Visualising Mutually Non-dominating Solution Sets in Many-objective Optimisation

David J. Walker, *Student Member, IEEE*, Richard M. Everson, *Member, IEEE*
and Jonathan E. Fieldsend, *Member, IEEE*

***Abstract*—As many-objective optimisation algorithms mature the problem owner is faced with visualising and understanding a set of mutually non-dominating solutions in a high dimensional space. We review existing methods and present new techniques to address this problem.**

We address a common problem with the well known heatmap visualisation, that the often arbitrary ordering of rows and columns renders the heatmap unclear, by using spectral *seriation* to rearrange the solutions and objectives and thus enhance the clarity of the heatmap. A multi-objective evolutionary optimiser is used to further enhance the simultaneous visualisation of solutions in objective and parameter space.

Two methods for visualising multi-objective solution objectives in the plane are introduced. First, we use RadViz and exploit interpretations of barycentric coordinates for convex polygons and simplices to map a mutually non-dominating set to the interior of a regular convex polygon in the plane, providing an intuitive representation of the solutions and objectives.

Second, we introduce a new measure of the similarity of solutions—the dominance distance—which captures the order relations between solutions. This metric provides an embedding in Euclidean space, which is shown to yield coherent visualisations in two dimensions.

The methods are illustrated on standard test problems and data from a benchmark many-objective problem.

I. Introduction

The recent trend towards investigating *many-objective* problems, problems with four or more conflicting objectives, has brought with it several difficulties [1]. One impediment to understanding the results of a many-objective optimisation is visualising the set of solutions produced. In a *multi-objective* context, problems consisting of two or three objectives, an intuitive visualisation is obtained from scatter plots of the solutions in objective space, allowing the decision maker to identify the trade-off between objectives. However, the relationships between even three variables can be difficult to comprehend when plotted in two dimensions, and it is not usually possible for the decision maker to comprehend four or more spatial dimensions visually. The goal of this paper is to introduce methods that permit a decision maker to explore the results of an evolutionary algorithm applied to solving a many-objective optimisation problem to aid the selection of a solution and illustrate relationships between solutions, as well as between objectives.

The authors are with the Department of Computer Science, College of Engineering, Mathematics and Physical Sciences, University of Exeter, EX4 4QF, UK. (email: {D.J.Walker, R.M.Everson, J.E.Fieldsend}@exeter.ac.uk).
David Walker was supported by an EPSRC CASE Studentship with Motorola during the preparation of this paper.

There are many multivariate data visualisation methods, some of which have been applied to the visualisation of many-objective solution sets. Some of these, for example parallel coordinate plots [2] and heatmaps [3], result in a visualisation from which all of the original data can be recovered, however often they can be difficult to interpret because solutions are overlayed or arbitrarily ordered. Other methods, such as principal component analysis [4] and Neuroscale [5] compress the dimensionality of the population into a 2-dimensional space so that it can be visualised with standard techniques, such as scatter plots [6]. With these techniques it is not possible, using the visualisation alone, to recover the original objective values of the solutions it illustrates, and so potentially useful information has been lost. Section II reviews the application of these methods, and others, to visualising many-objective populations. In [7] we suggested novel methods for ordering a many-objective population in order to facilitate a more comprehensible visualisation without the need to remove objectives. Here we introduce new visualisation techniques for displaying many-objective mutually non-dominating sets to facilitate understanding about the relationships between the constituent solutions. We begin by presenting techniques to enhance the clarity of the popular *heatmap* method and show how it can be used to visualise both objective and parameter space components of solutions without discarding any objectives.

Despite the loss of potentially important information, dimension reduction methods often produce a useful visualisation and it is especially appealing to visualise solutions in the plane as paper and computer screens are two-dimensional and humans are adept at interpreting planar diagrams. General dimension reduction methods, such as principal component analysis, Self Organising Maps [8] and Neuroscale, are ignorant of the mutually non-dominating nature of solutions lying on an estimate of the true Pareto front of a many-objective problem and in general do not preserve the dominance relations between individual solutions. We therefore examine two visualisation methods suited to mutually non-dominating data. In the first we use RadViz [9] to map the objective axes to the vertices of a planar polygon. We provide a new derivation of the RadViz projection, exploiting interpretations of barycentric coordinates for convex polygons and for simplices in many dimensions, which indicates how it is useful for mutually non-dominating sets. In the second planar visualisation method we introduce a new similarity measure—dominance distance—between solutions, based on their relative dominance. Using metric multi-dimensional scaling, dominance distance yields

an embedding of the solution set in Euclidean space which is linearly projected to a two- or three-dimensional visualisation space. The RadViz method preserves the identity of the objectives in the visualisation, while the second may be viewed as a method that clusters together solutions that have a similar relation to other solutions in the set.

Throughout this paper we are concerned with visualising the solutions to the M-objective minimisation problem which is succinctly stated as

$$\text{minimise} \quad \mathbf{y} = (f_1(\mathbf{x}), \ldots, f_M(\mathbf{x})), \tag{1}$$

where the functions $f_m(\mathbf{x})$ map the P-dimensional parameter vectors $\mathbf{x}$ to objectives (there may also be constraints). We assume that the parameters and objectives are real valued so that $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^P$ and $\mathbf{y} \in \mathbb{R}^M$. The result of a minimisation is generally a set of solutions $\mathcal{P} = \{\mathbf{x}_k\}_{k=1}^K$ and the corresponding objectives $\mathcal{F} = \{\mathbf{y}_k\}_{k=1}^K$ which approximate the Pareto set and Pareto front respectively. Regarded as a set of M-dimensional coordinate points, the elements of $\mathcal{F}$ are not dominated and thus *mutually non-dominating* and by extension the elements of $\mathcal{P}$ are also regarded as non-dominated and mutually non-dominating. We denote by y_{km} the value of the k-th solution on objective m.

The techniques we present in this paper are illustrated on sets of mutually non-dominating solutions for several many-objective problems. The first is a set of solutions to a 9-objective problem, optimising the design of a waveform for a Pulsed Doppler Radar, proposed as a prototypical many-objective problem [10]. We also use populations of solutions to test problems, visualising problem instances of DTLZ2 and DTLZ6 [11], as well as WFG3 and WFG8 from the Walking Fish Group suite of test problems [12]. These examples are introduced in detail later in the paper.

The remainder of the paper is structured as follows. Section II reviews a selection of visualisation methods, although we do not aim to provide an exhaustive review, concentrating instead on methods that have been used for the visualisation of estimated Pareto fronts. Section III introduces seriation for enhancing heatmaps, applying it to the reordering of solutions in Section III-A and for reordering objectives in Section III-C; Section IV demonstrates the joint seriation of two spaces so that heatmaps of both parameter and objective spaces can be visualised together. We then present two methods for reducing the dimensionality of a solution set. In Section V we use RadViz to map solutions to the interior of a polygon in the plane and in Section VI we use multi-dimensional scaling in conjunction with a new dominance-based metric. Conclusions are drawn in Section VII and future work is discussed.

II. Many-objective Set Visualisation

A variety of methods have been employed to cope with visualising the increasing number of objectives found in industrial and scientific optimisation problems. In this section, we review the principal methods.

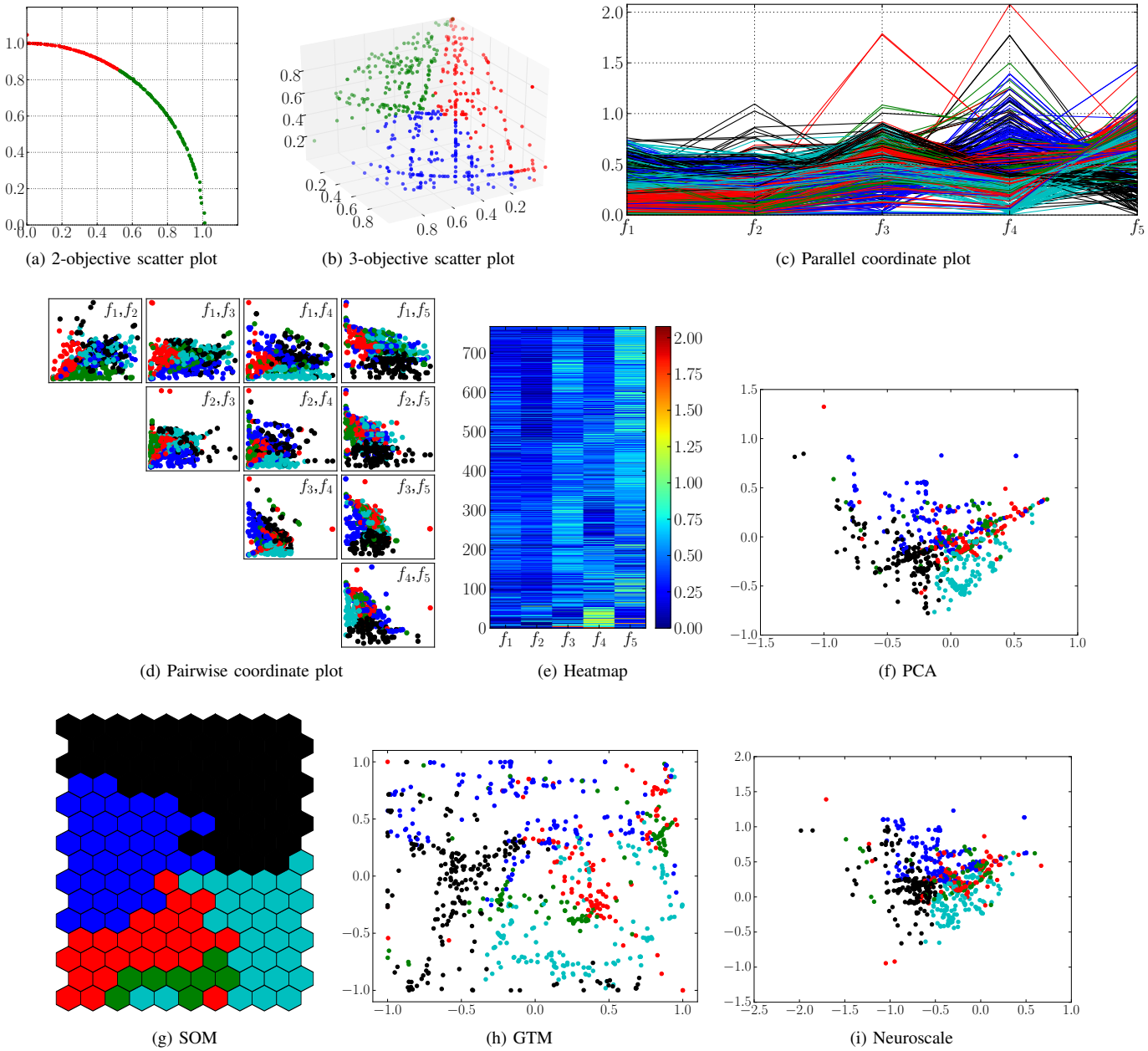
For illustration, Fig. 1 presents examples of some of the existing many-objective visualisation methods. The examples all show the objective-space mapping of solutions $\mathcal{F} = \{\mathbf{y}_k\}$ generated by running a basic population-based multi-objective evolutionary algorithm (MOEA) for 5,000 generations, maintaining an elite archive of non-dominated solutions throughout the process. Specifically, the algorithm was a $(\mu+\lambda)$–evolution strategy (ES) in which each of the μ parent solutions produced a single child solution at each generation; $\mu = 100$ and $\lambda = 100$. The child solutions were mutated with an additive Gaussian mutation ($\sigma = 0.1$, mutation probability $1/P$) and the archive was used to maintain the current approximation of the true Pareto front. We used the well-known test problem DTLZ2 [11], and generated results for $M = 2, 3, 5$ objective instances of the problem; the number of parameters was $P = 10+(M-1)$ as recommended by [11]. With the exception of Figs. 1(a) and 1(b), all of the examples use 768 solutions from 5-objective archive. By construction, the Pareto front is known to be the portion of a spherical shell of radius 1 lying in the positive orthant. The solutions are quite well converged; for the 2-objective instance the median distance from the true Pareto front is 5.48×10^{-4}. For the 3-objective archive the median distance is 3.15×10^{-3}, and for the 5-objective archive 5.38×10^{-2}.

Probably the most common method for visualising solutions in a *multi*-objective context is to produce a scatter plot on 2- or 3-dimensional axes, where each axis represents an objective. Examples are shown in Fig. 1(a) (two objectives) and 1(b) (three objectives)[1] and clearly show the spherical nature of the estimated front. Solutions have been coloured according to the objective m on which the solution is performing "best" as follows. In order to avoid biases due to the differing scales on which the objectives are measured, the solutions were ranked on each objective separately; we denote the rank of solution $\mathbf{y}_k$ on the m-th objective by r_{km} with 1 being the best rank and K the worst. Then solution $\mathbf{y}_k$ is coloured according to the objective for which r_{km} is minimum. As Fig. 1(b) shows, this colouring tends to colour neighbouring solutions in the same colour and gives some indication of the nature of a region of objective space. Although this provides relatively little additional information for 2 and 3 objectives, the same device considerably enhances the interpretability of many objective visualisations [6].

A. Many-objective Methods

Two of the earliest methods identified for use in many-objective optimisation were *parallel coordinate plots* [13], [14], [2] and *pairwise coordinate plots* [15]. A parallel coordinate plot presents each solution $\mathbf{y}_k$ as graph of y_{km} versus objective m with the values connected by lines. Whilst this is easily extended to any number of objectives, as Fig. 1(c) shows for the 5-objective DTLZ2 solutions, the plots are often too cluttered to be of use. A pairwise coordinate plot compares each pair of objectives as a 2-dimensional scatter plot, as shown in Fig. 1(d). This is useful for revealing correlated and anti-correlated pairs of objectives and provides information on the pairwise interactions between objectives. However, the

[1] In order to enhance the clarity of the visualisation, one solution has been omitted from Fig. 1(b). This solution was an outlier and scaled the axes so that the remainder of the solutions were difficult to see.



(a) 2-objective scatter plot (b) 3-objective scatter plot (c) Parallel coordinate plot

(d) Pairwise coordinate plot (e) Heatmap (f) PCA

(g) SOM (h) GTM (i) Neuroscale

Fig. 1: Examples of some of the visualisation methods reviewed in Section II. The solutions shown are for a 5-objective DTLZ2 archive (with the exception of 1(a) and 1(b), which are 2- and 3-objective problem instances, respectively) generated using a basic MOEA. The solutions represent the non-dominated archive after 5000 function evaluations. Solutions are coloured according to which objective has the highest rank: 1 = red, 2 = blue, 3 = green, 4 = cyan, 5 = black.

fact that the points representing a particular solution in each plot are not visually linked together means that it is generally difficult to perceive relations between solutions. While it is mechanically easy to extend to any number of objectives, the number of plots $M(M-1)/2$ rapidly becomes overwhelming.

Heatmaps are frequently used to visualise large multivariate datasets (see for example [16]) and have recently been used for multi-objective populations [3], [17], [18], [19], [20]. In a heatmap objectives are represented as columns, solutions by rows, and relative values as 'heat' represented by colour. Fig. 1(e) presents a heatmap of the 5-objective DTLZ2 archive. The arbitrary ordering of solutions means that it can be difficult to observe relationships between the various solutions and objectives. Schemes for reordering the rows and columns of a heatmap to present a clearer view of a multi-objective population, which we discuss later, have been proposed [3], [17]. In addition, in order to be of use, the objectives must be on the same scale. One way in which this is done is by normalising values to similar ranges; for example, in [17] linear scaling of the solutions for each objective to $[0, 1]$ before assigning colours is recommended. In Fig. 1(e) the objectives are all on roughly the same scale, so no scaling

was done before assigning colours; nonetheless, the heatmap is dominated by the cooler colours. We present methods which use the full range of colours and place similar solutions together to enhance a visualisation of both objective space and parameter space in Sections III–IV.

