Multivariate Analysis to Assist Decision-Making in Many-objective Engineering Optimization Problems

Francisco Santos and Lino Costa

Centro ALGORITMI, Universidade do Minho, 4710-057 Braga, Portugal
francisco_dos_santos@outlook.pt, lac@dps.uminho.pt

Abstract. Data processing (or the transformation of data into knowledge and/or information) has become an indispensable tool for decision-making in many areas of engineering. Engineering optimization problems with many objectives are common. However, the decision-making process for these problems is complicated since there are many trade-offs that are difficult to identify. Thus, in this work, multivariate statistical methods, Principal Component Analysis (PCA) and Cluster Analysis (CA), have been studied and applied to analyze the results of many objective engineering optimization problems. PCA reduces the number of objectives to a very small number, CA through the similarities and dissimilarities, creates groups of solutions, i.e., bringing together in the same group solutions with the same characteristics and behaviors. Two engineering optimization problems with many objectives are solved: a mechanical problem consisting in the optimal design of laminated plates, with four objectives and a problem of optimization of the radar waveform, with nine objectives. For the problem of the design of laminated plates through PCA allowed to reduce to two objectives and through CA it was possible to create three distinct groups of solutions. For the problem of optimization of the radar waveform, it was possible to reduce the objectives from nine to two objectives representing the greatest variability of the data, and CA defined three distinct groups of solutions. These results demonstrate that these tools are effective to assist the decision-making processes in the presence of a large number of solutions and/or objectives.

Keywords: Multivariate analysis · Multi-objective optimization · Dimensionality reduction · Decision-making

1 Introduction

Optimization refers to finding one or more solutions that correspond to extreme values or commitments to one or more objectives. Optimization plays a very important role in the various engineering areas where optimization problems arise, such as, for example, the design of new products, improvement of manufacturing and logistics processes. On the other hand, most of these engineering problems are multi-objective in nature. The simultaneous optimization of conflicting objectives allows to obtain an approximation to the Pareto optimal set. This set contains non-dominated solutions that represent...
different trade-offs. Then, the decision maker has to choose one of the alternatives according to his preferences. However, when the number of objectives and alternatives is high, the decision process is very difficult. These optimization problems with four or more objectives are, therefore, generally referred to as problems with many objectives.

Evolutionary algorithms are population-based algorithms commonly used in multi-objective optimization, in which one seeks to find approaches to different Pareto optimal solutions [1]. Very recently, these algorithms have been developed and applied to solve problems with a high number of objectives [2]. However, due to the large number of solutions and the high dimension of the objective space, the decision-making process becomes difficult and sometimes almost impractical [3]. Thus, the visualization and analysis of the solutions obtained by the algorithms are crucial for decision-making [4].

In [5], it is proposed a multi-objective evolutionary algorithm to solve problems with many objectives, trying to reduce the dimension of the Pareto front in terms of the number of objectives throughout the search. In practice, it is not always easy to identify the redundancy of objectives. This study highlights the usefulness of dimensionality reduction techniques, in the context of multi-objective optimization with many objectives. The use of PCA of responses in multi-objective problems can also be found in [6].

Brockhoff and Zitzler [7] also identified the difficulties of solving problems with a high number of objectives, both in terms of the quality of the approximations obtained to the optimal set of Pareto, as well as the execution time and also the complexity of the decision-making process. To overcome these difficulties, they proposed the reduction of objectives by identifying those that can be omitted while maintaining the dominance relationship between solutions. Brockhoff and Zitzler [8] defined a generalized notion of conflict between objectives, which specifies the necessary and sufficient conditions for the inclusion of the objectives to be optimized.

Principal component analysis was also used to reduce dimensionality in multi-objective problems by Costa and Oliveira [9] to assist the decision-making process. The use of graphic representations describing the relationship between objectives was also proposed, allowing the simultaneous visualization of all solutions with all objectives, thus facilitating the detection of conflicting objectives. Costa and Oliveira [4] argue that the notion of conflicts between objectives can be misleading, since the hypothesis of the existence of trade-offs between objectives implies that all objectives are important. Thus, even objectives that are not necessary to induce and preserve the dominance relationship between the solutions can contain some useful information to assist the decision-making process.

