Revisiting Pareto-Optimal Multi- and Many-Objective Reference Fronts for Continuous Optimization

Gabriela Cavalcante da Silva
IMD, Universidade Federal do Rio Grande do Norte
Natal, RN, Brazil
gabrielacavalcante@ufrn.edu.br

Elizabeth F. Wanner
DECOM
CEFET-MG
Belo Horizonte, Brazil
efwanner@cefetmg.br

Leonardo C. T. Bezerra
IMD, Universidade Federal do Rio Grande do Norte
Natal, RN, Brazil
leobezerra@imd.ufrn.br

Thomas Stützle
IRIDIA
Université Libre de Bruxelles
Belgium
stuetzle@ulb.ac.be

***Abstract*—The performance assessment of multi-objective heuristic algorithms is one of the most significant contributions from the evolutionary optimization algorithms community. By contrast, performance assessment in the context of many-objective optimization is still a challenging, open research field. Recent advances have demonstrated disagreements between Pareto-compliant performance metrics, and indicated that reference fronts produced by benchmark generators of Pareto-optimal fronts could be further improved. In this work, we investigate these reference fronts with the help of multi-dimensional visualization techniques and Pareto-monotonic archivers. Interestingly, reference fronts produced by benchmark generators for DTLZ and WFG continuous optimization problems show significant issues, even when only three objectives are considered. Furthermore, given that input solution sets for five-objective problems are not high-quality, archivers are unable to output reasonable approximation fronts. We conclude that the performance assessment of EMO algorithms needs to urgently address reference front generation.**

I. Introduction

Assessing the quality of approximation fronts produced by heuristic optimization algorithms is a critical subject in evolutionary multi-objective (EMO) research. Over the years, a solid theoretical ground has been set with the proposal of (i) Pareto-compliant performance indicators [1], intended to measure the desirable properties of approximation fronts, and; (ii) scalable, artificial benchmark problems for continuous optimization [2], [3], designed to assess EMO algorithms in the presence of different numbers of objectives and problem characteristics deemed representative of real-world scenarios.

More recently, novel investigations in the context of many-objective optimization have revealed that these approaches may still require improvements. For instance, Pareto-compliant performance measures have been shown to increasingly disagree as a function of the number of objectives and problem characteristics [4], [5]. Regarding artificial benchmark problem sets, some DTLZ [2] and WFG [3] problems have been shown to present issues when scaled to three or more objectives [3], [5], [6]. Moreover, reference fronts produced by these benchmark generators have shown room for improvement [5], [7].

Altogether, the concerns regarding the performance assessment of EMO algorithms in the context of many-objective optimization pose a real challenge for the community. In more detail, the most commonly adopted performance measures are unary indicators that have been recast as binary, requiring reference fronts for their computation. So far, these reference fronts have been generated either by aggregating the output of high-performing algorithms or by using benchmark generators themselves. The latter option has been considered advantageous over the former, as it is believed that these fronts are an ideal representations of the true Pareto front. In fact, no standard reference fronts for these problems were ever made available, under the assumption that producing high-quality fronts from the generators is a straightforward task.

In this work, we thoroughly investigate reference fronts produced by DTLZ and WFG problems for three and five objectives. We start by randomly sampling increasingly large fronts for selected problems and assess them visually with the help of multi-dimensional visualization [8]. Interestingly, reference fronts present significant issues even for three objectives. For instance, since generators sample points from a surface, fronts produced for multi-modal problems include dominated solutions sometimes to a very large extent. In addition, their distribution and spread are very poor for both benchmark sets, and even fronts with $100,000$ optimal solutions are unable to well approximate WFG problems. This is further worsened when five objectives are considered, and no DTLZ or WFG problem can be well approximated.

In the second part of this work, the reference fronts produced are filtered by Pareto-monotonic archivers. Our rationale is that large fronts would render measure computation expensive, and thus we reduce the size of the randomly sampled fronts using the multi-grid (MGA [9]) and the hypervolume (AA$_s$ [10]) archivers. The outcomes of the archivers are once again assessed with the help of multi-dimensional visualization. For the fronts that had been successfully approximated through random sampling, each archiver favors a different distribution in its outcome. For the remaining fronts, archivers are affected by the issues of the original input.

The insights observed in this work raise a significant concern with the assessment of EMO algorithms in the context of multi- and many-objective optimization. As discussed, Pareto-compliant measures directly depend on the quality of the reference fronts considered. Though good approximations can be found for a handful of problems, our findings indicate that reference front generation should be urgently addressed at the risk of assessing EMO algorithms from a biased perspective. This is even more important in many-objective optimization, where the performance assessment theory is still maturing.

The remainder of this work is organized as follows. Section II briefly reviews concepts, benchmark problems, and approaches related to this work. Next, we detail in Section III the experimental design we adopt in our assessment, and respectively discuss results from random sampling and archiver processing in Sections IV and V. Finally, we conclude and discuss future work in Section VI.

II. Background

In this section, we briefly review the most relevant concepts required to understand the remainder of this work. Initially, we discuss the importance of reference fronts and the desirable properties they should present. Next, we review the origins of artificial benchmark problems in the context of continuous optimization and briefly summarize their properties. We also review archivers as a means to reduce fronts in a Pareto-compliant way, detailing the archivers we consider in this work. Finally, we review multi-dimensional visualization approaches, describing the technique we apply in this work.