B. Dimension Reduction for Visualisation

Since scatter plots provide such an intuitive visualisation of multi-objective solutions, an obvious course of action is to project the solution archive into 2 or 3 dimensions and draw a scatter plot. We briefly review dimension reduction methods which have been used for visualisation of many-objective solutions in this manner.

Probably the most common linear dimension reduction technique is principal component analysis (PCA, [4]), which identifies the directions of objective space that capture the maximum amount of variance in the solutions. Fig. 1(f) shows the PCA projection of the 5-objective DTLZ2 solutions into the two-dimensional space spanned by the first two principal components. The projection has identified the two directions in objective space which retain the most variance and are therefore the best *linear* approximation to the original archive in a mean squared sense. However, as the colouring shows, the solutions are overlayed in such a way that it is difficult to distinguish among them and the coherence of neighbouring solutions evident in the multi-objective examples (Figs. 1(a) and 1(b)) is absent here. Of course, some information must be lost in projecting into a lower dimensional space, but we note that PCA is oblivious to the mutually non-dominating nature of these solutions.

Three nonlinear methods that have been used for visualising many-objective solutions are Self Organising Maps (SOM) [8], Generative Topographical Mapping (GTM) [21] and Neuroscale [5]. All three of these methods aim to preserve local structure between objective-space solutions.

The SOM [8] is a topographically-arranged network of interacting transformation functions, whose response (displayed as a degree of *excitation* of all the component nodes) varies depending upon the network input. In the basic formulation (as used in e.g. [22], [6]), the SOM defines a mapping from the input space onto a two-dimensional array of nodes. Each node in this array has an associated M-dimensional *reference vector* $\mathbf{w}$, and these nodes are compared to any input, $\mathbf{y}$, to the network, in a parallel fashion. Abstractly, the SOM seeks to find some best matching node to $\mathbf{y}$, denoted $\mathbf{w}_c$, whose response should be maximised given the input. Additionally, the learning algorithms incorporated in SOMs seek to instill a local relationship between neighbouring nodes, such that nodes that are spatially close to one another in the network topology, should also be concerned with adjacent regions of input space. One interpretation of this approach presented in [8] is to view the trained SOM as a nonlinear projection of the probability density function of the M-dimensional input into the two-dimensional display provided by the network.

Fig. 1(g) illustrates the reference vectors associated with a SOM of the 5-objective DTLZ2 archive. Each hexagon represents one of the reference vectors in the trained mapping, and each vector has been coloured according to the objective of the reference vector which has the best value (determined by comparing the reference vector to the training data, and seeing which of its objective values would have the best rank in this data). From the distribution of colours in the visualisation it is clear that the reference vectors have been distributed across the Pareto front and provides a coarse-grained spatially coherent representation.

The Generative Topographic Mapping (GTM) [21] is an alternative to the SOM which provides a generative probabilistic model allowing the likelihood of new data to be assessed and incorporated. It represents the data as a nonlinear mapping to the high-dimensional data space of a topographically ordered low-dimensional latent space. The data is then visualised as its projection into the latent space. The nonlinear mapping is achieved by a constrained mixture of radial basis functions and a Gaussian noise model accounts for discrepancies between the noise-free mapping from latent space to data space and the observed data. The likelihood corresponding to this generative model is then maximised using the expectation-maximisation algorithm in order to learn the model parameters and the latent visualisation. Fig. 1(h) shows the visualisation of the 5-objective DTLZ2 front by GTM; individuals have been broadly clustered into similar groups, as indicated by the colouring, but there is an imperfect separation into distinct, topographically coherent regions.

Neuroscale [5], [23] has also been used for many-objective visualisation [6], [24]. It also uses radial basis functions to form a nonlinear mapping projecting an M-dimensional individual $\mathbf{y}$ into a Q-dimensional individual $\hat{\mathbf{y}}$ using topographical information derived from considering the distances between solutions. The radial basis functions are arranged as a neural network whose inputs are the high-dimensional solutions and the outputs are the corresponding low-dimensional solutions. The network weights are adjusted in order to minimise the Sammon stress [25]:

$$\sum_{k}^{K}\sum_{j>k}^{K}\left(d_{kj}-\hat{d}_{kj}\right)^2, \tag{2}$$

in which d_{kj} is a distance between the individuals k and j in M-dimensional objective space and $\hat{d}_{kj}$ is a distance between the corresponding individuals in the Q-dimensional space. This metric is minimised when the distances between pairwise individuals in the original objective space and the embedded space are the same. Fig. 1(i) demonstrates the application of Neuroscale to the visualisation of the 5-objective DTLZ2 front. Solutions have been coloured by the objective on which they best perform, but like PCA, segregation into distinct regions is unclear. Closely related to Neuroscale is a method in which solutions are embedded in a 3-objective space by minimising the Sammon error [26]. This embedding is then presented to a decision maker as an interactive virtual environment that can be explored for knowledge discovery.

Clustering approaches have also been used, for example, to visualise the results of multi-objective nurse scheduling, Fuzzy C-Means Clustering was used to cluster solutions [27]. The axes onto which the data were then projected were identified

using Fuzzy Multiple Discriminant Analysis by finding the projection that maximises the ratio of within-class scatter and between-class scatter.

The *prosection* method [28] visualises a population by compressing the objectives using prosections. A prosection is the projection of individuals within a section of the objective space into a low-dimensional space; in two dimensions, solutions are projected onto a line running through the section and intersecting the origin. The projections of solutions onto the line defining the section are then rotated through the angle between the projection line and one of the axes so that a dimension is removed. Reducing the dimensionality in this way has the advantage that if one solution dominates another then its prosection projection dominates the others projection. However, two mutually non-dominating solutions may be projected so that one dominates the other. Additionally, it is currently only possible to visualise populations of four objectives or fewer. Another method [29] which seeks to preserve dominance relationships first projects the non-dominated solutions onto the positive quarter of a circle centred on the origin. Then a greedy procedure is used to find the position of each dominated solution that best preserves the dominance relationships of the original population.

Most of the visualisation methods reviewed in this section tend to suffer from one of two problems. They are either lossless and present the entire set of objectives, which often results in a lack of clarity, or they make a dimension reduction which loses information about the dominance relations between solutions. While the prosection method does not always suffer from these problems, it is currently only applicable to problems comprising four objectives or fewer. Methods for enhancing the clarity of (lossless) heatmaps by reordering the solutions have been investigated by [3] and [17], and in the following section we investigate spectral methods for reordering the solutions, objectives and parameters in a combined heatmap visualisation of both parameter and objective space. In later sections we investigate dimension reduction methods that attempt to minimise the loss of dominance information.

III. Seriation of Heatmaps

As illustrated in Fig. 1(e), a heatmap represents the data as a grid of pixels whose colours indicate values on a scale from maximal (hot) to minimal (cold). In [3] the use of a heatmap for presenting both the objective and parameter space views of solution sets in multi-objective optimisation is presented. More recently, a heatmap was incorporated into an interactive multi-objective particle swarm optimisation algorithm [18]. Heatmaps are a particularly useful method for visualising solution sets to many-objective problems because they allow the trade-off between objectives to be observed, providing important information to a decision maker. Their scalability, both in terms of solutions and objectives, means that they can visualise large populations of solutions to problems defined in terms of large numbers of objectives. In addition, this information is available without having to compress or discard objectives, meaning that no information is lost in the visualisation process and the original data is recoverable, unlike techniques such as PCA in which dimensionality compression discards potentially useful information.

Whilst heatmaps can convey useful information, we identify two problems: the presence of one or two larger objectives values in the data means that the full range of colours may not be used; and, more seriously, arbitrary ordering of the solutions and objectives in the heatmap hampers its interpretability. Hierarchical clustering has been proposed to ameliorate this [3]; in this paper, we apply a spectral method to seriate solutions and objectives, placing similar solutions and objectives close to each other.

The goal of *seriation* is to construct a permutation over individuals such that similar individuals are placed close together, and dissimilar individuals far apart. Seriation has a long history with early uses in archeology to establish a chronological ordering of artifacts based on the similarity of their features [30], and in sociology for grouping similar people together [31]. An important advance was made in [32], which introduced a spectral method for finding an approximate solution to the seriation problem. For an extensive historical review of seriation see e.g. [33]. Here we use seriation to reorder the set of non-dominated solutions to a many-objective problem in both objective space and parameter space, with the aim of placing similar solutions and objectives close to each other in the heatmap thus visualising the trends and exceptions from the trends present in the solutions.

Rather than visualise the objective values y_{km} themselves, we rank the solutions on each objective to obtain r_{km}, the rank of the k-th solution on the m-th objective. We denote the vector of M ranks for an individual solution by $\mathbf{r}_k = (r_{k1}, r_{k2}, \ldots, r_{kM})$. This ranking has two principal benefits. First, the values to be displayed by assigning a colour should be on a common scale [3]. Ranking the solutions brings them onto the same range, $1 \leq r_{km} \leq K$, without damaging the dominance relations between solutions, so that $\mathbf{r}_k \prec \mathbf{r}_j$ iff $\mathbf{y}_k \prec \mathbf{y}_j$. Second, provided there are no tied ranks, each ranked value occurs exactly M times, so that each colour in the heatmap occurs M times, thus using the full range of colours equally. This is equivalent to histogram equalisation (e.g., [34]) for each objective and avoids the problem with linear scaling of objectives, apparent in Fig. 1(e), that detail is lost because a few large values force the majority of the heatmap to be displayed in cool colours. Using ranks has the disadvantage that the ranks must be recomputed if the solution set changes. However, the cost of recomputation is small and in this paper we confine our deliberations to static solution sets with K solutions.

A. Seriation of Solutions

The similarity between solutions can be measured in a variety of ways, and we examine rank-based similarity measures below. For a set of K solutions we begin by constructing a $K \times K$ *similarity matrix* $\mathbf{A}$ describing the similarity between solutions $\mathbf{r}_k$ and $\mathbf{r}_j$ based on the squared difference between the corresponding ranks:

$$A_{kj} = 1 - \frac{1}{M(K-1)^2} \sum_{m=1}^{M} (r_{km} - r_{jm})^2 . \qquad (3)$$

Since the greatest difference in ranks, which occurs when two objectives are anti-correlated, is $K-1$, it is guaranteed that $\sum_{m=1}^{K}(r_{km}-r_{jm})^2 \leq M(K-1)^2$, so that $0 \leq A_{kj} \leq 1$.

In order to place similar solutions together, we seek a permutation over the solutions $\boldsymbol{\pi}$ that minimises:

$$g_{\mathbf{y}}(\boldsymbol{\pi}) = \sum_{j=1}^{K}\sum_{k=1}^{K} A_{ij}(\pi_k - \pi_j)^2. \tag{4}$$

The objective function $g_{\mathbf{y}}(\boldsymbol{\pi})$ is minimised when similar solutions are placed close together, and dissimilar solutions far apart. In general, this is NP-hard because the permutation is discrete [32]. Instead, [32] suggests an approximation obtained by relaxing the permutation $\boldsymbol{\pi}$ to a continuous variable $\mathbf{v}$ and minimising:

$$h(\mathbf{v}) = \sum_{j=1}^{K}\sum_{k=1}^{K} A_{kj}(v_k - v_j)^2 \tag{5}$$

with respect to $\mathbf{v}$. This relaxed objective is subject to two constraints. Firstly, to ensure that adding a constant to all v_k does not change the order of the individuals the constraint $\sum_k v_k = 0$ is imposed. Also, in order to avoid the trivial solution in which all $v_k = 0$, we demand that $\sum_k v_k^2 = 1$. The solution to the constrained problem can be found with linear algebra. Briefly, the problem is rewritten as $h(\mathbf{v}) = \mathbf{v}^T\mathbf{L}\mathbf{v}$, where $\mathbf{L}$ is the *graph Laplacian* [35], [36] of $\mathbf{A}$ defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$ with $\mathbf{D}$ the diagonal matrix whose elements are $D_{kk} = \sum_j A_{kj}$. Then eigenvectors of $\mathbf{L}$ correspond to stationary points of $h(\mathbf{v})$ [35]. The smallest eigenvalue of $\mathbf{L}$ is zero, with the corresponding eigenvector having equal elements. A discrete permutation is recovered from the *Fiedler vector* [35], the eigenvector corresponding to the smallest non-zero eigenvalue of $\mathbf{L}$, by ordering the individuals such that the individual with the k-th smallest value in the Fiedler vector occupies the k-th position in the permutation. At first sight locating the Fiedler vector requires a full eigendecomposition of $\mathbf{L}$, however, the Fielder vector may be identified as the eigenvector corresponding to the largest eigenvalue of the complementary graph Laplacian matrix, which may be efficiently found by the power method [37]. Although this method might be necessary for large populations, for the applications addressed here the matrix decomposition is computationally inexpensive (although $O(K^3)$) and very much faster than exhaustive search which requires $O(K!)$ operations, rendering it infeasible for $K \gtrapprox 10$.

B. Illustration

Throughout this paper, we use a running example dataset drawn from the domain of many-objective optimisation. Hughes applied the MSOPS algorithm [38] to the problem of designing an appropriate set of waveforms that can be transmitted by a Pulsed Doppler Radar to simultaneously measure the velocity and distance of a target [10]. To do this, Hughes optimised a schedule of Pulse Repetition Intervals (PRIs), which are the times between transmission of radar pulses. The P parameters $\mathbf{x}$ consist of a set of PRI values and Hughes has provided for results for $P = 4, 6, 8, 10$ and 12. These parameters map onto an objective vector $\mathbf{y}$ consisting of $M = 9$ objectives which characterise different aspects of the radar signal, together with a final objective which is the total transmission time for the waveform.

- Objectives 1 and 2 measure the median range and velocity before the schedule is not decodable.
- Objectives 3 and 4 measure the median range and velocity before the schedule has blind regions.
- Objectives 5 and 6 measure the minimum range and velocity before the schedule is not decodable.
- Objectives 7 and 8 measure the minimum range and velocity before the schedule has blind regions.
- Objective 9 is the time required to transmit the total waveform in milliseconds.

The first 8 objectives are to be maximised, and the 9th to be minimised, however in the datasets [39] all objectives have been organised for minimisation. We use 200 randomly sampled solutions of the 11938 solutions in the 12 PRI archive containing the current approximation of the Pareto front as identified by the MSOPS algorithm.

Fig. 2 shows a heatmap of the radar data. The left hand panel shows the solutions in their original order, where each solution, $\mathbf{r}_k$ in rank coordinates, comprises a row of the heatmap. The colour-scale extends between 1 and 200 because with 200 solutions the ranks lie in this range. The central panel shows the seriated heatmap: clearly similar solutions, shown by similar colours, have been grouped together aiding interpretation.