In this paper, multivariate statistical methods are used to assist the decision-making process for problems with many objectives, in particular, in the engineering area.

In Sect. 2, multivariate analysis techniques and graphical representations in the context of multi-objective optimization are presented. The results of application of the proposed method to two many-objective optimization engineering problems are presented and discussed in Sect. 3. Finally, in Sect. 4, some conclusions and future work are addressed.
2 Multivariate Analysis

In certain multi-objective problems, some objectives are not important to define the Pareto front. In this case, the dimensionality of the problems can be reduced and, consequently, the decision-making process becomes simpler and more treatable. On the other hand, it is interesting to identify the similarities between the solutions belonging to the sets of non-dominated solutions. For this purpose, multivariate statistical methods can be useful in situations where several variables must be measured simultaneously and the connections, similarities and dissimilarities must be identified.

The goal is to summarize, analyze and interpret data in which several response variables are evaluated. In the context of optimization with many objectives, it is intended to interpret and understand the relationships between the objectives as well as the similarities between solutions. For this purpose, multivariate statistical methods such as Principal Component Analysis (PCA) and Cluster Analysis (CA) as well as graphical representation tools such as biplots are of particular use and relevance.

2.1 Principal Component Analysis

Principal Component Analysis is a multivariate statistical method that makes it possible to transform a vector with the variables correlated with each other into a vector of unrelated (orthogonal) variables that are called principal components. In this way, these principal components are calculated in decreasing order of their importance, therefore, the first explains the maximum variance of the original data, the second the maximum variance not yet explained, and so on. Therefore, the last principal component will be the one that least contributes to the explanation of the variation of the initial data. PCA differs from other methods in that its objective is not to find correlation between variables, but to find mathematical functions in the initial variables that can explain as much of the variation that exists in the data and that allows them to describe and reduce them.

The objective of reducing dimensionality is to find a smaller number of objectives that have the relevant information. The reduction in dimension is obtained by transforming the original variables into a new set of variables, called principal components that are not correlated. Ordering is an important property, since it is this that justifies why the first principal components can be considered to have the greatest variability in the data set. However, a disadvantage is that it is difficult to deal with the existence of different types of nonlinearities in data [4].

2.2 Cluster Analysis

Cluster Analysis is used with the objective of grouping objects or individuals classified according to a set of characteristics or variables in groups called clusters. CA can be also useful to reduce the dimensionality by grouping individuals with similarities. In addition to defining a measure of similarity between the objects to be measured, it is also necessary to have a grouping method (hierarchical or non-hierarchical) and an algorithm that groups the objects into clusters according to the measure. Currently, several clustering algorithms are available in the literature [10]. One of the most used is
the k-means algorithm because it is, in general, effective and easy to implement. In this algorithm, data points are separated into k distinct groups of equal variation, minimizing the sum of the square of the distance between the points, i.e., the squared sum of the errors. The goal is to group individuals with similar characteristics in the same group, so it is important to define k, the number of clusters. Several methods, such as the elbow method and the gap method can be considered to define the optimal number of clusters.

2.3 Graphical Representations

In general, for optimization problems with many conflicting objectives, decision makers have to compare several alternatives and select the most preferred. The comparison of multidimensional vectors is very difficult, especially without any support tool. In order to support the decision maker to identify and understand the similarities and dissimilarities, several graphical visualization tools can be used. However, visualizing and interpreting graphs is not always simple. On the one hand, the graphs must be easy to understand and must not omit much information, on the other hand, no extra unintended information should be included [11].

Visualization techniques can be classified according to the principles used: representations involving bar diagrams, scatter diagrams or value paths, diagrams using circles and polygons, representations based on hierarchical clustering and projection-based. However, there are other advanced graphic representation techniques that are very useful in large spaces due to the large number of objectives and the number of solutions in the final set of approximations to the Pareto optimal solutions. In the presence of more than two objectives, one of the main difficulties is the visualization of the compromises between different objectives represented by each solution. The usual two-dimensional representation does not provide a better view of how a given solution is positioned in the different projections [4]. Therefore, advanced graphical representations can be used such as biplots and Andrews graphs.

