Increasing the Density of Multi-Objective Multi-Modal Solutions using Clustering and Pareto Estimation Techniques

R. Kudikala, I. Giagkiozis, and P.J. Fleming
Department of Automatic Control and Systems Engineering,
The University of Sheffield, Sheffield - S1 3JD, UK

Abstract— *For continuous multi-objective optimization problems there exists an infinite number of solutions on the Pareto-optimal front. A multi-objective evolutionary algorithm attempts to find a representative set of the Pareto-optimal solutions. In the case of multi-objective multi-modal problems, there exist multiple decision vectors which map to identical objective vectors on Pareto front. Many multi-objective evolutionary algorithms fail to find and preserve all of the multi-modal solutions in the non-dominated solutions set. Finding more of the available multi-modal solutions would give the decision maker a greater selection when choosing between solutions. In this paper, we present an extended version of the Pareto estimation method, to increase the density of the multi-objective multi-modal solutions. The method uses clustering analysis to identify and separate different clusters in the decision variables space which correspond to the multi-modal Pareto optimal solutions. Then Pareto estimation procedure is employed for these individual clusters, there by increasing the density of available multi-modal solutions. The proposed method has been tested on experimental test functions and is shown to be successful.*

Keywords: Multi-objective optimization, multi-objective multi-modal problems, cluster analysis, genetic algorithms.

1. Introduction

Many multi-objective evolutionary algorithms (MOEAs) fail to find and preserve all of the multi-modal solutions in the non-dominated solutions set [1]. Due to the incorporation of diversity operators in MOEA, they will assign low fitness values to solutions that are densely clustered in objective space, which will eventually lead to their elimination from the population. Hence they can identify only one set of decision vectors out of the multi-modal solutions and converge to any one of the global optima out of multiple global optima present in the multi-objective multi-modal problems. Finding the multi-modal solutions would allow the decision maker a greater choice when choosing between solutions. For example, in chemical process optimization the decision maker would want to know about different temperature settings for which the process can deliver the same results [2].

In this work, we present an extended version of the Pareto estimation method [3], which can be used to increase the number of multi-modal solutions. The method uses clustering analysis to identify and separate different clusters in the decision variables space which correspond to the multi-modal Pareto optimal solutions. Then Pareto estimation procedure is employed for these individual clusters, therefore increasing the density of available multi-modal solutions in multi-objective problems.

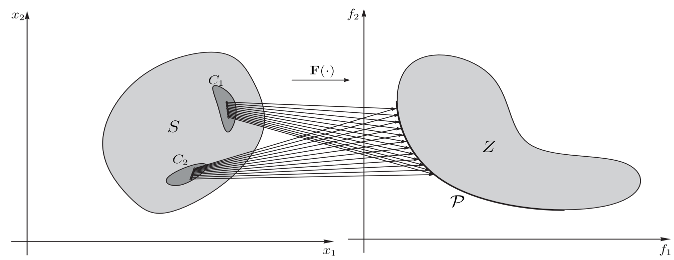

Fig. 1: A many to one objective function. The two sets C_1 and C_2 map to the same Pareto optimal solutions in the Pareto optimal set $\mathcal{P}$.

The remainder of this paper is organized as follows. In Section 2 a general definition of a multi-objective optimization problem and key concepts and definitions are provided. Section 3 presents Pareto estimation method and in Section 4 the extended Pareto estimation method with clustering is described for multi-objective multi-modal problems. In Section 5 the method is tested against a multi-objective multi-modal test problem with three cases and these tests are reported in Section 6. This paper is summarized and concluded in Section 7.

2. Problems with Multiple Global Optima

A multi-objective problem (MOP) is defined as:

$$\begin{aligned} &\min_{\mathbf{x}} \mathbf{F}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_k(\mathbf{x})), \\ &\text{subject to } \mathbf{x} \in S, \end{aligned} \tag{1}$$

where k describes the multiplicity of scalar objective functions $f(\cdot)$ and S is the *feasible region*. The vector of variables, $\mathbf{x}$, in this context is often referred to as decision vector while $\mathbf{z} = \mathbf{F}(\mathbf{x})$ is referred to as objective vector. An implicit assumption is that the individual scalar objective functions in (1) are mutually *competing*. The objective function described in (1) can in some cases be a many-to-one mapping. Namely, there exist $\mathbf{x}, \mathbf{y} \in S$ and $\mathbf{x} \neq \mathbf{y}$ that map to the same objective vector, $\mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{y})$. This can especially impact the optimization algorithm when the objective function is many-to-one in the domain of Pareto optimal solutions, see *Fig. 1*. In this case, the same Pareto optimal objective vector can be obtained for more than one decision vector.