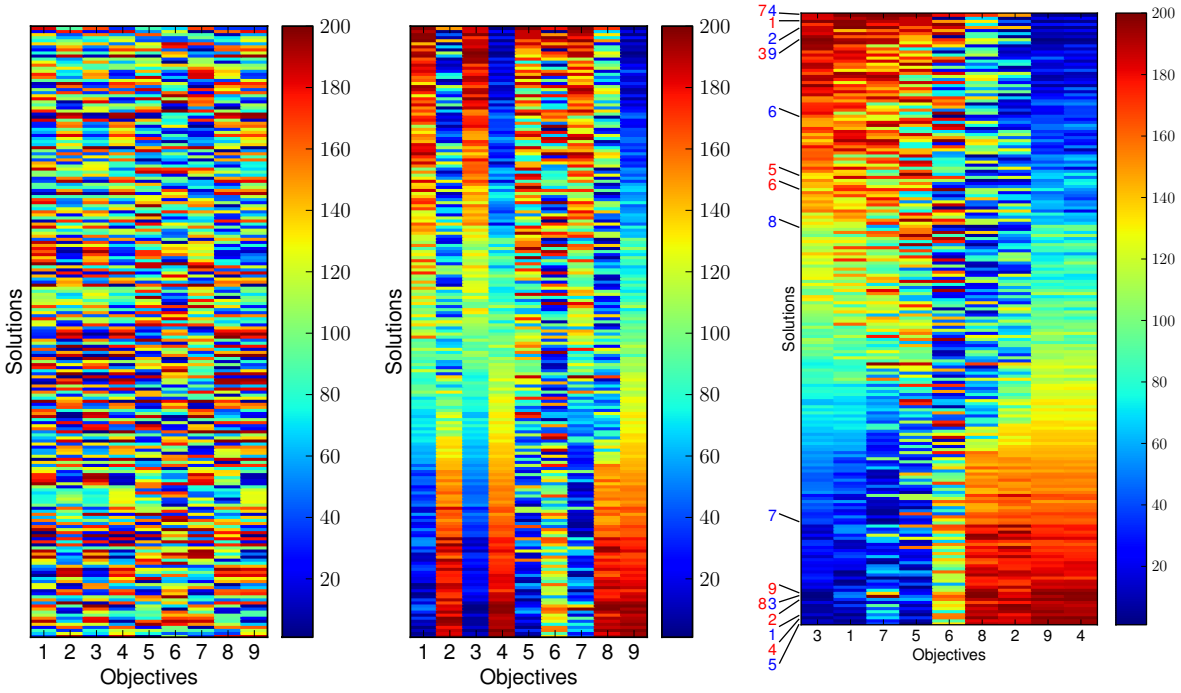
The similarity matrices $\mathbf{A}$ before seriation and after seriation, with the rows and columns permuted according $\boldsymbol{\pi}$, are shown in Fig. 3. In the seriated case, highly similar (red) solutions have been aligned along the diagonal, meaning that the permutation has placed them close together.

C. Seriation of Objectives

Although the seriation of individual solutions shown in Fig. 2(b) brings similar individuals together, the interpretability of the heatmap can be further enhanced by placing similar *objectives* together. To do this, we follow the same procedure, but using the following $M \times M$ similarity matrix, which measures how similar the m-th and n-th objectives are in terms of the average squared difference in ranks of the solutions on those two objectives:

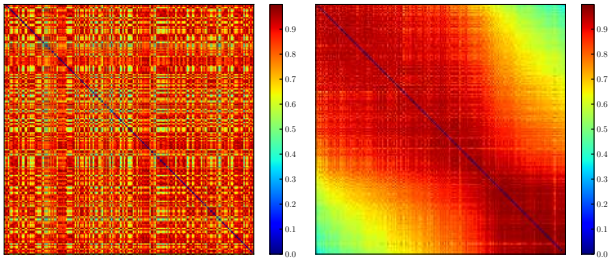
$$S_{mn} = 1 - \frac{1}{K(K-1)^2}\sum_{k=1}^{K}\left(r_{km} - r_{kn}\right)^2. \tag{6}$$

Fig. 2(c) shows the result of seriating the heatmap in Fig. 2(b) with respect to the objectives. Although in this case it would be feasible to exhaustively test all the $9!/2$ permutations, spectral seriation is significantly faster. As can be seen, similar objectives have been placed next to each other, grouping all of the range-based objectives (objectives 1, 3, 5, 7) on one side of the heatmap and the velocity-based objectives (2, 4, 6, 8) on the other. Interestingly objective 9, which measures the transmission and decoding time is placed with the velocity-based objectives and is clearly well-correlated with objectives 2 and 4. Also it is clear that objective 6 (minimum velocity



(a) Original (b) Seriation of (2a) w.r.t. Solutions (c) Seriation of (2b) w.r.t. Objectives

Fig. 2: Heatmaps of the radar dataset [10]. (a): objectives of the solution set in original order (converted to ranks); (b): solutions after seriation using similarity matrix $\mathbf{A}$, placing similar solutions together; and (c): solutions after seriation of objectives ordered to place similar objectives together using similarity matrix $\mathbf{S}$ (3). Note that seriating objectives has grouped those objectives relating to range (1, 3, 5, 7) together, as it has with those relating to velocity (2, 4, 6, 8). The numbering on the left hand side of (c) highlights the best (blue) and worst (red) solution for each of the 9 objectives.



(a) Original individual similarity (b) Seriated individual similarity

Fig. 3: Similarity matrices of the Pulsed Doppler Radar data during seriation. Left: original similarity matrix, $\mathbf{A}$. Right: $\mathbf{A}$ with rows and columns permuted according to the $\boldsymbol{\pi}$ found by seriation. Seriation collects the similar (dark red) solutions so that they lie close to the diagonal, and pushes the dissimilar solutions (cooler colours) to the edge.

before schedule cannot be decoded) is least well correlated with either group, with small values occurring in combination with large and small range and velocity objectives. We emphasise that neither of these observations about the character of the non-dominated solution set as a whole could have been made from the original heatmap (Fig. 2(a)).

D. Rank-based Objective Seriation

Rather than the squared Euclidean distance used in the similarity (6), natural metrics for measuring the similarity between permutations or ranks are Spearman's footrule [40], [41], [42] and Kendall's τ metric [43]; here we examine their use in seriation.

Note that the ranks of the solutions on a single objective m can be written as a K-dimensional vector $\boldsymbol{\rho}_m = (r_{1m}, \ldots, r_{Km})$, which is a permutation of the integers $1, \ldots, K$.[2] Spearman's footrule is the city block distance between two permutations [42]. Given permutations of the m-th and n-th objectives, $\boldsymbol{\rho}_m$ and $\boldsymbol{\rho}_n$, Spearman's footrule is the summed absolute difference between the positions of solutions in the two permutations:

$$D_{mn} = \sum_{k=1}^{K} |r_{km} - r_{kn}| \,. \tag{7}$$

The maximum possible value of the metric $D_{\max}$ is

$$D_{\max} = \begin{cases} K^2/2 & K \text{ even} \\ (K+1)(K-1)/2 & K \text{ odd.} \end{cases} \tag{8}$$

We therefore define a similarity between permutations $\boldsymbol{\rho}_m$ and $\boldsymbol{\rho}_n$ as:

$$S_{mn} = D_{\max} - D_{mn}. \tag{9}$$

[2]Note that $\boldsymbol{\rho}_m$ is K-dimensional vector of ranks corresponding to objective m, while $\mathbf{r}_k$ is the M-dimensional vector of ranks pertaining to solution $\mathbf{y}_k$.

TABLE I: Comparison of the permutations produced using spectral seriation with Euclidean distance, Kendall's τ and Spearman's footrule similarities, against optimum quality found by exhaustive search. Each row of the table shows the quality of the permutation found using a particular similarity matrix evaluated using the objective function $g(\boldsymbol{\pi})$ for the corresponding column. Bold entries indicate that the best possible quality has been located by the spectral method.

	$g_{\text{EUC}}(\boldsymbol{\pi})$	$g_{\text{TAU}}(\boldsymbol{\pi})$	$g_{\text{SPF}}(\boldsymbol{\pi})$
$\boldsymbol{\pi}_{\text{EUC}}$	$\mathbf{7.9038 \times 10^2}$	5.2483×10^6	$\mathbf{2.6398 \times 10^6}$
$\boldsymbol{\pi}_{\text{TAU}}$	7.9053×10^2	$\mathbf{5.2477 \times 10^6}$	2.6445×10^6
$\boldsymbol{\pi}_{\text{SPF}}$	$\mathbf{7.9038 \times 10^2}$	5.2483×10^6	$\mathbf{2.6398 \times 10^6}$

Kendall's τ metric considers how the pairwise ordering of individuals differs between permutations. If the relative ordering between individuals k and j on objectives m and n is unchanged ($r_{km} > r_{jm}$ and $r_{kn} > r_{jn}$ or $r_{km} < r_{jm}$ and $r_{kn} < r_{jn}$) then a cost is defined as $\tau_{kj}(\boldsymbol{\rho}_m, \boldsymbol{\rho}_n) = 0$; otherwise, if the pairwise relative order is different, $\tau_{kj}(\boldsymbol{\rho}_m, \boldsymbol{\rho}_n) = 1$. The costs are summed to arrive at the overall metric:

$$\tau_{mn} = \sum_{j=1}^{K} \sum_{k=1}^{K} \tau_{kj}(\boldsymbol{\rho}_m, \boldsymbol{\rho}_n). \tag{10}$$

Here we have assumed that there are no tied ranks, but alternative formulations are available to cope with tied ranks [44]. The maximum value of the τ metric, $\tau_{\max}$, occurs when $\boldsymbol{\rho}_m$ is the reverse ordering of $\boldsymbol{\rho}_n$, and is:

$$\tau_{\max} = K(K-1)/2. \tag{11}$$

Kendall's τ, like Spearman's footrule, is a metric and increases in value as the two permutations become more dissimilar. We therefore define a similarity:

$$S_{mn} = \tau_{\max} - \tau_{mn}. \tag{12}$$

To demonstrate the use of permutation metrics for computing similarity and seriation, we seriated the objectives in a variant of the radar data. The new dataset consisted of the original 9 objectives together with two additional objectives inserted at random positions in the data. The first additional objective y_A was produced by averaging two existing, well correlated, objectives (objectives 1 and 3): $y_{kA} = (y_{k1} + y_{k3})/2$. The second additional objective y_B consists of uniform random noise; this objective is therefore expected to be uncorrelated with the rest of the data. We expect that seriation will move the well-correlated objective close to the objectives from which it was constructed, while the uncorrelated objective is expected to be placed away from the correlated groups. The result of seriating this new dataset is shown in Fig. 4. As can be seen, in all cases the well-correlated objective (objective A) is placed next to both of the objectives upon which it is based. In addition, the uniform random objective (objective B) is placed in the middle of the heatmap, between the two groups of well correlated objectives.

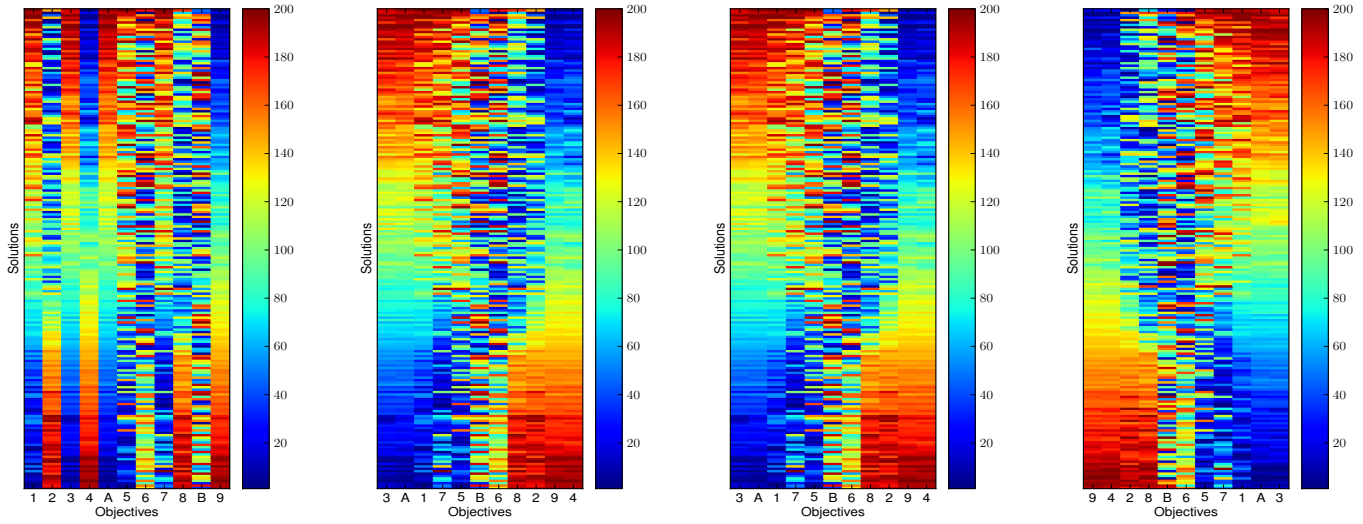
Returning to the original 9-objective radar archive, it is feasible to evaluate all of the $9!/2$ distinct permutations of objectives in order to examine the quality of the approximate solution under the three metrics. Table I presents the results of this analysis. If $g_{\mathbf{S}}(\boldsymbol{\pi}) = \sum_m \sum_n S_{mn}(\pi_m - \pi_n)^2$ is the quality function to be minimised, then each row of the table gives the quality of the permutation which was found by spectrally seriating with the similarity matrix for that row (equations (6), (12) and (9)), but with the quality $g(\boldsymbol{\pi})$ evaluated using the similarity matrix for the corresponding column; for example, on the first row, $\boldsymbol{\pi}_{\text{EUC}}$ is the permutation found by spectrally seriating according to the Euclidean distance metric. We then consider the quality of this permutation using the Euclidean similarity matrix and the similarities based on Kendall's τ and Spearman's footrule. Qualities highlighted in bold indicate that the permutation is the best found by exhaustive search. Reassuringly, as indicated by the bold diagonal entries, for each of the similarity matrices, the spectral seriation has located the best permutation. However, as the spectral method is an approximation this will not always be the case, especially for more objectives. Interestingly we note that the Euclidean and Spearman's footrule similarities perform identically and the only difference with the Kendall's τ seriation is the reversed order of objectives 9 and 4 (see Fig. 4). We observe that, as here, seriations according to Spearman's footrule and Kendall's τ are usually very similar, perhaps unsurprisingly in light of the inequalities $\tau_{mn} \leq D_{mn} \leq 2\tau_{mn}$ [42]. In general we recommend Spearman's footrule for its simplicity and speed of calculation.

IV. Joint Seriation of Decision and Objective Spaces

In the previous section we showed how spectral seriation can be used to reorder objective space heatmaps to enhance their interpretability. It is often of interest to view the parameter space solutions $\{\mathbf{x}_k\}$ alongside their objective space counterparts. Examples of visualisations that incorporate information from both parameter and objective space are [3] and [45]; the method proposed in [3], employing a heatmap, is of particular interest and we discuss it later in this section. In this section we show how to simultaneously optimise the parameter space and objective space views. We assume that the parameters are real valued $x_{km} \in \mathbb{R}$ which allows meaningful distances between parameters to be calculated. As with the objective space case, it is necessary to normalise the parameter values so that the heatmap uses the full range of colours to represent them.

A straightforward way of jointly seriating parameter and objective space is shown in Fig. 5 for a population of solutions to the test problem WFG8 from the standard Walking Fish Group test problem suite [12]. The population comprises 200 solutions to a 10-objective instance of the problem, where the number of parameters is 38. The solutions were sampled from the known Pareto optimal set. Here, the solutions $\mathbf{y}_k$ have been seriated in objective space with respect to objectives (reordering the columns) and then solutions (reordering the rows), which yields the heatmap shown in the lower right-hand panel of Fig. 5.[3] Then, solutions $\mathbf{x}_k$ in parameter space

[3]Note that the same result would have been obtained by seriating first with respect to solutions (rows), and then by objectives (columns).



