Analysis of Objectives Relationships in Multiobjective Problems Using Trade-Off Region Maps

Rodrigo L. Pinheiro  
ASAP Research Group  
School of Computer Science  
The University of Nottingham  
rzp@cs.nott.ac.uk

Dario Landa-Silva  
ASAP Research Group  
School of Computer Science  
The University of Nottingham  
jds@cs.nott.ac.uk

Jason Atkin  
ASAP Research Group  
School of Computer Science  
The University of Nottingham  
jaa@cs.nott.ac.uk

ABSTRACT
Understanding the relationships between objectives in many-objective optimisation problems is desirable in order to develop more effective algorithms. We propose a technique for the analysis and visualisation of complex relationships between many (three or more) objectives. This technique looks at conflicting, harmonious and independent objectives relationships from different perspectives. To do that, it uses correlation, trade-off regions maps and scatter-plots in a four step approach. We apply the proposed technique to a set of instances of the well-known multiobjective multidimensional knapsack problem. The experimental results show that with the proposed technique we can identify local and complex relationships between objectives, trade-offs not derived from pairwise relationships, gaps in the fitness landscape, and regions of interest. Such information can be used to tailor the development of algorithms.

Categories and Subject Descriptors
G.1.6 [Optimisation]; I.2.8 [Problem Solving, Control Methods, and Search]

Keywords
multiobjective fitness landscape analysis; trade-off region maps; fitness landscape visualisation

1. INTRODUCTION
It is important to understand the relationships between objectives in multiobjective optimisation problems (MOPs) because this can help to tailor the search according to the multiobjective fitness landscape. This is particularly when tackling large real-world MOPs with many objectives (more than two). In the multiobjective optimisation literature, the focus is often on MOPs that exhibit strong conflict relationships between objectives as this is part of the motivation for applying multiobjective techniques. However, the conflict relationship between objectives could be local rather than global. A global conflict relationship would hold throughout most, if not all of, the search space. On the other hand, a local conflict relationship would hold in a restricted region of the search space, e.g. the objectives could be in conflict early in the search but not conflicting later. This was discussed by Knowles and Corne [1] in the context of the multiobjective quadratic assignment problem. Castro-Gutierrez et al. [2] studied the objectives’ relationships in multiobjective vehicle routing problems.

As the number of objectives increases, composite relationships between objectives might emerge, i.e. relationships between objectives (conflicting or otherwise) that are not global but localised and more complex. Several techniques have been previously applied for the analysis and visualisation of relationships between objectives in MOPs. These include parallel coordinates, scatter-plots (which both involve graphical representations), Kendall correlation [3] (a quantitative metric), and statistical measures [4], among others. Purshouse and Fleming [5] discussed these techniques in their research into the relationships between objectives in MOPs. Other works that have used some of these techniques include Castro-Gutierrez et al. [2] on multiobjective vehicle routing problems, and Ishibuchi et al. [6] on many-objective problems with correlated objectives. One limitation of these techniques is that they are most suited to identify only pairwise relationships between objectives.

In this work, we propose a technique to analyse and visualise global and local relationships between many objectives in order to achieve a clearer understanding of the fitness landscape in a MOP. The technique requires a set of non-dominated solutions to be supplied (which can be obtained in any way) and uses Karnaugh maps [7] to visualise composite relationships between many objectives. The technique also uses correlation and scatter-plots to complement the analysis. The technique involves four steps: analyse global pairwise relationships between objectives in the given non-dominated set, estimate the range of values for each objective, identify objective trade-offs using Karnaugh region maps, and identify local relationships between objectives using scatter-plots. This technique seeks to analyse the objectives’ relationships from multiple perspectives in order to better understand the fitness landscape in many-objective combinatorial optimisation problems, which are well known for having irregular and difficult to assess fitness landscapes.

Section 2 surveys some related work. Section 3 provides the motivation for this work. Section 4 describes the proposed technique while experimental results applying the proposed technique to several instances of the multiobjective
multidimensional 0-1 knapsack problem \[8, 9\] are presented in Section 5. Our main observation is that different instances of the same problem may exhibit very different relationships between objectives. Finally, Section 6 concludes the paper.

