ACADEMIA
Accelerating the world's research.

An investigation on preference order ranking scheme for multiobjective evolutionary optimization

Dragan Savic

... , IEEE Transactions on

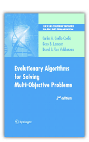
Cite this paper

Downloaded from Academia.edu

Get the citation in MLA, APA, or Chicago styles

Related papers

Download a PDF Pack of the best related papers



Evolutionary Algorithms for Solving Multi-Objective Problems
Vahid FARYAD

A Grid-Based Evolutionary Algorithm for Many-Objective Optimization
Shengxiang Yang

Incorporating the Notion of Relative Importance of Objectives in Evolutionary Multiobjective Optimiza…
Anis Rachmawati

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/3418906

An Investigation on Preference Order Ranking Scheme for Multiobjective Evolutionary Optimization

Article *in* IEEE Transactions on Evolutionary Computation · March 2007

DOI: 10.1109/TEVC.2006.876362 · Source: IEEE Xplore

CITATIONS
115

READS
140

3 authors, including:

Soon-Thiam Khu
University of Surrey

75 PUBLICATIONS **1,639** CITATIONS

SEE PROFILE

Dragan Savic
University of Exeter

423 PUBLICATIONS **6,874** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Saraswati View project

Project 2014-2017 SINATRA (Susceptibility of catchments to INTense RAinfall and flooding) View project

All content following this page was uploaded by Dragan Savic on 13 October 2013.

The user has requested enhancement of the downloaded file. All in-text references underlined in blue are added to the original document and are linked to publications on ResearchGate, letting you access and read them immediately.

An Investigation on Preference Order ï Ranking Scheme for Multi Objective Evolutionary Optimisation

Francesco di Pierro, Soon-Thiam Khu* and Dragan A. Savić

Centre for Water Systems, University of Exeter, North Park Road, Exeter, EX4 4QF, Devon, United Kingdom.
Email: {F.di-Pierro; S-T.Khu; D.Savic}@exeter.ac.uk
* Corresponding author.

ABSTRACT

It may be generalized that all Evolutionary Algorithms (EA) draw their strength from two sources: exploration and exploitation. Surprisingly, within the context of Multi-Objective (MO) optimization, the impact of fitness assignment on the exploration-exploitation balance has drawn little attention. The vast majority of MOEAs presented to date resort to Pareto dominance classification as a fitness assignment methodology. However, the proportion of Pareto optimal elements in a set P grows with the dimensionality of P. Therefore, when the number of objectives of a Multi-Objective Problem (MOP) is large, Pareto dominance-based ranking procedures become ineffective in sorting out the quality of solutions. This paper investigates the potential of using preference order-based approach as an optimality criterion in the ranking stage of Multi-Objective Evolutionary Algorithms (MOEAs). A ranking procedure that exploits the definition of Preference Ordering (PO) is proposed, along with two strategies that make different use of the conditions of efficiency provided, and it is compared to a more traditional Pareto dominance-based ranking scheme within the framework of NSGA-II. A series of extensive experiments is performed on four widely applied test functions, namely DTLZ1, DTLZ2, DTLZ3 and DTLZ5, for up to eight objectives. The results are analyzed through a suite of five performance metrics and indicate that the ranking procedure based on PO enables NSGA-II achieve better scalability properties compared to the standard ranking scheme and suggest that the proposed methodology could be successfully extended to other MOEAs.

KEYWORDS

Fitness assignment, multi-objective, ranking procedure, selective pressure.

INTRODUCTION

It may be generalized that all Evolutionary Algorithms (EA) draw their strength from two sources: exploration and exploitation. Genetic Algorithms are no exception. When exploration and exploitation are unbalanced, either the search stalls around unpromising areas of the objective space or it prematurely converges to some local optima.

