal{N}_{\mathcal{NS}}$ $(\mu \pm \sigma)$
2 05	0.15 ± 0.09	0.48 ± 0.63	1.93 ± 0.01	4.20 ± 0.17
2 10	0.19 ± 0.09	1.64 ± 0.75	1.99 ± 0.05	6.58 ± 0.13
2 20	0.20 ± 0.07	2.10 ± 0.58	1.91 ± 0.02	8.06 ± 0.17
2 30	0.22 ± 0.08	2.24 ± 0.47	1.98 ± 0.03	8.40 ± 0.17
2 50	0.23 ± 0.08	2.34 ± 0.34	1.99 ± 0.02	8.18 ± 0.22
3 05	0.08 ± 0.04	0.70 ± 0.61	1.48 ± 0.00	3.54 ± 0.02
3 10	0.15 ± 0.07	1.83 ± 0.59	1.57 ± 0.00	5.62 ± 0.03
3 20	0.17 ± 0.07	2.25 ± 0.03	1.59 ± 0.02	6.67 ± 0.01
5 10	0.14 ± 0.07	2.06 ± 0.34	1.38 ± 0.02	4.29 ± 0.02
5 20	0.15 ± 0.07	2.25 ± 0.31	1.37 ± 0.00	5.00 ± 0.01
7 10	0.16 ± 0.07	1.99 ± 0.38	1.27 ± 0.01	3.30 ± 0.01
7 20	0.16 ± 0.08	2.17 ± 0.38	1.28 ± 0.00	3.94 ± 0.02
DTLZ M				
1 05	170 ± 129	410 ± 141	480 ± 0	539 ± 6
1 15	313 ± 154	948 ± 165	321 ± 4	507 ± 2
1 25	346 ± 152	944 ± 207	269 ± 1	331 ± 6
2 05	0.09 ± 0.05	0.11 ± 0.06	1.16 ± 0.00	1.08 ± 0.02
2 15	0.23 ± 0.11	2.08 ± 0.45	0.99 ± 0.01	2.41 ± 0.00
2 25	0.22 ± 0.11	2.12 ± 0.49	0.77 ± 0.01	2.10 ± 0.00
3 05	709 ± 223	0738 ± 220	1168 ± 5	1047 ± 18
3 15	903 ± 225	1733 ± 356	0880 ± 7	1358 ± 08
3 25	908 ± 220	1808 ± 400	0669 ± 4	1187 ± 13
4 05	0.12 ± 0.06	0.13 ± 0.06	1.13 ± 0.05	1.14 ± 0.00
4 15	0.22 ± 0.09	2.20 ± 0.15	1.21 ± 0.00	2.80 ± 0.01
4 25	0.29 ± 0.13	2.18 ± 0.15	1.24 ± 0.01	2.63 ± 0.00
7 05	1.24 ± 0.10	1.17 ± 0.07	1.77 ± 0.03	2.09 ± 0.02
7 15	5.04 ± 0.57	2.94 ± 0.72	3.61 ± 0.05	5.50 ± 0.02
7 25	5.44 ± 0.56	3.12 ± 0.66	3.62 ± 0.04	5.28 ± 0.12

1) The POF is characterized by $g(X) = 0$ and $I_s = 1.0$: Given this, $\mathcal{N}_\epsilon$ is better than $\mathcal{N}_{\mathcal{NS}}$ in terms of the quality indicators, because Table X shows that:
   - For $\mathcal{N}_\epsilon$: $g(X) = 0.08 \pm 0.04$ and $I_s = 1.48 \pm 0.00$,
   - For $\mathcal{N}_{\mathcal{NS}}$: $g(X) = 0.70 \pm 0.61$ and $I_s = 3.54 \pm 0.02$.
2) $\mathcal{F}_\mathcal{T} = \{f_3, f_4, f_5\}$ and f_1–f_2–f_3 are non-conflicting (correlated) among themselves, where, $R_{12} = R_{13} = R_{23} = 1.0$. However:
   - $\mathcal{N}_{\mathcal{NS}}$ (Table IVa) reports $R_{12} = 0.73$, $R_{13} = 0.82$ and $R_{23} = 0.87$.
   - $\mathcal{N}_\epsilon$ (Table VIIa) reports $R_{12} = 0.98$, $R_{13} = 0.97$ and $R_{23} = 0.96$.

   This highlights the superiority of $\mathcal{N}_\epsilon$ over $\mathcal{N}_{\mathcal{NS}}$ in capturing the inter-relationships of the objective vectors–on the true POF. The same is also evident in Figure 7a for $\mathcal{N}_{\mathcal{NS}}$, where compared to $\mathcal{N}_\epsilon$ (Figure 7b), more solutions are wrongly in conflict with respect to $\{f_1, f_2, f_3\}$.

TABLE XI

PERFORMANCE OF DIFFERENT OBJECTIVE REDUCTION ALGORITHMS ON REDUNDANT TEST PROBLEMS, CORRESPONDING TO $\mathcal{N}_{\mathcal{NS}}$ AND $\mathcal{N}_\epsilon$ AND $\theta = 0.997$. THE NUMBERS IN THE TABLE INDICATE THE FREQUENCY OF SUCCESS IN IDENTIFYING THE TRUE I AND $\mathcal{F}_\mathcal{T}$, OUT OF 20 RUNS. THE DASHES (-) REPLACE 0 TO IMPLY INCONSEQUENTIAL ENTRIES WHEREVER THE PREREQUISITE OF ACCURATE I IS NOT MET. THE TABLE'S FOOTNOTE REPORTS THE PROBLEMS THAT REQUIRE MULTIPLE ITERATIONS OF THE ALGORITHM TO OBTAIN ACCURATE RESULTS, AS: P—AR(BI), IMPLYING THAT FOR THE PROBLEM P, A RUNS OUT OF 20, REQUIRED B ITERATIONS EACH

Test problems		Proposed approaches								Dominance relation preservation [31], [36]: δ-MOSS, 0% Error							
		NL-MVU-PCA				L-PCA				Greedy Approach				Exact Approach			
DTLZ5(I, M)		$\mathcal{N}_\epsilon$		$\mathcal{N}_{\mathcal{NS}}$		$\mathcal{N}_\epsilon$		$\mathcal{N}_{\mathcal{NS}}$		$\mathcal{N}_\epsilon$		$\mathcal{N}_{\mathcal{NS}}$		$\mathcal{N}_\epsilon$		$\mathcal{N}_{\mathcal{NS}}$	
I	M	I[b]	$\mathcal{F}_\mathcal{T}$	I[c]	$\mathcal{F}_\mathcal{T}$	I[d]	$\mathcal{F}_\mathcal{T}$	I	$\mathcal{F}_\mathcal{T}$	I	$\mathcal{F}_\mathcal{T}$	I	$\mathcal{F}_\mathcal{T}$	I	$\mathcal{F}_\mathcal{T}$	I	$\mathcal{F}_\mathcal{T}$
2	5	20	20	20	20	20	14	20	1	0	-	0	-	0	-	0	-
2	10	20	20	11	11	20	7	0	-	0	-	0	-	0	-	0	-
2	20	20	20	7	7	20	2	0	-	0	-	0	-	0	-	0	-
2	30	20	20	14	14	20	1	0	-	0	-	0	-	0	-	0	-
2	50	20	20	14	14	20	0	10	-	0	-	0	-	0	-	0	-
3	5	20	20	18	18	20	9	0	0	0	-	0	-	0	-	0	-
3	10	20	20	0	-	20	2	0	-	0	-	0	-	0	-	0	-
3	20	20	20	0	-	20	1	0	-	0	-	0	-	0	-	0	-
5	10	19	19	0	-	19	3	0	-	0	-	0	-	0	-	0	-
5	20	19	19	0	-	18	3	0	-	0	-	0	-	0	-	0	-
7	10	19	19	0	-	20	6	0	-	0	-	0	-	0	-	0	-
7	20	16	16	0	-	13	3	0	-	0	-	0	-	0	-	0	-
WFG3	5	20	20	20	20	20	19	20	4	0	-	0	-	0	-	0	-
	15	15	15	20	20	10	1	19	0	0	-	0	-	0	-	0	-
	25	9	9	20	20	6	0	20	0	0	-	0	-	0	-	0	-

[a] The tabulated results are obtained using the source codes at: http://www.tik.ee.ethz.ch/sop/download/supplementary/objectiveReduction/. The same codes are used for the results presented later in Tables XII and XVII.

[b] DTLZ5$(2, 30)$—1R(2I); DTLZ5$(5, 10)$—3R(2I); DTLZ5$(5, 20)$—8R(2I); DTLZ5$(7, 10)$—7R(2I); DTLZ5$(7, 20)$—9R(2I) and 1R(4I); WFG3(15)—6R(2I) and 3R(3I); WFG3(25)—5R(2I), 1R(3I) and 1R(4I).

[c] DTLZ5$(2, 10)$—11R (2I), 5R(3I) and 2R(4I); DTLZ5$(2, 20)$—7R(2I) iterations, 2R(4I), 2R(5I), 1R(6I) and 1R(7I); DTLZ5$(2, 30)$—3R(2I), 3R(3I), 4R(4I) and 4R(5I); DTLZ5$(2, 50)$—8R(2I) and 2R(3I); DTLZ5$(3, 5)$—9R(2I).

[d] For DTLZ5$(5, 20)$—1R(2I); DTLZ5$(7, 10)$—3R(2I); DTLZ5$(7, 20)$—1R(2I); WFG3(5)—11R(2I); WFG3(15)—1R(2I).

