Visualization of Pareto-Sets in Evolutionary Multi-Objective Optimization

Mario Köppen, Kaori Yoshida
Kyushu Institute of Technology
Dept. Artificial Intelligence
680-04 Kawazu, Iizuka, Fukuoka 820-8502, Japan
E-Mail: mkoeppen@ieee.org, kaori@ai.kyutech.ac.jp

Abstract

In this paper, a method for the visualization of the population of an evolutionary multi-objective optimization (EMO) algorrithm is presented. The main characteristic of this approach is the preservation of Pareto-dominance relations among the individuals as good as possible. It will be shown that in general, a Pareto-dominance preserving mapping from higher- to lower-dimensional space does not exist, so the demand to have as few wrong dominance relations after the mapping as possible gives an objective in addition to other mapping objectives like preserving nearest neighbor relations. Thus, the mapping itself poses a multi-objective optimization problem by itself, which is also handled by an EMO algorithm (NSGA-II in this case). The resulting mappings are shown for the run of a modified NSGA-II on the 15 objective DTLZ2 problem as an example. From such plots, some insights into evolution dynamics can be obtained.

1 Introduction

By the formulation of more than one optimization goal, or more than one decision criterion, the true nature of an increasing number of practical problems can be better reflected. The most common application field here is the engineering of design, but other fields are also gaining more and more advantage by taking this perspective, for example benchmarking techniques, fault tolerance, scheduling, networking and the more. This development is accompanied by the provision of improving techniques to handle such problems, and among them are the evolutionary multi-objective optimization algorithms (EMO).

A notable point of the multi-objective optimization is the higher degree of interactivity required. In contrary to classical optimization, where usually a single "best" point in the search space is solicited, in a typical multi-objective problem specification, a trade-off between conflicting objectives exists, and there is more than one solution. Before, afterward, or during the optimization, a "decision maker" has to intervene, since in a practical approach, only one of the many trade-offs can be used as a solution. In a typical EMO algorithm, like NSGA-II, SPEA, PAES, only the set of optima solution (the Pareto-front) is approached, while the final selection of a single solution is left to the user of the system, or its developper.

Visualization of the optimum set, so far, assisted the user of a system for making a final selection or decision from the Pareto-front. Several techniques have been developped in the past. The most simple one is the technique of "parallel lines," which is in practical use in the engineering design for decades already. Here, each line represents one objective by some scale, and a set of objectives is plotted by setting all points on all lines to their corresponding values. However, this is only helpful for the visualization of a small number of points. Other techniques employ features derived from the Pareto-front for visualization purposes, often with a statistical base (Interactive Decision Maps [6], Hyper-space Diagonal Counting [1], SOM-based [7]).

Such techniques are needed, and useful for the processing, once the Pareto-front has been found. However, before, they are not needed, or even not applicable. So, in this paper, we want to contribute to the visualization of the *search for the Pareto-front* by visualizing an evolving population of an EMO algorithm.

On a first glance, this seem to be an easy task, with a lot of background already existing. As a dataset in a higher dimensional space, classical methods of mapping into lower-dimensional spaces like Sammons mapping can be used to map the elements of an approximated Pareto-front into 2D- or 3D-space. Here, we object that such a mapping is not specific for the mapping of a Pareto-front, as it does not care for the preservation of the underlying Pareto-dominance relation among points in the high-dimensional space.

Several other problems are related to such an approach. One is the rapdidly decreasing number of Pareto-dominated points with increasing number of objectives [4]. From 5 or 8 objectives onward, the concept of Pareto-dominance seems to be lesser suitable to grasp the meaning of "better" in a way that can be employed in a heuristic search algorithm like any EMO. The second point is, and this will be shown in this paper, that there is no general Pareto-dominance preserving mapping from a higher-dimensional space to a lower-dimensional space. This brings, as a third point, the additional need for handling trade-offs in the mapping itself.

Based on this considerations, we are proposing a two-stage mapping of a higher-dimensional data set. First stage is the mapping of the Pareto-set (i.e. the Pareto-front in a multi-objective search problem) to a Pareto-set in 2D, and the second stage is the mapping of the dominated points. The first stage will include, in some cases, a multi-objective optimization by itself. Here, we are considering the two objectives *low number of wrong implicit dominance relations* (dominances appearing among the mapped points, but not among the original points), and *best represenation of distances*. This will appear as a multi-objective version of the Traveling Salesman Problem (TSP).