Most commonly, explorative properties are attributed to genetic operators such as recombination and mutation, while exploitative ones to selection. In all EA, the latter is referred to as the mechanism that considers the quality (fitness) of individuals in order to make a choice as to which chromosome to favor for reproductive purposes. Surprisingly, the process through which individuals are assigned a fitness value, generally referred to as fitness assignment, has been given little attention in the perspective of the quest for an exploration-exploitation balance in a Multi-Objective context.

Many real-world optimization problems involve multiple objectives that need be considered simultaneously. If these objectives are conflicting, as it is usually the case, a single solution optimal for all the objectives can not be found. As a consequence, the assessment of the quality (fitness) of a set of solutions is not straightforward (compared to single objective optimization) and calls for a method that provides a formal definition of the qualitative notion of compromise. The vast majority of MOEAs presented to date [2;8;14;19;23;25;26], solve this predicament through the concept of Pareto dominance, which is embedded in their ranking procedures and exploited to classify the solutions generated and, therefore, to drive the search for better ones. However, as it is theoretically shown in [12] and empirically evidenced in [6, pages 404-405], the proportion of Pareto optimal elements in a set P grows with the dimensionality of P. Therefore, when the number of objectives of a Multi-Objective Problem (MOP) is large, Pareto dominance-based ranking procedures become ineffective in sorting out the quality of solutions and selection operators can play a minor role to compensate for this effect, which inexorably translates into a lack of selective pressure.

Preference Ordering (PO) is a generalization of Pareto optimality that provides two definitions of efficiency more stringent than Pareto dominance to compare solutions to MOPs [3]. Therefore, a question that naturally arises is whether PO could substitute, or be exploited in conjunction with, Pareto ranking to classify solutions in a multi-objective framework in order to alleviate the aforementioned lack of effectiveness. While PO has been applied successfully in a recent study [18] as a post processing routine to select for further assessment a subset of high quality solutions from a set of Pareto optimal solutions, to the best knowledge of the authors it has never been exploited to drive the search of an MOEA.

This paper investigates the potential of using preference order-based approach as an optimality criterion in the ranking stage of Multi-Objective Evolutionary Algorithms (MOEAs). This study will demonstrate that, since PO is less affected than Pareto dominance by the progressive loss of effectiveness as the dimension of the objective space of an MOP increases, it could greatly enhance the scalability properties of multi-objective genetic algorithms.

A ranking procedure that exploits the definition of PO is proposed, along with two strategies that make different use of the conditions of efficiency provided, and it is compared to a more traditional Pareto dominance-based ranking scheme within the framework of NSGA-II [7]. A series of extensive experiments is performed on four widely applied test functions, namely DTLZ1, DTLZ2, DTLZ3 and DTLZ5 [10], for five different numbers of objectives (4, 5, 6, 7 and 8). The results are analyzed and the convergence and diversity preserving properties are compared by means of a suite of five performance metrics.

The paper is structured as follows: the first section gives the reader the theoretical background of Preference Ordering. The second section proposes a ranking procedure based on Preference Ordering along with two alternative strategies that exert different levels of selective pressure. Some empirical results on the complexity of the algorithm implemented are also given. The third and fourth sections introduce the 4 test functions used as Multi-Objective Problems (MOP) for the experiments of this study and the metrics to measure the performances of the algorithms respectively. The fifth section details the experimental setup. Finally, the results are discussed and unexplored research issues suggested.