The superiority of $\mathcal{N}_\epsilon$ over $\mathcal{N}_{\mathcal{NS}}$ is reflected in the following:
1) Higher accuracy for L-PCA corresponding to $\mathcal{N}_\epsilon$: L-PCA based on $\mathcal{N}_{\mathcal{NS}}$ could not even identify the true dimension of the POF ($I = 3$), let alone its composition ($\mathcal{F}_\mathcal{T}$). However, with $\mathcal{N}_\epsilon$, it accurately identifies I but not $\mathcal{F}_\mathcal{T}$. In this, f_1 with lower variance is erroneously picked as more important than f_3 with higher variance. This relates to the limitations of linear PCA approach discussed in Section III.
2) Higher efficiency for NL-MVU-PCA corresponding to $\mathcal{N}_\epsilon$: NL-MVU-PCA correctly identifies I and $\mathcal{F}_\mathcal{T}$ in case of both $\mathcal{N}_{\mathcal{NS}}$ and $\mathcal{N}_\epsilon$. However, the efficiency in latter case ($N_v = 3$ for $\mathcal{N}_\epsilon$) is higher than the former ($N_v = 4$, for $\mathcal{N}_{\mathcal{NS}}$).
3) Higher reliability of the reported errors: It can be realized that the reliability of the error measure (Equation 6) depends on the accuracy the three factors, namely, c_i^M (contribution of f_i across all V_j's); δ_{ij} (whether or not f_i and f_j are *identically* correlated); and R_{ij} (strength of correlation between f_i and f_j). The accuracy of these factors in turn depends on the interplay of how accurate an algorithm is (controlling c_i^M) and how well the underlying solutions (controlling δ_{ij} and R_{ij}) represent the true POF (for which $\mathcal{E}_t = 0$ corresponding to $\mathcal{F}_\mathcal{T} = \{f_3, f_4, f_5\}$).

In the wake of the results in Sections VI-B and VI-C and the observations in Items VI-D (1) and VI-D (2), the above argument explains why the reliability of the reported $\mathcal{E}_t$ improves in the following order: (i) L-PCA ($\mathcal{N}_{\mathcal{NS}}$, $\mathcal{E}_t = 0$), (ii) NL-MVU-PCA ($\mathcal{N}_{\mathcal{NS}}$, $\mathcal{E}_t = 0.0131$), (iii) L-PCA ($\mathcal{N}_\epsilon$, $\mathcal{E}_t = 0.0077$) and (iv) NL-MVU-PCA ($\mathcal{N}_\epsilon$, $\mathcal{E}_t = 0.0024$).

E. Benefit of objective reduction

It is worth highlighting here the benefit that objective reduction offers. Figure 1c illustrated the poor performance of NSGA-II as the number of objectives grew to five, and the poor performance was linked to the hinderance in search caused by the presence of redundant objectives. However, with NL-MVU-PCA identifying that f_1 and f_2 are redundant, an apparently many-objective problem is reduced to a three-objective problem. The improvement in the quality of the $\mathcal{N}_{\mathcal{NS}}$, accompanying this reduction can be assessed from Figure 8, where $\mathcal{N}_{\mathcal{NS}}$ conforms with the true POF.

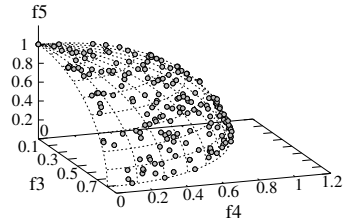


Fig. 8. DTLZ5-$(3, 5)$: Illustrating that the $\mathcal{N}_{\mathcal{NS}}$ (one run) obtained for the reduced problem, $\mathcal{F}_\mathcal{T} = \{f_3, f_4, f_5\}$, conforms with the true POF.

VII. EXPERIMENTAL RESULTS ON OTHER BENCHMARK PROBLEMS

This section presents the experimental results for a wide range of redundant and non-redundant problems, in Tables XI

TABLE XII
PERFORMANCE OF DIFFERENT OBJECTIVE REDUCTION ALGORITHMS ON NON-REDUNDANT TEST PROBLEMS, CORRESPONDING TO $\mathcal{N}_{\mathcal{NS}}$ AND $\mathcal{N}_\epsilon$ AND $\theta = 0.997$. THE NUMBERS IN THE TABLE INDICATE THE FREQUENCY OF SUCCESS IN IDENTIFYING THE $\mathcal{F}_\mathcal{T}$, OUT OF 20 RUNS. THE DASHES (-) DENOTE CASES WHICH WERE NOT TRIED DUE TO HUGE COMPUTATIONAL TIME

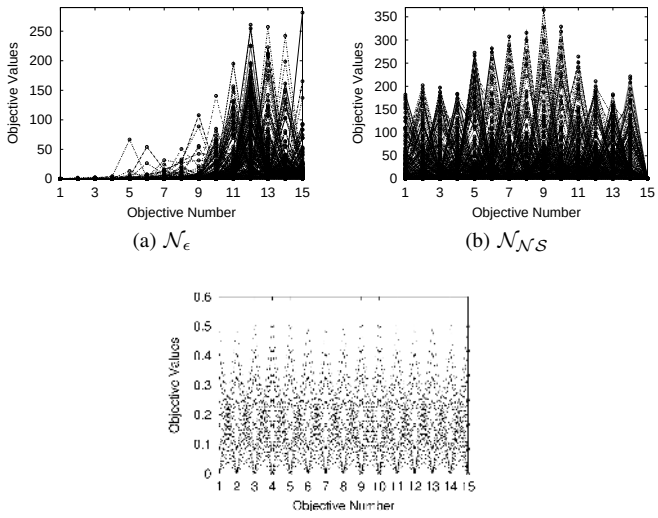
Test problems		Proposed approaches				Dominance relation preservation [31], [36]: δ-MOSS, 0% Error			
DTLZ(M)		NL-MVU-PCA		L-PCA		Greedy Approach		Exact Approach	
Name	M	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{NS}}$	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{NS}}$	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{NS}}$	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{NS}}$
DTLZ1	5	20	20	20	20	20	20	20	20
	15	6	20	14	20	8	0	7	0
	25	0	20	3	20	0	0	–	–
DTLZ2	5	20	20	20	20	20	20	20	20
	15	20	20	20	20	1	0	0	0
	25	16	20	17	20	0	0	–	–
DTLZ3	5	20	20	20	20	20	20	20	20
	15	11	20	20	20	8	0	2	0
	25	0	20	5	20	0	0	–	–
DTLZ4	5	20	20	20	20	14	20	20	20
	15	20	20	20	20	0	0	0	0
	25	20	20	20	20	0	0	–	–
DTLZ7	5	20	20	20	20	20	20	20	20
	15	19	17	20	20	1	0	0	0
	25	20	20	20	20	0	0	–	–

and XII, respectively. These results need to be interpreted in the wake of: (i) the quality of the underlying $\mathcal{N}_{\mathcal{NS}}$ and $\mathcal{N}_\epsilon$ summarized in Table X, and (ii) the resulting errors associated with NL-MVU-PCA and L-PCA (reported in the supplementary file provided with the paper).

Reference to the quality indicators for the DTLZ5(I, M) problems in Table X, reveals that $\mathcal{N}_\epsilon$ is superior than $\mathcal{N}_{\mathcal{NS}}$. The discussion in Section VI-D on DTLZ5$(3, 5)$ highlighted that $\mathcal{N}_\epsilon$ is also better than $\mathcal{N}_{NS}$, in terms of capturing the inter-relationships of the objective vectors–on the true POF. The same trend holds true for the other DTLZ5(I, M) problems, and explains the results in Table XI. In that, superior results are obtained by both NL-MVU-PCA and L-PCA in the case of $\mathcal{N}_\epsilon$ as against $\mathcal{N}_{\mathcal{NS}}$.

For interpretation of the results for non-redundant problems, refer to Tables X and XII. It can be seen that:
1) In contrast to the redundant problems, the performance of both NL-MVU-PCA and L-PCA is found to be better with $\mathcal{N}_{NS}$ as against $\mathcal{N}_\epsilon$. For example, all the three versions of the DTLZ1 problem are accurately solved in case of $\mathcal{N}_{NS}$. However, the accuracy with $\mathcal{N}_\epsilon$ falls as objectives increase from 5, 15 to 25. The explanation lies in the following. The true POF of DTLZ1(15), as shown in Figure 9c, is characterized by two main features: (i) all the objectives are conflicting, and (ii) the variance of all the objectives on the true POF is equal. For an accurate analysis, the proposed algorithms need a solution set that exhibits the above two features. Clearly, these features of the POF are better retained by $\mathcal{N}_{NS}$ (Figure 9b) than $\mathcal{N}_\epsilon$ (Figure 9a). In that, while the former exhibits conflict between all the objectives, the latter suggests that some of the objectives may be non-conflicting (such as f_1 to f_4). Hence, *from the perspective of the proposed algorithms*, $\mathcal{N}_{NS}$ is more representative of the true POF than $\mathcal{N}_\epsilon$ despite the fact that the latter is better in terms on convergence and diversity measures (Tables X).

To summarize, from the PCA based objective reduction perspective, the interpretation of the population quality based on convergence and diversity measures alone is inadequate. In that, how far the inter-relationships of the objective vectors on the true POF in terms of their conflict and relative variance are retained in a solution set, is critical. It may be noted that NL-MVU-PCA when applied to the analytically generated solution set (Figure 9c), accurately solved the problem. Based on these observations, the poor performance of NL-MVU-PCA corresponding to $\mathcal{N}_\epsilon$ could be attributed to the latter's mis-represention of the true POF.



(a) $\mathcal{N}_\epsilon$ (b) $\mathcal{N}_{\mathcal{NS}}$

(c) Analytically generated POF

Fig. 9. DTLZ1(15): Parallel coordinate plots for $\mathcal{N}_\epsilon$, $\mathcal{N}_{NS}$ (one run) and analytically generated POF.