2. RELATED WORK

Better understanding of fitness landscapes has been beneficial in multiobjective combinatorial optimisation (MOCO) problems. For example, Garrett and Dasgupta \[10, 11\] adapted single-objective landscape analysis techniques (distribution of optima, fitness distance correlation, ruggedness, random walk analysis and analysis of the geometry of the solution space) for tailoring multiobjective evolutionary algorithms. They applied such techniques to quadratic assignment and generalised assignment problems with two objectives. They concluded that the performance of hybrid algorithms benefits from using knowledge of the fitness landscape. Castro-Gutierrez et al. \[2\] used the objectives’ pairwise dependency correlation analysis proposed by Purshouse and Fleming \[5\] to assess the conflicting nature of objectives in multiobjective vehicle routing problems.

Brownlee and Wright \[12\] proposed a visualisation technique to evaluate the quality of a non-dominated set based on a ranking of the objectives. However, this technique may not be suitable for large solution sets as it is based on the individual analysis of solutions. Other visualisation techniques include objective wheels, bar graphs and colour stacks as explored by Anderson and Dror \[13\].

Verel et al. \[14\] adapted single-objective landscape analysis techniques to set-based multiobjective problems with objective correlation. Later, Verel et al. \[15\] conducted a study on the landscape of local optima in such problems. Verel et al. \[16\] proposed to carry out a priori analysis of a problem by evaluating the problem size, its epistasis, the number of objectives and the correlation values between objectives, to suggest the best way to tackle it. They concluded that, depending on the problem features, different types of algorithms (scalar or Pareto approach) and sizes of the solution archive should be employed.

Walker et al. \[17\] reviewed different methods (scatter plots, parallel coordinates and heat maps) to visualise solution sets for many-objective problems. They also proposed two techniques: a data mining visualisation tool to plot a convex graph, and a new similarity measure of solutions to plot them in a two-dimensional space.

Giagkiozis and Fleming \[18\] proposed a technique to estimate the Pareto front of a continuous optimisation problem, and then use the estimated front to obtain values for the decision variables of interesting solutions. They proposed using a multiobjective algorithm to obtain an initial solution set, which is then used to calculate a projection matrix of the optimal Pareto set. They tested their technique on convex benchmark problems.

Tusar and Filipic \[19\] presented a comprehensive survey and assessment on several visualisation techniques for many-objectives approximation sets. They also presented a visualisation method that uses orthogonal projections of a section and applied it to four-dimensional approximation sets.

It is clear that understanding the fitness landscape of multiobjective optimisation can help to develop better solution methods. It is also clear that the analysis and visualisation of objectives’ relationships, particularly in combinatorial landscapes with many objectives, is a topic of interest for researchers. The technique proposed in this paper seeks to make a contribution in this area.

3. OBJECTIVES RELATIONSHIPS IN MULTIOBJECTIVE OPTIMISATION

The focus of this research is to investigate the relationship between objectives in MOPs by analysing the non-dominated approximation set and its coverage of the solution space. We use the concepts of conflict, harmony and independence between objectives as proposed by Purshouse and Fleming \[5\].

Results from some existing techniques to assess the conflicting nature of objectives can be deceiving. The literature includes studies of pairwise relationships between objectives \[2, 6, 20\]. However, analysis techniques such as Kendall correlation \[3\] only manage to identify global relationships between objectives. Figure 1 shows a Pareto-front between two maximisation objectives, \(Z_1\) and \(Z_2\), in a scenario with three objectives (we omit the scatter-plots for \(Z_3\)). Clearly, when \(Z_1 < 0.5\), the objectives are conflicting while when \(Z_1 > 0.5\) the objectives are harmonious. However, if the number of solutions with \(Z_1\) lower than 0.5 is roughly the same as the number of solutions with \(Z_1\) higher than 0.5, simply applying the Kendall correlation technique would result in a correlation value close to 0. The conclusion could be drawn that the objectives are independent, when in reality there may be local relationships that could be exploited.

Some existing techniques might not reveal trade-offs which result from composite relationships between two and more objectives. Figure 2a lists a set of non-dominated fitness vectors for three maximisation objectives and Figure 2b shows their scatter-plot matrix and correlation values. We can observe that high values (31-100) appear simultaneously only in up to two of the three objectives. Two points in Figure 2a have only one good objective value. Four points are good for only \(Z_1\) and \(Z_2\). Three points are good for only \(Z_1\) and \(Z_3\), and the remaining points are good for only \(Z_2\) and \(Z_3\). No solution present has values higher than 50 for all three objectives simultaneously. The scatter plot and correlation values do not help us to appreciate the three-way trade-off. Likewise, we can see that the correlation values do not indicate any strong pairwise correlation.