(a) Seriated w.r.t. solutions (b) Euclidean distance (c) Spearman's footrule (d) Kendall's τ metric

Fig. 4: Seriation according to different metrics. The radar data has been modified to include an objective produced by summing objectives 1 and 3 (objective A) and an objective entirely comprised of uniform random noise (objective B). Prior to seriation of objectives, the solutions have been seriated to produce the population shown in (a).

have been reordered to match the order of the solutions in the seriated objective space (that is, the rows of the bottom lefthand panel), and finally the parameters (columns) were seriated to place similar parameters close to each other. As is common for many test problems, the parameters of WFG8 are grouped into two types. In this instance of the problem, 18 of the parameters are *position* parameters, which control the region of the Pareto front on which the solution lies. The remaining 20 are *distance* parameters, controlling the distance of the solution from the Pareto front. As can be seen in the figure, the distance parameters have been grouped together in the centre of the heatmap. The position parameters, which consist of a uniform spread of values, have been moved to the sides of the heatmap. In a similar manner to objective space seriation, the order of parameters was seriated by using spectral seriation to approximately minimise

$$g_{\mathbf{x}}(\boldsymbol{\pi}) = \sum_{j=1}^{K}\sum_{k=1}^{K} \Lambda_{kj}(\pi_k - \pi_j)^2. \tag{13}$$

The objective function $g_{\mathbf{x}}(\boldsymbol{\pi})$ requires a parameter space measure of the similarity of solutions. Since ranking solutions in parameter space is meaningless we measure the parameter space similarity of solutions $\mathbf{x}_k$ and $\mathbf{x}_j$ using their correlation or the well-known cosine similarity:

$$\Lambda_{kj} = \frac{\sum_p x_{kp}x_{jp}}{\sqrt{\sum_p x_{kp}^2 \sum_p x_{jp}^2}}, \tag{14}$$

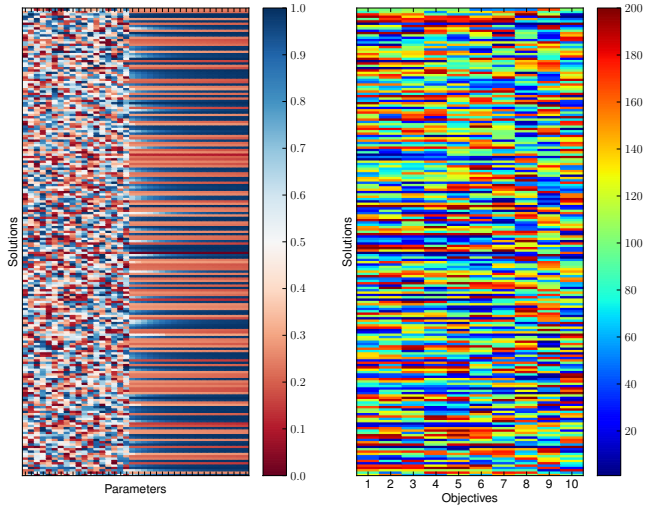
which is the cosine of the angle between $\mathbf{x}_k$ and $\mathbf{x}_j$. We also provide an example later using the negative mean difference of solutions

$$\Lambda_{kj} = -\left|\sum_{p=1}^{P}(x_{kp} - x_{jp})\right| \tag{15}$$

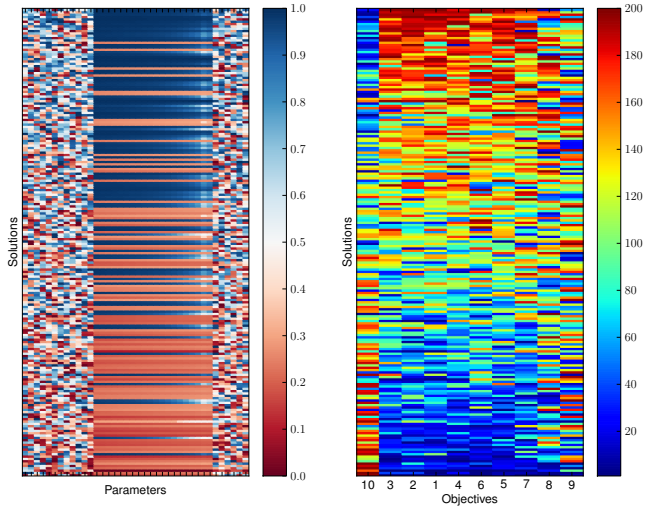
which has the effect of placing solutions with parameters of the same magnitude together. However, particular optimisation problems may suggest alternative measures of parameter space similarity to those used here.

As Fig. 5 shows, the resulting ordering of parameter-space solutions (induced by the objective space ordering of solutions) has in general placed those solutions with larger parameter values together at the top of the heatmap and the seriation has revealed a clear correlation of larger parameter values with solutions which optimise objective 10 well, while the remainder of the objectives are best optimised by small parameter values. Clearly, however, reordering the parameter space solutions by the objective space seriation has not induced the same improvement in clarity as we have previously demonstrated in objective space; in a perfect reordering, *all* of the large-parameter solutions would reside in the top half of the heatmap, and all of the small-parameter solutions would be placed near the bottom. We discuss an approach to resolving this later in this section. The seriation of parameters has, however, produced a good result. Later in this section we show a similar seriation for solutions to the radar data.

Fig. 6 presents the same data, this time using the visualisation method presented by [3]. Here, unlike the method we propose, the objectives and parameters are visualised with a single heatmap, and the columns of the heatmap are clustered so that both objectives and parameters are reordered together. Solutions are clustered with single linkage hierarchical clustering based on the Euclidean distance between their normalised objective values. Similarly, objectives and parameters are clustered together. As in Fig. 5, the distance parameters have been gathered together, however the position parameters and objectives are intermixed. This makes observation of the trade-off between objectives more difficult and we prefer a visualisation that keeps the two spaces separate. Additionally,



(a) Before seriation



(b) After seriation

Fig. 5: Seriation of both objective and parameter spaces for the WFG8 test problem. The upper panels show heatmaps of parameter space (left) and objective space (right). The lower panels show the result of first seriating the objectives, followed by the solutions according to their objective space similarity (bottom right). The resulting solution ordering is then applied to the solutions in parameter space, and the parameters themselves are seriated to yield the bottom left heatmap.

the need to cluster both spaces together requires that a common normalisation for parameters and objectives be found.

Whilst the example in Fig. 5 was seriated according to objective space solutions, after which parameter-space solutions were reordered to match the objective-space seriation, it could justifiably be done the other way around. However, there is clearly a trade-off between the quality of the solution orderings in parameter and objective spaces. The simultaneous clustering method [3] obscures the trade-off, giving unknown relative weights to parameters and objective spaces. Here we therefore seek to simultaneously optimise the ordering in both by using a two-objective evolutionary algorithm to locate the

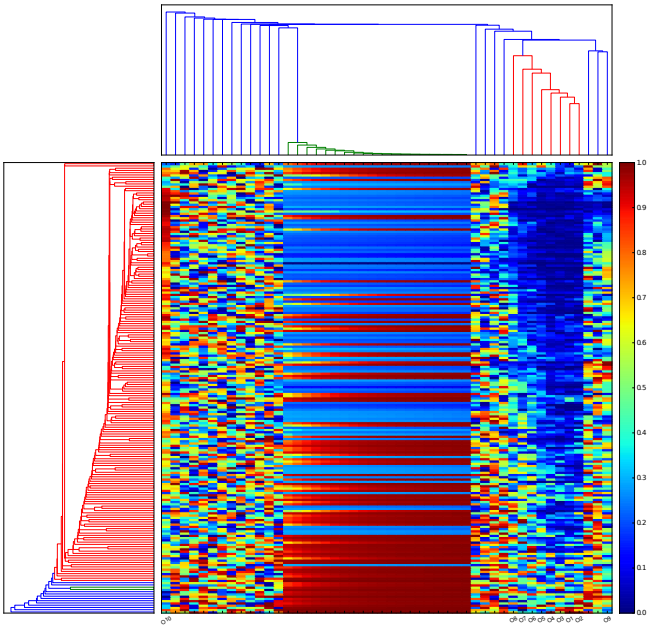


Fig. 6: A heatmap of the WFG8 parameters and objectives, reordered with hierarchical clustering as proposed by [3]. As in the other heatmaps presented here, a row represents a solution and a column is either an objective or a parameter (columns representing an objective are marked on the abscissa). Solutions were clustered with single linkage clustering based on the Euclidean distance between normalised objective values, and objectives and parameters are clustered together. Dendrograms show the clustering for solutions (on the left) and objectives and parameters (top).

approximate the Pareto front between $g_{\mathbf{y}}(\boldsymbol{\pi})$ and $g_{\mathbf{x}}(\boldsymbol{\pi})$ (cf. (4) and (13)).

We use a $(\mu + \lambda)$–evolution strategy with a passive elite archive to explore the trade-off between the quality of the ordering in the two spaces as described in Algorithm 1.[4] A population of μ permutations is maintained and at each generation is mutated to generate λ offspring, which are evaluated under the two objective functions (lines 5 and 6). Non-dominated permutations are added to the elite archive E and any permutations which the new entrants dominate are removed (line 7). The best μ permutations of a Pareto sorting [47] of the union of the parent and child populations are retained to form the parent population for the next generation (line 8).

Mutation was achieved by block transposition, in which a block of elements are swapped with a second block of elements, and shuffle transposition, in which the elements in a block are shuffled at random [48]. Half of the time a single randomly-chosen method was used; otherwise both methods were used in a randomly chosen order. In all cases the block length was randomly chosen in the range $[1, \lfloor K/10 \rfloor]$.

The population was initialised with permutations generated by seriating using a convex combination of the objective and parameter space similarities:

$$\mathbf{S}_\eta = \eta\mathbf{A} + (1 - \eta)\mathbf{\Lambda} \tag{16}$$

[4]This differs from the $(\mu + \lambda)$–evolution strategy variants of the popular Pareto Archived Evolution Strategy [46], which use an active archive.

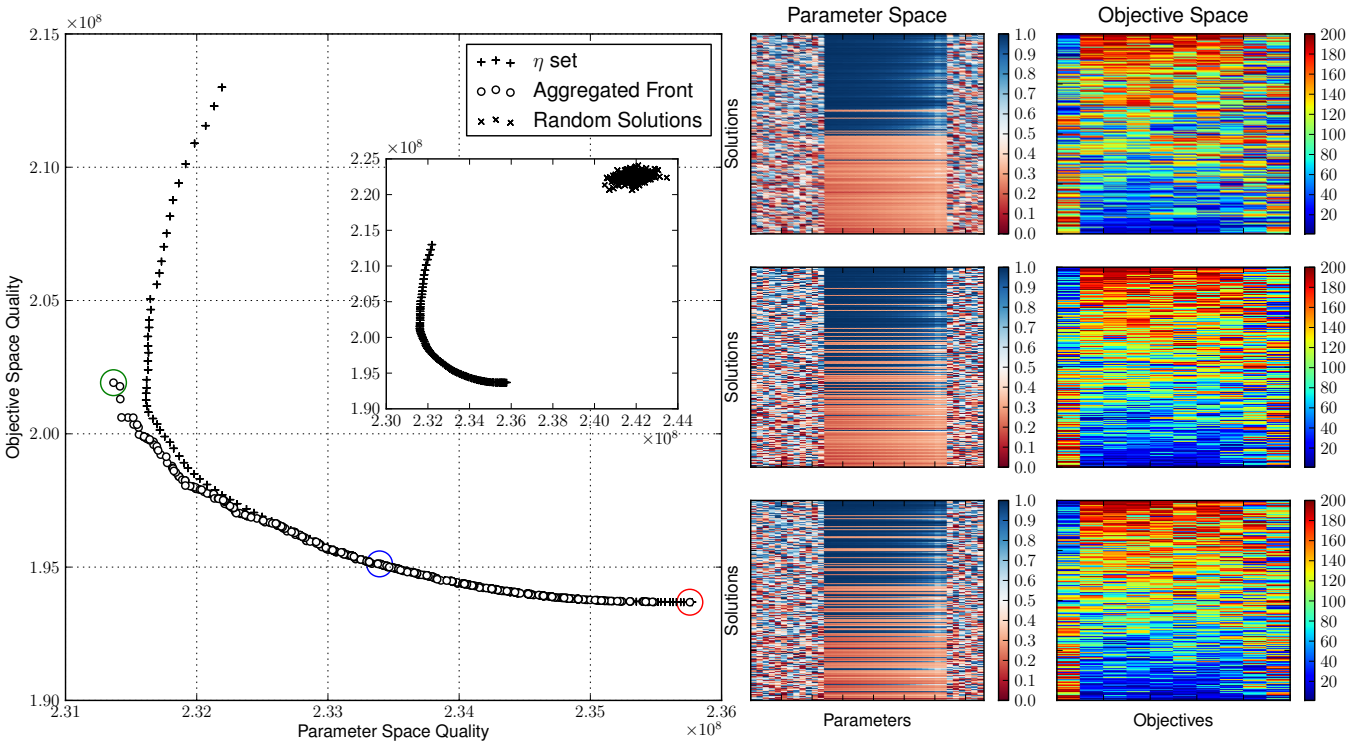


Fig. 7: Optimised trade-off between seriation quality in parameter space $g_{\mathbf{x}}(\boldsymbol{\pi})$ and objective space $g_{\mathbf{y}}(\boldsymbol{\pi})$ for the WFG8 solution set. The main panel shows the initial solutions from seriation using the similarity matrix $\mathbf{S}_\eta$ (16) as plus symbols (also shown in the insert) and the combined estimated Pareto front from 10 runs of the evolutionary optimiser. The heatmaps on the right-hand side of the figure correspond to optimised solutions marked on the main panel by large circles. The top and bottom visualisations are the solution orderings found by the MOEA which best optimise the parameter space ordering and objective space ordering respectively. The central heatmaps show a solution towards the centre of the Pareto front. Objectives and parameters have also been seriated independently. Parameter space seriation has grouped parameters $\mathbf{x}$ roughly into those which control the distance of a solution from the true Pareto front and those which control the angular location of a solution on the front.

Algorithm 1 Multi-objective $(\mu + \lambda)$–evolution strategy with a passive elite archive for seriation.

1 :	$\{\boldsymbol{\pi}_i\}_{i=1}^{\mu} :=$ `initialise_permutations`()	*Initialise μ permutations*
2 :	$\{\mathbf{v}_i\}_{i=1}^{\mu} := \{(g_{\mathbf{x}}(\boldsymbol{\pi}_i), g_{\mathbf{y}}(\boldsymbol{\pi}_i))\}_{i=1}^{\mu}$	*Evaluate the permutations in terms of the seriation qualities*
3 :	$E :=$ `extract_nondominated`$(\{(\mathbf{v}_i, \boldsymbol{\pi}_i)\}_{i=1}^{\mu})$	*Initial estimate of the Pareto set*
4 :	for $t := 1 : s$	*For s generations*
5 :	$\quad\{\boldsymbol{\pi}'_i\}_{i=1}^{\lambda} :=$ `mutate`$(\{\boldsymbol{\pi}_i\}_{i=1}^{\mu})$	*Mutate parents to create children*
6 :	$\quad\{\mathbf{v}'_i\}_{i=1}^{\lambda} := \{(g_{\mathbf{x}}(\boldsymbol{\pi}'_i), g_{\mathbf{y}}(\boldsymbol{\pi}'_i))\}_{i=1}^{\lambda}$	*Evaluate children*
7 :	$\quad E :=$ `update_elite_archive`$(E, \{\mathbf{v}'_i\}_{i=1}^{\lambda}, \{\boldsymbol{\pi}'_i\}_{i=1}^{\lambda})$	*Update estimate of the Pareto set*
8 :	$\quad\{(\mathbf{v}_i, \boldsymbol{\pi}_i)\}_{i=1}^{\mu} :=$ `sort_and_extract`$(\{(\mathbf{v}_i, \boldsymbol{\pi}_i)\}_{i=1}^{\mu} \cup \{(\mathbf{v}'_i, \boldsymbol{\pi}'_i)\}_{i=1}^{\lambda}, \mu)$	*Combine the populations, sort, and extract the μ highest ranked*
9 :	end	

with η chosen at equal intervals from $\eta = 0$ to $\eta = 1$, so that a range of permutations is produced from those focusing entirely on parameter space similarity to those focusing entirely on objective space. In the results presented here the population size was 100, and we set $\mu = \lambda$, so each parent is mutated into a single child.