2) In contrast to the redundant problems, the performance of L-PCA based on $\mathcal{N}_\epsilon$, is better than NL-MVU-PCA. This observation is highly *misleading*, as explained below. If $\mathcal{N}_\epsilon$ for DTLZ1(15), shown in Figure 9a, is to be treated as the representative of the true POF, it suggests that: (i) some objectives may be non-conflicting (such as f_1 to f_4),

TABLE XIII

EFFECT OF THE VARIANCE THRESHOLD ($\theta = 0.682$) ON THE PERFORMANCE OF THE PROPOSED L-PCA AND NL-MVU-PCA ALGORITHMS ON BOTH REDUNDANT AND NON-REDUNDANT TEST PROBLEMS, CORRESPONDING TO $\mathcal{N}_{\mathcal{NS}}$ AND $\mathcal{N}_\epsilon$. THE NUMBERS IN THE TABLE INDICATE THE FREQUENCY OF SUCCESS IN IDENTIFYING THE TRUE I AND/OR $\mathcal{F}_\mathcal{T}$, OUT OF 20 RUNS. THE DASHES (-) REPLACE 0 TO IMPLY INCONSEQUENTIAL ENTRIES, AS THE PREREQUISITE I IS NOT MET. THE TABLE'S FOOTNOTE REPORTS THE PROBLEMS THAT REQUIRE MULTIPLE ITERATIONS OF THE ALGORITHM TO OBTAIN ACCURATE RESULTS, AS: P—AR(BI), IMPLYING THAT FOR THE PROBLEM P, A RUNS OUT OF 20, REQUIRED B ITERATIONS EACH

Test problems		NL-MVU-PCA				L-PCA				Test Problems					
DTLZ5(I, M)		$\mathcal{N}_\epsilon$		$\mathcal{N}_{\mathcal{NS}}$		$\mathcal{N}_\epsilon$		$\mathcal{N}_{\mathcal{NS}}$		DTLZ(M)		NL-MVU-PCA		L-PCA	
I	M	I[a]	$\mathcal{F}_\mathcal{T}$	I[b]	$\mathcal{F}_\mathcal{T}$	I[c]	$\mathcal{F}_\mathcal{T}$	I[d]	$\mathcal{F}_\mathcal{T}$	Name	M	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{NS}}$	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{NS}}$
2	5	20	20	20	20	20	1	20	18	DTLZ1	5	17	18	15	19
2	10	20	20	12	12	20	0	7	4		15	6	18	10	20
2	20	20	20	8	8	20	0	2	1		25	0	20	3	20
2	30	20	20	14	14	20	0	10	1	DTLZ2	5	19	18	17	17
2	50	20	20	19	19	20	0	20	2		15	19	19	20	17
3	5	20	20	18	18	20	5	1	1		25	14	20	15	20
3	10	20	20	0	-	20	1	0	-	DTLZ3	5	16	17	19	18
3	20	20	20	0	-	20	0	0	-		15	6	20	12	19
5	10	17	17	0	-	18	3	0	-		25	0	20	2	20
5	20	19	19	0	-	10	0	0	-	DTLZ4	5	15	20	19	18
7	10	16	15	0	-	18	7	0	-		15	19	20	20	20
7	20	13	13	0	-	9	1	0	-		25	20	20	20	20
	5	20	20	20	20	20	14	20	1	DTLZ7	5	20	20	19	20
WFG3	15	15	15	20	20	3	1	20	0		15	19	17	20	20
	25	9	9	20	20	0	-	20	0		25	20	20	20	20

[a] DTLZ5$(2, 30)$—1R(2I); DTLZ5$(5, 10)$—2R(2I); DTLZ5$(5, 20)$—6R(2I); DTLZ5$(7, 10)$—4R(2I); DTLZ5$(7, 20)$—5R(2I) and 1R(4I); WFG3$(3, 10)$—6R(2I) and 3R(3I); WFG3$(3, 20)$—5R(2I), 1R(3I) and 1R(4I).

[b] DTLZ5$(2, 10)$—6R(2I), 5R(3I) and 1R(4I); DTLZ5$(2, 20)$—1R(2I), 4R(3I), 1R(4I), 1R(5I) and 1R(7I); DTLZ5$(2, 30)$—4R(2I), 4R(3I), 5R(4I) and 1R(5I); DTLZ5$(2, 50)$—11R(2I), 2R(3I) and 2R(4I); DTLZ5$(3, 5)$—10R(2I).

[c] DTLZ5$(7, 10)$—3R(2I); WFG3(5)—9R(2I); WFG3(15)—1R(2I).

[d] DTLZ5$(2, 10)$—3R(2I) and 1R(3I); DTLZ5$(2, 20)$–1R(5I); DTLZ5$(2, 30)$—1R(3I); DTLZ5$(2, 50)$—2R(2I).

(ii) different objectives have different degrees of conflict, and (iii) the variance of the objectives on the POF is significantly different. These characteristics imply that the DTLZ1(15) is a redundant problem, and hence an accurate objective reduction algorithm should bring out this fact. That is what NL-MVU-PCA does but L-PCA fails to do. However, as this $\mathcal{N}_\epsilon$ is highly mis-representative of the true POF (Figure 9c), the reduction of objectives by NL-MVU-PCA ironically shows up as its inaccuracy in Table XII. Similarly, the inability of L-PCA to reduce the objectives in a data set like in Figure 9a ironically shows up as its accuracy.

A. Effect of Threshold Parameter

To assess the sensitivity of the proposed algorithms on θ, experiments are also performed with[12] $\theta = 0.682$ and $\theta = 0.954$. Based on the discussions in Section IV-G, a reduction in θ could expectedly have the following effects:
1) For problems with high redundancy: Only a few of the top principal components are likely to reveal the essential objectives. Hence, the deterioration in the performance with a reduction in θ (smaller N_v) should not be too drastic for an accurate algorithm (such as NL-MVU-PCA) when applied on a data set that approximates the true POF with reasonable accuracy (such as $\mathcal{N}_\epsilon$). However, the performance deterioration may be more significant in the case of either an inaccurate algorithm (such as L-PCA) or a poor data set (such as $\mathcal{N}_{\mathcal{NS}}$).
2) For non-redundant problems where all objectives are essential: Reduction in θ (smaller N_v) implies a lesser chance for all the objectives to be picked as essential by the eigenvalue analysis. Hence, the performance deterioration of the algorithms is likely to be more significant than that in the case of redundant problems.

It can be seen that the results presented in Table XIII for $\theta = 0.682$ comply[13] with the expected trend discussed above.

B. Effect of Population Size

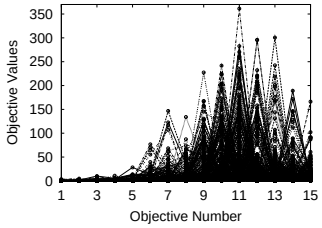
Acknowledging that the approximation of the entire POF for a many-objective problem requires a larger population size (N), this section investigates if the performance of the proposed algorithms improves with an increase in N. Towards it, some of the poor performing cases in Table XII, such as the 15- and 25-objective versions of DTLZ1 and DTLZ3 problems, corresponding to $\mathcal{N}_\epsilon$ are solved with increased $N = 400$. Figure 10a shows a marked improvement in the results for the 15-objective versions along with a lesser improvement in the 25-objective instances. The poor results for these cases, with $N = 200$, were earlier explained through Figure 9 in the context of the disparity in the convergence levels of different objectives. Figure 10b shows that with an increase in N, this disparity has moderated down, leading to improved results. The same trend is observed for other problems, though not shown for brevity.

[12]This is in analogy with Gaussian distributions where $\pm\sigma$ and $\pm 2\sigma$, account for 68.2% and 95.4% variance, respectively. While, the results for $\theta = 0.682$ are presented in Table XIII, those with $\theta = 0.954$ are presented in the supplementary file provided with this paper.

[13]The same holds for the results corresponding to $\theta = 0.954$, presented in the supplementary file provided with this paper.

Problem (M)	NL-MVU-PCA
DTLZ1	
15	16
25	6
DTLZ3	
15	20
25	3

(a) These results indicate the successful cases in 20 random runs



(b) Parallel coordinate plot of DTLZ1(15), for one run.

Fig. 10. Sample results for $\mathcal{N}_\epsilon$ with $N = 400$ and $\theta = 0.997$.

C. Inferences Drawn from the Experimental Results

The experimental results presented above suggest that:

1) The accuracy of the proposed algorithms depends on how closely the given non-dominated solution set represents the true POF. In that: (i) reference to convergence and diversity indicators alone is inadequate, and (ii) how far the given solution set is characterized by the inter-relationships of the objective vectors on the true POF, in terms of their conflict and relative variance, is critical.
2) A high value of θ ($\theta = 0.997$) makes the proposed algorithms more robust in terms of their performance on both the redundant and non-redundant problems.
3) The accuracy of the proposed algorithms increases with an increase in the size of the non-dominated set.
4) The DRP based greedy and exact algorithms, operating on either $\mathcal{N}_\epsilon$ or $\mathcal{N}_{\mathcal{NS}}$, could not accurately identify the true dimension (I) of the POF for the redundant problems (for reasons discussed in the Section VIII). Even in the case of non-redundant problems, the performance of DRP based algorithms was poor, except for the 5-objective problems.

VIII. Comparative Analysis of Different Objective Reduction Algorithms

The results presented above highlight that the accuracy of both L-PCA and NL-MVU-PCA, in terms of identifying an *essential* objective set, is higher than that of the DRP based exact and greedy algorithms. In that, the latter, *with* $\delta = 0$: (i) fail to identify the true dimension of the POF for the redundant DTLZ5(I, M) problems (Table XI), and (ii) over-reduce the number of objectives in the non-redundant problems (except for $M = 5$) as in Table XII. These results conform with those presented in the original sources for the DRP approach, as highlighted in Table XIV.