Hence, to better analyse and visualise the multiobjective nature of optimisation problems, we need techniques that help us to identify global, local and composite relationships.
Step 1 – Global Pairwise Relationship Analysis:

First, the Kendall correlation values [3] are calculated, as in [5], to identify global pairwise relationships. Strongly conflicting correlations (values < −0.5) immediately indicate that a trade-off surface exists, whilst strongly harmonious correlations (values > 0.5) indicate that objectives could be aggregated or clustered. Correlation values showing that objectives are independent indicate that the data is not globally dependent, but do not imply the absence of local trade-offs in the fitness landscape. If independence is detected, the problem could be decomposed by separating the decision variables according to the objectives, to solve each objective (or groups of objectives) separately as such an approach provides improved performance [23].

Step 2 – Objective Range Analysis:

Here, the range (difference between best and worst values) is calculated for each objective in the given approximation set. Then, using problem domain knowledge the objectives which are interesting for further exploration are identified. A meaningful objective is an objective with a range which is large enough so that solutions can be classified into different quality categories regarding that objective value (e.g. good to bad, high to low, etc). Analogously, a non-meaningful objective is an objective with a range so small that the variability in the solution quality regarding that objective is considered negligible, thus not worth exploring further.

One way to deal with non-meaningful objectives is to ignore them during the optimisation. It is possible that by optimising the other objectives, the non-meaningful ones will present acceptable values within their small range anyway. Having a small range on the given approximation set does not mean that the range will be small across the entire solution space. The non-meaningful objective could be removed, the MOA re-executed and the ranged recalculated for the excluded objective. If the new solution set still exhibits only a small range for the excluded objective, it can be safely ignored. Another way to deal with non-meaningful objectives is to combine or cluster them [20].

Step 3 – Trade-off Regions Analysis:

We apply a quantitative method, namely Karnaugh maps [7], to classify the solution space into regions to help with the identification of trade-offs and the complex relationships between objectives. A Karnaugh map is a method for simplifying boolean algebra expressions using a truth table. The map has $2^n$ cells where $i$ is the number of variables. The cells are labelled with binary numbers following the Grey code, meaning that any two adjacent cells differ in one bit. Hence, in a three variable scenario, the cells adjacent to cell 0 (0002) are cells 1 (0012), 2 (0102) and 4 (1002). Karnaugh maps make it easy to visualise patterns that are used to group boolean variables.

In this step, for each objective $Z_i$ where ($i = 1, 2, \ldots, m$), we define a threshold $t_i$ such that values above $t_i$ (maximisation problem) are considered good or acceptable, and values below $t_i$ are considered inadequate. These $t_i$ can be set using some knowledge of the problem domain or empirically – for example, using the average value for each objective as its threshold. Next, we classify each objective value in each
solution as good (↑) or bad (↓). Finally, we draw a region map similarly to a Karnaugh map, but showing the count of solutions in each region rather than 0-1 output variables, and using solutions classifications (labelled ↑ or ↓) instead of the input variable values. The map is built with $2^m$ regions such that each region represents a single combination of good and/or bad objectives. We number the regions $Z_k$ using a binary encoding such that the least significant digit represents $Z_1$ and the most significant digit is $Z_m$, $\uparrow = 0$ and $\downarrow = 1$. For instance, the region $Z_4^k$, $Z_2^k$ and $Z_1^k$, would be region $r_5$, since binary 1011 is 5.

![Region Map Schematic](image)

**Figure 3: Region map schematics.**

Figure 3 presents the region map schematics for 3, 4 and 5 objectives. Each region is identified with the $r_k$. Regions with the same number of good solutions are highlighted with the same shade of grey in such a way that lighter tones represent a higher number of good solutions while darker tones represent fewer good solutions.