If the above assumptions hold then only a partial ordering can be defined unambiguously. Namely, when comparing two decision vectors $\mathbf{x}, \tilde{\mathbf{x}} \in S$, it can so happen that their corresponding objective vectors are incomparable. In practice, this situation is resolved by a decision maker who will select one solution over all others, thus inducing a form of complete ordering. However this ordering is mostly subjective, even in the case that utility functions [4] are used to ease the work of the DM. In the absence of a DM a usual assumption is that the relative importance of the objectives, f_i, is unknown hence it is reasonable to obtain several non-comparable solutions. The problem of inducing partial ordering in Euclidean spaces was initially studied by Edgeworth [5], and later further expanded by Pareto [6]. The relations introduced by Pareto are defined as follows for a minimization problem:

Definition 1: A decision vector $\mathbf{x}^\star \in S$ is said to **weakly dominate** a decision vector $\mathbf{x}$ iff $f_i(\mathbf{x}^\star) \leq f_i(\mathbf{x}),\ \forall i \in \{1, 2, \ldots, k\}$ and $f_i(\mathbf{x}^\star) < f_i(\mathbf{x}),$ for at least one $i \in \{1, 2, \ldots, k\}$ then $\mathbf{x}^\star \preceq \mathbf{x}$.

Definition 2: A decision vector $\mathbf{x}^\star \in S$ is said to **dominate** a decision vector $\mathbf{x}$ iff $f_i(\mathbf{x}^\star) < f_i(\mathbf{x}),\ \forall i \in \{1, 2, \ldots, k\}$ then $\mathbf{x}^\star \prec \mathbf{x}$.

Definition 3: A decision vector $\mathbf{x}^\star \in S$ is said to be **Pareto optimal** if there is no other decision vector $\mathbf{x} \in S$ such that $f_i(\mathbf{x}) \leq f_i(\mathbf{x}^\star),\ \forall i \in \{1, 2, \ldots, k\}$ and $f_i(\mathbf{x}) < f_i(\mathbf{x}^\star),$ for at least one $i \in \{1, 2, \ldots, k\}$.

Definition 4: Let $\mathbf{F} : S \rightarrow Z$, with $S \in \mathbb{R}^n$ and $Z \in \mathbb{R}^k$. If S is the feasible region then the set Z is the feasible region in objective space. Given a set $\mathbf{A} \subset Z$, the **non-dominated set** is defined as $\mathcal{P} = \{\mathbf{z} : \nexists \tilde{\mathbf{z}} \preceq \mathbf{z}, \forall \tilde{\mathbf{z}} \in \mathbf{A}\}$. If $\mathbf{A}$ is the entire feasible region in the objective space, Z, then the set $\mathcal{P}$ is called the **Pareto optimal set** (PS) or **Pareto Front** (PF). Any element $\mathbf{z} \in Z$ is referred to as **objective vector**.

Definition 5: The **ideal objective vector**, $\mathbf{z}^\star$, is the vector with elements $(\inf(f_1), \ldots, \inf(f_k))$ [7, pp. 16].

Definition 6: The **nadir objective vector**, $\mathbf{z}^{nd}$, is the vector with elements $(\sup(f_1), \ldots, \sup(f_k))$, subject to f_i be elements of objective vectors in the Pareto optimal set [7, pp. 16].

Definition 7: The **convex hull** [8, pp. 24] of the set $C = \{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$, denoted as $\mathbf{conv}\, C$, where $\mathbf{e}_i$ is a $k \times 1$ vector of zeros with 1 on the i^{th} position, is referred to as $\mathbf{CH_I}$.

Definition 8: The **extended convex hull** ($\mathbf{EH_I}$) of the set C, is the union of CH_I and the points in the affine space of the set C produced by the projection of a Pareto optimal front, with ideal vector $\mathbf{0}$ and nadir vector $\mathbf{1}$, onto the hyper-surface of C.

Definition 9: Two decision vectors $\mathbf{x}, \mathbf{y} \in S$ are said to be **multi-modal solutions** if they satisfy $\mathbf{x} \neq \mathbf{y}$, and $\mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{y})$ for all $i = 1, \ldots, k$.

3. Pareto Estimation Method

3.1 Motivation

Consider the following problem. At the end of an optimization run on a multi-objective optimization problem we have a set of solutions that approximate the Pareto optimal front. Subsequently, these solutions are presented to a decision maker (DM) who can identify a few candidate solutions that are of interest, however, he would prefer a solution in the vicinity of the aforementioned solutions. In this case the analyst does not have many options and would either restart the optimization in hope that the preferred solution of the decision maker is obtained. An alternative is to use an interactive method such as, progressive preference articulation [9]. These alternatives present a number of difficulties of which the most obvious one is that the computational load is increasing disproportionately to the expected gain as there is absolutely no guarantee that the preferred solutions will be obtained. This consideration may lead the decision maker to abandon all the above scenarios and simply select one solution from the already existing Pareto set approximation.

The Pareto estimation method (PE) initially introduced in [3] resolves, to some extent, this issue by allowing the decision maker to explore more solutions in the vicinity of already obtained ones without resorting to further optimization. Specifically, Pareto estimation gives positive answer to the question: "Given a set of Pareto optimal solutions, obtained by any optimization algorithm, can specific solutions on the Pareto front be obtained that are not part of the initially obtained Pareto set?".