TABLE XIV
Results for the DRP based greedy algorithm, as reported in [31], [36], [49], for the IBEA population (100 generations)

Test problems	δ	M:Population size		
		5:100	15:200	25:300
DTLZ2	0	5	13	18
DTLZ7	0	5	10	11

The performance of the DRP based algorithms needs to be interpreted in the wake of the goal that the DRP approach pursues and the fundamental limitations that its implementation in [31], [36], [49] suffers from. Following the discussions in Sections II and III, it is clear that the non-dominated solutions obtained from most existing MOEAs can be viewed as a combination of *unnoised* signal (representing the intrinsic dimensionality m) and *noised* signal (which may have dimensionality higher than m). Hence, the DRP approach which aims at preserving the dominance relations will be influenced by the dimension of *noised* signal. This explains why a δ-minimum set with $\delta = 0$ may not coincide with an $\mathcal{F}_\mathcal{T}$. This argument could be realized for DTLZ5$(3, 5)$ in Figures 1c and 7a (where some solutions are conflicting in f_1–f_2–f_3, even though the POF is characterized by non-conflicting f_1–f_2–f_3). Hence, for realistic data sets, the merit of preserving the dominance relations is questionable—if the aim is to identify an $\mathcal{F}_\mathcal{T}$ for a given problem. However, the DRP approach could be useful when the optimal solutions are available (all characterized by the same dimension) and the decision maker is interested in knowing smaller objective sets corresponding to different degrees of error in the dominance relations. While the above argument partly explains the DRP based results reported in this paper, the other reasons could be attributed to the assumptions inherent in the implementation of the DRP approach, discussed below.

A. On the limitations of the DRP based algorithms

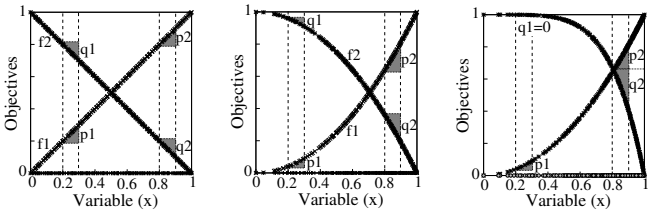
The limitations of DRP based algorithms emanate from the assumptions underlying the definition of δ error. Consider: (i) a certain set (A) of the non-dominated solutions (X) corresponding to the original objective set $\mathcal{F}$, i.e., $A \subset X$, and (ii) $\mathcal{F}' \subset \mathcal{F}$. Then, δ_{max} signifies the maximum error incurred by wrongly assuming that $\vec{x}$ weakly dominates $\vec{y}$ ($\vec{x}, \vec{y} \in A$) with respect to $\mathcal{F}'$ and is given by Equation 10 [49]. In particular, if $\delta_{max} = 0$, it implies that the objectives in $\mathcal{F} \setminus \mathcal{F}'$ are redundant.

$$\delta_{max}(A, \mathcal{F}', \mathcal{F}) = \max_{\substack{\vec{x}, \vec{y} \in A \\ \vec{x} \not\preceq_{\mathcal{F}} \vec{y} \\ \vec{x} \preceq_{\mathcal{F}'} \vec{y}}} \left\{ \max_{f_i \in \mathcal{F}} \{f_i(\vec{x}) - f_i(\vec{y})\} \right\} \quad (10)$$

Equation 10 is based on two stringent assumptions, that:
1) δ error across all the solutions ($\vec{x}, \vec{y} \in A$) is comparable: It is assumed in [36][14] that *an error made close to the Pareto-optimal front is of the same importance as the same error made far away from the Pareto-optimal front*. For this to hold true, the solutions on the non-dominated front need to be *equally* distributed, else the importance of δ in more dense regions which in general signify more important regions, will be higher. It implies that the δ measure is likely to be erroneous in situations where the density of solutions is not equal, such as in distributions like Gaussian (more midrange solutions than at extremes) or non-Gaussian (more extreme solutions than at midrange). This is corroborated in [36] where it is stated (not necessarily in the same sequence) that *situations where the objective function values are not equally distributed are not considered in the study, for example: (i) situations where*

[14]In footnote on Page 143.

the decision maker prefers extremal solutions with maximal objective function values, or (ii) situations where the solutions close to extremal values are more unlikely than ones with midrange values.



(a) $f_1 = x$ and $f_2 = 1 - x$ (b) $f_1 = x^2$ and $f_2 = 1 - x^2$ (c) $f_1 = x^2$ and $f_2 = 1 - x^5$

Fig. 11. Illustration: Unless all the objectives are (a) linear, or (b) identically nonlinear, the DRP approach's assumption of comparable δ across all the objectives, over the entire non-dominated front will not hold. For all three cases, $x \in [0, 1]$. p_i and q_i show the variation in f_1 and f_2, respectively, over different regions of the front.

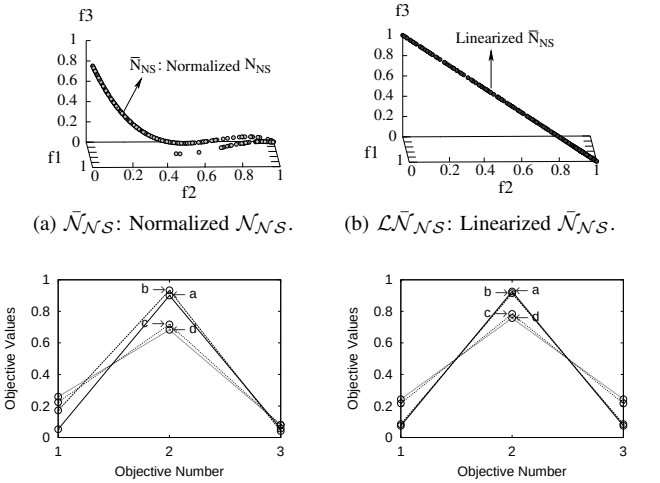
2) δ error is comparable across all the objectives for any two given solutions ($\max_{\vec{x},\vec{y} \in A} \{\max_{f_i \in \mathcal{F}} \{f_i(\vec{x}) - f_i(\vec{y})\}\}$): It is assumed in [36] that *all objective values have the same scale and reference point such that the small errors δ are comparable among the objectives*. This assumption limits the scope of this approach to *linear* objective reduction, because it can hold true only when all the objectives are either linear (Figure 11a) or identically nonlinear (same degree of nonlinearity, Figure 11b). It can be seen in Figure 11c, where the objectives have different degree of nonlinearity, that the δ error across the objectives is not comparable, at any region of the non-dominated front. This is corroborated in [36] where it is stated that *an incorporation of nonlinear objective functions would be extremely useful but remains future work*.

An example problem defined by Equation 11, has been considered below, to demonstrate that the scope of the DRP based algorithms is limited to *linear* objective reduction.

$$\left.\begin{array}{l} f_1 = x^2 + 1; \quad f_2 = -x^2 + x + 3; \quad f_3 = -(f_1 + f_2^3), \\ \text{where } x \in [-2, 2] \end{array}\right\} \quad (11)$$

As the second assumption (above) requires all the objectives to have the same scale and reference points, $\bar{\mathcal{N}}_{\mathcal{NS}}$ (normalized $\mathcal{N}_{\mathcal{NS}}$) corresponding to a population size of 200 and 2000 generations, shown in Figure 12a, is used. Given the disparate degree of nonlinearity in f_3, compared to f_1 and f_2, the assumption that the δ across all the objectives, over the entire non-dominated front is comparable, is violated (as in Figure 11c, but not shown for brevity).

This is a single-variable problem, hence, the number of essential objectives should be equal to two (regularity property [4]). The results in Table XV corresponding to $\bar{\mathcal{N}}_{\mathcal{NS}}$ show that unlike NL-MVU-PCA and L-PCA, the DRP based algorithms fail to identify the redundancy of one objective. To relate the failure of the latter with its limited scope of linear reduction, the $\bar{\mathcal{N}}_{\mathcal{NS}}$ is unfolded or *linearized* using the kernel matrix (K) learnt by the MVU principle. Let the linearized $\bar{\mathcal{N}}_{\mathcal{NS}}$ be referred as $\mathcal{L}\bar{\mathcal{N}}_{\mathcal{NS}}$, where $\mathcal{L}\bar{\mathcal{N}}_{\mathcal{NS}} = K\bar{\mathcal{N}}_{\mathcal{NS}}$. While the $\mathcal{L}\bar{\mathcal{N}}_{\mathcal{NS}}$ is shown in Figure 12b, Table XV shows that the



(a) $\bar{\mathcal{N}}_{\mathcal{NS}}$: Normalized $\mathcal{N}_{\mathcal{NS}}$. (b) $\mathcal{L}\bar{\mathcal{N}}_{\mathcal{NS}}$: Linearized $\bar{\mathcal{N}}_{\mathcal{NS}}$.

(c) Four solutions from $\bar{\mathcal{N}}_{\mathcal{NS}}$. (d) The same solutions as in (c) from $\mathcal{L}\bar{\mathcal{N}}_{\mathcal{NS}}$.

Solutions	$\bar{\mathcal{N}}_{\mathcal{NS}}$			Linearized $\bar{\mathcal{N}}_{\mathcal{NS}}$		
	f_1	f_2	f_3	f_1	f_2	f_3
a	0.053	0.901	0.056	0.074	0.926	0.074
b	0.172	0.934	0.041	0.085	0.915	0.085
c	0.224	0.717	0.080	0.217	0.783	0.217
d	0.259	0.683	0.079	0.243	0.757	0.243

(e) The objective values for the solutions from $\bar{\mathcal{N}}_{\mathcal{NS}}$ and $\mathcal{L}\bar{\mathcal{N}}_{\mathcal{NS}}$.

Fig. 12. An example to illustrate that DRP based approach offers *linear* objective reduction. The plots are for one NSGA-II run.

