Objective Dimension and Problem Structure in Multiobjective Optimization Problems

Ramprasad Joshi(✉), Bharat Deshpande, and Paritosh Gote

BITS, Pilani - K K Birla Goa Campus, Zuarinagar 403726, Goa, India
{rsj,bmd}@goa.bits-pilani.ac.in, paritosh.gote@gmail.com

Abstract. Multiobjective optimization seeks simultaneous minimization of multiple scalar functions on $\mathbb{R}^n$. Unless weighted sums are made to replace the vector functions arising thus, such an optimization requires some partial- or quasi-ordering of points in the search space based on comparisons between the values attained by the functions to be optimized at those points. Many such orders can be defined, and search-based (mainly heuristic) optimization algorithms make use of such orders implicitly or explicitly for refining and accelerating search. In this work, such relations are studied by modeling them as graphs. Information apparent in the structure of such graphs is studied in the form of degree distribution. It is found that when the objective dimension grows, the degree distribution tends to follow a power-law. This can be a new beginning in the study of escalation of hardness of problems with dimension, as also a basis for designing new heuristics.

1 Introduction

Multiobjective optimization requires various nontrivial choices of the algorithm designer as well as solution deployer. Acceptable solution criteria themselves are subject to complicated choices affecting many other decisions down the line. Design of evolutionary algorithms for multiobjective optimization involves choosing the search heuristic, designing appropriate representation, designing appropriate variation operators, defining ordering relations, designing selection strategies, and possibly designing adaptation among one or several of all these parameters. Because of the complexity of these choices and designs, and because slight variations in them can produce widely varying behaviours and performances, analysing problem hardness or even defining a problem hardness notion that is not dependent on the semantics and the intuition behind algorithm designs has been a vexatious exercise. We argue here that making the geometric intuition that usually underlies algorithm designs also the basis of analysing problem structure will go a long way in the prediction of problem hardness with respect to specific design primitives in algorithms. For this purpose, studying the partial orders induced by the geometry of problems, and applying probability measure theory, can be a starting point.

A.I. Esparcia-Alcázar et al. (Eds.): EvoApplications 2014, LNCS 8602, pp. 639–650, 2014.
DOI: 10.1007/978-3-662-45523-4_52

1.1 Moraglio et al.'s Geometric View of Variation Operators

Moraglio et al., in a series of works (e.g.[4],[3],[6]) have investigated and established a geometric-topological view of the search performed by evolutionary and other population-based heuristic algorithms. They unify the heuristic ideas behind the varied designs of variation (mutation and crossover) operators, and demonstrate that (most) evolutionary algorithms perform convex search, convex in the geometry induced by the neighbourhood structure and metric imposed on the search space by the algorithms' operators.

1.2 A Similar View for Selection Operators

In order to combine such a unified and powerful framework with analyses of problems so that a composite theory of evolutionary computation (of algorithms and problems) can be developed, we propose to use probability measure theory on the spaces of partial- and quasi-order relations that are imposed by the comparison operators and used by the selection operators of evolutionary and other heuristic algorithms. We unify the discrete nature of computed sequences and orders of sampled search points with probability measure theory on the most general continuous search spaces using simple graph models.

The rest of the paper is organized as follows. In Section 2 the basic definitions and their motivation are discussed. Subsequent Section 3 develops elementary tools of analysis of problem structure, especially from probability theory. Section 4 follows up on this development to make a conjecture, which is substantiated by computational experiments described in Section 5. Conclusions (in Section 6) sum up the paper.

1.3 Discovery of the Power Law

Power law distributions (see Clauset et al.[1]) arise in many natural as well as social mass processes, such as the World Wide Web. Among other things, a power law distribution over degrees in a graph indicate a certain scale-freeness[5]. Below (Section 5) we provide evidence that initial populations for optimization problems of many objectives tend to have a power-law distribution over the counts of points dominated by each point, indicating that variation and selection operators that depend on dominance relationships among individuals in the population (e.g. tournament selection) will not be able to distinguish between different solutions and identify niche areas. We investigated only the initial population genrated by uniform random sampling, but it opens up a new way of examining the properties of graphs arising in an optimization by heuristic search process induced by the dominance relationships and following various generative distributions, thereby providing useful information about the hardness of a problem or about tunability of algorithm performance.