A. Performance Assessment

In the context of multi- and many-objective optimization, reference fronts play a critical role due to their application to measure computation. In more detail, the most desirable property for performance measures in EMO assessment is *Pareto-compliance*. If a measure $\mathcal{I}$ is Pareto-compliant and it favors a given front $\mathcal{A}$ over another given front $\mathcal{B}$, then we are assured that $\mathcal{A}$ cannot be dominated by $\mathcal{B}$. Thus far, three performance measures have been proven Pareto-compliant, namely the hypervolume, the ϵ-indicator family (additive and multiplicative), and more recently the inverted generational distance plus. Since their proposal, the first two have been extensively adopted in the context of multi-objective optimization, and the latter has been increasingly used in the context of many-objective optimization.

Though the hypervolume indicator is originally unary, it is most commonly employed recast as a *binary indicator*, such as the ϵ-indicators and the inverted generational distance plus. In more detail, binary indicators are computed in respect to a reference front that should be a good approximation of the actual Pareto front. A high-quality reference front is thus critical to performance measures, and should present the following desirable properties:

Convergence. Refers to the Pareto-optimality of the solutions. A reference front should not contain dominated solutions. If solutions in a reference front are produced by a benchmark generator, they are already expected to be Pareto-optimal. Conversely, reference fronts computed by aggregating outputs from EMO algorithms cannot be guaranteed to be optimal.

Spread. Refers to the distance between extreme solutions. Ideally, the optimal solution for each objective should be included in reference fronts, as well as the points that are placed along the edges of the surface that the problem represents.

problem	separability	modality	bias	geometry
DTLZ2	✓	uni	–	concave
DTLZ4	✓	uni	polynomial	concave
DTLZ7	✓	uni, multi	–	disconnected
WFG1	✓	uni	polynomial	convex, mixed
WFG2	–	uni, multi	–	convex, disconnected
WFG4	✓	multi	–	concave

TABLE I
SUMMARIZED DESCRIPTION OF BENCHMARK PROBLEMS CONSIDERED.

Distribution. Refers to the evenness in spacing between solutions over the objective space. High-quality reference fronts should not present gaps, missing out entire regions of the true Pareto front.

Recent works have empirically demonstrated that even Pareto-compliant metrics may value spread and distribution differently as a function of problem characteristics and the number of objectives considered [5]. Nonetheless, perfectly spread and distributed reference fronts are generally regarded ideal in EMO research.

B. Artificial Benchmark Problems

A priori knowledge of problem characteristics and optimal solutions are desirable assets in the context of EMO. Since problems that demand heuristic optimization are challenging for exhaustive enumeration of optimal solutions (and hence understanding problem characteristics), EMO researchers have devised several artificial benchmark problem sets, specially for continuous optimization. In particular, the DTLZ [2] and WFG [3] sets are the most widely used in benchmarking, mostly due to their scalability w.r.t. number of variables and objectives. In common, these artificial problem sets comprise distance- and position-related variables, which respectively affect only convergence (distance) or spread and distribution (position). More importantly, this separation serves as a backdoor to produce Pareto-optimal solutions. Specifically, if distance-related variables are set to pre-specified values, sampling position-related variables results in different solutions expected to be Pareto-optimal.

Though scalable, a common issue with works that benchmark EMO algorithms using the DTLZ and WFG problem sets is that researchers often employ them with the original number of variables and ratio between distance- and position-related variables proposed by their authors. This greatly hinders the benefits of their configurable and scalable design. In fact, over the years some of these problems were reported to present issues as the number of objectives or variables are increased [3], [5], [6]. For this reason, in this work we restrict our analysis to DTLZ2, DTZL4, and DTLZ7, WFG1, WFG2, and WFG4. Table I lists the most important properties from each problem, where we remark that WFG4 is representative of the remaining concave WFG problems (WFG5–9).

It is important to highlight that the multi-modality nature of the functions that comprise DTLZ7 and WFG2 implicate that their fronts are disconnected. This is shown in Figure 1, provided in the original DTLZ paper [2] to illustrate the objective space of DTLZ7 with three objectives. In a minimization problem like this, the Pareto-optimal solutions concentrate in patches that correspond to the valleys of the functions. By contrast, solutions in the hills of the surface are dominated by the solutions in the valleys.

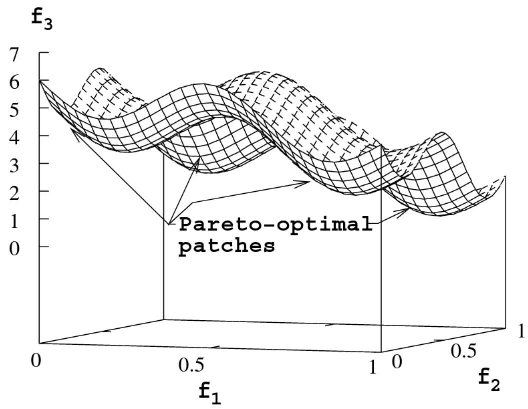

Fig. 1. Illustration of a multi-modal problem provided in [2], which presents a disconnected Pareto front.