DRP based algorithms applied to $\mathcal{L}\bar{\mathcal{N}}_{\mathcal{NS}}$ identify the presence of a redundant objective. For a deeper understanding of the effect of data *linearization* on the DRP based algorithms, four solutions picked from $\bar{\mathcal{N}}_{\mathcal{NS}}$ and their *linearized* counterparts in $\mathcal{L}\bar{\mathcal{N}}_{\mathcal{NS}}$, are shown in Figures 12c, 12d and 12e. It can be verified that while the dominance relations are violated on elimination of f_3 in the case of four $\bar{\mathcal{N}}_{\mathcal{NS}}$ solutions, they remain preserved in the case of the $\mathcal{L}\bar{\mathcal{N}}_{\mathcal{NS}}$ solutions.

TABLE XV
RESULTS FOR THE EXAMPLE PROBLEM IN EQUATION 11, CORRESPONDING TO 20 RUNS OF NSGA-II. THE ESSENTIAL OBJECTIVE SETS OBTAINED ARE SHOWN, WITH THE FREQUENCY OF OCCURRENCE BELOW

Data	Proposed approaches		DRP [31], [36] based δ-MOSS, 0% Error	
	NL-MVU-PCA	L-PCA	Exact	Greedy
$\bar{\mathcal{N}}_{\mathcal{NS}}$	$\{f_1, f_2\}$ (20)	$\{f_1, f_2\}$ (20)	$\{f_1, f_2, f_3\}$ (20)	$\{f_1, f_2, f_3\}$ (20)
$\mathcal{L}\bar{\mathcal{N}}_{\mathcal{NS}}$	$\{f_1, f_2\}$ (20)	$\{f_1, f_2\}$ (20)	$\{f_1, f_2\}$ (20)	$\{f_1, f_2\}$ (20)

This example establishes that while the dominance relation preservation approach is promising, the manner in which it is implemented in [31], [36], [49] by ignoring the effect of nonlinearity leads to erroneous results. This basically links to the use of a distance measure that is inadequate to account for nonlinearity (as discussed in Section II-A). The current implementation relies on the absolute difference (δ) in the values of two solutions on a *particular* objective vector.

The assumptions stated above basically define the situations (*linearity*) in which this δ may be valid.

B. On the computational complexities of different algorithms

It is also important to compare the objective reduction approaches based on their computational complexity, summarized in Table XVI. Focusing on the DRP approach it can be seen that: (i) the exact algorithm is almost impractical to use, since it is exponential in M and quadratic in N; and (ii) given that, in the case of many-objective problems $N \gg M$, even the greedy algorithm is likely to be more expensive than NL-MVU-PCA (worst case being $O(M^6)$, with $k = M - 1$).

TABLE XVI
COMPUTATIONAL COMPLEXITY OF OBJECTIVE REDUCTION ALGORITHMS. N IS THE SIZE OF THE NON-DOMINATED SET; N_g THE GENERATIONS FOR AN MOEA; M THE NUMBER OF OBJECTIVES

Approaches	Computational complexity
A. Dom. rel. preservation (δ-MOSS)	
(i) Exact Algorithm	$O(N^2M2^M)$
(ii) Greedy Algorithm	$O(min\{N^2M^3, N^4M^2\})$
B. Unsupervised feature selection[a]	$O(NM^2)+$ clustering overhead
C. Removal of data dependencies	
(i) PCA based reduction[a]	$O(NM^2 + M^3)$
(ii) MVU-PCA based reduction	$O(M^3q^3)$ where q is the *neighborhood* size[b]

[a] In [32], [50] and [51], the complexity of clustering has not been included. Furthermore, the complexity of PCA based reduction is incorrectly cited as $O(NM^2 + M^3 + N^2MN_g)$. Notably, each of the above approaches operate on the non-dominated set, hence, their is no rationale for adding the computational complexity of obtaining this non-dominated set, only to the complexity of PCA based reduction ($O(N^2MN_g)$ for NSGA-II).
[b] In the most constrained case, $q = O(M)$.

To conclude, the following remarks can be made:

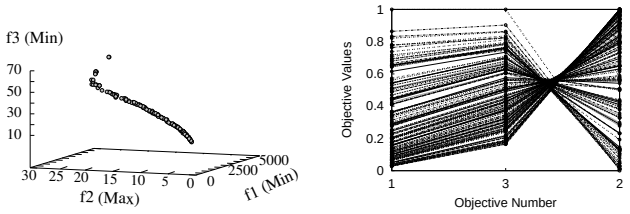
- Unlike the NL-MVU-PCA proposed in this paper, the scope of the DRP based algorithms is limited to *linear* objective reduction and *equally* spaced solutions. These are major limitations because the real-world problems may have objectives with different degree of nonlinearity, in which case the results based on *linear* objective reduction could be misleading. Furthermore, it is unlikely that the non-dominated solutions obtained from an MOEA in the case of many-objective problems will be *equally* spaced (as seen in Figure 1c). However, any departure from *linearity* or *equally* spaced solutions is a matter of *degree* and the performance of the these algorithms will depend on the interplay of these two factors. For example, these algorithms: (i) could solve the 5-objective instances of the non-redundant problems (Table XII) but not those with $M > 5$, and (ii) the performance in DTLZ1(15) with *linear* POF is better than DTLZ2(15) with *nonlinear* POF.
- The use of the DRP based exact algorithm which guarantees a δ-minimum objective set (subject to its assumptions holding true) becomes impractical for large problems, owing to its computational complexity. Furthermore, the greedy algorithm does not guarantee the δ-minimum objective set for a given problem. In other words, the aim of finding the δ-minimum objective set based on dominance relation preservation is practically difficult to realize for large problems, which is also the limitation of the algorithms proposed in this paper.

IX. REAL-WORLD PROBLEMS

This section considers two real-world problems.

A. Multi-Speed Gearbox Design Problem

This is a multi-speed gearbox design problem, comprising of three objectives. It does not belong to the many-objective domain, yet, it is included here for a comparative analysis of the proposed L-PCA and NL-MVU-PCA vis-à-vis the DRP based algorithms. While the problem formulation[15] can be found in [52], it may be noted that the three objectives relate to (i) minimization of (f_1) overall volume of gear material used (which is directly related to the weight and cost of the gearbox), (ii) maximization of (f_2) power delivered by the gearbox, and (iii) minimization of (f_3) the center distance between input and output shafts.



(a) Non-dominated front (b) Parallel-coordinate plot (order of f_2 and f_3 is swapped)

Fig. 13. Multi-speed gearbox design problem. The plots correspond to one run of NSGA-II.

The $\mathcal{N}_{\mathcal{NS}}$ corresponding to a population size of 200 and 2000 generations is as shown in Figure 13a. Here, it can be seen that f_1 and f_3 are non-conflicting among themselves, while each is in conflict with f_2. This is also affirmed by Figure 13b. These observations are physically justifiable because for a fixed number of gears, the lower the center distance between the input and output shafts (f_3), the smaller the size of each gear will be, resulting in lower overall volume of gear material (f_1). It can be seen from the Table XVII that while L-PCA and NL-MVU-PCA identify the redundancy in the problem, the DRP based exact and greedy algorithms fail to capture it. This can be explained through Figure 15a. In that, while L-PCA and NL-MVU-PCA exploit the positive correlation between f_1 and f_3 for objective reduction, the DRP based algorithms fail owing to the nonlinearity (discussed in detail, in Section VIII).

B. Storm Drainage System Problem

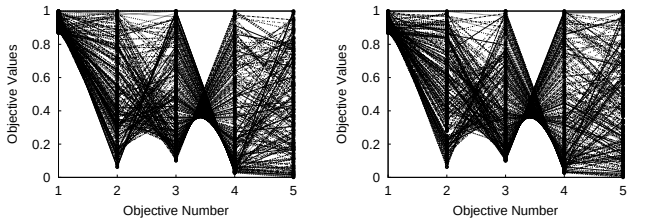
Originally described in [53], this is a five-objective, seven-constraint problem which relates to optimal planning for a storm drainage system in an urban area. The results for this problem, presented in Table XVII, show that all the algorithms identify either f_1 or f_3 as redundant. Both these sets of results are correct because: (i) f_1 and f_3 are positively

[15] The version where the gear thickness, number of teeth, power and module are all kept as variables.

TABLE XVII
PERFORMANCE OF OBJECTIVE REDUCTION ALGORITHMS ON TWO REAL-WORLD APPLICATION PROBLEMS. THESE RESULTS CORRESPOND TO 20 NSGA-II RUNS WITH UNIFORMLY DISTRIBUTED SEEDS; EACH RUN CORRESPONDING TO 200 POPULATION SIZE AND 2000 GENERATIONS

	Proposed approaches		Dominance relation preservation [31], [36]: δ-MOSS, 0% Error	
Real-world Problems	NL-MVU-PCA	L-PCA	Greedy Approach	Exact Approach
(a) Multi-speed gearbox[a]	$\{f_1, f_2\}$	$\{f_1, f_2\}$	$\{f_1, f_2, f_3\}$	$\{f_1, f_2, f_3\}$
(b) Storm drainage system[b]	$\{f_2, f_3, f_4, f_5\}$	$\{f_1, f_2, f_4, f_5\}$[c]	$\{f_2, f_3, f_4, f_5\}$	$\{f_2, f_3, f_4, f_5\}$

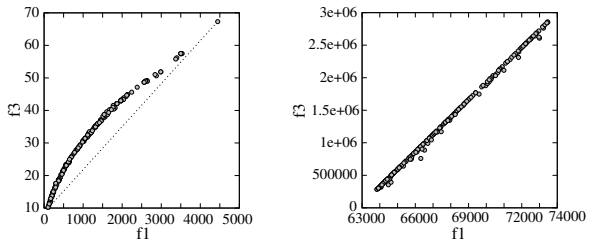
[a] The error associated with NL-MVU-PCA and L-PCA is 0.00262 ± 0.00035 and 0.00559 ± 0.00074, respectively.
[b] The error associated with NL-MVU-PCA and L-PCA is 0.00002 ± 0.00001 and 0.00021 ± 0.00012, respectively.
[c] In 18 out of the 20 runs, L-PCA finds $\mathcal{F}_s = \{f_1, f_2, f_4, f_5\}$, while twice it finds $\mathcal{F}_s = \{f_2, f_3, f_4, f_5\}$.