The main advantage of the region map is that we can easily identify which objectives simultaneously present good values and the existence of trade-offs. If the region $r_0$ is not empty, then we have solutions with acceptable values in all objectives, meaning that the problem could potentially be tackled with single-objective algorithms. A range analysis on the solutions in this region could provide additional information on which approach is appropriate. On the contrary, when most solutions fall into region $r_{2^{m-1}}$, it means that the thresholds may have been set too high and should be lowered for more accurate results. When there are no solutions in $r_0$, but there are solutions scattered across the regions, there are trade-offs and the map can be used to visualise them.

**Step 4 – Multiobjective Scatter-plot Analysis:**

The last step is an analysis using scatter plots. First, for each instance we normalise the values of all objectives. Then, we select an objective and draw a scatter plot of all remaining objectives against the selected objective. Finally, we can combine all scatter plots into a single one. By visually inspecting this combined graph we can identify local relationships (conflicts and harmony), interesting patterns, gaps in the solution space, and well-spread trade-offs or isolated regions. This information can help us to tailor a solution algorithm by directing the search towards the regions of interest. When the landscape of the solution space is consistent throughout all instances analysed in this way, we could have a clearer idea of what type of solutions to expect when solving unseen instances.

When picking an objective for this process, it is preferable to select one that has a wider range of values rather than being concentrated in only a small range, otherwise the resulting graph may be more difficult to read. It may be interesting to test different objectives in order to spot which provide more useful information, or, if multiple objectives provide different insights, all of them could be considered instead of only one.

The next section presents experimental results from applying the proposed analysis and visualisation technique to five sets of benchmark instances of the multiobjective multidimensional knapsack problem.

## 5. SAMPLE ANALYSIS

In order to illustrate the analysis technique we apply it to different scenarios of the multiobjective multidimensional knapsack problem (MOMKP) [8]. We aim to show that within the same problem, the proposed technique can identify multiple scenarios with distinct multiobjective natures.

In the MOMKP, we have a set of items $\{i = 1, \ldots, n\}$ with $m$ weights $w_i^j$ ($j = 1, \ldots, m$) and profits $c_i^k$ ($k = 1, \ldots, p$). A set of items must be selected to maximise the $p$ profits while not exceeding the capacities $W_j$ of the knapsack. This problem can be formulated as follows:

$$\begin{align*}
\text{maximise} & \quad \sum_{i=1}^{n} c_i^k x_i & \quad k = 1, \ldots, p \\
\text{subject to} & \quad \sum_{i=1}^{n} w_i^j x_i \leq W_j & \quad j = 1, \ldots, m \\
& \quad x_i \in \{0, 1\} & \quad i = 1, \ldots, n
\end{align*}$$

We considered five MOMKP datasets, each with five instances, all with $m = 4$, $p = 4$, $n = 1000$ and $W_j = 50000$. The first four datasets were generated following the guidelines in Bazgan et al. [9] and are as follows:

- **Set A**: Independent random instances where $w_i^k \in [1, 1000]$ and $c_i^k \in [1, 1000]$.
- **Set B**: Uncorrelated harmonious instances where $w_i^k \in [1, 1000]$, $c_i^k \in [1, 1000]$ and $c_i^k \in [\max\{c_i^k - 100, 1\}, \min\{c_i^k + 100, 1000\}]$ for $k = (2, 3, 4)$.
- **Set C**: Uncorrelated conflicting instances where $w_i^k \in [1, 1000]$, $c_i^k \in [1, 1000]$ and $c_i^k \in [\max\{900 - c_i^k, 1\}, \min\{1100 - c_i^k, 1000\}]$ for $k = (2, 3, 4)$.
- **Set D**: Correlated conflicting instances where $w_i^k \in [\max\{900 - |c_i^k - c_i^k|, 1\}, \min\{1100 - |c_i^k - c_i^k|, 1000\}]$, $w_i^k \in [\max\{900 - c_i^k - c_i^k, 1\}, \min\{1100 - c_i^k - c_i^k, 1000\}]$, $c_i^k \in [1, 1000]$ and $c_i^k \in [\max\{900 - c_i^k - c_i^k, 1\}, \min\{1100 - c_i^k - c_i^k, 1000\}]$ for $k = (2, 3, 4)$. 


Set A contains only independent objectives. In set B all objectives are harmonious. Set C contains three pairs of conflicting objectives, \((Z_2, Z_1), (Z_3, Z_2)\) and \((Z_3, Z_3)\), while the weights are uncorrelated. Set D has conflicting objectives, as set C, but the weights are correlated to the objective values. The fifth set \(X\) was generated using data from a real-world home-healthcare scheduling problem.