A biplot is a projection-based technique, related to scatter plots. In the context of multi-objective optimization, PCA can be used to find a plan (two dimensions) where the criteria can be projected. In fact, the prefix “bi” refers to the joint display of rows and columns of the original data matrix that contains the alternatives. The biplot is a scatter plot that represents a two-dimensional matrix, both by lines (points) and columns (vectors). This plot allows the simultaneous visualization of all solutions with all objectives, thus facilitating the identification of conflicting objectives and trade-offs [4].

The Andrews graph is one of the solutions for representing and interpreting multivariate data [12]. This graph is a representation that transforms criteria vectors into two-dimensional curves with the aid of Fourier series [11]. All alternatives can be graphically represented in the same coordinate system. The Euclidean distance in the transformation is preserved. Therefore, the criterion vectors next to each other are transformed into curves that are not far from each other.
2.4 Non-dominated Sets Analysis

This section briefly describes the proposed method for analyzing sets of non-dominated solutions. The ultimate goal is to assist the decision-making process. This analysis combines and integrates the multivariate techniques described in the previous sections. Two distinct phases are implemented in order to facilitate the interpretation and to achieve the best results.

In the first phase, the aim is to identify possible redundant objectives. PCA is applied to the set of non-dominated solutions, involving three main steps: preparation of the data, determination of the correlations between objectives, computation of the explicability and the relationship between objectives. Finally, the graphical representation, the objectives and the solutions are presented graphically using, for instance, biplots or the Andrews graph.

In a second phase, CA is applied to the set of non-dominated solutions. This is done considering the results of PCA. The identification of the number of clusters is crucial for the use of the k-means algorithm. The graphic representation of the clusters can be done in terms of the two-dimensional projections of the Pareto front or other multivariate graphics such as biplots, allowing the decision maker to identify trade-offs. The two phases of the proposed method to analyze a set of non-dominated solutions can also be used interactively with the decision maker.

3 Results

This section presents the results of the application of the proposed method to analyze sets of non-dominated solutions for two multi-objective optimization problems of the areas of mechanical engineering and electronic engineering: a laminated plate design optimization problem and a radar performance optimization problem. The methods to analyze these engineering problems were implemented in R language [13].

3.1 Laminated Plate Design Optimization

This multi-objective optimization problem is related to the design of a laminated plate consisting of several layers or lamina of different materials that can have different thickness. Each layer is made of a single material. Thus, the decision variables are the thicknesses, the Young’s modulus and the Poisson’s ratio. The thicknesses of the layers can take any value from a continuous range and there are a finite number of different materials available. Therefore, this is a mixed integer optimization problem because there are continuous and discrete variables. The problem can be considered an optimization problem with many objectives where the objectives to be minimized are the compliance that is inversely related to the rigidity, the total price, the total mass and the total thickness of the plate. More details about this problem are presented by Costa et al. [14].

The data set comprises a total of 337 non-dominated solutions that are strongly correlated with each other in terms of the four objectives of compliance, total price, total mass and total thickness.
Table 1 shows the correlations between the objectives. There is a negative correlation between complacency and all other goals. On the other hand, there is a positive correlation between price, mass and thickness.

Table 1. Correlations of objectives for the laminated plates design optimization problem.

|                | Compliance | Price | Mass  | Thickness |
|----------------|------------|-------|-------|-----------|
| Compliance     | 1.0000     | −0.5381| −0.5327| −0.3114   |
| Price          | −0.5381    | 1.0000| 0.8740| 0.6641    |
| Mass           | −0.5327    | 0.8740| 1.0000| 0.7145    |
| Thickness      | −0.3114    | 0.6641| 0.7145| 1.0000    |

The goal of using PCA is to find the objectives that have less variability in the data and that, possibly, can be omitted by losing the minimum quantity of information. Table 2 presents the summary of the principal components (PCs) obtained by PCA. The criteria considered it to retain PCs with variance greater than one and a cumulative proportion of explicability superior to 80%. The variance of PC1 is superior to 1 and the cumulative proportion of PC1 and PC2 is 89.09%. Thus, PC1 and PC2 are retained.