(a) $\mathcal{N}_{\mathcal{NS}}$ with all five objectives (b) f_1 reconstructed from $\mathcal{N}_{\mathcal{NS}}$ obtained with $\{f_2, f_3, f_4, f_5\}$

Fig. 14. Storm drainage problem: Parallel coordinate plots (normalized), corresponding to one run of NSGA-II.

correlated (Figure 15b), and (ii) $\mathcal{N}_{\mathcal{NS}}$ obtained with either $\{f_2, f_3, f_4, f_5\}$ or $\{f_1, f_2, f_4, f_5\}$ conforms with that for the original problem (for brevity, only the former's conformance is shown in Figure 14).



(a) Nonlinearity in f_1–f_3 subspace of gearbox design problem (b) Linearity in f_1–f_3 subspace of storm drainage problem

Fig. 15. Highlighting the nonlinear and linear characteristics of the two real-world problems. These plots correspond to one run of NSGA-II.

The success of the DRP based algorithms can be attributed to the linearity of this problem, as evident in Figure 15b (consistent with the inferences from Section VIII). Furthermore, among the positively correlated f_1 and f_3, NL-MVU-PCA always picks f_3 as important owing to its larger variance (Figure 14) than f_1–a fact that L-PCA could identify only twice out of the 20 runs.

X. CONCLUSIONS

This paper has proposed a framework for both linear and nonlinear objective reduction algorithms, namely, L-PCA and NL-MVU-PCA, respectively. The performance of these algorithms has been studied for 30 test instances (from both redundant and non-redundant problems) and for two real-world problems. The sensitivity of these algorithms on the critical parameters involved and on the quality of the underlying non-dominated sets has been discussed. An error measure has also been proposed to assess the obtained results. A comparative analysis of the proposed algorithms vis-à-vis the dominance relation preservation based algorithms, has been done in terms of their scope of application and the computational complexity. For future work, the endeavor of the authors will be to demonstrate the application of the proposed algorithms for online objective reduction, δ-MOSS problems with different δs and on k-EMOSS problems. The future work will also focus on testing the framework on more real-world problems and problems with high-dimensional and complicated Pareto-set shapes.

REFERENCES

[1] V. Khare, X. Yao, and K. Deb, "Performance scaling of multi-objective evolutionary algorithms," in *Proceedings of the Second Evolutionary Multi-Criterion Optimization (EMO-03) Conference (LNCS 2632)*, 2003, pp. 376–390.
[2] E. Hughes, "Evolutionary many-objective optimisation: many once or one many?" in *IEEE Congress on Evolutionary Computation*, vol. 1. IEEE, Sept. 2005, pp. 222–227.
[3] J. Knowles and D. W. Corne, "Quantifying the effects of objective space dimension in evolutionary multiobjective optimization," in *Proceedings of the Fourth International Conference on Evolutionary Multi-Crietrion Optimization*, S. Obayashi, K. Deb, C. Poloni, T. Hiroyasu, and T. Murata, Eds. Matshushima, Japan: Springer. Lecture Notes in Computer Science Vol. 4403, 2007, pp. 757–771.
[4] Q. Zhang, A. Zhou, and Y. Jin, "Rm-meda: A regularity model-based multiobjective estimation of distribution algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 1, pp. 41–63, February 2008.
[5] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A fast and elitist multi-objective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, Apr. 2002.
[6] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength pareto evolutionary algorithm for multiobjective optimization," in *Evolutionary Methods for Design, Optimization, and Control*, K. Giannakoglou, D. Tsahalis, J. Periaux, K. Papailiou, and T. Fogarty, Eds. CIMNE, Barcelona, Spain, 2002, pp. 19–26.
[7] J. D. Knowles and D. W. Corne, "Approximating the nondominated front using the pareto archived evolution strategy," in *Evolutionary Computation*, vol. 8, no. 2, 2000, pp. 149–172.
[8] Q. Zhang and H. Li, "MOEA/D: A multi-objective evolutionary algorithm based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712–731, December 2007.
[9] H. Ishibuchi, Y. Hitotsuyanagi, H. Ohyanagi, and Y. Nojima, "Effects of the existence of highly correlated objectives on the behavior of moea/d," in *Evolutionary Multi-Criterion Optimization*, ser. Lecture Notes in Computer Science, R. Takahashi, K. Deb, E. Wanner, and S. Greco, Eds. Springer Berlin / Heidelberg, 2011, vol. 6576, pp. 166–181.
[10] E. Zitzler and S. Künzli, "Indicator-based selection in multiobjective search," in *PPSN VIII*, vol. 3242. Springer LNCS, 2004, pp. 832–842.
[11] T. Wagner, N. Beume, and B. Naujoks, "Pareto-, aggregation-, and indicator-based methods in many-objective optimization," in *Proc. 4th Int. Conf. Evol. Multi-Criterion Optimization*. New York: Springer Berlin / Heidelberg, Mar. 2007, vol. 4403, pp. 742–756.
[12] E. Zitzler, "Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications," Ph.D. dissertation, ETH Zurich, Switzerland, 1999.

[13] L. While, P. Hingston, L. Barone, and S. Huband, "A faster algorithm for calculating hypervolume," *Evolutionary Computation, IEEE Transactions on*, vol. 10, no. 1, pp. 29 – 38, Feb. 2006.

[14] C. Fonseca, L. Paquete, and M. Lopez-Ibanez, "An improved dimension-sweep algorithm for the hypervolume indicator," in *Evolutionary Computation, 2006. CEC 2006. IEEE Congress on*, July 2006, pp. 1157 – 1163.

[15] N. Beume, "S-metric calculation by considering dominated hypervolume as klee's measure problem," *Evolutionary Computation*, vol. 17, no. 4, pp. 477–492, 2009.

[16] A. V. Lotov, V. A. Bushenkov, and G. K. Kamenev, *Interactive Decision Maps*. Boston, MA: Kluwer, 2004.

[17] S. Obayashi and D. Sasaki, "2003, visualization and data mining of pareto solutions using self-organizing map," in *Deb and REFERENCES 202 L. Thiele (eds), Proceedings of the Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003*. Springer, 2003, pp. 796–809.

[18] H. Li and Q. Zhang, "Multiobjective Optimization Problems With Complicated Pareto Sets, MOEA/D and NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 2, pp. 284–302, April 2009.

[19] H. Sato, H. E. Aguirre, and K. Tanaka, *Controlling Dominance Area of Solutions and Its Impact on the Performance of MOEAs*. Springer Berlin Heidelberg, May 2007, vol. 4403/2007, ch. 5, pp. 5–20.

[20] N. Drechsler, R. Drechsler, and B. Becker, "Multi-objective optimisation based on relation favour," in *Proc. 1st Int. Conf. Evol. Multi-Criterion Optimization*, ser. Lecture Notes in Computer Science. New York: Springer Berlin / Heidelberg, 2001, vol. 1993, pp. 154–166.

[21] D. W. Corne and J. D. Knowles, "Techniques for highly multiobjective optimisation: some nondominated points are better than others," in *Genetic and Evolutionary Computation Conference (GECCO)*, 2007, pp. 773–780.

[22] S. Kukkonen and J. Lampinen, "Ranking-dominance and many-objective optimization," in *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on*. IEEE, Sep. 2007, pp. 3983–3990.

[23] M. Köppen and K. Yoshida, "Substitute Distance Assignments in NSGA-II for Handling Many-objective Optimization Problems," in *Evolutionary Multi-Criterion Optimization*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, vol. 4403, ch. 55, pp. 727–741.

[24] H. Ishibuchi, T. Doi, and Y. Nojima, "Incorporation of scalarizing fitness functions into evolutionary multiobjective optimization algorithms," in *Parallel Problem Solving From Nature IX (PPSN-IX)*. Springer Berlin / Heidelberg, Sep. 2006, vol. 4193, pp. 493–502.

[25] H. Ishibuchi, N. Tsukamoto, and Y. Nojima, "Iterative approach to indicator-based multiobjective optimization," in *IEEE Congress on Evolutionary Computation (CEC 2007)*. IEEE, Sep. 2007, pp. 3967–3974.

[26] J. Branke and K. Deb., "Integrating user preferences into evolutionary multi-objective optimization," in *Knowledge Incorporation in Evolutionary Computation*, Y. Jin, Ed. Hiedelberg, Germany: Springer, 2004, pp. 461–477.

[27] P. J. Fleming, R. C. Purshouse, and R. J. Lygoe, "Many-objective optimization: An engineering design perspective," in *Evolutionary Multi-Criterion Optimization-EMO*. Springer Verlag, 2005, pp. 14–32.

[28] K. Deb and J. Sundar, "Preference point based multi-objective optimization using evolutionary algorithms," in *Genetic and Evolutionary Computation Conference (GECCO)*, 2006, pp. 635–642.

[29] L. Thiele, K. Miettinen, P. J. Korhonen, and J. Molina, "A preference-based interactive evolutionary algorithm for multiobjective optimization," Helsinki School of Economics, Working Paper W-412, 2007.

[30] H. Ishibuchi, N. Tsukamoto, and Y. Nojima, "Evolutionary many-objective optimization: A short review," in *IEEE Congress on Evolutionary Computation*, 2008, pp. 2424–2431.

[31] D. Brockhoff and E. Zitzler, "Are all objectives necessary? on dimensionality reduction in evolutionary multiobjective optimization," in *Conference on Parallel Problem Solving from Nature (PPSN IX)*, ser. LNCS, T. P. Runarsson *et al.*, Eds., vol. 4193. Berlin, Germany: Springer, 2006, pp. 533–542.

[32] A. L. Jaimes, C. A. C. Coello, and D. Chakraborty, "Objective reduction using a feature selection technique," in *Genetic and Evolutionary Computation Conference (GECCO)*, 2008, pp. 673–680.