2 The Structure of the Explored Search Space

Heuristic (including evolutionary) as well as classical (Newton-like) algorithms explore the search space in an iterative manner: beginning with some initial set of points, they try to figure out, in either geometric, or algebraic, or analytical manner, the next set of points which potentially may be better in the previous set. Similar to the geometric-topological view of Moraglio et al. we here look at the informative structure contained in the explored set of points (either the set under consideration in one iteration or all the points explored till some iteration) by examining the structure of the (transitive) graph that represents the transitive partial order on these points obtained by a strict dominance relation.

2.1 The Search Space

For simplicity, we take a closed bounded Euclidean space $\mathbb{X} \subsetneq \mathbb{R}^n, n \geq 1$ as the *search* space, and a bounded continuous function $f : \mathbb{X} \mapsto \mathbb{R}^m, m \geq 1$ as the multiobjective optimization (minimization) problem. We call $f(\mathbb{X})$ the *objective* space.

2.2 The Partial Order

The partial order we consider is $\prec \subsetneq \mathbb{X} \times \mathbb{X} : x \prec y \Leftrightarrow f_i(x) < f_i(y), i \in \{1, 2, \ldots, m\}$. It is obvious that $\prec$ is a transitive, irreflexive, antisymmetric relation. The transitivity is important to our analysis, in a practical way: it renders making graphs and computing their properties easier. However, it does not take away much of generality: the usual dominance relation that is used extensively in EC literature $x \preccurlyeq y \Leftrightarrow (\forall i \in \{1, 2, \ldots, m\} f_i(x) \leq f_i(y)) \wedge (\exists i \in \{1, 2, \ldots, m\} f_i(x) < f_i(y))$ considers for any given x additionally (to our $\prec$) only a null set of points $y, x \preccurlyeq y$ that has measure 0 as long as the measure is absolutely continuous with respect to the Lebesgue measure. Moreover, in practical floating-point calculations, strict equality comparison does not yield more accurate resutls; it can be counterproductive on the contrary.

2.3 The Graph

We consider the simple directed acyclic graph $G = (V, E)$, for $V \subset \mathbb{X}, |V| < \infty$, induced by the $\prec$ relation: $\forall x \neq y \in V, (x, y) \in E \Leftrightarrow x \prec y$. There are no self loops in G because $\prec$ is irreflexive. G is transitively closed. Such a graph is depicted in figure 1, in which the points are numbered. Thus, in the figure,

$$1 \prec 5, 1 \prec 6; 2 \prec 7, 2 \prec 8 \prec 12 \prec 15; 3 \prec 9 \prec 13, 3 \prec 10 \prec 14; 4 \prec 10 \prec 14, 4 \prec 11.$$

Inset is a possible scenario in a 2-D objective space that can give rise to this graph partially.