In the next section, we will give the reasoning guiding to these optimization goals in visualization of approximated Pareto-fronts. Section 3 then will detail the mapping procedure, section 4 give an example and also demonstrate, what kind of inference can be done from this way of visualizing. The paper ends with a conclusion, and the listing of open problems and future tasks.

2 Background

For yielding a Pareto-dominance preserving mapping Π of a set $P_i \in R_{D_1}$ with $i = 1, \cdots, N$ of N points from a higher dimension D_1 to a set of points $p_i \in R_{D_2}$ of lower dimension $D_2 < D_1$, we would require the following property:

$$\forall_{1 \le i,j \le N} : P_i >_D P_j \longleftrightarrow p_i >_D p_j \tag{1}$$

where $p_i = \Pi(P_i)$. While such a mapping is easy to compose in the right-to-left direction (by using any function Π that is monotone in each argument), the other direction comes out to be impossible. This is established by the following theorem.

Theorem 1. *If $D_1 > D_2$, then there is no Pareto-dominance preserving mapping of all points of R_{D_1} into R_{D_2}.*

The proof will be given in the appendix. Here, we just want to scetch the basic idea of the proof. Consider the three points $P_1 = (1, 2, 3), P_2 = (2, 3, 1)$ and $P_3 = (3, 1, 2)$ from R_3 and assume there is a mapping of these three points into three points p_1, p_2 and p_3 in R_2. As indicated by figure 1, in such a mapping, always one of the three points is located "between" the two other points (for a more formal specification of this property see the proof in the appendix). In the case of the example above, p_2 is assumed to be located between p_1 and p_3. What's essentially making the Pareto-dominance preserving mapping impossible is the fact that for a lower dimension, there are dependencies between the dominance relations that do not need to exist among the original points. In the given example, for any point in R_2 that is dominated by p_1 and p_3, it will also be dominated by p_2 (the grey area in fig. 1). However, if we consider the point $(1, 1, 2)$ from R_3, it is actually max-dominated by P_1 and P_3, and thus it should be mapped into the area that is dominated by p_1 and p_3 (again, the grey area in fig. 1). However, the point $(1, 1, 2)$ is *not* dominated by P_2, but if P_2 is mapped between p_1 and p_3, then the projection of $(1, 1, 2)$ will be dominated by p_2. Therefore, p_2 cannot be located between p_1 and p_3. By considering the points $(2, 1, 1)$ and $(1, 2, 1)$ in a similar manner, it can also be seen that p_1 cannot be located between p_2 and p_3, and that p_3 cannot be located between p_1 and p_2. Together this gives the impossibility to map the Pareto-dominance relations among the six points $(1, 2, 3), (2, 3, 1), (3, 1, 2), (1, 1, 2), (1, 2, 1)$ and $(2, 1, 1)$ into R_2 in a Pareto-dominance preserving manner. The proof in the appendix basically generalises this fact to any dimensions.

From this fact, we can see that there is no direct way of mapping a Pareto-front to a dimension open for human comprehension (one to three dimensions) in the way we would like. The only possible way is to keep the number of wrongly indicated dominance relations at a minimum.

3 Two-stage visualization procedure

The proposed procedure uses different means of mapping the dominated and non-dominated elements of a set. The mapping of the dominated set is following

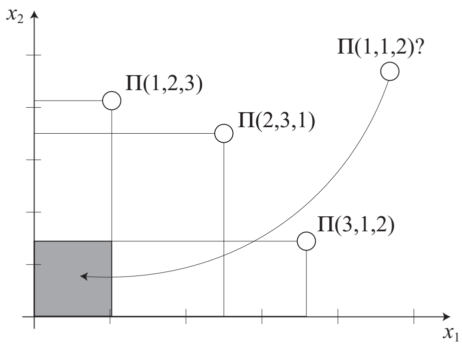

Figure 1. The impossibility to project 3D points into 2D, while preserving Pareto-dominance. Assume the given constellation of the mapped points, than it can be seen that there is no place to map the point $(1, 1, 2)$, since this point is dominated by $(1, 2, 3)$ and $(3, 1, 2)$, but not by $(2, 3, 1)$.

two main principles: it is a mapping onto a Pareto-set in 2D, where variations can be used to visualize characteristics of the high-dimensional data sets; and it is optimized in a manner that the following mapping of the dominated points makes less errors, but also fulfills other optimality criteria of the mapping, like distance preservation in this case.