Among the problems we select, we remark that it is possible to produce perfectly spread and distributed reference fronts for DTLZ2 and DTLZ4 using exact approaches such as Das and Dennis' method for weight generation [11]. However, this is not trivial for the remaining problems. More importantly, in the context of many-objective optimization weight generation itself is ongoing research, with different alternatives leading to different compromises between spread and distribution.

C. Solution Archivers

Similar to solution populations in EMO algorithms, reference fronts must be limited in size but high-quality approximations of the actual Pareto front. For this reason, it is possible to use the same archiving techniques proposed for EMO populations in the context of reference front truncation. Among the most important properties, archivers should present is Pareto-monotonicity, meaning that the inclusion of novel solutions in the archiver will not lead to deterioration. So far, only two archivers have been proven to present this property [12], namely the multi-grid (MGA [9]) and the hypervolume (AA$_s$ [10]) archivers:

MGA uses the concept of box-dominance to discretize the objective space into cell grids and compute solution density. The grid is dinamically computed as a function of the extreme solutions provided in the input front. Besides being Pareto-monotonic, the MGA is the only archiver expected to scale polynomially as the number of objectives is increased.

AA$_s$ evaluates solutions based on their contribution to the hypervolume measure. Similarly to the MGA, the AA$_s$ is also Pareto-monotonic and dynamic, with the reference point used for hypervolume computation being a function of the extreme solutions provided in the input front. In contrast to the MGA, though, the computational cost of the AA$_s$ scales exponentially with the increase in the number of objectives.

Though no work has graphically discussed the expected shapes of fronts processed by MGA, for AA$_s$ we expect that its behavior be similar to the hypervolume indicator. In [4], authors discuss the effect of problem geometry and the number of objectives on the preferences captured by different performance measures. For the hypervolume, distribution is favored over spread on convex problems, such as WFG1 and WFG2. Conversely, on concave problems like DTLZ2, DTLZ4, and the WFG problems we represent with WFG4, the hypervolume favors spread over distribution.

D. Multi-dimensional Visualization

Visualizing solution sets for multi- and many-objective optimization problems is paramount for estimating location, distribution, and shape of the fronts, among other important characteristics. Several multi-dimensional visualization techniques can be identified in the EMO and multi-criteria decision making (MCDM) literature [13], [15]. Below, we briefly describe four of the most relevant and/or recent:

Prosection uses projection and rotation to produce 3D representations of sections from a 4D front [13]. It is able to capture information about Pareto dominance between solutions, and front shape and distribution. Although prosections can deal with large approximation sets in small computational time, this approach is currently only applicable to 4D fronts.

Sammon mapping projects the original data into a lower dimensional space while trying to preserve inter-point distance [14]. This is achieved by modelling the error as a stress function and employing an optimization method to iteratively refine an initial projection (usually a principal component). Though Sammon plots are able to preserve distribution, the values of the original objectives cannot be indivually assessed using this kind of representation.

Radial coordinate visualization (RadViz) maps m-dimensional points into a two-dimensional, radial coordinate plane. Objectives are modeled as dimensional anchors placed along equidistant angles, and the location of a given point represents the equilibrium between the strength of each anchor [8]. The advantage of RadViz is its reproducible anchor displacement, and that the distribution of a large set of points is captured in an easy and scalable way. However, RadViz does not preserve Pareto dominance relation between points nor the geometry of the front being assessed.

PaletteViz decomposes a RadViz visualization into layers that move progressively from front boundaries to its center [16]. Points are colored to evidence their distance to the boundaries, and stacked layers enable better understanding of front structure and neighborhood information. Moreover, stacking enables decision makers to navigate different trade-off levels among the objectives. While being very effective as a visualization tool, PaletteViz is yet not available as a software.

In the next section, we describe the experimental setup we adopt in this work, where multi-dimensional visualization helps us assess the quality of reference fronts produced from benchmark generators for the problems discussed in this section, before and after archive truncation.

III. Experimental Setup

As discussed in the previous section, DTLZ and WFG are configurable benchmark problems that provide Pareto-optimal solution generators. In this section, we design an experimental setup to evaluate the approximation fronts produced by these generators before and after archiver truncation. The most relevant details of the setup adopted are given below. Since archivers can be computationally demanding, we also list hardware specifications for the batch of experiments we run on a computational cluster:

Problem setup. Following [7], each problem comprises 50 variables to be optimized. We adopt problem-specific parameter settings and the ratio between position- and distance-related variables proposed in the DTLZ and WFG papers, with the adaptations discussed in [7] to meet problem constraints.

Reference front generation. For each problem selected, we uniformly randomly sample $n \in \{1\,000, 10\,000, 100\,000\}$ solutions. As discussed, this is done by setting distance-related variables to pre-specified optimal values, and uniformly randomly sampling values for the position-related variables within their pre-defined domains. For each n, we adopt 10 different seeds for random number generation, which are paired for different front sizes. Reference fronts are produced both for $M = \{3, 5\}$ objectives, to assess the impact of the increase in the number of objectives.

Archiver truncation. We adopt the MGA and AA_s implementations respectively provided by [9] and [10]. Archivers take as input a maximum output size, a given reference front, and the seed used for its generation, in case a seed is required. Truncated fronts present a maximum size $n = 100$ when $M = 3$ and $n = 1\,000$ when $M = 5$, following [7].