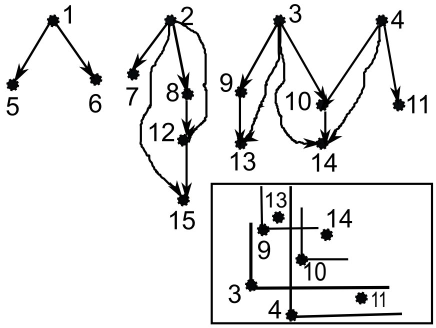


Fig. 1. A Transitive Graph induced by $\prec$

2.4 The Properties of $\prec$ and G Relevant to the Search Space Structure

Any EA (or, many other population-based heuristics too) will make decisions (viz. parental selection, survival selection, variation operators' specific geometry) based on some dominance relationships among the set of points under consideration in one iteration (e.g. a population in a generation in an EA run). "Differentiation" among the population in objective space is a major theme in EA design and performance, as also nearness or similarity between points in the search space. "Locating the pareto-optimal front" means identifying the nearness criteria among the population that lead to differentiation (towards more dominance) of similar or near points from the rest of the search space. For differentiation, one of the criteria used is the "hypervolume", or a rough estimate of the measure $\nu_{\prec}(x) = \nu(H_x = \{y \in \mathbb{X} : x \prec y\})$ of points H_x dominated by each point x where ν is some volumetric measure, usually taken to be the usual Euclidean volume (Lebesgue measure). Although densities in the search and objective space can be quite different, though related, and can be nonuniform throughout the objective space, the main technique used is taking a reference point in $\mathbb{R}^m$, not necessarily in $f(\mathbb{X})$, and take the Euclidean volume (the Lebesgue measure) of the hypercube defined by the two corners, one the image $f(x)$ of the point x, and the other the reference point, as the hypervolume $hyp(x)$ dominated by x. It is obvious that the $\prec$ relation respects hypervolumes (for a suitable reference point not in the interior of the objective space):

$$H_x \supsetneq H_y \Leftrightarrow x \prec y \Rightarrow hyp(x) > hyp(y). \tag{1}$$

However, we must also take into consideration the fact that algorithms in practice tend to calculate hypervolumes in the objective space. Densities in search and objective spaces can be quite different. We need to take up the question of the relation between the graph G and hypervolumes as calculated by the existing algorithms. Observe that hypervolumes are computed using a reference point in such a way that the putatively dominated sets contain the actual dominated sets, most often properly. The reference point itself must be dominated by every point that dominates anything else in $f(\mathbb{X})$. In other words, the reference point must be well-nigh "high above and outside" the objective space $f(\mathbb{X})$. This makes hypervolume calculations easier. Therefore, in this setting $\forall x \in \mathbb{X}, hyp(x) \geq \nu(H_x)$ where ν is the Lebesgue measure normalized over the search space. Still, because inclusion is equivalent to $\prec$ relation for the dominated sets, and the reference point does not change that inclusion in the first part of (1), the implication in its second part must hold too. Now we can take the $\prec$ relation and its induced graph to represent information that practical heuristic algorithms in EC use for decisions, whether based on ranking or ordering, or based on hypervolumes. For maintaining rigour, however, we confine the discussion to the actual measures of dominated sets.

From the foregoing, it is clear that the hypervolume dominated by a given graph G induced by $\prec$ over f over $\mathbb{X}$, is the total hypervolume (of a union discounting intersections) dominated by the nondominated points, or the vertices of G that have indegree 0. Such points are easily identified by a depth-first (DFS) traversal of G, that follows a directed path in the graph until a potential cycle or a dead-end is visited, restarting at unvisited vertices and down unexplored paths. Such a traversal results in a set of (possibly several, disjoint) tree, in which cycle-forming edges are omitted, and intersecting paths are explored short of the intersecting edge. This set is called a DFS-forest. The nondominated points in a given V of $G = (V, E)$ will be the roots of the trees in this forest. In figure 1, vertices 1,2,3,4 are the roots of DFS trees in and DFS run (regardless of the sequence of vertices taken). If the usual order of natural numbers is taken, then such a DFS run on this graph will yield a DFS forest that is the whole graph except the edge $(4 \rightarrow 10)$.

Let us call the paths in G that go across two disjoint trees in this DFS-forest as *bridges*. Thus in figure 1, the edge $(4 \rightarrow 10)$ is the only bridge. If there are too many bridges in G itself, then the sets dominated by the nondominated points are also intersecting too often. When the bridges are near the roots of the DFS-forest trees, the intersection sets are large too. It can be seen now that the more the disjoint paths in G, the closer (from below) is the total hypervolume dominated by G to the simple sum of the hypervolumes dominated by the nondominated points, because there will be fewer and smaller intersections among the dominated sets. Of course, there is the possibility that the chosen V is such that intersections among the dominated sets are not reflected in the intersecting paths. How is V to be chosen such that this probability is negligible? We address this question in Section 3.