For each instance we run a single-objective genetic algorithm on each objective alone, then both NSGAII [24] and MOEA/D [25] algorithms on each pair and triple of objectives. We then combined all the obtained non-dominated solutions into an archive. We randomly drew from the archive in the MOEA/D and three runs for the NSGA-II with all objectives. We then combined all the obtained non-dominated solutions into an archive. We randomly drew from the archive in the MOEA/D and three runs for the NSGA-II with all objectives.

<table>
<thead>
<tr>
<th>Set</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Z_1, Z_2)</td>
<td>-0.15</td>
<td>-0.33</td>
<td>-0.21</td>
<td>-0.39</td>
<td>-0.49</td>
<td>0.314</td>
</tr>
<tr>
<td>(Z_1, Z_3)</td>
<td>-0.02</td>
<td>0.09</td>
<td>-0.43</td>
<td>0.34</td>
<td>-0.11</td>
<td>-0.118</td>
</tr>
<tr>
<td>(Z_1, Z_4)</td>
<td>-0.21</td>
<td>0.09</td>
<td>0.18</td>
<td>-0.20</td>
<td>0.07</td>
<td>0.006</td>
</tr>
<tr>
<td>(Z_2, Z_3)</td>
<td>-0.30</td>
<td>-0.37</td>
<td>0.17</td>
<td>0.01</td>
<td>-0.42</td>
<td>-0.182</td>
</tr>
<tr>
<td>(Z_2, Z_4)</td>
<td>0.06</td>
<td>-0.22</td>
<td>-0.17</td>
<td>-0.41</td>
<td>-0.09</td>
<td>-0.166</td>
</tr>
<tr>
<td>(Z_3, Z_4)</td>
<td>-0.25</td>
<td>-0.20</td>
<td>-0.22</td>
<td>-0.26</td>
<td>0.19</td>
<td>-0.148</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Set</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Z_1, Z_2)</td>
<td>-0.96</td>
<td>-0.97</td>
<td>-0.98</td>
<td>-0.98</td>
<td>-0.98</td>
<td>-0.974</td>
</tr>
<tr>
<td>(Z_1, Z_3)</td>
<td>0.93</td>
<td>0.95</td>
<td>0.95</td>
<td>0.96</td>
<td>0.96</td>
<td>0.950</td>
</tr>
<tr>
<td>(Z_1, Z_4)</td>
<td>-0.93</td>
<td>-0.92</td>
<td>-0.92</td>
<td>-0.97</td>
<td>-0.95</td>
<td>0.938</td>
</tr>
<tr>
<td>(Z_2, Z_3)</td>
<td>-0.96</td>
<td>-0.96</td>
<td>-0.97</td>
<td>-0.97</td>
<td>-0.96</td>
<td>0.964</td>
</tr>
<tr>
<td>(Z_2, Z_4)</td>
<td>0.95</td>
<td>0.92</td>
<td>0.93</td>
<td>0.96</td>
<td>0.94</td>
<td>0.940</td>
</tr>
<tr>
<td>(Z_3, Z_4)</td>
<td>-0.98</td>
<td>-0.94</td>
<td>-0.95</td>
<td>-0.97</td>
<td>-0.97</td>
<td>0.962</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Set</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Z_1, Z_2)</td>
<td>-0.92</td>
<td>-0.92</td>
<td>-0.94</td>
<td>-0.94</td>
<td>-0.94</td>
<td>-0.932</td>
</tr>
<tr>
<td>(Z_1, Z_3)</td>
<td>0.88</td>
<td>0.84</td>
<td>0.88</td>
<td>0.87</td>
<td>0.87</td>
<td>0.868</td>
</tr>
<tr>
<td>(Z_1, Z_4)</td>
<td>-0.84</td>
<td>-0.84</td>
<td>-0.87</td>
<td>-0.87</td>
<td>-0.87</td>
<td>-0.858</td>
</tr>
<tr>
<td>(Z_2, Z_3)</td>
<td>-0.87</td>
<td>0.90</td>
<td>-0.91</td>
<td>-0.90</td>
<td>-0.90</td>
<td>0.896</td>
</tr>
<tr>
<td>(Z_2, Z_4)</td>
<td>0.82</td>
<td>0.86</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.858</td>
</tr>
<tr>
<td>(Z_3, Z_4)</td>
<td>-0.92</td>
<td>-0.93</td>
<td>-0.93</td>
<td>-0.92</td>
<td>-0.92</td>
<td>0.924</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Set</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Z_1, Z_2)</td>
<td>-0.26</td>
<td>-0.29</td>
<td>-0.13</td>
<td>-0.24</td>
<td>-0.32</td>
<td>-0.246</td>
</tr>
<tr>
<td>(Z_1, Z_3)</td>
<td>-0.13</td>
<td>-0.24</td>
<td>-0.27</td>
<td>-0.25</td>
<td>-0.27</td>
<td>0.232</td>
</tr>
<tr>
<td>(Z_1, Z_4)</td>
<td>0.29</td>
<td>0.12</td>
<td>-0.29</td>
<td>-0.21</td>
<td>0.07</td>
<td>-0.196</td>
</tr>
<tr>
<td>(Z_2, Z_3)</td>
<td>0.25</td>
<td>-0.13</td>
<td>-0.26</td>
<td>-0.23</td>
<td>-0.19</td>
<td>-0.212</td>
</tr>
<tr>
<td>(Z_2, Z_4)</td>
<td>0.13</td>
<td>-0.28</td>
<td>0.22</td>
<td>-0.11</td>
<td>-0.27</td>
<td>-0.202</td>
</tr>
<tr>
<td>(Z_3, Z_4)</td>
<td>0.24</td>
<td>-0.23</td>
<td>0.14</td>
<td>-0.21</td>
<td>-0.18</td>
<td>0.200</td>
</tr>
</tbody>
</table>