Multi-dimensional visualization. We employ RadViz for its simplicity of use, scalability, and availability as open source software.We argue that, since our goal is to visualize Pareto-optimal reference fronts in multi- and many-objective optimization problems, not preserving Pareto dominance between solutions is not an issue.

We adopt two different implementations of RadViz. In preliminary experiments, we employed the implementation provided in the Orange3 data mining suitefor interactive visualization. The most relevant insights were selected to be provided as plots in this paper and as supplementary material [17], and are produced using the RadViz implementation provided in the *pandas* Python data science programming library. Reproducible notebooks for plotting are also provided as supplementary material, along with the data produced [17].

Performance measure computation. To complement the visual analysis, we compute the unary hypervolume indicator of the truncated reference fronts. As traditional in the literature, we use $[2.1]^M$ as reference point, after normalizing values to the $[1, 2]$ interval. For all problems, we take the origin of the axes as lower bound for normalization. Additionally, for DTLZ problems we consider 1.1 as upper bound. However, the upper bound for the M-th objective of DTLZ7 is $2.2M$, given the domain defined for this objective. Similarly, for WFG problems upper bound for objective i figures $2.2i$.

Hardware. Computationally demanding jobs are executed on a cluster running CentOS 6.2. Each job runs on a single core from a rack node containing 2 Intel Xeon E5-2680 v3 (12 cores each, 2.5GHz, 2x 16MB L2/L3 cache), with 2.4Gb RAM available per job.

In the following sections, we discuss results obtained from experiments before (Section IV) and after archiver truncation (Section V).

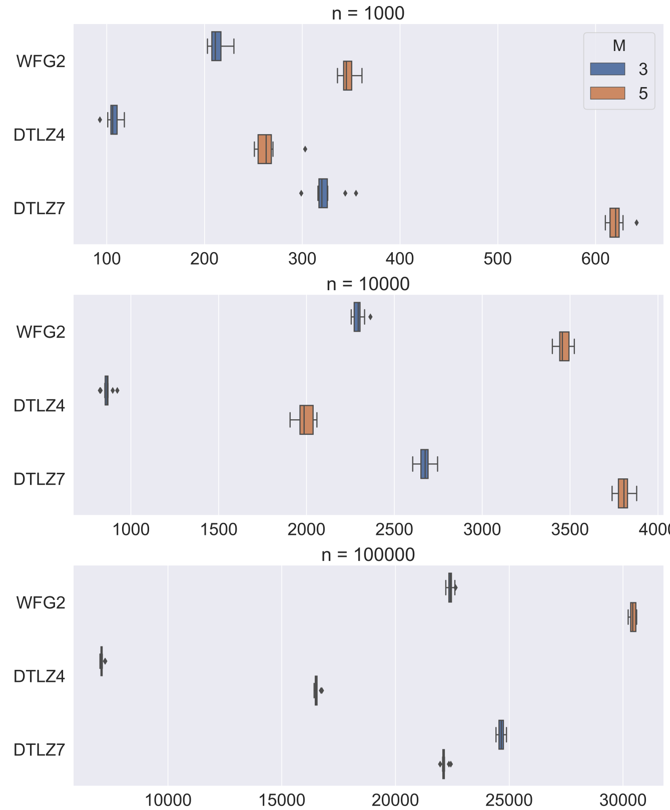

Fig. 2. Boxplots depicting reference front sizes (x axis) after dominance filtering for increasing numbers of randomly sampled solutions. From top to bottom: $n \in \{1\,000, 10\,000, 100\,000\}$. Boxplots are grouped by problem (y axis) and M (color).

IV. Randomly Sampled Reference Fronts

We start our analysis with the assessment of the reference fronts uniformly randomly sampled from the DTLZ and WFG benchmarks. The most relevant insights observed in this analysis can be grouped as M-dependent or not. We initially discuss the most relevant M-independent insight, namely the effect of dominance filtering. Next, we individually discuss the insights observed for the different M considered in this work.

A. Dominance Filtering

A preliminary dominance filtering analysis revealed that benchmark generators may provide solutions that are not Pareto-optimal. Hence, we start our analysis with an assessment of the percentage of solutions that get filtered out for the problems selected, given as boxplots in Figure 2. From top to bottom, boxplots depict filtered front sizes when reference fronts are uniformly randomly sampled with increasing n. For brevity, boxplots include only problems for which dominance filtering had an impact, namely problems that present multimodality (WFG2 and DTLZ7) or bias (DTLZ4).

Results show that a large extent of the solutions randomly sampled from benchmark generators for these problems are in fact dominated. For multi-modal problems, this is rather striking and reveals that solutions for which distance-related parameters are set to optimal can still be dominated. As discussed in Section II, this happens because sampling is done along the whole multi-modal surface that comprises the