Table 2. Summary of PCs for the laminated plate design optimization problem.

|                | PC1      | PC2      | PC3      | PC4      |
|----------------|----------|----------|----------|----------|
| Standard deviation | 1.6900   | 0.8411   | 0.5607   | 0.3494   |
| Proportion of variance | 0.7140   | 0.1769   | 0.0786   | 0.0305   |
| Cumulative proportion | 0.7140   | 0.8909   | 0.9695   | 1.0000   |

Figure 1 shows the correlation cycle between the objectives. In this graph, it is possible to observe that:

- the participation of each objective in PC1 and PC2, where the objectives furthest from the center of the circle are best represented with the contribution value closest to one;
- the correlation between the objectives, where all objectives facing the same direction correlate positively and negatively with the objectives facing the opposite side;
- the contribution of each objective (the red corresponding to the highest contribution and the blue to the lowest contribution);
- the quality of representation of each objective (greater intensity corresponds to greater quality).

Thus, it is possible to observe that the compliance objective has a better contribution and quality in the two PCs. On the other hand, price, mass and thickness objectives are correlated positively with each other and all of them negatively with the compliance. Only the results for PC1 and PC2 are presented as they are the most important for this analysis.
Figure 2 shows the quality of representation of the solutions in terms of PC1 and PC2 where the red color corresponds to the best represented solutions and the blue one the least well represented. For example, solutions close to the origin of the graph will be less well represented by components. On the other hand, solutions 187 and 20 are better represented in relation to the others, in particular, in the objective compliance (see also Fig. 1). Due to the large number of solutions, CA may play an important role here to identify similarities between solutions and the trade-offs they represent.

The objectives are highly correlated. This fact is important for the analysis of clusters. The purpose of cluster analysis is to group solutions with similar characteristics in the same group. In order to define the optimal number of clusters, the elbow and gap methods were used. The elbow method indicates that the major contributions are found in the first four clusters. On the other hand, the gap method suggests only three clusters. Since the aim is to minimize the number of clusters, the latter solution was adopted and computed using the k-means algorithm. Considering three clusters as indicated by the gap method, it can be observed in Fig. 3 that the clusters are well defined.

The Andrews graph can be represented for each cluster. For instance, observing the Andrews graph for cluster3 (Fig. 4), there are three groups of curves separated into three different colors (red, black and purple). These three colors group solutions with similar characteristics. These solutions are closer together and have very close objective values.
Figure 5 shows the two-dimensional projections of the Pareto front with the well-defined clusters represented by the colors red, green and black. Among these objectives there are high correlations facilitating the definition of different groups. Considering this information, the decision maker is able to choose which of the clusters best matches his/her preferences.

**Fig. 2.** Quality of representation of the solutions in PC1 and PC2 for the laminated plate design optimization problem (Color figure online).

**Fig. 3.** Representation of the three clusters in terms of PC1 and PC2 for the laminated plate design optimization problem.
Fig. 4. Andrews graph for cluster3 for the laminated plate design optimization problem (Color figure online).

Fig. 5. Two-dimensional projections of Pareto front with clusters for the laminated plate design optimization problem (Color figure online).
3.2 Radar Performance Optimization

This multi-objective optimization problem involves nine objectives [15]. It refers to the design of a waveform for a Pulsed Doppler radar, typical of many aerial combat systems. Radar systems are important for measuring both the distance and the speed of targets. Unfortunately, for very long distances and very high speeds, using a simple waveform it is only possible to measure: i) distance without ambiguity, but ambiguous speed; ii) speed without ambiguity, but with ambiguous distance or; iii) ambiguity in both distance and speed. In order to allow error-free transmissions, a simple set of waves is transmitted, each one subtly different from the other. The results of these multiple waveforms are combined to resolve ambiguities. In fact, the problem is how to choose the optimal set of simple waveforms. More details about this difficult problem are presented by Hughes [15].