The points of the Pareto-set are mapped onto points on a quarter-circle in the first quadrant, i.e. onto points (x, y) fulfilling $x, y \geq 0$ and $x^2 + y^2 = r^2$. The distribution of the points on the quarter-circle, and its radius r are subject to representing characteristics of the Pareto-set. In the approach presented here, we distribute the points according to match their nearest-neighbour distances, and for the radius to reflect the average point norm (it was considered to derive r from the hyper-volume [3], but despite of the existence of faster algorithms, for this purpose the computational load is still too high to be used in practice).

Once the Pareto-set itself is mapped, the remaining points, all dominated, are mapped following a greedy procedure. We use the notion of a *minimum point* of a set of points.

Definition 1. *Given a set S of N points P_i from R_n. Then, the minimum point* $\min(S)$ *of S is the point P with components*

$$P[i] = \min_{j=1,\cdots,N}(P_j[i]) \tag{2}$$

for $i = 1, \cdots, n$.

For short, any dominated point is mapped to the minimum point of the mappings of all points of the Pareto-front, by which it is dominated. This assignment can do some errors due to what we call *implicit dominance*: any Pareto-set in 2D is equivalent to an ordering of its points. The point with one objective lowest, the other highest, comes first with ordering number 1, and the one with the first objective highest, and the other lowest comes last. The points can be sorted by the one objective increasing, or the other decreasing. So, if there is a point that is dominated by two points A and B in the original space, and A and B gets mapped to points a and b with the ordering numbers i and j respectively, then in 2D, the point will be also dominated by all points with ordering numbers between i and j. This is not necessarily the case in the original space. The genral impossibility of a Pareto-dominance preserving mapping is based on the fact that wrong implicit dominances can never be avoided in general. However, different orderings oft the mapped points may result in different number of wrong implicite dominances, including also making no such error.

Taking this as an objective, the first stage mapping can also be considered as a multi-objective optimization problem, which is a multi-objective version of the Euclidian-TSP:

Given a set of n non-dominated points P_i that is mapped to a set of non-dominated points p_i in two-dimensional Euclidian space, and be $p_{(i)}$ the point among the p_i with the i-th highest first component (or $(n - i)$-th lowest second component). There is also a set of dominated points D_l. The mapping $i \to (i)$ is a permutation Γ of the set $(1, 2, \cdots, n)$. With $d(P_i, P_j)$ we denote the Euclidian distance between the points P_i and P_j. Then, the search is for a permutation Γ minimizing the following objectives:

$$O_1(\Gamma) = \sum_{i=1}^{n-1} d(P_{(i)}, P_{(i+1)}) \tag{3}$$

$$O_2(\Gamma) = ||\{k \mid \exists 1 \leq i < k < j \leq n, D_l \text{ with } P_i >_D D_l, P_j > D_l \text{ but not } P_k >_D D_l\}|| \tag{4}$$

After this overview, the steps in more detail.

1. Map the set of points (e.g. population of an EMO algorithm) by some suitable monotone and component-wise transfer function into the set of points ${Q_i}$[1]. Q is decomposed into a set of non-dominated points P_i and a set of dominated points D_i.

[1]This in case that the initial data range is complicating the direct visualization

2. Compute r as the average norm of the P_i, multiply by some scaling factor.

3. Find a permutation Γ for the ordering of the points P_i by multi-objective optimization for minimizing the two objectives given by eqns. (3) and (4).

4. Decide on a small border offset angle δ and distribute points p_i on the quarter-circle in the first quadrant with radius r between the arcs δ and $\Pi/2-\delta$ proportionally to the distances $d(P_{(i)}, P_{(i+1)})$, where (i) is the ordering index given by the permutation Γ for i. Now, P_i is mapped onto $p_{(i)}$, and this concludes the first stage, the mapping of the non-dominated set.

