The Supplement of the Paper “Multi-Line Distance Minimization: A Visualized Many-Objective Test Problem Suite”

Miqing Li, Crina Grosan, Shengxiang Yang, Senior Member, IEEE, Xiaohui Liu, and Xin Yao, Fellow, IEEE

This supplement consists of three sections. Section I and Section II are the proof of Theorem 2 and Theorem 3, respectively. Section III first discusses the relation between existing test problems and the proposed ML-DMP, and then presents same/similar observations of the ML-DMP to existing test problems as well as also new findings obtained on the ML-DMP.

I. PROOF OF THEOREM 2

Theorem 2. For an ML-DMP \( \Omega = \mathbb{R}^2 \) with a regular polygon of \( m \) vertexes \( \{A_1, A_2, \ldots, A_m\} \), points inside the polygon (including the boundary points) are the Pareto optimal solutions. In other words, for any point in the polygon, there is no point \( c \in \mathbb{R}^2 \) that dominates it.

Proof: Let us first consider the boundary points. It is clear that the \( m \) vertex points of the polygon are Pareto optimal since they are the intersection point of two target lines and have minimum distance (i.e., 0) to these two lines. Non-vertex boundary points have the minimum distance (0) to only one target line. Consider a five-objective ML-DMP in Fig. 1, where boundary point \( P_1 \) is located on the target line \( A_1A_2 \) and has the best value on this objective. Clearly, the point that is able to dominate such a boundary point should be located on the same target line, if existing. Without the loss of generality, assume that \( P_2 \) is the point dominating \( P_1 \). According to the definition of the Pareto dominance, \( d(P_2, A_2A_3) \leq d(P_1, A_2A_3) \). However, since \( d(P_2, A_2) > d(P_1, A_2) \), we have \( d(P_2, A_2A_3) > d(P_1, A_2A_3) \), therefore, a contradiction.

Now consider the points inside the polygon. For an interior point \( P_3 \) in Fig. 1, assume there is a point \( P_4 \) such that \( P_4 \prec P_3 \). Draw a semi-straight line starting from \( P_4 \) and passing through \( P_3 \) (i.e., \( P_4P_3 \)). Since \( P_3 \) is inside the polygon, there must be an intersection point of \( P_4P_3 \) and the polygon boundary. This means that there exists at least one target point that is able to dominate such a boundary point should be located on the same target line, if existing. Without the loss of generality, assume that \( P_2 \) is the point dominating \( P_1 \). According to the definition of the Pareto dominance, \( d(P_2, A_2A_3) \leq d(P_1, A_2A_3) \). However, since \( d(P_2, A_2) > d(P_1, A_2) \), we have \( d(P_2, A_2A_3) > d(P_1, A_2A_3) \), therefore, a contradiction.

II. PROOF OF THEOREM 3

Theorem 3. Considering an ML-DMP with a regular polygon of \( m \) vertexes \( \{A_1, A_2, \ldots, A_m\} \), the feasible region \( \Omega = \Phi \land S \), where \( \Phi \) is the union set of all the constrained polygons and \( S \) is a two-dimensional rectangle space in \( \mathbb{R}^2 \) (i.e., the rectangle constraint defined by the marginal values of decision variables). Then, the points inside the regular polygon (including the boundary) are the sole Pareto optimal solutions of the ML-DMP.

Proof: This theorem consists of two parts: 1) every point inside the polygon is Pareto optimal (which we have seen in Theorem 2), and 2) for any feasible point outside the polygon, there exists at least one interior point that dominates it.

To prove the second part (i.e., for any given point outside the polygon, to find a point inside the polygon that dominates it), we can draw \( m \) lines passing through the given point such that they parallel the \( m \) target lines, respectively. This naturally leads to two situations: 1) there is at least one of these parallel lines intersecting with the regular polygon, and 2) there is no intersection of these parallel lines and the polygon. Next, we consider these two situations separately.

For the first situation, let us consider the example in Fig. 2, where \( A_1 \) to \( A_5 \) are the five vertexes of the regular pentagon and the five “rectangular wings” of the pentagon are constrained areas. Considering feasible point \( P \) in the figure,
\( P' \) is the intersection point of two straight lines \( \overrightarrow{A_1A_2} \) and \( \overrightarrow{P''P} \), where \( \overrightarrow{PP} \) parallels \( \overrightarrow{A_5A_1} \). Now we prove that \( P' < P \).