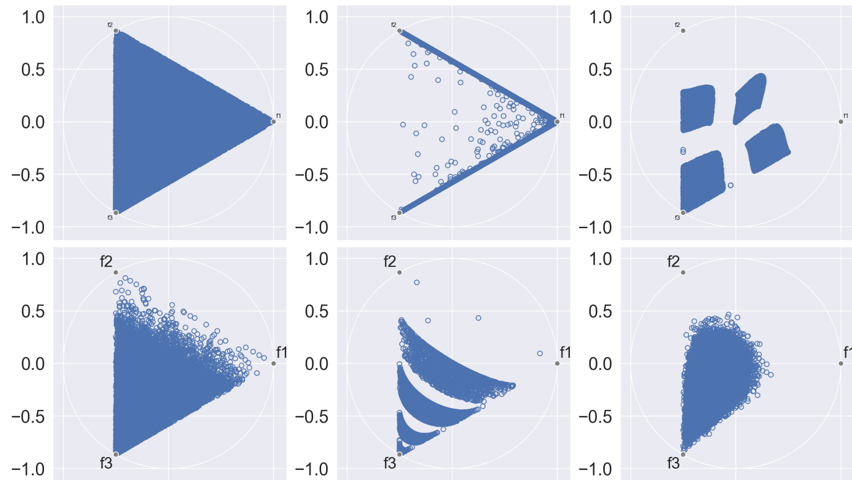

Fig. 3. RadViz plot of selected problems with three objectives and 50 variables for 100 000 randomly sampled points. Top, from left to right: DTLZ2, DTLZ4, and DTLZ7. Bottom, from left to right: WFG1, WFG2, WFG4. For brevity, a single seed is depicted, representative of the remaining.

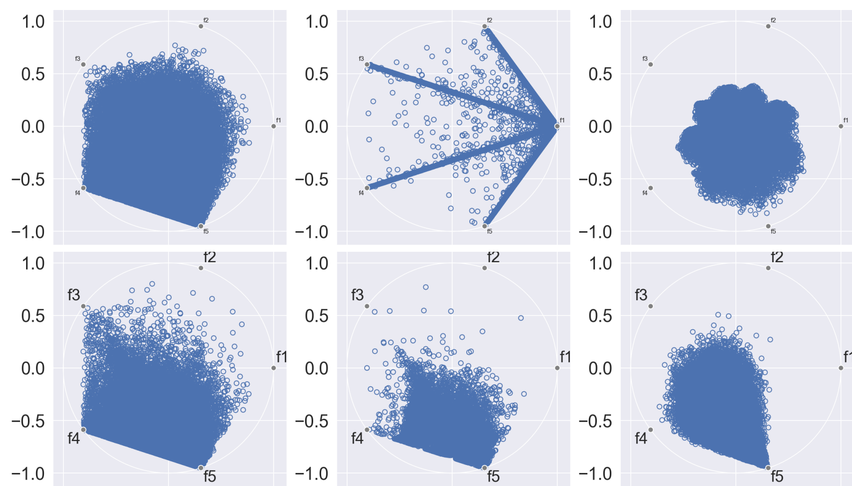

Fig. 4. RadViz plot of selected problems with five objectives and 50 variables for 100 000 randomly sampled points. Top, from left to right: DTLZ2, DTLZ4, and DTLZ7. Bottom, from left to right: WFG1, WFG2, WFG4. For brevity, a single seed is depicted, representative of the remaining.

objective function, and hence points in hills are dominated by points in valleys. For problems that present bias, solutions tend to be sampled from the same region of the objective space, increasing the chances of repeated solutions.It is also important to remark that bias is present in WFG1 to a lesser degree than in DTLZ4, and hence dominance filtering effects are not as evident for the former.

B. Results for $M = 3$

RadViz plots of reference fronts sampled with $n = 100\,000$ and $M = 3$ are given on Figure 3. For brevity, plots for remaining sizes are given as supplementary material, since they are unable to approximate well the actual Pareto fronts for any of the problems considered. Figure 3 shows that reference fronts for DTLZ problems (top row) are high-quality approximations of the true Pareto fronts for DTLZ2 (left) and DTLZ7 (right). As for DTLZ4 (center), solutions concentrate on the biased edges, whereas a high-quality approximation should look similar to the front found for DTLZ2.

Concerning WFG problems (bottom row), reference fronts appear biased even if the original problems are not, missing entire regions of the objective space. For instance, WFG1 (left) and WFG4 (right) should resemble the front for DTLZ2, but specially for WFG4 the sampled points present a very poor spread. Concerning WFG2 (center), increasing the number of sampled solutions from $n = 1\,000$ to $n = 100\,000$ produces fronts that find more solution pools, but an entire pool appears to still be missing.

C. Results for $M = 5$

Figure 4 depicts results for fronts with $n = 100\,000$ and $M = 5$. Once again, top row plots illustrate DTLZ fronts (respectively DTLZ2, DTLZ4, and DTLZ7, from left to right), whereas bottom row plots illustrate WFG fronts (respectively WFG1, WFG2, and WFG4, from left to right). A pentagon pattern is expected for DTLZ2, DTLZ4, WFG1, and WFG4. Yet, none of the fronts produced for these problems simultaneously present good spread and distribution.

Fig. 5. RadViz plot of selected problems with $M = 3$ and 50 variables, truncated using MGA of size 100. Top, from left to right: DTLZ2, DTLZ4, and DTLZ7. Bottom, from left to right: WFG1, WFG2, WFG4. For brevity, a single seed is depicted.