3.2 Overview

In [3] it was shown that using the Pareto estimation method the number of Pareto optimal solutions can be increased in specific regions of interest. Pareto estimation was applied to a 3-objective portfolio optimization problem successfully targeting two regions where the optimization algorithm used could not obtain solutions across 20 optimization runs. However, one of the assumptions in [3] was that the objective function is one-to-one, or at least that this condition obtains for the mapping between the Pareto set in decision and objective space. If this condition doesn't hold the artificial neural network used would face difficulties as for the same objective vector it would have to produce two or more output vectors simultaneously, see *Fig. 3*. In the rest of this section we briefly describe the Pareto estimation method and then explain the motivation for the extension introduced in this work. For a more complete description of the original version of the Pareto estimation method the reader is referred to [3].

A major motivation for the introduction of PE has been that Pareto optimal solutions can be obtained in specific regions of the Pareto front without the need to resort to additional optimization runs. Although there is no guarantee that such solutions will be produced the success rate of PE on a set of difficult test problems illustrated that the relative cost of

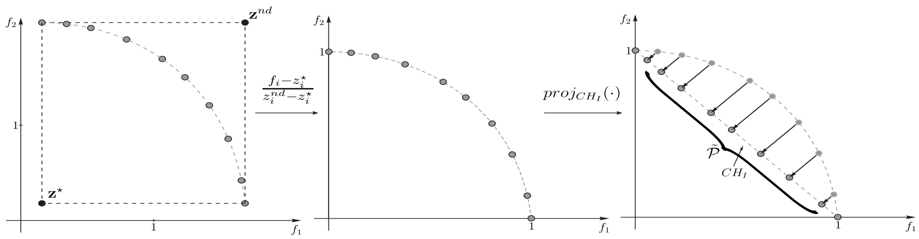

Fig. 2: Illustration of the Π^{-1} mapping for a hypothetical Pareto set $\mathcal{P}$.

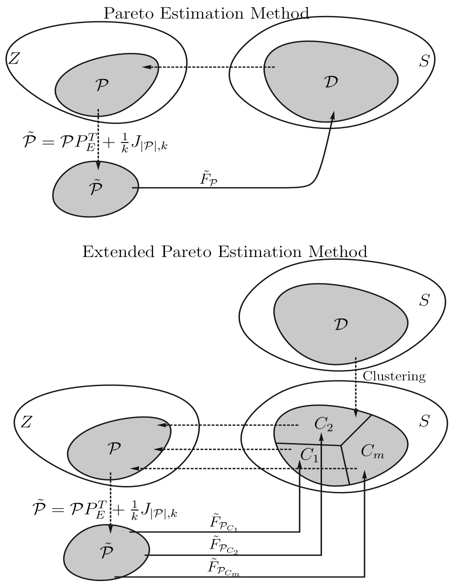

Fig. 3: The original version of the Pareto estimation method (top). The extended Pareto estimation method first clusters the Pareto optimal decision variable vectors and identifies a NN for every one (bottom).

applying PE before another method is justified [3]. PE depends on the ability to identify a relationship (mapping) from Pareto optimal solutions in objective space to decision space. This relationship can then be manipulated to produce solutions in specific parts of the Pareto front. We refer to this mapping as, $F_{\mathcal{P}}$, whose domain of definition is the set of Pareto optimal objective vectors, $\mathcal{P}$, and its range their corresponding decision variable vectors $\mathcal{D}$,

$$F_{\mathcal{P}} : \mathcal{P} \rightarrow \mathcal{D}. \tag{2}$$

We elected originally to use a radial basis neural network as it has been shown that it has competitive performance compared to the alternatives, see [10]. The theoretical argument that supports PE is was presented initially in [11], [12] and was later used by Zhang et al. [13] to create RM-MEDA, a regularity-based estimation of distribution algorithm. The argument is that for continuous multi-objective problems the Pareto optimal set is a piecewise continuous manifold in decision space. This effectively enables the identification of the mapping in (2).

3.3 Pareto Estimation - General Procedure

Pareto estimation is comprised of three main parts:

- A transformation of the Pareto optimal solutions in objective space, $\Pi^{-1} : \mathcal{P} \rightarrow \tilde{\mathcal{P}}$.
- The identification of the relationship, $\tilde{F}_{\tilde{\mathcal{P}}} : \tilde{\mathcal{P}} \rightarrow \mathcal{D}$, where $\mathcal{D}$ are decision vectors corresponding to Pareto optimal objective vectors, $\mathcal{P}$.
- The generation of a set, $\mathcal{E}$, and its use with the $F_{\tilde{\mathcal{P}}}$ mapping to generate a set of estimated decision vectors, $\mathcal{D}_{\mathcal{E}}$.