Fig. 7 shows the combined non-dominated permutations from 10 runs for the 200 WFG8 solutions and objectives used previously. As the main panel shows, the solutions after optimisation are very close to the initial solutions found via equation (16).

The inset panel shows the objectives resulting from the seriations of $\mathbf{S}_\eta$ used to initialise the evolutionary population, together with the $g_{\mathbf{x}}(\boldsymbol{\pi})$ and $g_{\mathbf{y}}(\boldsymbol{\pi})$ corresponding to 200 randomly chosen permutations. Clearly the initialisation using $\mathbf{S}_\eta$ provides a very good approximation to the trade-off between $g_{\mathbf{x}}(\boldsymbol{\pi})$ and $g_{\mathbf{y}}(\boldsymbol{\pi})$ as identified by the MOEA. Although the MOEA has improved the front slightly, its main effect in this case has been to remove dominated members of the seeded initialisation set and to fill in the gaps in the initialisation set. Whilst it is useful for a decision maker to have a full Pareto front on which to base their selection of

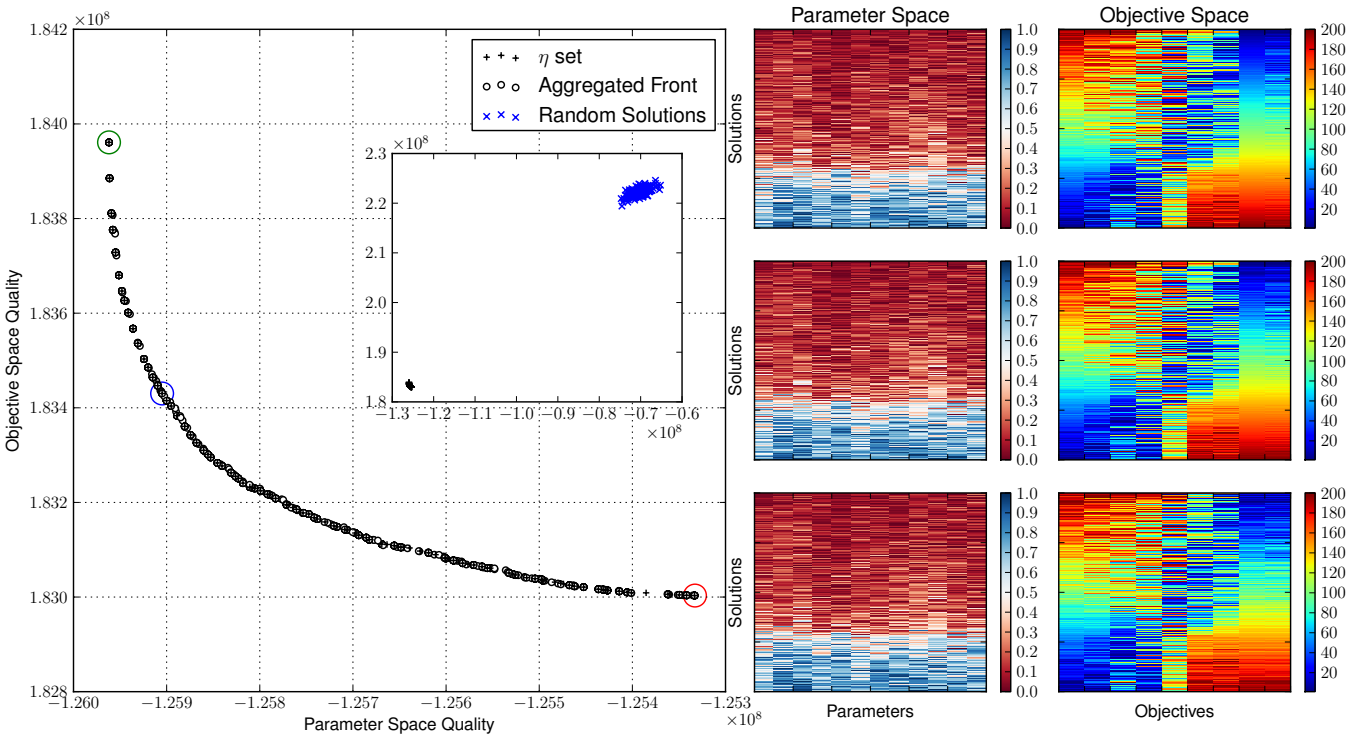


Fig. 8: Optimised trade-off between seriation quality in parameter space $g_{\mathbf{x}}(\boldsymbol{\pi})$ and objective space $g_{\mathbf{y}}(\boldsymbol{\pi})$ for the radar data. The main panel shows the initial solutions from seriation using the similarity matrix $\mathbf{S}_\eta$ (16) as pluses (also shown in the insert) and the combined estimated Pareto front from 10 runs of the evolutionary optimiser. The heatmaps on the right-hand side of the figure correspond to optimised solutions marked on the main panel by large circles. The top and bottom visualisations are the solution orderings found by the MOEA which best optimise the parameter space ordering and objective space ordering respectively. The central heatmaps show a solution towards the centre of the Pareto front. Objectives and parameters have also been seriated independently.

operating point, the MOEA has failed to uncover much beyond the original initialisation set, although of course the dominated solutions corresponding to low objective-space qualities $g_{\mathbf{y}}(\boldsymbol{\pi})$ have been eliminated. The heatmaps on the righthand side of the main panel show the seriations produced by solutions along the Pareto front approximation. The top heatmaps represent the solution highlighted at the top of the approximate front which is the best ordering with respect to parameter space. The bottom heatmaps show the solution highlighted at the bottom of the approximate front, namely the best objective space ordering found and are essentially the heatmaps shown in the bottom row of Fig. 5. The middle heatmaps represent a compromise between the objective space and parameter space coherence. There is a clear reduction in seriation quality when parameter space is ordered in terms of objective space similarity and vice versa. However, it may be worth accepting this compromise in order to be able to view the two spaces together.

We also demonstrate the result of optimising the joint seriation of parameters and objective vectors for the radar data. Fig. 8 presents the result of this optimisation, and was produced by following the same procedure as that used to produce Fig. 7 – this time utilising the negative mean difference as the parameter space solution similarity (15). As before, the MOEA was seeded with 100 permutations obtained from seriating $\mathbf{S}_\eta$ for linearly spaced η and the non-dominated permutations from the union of 10 runs are shown. Like the WFG8 population, the MOEA has found permutations which have only a marginal improvement over the initialised seriations of $\mathbf{S}_\eta$. For this particular problem, however, it is clear to see that seriating solutions in objective space also leads to a good ordering of solutions in parameter space (and vice versa) and that either seriation is difficult to improve upon. This indicates a strong correlation between solutions and objective vectors, providing useful information to the problem owner and we emphasise that this radar problem is a real problem rather than a synthetic test problem. We also draw attention to the marked improvement over random permutations when using spectral seriation with any of the $\mathbf{S}_\eta$.

It is not always possible to achieve an ordering that simultaneously groups like solutions in parameter space and objective space. Nonetheless these two examples illustrate that a seriation which compromises between parameter space grouping and objective space grouping can be a helpful visualisation, particularly as it allows the investigator to assess objectives and parameters together.

Although on a limited number of examples, these results indicate that seriation of convex combinations of $\mathbf{A}$ and $\boldsymbol{\Lambda}$ provides a very good indication of the approximation to the Pareto front achieved by the MOEA. We note that there may be examples where a MOEA can improve more significantly on the convex combination defined in (16). Here we have

used the cosine similarity and negative mean difference for measuring the proximity of parameter vectors, however, the choice of similarity is less clear than for objective space, where conversion to ranks simplifies the choice. Other similarity measures may be more useful for other specific problems, particularly for categorical parameters for which there is no natural ordering.

Visualising both the parameter and objective spaces together can be useful for a decision maker, and this section has presented a lossless technique for producing such a visualisation. As heatmaps are not restricted in terms of dimensionality or the number of solutions that they can represent, they are an important tool for the many-objective optimisation community, and we have outlined a simple technique for enhancing their clarity to support their use in decision making. In addition to a relatively low computational complexity, the spectral seriation method is flexible in that it allows the choice of a similarity measure that specifically suits the type of optimisation to be visualised.

V. Visualisation in the Plane

Visualising solutions as points in the two-dimensional plane is appealing and intuitive. People are well adapted to understanding maps and diagrams in which physical proximity of points indicates that the points are similar, and Gestalt theories suggest how the human perceptual system can organise groupings of points into perceptually significant clusters. Methods such as PCA, SOM, Neuroscale and the GTM are all general purpose dimension reduction techniques that can be used to represent M-dimensional objective vectors as coordinates in the plane; they all attempt, explicitly or implicitly, to preserve in the plane the distances between points in the original M dimensions. These methods for general data are however ignorant of the mutually non-dominating nature of the solutions to multi-objective problems. In this section and section VI we examine two techniques for obtaining planar representations of a set of solutions which use the mutually non-dominating nature of the solutions.

The first method is closely related to RadViz [9], [49] which, along with related methods, has been extensively used for visualising multivariate data in "radial" coordinates [50], [51]; see [52] for an insightful derivation of the properties of RadViz and related "normalised radial visualisations". A frequently used analogy for understanding RadViz involves "anchor points" which are chosen on the circumference of a circle in the plane. Each of these anchor points is associated with a coordinate in the high dimensional space. The point to be visualised is imagined to be connected to the anchor points by springs whose stiffness is proportional to the point's corresponding high-dimensional coordinate; it is visualised in the plane by the location at which the spring forces are in equilibrium. To our knowledge RadViz has not be used for visualising sets of mutually non-dominating solution objectives. In this section we therefore give a new derivation of the RadViz algorithm using barycentric coordinates and emphasise why it may be useful for visualising mutually non-dominating sets.

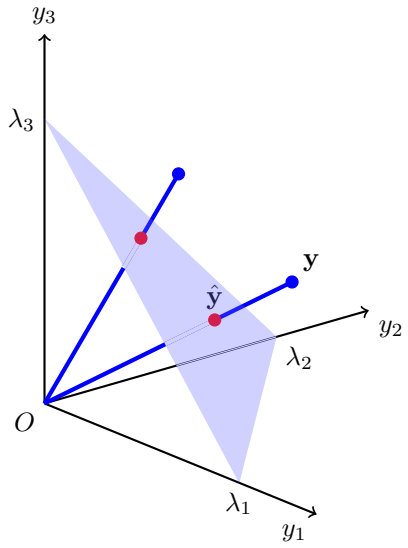


Fig. 9: Projection of a point $\mathbf{y}$ onto the simplex defined by $\{\lambda_1, \lambda_2, \lambda_3\}$; $\mathbf{y}$ is projected to $\hat{\mathbf{y}}$.

Put simply, barycentric coordinates are used to map solutions from M dimensions to the interior of a regular planar polygon with M vertices. Without loss of generality, we assume that the solutions $\mathbf{y}_k$ to be visualised are all non-negative ($y_{km} \geq 0$ for all k, m). Then the *simplex* defined by the numbers $\{\lambda_m : \lambda_m > 0\}_{m=1}^{M}$ is the portion of the (hyper-) plane that lies in the positive orthant and which intersects the coordinate axes at distances λ_m from the origin, as illustrated in Fig. 9. The simplex is therefore the segment of the plane in the positive orthant defined by

$$\mathbf{n} \cdot \mathbf{y} = d \qquad y_m \geq 0, \ m = 1, \ldots, M \tag{17}$$

where the elements of the unit vector $\mathbf{n}$ normal to the simplex are $n_m = d/\lambda_m$ and the perpendicular distance to the origin d can be found as

$$d^{-2} = \sum_{m=1}^{M} \lambda_m^{-2}. \tag{18}$$

We project the mutually non-dominating set of solutions onto the simplex by:

$$\hat{\mathbf{y}}_k = \mathbf{y}_k / (\mathbf{y}_k \cdot \mathbf{n}). \tag{19}$$

Importantly, note that a mutually non-dominating set remains mutually non-dominating after projection onto the simplex, so dominance relations are preserved by this projection. However, since all points on the ray from the origin through $\mathbf{y}$ are projected to $\hat{\mathbf{y}}$ this projection obscures the relations in general sets of points.

The $\hat{\mathbf{y}}_k$ are now mapped to the plane using barycentric coordinates as follows. Barycentric coordinates are well known for triangles, but the idea is readily generalised to convex polygons in the plane and to simplices in more than two dimensions such as the vertices of the simplex described above [53], [54]. Let $\boldsymbol{\lambda}_m = (0, \ldots, \lambda_m, \ldots, 0)$ be the position vector of the m-th vertex. Then the barycentric coordinates of $\hat{\mathbf{y}}$ are

defined as the weights ω_m so that

$$\hat{\mathbf{y}} = \sum_{m=1}^{M} \omega_m \boldsymbol{\lambda}_m \tag{20}$$

with the constraints that $\omega_m \geq 0$ and $\sum_m \omega_m = 1$. When $\hat{\mathbf{y}}$ is close to vertex $\boldsymbol{\lambda}_m$ then ω_m is large and the other ω_n, $n \neq m$, are small. If, instead, the $\{\boldsymbol{\lambda}_m\}_{m=1}^{M}$ define the vertices of a convex polygon in the plane, then the ω_m correspond to the weights that must be placed at the vertices in order to balance the polygon on a spike at $\hat{\mathbf{y}}$. Note that there is a redundancy in the M barycentric coordinates because $M-1$ of them completely determine the remaining one through the constraint that they sum to 1; in an M-dimensional space this may be viewed as arising from the constraint that $\hat{\mathbf{y}}$ is confined to the $(M-1)$-plane of the simplex.

The vector of the first $M-1$ barycentric coordinates corresponding to $\hat{\mathbf{y}}$ is found as

$$\boldsymbol{\omega}_{1:M-1} = \mathbf{B}^{\dagger}(\hat{\mathbf{y}} - \boldsymbol{\lambda}_M) \tag{21}$$

where $\mathbf{B}$ is the $M \times (M-1)$ matrix whose m-th column is $\boldsymbol{\lambda}_m - \boldsymbol{\lambda}_M$, and $\mathbf{B}^{\dagger} \equiv (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}$ denotes the pseudo-inverse of $\mathbf{B}$. Then $\omega_M = 1 - \sum_{m=1}^{M-1} \omega_m$.

Let $\mathbf{v}_m$, $m = 1, \ldots, M$ be the vertices of a regular polygon $\mathbb{P}$ in the plane, centred on the origin:

$$\mathbf{v}_m = \begin{bmatrix} \cos(2\pi(m-1)/M) \\ \sin(2\pi(m-1)/M) \end{bmatrix}. \tag{22}$$

An M-dimensional point in the simplex $\hat{\mathbf{y}}$ is mapped to the point $\mathbf{z}$ in $\mathbb{P}$ that has the same barycentric coordinates $\boldsymbol{\omega}$ as $\hat{\mathbf{y}}$, namely

$$\mathbf{z} = \mathbf{V}\boldsymbol{\omega} \tag{23}$$

where $\mathbf{V}$ is the matrix whose columns are the vectors $\mathbf{v}_m$. The vertices of $\mathbb{P}$ are identified with the objectives to be minimised. Thus a solution $\mathbf{y}_k$ that has a large value y_{km} for the m-th objective will be mapped close to the m-th vertex $\mathbf{v}_m$. Solutions which have approximately equal values on all objectives will be mapped close to the centre of the polygon.