[33] 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. 99, pp. 1–18, 2011.

[34] K. Deb and D. K. Saxena, "Searching for pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems," in *IEEE Congress on Evolutionary Computation*, 2006, pp. 3353–3360.

[35] D. K. Saxena and K. Deb, "Non-linear Dimensionality Reduction Procedures for Certain Large-Dimensional Multi-objective Optimization Problems: Employing Correntropy and a Novel Maximum Variance Unfolding," in *Evolutionary Multi-Criterion Optimization, 4th International Conference, EMO 2007*, S. Obayashi, K. Deb, C. Poloni, T. Hiroyasu, and T. Murata, Eds. Matshushima, Japan: Springer. Lecture Notes in Computer Science Vol. 4403, 2007, pp. 772–787.

[36] D. Brockhoff and E. Zitzler, "Objective Reduction in Evolutionary Multiobjective Optimization: Theory and Applications," *Evolutionary Computation*, vol. 17, no. 2, pp. 135–166, 2009.

[37] J. Shlens, "A tutorial on principal component analysis," Center for Neural Science, New York University, available at:http://www.snl.salk.edu/∼shlens/pca.pdf (accessed: May 2011), Tech. Rep., April 2009.

[38] B. Scholkopf, A. Smola, and K. R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[39] L. K. Saul, K. Q. Weinberger, J. H. Ham, F. Sha, and D. D. Lee, "Spectral methods for dimensionality reduction," in *Semisupervised Learning*, O. C. B. Schoelkopf and A. Zien, Eds. MIT Press, Cambridge, MA, 2006.

[40] K. Q. Weinberger and L. K. Saul, "Unsupervised learning of image manifolds by semidefinite programming," *International Journal of Computer Vision*, vol. 70, no. 1, pp. 77–90, 2006.

[41] J.F.Sturm, "Using sedumi 1.02, a matlabtoolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11, no. 1, pp. 625–653, 1999.

[42] B.Borchers, "Csdp, a c library for semidefinite programming," *Optimization Methods and Software*, vol. 11, no. 1, pp. 613–623, 1999.

[43] J. Cohen, *Statistical power analysis for the behavioral sciences*. 2nd ed, Hillsdale, NJ: Erlbaum, 1988.

[44] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable Test Problems for Evolutionary Multi-Objective Optimization," in *Evolutionary Multiobjective Optimization: Theoretical Advances and Applications*, A. Abraham, R. Jain, and R. Goldberg, Eds. Springer, 2005, pp. 105–145.

[45] 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.

[46] K. Deb, M. Mohan, and S. Mishra, "A fast multi-objective evolutionary algorithm for finding well-spread pareto-optimal solutions," Kanpur Genetic Algorithms Laboratory, Indian Institute of Technology Kanpur, KanGAL Report 2003002, 2003.

[47] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler, "Combining convergence and diversity in evolutionary multi-objective optimization," in *Evolutionary Computation*, vol. 10, no. 3, 2002, pp. 263–282.

[48] R. C. Purshouse and P. J. Fleming, "Evolutionary many-objective optimization: An exploratory analysis," in *IEEE Congress on Evolutionary Computation*, 2003, pp. 2066–2073.

[49] D. Brockhoff, "Many-Objective Optimization and Hypervolume Based Search," Ph.D. dissertation, ETH Zurich, Aachen, Germany, 2009.

[50] A. L. Jaimes and C. A. C. Coello, "Some techniques to deal with many-objective problems," in *Genetic and Evolutionary Computation Conference (GECCO)*, 2009, pp. 2693–2696.

[51] A. L. Jaimes, C. A. C. Coello, and J. E. U. Barrientos, "Online objective reduction to deal with many-objective problems," in *Evolutionary Multi-Criterion Optimization-EMO*. Springer Verlag, 2009, pp. 423–437.

[52] P. Jain and A. Agogino, "Theory of design: An optimization perspective," *Mech. Mach. Theory*, vol. 23, no. 3, 1990.

[53] K. Musselman and J. Talavage, "A trade-off cut approach to multiple objective optimization," *Operations Research*, vol. 28, no. 6, pp. 1424–1435, 1980.

Supplement on Parameter Sensitivity, Data Quality and Resulting Errors

TABLE I
DERIVATIONS FOR $\mathcal{D}_{\mathcal{T}}$ (DISPERSAL OF SOLUTIONS ON THE TRUE POF) FOR THE USED TEST PROBLEMS. $\mathcal{D}_{\mathcal{T}}$ ALLOWS FOR THE COMPUTATION OF NORMALIZED MAXIMUM SPREAD INDICATOR ($I_s = \mathcal{D}_{\mathcal{A}}/\mathcal{D}_{\mathcal{T}}$)

Problem	POF Condition	f^{max}	f^{min}	$\mathcal{D}_{\mathcal{T}}^2 = \sum_{i=1}^{M} \left(f_i^{max} - f_i^{min}\right)^2$
DTLZ1	$g = 0$	$f_i = \frac{1}{2},\ \forall i = 1, \ldots, M$	$f_i = 0,\ \forall i = 1, \ldots, M$	$\sum_{i=1}^{M} \left(\frac{1}{2}\right)^2 = 0.25M$
DTLZ2 DTLZ3 DTLZ4	$g = 0$	$f_i = 1,\ \forall i = 1, \ldots, M$	$f_i = 0,\ \forall i = 1, \ldots, M$	$\sum_{i=1}^{M} 1 = M$
DTLZ5(I,M)	$g = 0$	$f_1 = \left(\frac{1}{\sqrt{2}}\right)^{M-I}$ $f_i = \left(\frac{1}{\sqrt{2}}\right)^{M-I+2-i},\ \forall i = 2, \ldots, M-I+1$ $f_i = 1,\ \forall i = M-I+2, \ldots, M$	$f_i = 0,\ \forall i = 1, \ldots, M$	Note[a]
DTLZ7	$g = 1$	$f_i = 1,\ \forall i = 1, \ldots, M-1$ $f_M = 2M$	$f_i = 0,\ \forall i = 1, \ldots, M-1$ $f_M = 2M - t_{max}$	$M - 1 + t_{max}^2$, Note[b]

[a] $\left(\frac{1}{2}\right)^{M-I} + \sum_{i=2}^{M-I+1} \left(\frac{1}{2}\right)^{M-I+2-i} + \sum_{i=1}^{I-1} 1$
[b] $t = \sum_{i=1}^{M-1} x_i(1 + \sin(3\pi x_i))$ and $t_{max} \approx 0.8594009(M-1)$

TABLE II
ERRORS (E_r) ASSOCIATED WITH NL-MVU-PCA AND L-PCA FOR REDUNDANT AND NON-REDUNDANT TEST PROBLEMS, WITH $\theta = 0.997$. THE MEAN (μ) AND STANDARD DEVIATION (σ) IS OBTAINED FOR 20 RANDOM RUNS

Test problems		NL-MVU-PCA		L-PCA	
DTLZ5(I,M)[a]		$\mathcal{N}_\epsilon$	$\mathcal{N}_{NS}$	$\mathcal{N}_\epsilon$	$\mathcal{N}_{NS}$
I	M	$(\mu \pm \sigma)$	$(\mu \pm \sigma)$	$(\mu \pm \sigma)$	$(\mu \pm \sigma)$
2	5	0.00747 ± 0.00435	0.01541 ± 0.01271	0.05032 ± 0.02293	0.03306 ± 0.02591
2	10	0.00741 ± 0.00391	0.01813 ± 0.01805	0.05589 ± 0.02835	0.01387 ± 0.02214
2	20	0.00956 ± 0.00481	0.02437 ± 0.02425	0.05882 ± 0.02542	0.01173 ± 0.01836
2	30	0.01511 ± 0.00884	0.05226 ± 0.02069	0.07751 ± 0.04151	0.08508 ± 0.03425
2	50	0.01565 ± 0.01095	0.06615 ± 0.02003	0.10391 ± 0.04306	0.21139 ± 0.05023
3	5	0.00113 ± 0.00085	0.01382 ± 0.00566	0.00462 ± 0.00289	0.00000 ± 0.00000
3	10	0.00307 ± 0.00202	0.00000 ± 0.00000	0.03261 ± 0.01764	0.00000 ± 0.00000
3	20	0.00547 ± 0.00394	0.00020 ± 0.00054	0.05459 ± 0.02273	0.00044 ± 0.00113
5	10	0.00159 ± 0.00092	0.00000 ± 0.00000	0.01454 ± 0.00624	0.00000 ± 0.00000
5	20	0.00258 ± 0.00127	0.00000 ± 0.00000	0.02955 ± 0.01351	0.00000 ± 0.00000
7	10	0.00179 ± 0.00086	0.00000 ± 0.00000	0.01052 ± 0.00465	0.00000 ± 0.00000
7	20	0.00218 ± 0.00077	0.00000 ± 0.00000	0.03391 ± 0.01248	0.00000 ± 0.00000
WFG3[a]	5	0.00725 ± 0.00114	0.00261 ± 0.00074	0.09986 ± 0.04019	0.03722 ± 0.01396
	15	0.00962 ± 0.00296	0.00293 ± 0.00112	0.24881 ± 0.18592	0.21041 ± 0.05908
	25	0.01002 ± 0.00511	0.00266 ± 0.00177	0.13044 ± 0.09724	0.19117 ± 0.05063
DTLZ1[b]	5	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
	15	0.00058 ± 0.00070	0.00000 ± 0.00000	0.00167 ± 0.00321	0.00000 ± 0.00000
	25	0.00140 ± 0.00077	0.00000 ± 0.00000	0.01728 ± 0.01241	0.00000 ± 0.00000
DTLZ2	5	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
	15	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
	25	0.00002 ± 0.00007	0.00000 ± 0.00000	0.00044 ± 0.00107	0.00000 ± 0.00000
DTLZ3	5	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
	15	0.00022 ± 0.00033	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
	25	0.00116 ± 0.00045	0.00000 ± 0.00000	0.01167 ± 0.01102	0.00000 ± 0.00000
DTLZ4	5	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
	15	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
	25	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
DTLZ7	5	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
	15	0.00028 ± 0.00123	0.00236 ± 0.00692	0.00000 ± 0.00000	0.00000 ± 0.00000
	25	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000

[a] Redundant Problems: In Section VI.D of the main paper, it has been discussed that the error measures could be reliable only if the following conditions are simultaneously met: (i) the objective reduction algorithm is accurate, and (ii) the solution set on which an algorithm operates is representative of the true POF. In the current context, four different combinations of accurate/inaccurate algorithms and good/bad solution sets are possible. For an accurate–good combination like NL-MVU-PCA operating on $\mathcal{N}_\epsilon$, the error measures should be reliable. This can be seen to be true above. Other error measures need to be interpreted in the light of algorithm–population combinations and the results in Table XI.