5. For each dominated point D_l, $dom(D_l)$ indicates the subset of all P_i that dominate D_l, and $\Pi(dom(D_l))$ the mappings of these points achieved in the foregoing step. Then, D_l is mapped onto the minimum vector of this set $\min(\Pi(dom(D_l)))$. This concludes the second stage, the mapping of all dominated points.

4 Example and discussion

In this section, we want to exemplify the procedure for an optimization of the DTLZ2 function, using an optimization run of the modified NSGA-II as described in [5]. Since the optimization itself is not the primary goal here, and also for space limitations, we will not give details here. The reader is kindly referenced to [2] for the DTLZ functions, and to [5] for details of the used algorithm.

The used algorithm was NSGA-II, with using $-\epsilon$-dominance as secondary ranking assignment, and it was applied to the DTLZ2 problem with 15 objectives, but only 6 were used for visualization (otherwise, there would be nearly not a single dominance case). Population size of the NSGA-II was 20, and the visualization went over 2000 generations. For other settings see [5]. Each component x of the objective vector was mapped to $1-\ln(x)$, to account for the low order of magnitude of the objective values. For each generation, the radius was computed as given above, and as scaling factor 200 was used, to generate output images of size 500×500 pixels.

The optimization of the permutation Γ was performed by a "standard" NSGA-II, i.e. using crowding distance as secondary ranking assignment, population size 10, 100 generations, and the other settings as in [5]. For encoding a permutation, *real-valued sorting encoding* was used, i.e. the individual is a real vector,

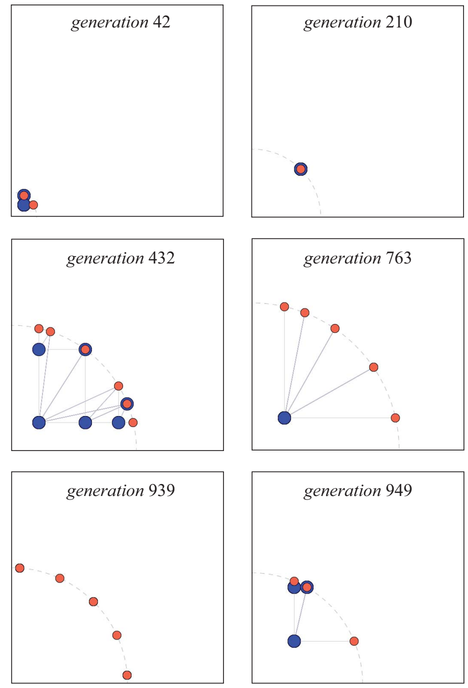

Figure 2. Visualization of the parent population of a modified NSGA-II on the DTLZ2 problem with 15 objectives for selected generation numbers.

which encodes a sorting by the sorting of its components (e.g. the vector $(0.7, 0.4, 0.5)$ encodes the permutation $(1, 3, 2)$ of $(1, 2, 3)$). This optimization was only initiated, when there were points dominated by more than one other point. Usually, the optimization was able to remove all wrong implicit dominances, since the number of dominated points is usually small.

Some plots generated this way are shown in fig. 2. These plots also show typical patterns that often appear in such plots. The red circles indicates the non-dominated elements, the blue the dominated ones. Blue circles "behind" red ones are points that are only dominated by one element of the approximated Pareto-front. The connections between red and blue circles indicates dominance relations. Note that usually, not all 20 points are visible, since the mapping often projects points onto the same position. The plot of generation 42 shows the situation shortly after initialization, the average norm is still small. A pattern like in the plot of generation 210 is often appearing in the evolution: a single point dominates the reminder of the population, but the progress in average norm (or any other indicator) is usually small. Generation 432 gives a more complex dominance pattern. Often, the appearance of such patterns for a few generations is followed by an increase in the radius, while the situations shown in the plots for generations 763 and 939, where nearly all the population is on the Pareto-front of the population, are usually followed by a loss in average norm. The plot of the lower right subfigure for generation 949 shows that the radius has decreased 10 generations after the whole population was on the Pareto-front.