The data comprises 11,938 non-dominated solutions. Most of the correlations between the nine objectives are high, which indicate the existence of significant associations. Table 3 summarizes those correlations where the correlations appear ordered according to its value and in bold if greater than 0.5.

| Objectives | Positive correlations | Negative correlations |
|------------|-----------------------|-----------------------|
| f1         | f3 > f7 > f5          | f9 > f4 > f2 > f8 > f6 |
| f2         | f9 > f4 > f8 > f6     | f3 > f1 > f7 > f5     |
| f3         | f1 > f7 > f5          | f9 > f4 > f2 > f8 > f6 |
| f4         | f9 > f2 > f8 > f6     | f1 > f3 > f7 > f5     |
| f5         | f1 > f3 > f7          | f9 > f4 > f2 > f8 > f6 |
| f6         | f8 > f2 > f9 > f4     | f3 > f1 > f5 > f7     |
| f7         | f1 > f3 > f5          | f4 > f9 > f2 > f8 > f6 |
| f8         | f2 > f9 > f4 > f6     | f3 > f1 > f5 > f7     |
| f9         | f4 > f2 > f8 > f6     | f3 > f1 > f7 > f5     |

The summary of the principal components resulting from PCA is given in Table 4. It can be seen that the first two principal components already accumulate 84% of the variance. Thus, it can be considered that PC1 and PC2 are sufficient to explain the variability of the initial data. In this problem, even without the remaining seven PCs, it is possible to explain most of the data variability.

| PC1  | PC2  | PC3  | PC4  | PC5  | PC6  | PC7  | PC8  | PC9  |
|------|------|------|------|------|------|------|------|------|
| 2.55 | 1.02 | 0.81 | 0.69 | 0.40 | 0.32 | 0.19 | 0.17 | 0.04 |

Table 3. Correlations of objectives for the radar performance optimization problem.

Table 4. Summary of PCs for the radar performance optimization problem.
The correlation cycle for the first two principal components PC1 and PC2 is shown in Fig. 6. In addition to the correlations between the objectives, the graph also represents the contribution of the objective to the respective PC (where the red color has the largest representation, intermediate yellow color and blue color the lowest representation). On the other hand, vectors oriented in the same direction indicate that the objectives are positively correlated and are negatively correlated when in opposite directions. So, between objectives f1, f3, f5, f7 there is a positive and very strong correlation, and they are negatively correlated with objectives f2, f4, f8, f9, these being positively correlated between them. The objective f6 can be considered independent of the remaining objectives.

![Fig. 6. Correlation cycle between the objectives for the radar performance optimization problem (Color figure online).](image)

Given the very large number of non-dominated solutions, it is not easy to identify trade-offs. In this problem, cluster analysis is crucial. The elbow method indicates three clusters and the gap method suggests nine clusters. Since the aim is to obtain a reduced number of groups of solutions, three clusters were considered and computed by the k-means algorithm. In Fig. 7, the three clusters are represented in terms of PC1 and PC2. Note that points of each cluster almost do not overlap between them, which suggests that the number of clusters is adequate.

Figure 8 shows the two-dimensional projections of the Pareto front for the different combinations of the objectives with the three clusters in different colors (black, red and green). It can be observed that there are some overlaps of the solutions due to the large number of solutions and objectives that difficulties the separation of solutions in groups. For each cluster, it is possible to select one or more representative solutions, for example, those closest to the respective centroids.
Fig. 7. Representation of the three clusters in terms of PC1 and PC2 for the radar performance optimization problem.

Fig. 8. Two-dimensional projections of Pareto curve with clusters for the radar performance optimization problem (Color figure online).
In order to find representative solutions for each cluster, the solution closest to the centroid was selected. In Fig. 9, it is plotted the Andrews graph where each curve corresponds to the representative solutions of each cluster. The light red curve and the darkest red curve are closer and correspond to more similar representative solutions in terms of objectives. The black curve presents a different behavior in relation to the other curves, taking negative positions when the others are positive.

4 Conclusions

In this work, it is proposed to use multivariate statistical methods, such as Principal Components Analysis and Cluster Analysis to help the decision-making process in many-objective optimization problems. When the number of non-dominated solutions and/or objective is large, it is very difficult the visualization of trade-offs as well as the relationships between objectives. Therefore, the proposed method allows to inspect associations between objectives, identify redundant objectives, and similarities between solutions by defining clusters. Clustering is particularly interesting since allow to select one or more representative solutions, for example, those closest to the cluster’s centroid. Moreover, graphical tools are critical for the visualization of results. This way, the decision-making process is substantially facilitated.