3 Choosing V to Minimize Intersection Without Bridges

The simplest scheme to choose V so that there is a fair correspondence between the number of bridges and intersections of dominated sets is to choose it uniformly randomly. Our next simple proposition tells that the proportion of vertices in V sampled uniformly from any closed connected set of nonzero measure in $\mathbb{X}$ is sharply concentrated around its measure by the uniform probability measure.

Proposition 1. *Let $X \subset \mathbb{R}^n$ be a bounded measurable set and ν be the Lebesgue measure normalized on and restricted to X, such that $\nu(X) = 1$. If V is sampled uniformly at random from X, with $|V|= q < \infty$, then for $Y \subset X$ Borel,*

$$\mathbb{P}[|\{v \in V \cap Y\}| \geq q\nu(Y) + t] \leq \exp\left(\frac{-t^2}{2(q\nu(Y) + t/3)}\right)$$

and

$$\mathbb{P}[|\{v \in V \cap Y\}| \leq q\nu(Y) - t] \leq \exp\left(\frac{-t^2}{2(q\nu(Y))}\right).$$

Proof: Observe that when sampled uniformly, $|\{v \in V \cap Y\}|$ is a binomial random variable that is the sum of the Bernoulli trials over $\mathbf{1}_Y$ with

$$p = \mathbb{P}[x \in Y] = \nu(Y);\ \mathbb{P}[x \notin Y] = 1 - p.$$

So $\mathbb{E}[|\{v \in V \cap Y\}|] = qp = q\nu(Y)$. The result follows from the direct application of Chernoff bounds. □

Let $u, v \in V$ be two uniformly randomly chosen points, let $\nu(H_u \cap H_v) = h$ and let the number of points in V that are descendents of u, v both be denoted by the random variable N. That means $|\{w \in V \cap H_u \cap H_v\}|= N$. Then by Proposition 1, with $|V|= q$, if $h \neq 0$, $\mathbb{P}[N = 0] \leq e^{\frac{-qh}{2}}$. This precisely is an upper bound on the probability of an intersection among dominated sets not being represented by any bridge in the graph G; and this is tight (upto multiplicative fractional constants) by Chernoff bounds. For a large graph, this rapidly diminishes. Hence we can conclude that

Proposition 2. *When there is no bridge in G between DFS trees rooted in two vertices $u, v \in V$, then $H_u \cap H_v = \Phi$, with a high probability $\geq 1 - \epsilon$, wherein $\epsilon \downarrow 0$ as $q \uparrow \infty$.* □

4 Degree Distribution in a Graph with No Bridges

In each DFS tree (on the graph G obtained as in Section 3 above) containing q_r vertices, the (out-)degrees (in G) of the vertices in the tree are distributed as follows. For each degree in $\{0, 1, \ldots, q_r - 1\}$, the number of vertices with that degree diminishes as the degree rises. With degree $q_r - 1$, there is exactly 1 vertex in the tree, and if there are no bridges, then there is exactly one vertex of degree $q_r - 1$ in G for each DFS tree with q_r vertices. Take $r \in [0, 1]$ and $q_r = qr$. Suppose $r_1, r_2, \ldots, r_k \in [0, 1]$ are the fractions associated with all the k DFS trees of sizes qr_1 etc. in the bridgeless graph G. Then $\sum_i r_i = 1$, and each $r_i = h_i \pm \frac{t_i}{q}$ for some small t_i, where $h_i = \nu(H_{x_i})$, x_i the root of the i^{th} tree. Thus $\sum_i (h_i \pm \frac{t_i}{q}) = 1$. By Proposition 1, the set dominated by the root x_i of a q_{r_i}-tree has this bound:

$$\mathbb{P}[qh_i - t_i < qr_i - 1 < qh_i + t_i] \geq 1 - \exp\left(\frac{-t_i^2}{2(qh_i + t_i/3)}\right) - exp\left(\frac{-t_i^2}{2qh_i}\right).$$

Rearranging and simplifying the inequalities, we get

$$\mathbb{P}\left[r_i - \frac{1+t_i}{q} < h_i < r_i - \frac{1-t_i}{q}\right] \geq 1 - 2\exp\left(\frac{-t_i^2}{2(qh_i + t_i/3)}\right).$$