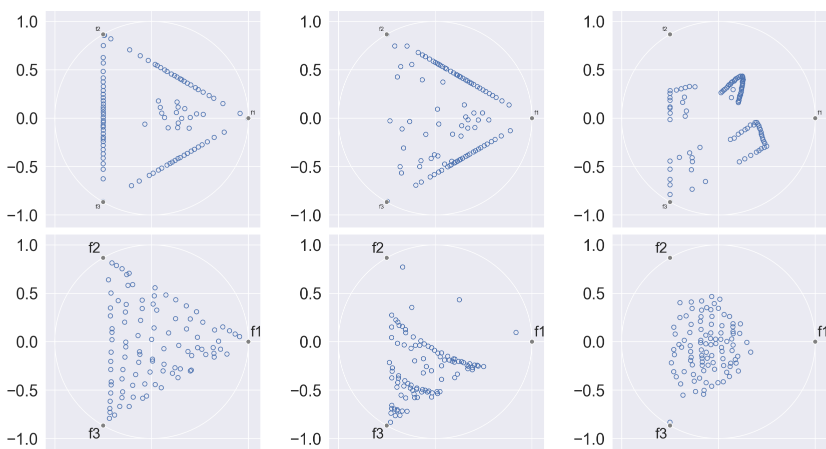

Fig. 6. RadViz plot of selected problems with $M = 3$ and 50 variables, truncated using AA_s of size 100. Top, from left to right: DTLZ2, DTLZ4, and DTLZ7. Bottom, from left to right: WFG1, WFG2, WFG4. For brevity, a single seed is depicted.

Concerning multi-modal problems (DTLZ7 and WFG2), we see quite different outcomes. For DTLZ7, RadViz limitations do not allow us to investigate the center region of the solution pools, but boundary regions indicate that distribution could still be improved in the region that appears closer to objective f_5. For WFG2, distribution is even worse, but concentrates around the edge formed between objectives f_4 and f_5. Finally, DTLZ4 presents the same pattern observed previously, with solutions concentrated on the biased edges.

D. Discussion

The assessment of the uniformly randomly sampled reference fronts revealed striking results. Besides the presence of solutions that were not Pareto-optimal (sometimes to a very large extent of the fronts produced), the spread and distribution for some DTLZ and all WFG problems are very poor already with three objectives. In fact, it is rather remarkable that such results have not yet been reported. More importantly, they reveal an urgent need for standardized reference fronts and for a method for generating them without the spread and/or distribution issues observed in this section.

Finally, we remark that results for $n < 100\,000$ proved even poorer than the results for $n = 100\,000$. In this context, the assessment we conduct in the next section on the effects of archiver truncation becomes even more relevant, as providing reference fronts with such a large number of solutions would be impractical for several reasons.

V. Archiver-Truncated Fronts

As previously discussed, too large reference sets would greatly increase the overhead for performance measure computation, specially in many-objective optimization. In this section, we assess fronts truncated by MGA and AA_s, grouping insights by the number of objectives considered. Yet, we remark that very often MGA was unable to process DTLZ4 fronts even with a week-long cutoff time. We conjecture that this long processing times are a factor of the dynamically adjusting grid, specially given solution concentration along the biased edges discussed in the previous section.

A. Results for $M = 3$

Figures 5–6 respectively show $M = 3$ fronts truncated by MGA and AA_s with maximum capacity 100 when provided $100\,000$ uniformly randomly sampled solutions. In common, results for DTLZ4, WFG1, and WFG4 change very little. Regarding the remaining problems, MGA generally presents

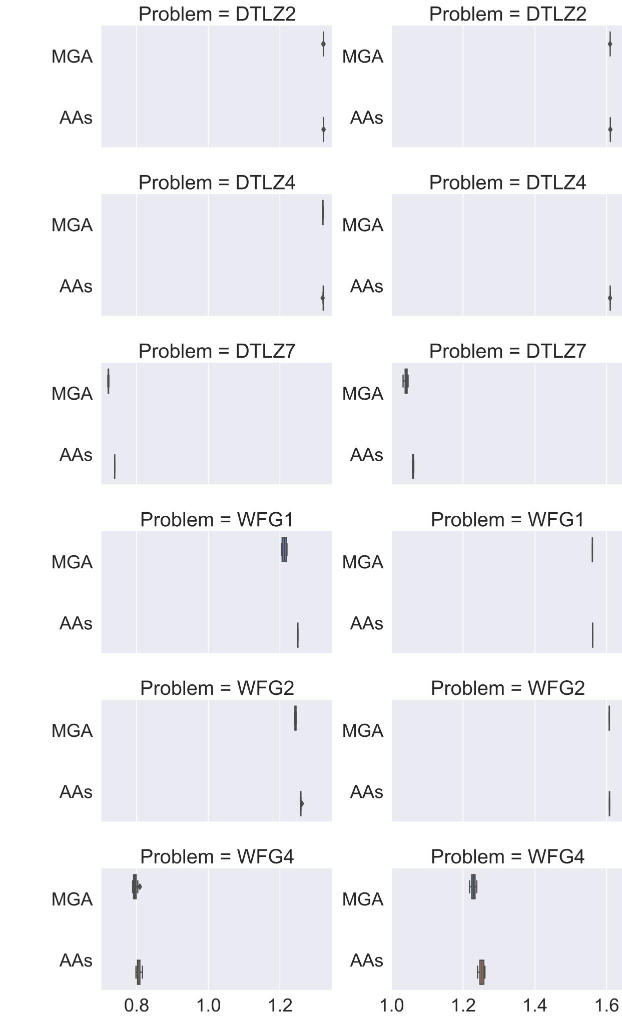

Fig. 7. Boxplots depicting hypervolume values (x axis) achieved by truncated reference fronts using different archivers (y axis). Left: $M = 3$; right: $M = 5$.

a better distribution than AA_s. Indeed, the poor distribution of the outputs from AA_s is often striking, but is an expected consequence of (i) the low-quality inputs provided, specially in the case of DTLZ4 and WFG2, and; (ii) how the hypervolume favors spread over distribution.