There remains an additional degree of freedom because the identification of which vertices of $\mathbb{P}$ correspond to which objectives has not been made. If $\mathbf{P}$ is an M-dimensional permutation matrix, then (23) may be modified to

$$\mathbf{z} = \mathbf{V}\mathbf{P}\boldsymbol{\omega} \tag{24}$$

where we are at liberty to choose the permutation. Since the vertices may be considered to lie on a circle and clockwise and anti-clockwise permutations are equivalent there are $(M-1)!/2$ distinct permutations. Methods such as VizRank [55], which are used for visualising data labelled into classes, use heuristic search to locate permutations which give good separations of the classes in the visualisation plane. Here solutions do not belong to naturally defined classes[5] although it would be possible to search for permutations which yield good clusters in the visualisation. A further possibility to determine the order of the objectives around the vertices of $\mathbb{P}$ is to minimise a stress measuring how well near-neighbour distances are preserved in the projection into $\mathbb{P}$; see [5] and [56] for examples of this approach in other contexts. Here we elect to order the objectives around the polygon so that similar objectives are placed near to each other. Empirically we find that this tends to place solutions which are near neighbours in objective space close to each other in the planar representation. As above, we use Spearman's footrule to measure the similarity of two objectives, f_m and f_n, by the sum of the absolute differences between the ranks of solutions on those two objectives:

$$S_{mn} = D_{\max} - \sum_{k=1}^{K} |r_{km} - r_{kn}| \tag{25}$$

This periodic seriation problem may be solved in a similar manner to the linear seriation already described [57].

In summary, the visualisation is found by projecting $\mathbf{y}$ to $\hat{\mathbf{y}}$ on the simplex using (19), which allows barycentric coordinates $\boldsymbol{\omega}$ to be found via (21), after which the coordinates in the plane $\mathbf{z}$ are calculated with (24). The order of the vertices is chosen by periodic seriation using (25) to measure the similarity of objectives.

The computational complexity of the projection (dominated by finding the pseudo-inverse (21)) and seriation is $O(M^3)$. For the numbers of objectives typically involved in many-objective optimisations the computation is sufficiently fast to be incorporated in interactive tools.

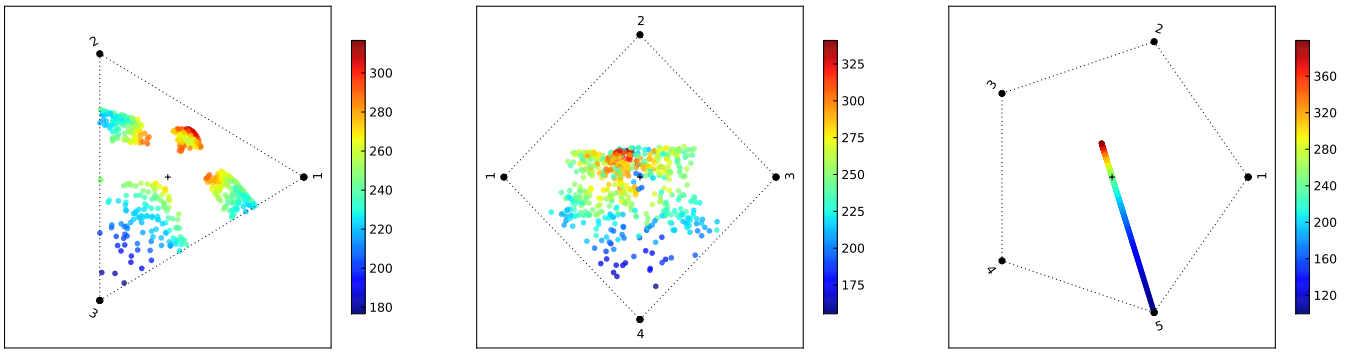
Fig. 10 shows the visualisation of three examples from well-known test problems. The problem DTLZ6 [11] has a Pareto front comprising several disconnected "cushions"; the four clusters are evident in the visualisation of 500 solutions on the true Pareto front shown in Fig. 10(a) for $M = 3$ objectives. If the planar representation $\mathbf{z}_k$ of a solution $\mathbf{y}_k$ lies close to vertex $\mathbf{v}_m$ then the objective y_{km} tends to be large, or the objectives on the *opposite* vertices are small. Solutions with similar values for all objectives tend to be mapped close to the centre of the polygon.

The solutions are coloured by their average rank:

$$\bar{r}_k = \frac{1}{M}\sum_{m=1}^{M} r_{km}. \tag{26}$$

The visualisation reveals the symmetry between objectives y_1 and y_2. It is clear that very good (low, blue) average rank solutions are achieved only at the expense of large y_3; likewise intermediate average rank solutions (average rank ≈ 230, coloured cyan) are found close to the $y_3 - y_1$ and $y_3 - y_2$ edges which correspond to small values of y_2 and y_1 respectively. However, the projective nature of the visualisation means that it is unable to uncover the distinct clusters in the DTLZ6 problem with $M = 4$ objectives, as shown in Fig. 10(b) (none of the alternative permutations of the vertex order are more successful). Nonetheless, it does reveal the symmetry between objectives y_1 and y_3 and the distinguished axis y_4, and in conjunction with the average rank colouring, permits the decision maker to explore regions of low and high rank solutions. As a final test problem example we show in Fig. 10(c) the visualisation of the degenerate front for the

[5] One possibility would be to assign labels according to the objective on which the solution has best rank, as done to colour solutions in Fig. 1.



(a) DTLZ6: $M = 3$ objectives (b) DTLZ6: $M = 4$ objectives (c) WFG3: $M = 5$ objectives

Fig. 10: Visualisation in the plane using RadViz. Vertices of the polygon are labelled by the objective to which they correspond. A solution close to a vertex m tends to have a large y_m, while solutions with approximately equal objectives lie close to the centre of the polygon. Solutions are coloured by their average rank.

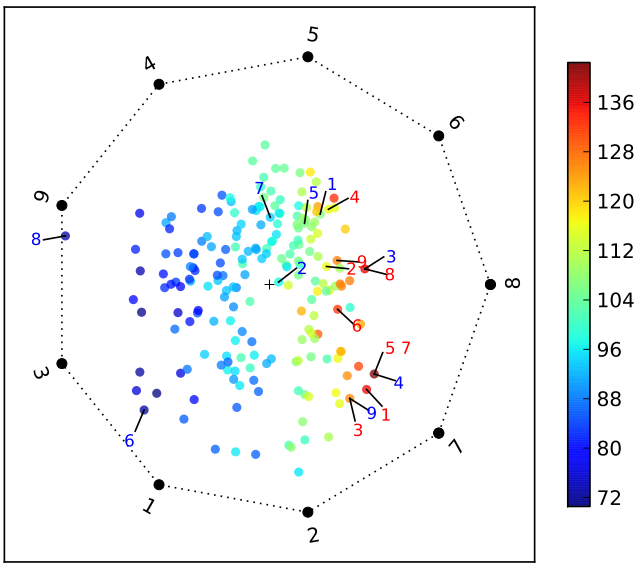


Fig. 11: Visualisation in the plane of the radar data solutions. Vertices of the polygon are labelled by the objective to which they correspond. Solutions are coloured by their average rank. The best and worst solutions for each objective are labelled in blue and red respectively. A solution close to a vertex m has a large y_m.

Walking Fish Group problem WFG3 with 5 objectives. As the visualisation clearly shows this front is a one-dimensional curve through objective space, terminating on the y_5 axis. The visualisation maps straight lines in objective space to straight lines in the visualisation plane [52].

Fig. 11 shows the 200 solutions of the 9-objective radar data used earlier mapped to the interior of a nonagon. As before solutions are coloured according to their average rank, and the best and worst solutions for each objective are labelled in blue and red respectively. Clearly there is some information loss resulting from the projection onto the plane and the y_{km} themselves cannot be simply read from this diagram, but the representation places solutions with similar objective values close together in the plane. As before, solutions which have a poor value on objective m tend to be placed close to vertex $\mathbf{v}_m$, while solutions which have a good value for an objective tend to be placed opposite that objective, for example in this plot the best solutions for y_6 and y_8, which also have low average rank. We note also that solutions with extreme values of the objectives tend to be placed around the "edges" of the visualisation in keeping with the intuition that the extremes lie on the edges. The grouping of similar objectives together with periodic seriation further enhances interpretability because if a solution tends to have large values for a group of objectives it can be placed close to that group.

The choice of values for the λ_m determining the simplex in objective space clearly affects the visualisation, however we find that it is insensitive to the precise choice; for example, setting λ_m to be the mean or median of y_{km} makes the visualisation relatively invariant to scaling of the objectives. Rather than a linear scaling, however, we prefer to transform the set to "rank coordinates" r_{km} as described above for heatmaps. This was done to produce Figs. 10 and 11. Choosing the $\lambda_m = \lambda$ for all m then treats objectives equally and the precise value of λ is irrelevant.

Diagrams such as Fig. 11 have their greatest utility in interactive tools that allow the decision maker to interrogate particular solutions.[6] Understanding and choice of a particular solution is also be facilitated by colouring the solutions as in the figures, where solutions are coloured by their average rank (26). Low average rank tends to indicate those solutions which achieve relatively good objective values on all objectives, and may be promising candidates for decision makers seeking a trade-off between objectives. We have investigated colouring with a variety of alternative measures of solution quality, such as preference ordering [58] and the power index [59], [60], [61], [62] adapted for use with many-objective populations [7]. Overall we find that the average rank, which is equivalent to the outflow ordering [63], [7], is simple to compute and provides the same qualitative information as alternatives.

[6] An interactive tool is available from http://emps.exeter.ac.uk/staff/reverson

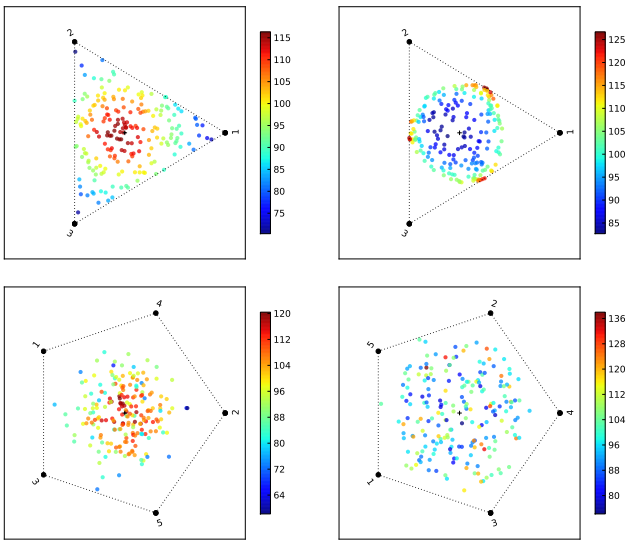


Fig. 12: Planar visualisation of convex outwards (left) and concave (right) sets with 3 (top) and 5 (bottom) objectives. Solutions are coloured by average rank.

The average rank also provides an interesting interpretation of the projection of the rank coordinates $\mathbf{r}_k$ onto the simplex. As described above (cf. Fig. 9 and equation (19)), a solution $\mathbf{y}_k$ in ranked coordinates $\mathbf{r}_k$ is projected onto the simplex to $\hat{\mathbf{r}}_k$ by contracting it by a factor $\gamma_k = \|\hat{\mathbf{r}}_k\|/\|\mathbf{r}_k\| = \mathbf{r}_k \cdot \mathbf{n}$ where $\mathbf{n}$ is the unit vector normal to the simplex. If, as suggested above for rank coordinates, the simplex is chosen with all the λ_m equal, then $\mathbf{n} = \mathbf{1}/\sqrt{M}$, where $\mathbf{1}$ is a vector of 1s. Consequently $\gamma_k = \mathbf{r}_k \cdot \mathbf{1}/\sqrt{M} \propto \bar{r}_k$, the average rank. Thus the average rank of a solution is proportional to the factor by which the solution (in rank coordinates) must be contracted in order that it lies on the simplex. Solutions with low average rank are close to the simplex, whereas solutions with high average rank are distant from the simplex.

This observation can be used to infer some information about the configuration the solutions. Solution sets which are convex outwards (such as the Pareto front for the DTLZ2 problem shown in Fig. 1(b)) have low average rank solutions towards the edge of the set, while low average rank solutions tend to be on the periphery when the set is concave. This is illustrated in Fig. 12 for 3 and 5-dimensional solutions lying on convex and concave fronts:[7] clustering of the high average rank solutions in the centre of the convex is evident, while high average rank solutions tend to be closer to the edges for the concave case.

VI. Dominance Distance for Multidimensional Scaling

The planar visualisation method presented in the previous section is an explicitly geometric construction that transforms the corners of the simplex in objective space to the vertices of a polygon. Although this method, unlike general dimension reduction methods such as PCA, GTM, Neuroscale and the SOM, takes account of the mutually non-dominating nature of the solutions, it does not take into account the relations between objectives. Here we define a new measure of the similarity of solutions, which attempts to capture the degree of dominance between solutions. This "dominance distance" is then used to embed the solutions in Euclidean space via metric multidimensional scaling (MDS) [64], [65].

Although the solution sets to multi- and many-objective problems comprise mutually non-dominating solutions, for generality we consider sets of points $\{\mathbf{y}_k\}$, some of which may dominate others. We regard two points $\mathbf{y}_k$ and $\mathbf{y}_j$ as similar if they both dominate a third point $\mathbf{y}_p$ or are both dominated by $\mathbf{y}_p$ or are both mutually non-dominating with $\mathbf{y}_p$. Refining this idea, we define the similarity of $\mathbf{y}_j$ and $\mathbf{y}_k$ relative to $\mathbf{y}_p$ as proportional to the number of objectives on which $\mathbf{y}_j$ and $\mathbf{y}_k$ have the same relation (greater than, less than, or equal) to $\mathbf{y}_p$. Thus

$$S(\mathbf{y}_k, \mathbf{y}_j; \mathbf{y}_p) = \frac{1}{M} \sum_{m=1}^{M} \Big[I((y_{pm} < y_{km}) \wedge (y_{pm} < y_{jm})) + I((y_{pm} = y_{km}) \wedge (y_{pm} = y_{jm})) + I((y_{pm} > y_{km}) \wedge (y_{pm} > y_{jm})) \Big] \quad (27)$$

where $I(p)$ is the indicator function that is 1 when the proposition p is true and 0 otherwise. The second term in (27) accounts for exact equality on an objective; although this occurrence is very rare with real-valued objectives it may arise with integers or as the result of rounding during measurement, see for example [66]. We define the distance relative to $\mathbf{y}_p$ as:

$$D(\mathbf{y}_k, \mathbf{y}_j; \mathbf{y}_p) = 1 - S(\mathbf{y}_k, \mathbf{y}_j; \mathbf{y}_p). \quad (28)$$

The *dominance distance* is obtained by averaging $D(\mathbf{y}_k, \mathbf{y}_j; \mathbf{y}_p)$ across all the elements of the set:

$$D(\mathbf{y}_k, \mathbf{y}_j) = \frac{1}{K-2} \sum_{\substack{p=1 \\ p \notin \{k,j\}}}^{K} D(\mathbf{y}_k, \mathbf{y}_j; \mathbf{y}_p). \quad (29)$$

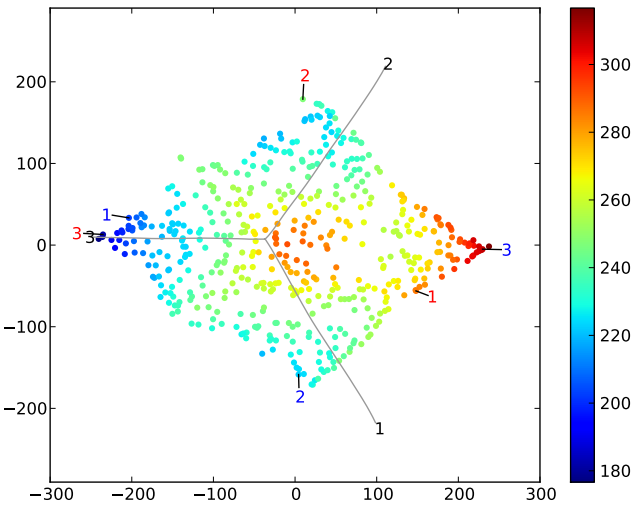
Theorem 1: The dominance distance (29) is a metric.