The first part is essentially a projection of the set $\mathcal{P}$ onto the CH_I. This is essential as it simplifies the task of generating the set $\mathcal{E}$ to a large degree, see *Fig. 2*. Prior to application of the projection the set $\mathcal{P}$ is normalized using,

$$\tilde{f}_i = \frac{f_i - \mathbf{z}_i^\star}{\mathbf{z}_i^{nd} - \mathbf{z}_i^\star}, \tag{3}$$

where $\mathbf{z}^\star$ and $\mathbf{z}^{nd}$ can be estimated from the Pareto optimal set $\mathcal{P}$. Using the normalization in (3) the objectives in $\mathcal{P}$ are restricted in the range $[0, 1]$ as shown in *Fig. 2*. Consequently the normalized objective vectors are projected onto CH_I as follows,

$$\tilde{\mathcal{P}} = \mathcal{P}P_E^T + \frac{1}{k}J_{|\mathcal{P}|,k}. \tag{4}$$

The matrix $J_{|\mathcal{P}|,k}$ is a $|\mathcal{P}| \times k$ unit matrix and P_E is a projection matrix obtained as:

$$\begin{aligned} P_E &= H(H^T H)^{-1} H^T, \\ H &= \left(\mathbf{e}_1 - \frac{1}{k}\mathbf{1} \cdots \mathbf{e}_{k-1} - \frac{1}{k}\mathbf{1}\right), \end{aligned} \tag{5}$$

where $\mathbf{e}_i$ is a vector of zeros with its i^{th} element set to 1. Next the artificial neural network (ANN) which is employed to identify the mapping $\tilde{F}_P$, is created (see [3]), using $\tilde{\mathcal{P}}$ and $\mathcal{D}$ as the training inputs and outputs respectively.

When the ANN has been trained, it then can be used for creating more Pareto optimal solutions in specific regions on the PF given a set, $\mathcal{E}$, is supplied as input. $\mathcal{E}$ can be generated in one of two ways:

- In a specific region, presumably that is of interest to the decision maker.
- On the entire CH_I, which if PE is successful will cover the entire Pareto front.

In this paper, we employ the second method as it illustrates the ability of PE and its extended version presented here, to successfully identify Pareto optimal solutions across the entire front. It should be noted however that we envisage that the usage of PE would be to target specific parts of the PF as seen in [3].

4. Clustering and Pareto Estimation for Multi-Objective Multi-Modal Solutions

In the case of multi-objective multi-modal problems, there exists multiple decision vectors which result in identical objective vectors on Pareto front as shown in *Fig. 1*. This corresponds to the many-to-one mapping of the multiple decision vectors in $\mathcal{D}$ to the objective vectors in $\mathcal{P}$. The decision vectors corresponding to each multi-modal optimal fronts (Pareto front) originate from different clusters $\mathcal{C}_m$ in decision variable space $\mathcal{D}$. The ANN relationship will fail to produce the one-to-many mapping of $\tilde{F}_{\tilde{\mathcal{P}}} : \tilde{\mathcal{P}} \to \mathcal{D}$. It will generate any one but not all of the multi-modal solutions. In order to overcome this problem, the different clusters $\mathcal{C}_m$ of multi-modal solutions present in the non-dominated set can be identified and separated using a clustering algorithm. The obtained clusters of decision vectors $\mathcal{C}_m$ and corresponding objective vectors in $\mathcal{P}$ will have one-to-one mapping between decision variable space and objective space for the Pareto front. Once the different clusters of decision vectors $\mathcal{C}_m$ are separated, the ANN can be trained for the individual cluster of solutions $\mathcal{C}_m$ and $\tilde{\mathcal{P}}$ to identify number of one-to-one mappings $\tilde{F}_{\tilde{\mathcal{P}}_{\mathcal{C}_m}} : \tilde{\mathcal{P}} \to \mathcal{C}_m$.

Most clustering algorithms need the number of output clusters to be pre-specified as an input to the algorithm. In general we do not know a priori the number of clusters available in the data set. Bezdek and Hathaway developed a visual assessment of cluster tendency (VAT) method [14], to identify potential clusters in a data set. Here the pair-wise dissimilarities between the n individuals of the data set are estimated and reordered,so that all the neighbouring individuals are consecutively ordered. The reordered $n \times n$ matrix of pair-wise dissimilarities is displayed as an intensity image with $n \times n$ pixels. Clusters are indicated by dark blocks of pixels along the diagonal of the image. However, the VAT method is too computationally costly for larger data sets. Wang et al., [15], proposed an improved VAT (iVAT) and an automated VAT (aVAT) methods to automatically determine the number of clusters and cluster separation based on the difference between diagonal blocks and off-diagonal blocks in the image of the reordered dissimilarity matrix. In this paper, the iVAT and aVAT [15] methods are used for identifying different clusters of decision vectors $\mathcal{C}_m$ in $\mathcal{D}$, which correspond to multi-modal solutions in the objective space $\mathcal{P}$. The steps involved in clustering and Pareto estimation of multi-objective multi-modal solutions are summarised as follows:

Step 1 Extract, $\mathcal{P}$, the non-dominated individuals obtained at the end run of an optimization algorithm, and, $\mathcal{D}$ the associated decision vectors.

Step 2 Perform clustering analysis on the obtained decision variable vectors $\mathcal{D}$ using a clustering algorithm.