To complement the visual analysis, Figure 7 gives boxplots of the hypervolumes achieved by the truncated reference fronts (x-axis) using the different archivers (y-axis), grouped by benchmark problem (rows) and M (columns). Concerning the analysis for $M = 3$ problems (left-most column), the hypervolumes for fronts truncated by AA_s are always better than for fronts truncated by MGA, though differences for the concave DTLZ problems (DTLZ2 and DTLZ4) are negligible. Finally, we highlight that the only problem for which the hypervolumes are much smaller than expected is WFG4, reflecting the issues with the input reference fronts.

Having the hypervolume favoring the AA_s is an expected result, but it confirms the need for multiple performance indicators being taken into consideration when designing EMO algorithms and components already for three objectives. In more detail, from the distribution perspective the fronts truncated by MGA are better than the fronts truncated by AA_s; yet, an analysis based solely on the hypervolume misses this.

B. Results for $M = 5$

Figures 8–9 respectively show $M = 5$ fronts truncated by MGA and AA_s with maximum capacity $1\,000$ when provided $100\,000$ randomly sampled solutions. This time, differences are only observable for (i) DTLZ2, for which MGA favors the central region w.r.t. AA_s, and; (ii) DTLZ4, for which MGA is unable to finish processing any $n = 100\,000$ front. Overall, RadViz allows us to see differences when they are striking, but a more advanced approach such as PaletteViz would be required for a deeper contrast between MGA and AA_s.

Boxplots for $M = 5$ problems are given in the right-most column of Figure 7. Note that the ranges of the hypervolumes change as a function of M, but the patterns in both columns is remarkably similar. The only differences concern (i) DTLZ4, for which MGA is unable to process reference fronts, as discussed, and; (ii) the convex WFG problems (WFG1 and WFG2), for which the difference between archivers is now also negligible. In a sense, the poor inputs provided to the archivers renders the truncated fronts very similar, and so future work should address this issue from the perspective of better-produced reference fronts.

VI. Conclusion

Heuristic optimization algorithms directly depend on sound theoretical and experimental performance assessment methodology. In the context of multi- and many-objective optimization, the evolutionary multi-objective (EMO) community has been instrumental to this end. Nonetheless, the recent observations from many-objective optimization works indicate that even contributions that were deemed mature must be improved.

One such topic is Pareto-optimal reference front generation, generally overlooked in EMO assessment as evidenced by the absence of standard reference fronts in the community. This work built on existing investigations that indicated that Pareto-optimal reference fronts in the context of continuous optimization present room for improvement. Leveraging multidimensional visualization and solution archiving, we have demonstrated that the underlying issues with randomly sampled reference fronts from benchmark generators are even more serious than anticipated. Concretely, the presence of dominated solutions and poor spread and distribution are observed even when a large number of solutions is produced and already for three-objective problems.

Future work should urgently address these issues. Specifically, artificial benchmark sets became popular precisely for the possibility of Pareto-optimal reference fronts. More importantly, binary Pareto-compliant measures can produce absolutely biased conclusions if the reference fronts adopted are not high-quality. Furthermore, methods produced for reference front generation should value the configurable nature of benchmark sets, allowing to explore the effects of increasing number of variables, the ratio between distance- and position-related variables, and other eventual configurable parameters.

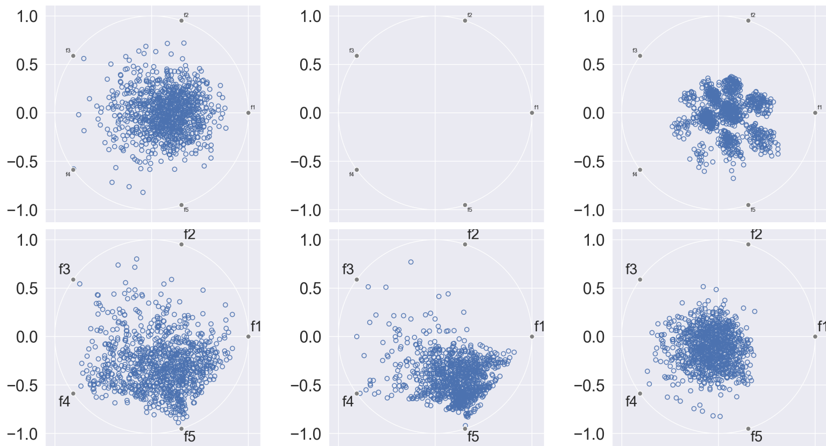

Fig. 8. RadViz plot of selected problems with $M = 5$ and 50 variables, truncated using MGA of size 100. Top, from left to right: DTLZ2, DTLZ4, and DTLZ7. Bottom, from left to right: WFG1, WFG2, WFG4. For brevity, a single seed is depicted.