\( P' \) dominating \( P \) means that \( P' \) is closer to all the target lines than \( P \) (or being equal on some target lines). For target lines \( \overrightarrow{A_5A_1} \) and \( \overrightarrow{A_1A_2} \), we have \( d(P',\overrightarrow{A_5A_1}) = d(P,\overrightarrow{A_5A_1}) \) and \( d(P',\overrightarrow{A_1A_2}) < d(P,\overrightarrow{A_1A_2}) \). Now we consider the remaining target lines \( \overrightarrow{A_2A_3}, \overrightarrow{A_3A_4} \) and \( \overrightarrow{A_4A_5} \), which can be divided into two groups. One group corresponds to those intersecting with \( \overrightarrow{PP} \) at a point above the target line \( \overrightarrow{A_1A_2} \) (i.e., \( \overrightarrow{A_3A_4} \) and \( \overrightarrow{A_4A_5} \)), and the other corresponds to those below \( \overrightarrow{A_1A_2} \) (i.e., \( \overrightarrow{A_2A_3} \)). For the target lines of the first group, it is clear that their distance to \( P' \) is shorter than that to \( P \) because the intersection point is on the extended line of segment \( PP' \) with the direction from \( P \) to \( P' \). For the target lines of the second group, namely \( \overrightarrow{A_2A_3} \) here, since \( \overrightarrow{A_2A_3} \) and \( \overrightarrow{A_5A_1} \) are on either side of point \( P' \), the intersection point (denoted as \( P'' \)) of \( \overrightarrow{A_5A_1} \) and \( PP' \) is inside the constrained area determined by \( \overrightarrow{A_2A_3} \) and \( \overrightarrow{A_5A_1} \). Then, it holds that \( |\overrightarrow{PP''}| < |\overrightarrow{PP'}| \); otherwise \( P \) would be inside the constrained area (i.e., an infeasible point). So, \( d(P',\overrightarrow{A_2A_3}) < d(P,\overrightarrow{A_2A_3}) \). This proves that \( P' < P \).

In the above example, target line \( \overrightarrow{A_2A_3} \) is on the other side of point \( P' \) relative to the target line (i.e., \( \overrightarrow{A_5A_1} \)) that parallels \( PP' \). One may ask what will happen if they are on the same side of the intersected point. Fig. 2 also gives an example of this situation, where \( Q' \) is the intersection point of \( \overrightarrow{A_5A_1} \) and \( QQ' \), and \( QQ' \) parallels \( \overrightarrow{A_1A_2} \). Similar to the case of \( P \), we can prove that \( d(Q',\overrightarrow{A_5A_1}) < d(Q,\overrightarrow{A_5A_1}) \) and \( d(Q',\overrightarrow{A_1A_2}) < d(Q,\overrightarrow{A_1A_2}) \). However, for target line \( \overrightarrow{A_2A_3} \) which is on the same side of point \( Q' \) relative to \( \overrightarrow{A_1A_2} \), \( d(Q',\overrightarrow{A_2A_3}) \) could be larger than \( d(Q,\overrightarrow{A_2A_3}) \), as shown in the example. To deal with this, we can draw a line \( L \) parallel to \( \overrightarrow{A_2A_3} \) such that their distance is the same as that of \( Q \) to \( \overrightarrow{A_2A_3} \). We denote \( Q'' \) as the intersection point of \( L \) and \( \overrightarrow{A_2A_3} \). Now we prove that \( Q'' < Q \).

First, we can easily know that \( d(Q',\overrightarrow{A_2A_3}) = d(Q,\overrightarrow{A_2A_3}) \), \( d(Q'',\overrightarrow{A_2A_3}) < d(Q,\overrightarrow{A_2A_3}) \) and \( d(Q'',\overrightarrow{A_1A_2}) < d(Q,\overrightarrow{A_1A_2}) \). For the remaining target lines \( A_4A_5 \) and \( A_1A_2 \), they intersect with line \( QQ'' \) at a point below \( \overrightarrow{A_2A_3} \). Thus, it holds that \( d(Q',\overrightarrow{A_2A_3}) < d(Q,\overrightarrow{A_2A_3}) \) and \( d(Q'',\overrightarrow{A_2A_3}) < d(Q,\overrightarrow{A_2A_3}) \). This proves that \( Q'' < Q \). Now one may ask if in other ML-DMPs with more objectives (i.e., more target lines) there exists one target line that intersects with \( QQ'' \) at a point above \( \overrightarrow{A_2A_3} \). In fact, the answer is no; otherwise \( Q \) would be inside the constrained area determined by that line and target line \( \overrightarrow{A_2A_3} \) since \( d(Q',\overrightarrow{A_2A_3}) > d(Q,\overrightarrow{A_2A_3}) \).