Step 3 The obtained clusters of decision vectors $\mathcal{C}_m$ and corresponding objective vectors $\mathcal{P}_{\mathcal{C}_m}$ will have one-to-one mapping between decision variable space and objective space.

Step 4 For each individual cluster normalize $\mathcal{P}$ using (3).

Step 5 Project the normalized $\mathcal{P}$ onto the the $k-1$ hyperplane defined by the set of vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_{k-1}\}$ using (5) and (4), to produce $\tilde{\mathcal{P}}$.

Step 6 Identify the mapping $\tilde{F}_{\tilde{\mathcal{P}}_{\mathcal{C}_m}} : \tilde{\mathcal{P}} \to \mathcal{C}_m$ using $\tilde{\mathcal{P}}$ and $\mathcal{C}_m$ as inputs and outputs, respectively, and use these to train the ANN.

Step 7 Create the set $\mathcal{E}$. In this work this is a set of evenly spaced convex vectors.

Step 8 Use the set $\mathcal{E}$ as inputs to the ANN created in Step 5, to obtain estimates of decision vectors $\mathcal{C}_{\mathcal{E}}$.

Step 9 All the sets $\mathcal{C}_{\mathcal{E}}$ can be used with the objective function $\mathbf{F}(\cdot)$ to verify that the produced solutions are non-dominated and acceptable.

5. Experimental Setting

We employed the following multi-objective multi-modal test functions as seen in [1].

$$\begin{aligned} \mathbf{F}(\mathbf{x}) &= (f_1(\mathbf{x}), f_2(\mathbf{x})) \\ &= \left(\sum_{i=1}^{n} \sin(\pi x_i), \sum_{i=1}^{n} \cos(\pi x_i) \right) \\ & x_i \in [0, 6], i = 1, 2, ..., n. \end{aligned} \tag{6}$$

The above objective functions are chosen since both objectives are in conflict with each other and will have a trade-off in the objective space. For the minimization case, the above problem will have a known Pareto front which varies between $-\sum_{i=1}^{n} i$ to 0, where i is the number of decision variables chosen. The above problem is also a multi-objective multi-modal problem. The two objective functions are periodic functions with a period of 2. They will have efficient frontiers which correspond to the Pareto-optimal solutions for all the decision variable values varying in the ranges $x_i \in [2r+1, 2r+3/2]$, where r is an integer.

Deb and Tiwari [1] developed a generic evolutionary algorithm: Omni-optimizer, which incorporates restricted selection and crowding measure utilizing both objective and variable space information to find and preserve a well distributed multi-modal solutions. Here we use Omni-optimizer for solving the multi-objective multi-modal test problem in all three cases. Also we employ the ratio of the inverted generational distance, $D_R(\cdot,\cdot)$ and the ratio of the mean nearest neighbour distance $S_R(\cdot,\cdot)$ as well as the C-Metric. Due to space limitations we cannot include a description of these metrics, the reader is referred to [3].

6. Results and Discussion

In this paper, we are considering three test cases of the multi-objective multi-modal optimization problem (6) with different numbers of variables and population sizes in the optimization.

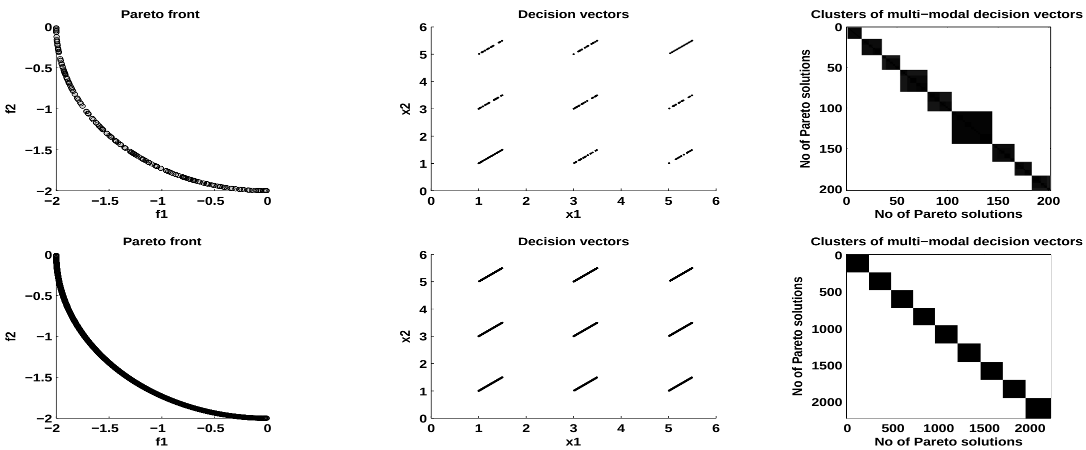

Fig. 4: Case I: Non-dominated solutions obtained with Omni-optimizer (top) and extended Pareto estimation methods (bottom) in objective space, decision variable space and image of clusters.

