Visualization of The Boundary Solutions of High Dimensional Pareto Front from A Decision Maker's Perspective

AKM Khaled Ahsan Talukder
Computational Optimization and Innovation (COIN) Laboratory
Michigan State University
East Lansing, Michigan, USA
talukde1@msu.edu

Kalyanmoy Deb
Computational Optimization and Innovation (COIN) Laboratory
Michigan State University
East Lansing, Michigan, USA
kdeb@egr.msu.edu

Julian Blank
Computational Optimization and Innovation (COIN) Laboratory
Michigan State University
East Lansing, Michigan, USA
blankjul@egr.msu.edu

ABSTRACT

For the last couple of years, the development of many-objective optimization problems opened new avenues of research in the evolutionary multi-objective optimization domain. There are already a number of algorithms to solve such problems, now the next challenge is to interpret the results produced by those algorithms. In this paper, we propose an alternative way to visualize high-dimensional Pareto fronts where the goal is to present the Pareto front in terms of a decision maker's perspective. A decision maker is more interested in the different aspects of the end results instead of the convergence and spread of a Pareto front solutions. They are interested in Pareto-optimal solutions that offer the most trade-off. They are also interested to know the boundary solutions of a Pareto front. In this paper, we present a way to visualize the Pareto front in high dimension by keeping those criteria in mind.

CCS CONCEPTS

• **Applied computing** → **Multi-criterion optimization and decision-making**; • **Human-centered computing** → *Visualization theory, concepts and paradigms*;

KEYWORDS

Many-objective Optimization, Pareto-front, Visualization, Decision Making, Neighborhood Embedding, Boundary Solutions, Cluster

ACM Reference Format:
AKM Khaled Ahsan Talukder, Kalyanmoy Deb, and Julian Blank. 2018. Visualization of The Boundary Solutions of High Dimensional Pareto Front from A Decision Maker's Perspective. In *GECCO '18 Companion: Genetic and Evolutionary Computation Conference Companion, July 15–19, 2018, Mitaka, Japan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3205651.3205782

GECCO '18 Companion, July 15–19, 2018, Kyoto, Japan

ACM ISBN 978-1-4503-5764-7/18/07.
https://doi.org/10.1145/3205651.3205782

1 INTRODUCTION

With the recent advancement in many-objective (i.e. problems with four or more conflicting objective functions) optimization algorithms, there have been a number of post-optimization issues that need to be addressed. One of the most challenging problem is to understand the results produced by a many-objective optimization (MOOP) solver. In an MOOP scenario, if the target problem is consisted of three objectives, a scatter plot is the most intuitive way to visualize the results, which allows the *decision maker (DM)* to easily understand the trade-off between the objectives, robustness of a solution (i.e. risk assessment) or analyzing the neighborhood of a particular solutions etc. Naturally, since it is not possible for a DM to comprehend four or more spatial dimensions visually, we try to interpret data points from higher dimensions by mapping them over a lower dimensional space.

The existing approaches [3] do not take into account the practical aspects of a PF solutions for decision making. For example, one might want to know what are the extreme solutions that optimizes only one objective. Another instance being which solution offer most trade-off? If there are multiple clusters in the PF then what solutions compose the boundaries in those cluster? And how are we going to visualize them? These questions are not generally addressed in the current multivariate data visualization techniques.

In this paper, we are going to present an alternative idea to visualize the Pareto front in high dimension by keeping those questions in mind, specially the boundary solutions of a high-dimensional Pareto front.

2 VISUALIZATION METHOD

From our previous experiences with industry collaborations, we have seen that decision making procedure requires a completely different set of criteria than an EMO developer might pursue. For example, a DM might want to know the where the boundary solutions of a Pareto front exist.

2.1 Finding Boundary Solutions

The boundary solutions are important if the PF is consisted of multiple isolated clusters and the boundary solutions reside on the boundary of a cluster. Hence they are helpful to identify robust solutions in the PF. The boundary point extraction algorithm works as follows:

Algorithm 1 Extract $\mathcal{PF}$ Boundaries

Require: $\mathcal{C} \leftarrow \{C_1, C_2, \ldots, C_m\}$ clusters found from applying DBSCAN [1] on the PF solutions.
Require: $K \leftarrow$ number of boundary solutions required by the DM for each cluster.
$\mathcal{B} \leftarrow \emptyset$, set of boundary points
for each $C_i \in \mathcal{C}$ **do**
  Find eigenvectors ($\mathbf{v}$) and eigenvalues (λ) of C_i
  $E \leftarrow \{(\mathbf{v}_1, \lambda_1), (\mathbf{v}_2, \lambda_2), \ldots, (\mathbf{v}_n, \lambda_n)\}$
  $E \leftarrow$ sort E according to descending λ
  for j from 1 to n **do**
    $V \leftarrow \{\mathbf{v}_j, -\mathbf{v}_j, \mathbf{u}_j, -\mathbf{u}_j\}$ s.t. $\mathbf{u}_j \cdot \mathbf{v}_j = 0$
    for each $\mathbf{w} \in V$ **do**
      $\mathbf{x} \leftarrow \min_{\mathbf{y} \in C_i} \frac{\mathbf{y} \cdot \mathbf{w}}{\|\mathbf{y}\| \|\mathbf{w}\|}$
      if $|\mathcal{B}| < K$ **then**
        $\mathcal{B} \leftarrow \mathcal{B} \cup \mathbf{x}$
      else
        return $\mathcal{B}$
      end if
    end for
  end for
end for

The idea is to find the furthest points on the cluster (clusters of data points can be found using standard algorithms, such as DBSCAN [1]) from the centroid along a number of directions rotated at a specific interval starting from the principal axis. We draw a line from the cluster centroid along the direction of the eigenvector (with the highest eigenvalue) and find the points on the cluster that is the furthest from the centroid. The overall algorithm is presented in Algorithm 1.

2.2 Mapping onto A Lower Dimensional Space

For the visualization and representation, our method relies on the mapping of high dimensional PF objective values onto lower dimensional space (preferable on a two-dimensional space). In order to achieve this, we need to have a mapping that preserves most of the the local structure of the data points from the higher dimensional space, properties like solution trade-offs, boundaries and outliers can be visualized in an intelligible way. We utilize a method called *Stochastic Neighborhood Embedding (SNE)* [2]. SNE is a probabilistic approach that can place data points, described by high-dimensional vectors or by pairwise dissimilarities, in a low-dimensional space in a way that preserves neighborhood relations. One such visualization result is presented in the Figure 1.

3 CONCLUSIONS AND FUTURE WORK

In this paper, we have demonstrated an alternative approach to address the issue of high dimensional Pareto front visualization. Our approach is more practical in a sense that it takes account of the DM's perspective in the visualization mechanism. Using this same technique, we can also represent solutions with lowest/highest trade-offs (i.e. knee points) using the metric introduced in [5]. In spite of some limitations, we present our idea as an initial "proof of concept". Our approach is mainly based on the SNE mapping which does not retain the global topological relations among the data points. In the future we would like to address this issue, by adopting/modifying other topologically consistent neighborhood embedding techniques.

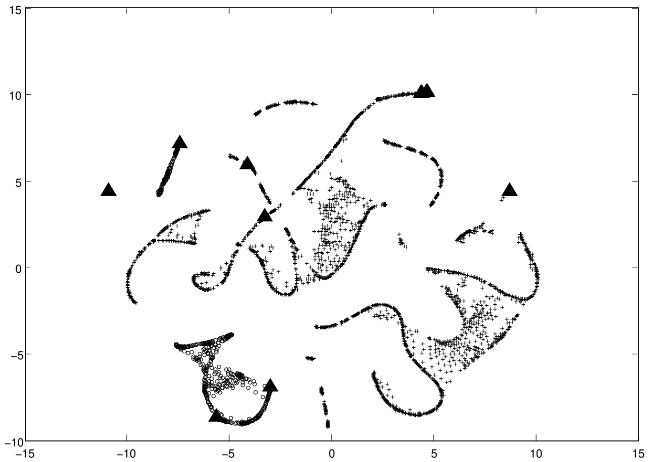

Figure 1: The SNE mapping of an example Pareto front generated from the CAR-CRASH problem [4]. The clusters are shown in different markers and the filled triangles are the boundary points.

REFERENCES

[1] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. A Density-based Algorithm for Discovering Clusters a Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96)*. AAAI Press, Palo Alto: CA, 226–231.
[2] Geoffrey E Hinton and Sam T. Roweis. 2003. Stochastic Neighbor Embedding. In *Advances in Neural Information Processing Systems 15*, S. Becker, S. Thrun, and K. Obermayer (Eds.). MIT Press, Cambridge, MA, 857–864.
[3] A. Inselberg. 2009. *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications*. Springer, Berlin, Germany.
[4] Xingtao Liao, Qing Li, Xujing Yang, Weigang Zhang, and Wei Li. 2007. Multiobjective optimization for crash safety design of vehicles using stepwise regression model. *Struct Multidisc Optim* 35, 6 (nov 2007), 561–569.
[5] Lily Rachmawati and Dipti Srinivasan. 2009. Multiobjective Evolutionary Algorithm With Controllable Focus on the Knees of the Pareto Front. *IEEE Transactions on Evolutionary Computation* 13, 4 (Aug 2009), 810–824.