Figure 3 also shows the effect of the optimization. Subfigure (a) shows the plot of generation 1572 without optimization, subfigure (b) after optimization. The structure of the dominance relations among the individuals has remarkably changed. But in general it has to be said that a pattern like the one for generation 210 in fig. 2 is much more pre-dominant than dominance-rich patterns.

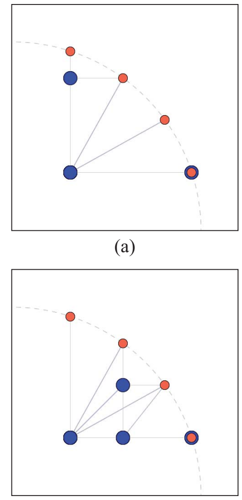

Figure 3. Example for the correction of the non-dominated point order in generation 1572. After the rearrangement (b), the figure shows all dominance relations correctly.

5 Conclusions

In this paper, the visualization of the population of an evolutionary multi-objective optimization algorithm (EMOA) was studied. The main goal was to have a visualization that gives insight into the Pareto-dominance relations among the individuals. It could be shown that in general, there is no Pareto-dominance preserving mapping from a higher-dimensional space to a lower-dimensional space. This is due to the appearance of wrong implicit dominance relations in the lower-dimensional space that are not present among the mapped points. Thus, the mapping of a population into a visual dimension like 2D requires some kind of trade-off, to have the mapping reflecting the distance relations among the individuals as good as possible, but also avoiding such wrong implicit dominances as much as possible. A procedure for visualizing a population was proposed, where this trade-off is solved by another multi-objective optimization algorithm. The resulting visualization procedure comes out to provide insight into the run of an EMOA, exemplified by DTLZ2 optimization with a modified NSGA-II.

Future work will focus on the employment of further characteristics of the plots, like size, shape and color of the circles, and shape of the mapped Pareto-set. In the latter case, for example, “bumpiness” could help to indicate concave parts of the Pareto-set. Also further work is required to ensure more smooth transitions between the plots of subsequent generations.

Acknowledgment

This research was partially supported by the Japan Society for the Promotion of Science, Grant-in-Aid for JSPS Fellows, 18.06094.

References

[1] G. Agrawal, C. Bloebaum, and K. Lewis. Intuitive design selection using visualized n-dimensional pareto frontier. In *46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, pages 1813–1826, 2005.

[2] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. Scalable Multi-Objective Optimization Test Problems. In *Congress on Evolutionary Computation (CEC'2002)*, volume 1, pages 825–830, Piscataway, New Jersey, May 2002. IEEE Service Center.

[3] M. Fleischer. The Measure of Pareto Optima. Applications to Multi-objective Metaheuristics. In C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, editors, *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, pages 519–533, Faro, Portugal, April 2003. Springer. Lecture Notes in Computer Science. Volume 2632.

[4] M. Köppen, R. Vicente-Garcia, and B. Nickolay. Fuzzy-Pareto-Dominance and Its Application in Evolutionary Multi-objective Optimization. In C. A. Coello Coello, A. Hernández Aguirre, and E. Zitzler, editors, *Evolutionary Multi-Criterion Optimization. Third International Conference, EMO 2005*, pages 399–412, Guanajuato, México, March 2005. Springer. Lecture Notes in Computer Science Vol. 3410.

[5] M. Köppen and K. Yoshida. Substitute distance assignments in NSGA-II for handling many-objective optimization problems. In S. Obayashi, K. Deb, C. Poloni, T. Hiroyasu, and T. Murata, editors, *Evolutionary Multi-Criterion Optimization, 4th International*

Conference, EMO 2007, Matsushima, Japan, March 2007. Proceedings, LNCS 4403, pages 727–741. Springer Berlin, Heidelberg, 2007.

[6] A. Lotov. Approximation and visualization of pareto frontier in the framework of classical approach to multi-objective optimization. In J. Branke, K. Deb, K. Miettinen, and R. E. Steuer, editors, *Practical Approaches to Multi-Objective Optimization*, number 04461 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany, 2005. <http://drops.dagstuhl.de/opus/volltexte/2005/235> [date of citation: 2005-01-01].