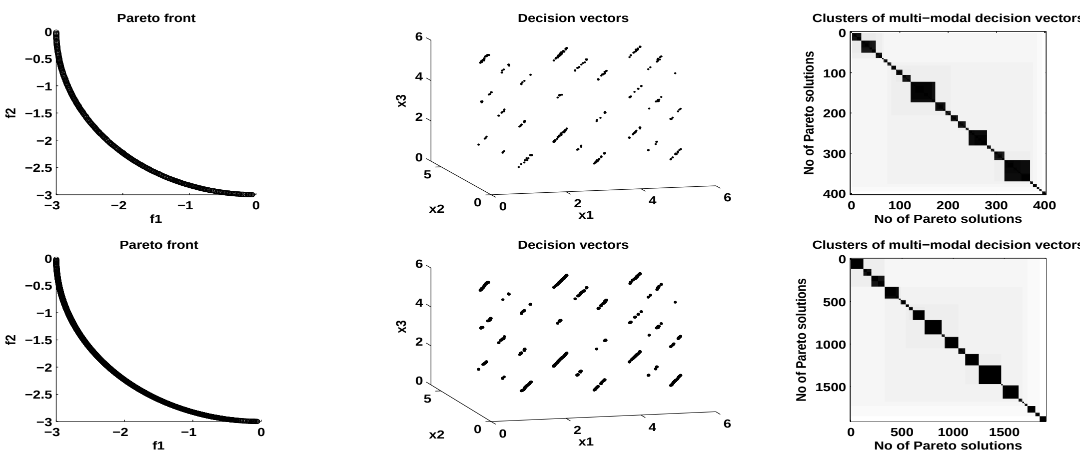

Fig. 5: Case II: Non-dominated solutions obtained with Omni-optimizer (top) and extended Pareto estimation methods (bottom) in objective space, decision variable space and image of clusters.

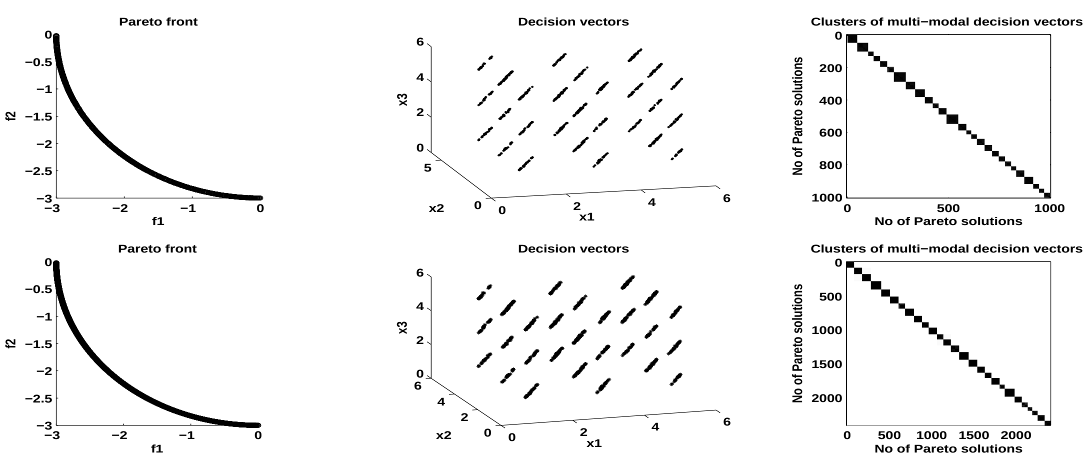

Fig. 6: Case III: Non-dominated solutions obtained with Omni-optimizer (top) and extended Pareto estimation methods (bottom) in objective space, decision variable space and image of clusters.

6.1 CASE I

In the Case I, we have chosen two decision variables $x_i \in [0, 6], i = 1, 2.$ for the optimization problem (6). Within the range of these two variables the problem will have nine multi-modal optimal fronts. We have set the population size to be 200 and run the optimizer for 500 generations. For the optimizer, finding the global Pareto front is not so difficult in this problem. However, finding all the multi-modal global Pareto fronts with good distribution of solutions in corresponding decision vector ranges is very difficult. In this particular instance, the Omni-optimizer [1] is able to find all nine multi-modal optimal fronts with 200 Pareto solutions with a good distribution of decision vectors in all ranges of $x_i \in [2r + 1, 2r + 3/2]$, where $i = 1, 2$ and $r = 0, 1, 2$. *Fig. 4* shows the obtained Pareto optimal solutions in top three sub-plots.

Cluster analysis using iVAT and aVAT [15] methods is performed for the obtained Pareto optimal decision vectors. The reordered dissimilarity matrix of 200 decision vectors is displayed as 200 x 200 image with gray scaling in right hand side sub-plot. The dark blocks appearing on the diagonal of the image represent individual clusters; the size of each dark block, represent number of individuals present in each cluster. It can be seen from this plot, that each cluster (dark block) has atleast 18 individual solutions. After separating these clusters of decision vectors, the procedure for Pareto estimation is executed. For each cluster, the ANN is trained to find the one-to-one mapping between objective space and decision vector space. Then this ANN is used to estimate 300 solutions in each cluster.