Now, if k is large and r_i not varying much, then each h_i has to be small. But then the lack of bridges means that the corresponding H_{x_i} are all pairwise disjoint and cover most of the search space, which, with each h_i small, is possible only if the overwhelming majority of x_i lie on the Pareto-optimal front, and their images are as distant as possible in the objective space. This argument needs to be made more rigorous, but we are justified here in claiming that

Conjecture 1. A bridgeless forest of a large number of trees of sizes that do not vary much indicates that a good approximation to the Pareto-optimal front is contained in it.

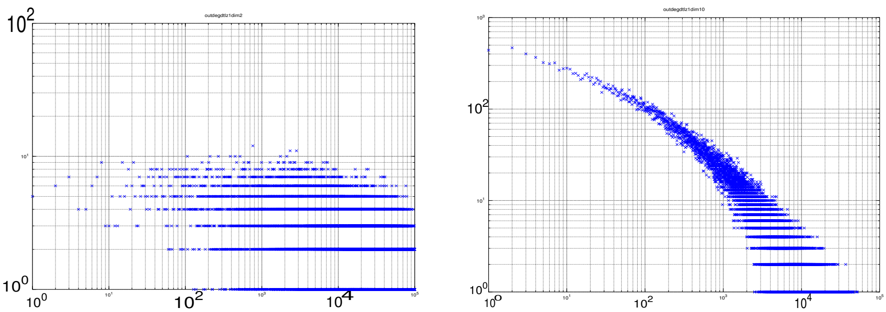
For an initial graph generated by uniform random sampling, this occurrence is highly unlikely for large m. But for small m, this is plausible. In case of large m, we can expect the more likely scenarios of a large variation in r_i, over small or large k, with a small or large number of bridges. Then the more the bridges, the more uniform is the degree distribution. As the dimension m grows, the variation in r_i will be larger, k larger, and the number of bridges smaller. Trees with large r_i will be less in proportion, and vice versa. Progressively this should lead to a situation that sees a rapid decrease in the number of trees with large size, hence a rapid decrease in the number of vertices with large degree. One would suspect a power-law distribution lurking here. In computational experiments on the DTLZ suite, we found this to be the case. We take a look at those results in Section 5. Note that the discussion of the out-degree distribution carries over with little change to in-degree distribution. Our computational results too confirm this, though we have omitted the graphs due to space constraints here.

5 Computational Experiments and Results

For the DTLZ suite of scalable test problems[2], we generated the graphs G as described in Section 3 above, for 30,000 points chosen uniformly randomly, for each objective dimension 2 through 10. The degree distributions were plotted in a log-log graph to see if power-law behaviour is apparent, which was found to be the case. The out-degree distribution graphs for dimensions 2 and 10, for problems DTLZ1, DTLZ2, DTLZ3, DTLZ4 are shown below (figures 2-5). In each graph, on the x-axis is the logarithm of the out-degree counts (strictly speaking an offset of 1 added, in order to avoid logarithm of 0), going from 10^0 through $10^{4.48}$ for out-degree counts going from 0 to 30,000. The almost-straight line of slope -1 except for DTLZ4 shows the power-law behaviour. The DTLZ4 exception needs explanation, which follows in the next paragraph. It is noteworthy that the graphs gradually take the power-law shapes as the objective dimension grows, though we cannot show all the graphs here. The programs used to carry out this data generation and analysis and the generated graphs are all available with the first author.