Proof: It is clear from (27) that $D(\mathbf{y}_k, \mathbf{y}_j) = D(\mathbf{y}_j, \mathbf{y}_k)$. Since the maximum value of the sum in (27) is 1, $0 \leq S(\mathbf{y}_k, \mathbf{y}_j; \mathbf{y}_p) \leq 1$ and therefore $D(\mathbf{y}_k, \mathbf{y}_j; \mathbf{y}_p) \geq 0$ and $D(\mathbf{y}_k, \mathbf{y}_j) \geq 0$. It is easily checked by direct substitution in (27) that $S(\mathbf{y}, \mathbf{y}; \mathbf{y}_p) = 1$ for all $\mathbf{y}_p$, so $D(\mathbf{y}, \mathbf{y}) = 0$. Conversely, $S(\mathbf{y}_j, \mathbf{y}_k; \mathbf{y}_p) = 1$ for all $\mathbf{y}_p$ only if $y_{km} = y_{jm}$ for all m. Thus $D(\mathbf{y}_k, \mathbf{y}_j) = 0$ iff $\mathbf{y}_k = \mathbf{y}_j$.

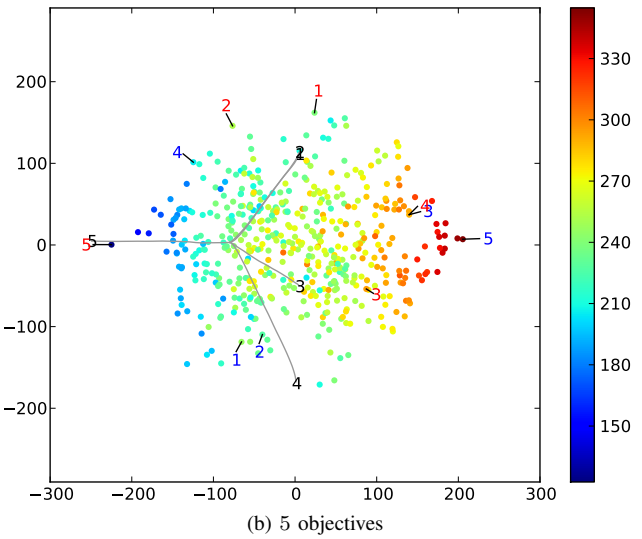
To see that $D(\cdot, \cdot)$ obeys the triangle inequality we associate with $\mathbf{y}_j$ and $\mathbf{y}_k$ strings $\mathbf{b}_j$ and $\mathbf{b}_k$ of length M on an alphabet of the symbols $\{-1, 0, +1\}$, so that a -1 in position m of the string for $\mathbf{y}_k$ indicates that $y_{km} < y_{pm}$, a $+1$ if $y_{km} > y_{pm}$, and 0 if $y_{km} = y_{pm}$. For example, with $M = 7$ objectives:

m	1	2	3	4	5	6	7
$\mathbf{b}_j$	-1	-1	$+1$	$+1$	-1	$+1$	0
$\mathbf{b}_k$	-1	$+1$	-1	$+1$	$+1$	$+1$	-1

[7] Solutions on the convex front in three dimensions lie on the positive octant of a spherical shell centred on the origin. Solutions on the concave front lie on the negative octant of a spherical shell centred on the origin and then translated by $(1, 1, 1)$. Likewise for 5 objectives.

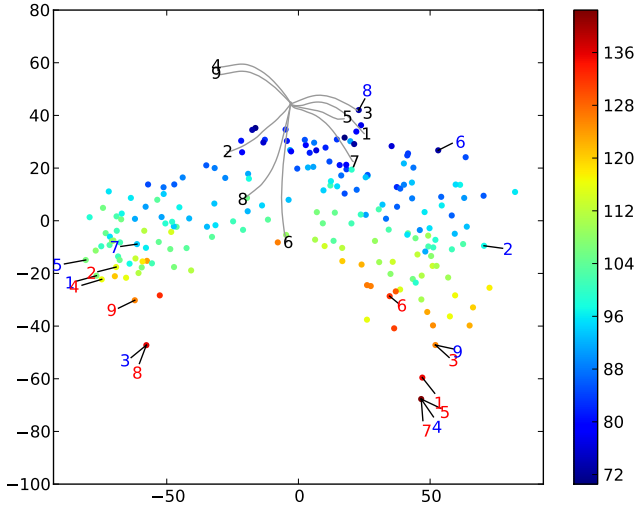


(a) 3 objectives

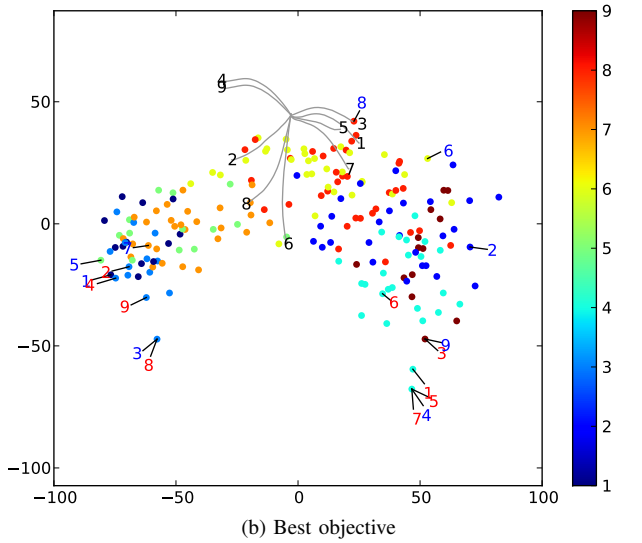


(b) 5 objectives

Fig. 13: Dominance distance MDS visualisations of the DTLZ6 test problem. Solutions are coloured by their average rank. The grey lines indicate the edges of the axis-parallel bounding box of the data which meet at the global best point. The best and worst solutions on each objective are labelled by that objective in blue and red respectively.



(a) Average rank



(b) Best objective

Fig. 14: Dominance distance MDS visualisations of the radar data. In Fig. 14a solutions are coloured by their average rank, whilst in Fig. 14b they are coloured by the objective for which they have the best rank, as were solutions in the example visualisations shown in Fig. 1. The grey curves indicate the edges of the axis-parallel bounding box of the data which meet at the global best point. The best and worst solutions on each objective are labelled by that objective in blue and red respectively.

Here $\mathbf{y}_j$ is greater than $\mathbf{y}_p$ on objectives $3, 4$ and 6, and $y_{j7} = y_{p7}$, while $\mathbf{y}_k$ is greater that $\mathbf{y}_p$ on objectives $2, 4, 5$ and 6. Then $M \times D(\mathbf{y}_k, \mathbf{y}_j; \mathbf{y}_p)$ is the Hamming distance between the strings $\mathbf{b}_j$ and $\mathbf{b}_k$, namely the number positions in which their symbols disagree. In the example $D(\mathbf{y}_j, \mathbf{y}_k; \mathbf{y}_p) = 4/7$. The Hamming distance is well known to be a metric which shows that $D(\cdot, \cdot)$ is also a metric. ■

A further characterisation of the dominance distance is provided by noting that with one objective ($M = 1$) the distance is just the sum of the difference in the ranks: $D(y_k, y_j) = |r_k - r_j|$. Consequently

$$D(\mathbf{y}_k, \mathbf{y}_j) = \frac{1}{M} \sum_{m=1}^{M} |r_{km} - r_{jm}| \tag{30}$$

so the distance between individuals is measured by the average magnitude of the difference in their ranks on each objective. Note that this is not the magnitude of the difference of their average ranks. Equation (30) provides an efficient method of calculating $D(\mathbf{y}_k, \mathbf{y}_j)$ compared with a straightforward application of (27) and (29).

Since $D(\cdot, \cdot)$ is a metric, obeying the triangle inequality, the matrix with elements $D_{kj} \equiv D(\mathbf{y}_k, \mathbf{y}_j)$ is a Euclidean distance matrix [67], [68] and there exists a set of points $\mathbf{z}_k \in \mathbb{R}^K$ separated by Euclidean distances $\|\mathbf{z}_k - \mathbf{z}_j\| = D_{kj}$.

If $\mathbf{F} = \mathbf{Z}\mathbf{Z}^T$ is a decomposition of

$$\mathbf{F} = -\frac{1}{2}\left(I - \frac{\mathbf{1}\mathbf{1}^T}{K}\right)\mathbf{D}\left(I - \frac{\mathbf{1}\mathbf{1}^T}{K}\right) \tag{31}$$

then the rows of $\mathbf{Z}$ are coordinates of the points that generate $\mathbf{D}$. Metric multi-dimensional scaling [69], [64], [65] finds a spectral decomposition of $\mathbf{F}$, which is positive semi-definite, and projects the embedding onto the principal eigenvectors of $\mathbf{F}$, thus retaining the best linear approximation (in a least squares sense) to the full embedding. Spectral decomposition of $\mathbf{F}$ has a computational complexity of $O(K^3)$, however, projections of several hundred points can easily be achieved in a second and if necessary the procedure might be made more efficient by finding only the principal few eigenvectors of $\mathbf{F}$.

Fig. 13(a) shows 500 3-objective DTLZ6 solutions projected onto the principal two eigenvectors of $\mathbf{F}$. As before solutions are coloured by their average rank. Also marked are the best and worst solutions for each objective and the edges of the axis-parallel bounding box which contains the solutions which meet at the global best point [70], namely $(\min_k(y_{k1}), \min_k(y_{k2}), \ldots, \min_k(y_{kM}))$; these edges are parallel to the coordinate vectors in the M-dimensional space. The visualisation reveals the symmetry between objectives y_1 and y_2, with y_3 a distinguished objective. Note that the worst solutions on each objective are mapped close to the ends of bounding box axes, while the best solutions are opposite these ends, indicating that the visualisation is providing a topographic representation from the dominance distance, which is itself based only on the greater than or less than relations between solutions. It is clear from the visualisation that good average rank (dark blue) corresponds to poor values of y_3 and good y_3 solutions are only obtained by having a poor average rank. Note that although the distinct clusters are not evident, the average rank shows that there are isolated patches of high and low rank. The original clusters are not visible because the dominance distance discards raw distance information, favouring instead information about the relative quality of solutions.

Fig. 13(b) shows solutions from the 5-objective version of the test problem. It is apparent that the visualisation has identified a remarkably similar structure and symmetry inherent in the solution set despite the additional two objectives. The last objective y_5 is again distinguished and there is a good correlation between objectives y_1 and y_2 (the bounding box axes are mapped almost on top of one another), and between y_3 and y_4.

The dominance distance visualisation of the degenerate WFG3 solutions (cf. Fig. 10(c)) maps the solutions to a single line along the coordinate axis with solutions arranged by average rank along it, and there is only a single non-zero eigenvalue of $\mathbf{F}$.

Fig. 14 shows the dominance distance visualisation of 200 radar data solutions. As indicated by the bands of colour, which represent bands of similar average rank, in Fig. 14(a) the method has produced a diagram which groups similar solutions together, allowing a decision maker to identify groups of solutions which have similar relations to other solutions in the set. Here the low-average rank solutions are located along the top of the "crescent" of solutions. There are bands of solutions with similar average rank running along the length of the "crescent". Those at one end of the "crescent" are related to other solutions in ways more similar to each other than those at the other end; that is, solutions at one end all tend to be greater than or less than other solutions on the same objectives. Note that the bounding box axes, which are more distant from the front than in the DTLZ6 front, are grouped into those associated with range (objectives $1, 3, 5$ and 7) and those associated with velocity (objectives $2, 4, 6$ and 8). It is interesting to note that the axis for y_9 has been placed close to y_4 and near to the y_2, y_8 and y_6 axes, which is also the result of seriating the objectives for heatmap visualisation (cf. Fig. 2(c)). The visualisation shows that the low average rank solutions are associated with good values for y_2, y_6 and y_8, while the best solutions for the other objectives, located near the horns of the crescent, have high average rank and are close to (or identical to) solutions which are very poor on other objectives.

We emphasise that this spatial arrangement in the visualisation plane reflects the similarity of order relations among solutions, rather than their spatial configuration in objective space. Nonetheless, as the colouring by best objective shows in Fig. 14(b), visualisation by dominance distance tends to group solutions that are close in objective space.

Here 54.3% of the variance in the K-dimensional embedding is retained in the 2-dimensional projection onto the plane. Projection onto the third eigenvector of $\mathbf{F}$ captures an additional 5.5%, but visualisation and interpretation of the three-dimensional representation is more cumbersome. One avenue that might be explored to further reduce the dimensionality is to use the Isomap [71] or similar nonlinear dimension reduction methods to look for non-linear manifolds in the data.

VII. Conclusion

We have presented a variety of methods for visualising the many-objective mutually non-dominating sets being produced by current multi- and many-objective evolutionary algorithms.

Heatmaps are a standard, well-understood visualisation method used in many areas. Here we have shown how their interpretability may be greatly enhanced by spectral seriation of both the objectives and the solutions in order to place similar objectives and similar solutions together. Seriation in parameter space facilitates understanding of the effect of decision variables on solutions. A straightforward multi-objective evolutionary algorithm gives small improvements over spectral seriation and suggests that in many cases the convex combinations of the objective and parameter space similarity matrices will yield orderings close to the seriations found by the multi-objective evolutionary algorithm, which will be important for using the method to monitor the progress of optimisations.

Throughout this paper we have emphasised the use of ranks rather than the raw objective values themselves. Use of rank coordinates captures the dominance relations between solutions and removes unknown relative scaling information in a natural way. It also means the visualisation produced is much less fragile to the insertion or removal of an individual because the net effect is to vary the range of the ranks by one rather than the potentially large range and scaling shifts with raw

values. Furthermore the relative order of a ranked solution is only affected at the insertion point, unless it is directly adjacent to the new solution or removed solution, its neighbours will be unchanged, and the largest rank shift any solution will experience on an objective is one. That said, we point out that if the solutions all have nearly equal values, ranking magnifies the small differences between them; in this case it will be important to ensure that the differences are significant. As a by-product ranking also performs histogram equalisation of the objectives so that the full colour scale is used. Use of ranks leads naturally to measures such as Spearman's footrule and Kendall's τ for measuring the similarity of objectives; we prefer Spearman's footrule for its computational simplicity and ease of interpretation.

Two planar visualisation methods of the solutions were presented. RadViz, exploiting interpretations of barycentric coordinates in objective space and the visualisation plane, provides an intuitive visualisation of the objective locations, and places solutions in relation to them. With many objectives the information compression that must occur to map the solutions onto the interior of a polygon is severe and can lead to potentially confusing placement of the solutions. However, the relationship between solutions and objectives is often easier to comprehend than in other point-based representations.

The introduction of the novel measure of solution similarity based on order relations with other solutions in the set permits an embedding of the solution set in Euclidean space. Standard linear and non-linear dimension reduction methods can then be used to visualise the solutions in two or three dimensions. We emphasise that the dominance distance measures how similarly two solutions relate to the rest of the set and thus provides a key to further analysis. The dominance distance can be used to cluster solutions and, in sets containing individuals that dominate other individuals, it identifies an axis describing the Pareto ranking structure [66]. Current work is exploring its use for identifying structure, such as outliers, within the solution set.

Visualising many-objective solutions in two or three dimensions inevitably necessitates a loss of information and it is likely particular methods or combinations of them will be more effective for particular problems; indeed it is very unlikely that a single method will work well for all problems. The methods presented here are designed to put a suite of tools at the many-objective investigator's disposal. Spectral seriation and the planar visualisation methods rely on linear algebra and matrix eigendecomposition. They are thus computationally cheap, especially in comparison with the computational effort required to obtain the solutions, and are sufficiently fast to be incorporated in interactive tools.