The quality of mapping estimated by ANN is highly dependent on the supplied training decision vectors. If the training data has a sufficient number of vectors, well distributed, then the ANN will estimate a better mapping, otherwise, the mapping estimated by ANN will be deceptive and may not generate good solutions in the Pareto estimation process. After combining all the solutions obtained from individual Pareto estimations, we perform non-dominated sorting to remove any dominated solutions from the set. At the end we have obtained around $2\,300$ non-dominated solutions out of $2\,700$ solutions estimated. In *Fig. 4* the bottom sub-plots show plots for objective vectors, decision vectors and gray scale image of the dissimilarity matrix of estimated solutions. It can be seen from these sub-plots, that Pareto estimation along with clustering is successfully able to find many solutions for the multi-objective multi-modal problem.

6.2 CASE II

In Case II, we have chosen the same objective functions with three decision variables $x_i \in [0, 6], i = 1, 2, 3.$. Within the range of these three variables the problem will have 27 multi-modal optimal fronts each one corresponding to $x_i \in [2r+1, 2r+3/2]$, where $i = 1, 2, 3$ and $r = 0, 1, 2$.

We have set 400 as the population size and run the optimizer for 500 generations. The obtained 400 non-dominated solutions are shown in the top three sub-plots *Fig. 5*. The optimizer has found all 27 multi-modal Pareto fronts, however, it is not able to obtain good distribution of solutions in all the corresponding decision vectors ranges. This can be easily observed from the image of clusters, in which some clusters have more than 70 solutions, where as some clusters got very less number of decision vectors around 2 or 3.

We applied Pareto estimation to each cluster of solutions, tried to estimate 200 solutions for each cluster. After combining the solutions estimated from all the clusters, non-dominated sorting is performed to get the non-dominated solutions. Around $1\,950$ solutions found to be non-dominated out of $5\,400$ estimated solutions. The method is able to estimate around 150 to 200 non-dominated solutions in some clusters, but failed to estimate more solutions in clusters where there are insufficient number of solution used for training the ANN. In *Fig. 5* the bottom three sub plots show the estimated non-dominated solutions, in objective space, decision variable space and image of clusters. (Note: Here the order of clusters in top and bottom images is not same.)

6.3 CASE III

In Case III, we have chosen the same objective functions with three decision variables $x_i \in [0, 6], i = 1, 2, 3.$, but now increase the population size to $1\,000$ in optimization and run the optimizer for 500 generations. The obtained $1\,000$ non-dominated solutions are shown in *Fig. 6* top three subplots. The optimizer has found all 27 multi-modal Pareto fronts and is now able to obtain a good distribution of solutions in all the corresponding decision vectors ranges. This can be easily observed from the image of clusters, in which clusters have solutions in the range of 20 to 60 solutions per cluster.

We applied Pareto estimation to each cluster of solutions and tried to estimate 150 solutions for each cluster. After combining the solutions estimated from all the clusters, non-dominated sorting is performed to get the non-dominated solutions. Around $3\,000$ solutions were found to be non-dominated out of $4\,050$ estimated solutions. The method is able to estimate around 100 to 150 non-dominated solutions in each cluster. In *Fig. 6* the lower three sub-plots show the estimated non-dominated solutions, in objective space, decision variable space and image of clusters. It can be seen that, a very good distribution of non-dominated solutions is obtained from the Pareto estimation. (Note: Here the order of clusters in top and bottom images is not same.)

Tables 1 and 2 summarize the various test metric computed for the non-dominated solutions before and after the Pareto estimation in all the three cases I, II, and III. These metrics indicated that the proposed method is able to estimate well distributed non-dominated solutions close to the true Pareto front, when compare to non-dominated solutions obtained from the Omni-optimizer.

Table 1: $D_R(\mathcal{P}, \mathcal{P}_\mathcal{E})$ and $S_R(\mathcal{P}, \mathcal{P}_\mathcal{E})$ values of the obtained solutions by OMNI-optimizer, $\mathcal{P}$, and the estimated set, $\mathcal{P}_\mathcal{E}$, by the extended Pareto estimation method.

	IGD Ratio			ESSm Ratio		
Problem	min	mean	std	min	mean	std
Case I	5.2689	6.3300	0.5565	9.1278	11.3150	1.3969
Case II	1.8721	2.3615	0.2722	5.0175	5.6117	0.3843
Case III	1.8316	2.2347	0.1524	2.4162	2.9601	0.2344

Table 2: C-Metric values of the solutions obtained by OMNI-optimizer, $\mathcal{P}$, and the estimated set, $\mathcal{P}_\mathcal{E}$, using the extended Pareto estimation method.