Table 1: Results for the pairwise relationship analysis (1.0 is completely harmonious, -1.0 is completely conflicting).

For each instance we run a single-objective genetic algorithm on each objective alone, then both NSGAII [24] and MOEA/D [25] algorithms on each pair and triple of objectives. We then combined all the obtained non-dominated solutions into an archive. We randomly drew from the archive in the MOEA/D and three runs for the NSGA-II with all objectives. The approximation non-dominated set was formed with all non-dominated solutions which were found in the process. Both NSGAII and MOEA/D used a population of 200 individuals, binary tournament selection and half uniform crossover [26] for 1500000 function evaluations. Overall, we obtained non-dominated sets with approximately 900 solutions for set A, 3 for set B, 550 for set C, 1500 for set D and 2500 solutions for set X.

Only a few non-dominated solutions were found for set B, the objectives there are strongly harmonious. Therefore, when maximising one of the objectives, the other objectives are also maximised. The data is therefore not enough for some of the analysis steps. However, the results are presented for completeness and illustrate that the number of solutions obtained can have a major impact on the analysis and that, it is important to have a comprehensive set, with enough well-spread solutions.

5.1 Application of the Proposed Technique

Step 1 - Global Pairwise Relationships Analysis

The results for this step are in Table 1. Note that coeffi-
for fully independent objectives, set $A$ has values close to 0. The values for sets $C$ and $D$ are also predictably close to either 1 or −1 indicating global conflicting or harmonious relationships. Set $X$ has values similar to $A$ – they do not reveal a strong global relationship between objectives as they are closer to 0 than to 1 or −1. Also, in set $X$ it is not possible to decompose the decision variables according to the objectives, as every item has all weights and values above zero.

**Step 2 - Objective Range Analysis.** The results for this step are in Table 2. Considering set $B$, it can be seen, that the set presents small ranges of less than 0.3% on average. In this dataset all objectives are harmonious and the solutions found are all located in a small region of the solution space. These few solutions dominated all other solutions explored. Both sets $C$ and $D$ present similar results to each other with large ranges for each objective (over 60%). This is expected since these are instances with conflicting objectives and present global trade-offs. The large ranges mean that while we have solutions with good values for a given objective, at least one other objective has poor value.

Finally, we highlight that while the global pairwise relationship analysis (step 1) hinted that sets $A$ and $X$ were similar, the difference between them now becomes clear with the results from step 2. In set $A$, each objective range is around 24.0% of the maximum value – the smallest ranges excluding the harmonious instances – whilst in set $X$ the ranges go up to 84.9%, the largest range found. Thus, we can see that the ranges for set $X$ are closer to those for set $D$, a conflicting scenario with global trade-offs.