FROM PARETO DOMINANCE TO PREFERENCE ORDERING

The multi-objective optimization problem (MOP) may be stated as a minimization problem (without loss of generality) as [1;4]:

Definition (MOP). Find a vector $\boldsymbol{x}^* = [x_1^*, x_2^*, ..., x_n^*]^{\mathrm{T}}$ such that:

$$\begin{aligned} \boldsymbol{f}(\boldsymbol{x}^*) = \min_{\boldsymbol{x}\in D} \boldsymbol{f}(\boldsymbol{x}) = \min_{\boldsymbol{x}\in D}[f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})]^{\mathrm{T}} \\ \boldsymbol{g}_{\boldsymbol{i}}(\boldsymbol{x}) \geq 0 \quad i = 1, 2, \ldots, l \\ \boldsymbol{h}_{\boldsymbol{i}}(\boldsymbol{x}) = 0 \quad i = 1, 2, \ldots, L \end{aligned} \tag{1}$$

where n is the dimension of the decision variables vector, m is the number of objectives for the MOP, l is the number of inequality constraints $\boldsymbol{g}(\boldsymbol{x})$, L the number of equality constraints $\boldsymbol{h}(\boldsymbol{x})$. D represents the *search space* (also referred to as *decision space*), i.e. the Euclidean space that is defined by the set of all n-tuples of real numbers, denoted by R^n. This definition could be easily extended to situations where the variables take values on other sets. The constraints $\boldsymbol{g}(\boldsymbol{x})$ and $\boldsymbol{h}(\boldsymbol{x})$ act on D as to defining a subset q, which represent the *feasible region* for the MOP. Any point $\boldsymbol{x}$ in q represents a feasible solution; the function $\boldsymbol{f}$ maps q into the $\mathcal{F}$ that is a subset of the *objective space,* denoted by R^m.

Definition (Ideal solution). A point ${}^{U}\boldsymbol{x}^* = [{}^{U}x_1^*, {}^{U}x_2^*, ..., {}^{U}x_n^*]^{\mathrm{T}}$ such that:

$$\boldsymbol{f}({}^{U}\boldsymbol{x}^*) = \min_{\boldsymbol{x}\in \boldsymbol{D}} \boldsymbol{f}({}^{U}\boldsymbol{x}) = \left[\min_{\boldsymbol{x}\in \boldsymbol{D}} f_1({}^{U}\boldsymbol{x}), \min_{\boldsymbol{x}\in \boldsymbol{D}} f_2({}^{U}\boldsymbol{x}), \ldots, \min_{\boldsymbol{x}\in \boldsymbol{D}} f_m({}^{U}\boldsymbol{x})\right]^{\mathrm{T}} \tag{2}$$

is the ideal solution and $\boldsymbol{f}({}^{U}\boldsymbol{x}^*)$ is the ideal vector (also referred to as the utopia point or shadow minimum) for the MOP. In words, a point ${}^{U}\boldsymbol{x}^*$ is an ideal solution for an MOP, and consequently $\boldsymbol{f}({}^{U}\boldsymbol{x}^*)$ is the utopia point, if all the individual objectives are simultaneously optimized.

The problem defined by Eqn. (1) is well posed only if the objective functions f_1,é , f_n are non conflicting. When this holds, if the ideal solution lies in the feasible region, i.e., if ${}^{U}\boldsymbol{x}^* \in \Omega$, then it is the solution to the optimization problem. Unfortunately, this is seldom the case. Most engineering problems naturally lead to the concurrent optimization of a pool of conflicting objectives. In such situations, a single (utopical) solution is not attainable; instead, one is constantly required to find a set of good compromises (ñtrade-offsò). Pareto optimality provides a formal definition of the intuitive yet qualitative notion of compromise.

Definition (Pareto Dominance and Pareto Optimality). The vector $\boldsymbol{f}(\boldsymbol{x})$ is said to dominate the vector $\boldsymbol{f}(\boldsymbol{x})$, denoted $\boldsymbol{f}(\boldsymbol{x}) \prec \boldsymbol{f}(\boldsymbol{x})$, if $f_i(\boldsymbol{x}) \leq f_i(\boldsymbol{x})$ for all $i \in$ {1,2,é ,n} and $f_j(\boldsymbol{x}) < f_j(\boldsymbol{x})$ for some $j \in$ {1,2,é ,n}. A point $\boldsymbol{x}^*$ is said to be Pareto optimal or Pareto non-dominated for the MOP if and only if there does not exist $\boldsymbol{x} \in$ q such that $\boldsymbol{f}(\boldsymbol{x}) \prec \boldsymbol{f}(\boldsymbol{x}^*)$. In words, a point $\boldsymbol{x}^*$ is Pareto efficient if there does not exist a point $\boldsymbol{x}$ in Ω that would achieve a better value for one of the objectives without worsening at least another.