	$C(\mathcal{P}_\mathcal{E}, \mathcal{P})$			$C(\mathcal{P}, \mathcal{P}_\mathcal{E})$		
Problem	min	mean	std	min	mean	std
Case I	0.4900	0.5848	0.0433	0.0000	0.0100	0.0042
Case II	0.4525	0.5200	0.0273	0.0000	0.0000	0.0000
Case III	0.5900	0.6154	0.0216	0.0000	0.0000	0.0000

7. Conclusions

For continuous multi-objective optimization problems, there exist an infinite number of solutions on the Pareto-optimal front. A multi-objective evolutionary algorithm (MOEA) attempts to find a representative set of the Pareto-optimal solutions. If the decision maker is not satisfied with the representative set found by the MOEA, and wants to explore different solutions available on the Pareto front, the MOEA needs to be re-run, which will increase the number of function evaluations without providing any guarantees that a suitable solution will be identified. In this case, the Pareto estimation method can prove useful in order to increase the density of available non-dominated solutions in particular regions or the entire Pareto front. However, in the case of multi-objective multi-modal problems, the Pareto estimation method is not able to identify the one-to-many mapping of the objective vectors to decision variables vectors.

In this paper, we have introduced an extended version of the Pareto estimation method, to increase the density of multi-objective multi-modal solutions. The method uses clustering analysis to identify and separate different clusters in the decision variable space which correspond to the multi-modal Pareto optimal solutions. These individual clusters are then used to estimate the relation between objective space to decision space using an ANN. Instead of a single network, as is the case in the Pareto estimation method, we employ a network for each individual cluster. These are then employed to estimate more solutions for the selected cluster, presumably by the decision maker, or all cluster. For testing purposes we have employed the latter method in this work.

The proposed method has been tested on experimental test functions, with three different case studies. We have used Omni-optimizer [1] to solve the test problem in three cases, and iVAT and aVAT [15] methods to identify different clusters of solutions in the cluster analysis. In all cases, the extended Pareto estimation method has successfully found many non-dominated solutions corresponding to different multi-modal solutions. The success of the proposed method highly depends on the number of solutions available in an individual cluster for training and estimating the one-to-one mapping between objective space and decision vector space. In case II, the method failed to improve density of the solutions in the clusters with a small number of individuals. However, it has improved the density of solutions in the remaining clusters representing multi-modal solutions to the test problem. We leave for future work the evaluation of the proposed method on a real-world system architecture design problems, which have a tendency to have multi-model solutions.

8. Acknowledgements

The first author wishes to acknowledge the financial support of Rolls-Royce plc and EPSRC through a Dorothy Hodgkin Postgraduate Award (DHPA).

References

[1] K. Deb and S. Tiwari, "Omni-optimizer: A generic evolutionary algorithm for single and multi-objective optimization," *European Journal of Operational Research*, vol. 185, no. 3, pp. 1062–1087, 2008.

[2] A. Tarafder, G. Rangaiah, and A. K. Ray, "A study of finding many desirable solutions in multiobjective optimization of chemical processes," *Computers & chemical engineering*, vol. 31, no. 10, pp. 1257–1271, 2007.

[3] I. Giagkiozis and P. Fleming, "Increasing the Density of Available Pareto Optimal Solutions," Department of Automatic Control and Systems Engineering, The University of Sheffield, Research Report No. 1028, November 2012.

[4] C. Vira and Y. Haimes, *Multiobjective Decision Making: Theory and Methodology*. North-Holland, 1983, no. 8.

[5] F. Edgeworth, *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. CK Paul, 1881, no. 10.

[6] V. Pareto, "Cours D'Économie Politique," 1896.

[7] K. Miettinen, *Nonlinear Multiobjective Optimization*. Springer, 1999, vol. 12.

[8] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[9] C. Fonseca and P. Fleming, "Multiobjective Optimization and Multiple Constraint Handling with Evolutionary Algorithms. I. A Unified Formulation," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 28, no. 1, pp. 26–37, 1998.

[10] Y. Jin, "A Comprehensive Survey of Fitness Approximation in Evolutionary Computation," *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, vol. 9, pp. 3–12, 2005.

[11] Y. Jin and B. Sendhoff, "Connectedness, Regularity and the Success of Local Search in Evolutionary Multi-Objective Optimization," in *Congress on Evolutionary Computation*, vol. 3, dec. 2003, pp. 910 – 1917.

[12] O. Schütze, S. Mostaghim, M. Dellnitz, and J. Teich, "Covering Pareto Sets by Multilevel Evolutionary Subdivision Techniques," in *Evolutionary Multi-Criterion Optimization*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2003, vol. 2632, pp. 10–10.

[13] Q. Zhang, A. Zhou, and Y. Jin, "RM-MEDA: A Regularity Model-Based Multiobjective Estimation of Distribution Algorithm," *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 1, pp. 41–63, 2008.

[14] J. C. Bezdek and R. J. Hathaway, "Vat: A tool for visual assessment of (cluster) tendency," in *Proceedings of the 2002 International Joint Conference on Neural Networks, 2002. IJCNN'02.*, vol. 3. IEEE, 2002, pp. 2225–2230.

[15] L. Wang, U. Nguyen, J. Bezdek, C. Leckie, and K. Ramamohanarao, "iVat and aVat: Enhanced visual analysis for cluster tendency assessment," *Advances in Knowledge Discovery and Data Mining*, pp. 16–27, 2010.