The above proved that for any point (outside the regular polygon) which has at least one parallel line intersecting with the polygon, we can find a point inside the polygon that dominates it. Next, we consider the second situation – there is no intersection of the given point’s parallel lines and the polygon. However, there may exist that the symmetric line of one (or more) of the parallel lines (with respect to the corresponding target line) intersects with the polygon. For example, for point \( R \) in Fig. 2, the symmetric line of \( L' \) and \( L'' \) with respect to \( \overrightarrow{A_1A_2} \) and \( \overrightarrow{A_4A_5} \), respectively, intersects with the polygon. According to the number of such intersected symmetric lines, we can further divide the second situation into four sub-cases: 1) no intersection, 2) one intersected line, 3) two intersected lines, and 4) more than two intersected lines, and then consider them separately.

For sub-case 1, it is clear that all the points inside the polygon dominate the given point. For sub-case 2, any interior points in the area constructed by that intersected symmetric line and the corresponding target line dominate the given point. For sub-case 3 (see the point \( R \) example in the figure), we have two intersected lines, and for each line there exists one “dominating” area. Thus, the points located in the intersection part of the two areas dominate the given point (there must exist overlapping part of these two areas; otherwise the given point will be inside the associated constrained area). Now consider sub-case 4. In fact, there do not exist three (or more) of the symmetric lines intersecting with the polygon. To explain this, we consider the reduction to absurdity method. Assume that there are three (or more) of such intersected lines. It is clear that any pair of them has an overlapping area. For a pair of the intersected lines, the given point is located inside the extension of their overlapping area. When there are three (or more) of such intersected lines, this implies that the given point is located on both sides of at least one of the lines, which is a contradiction. Therefore, we complete the proof for the second situation, and now the theorem is proved.

III. COMPARISON WITH EXISTING TEST PROBLEMS

Test problems play an important role in understanding the strengths and weaknesses of EMO algorithms. In many-objective optimization, several test problem suites have been
widespread use, such as DTLZ [39], WFG [40], Knapsack [41], TSP [42], and MNK-Landscapes [43] suites. The DTLZ suite consists of seven continuous test problems, which are scalable to the number of objectives and decision variables. The WFG suite has nine continuous test problems, which are scalable in the objective and decision variable dimensions. In contrast to the DTLZ suite, the WFG suite introduces a wide variety of problem attributes, e.g., the separability/non-separability, uni-modality/multi-modality, and concavity/convexity/mixture. In WFG, solutions contain k position and l distance parameters, which determine their distribution and their distance to the Pareto front, respectively. Knapsack, TSP, and MNK-Landscapes are three typical combinatorial optimization problems which are extended from single-objective optimization. Recently, researchers have also presented some new MOPs for many-objective optimization [44]–[46]. They either emphasize the complexity of the geometrical shape of the Pareto front/set, or consider correlation between decision variables and objective functions. Like the above test problems, the ML-DMP has its own specific properties, such as having a degenerate Pareto front when the number of objectives is larger than three, a lot of the dominance resistant solutions, and several pairs of completely conflicting objectives in a certain region. However, the salient feature of the ML-DMP is the visualization property. That is, its Pareto optimal solutions in the two-dimensional decision space have the geometrical similarity to their images in the high-dimensional objective space. This thus allows us to observe the search behavior of algorithms; for example, an algorithm tends to lead their solutions towards a certain area of the optimal front, and an algorithm prefers a set of solutions distributed regularly but not uniformly over the optimal front. Such information may not be able to be provided by performance indicators (via returning a scalar value to assess the algorithm’s solution set on a given test problem) in many-objective optimization.

In view of this, an extensive experimental study had been carried out. From this, it has been found that most of the observations (conclusions) obtained on the ML-DMP were the same as (or similar to) those on the proven test problems in the area. Table I summarizes these observations. It consists of the specific algorithm behavior, what problem the algorithm was tested on, and what paper this observation was reported from.

On the other hand, we have also obtained several new findings, some of which had not been observed on existing test problems. For example, conventional Pareto-based algorithms may completely fail on a four-objective MOP. A combination of decomposition-based (or indicator-based) algorithms with Pareto dominance seems to be promising, especially on low-dimensional MOPs which have a huge search space. The algorithm ε-MOEA can struggle to maintain the uniformity on some many-objective problems which are easy to converge. This is in contrary to some previous studies [14], [47], where the ε-dominance had been found to work well in this respect in the high-dimensional objective space. HypE may struggle to diversify its solutions over the boundary of the optimal front in a particular many-objective problem. This finding is interesting and different from the previous experience that hypervolume-based algorithms typically prefer the boundary solutions to the central ones [3], [48], [49].

### REFERENCES