In this paper we have discussed the visualisation of static populations and important future work will be to develop methods for the effective visualisation of changing populations as an optimisation evolves. As a first step the methods presented here are generally cheap enough to be recomputed afresh each time the population changes. Also, small changes in the population produce small changes in the visualisation and both the RadViz and dominance distance methods can straightforwardly incorporate new solutions that were not used in determining the original visualisation. However, additional work is required on effective dynamic visualisation methods that link the temporal evolution of individuals.

Finally, we point out that although these visualisations may aid understanding of the global structure of the solution set, work remains to be done on characterising the local structure, for example: where are the "knees" in many-objective Pareto fronts?

REFERENCES

[1] H. Ishibuchi, N. Tsukamoto, and Y. Nojima, "Evolutionary Many-objective Optimization: A Short Review," in *Proceedings of the IEEE Congress on Evolutionary Computation*, June 2008, pp. 2419–2426.

[2] C. M. Fonseca and P. J. Fleming, "Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization," in *Proceedings of the Fifth International Conference on Genetic Algorithms*. Morgan Kauffman, 1993, pp. 416–423.

[3] A. Pryke, S. Mostaghim, and A. Nazemi, "Heatmap Visualization of Population Based Multi Objective Algorithms," in *Proceedings of the 4th International Conference on Evolutionary Multi-criterion Optimization*. Springer-Verlag, 2007, pp. 361–375.

[4] I. Jolliffe, *Principal Component Analysis*. Springer, 2002.

[5] D. Lowe and M. E. Tipping, "Neuroscale: Novel Topographic Feature Extraction Using RBF Networks," in *Proceedings of Advances in Neural Information Processing Systems 9*, ser. NIPS'96, 1996, pp. 543–549.

[6] J. E. Fieldsend and R. M. Everson, "Visualisation of Multi-class ROC Surfaces," in *Proceedings of the ICML 2005 Workshop on ROC Analysis in Machine Learning*, 2005, pp. 49–56.

[7] D. J. Walker, R. M. Everson, and J. E. Fieldsend, "Visualisation and Ordering of Many-objective Populations," in *Proceedings of the IEEE Congress on Evolutionary Computation*, July 2010, pp. 3664–3671.

[8] T. Kohonen, *Self-organising Maps*. Springer, 1995.

[9] P. Hoffman, G. Grinstein, K. Marx, I. Grosse, and E. Stanley, "DNA Visual and Analytic Data Mining," in *VIS'97: Proceedings of the 9th Conference on Visualization*. Los Alamitos, CA, USA: IEEE Computer Society Press, 1997, pp. 437–441.

[10] E. J. Hughes, "Radar waveform optimisation as a many-objective application benchmark," in *Proceedings of the 4th International Conference on Evolutionary Multi-criterion Optimization*. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 700–714.

[11] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable Multi-Objective Optimization Test Problems," in *Proceedings of the IEEE Congress on Evolutionary Computation*, vol. 1, May 2002, pp. 825–830.

[12] S. Huband, L. Barone, L. While, and P. Hingston, "A Scalable Multi-objective Test Problem Toolkit," in *Proceedings of the 3rd International Conference on Evolutionary Multi-criterion Optimization*, C. A. Coello Coello, A. Hernández Aguirre, and E. Zitzler, Eds. Guanajuato, México: Springer, March 2005, pp. 280–295.

[13] M. D'Ocagane, *Coordonnées parallles et axiales: Méthode de transformation géométrique et procédé nouveau de calcul graphique déduits de la considération des coordonnées parallèlles*. Gauthier-Villars, reprinted by Kessinger Publishing, 1885.

[14] A. Inselberg, "N-dimensional Coordinates," in *Picture Data Description & Management, Transactions on Pattern Analysis and Machine Intelligence*, 1980, p. 136.

[15] K. Deb, *Multi-Objective Optimization using Evolutionary Algorithms*, ser. Wiley-Interscience Series in Systems and Optimization. John Wiley & Sons, Chichester, 2001.

[16] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, "Cluster Analysis and Display of Genome-wide Expression Patterns," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 25, pp. 14 863–14 868, December 1998.

[17] A. Nazemi, A. H. Chan, and X. Yao, "Selecting Representative Parameters of Rainfall-runoff Models Using Multi-objective Calibration Results and a Fuzzy Clustering Algorithm," in *British Hydrological Society 10th National Hydrology Symposium, Exeter*, 2008.

[18] J. Hettenhausen, A. Lewis, and S. Mostaghim, "Interactive Multi-objective Particle Swarm Optimization with Heatmap-visualization-based User Interface," *Engineering Optimization*, vol. 42, pp. 119–139, 2010.

[19] U. Biswas, U. Maulik, A. Mukhopadhyay, and M. Naskar, "Multiobjective Evolutionary Approach to Cost-effective Traffic Grooming in Unidirectional SONET/WDM Rings," *Photonic Network Communications*, vol. 18, pp. 105–115, 2009.

[20] E. Kiesling, J. Gettinger, C. Stummer, and R. Vetschera, "An Experimental Comparison of Two Interactive Visualization Methods for Multicriteria Portfolio Selection," in *Portfolio Decision Analysis*, A. Salo, J. Keisler, and A. Morton, Eds. Springer New York, 2011, vol. 162, pp. 187–209.

[21] C. Bishop, M. Svensén, and C. Williams, "GTM: The Generative Topographic Mapping," *Neural Computation*, pp. 215 – 235, 1998.

[22] S. Obayashi, "Pareto Solutions of Multipoint Design of Supersonic Wings using Evolutionary Algorithms," in *Adaptive Computing in Design and Manufacture V*. Springer-Verlag, 2002, pp. 3–15.

[23] D. Lowe and M. E. Tipping, "Feed-Forward Neural Networks and Topographic Mappings for Exploratory Data Analysis," *Neural Computing and Applications*, vol. 4, no. 2, pp. 83–95, 1996.

[24] R. Everson and J. Fieldsend, "Multi-class ROC Analysis from a Multi-objective Optimisation Perspective," *Pattern Recognition Letters*, vol. 27, pp. 531–556, 2006.

[25] I. Nabney, *NETLAB: Algorithms for Pattern Recognition*. Oxford University Press, 2004.

[26] J. J. Valdés and A. J. Barton, "Visualizing High Dimensional Objective Spaces for Multi-objective Optimization: A Virtual Reality Approach," in *Proceedings of the IEEE Congress on Evolutionary Computation*, 2007, pp. 4199–4206.

[27] T. Yoshikawa, D. Yamashiro, and T. Furuhashi, "A Proposal of Visualization of Multi-objective Pareto Solutions - Development of Mining Technique for Solutions," in *Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Multicriteria Decision Making (MCDM 2007)*, 2007.

[28] T. Tušar and B. Filipič, "Visualizing 4D Approximation Sets of Multiobjective Optimizers with Prosections," in *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, ser. GECCO '11. ACM, 2011, pp. 737–744.

[29] M. Köppen and K. Yoshida, "Visualization of Pareto-sets in evolutionary multi-objective optimization," in *Proceedings of the 7th International Conference on Hybrid Intelligent Systems*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 156–161.

[30] W. S. Robinson, "A Method for Chronologically Ordering Archaeological Deposits," *American Antiquity*, vol. 16, no. 4, pp. 293–301, April 1951.

[31] E. Forsyth and L. Katz, "A Matrix Approach to the Analysis of Sociometric Data: Preliminary Report," *Sociometry*, vol. 9, no. 4, pp. 340–347, 1946.

[32] J. E. Atkins, E. G. Boman, and B. Hendrikson, "A Spectral Algorithm for Seriation and the Consecutive Ones Problem," *SIAM Journal on Computing*, vol. 28, no. 1, pp. 297–310, 1998.

[33] I. Liiv, "Seriation and Matrix Reordering Methods: An Historical Overview," *Stat. Anal. Data Min.*, vol. 3, pp. 70–91, April 2010.

[34] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Prentice Hall, 2007.

[35] M. Fiedler, "Algebraic Connectivity of Graphs," *Czechoslovak Mathematical Journal*, vol. 23, no. 98, pp. 298–305, 1973.

[36] F. R. K. Chung, *Spectral Graph Theory*. American Mathematical Society, 1997.

[37] A. Kaveh and H. Rahimi Bondarabady, "Finite Element Mesh Decomposition Using Complementary Laplacian Matrix," *Commun. Numer. Meth. Engng*, vol. 16, no. 379-389, 2000.

[38] E. J. Hughes, "Evolutionary Many-objective Optimisation: Many Once or One Many?" in *Proceedings of the IEEE Congress on Evolutionary Computation*, September 2005, pp. 222–227.

[39] ——, "Best Known Non-dominated Results of Radar Waveform Optimisation," code.evanhughes.org, 2007.

[40] C. Spearman, "The Proof and Measurement of Association Between Two Things," *American Journal of Psychology*, vol. 15, pp. 72–101, 1904.

[41] ——, "'Footrule' for Measuring Correlation," *British Journal of Psychology*, vol. 2, pp. 89–108, 1906.

[42] P. Diaconis and R. L. Graham, "Spearman's Footrule as a Measure of Disarray," *Royal Statistical Society Series B*, vol. 32, no. 24, pp. 262–268, 1977.

[43] M. G. Kendall, "A New Measure of Rank Correlation," *Biometrika*, vol. 30, no. 1/2, pp. pp. 81–93, 1938.

[44] L. M. Adler, "A modification of Kendall's Tau for the case of arbitrary ties in both rankings," *Journal of the American Statistical Association*, vol. 52, no. 277, pp. 33–35, 1957.

[45] F. Kudo and T. Yoshikawa, "Knowledge Extraction in Multi-objective Optimization Problem based on Visualization of Pareto Solutions," in *Proceedings of the IEEE Congress on Evolutionary Computation*, June 2012, pp. 860–865.

[46] J. D. Knowles and D. W. Corne, "Approximating the Nondominated Front Using the Pareto Archived Evolution Strategy," *Evolutionary Computation*, vol. 8, no. 2, pp. 149–172, 2000.

[47] K. Deb, S. Agrawal, A. Pratab, and T. Meyarivan, "A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II," Indian Institute of Technology, Kanpur, India, KanGAL report 200001, 2000.

[48] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computation*. Springer, 2003.

[49] P. Hoffman, "Table Visualisation: a Formal Model and its Applications," Ph.D. dissertation, University of Massachuesetts Lowell, 1999.

[50] J. Sharko, G. Grinstein, and K. Marx, "Vectorized Radviz and its Application to Multiple Cluster Datasets," *IEEE Transactions on Visualisation and Computer Graphics*, vol. 14, pp. 1444–1451, 2008.

[51] E. Kandogan, "Star Coordinates: A Multidimensional Visualization Technique with Uniform Treatment of Dimensions," in *Proceedings of the 7th ACM International Conference on Knowledge Discovery and Data Mining*, 2000.

[52] K. Daniels, G. Grinstein, A. Russell, and M. Glidden, "Properties of Normalized Radial Visualisations," *Information Visualisation*, 2012.

[53] T. Ju, P. Liepa, and J. Warren, "A General Geometric Construction of Coordinates in a Convex Simplicial Polytope," *Computer Aided Geometric Design*, vol. 24, no. 3, pp. 161–178, 2007.

[54] J. Warren, S. Schaefer, A. N. Hirani, and M. Desbrun, "Barycentric Coordinates for Convex Sets," *Advances in Computational Mathematics*, vol. 27, no. 3, pp. 319–338, 2007.

[55] G. Leban, B. Zupan, G. Vidmar, and I. Branko, "Vizrank: Data Visualization Guided by Machine Learning," *Data Mining and Knowledge Discovery*, vol. 13, pp. 119–136, 2006.

[56] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," *Neural Comput.*, vol. 15, pp. 1373–1396, June 2003.

[57] P. Grindrod, D. Higham, and G. Kalna, "Periodic Reording," *IMA Journal of Numerical Analysis*, vol. 30, no. 1, pp. 195–207, 2010.

[58] F. di Pierro, S. T. Khu, and D. Savic, "An Investigation on Preference Ordering Ranking Scheme in Multiobjective Evolutionary Optimization," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 1, pp. 17–45, 2007.

[59] T. Wei, "The Algebraic Foundation of Ranking Theory," Ph.D. dissertation, Cambridge University, Cambridge, UK, 1952.

[60] M. Kendall, "Further Contributions to the Theory of Paired Comparisons," *Biometrics*, vol. 11, pp. 43–62, 1955.

[61] C. Berge, *The Theory of Graphs*. London: Methuen, 1962.

[62] J. P. Keener, "The Perron-Frobenius Theorem and the Ranking of Football Teams," *SIAM Rev.*, vol. 35, no. 1, pp. 80–93, 1993.

[63] R. van den Brink and R. P. Gilles, "The Outflow Ranking Method for Weighted Directed Graphs," *European Journal of Operational Research*, vol. 193, pp. 484–491, 2009.

[64] J. W. Sammon, "A Nonlinear Mapping for Data Structure Analysis," *IEEE Transactions on Computers*, vol. 18, no. 5, pp. 401–409, 1969.

[65] A. R. Webb, *Statistical Pattern Recognition, 2nd Edition*. John Wiley & Sons, October 2002.

[66] R. M. Everson, D. J. Walker, and J. E. Fieldsend, "League Tables: Construction and Visualisation from Multiple Key Performance Indicators," The University of Exeter, Tech. Rep., 2012.

[67] I. Schoenberg, "Remarks to Maurice Frechet article Sur la définition axiomatique d'une classe d'espaces distanciés vectoriellement applicable sur l'espace de Hilbert," *Annals of Mathematics*, vol. 38, pp. 724–732, 1935.

[68] J. Gower, "Euclidean Distance Geometry," *Math. Sci.*, vol. 1, pp. 1–14, 1985.

[69] J. Kruskal, "Multidimensional Scaling by Optimizing Goodness-of-Fit to a Nonmetric Hypothesis," *Psychometrica*, vol. 29, pp. 1–27, 1964.

[70] M. Garza-Fabre, G. Toscano-Pulido, and C. Coello, "Two Novel Approaches for Many-objective Optimization," in *Proceedings of the IEEE Congress on Evolutionary Computation*, July 2010, pp. 4480–4487.

[71] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

David Walker (S'10) graduated with a first class honours degree in Computer Science from the University of Exeter in 2007. Since late 2008 he has been studying for a PhD in Computer Science at Exeter. His research involves the visualisation and understanding of many-objective populations. Other current research interests are novel applications of evolutionary algorithms and the identification of preference information in data.

Richard Everson graduated with a degree in Natural Sciences (Physics and Theoretical Physics) from Cambridge University in 1983 and a PhD in Applied Mathematics from Leeds University in 1988. He worked at Brown and Yale Universities on fluid mechanics and data analysis problems until moving to Rockefeller University, New York to work on optical imaging and modelling of the visual cortex. After working at Imperial College, London, he joined the Computer Science department at Exeter University where he is now an Associate Professor of Machine Learning. Current research interests are in statistical pattern recognition, multi-objective optimisation and the links between them.

Jonathan Fieldsend (S'00-M'02) received the B.A. degree (with honours) in Economics from the University of Durham in 1998, the M.Sc. degree in Computational Intelligence from the University of Plymouth in 1999 and the Ph.D. degree in Computer Science from the University of Exeter in 2003. Between 2002-2006 Dr Fieldsend was employed as a Research Fellow in the University of Exeter. He is now a Lecturer in Computer Science at the same institution. His current research interests include multi-objective optimisation, optimisation with uncertainty, pattern recognition, machine learning and data visualisation. Dr Fieldsend is a Member of the IEEE Computational Intelligence Society, a Fellow of the Higher Education Academy, and sits on the South West Branch of the BCS.