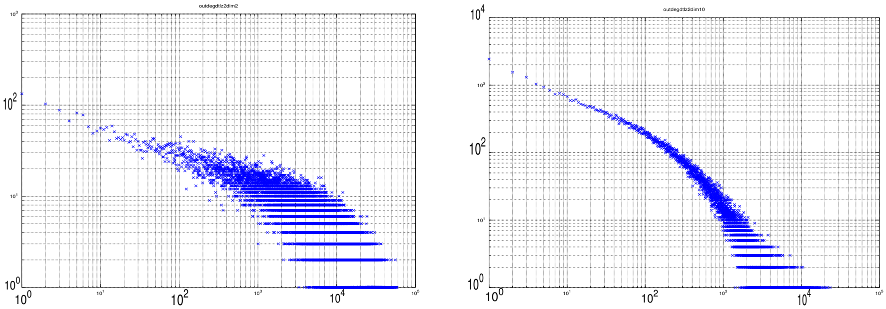
In their original paper describing the design of the DTLZ test problems, Deb et al.[2] have explained the goal in the design of DTLZ4 as testing "an MOEA's ability to maintain a good distribution of solutions", resulting in a modification of DTLZ2 that allows "a dense set of solutions to exist near" the plane of intersection of two dimensions in the objective space. This requires good diversity in the initial population itself, and therefore the performance of an MOEA on DTLZ4 in terms of quality of solutions depends sensitively on many parameters chosen at the time of a run. In a product measure absolutely continuous with the Lebesgue measure on the Euclidean space, the measure of points in this intersection region will be null because of the mapping, but their inverse image will be non-null. This will affect the degree distribution in a unique way, because a dense set in the search space will be a set of mutually indifferent points. In figure 6, the dimension 10 in-degree distribution is shown in a similar log-log plot for DTLZ2 and DTLZ4. DTLZ4 is a variation on DTLZ2, and the outlier in the plot for DTLZ4 (near [30,000,8]) shows the effect of the variation, due to the dense set of solutions depicted here, seen in the right-bottom corner.

The specificity of problem structure is even more apparent in the degree-difference graphs shown in the figures 7-10. Here the difference is out-degree minus in-degree. The difference (-30,000 to +30,000) is plotted on the x-axis, and the counts of points having that difference between their out- and in-degrees are plotted on the y-axis. These are **not** log-log plots, and the sharp concentration around the 0 difference is very obvious for the higher dimension. What is remarkable is that DTLZ4 has this concentration even more prominent, and in both lower and higher dimensions.