The proposed method was applied to two difficult many-objective engineering optimization problems. For the problem of the design of laminated plates through PCA allowed to reduce to two objectives and through CA it was possible to create three distinct groups of solutions. For the problem of optimization of the radar waveform, it was possible to reduce the objectives from nine to two objectives representing the greatest variability of the data, and CA defined three distinct groups of solutions. Moreover, it was possible the identification and selection of representative non-dominated solutions.

Fig. 9. Andrews graph with the representative solutions for the radar performance optimization problem (Color figure online).
These results obtained with the engineering problems demonstrate the validity and usefulness of the approach to assist the decision maker.

Future work comprises the development of multiphase strategies to provide interaction with the decision maker as well as the study of other multivariate statistical methods. Strategies to incorporate different levels of importance or preferences of criteria will be also investigated.

Acknowledgements. This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020.

References

1. Deb, K.: Multi-objective Optimization Using Evolutionary Algorithms. Wiley, Chichester (2001)
2. Hughes, E.J.: MSOPS-II: A general-purpose many-objective optimiser. In: 2007 IEEE Congress on Evolutionary Computation, pp. 3944–3951. IEEE (2007)
3. Purshouse, R.C., Fleming, P.J.: Evolutionary many-objective optimisation: An exploratory analysis. In: 2003 Congress on Evolutionary Computation – CEC 2003, vol. 3, pp. 2066–2073. IEEE (2003)
4. Costa, L., Oliveira, P.: Biplots in offline multiobjective reduction. In: 2010 IEEE Congress on Evolutionary Computation - CEC 2010, pp. 18–23. IEEE (2010)
5. Deb, K., Saxena, D.K.: On finding pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems. In: 2005 IEEE Congress on Evolutionary Computation – CEC 2005. IEEE (2005)
6. Reis, M.S., Pereira, A.C., Leça, J.M., Rodrigues, P.M., Marques, J.C.: Multi-response and multi-objective latent variable optimization of modern analytical instrumentation for the quantification of chemically related families of compounds: Case study - Solid Phase Microextraction (SPME) applied to the quantification of analytes with impact on wine aroma. J. Chemom. 33(3), e3103 (2019)
7. Brockhoff, D., Zitzler, E.: Dimensionality reduction in multiobjective optimization with (partial) dominance structure preservation: Generalized minimum objective subset problems. TIK Report, 247 (2006)
8. Brockhoff, D., Zitzler, E.: On objective conflicts and objective reduction in multiple criteria optimization. Peabody J. Educ. 0161956X, 81(243), 180–202 (2006)
9. Costa, L., Oliveira, P.: Multiobjective optimization: Redundant and informative objectives. In: 2009 IEEE Congress on Evolutionary Computation - CEC 2009, pp. 2008–2015. IEEE (2009)
10. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: A review. ACM Comput. Surv. 31(3), 264–323 (1999)
11. Miettinen, K.: Survey of methods to visualize alternatives in multiple criteria decision making problems. OR Spectrum 36(1), 3–37 (2012). https://doi.org/10.1007/s00291-012-0297-0
12. Khattree, R., Naik, D.: Andrews plots for multivariate data: Some new suggestions and applications. J. Stat. Plan. Infer. 100, 411–425 (2002)
13. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2019). https://www.R-project.org
14. Costa, L., Fernandes, L., Figueiredo, I., Júdice, J., Leal, R., Oliveira, P.: Multiple- and single-objective approaches to laminate optimization with genetic algorithms. Struct. Multidiscip. Optim. 27(1–2), 55–65 (2004)

15. Hughes, E.J.: Radar waveform optimisation as a many-objective application benchmark. In: Obayashi, S., Deb, K., Poloni, C., Hiroyasu, T., Murata, T. (eds.) EMO 2007. LNCS, vol. 4403, pp. 700–714. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-70928-2_53