[7] S. Obayashi and D. Sasaki. Visualization and Data Mining of Pareto Solutions Using Self-Organizing Map. In C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, editors, *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, pages 796–809, Faro, Portugal, April 2003. Springer. Lecture Notes in Computer Science. Volume 2632.

Appendix

A. Proof of Theorem 1

Here, we are considering the maximization case, and also assume $D_1 = D_2 + 1$. The proof can be applied to the minimization case in the same manner, and extended to any dimension greater than D_2+1 by adding a number of constant components to the points that are used in the proof.

Before we can prove theorem 1, we need the following proposition.

Proposition 1. *Any set S of $(n+1)$ points P_i from R_n contains at least one point P_l that is weakly Pareto-dominating[2] the minimum point of the set[3] $S - P_l$.*

Proof. Assume that for a point P_k from S, P_k is not weakly Pareto-dominating the minimum point of $S - P_k$. This means that for at least one index i, $P_k[i] < min(S - P_k)[i]$. This also means that for such an index i, one and only one point P_l can have the property $P_l[i] = \min(S)[i]$. If the minimum of the i-th component would appear more than once as the i-th component of a point from S, then for any P_l and P_k from S with $k \neq l$, $P_l[i] \geq min(S - P_k)[i]$ would hold.

Now consider the following substitution of component values: replace the components of each point by their corresponding component-wise ranking values, i.e. replace the smallest first component of all points in the points, where it appears by the value "1," then replace by "2" all appearances of the second-smallest first component and so on, and do the same for the second and all other components, index by index. For example, if the set of points is $P_1 = (2, 8), P_2 = (3, 6)$ and $P_3 = (4, 4)$, then it will be replaced by the set of points $(1, 3), (2, 2)$ and $(3, 1)$. Now, according to the remark given at the beginning of the proof, from the set $(1, 2, \cdots, n)$, we remove all indices, where a "1" is appearing more than once and restrict our attention only to indices, where "1" is appearing exactly once. Since there are maximal n indices remaining after excluding double occurrences of minima, but $(n+1)$ points, there must be at least one point that does not contain a "1" at one of the remaining index positions. Each of these points than will not have the minimum component alone anywhere and thus fulfill the claim of the corollary. Also note that for any set, where the minimum component never appears alone, each point has the claimed property. This completes the proof of the proposition. □

[2] Note that "weakly" dominating includes equality.

[3] This gives a definition for what was called to be "between" the other points before.

Now we can proof the theorem by considering a mapping $p = \Pi(P)$ from points $P \in R_{(n+1)}$ to points $p \in R_n$.

Proof. Assume that there is such a Pareto-dominance preserving mapping Π from $R_{(n+1)}$ to R_n and consider the following set S of points in $R_{(n+1)}$: $P_1 = (1, 2, \cdots, n, (n+1)), P_2 = (2, 3, \cdots, n, (n+1), 1), \cdots, P_n = ((n+1), 1, 2, \cdots, n)$, or in general form $P_i[j] = (i + j - 1) \mod (n+1)$, where $P_i[j]$ is the j-th component of the i-th point. Be $p_i = \Pi(P_i)$ the Pareto-dominance preserving mapping of the point P_i to a point p_i in R_n. There are $(n+1)$ points p_i now selected in R_n, thus they fulfill the requirements of proposition 1, and there is one of them, say p_l, which is weakly dominating the minimum point of the other points. This means that any point, which is dominated by all mapped points but p_l also has to be dominated by p_l. Since Π was assumed to be Pareto-dominance preserving, this means that any point dominated by all points but P_l in $R_{(n+1)}$ is also dominated by P_l.

Now consider the minimum point Q of $S - P_l$, where all components besides of component l have the value "1," and component l has the value "2" (remember that always $P_i[i] = 1$). Obviously, any point of $S - P_l$ dominates Q, but P_l is not dominating Q since its l-th component is "1" and thus smaller than the l-th component of Q, which is "2." This contradicts with the fact that the mapping of P_l (weakly) dominates the mapping of Q, since Q has to be mapped into a region that is dominated by the mappings of all other points, i.e. (weakly) dominated by their minimum point. Therefore, such a mapping cannot exist. □