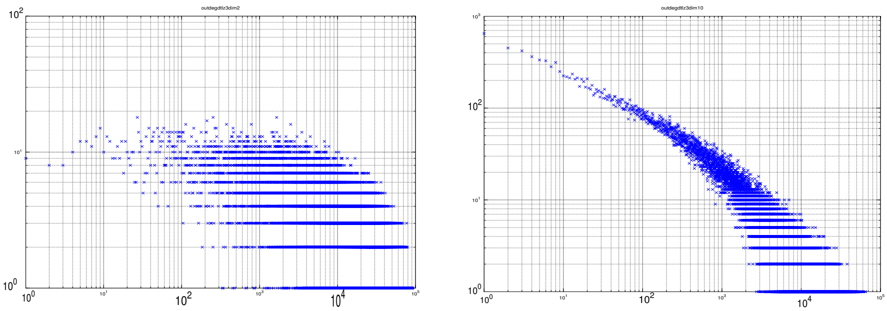
(a) Dimension 2 **(b)** Dimension 10

Fig. 2. DTLZ1, Out-degree distribution



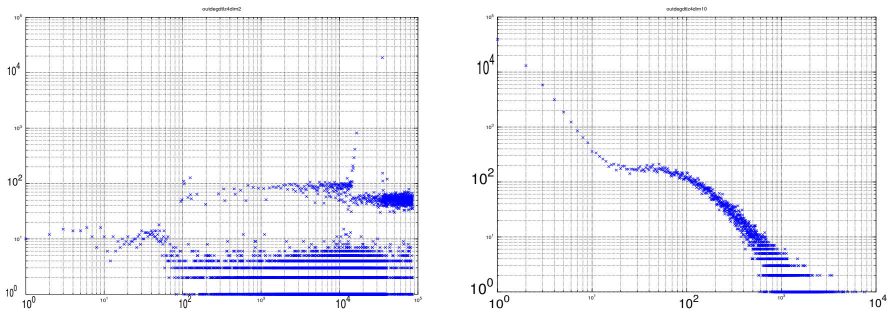
(a) Dimension 2 **(b)** Dimension 10

Fig. 3. DTLZ2, Out-degree distribution



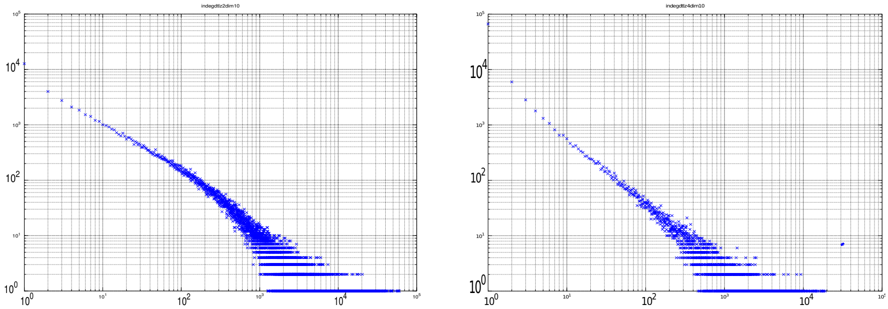
(a) Dimension 2 **(b)** Dimension 10

Fig. 4. DTLZ3, Out-degree distribution



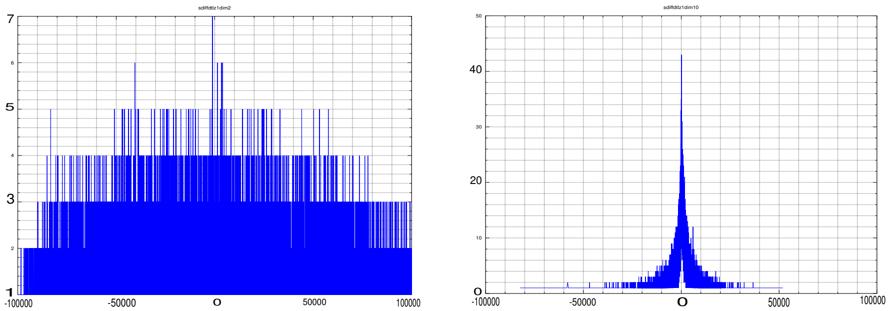
(a) Dimension 2 **(b)** Dimension 10

Fig. 5. DTLZ4, Out-degree distribution



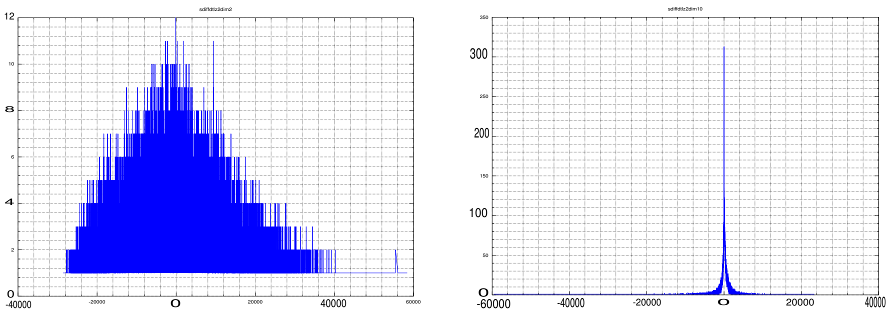
(a) DTLZ2, Dimension 10 **(b)** DTLZ4, Dimension 10