[b] Non-Redundant Problems: In Section VII of the main paper, it has been discussed that $\mathcal{N}_\epsilon$ for these problems is highly mis-representative of the true POF. In that, it shows characteristics of redundant problems, given which NL-MVU-PCA reduces the number of objectives, while L-PCA fails to. Ironically, this reflects as the inefficiency of NL-MVU-PCA and efficiency of L-PCA in Table XII of the main paper. The above error measures need to be interpreted in the wake of this fact.

TABLE III

EFFECT OF THE VARIANCE THRESHOLD ($\theta = 0.954$, WITH $q = M - 1$) AND NEIGHBORHOOD SIZE ($q = \sqrt{M}$, WITH $\theta = 0.997$) ON THE PERFORMANCE OF THE PROPOSED L-PCA AND NL-MVU-PCA ALGORITHMS ON REDUNDANT TEST PROBLEMS, CORRESPONDING TO $\mathcal{N}_{\mathcal{NS}}$ AND $\mathcal{N}_\epsilon$. THE NUMBERS IN THE TABLE INDICATE THE FREQUENCY OF SUCCESS IN IDENTIFYING THE TRUE I AND $\mathcal{F}_\mathcal{T}$, OUT OF 20 RUNS. THE DASHES (-) REPLACE 0 TO IMPLY INCONSEQUENTIAL ENTRIES, AS THE PREREQUISITE I IS NOT MET

Test problems		$\theta = 0.954$ and $q = M - 1$								$\theta = 0.997$ and $q = \sqrt{M}$							
		NL-MVU-PCA				L-PCA				NL-MVU-PCA				L-PCA			
DTLZ5(I, M)		$\mathcal{N}_\epsilon$		$\mathcal{N}_{\mathcal{NS}}$		$\mathcal{N}_\epsilon$		$\mathcal{N}_{\mathcal{NS}}$		$\mathcal{N}_\epsilon$		$\mathcal{N}_{\mathcal{NS}}$		$\mathcal{N}_\epsilon$		$\mathcal{N}_{\mathcal{NS}}$	
I	M	I[a]	$\mathcal{F}_\mathcal{T}$	I[b]	$\mathcal{F}_\mathcal{T}$	I[c]	$\mathcal{F}_\mathcal{T}$	I	$\mathcal{F}_\mathcal{T}$	I[d]	$\mathcal{F}_\mathcal{T}$	I[e]	$\mathcal{F}_\mathcal{T}$	I[f]	$\mathcal{F}_\mathcal{T}$	I	$\mathcal{F}_\mathcal{T}$
2	5	20	20	20	20	20	4	20	1	20	20	20	20	20	14	20	1
2	10	20	20	11	11	20	0	0	-	20	20	15	14	20	7	0	-
2	20	20	20	7	7	20	0	0	-	20	20	18	16	20	2	0	-
2	30	20	20	14	14	20	1	0	-	20	20	17	17	20	1	0	-
2	50	20	20	14	14	20	0	9	0	20	20	18	18	20	0	10	0
3	5	20	20	18	18	20	8	0	-	20	20	20	20	20	9	0	-
3	10	20	20	0	-	20	4	0	-	20	20	1	0	20	2	0	-
3	20	20	20	0	-	20	0	0	-	20	20	0	-	20	1	0	-
5	10	19	19	0	-	20	2	0	-	20	19	0	-	19	3	0	-
5	20	20	20	0	-	19	0	0	-	20	19	0	-	18	0	0	-
7	10	19	19	0	-	17	4	0	-	20	16	0	-	20	6	0	-
7	20	16	16	0	-	16	3	0	-	16	16	0	-	13	3	0	-
	5	20	20	20	20	20	19	20	3	20	20	20	20	20	19	20	5
WFG3	15	15	15	20	20	9	0	19	0	16	16	20	20	10	0	19	0
	25	9	9	20	20	4	1	19	0	11	11	20	20	6	0	20	0

[a] DTLZ5(2, 30)—1R (2I); DTLZ5(5, 10)—3R (2I); DTLZ5(5, 20)—6R (2I); DTLZ5(7, 10)—7R (2I); DTLZ5(7, 20)—8R (2I) and 1R (4I); WFG3(3, 15)—6R (2I) and 3R (3I); WFG3(3, 25)—5R (2I), 1R (3I) and 1R (4I).

[b] DTLZ5(2, 10)—4R (2I), 5R (3I) and 3R (4I); DTLZ5(2, 20)—1R (2I), 2R (4I), 2R (5I), 1R (6I), 1R (7I); DTLZ5(2, 30)—3R (2I), 3R (3I), 4R (4I) and 4R (5I); DTLZ5(2, 50)—8R (2I) and 2R (3I); DTLZ5(3, 5)—9R (2I).

[c] DTLZ5(7, 20)—2R (2I); WFG3(5)—11R (2I); WFG3(15)—1R (3I).

[d] DTLZ5(2,30)—1R (2I); DTLZ5(5,10)—3R (2I); DTLZ5(5,20)—8R (2I); DTLZ5(7,10)—4R (2I); DTLZ5(7,20)—9R (2I) and 1R (4I); WFG3(15)—7R (2I) and 2R (3I); WFG3(25)—3R (3I) and 1R (4I).

[e] DTLZ5(2,10)—9R (2I), 4R (3I) and 1R (4I); DTLZ5(2,20)—2R (2I), 7R (3I) and 7R (4I); DTLZ5(2,30)—3R (2I), 5R (3I), 5R (4I), 2R (5I) and 2R (6I); DTLZ5(2,50)—11R (2I), 5R (3I), 1R (4I) and 1R (6I).

[f] DTLZ5(7,10)—2R (2I); DTLZ5(7,20)—1R (2I); WFG3(5)—11R (2R).

TABLE IV

EFFECT OF THE VARIANCE THRESHOLD ($\theta = 0.954$, WITH $q = M - 1$) AND NEIGHBORHOOD SIZE ($q = \sqrt{M}$, WITH $\theta = 0.997$) ON THE PERFORMANCE OF THE PROPOSED L-PCA AND NL-MVU-PCA ALGORITHMS ON NON-REDUNDANT TEST PROBLEMS, CORRESPONDING TO $\mathcal{N}_{\mathcal{NS}}$ AND $\mathcal{N}_\epsilon$. THE NUMBERS IN THE TABLE INDICATE THE FREQUENCY OF SUCCESS IN IDENTIFYING THE $\mathcal{F}_\mathcal{T}$, OUT OF 20 RUNS. THE DASHES (-) REPLACE 0 TO IMPLY INCONSEQUENTIAL ENTRIES, AS THE PREREQUISITE I IS NOT MET

Test Problems		$\theta = 0.954$ and $q = M - 1$				$\theta = 0.997$ and $q = \sqrt{M}$			
DTLZ(M)		NL-MVU-PCA		L-PCA		NL-MVU-PCA		L-PCA	
Name	M	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{NS}}$	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{NS}}$	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{NS}}$	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{NS}}$
DTLZ1	5	20	18	20	20	20	19	20	20
	15	6	20	15	20	4	18	15	20
	25	0	20	4	20	0	20	4	20
DTLZ2	5	20	20	19	20	20	20	19	20
	15	20	20	20	20	17	18	20	20
	25	16	20	16	20	16	17	17	20
DTLZ3	5	19	20	20	20	20	20	20	20
	15	11	20	18	20	9	14	18	20
	25	0	20	4	20	0	18	4	20
DTLZ4	5	20	20	20	19	20	20	20	20
	15	20	20	20	20	19	15	20	20
	25	20	20	20	20	19	19	20	20
DTLZ7	5	20	20	20	20	20	4	20	20
	15	19	17	20	20	19	17	20	20
	25	20	20	20	20	20	20	20	20

Effect of Threshold Parameter (θ)

- The results presented above for both the redundant and the non-redundant problems can easily be interpreted in the wake of the arguments presented in Sections IV.G and VII.A in the main paper. In that, a lower θ is likely to have little effect on the accuracy of the algorithms in the case of highly redundant problems, while their accuracy is likely to fall as the degree of redundancy in the problems reduces. The same hold true for the results presented above.

Effect of Neighborhood size ($q = \sqrt{M}$)

- The question of how to fix the parameter q which controls neighborhood size in MVU, is an open question. Based on empirical evidence $q = 4$ is most often used in literature. However, in this paper, one of the goals being control parameter reduction, $q = M - 1$ has been recommended and used. The rationale for this choice of q lies in giving preference to retention of local isometry. The results presented above and those in Tables XI and XII in the main paper show that with the use of $q = \sqrt{M}$, the performance in the case of redundant problems marginally improved, while it marginally deteriorated for the redundant problems. As the shift in performance is not drastic, the choice of $q = M - 1$ can be justified, as it offers reasonably high accuracy, along with the advantage of a reduction in one control parameter.