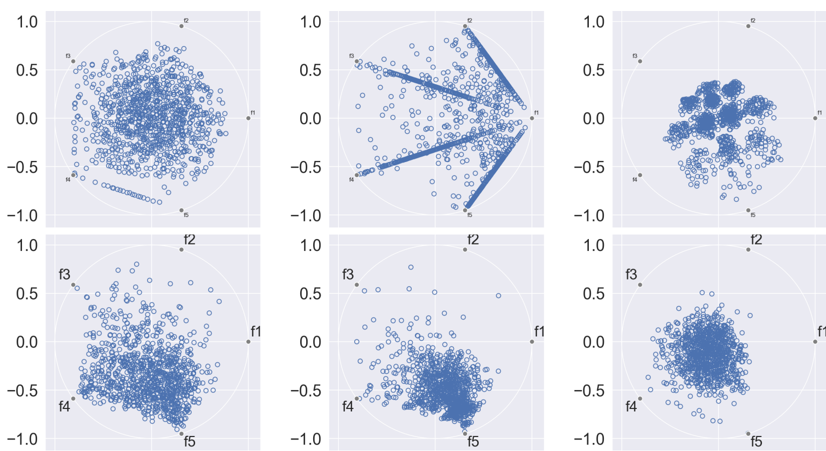

Fig. 9. RadViz plot of selected problems with $M = 5$ and 50 variables, truncated using AA_s of size 100. Top, from left to right: DTLZ2, DTLZ4, and DTLZ7. Bottom, from left to right: WFG1, WFG2, WFG4. For brevity, a single seed is depicted.

ACKNOWLEDGEMENT

T. Stützle acknowledges support from the Belgian F.R.S.-FNRS, of which he is a research director.

REFERENCES

[1] M. Fleischer. “The Measure of Pareto Optima. Applications to Multiobjective Metaheuristics,”. EMO, LNCS, pp. 519–533, Springer, 2003.

[2] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. “Scalable Test Problems for Evolutionary Multi-Objective Optimization, ” in EMO, LNCS, pp. 105-145, Springer, 2005.

[3] S. Huband, L. Barone, L. While, and P. Hingston. “A Scalable Multi-objective Test Problem Toolkit,”. EMO, LNCS, v. 3410. pp. 280-295, Springer, 2005.

[4] S. Jiang, Y. Ong, J. Zhang and L. Feng, “Consistencies and Contradictions of Performance Metrics in Multiobjective Optimization,” in IEEE Trans. on Cybern., vol. 44, no. 12, pp. 2391-2404, Dec. 2014.

[5] L. C. T. Bezerra, M. López-Ibàñez, M., and T. Stützle. “A large-scale experimental evaluation of high performing multi- and many-objective evolutionary algorithms, ”in Evol. Comput., 26(4), pp. 621–656, 2020.

[6] H. Ishibuchi, H. Masuda, and Y. Nojima. “Pareto fronts of many-objective degenerate test problems, ” in IEEE Trans. Evol. Comput., 20(5), pp. 807–813, 2016.

[7] L. C. T. Bezerra, M. López-Ibàñez, M., and T. Stützle. “Archiver Effects on the Performance of State-of-the-Art Multi- and Many-Objective Evolutionary Algorithms,” in GECCO, pp. 620–628, 2019.

[8] P. Hoffman, G. Grinstein, K. Marx , and I. Grosse. “DNA Visual And Analytic Data Mining,” in Proc. IEEE Vis. Conf., pp. 437–441, 1997.

[9] M. Laumanns and R. Zenklusen. “Stochastic convergence of random search to fixed size Pareto set approximations,” in Eur. J. of Oper. Res., 213(2), p. 414-421, 2011.

[10] J. D. Knowles. “Local-Search and Hybrid Evolutionary Algorithms for Pareto Optimization, ” in Ph.D. thesis, University of Reading, UK, 2002.

[11] I. Das and J. E. Dennis. “Normal-Boundary Intersection: A New Method for Generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems,” in SIAM J. on Optim., 8(3), p. 631-657, 1998.

[12] M. López-Ibàñez, J. Knowles, and M. Laumanns. “On Sequential Online Archiving of Objective Vectors,” EMO, LNCS, v. 6576, pp. 44-60, 2011.

[13] T. Tušar and B. Filipič. “Visualization of Pareto Front Approximations in Evolutionary Multiobjective Optimization: A Critical Review and the Prosection Method,” in IEEE Trans. on Evol. Comput., 19(2), pp. 225-245, 2015.

[14] J.J. Valdes and A. Barton, “Visualizing high dimensional objective spaces for multi-objective optimization: A virtual reality approach,” in IEEE CEC, pp. 4199-4206, 2007.

[15] Z. He and G. G. Yen. “Visualization and performance metric in many-objective optimization, ” in IEEE Trans. Evol. Comput., 20(3), pp. 386-402, 2016.

[16] A. K. A. Talukder and K. Deb. “PaletteViz: A Visualization Method for Functional Understanding of High-Dimensional Pareto-Optimal Data-Sets to Aid Multi-Criteria Decision Making,” in IEEE Comput. Intell. Mag., 15(2), pp. 36-48, 2020.

[17] G. C. da Silva, E. F. Wanner, L. C. T. Bezerra, and T. Stützle. “Supplementary material for Revisiting Pareto-optimal multi- and many-objective reference fronts for continuous optimization”, in https://gabicavalcante.github.io/multimanyreferencefronts, 2020.