### Table 3: Results for the trade-off regions analysis.

<table>
<thead>
<tr>
<th>Set</th>
<th>$z_2^*$</th>
<th>$z_3^*$</th>
<th>$z_4^*$</th>
<th>$z_5^*$</th>
<th>$z_6^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>01.2%</td>
<td>05.9%</td>
<td>07.9%</td>
<td>06.3%</td>
<td></td>
</tr>
<tr>
<td>$B$</td>
<td>04.6%</td>
<td>06.9%</td>
<td>07.2%</td>
<td>08.0%</td>
<td></td>
</tr>
<tr>
<td>$C$</td>
<td>11.2%</td>
<td>03.5%</td>
<td>08.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D$</td>
<td>02.8%</td>
<td>06.7%</td>
<td>08.7%</td>
<td>10.0%</td>
<td></td>
</tr>
<tr>
<td>$E$</td>
<td>00.0%</td>
<td>00.0%</td>
<td>05.0%</td>
<td>00.0%</td>
<td></td>
</tr>
</tbody>
</table>

**Step 3 - Trade-Off Regions Analysis.** The results for this step are in Table 3. For each instance we computed the number of solutions in each region and the map shows the average percentage for each set. We set the range threshold to 1% above the mean value found for each objective, thus considering a value slightly above the average to be good. We can observe that on set $A$ the front is well distributed as we have solutions in all regions, scattered throughout the solution space, as a result of the independent objectives. Additionally, note that we have solutions both in $r_0$ and $r_{15}$. This is due to the map presenting the combined results for all five instances in that set, and in some instances we have solutions in $r_0$ only and in other instances in $r_{15}$ only.

In sets $C$ and $D$ we clearly identify the global relationships. There are no solutions with good values in all objectives and most instances present no solution with good values in three objectives. The majority of the solutions are situated where $Z_1$ and $Z_3$ alone have good values and where $Z_1$ and $Z_4$ alone have good values, as these are the harmonious pairs. Additionally, we can observe that almost no solutions are present in conflicting areas. For instance, where $Z_1$ and $Z_2$ present good values simultaneously. Moreover, solutions in conflicting areas should be close to the chosen threshold.

The set $X$ does not contain solutions in $r_0$ and there are no solutions in regions $r_1$, $r_2$, $r_4$ and $rs$, meaning that no high values can be simultaneously found for three or more objectives. We can find good values simultaneously only for up to two objectives. The map for set $X$ resembles the ones for sets $C$ and $D$ in the sense that we can clearly see that there are several regions without solutions. Thus, we have trade-offs to present to the decision-maker. This means that the decision-maker has to choose between up to two good objective values to the detriment of the remaining objectives, since all of the regions containing solutions have at most two simultaneously good values.

**Step 4 - Multiobjective Scatter Plot Analysis.** This analysis was performed for each instance and Figure 4 presents the results for all instances of each set combined. We can see in Figure 4a that although instances in dataset $A$ are completely random, all of them show similar landscapes with a high concentration of solutions towards the (1,1) corner. Moreover, no local relationships can be identified, which is expected as the data is completely random.

On sets $C$ and $D$ we can clearly see the trade-off regions. Also, there is a noticeable gap in the solution space when $Z_1$ is in the range from 0 to 0.5 and when the remaining objectives are in the range from 0 to 0.4, approximately. Moreover, the landscape of the solution space appears to be similar for all instances of each of the sets $C$ and $D$.

Since the data was uniformly generated (these gaps are unlikely to arise from the data itself) and could represent limitations in the solution algorithms, indicating that they did not explore the entire front. It is well known that the performance of some MOEAs is limited when the number of objectives is more than three [27].

Set $X$ presents a unique scenario, as we can identify patterns and gaps in the solution space. The first feature worth...
noticing is that there is a lack of solutions with values within [0.85, 1]. Again, this is due to limitations of the solution algorithms. However, we can see that the size of the gap is small, confirming that instances with strongly conflicting objectives present a bigger challenge to the algorithms. We can also identify several local relationships. When $Z_1$ ranges from 0.5 to 0.8, the three remaining objectives simultaneously conflict and harmonise. Knowing that only two simultaneous objectives present high values (from the region map analysis), we can conclude that whenever $Z_1$ increases, only one of the other objectives simultaneously increases too.