Fig. 6. DTLZ2 and DTLZ4, In-degree distribution



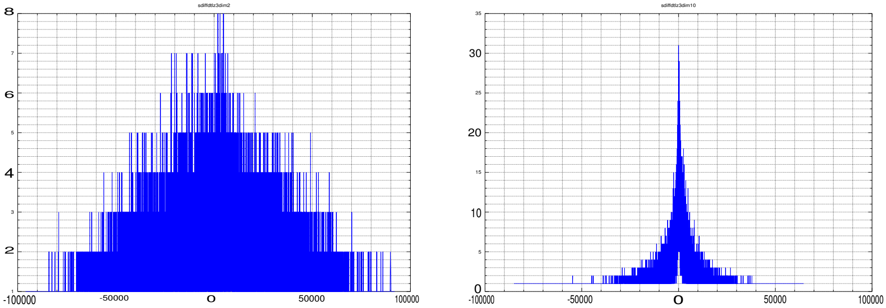
(a) Dimension 2 **(b)** Dimension 10

Fig. 7. DTLZ1, Degree-difference distribution



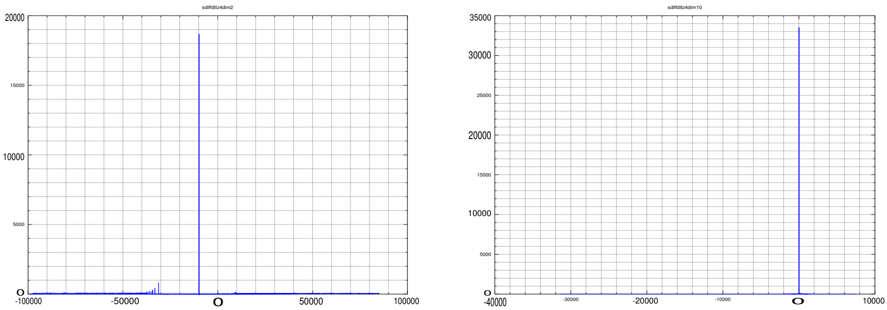
(a) Dimension 2 **(b)** Dimension 10

Fig. 8. DTLZ2, Degree-difference distribution



(a) Dimension 2 **(b)** Dimension 10

Fig. 9. DTLZ3, Degree-difference distribution



(a) Dimension 2 **(b)** Dimension 10

Fig. 10. DTLZ4, Degree-difference distribution

6 Conclusions and Future Work

In the graphs, it is apparent that specific problem strucure becomes progressively less important in the degree distribution as the dimension grows. The more the generated points, say 100,000, the more the behaviour is sharply tending towards power-law distribution as the dimension grows. This is not shown here yet. However, even in this there is a variation seen between DTLZ1,2,3 on the one hand and DTLZ4 on the other. This can be a starting point in separating problem-specific and class-general features of test problems and suites.

Examining graphs arising in a similar way but on populations generated by sampling other distributions than the uniform will be a new direction in analysing EA behaviour. Computational experiments with existing MOEAs and analytical framework for a family of distributions progressively sampled by such algorithms will open up the possibility of a new perspective on problem hardness and algorithm performance. The insights thus obtained can be useful in tuning algorithms by assessing the performance during a run.

This is an ongoing work, in which the present paper serves only as a proof-of-concept. Rigorous analysis of the conditions necessary and/or sufficient for obtaining various distributions in the degrees of the partial-order graphs is ongoing, in which other aggregate properties of the graphs are also being considered. For various orders, generated under various conditions such as adaptive or fixed sampling distributions for choosing points, which aggregate properties are preserved will be an interesting question for investigation. Combining this direction of work with Moraglio et al.'s work on algorithms ought to be the main goal in the long run.

References

1. Clauset, A., Shalizi, C.R., Newman, M.E.J.: Power-law distributions in empirical data. CoRR abs/0706.1062v2 (February 2009)
2. Deb, K., Thiele, L., Laumanns, M., Ziztler, E.: Scalable multi-objective optmization test problems. In: Proceedings of the 2002 Congress on Evolutionary Computation, vol. 1, pp. 825–830. IEEE (2002)
3. Moraglio, A.: Abstract convex evolutionary search. In: FOGA. pp. 151–162 (2011)
4. Moraglio, A.: Geometry of evolutionary algorithms. In: GECCO (Companion), pp. 1317–1344 (2012)
5. Newman, M.E.J.: The structure and function of complex networks. CoRR abs/cond-mat/0303516v1 (March 2003)
6. Yoon, Y., Kim, Y.H., Moraglio, A., Moon, B.R.: A mathematical unification of geometric crossovers defined on phenotype space. CoRR abs/0907.3200 (2009)