5.2 Discussion

With the proposed analysis and visualisation technique, we can better understand the multiobjective nature of the problem instances considered. Looking at only the correlation coefficients, we could conclude that: sets $A$ and $X$ do not present interesting multiobjective traits, that set $B$ is inconclusive and that sets $C$ and $D$ present conflicting and harmonious objectives. However, by applying our analysis and visualisation technique we can reach a more comprehensive understanding of these instance sets.

As fully random instances, dataset $A$ does not present relevant global or local pairwise relationships according to the global pairwise analysis (step 1) and the multiobjective scatter plot analysis (step 4). Additionally, the objective range analysis (step 2) shows that even though there is a large set of non-dominated solutions, these are concentrated in a reasonably small area of the search space. For this dataset we can use the information from the trade-off region maps to interact with the decision-maker to identify which regions are of more interest and then use single-objective optimisation algorithms to find solutions in that region. Since we have solutions in all of the regions of the map, any objective vector could provide an adequate solution.

Set $B$ presents a completely harmonious case and by analysing the ranges and bearing in mind that the algorithms found just a handful of solutions, we can assume that a single-objective algorithm aiming to maximise any of the objectives could provide a reasonable good solution. Sets $C$ and $D$ present similar scenarios, hence the correlation between weights and coefficients does not impact on the nature of the problem. The entire solution set represents a huge trade-off. We also notice that the algorithms found it very difficult to expand along the front and that they mainly explored the region surrounding the intersection of the trade-off. Nonetheless, by perceiving that all instances in these sets have similar landscapes and by knowing the approximate boundaries of each objective (by applying single-objective algorithms to each objective alone), we can estimate the landscape of solutions for other instances in those sets. Therefore, we could direct the search to the regions of interest after presenting the expected trade-offs to the decision-maker. However, if it is imperative to use an a posteriori approach, the global pairwise analysis and the scatter plots provide sufficient information to make feasible the grouping of harmonious objectives.

Finally, set $X$ presents a quite different picture. By only evaluating the global pairwise analysis (step 1) we conclude that there is no strong pairwise relationship between objectives. However, the objectives range analysis (step 2) shows that in fact we have non-dominated solutions that vary greatly in quality. This is an indication of the existence of trade-offs (as we can see by comparing this set with sets $C$ and $D$). The trade-off region analysis (step 3) showed the existence of overall trade-offs as it is not possible to have solutions with good values in more than two objectives simultaneously. Finally, the multiobjective scatter plot analysis (step 4) identified local relationships between objectives and gaps in the solution space, pointing to the existence of local conflicts. Therefore, instances in dataset $X$ exhibit a distinctive multiobjective nature perhaps with interesting options for a decision-maker. A sound possibility to tackle this problem would be to use the region map to identify the regions of interest and then locate those regions in the scatter plot. In case a selected region contains a local conflict, we can use the algorithm proposed by [1] to reach the trade-off front and then expand through it.

6. CONCLUSION

We proposed a technique that uses correlation, trade-off region maps and scatter-plots as tools for the analysis and visualisation of objectives’ relationships in multiobjective optimisation problems. The technique consists of four steps: 1) evaluate the global correlation values, 2) compute the range of values for all objectives, 3) compute the distribution of solutions in the different trade-off regions, and 4) conduct a scatter-plot analysis of the objectives.

We applied the proposed technique to five sets of instances for the multiobjective multidimensional knapsack problem. The proposed technique helps to identify features such as local and complex relationships between objectives, trade-offs not derived from pairwise relationships, gaps in the fitness landscape and regions of interest. Different conclusions can be reached about the objectives’ relationships, for different instances even though they are scenarios from the same problem. We also discussed how the analysis and visualisa-
tion technique could be used to better understand the fitness landscape of the problem in hand.

Future work includes applying the proposed technique to other optimisation problems to validate further. It is also important to study the impact that the initial approximation set provided has on the accuracy of the analysis. Finally, we intend to investigate how the components of this technique, such as the trade-off region map, could be employed during the optimisation process for a many-objectives algorithm, to direct the search towards regions of interest.

7. REFERENCES