It is easy to show that Pareto dominance induces a partition onto the objective space. Without loss of generality we restrict our reasoning to the feasibility region Ω. Suppose, for instance, that within Ω there exists a point A; its coordinates in Ω define the regions where points that dominate A and that are dominated by A could lie. We denote these regions Ω_A^+ and Ω_A^- respectively. Figure 1 is a pictorial representation of Ω_A^+ and Ω_A^- for two and three dimensions.

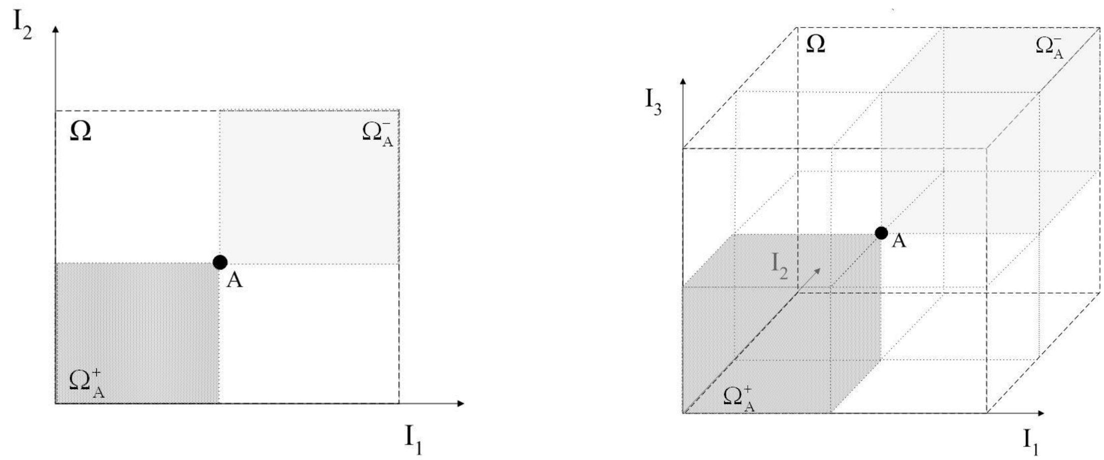


Figure 1: Representation of the partitioned induced by Pareto dominance onto the two and three-dimension objective spaces $\{I_1, I_2\}$ and $\{I_1, I_2, I_3\}$.

Definition (Pareto optimal set). For a given MOP, the Pareto optimal set denoted $\mathcal{P}^*$, is defined as:

$$\mathcal{P}^* := \{\boldsymbol{x} \in \Omega \mid \neg\exists\, \boldsymbol{x'} \in \Omega\ \boldsymbol{f}(\boldsymbol{x'}) \prec \boldsymbol{f}(\boldsymbol{x})\}$$

Definition (Pareto front). For a given MOP and a given Pareto optimal set $\mathcal{P}^*$, the Pareto front is defined as:

$$\mathcal{PF}^* := \{\boldsymbol{f} = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), ..., f_n(\boldsymbol{x})) \mid \boldsymbol{x} \in \mathcal{P}^*\}$$

As the number of objectives increases, the $\mathcal{PF}^*$ of an MOP quickly becomes vast. An intuitive explanation is suggested by Figure 1: moving from a two to a three-dimension objective space, $(\Omega_A^- + \Omega_A^-)$ grows at a lower rate than Ω.

This phenomenon has a great impact on the performance of the optimization algorithms that rely on Pareto optimality principle to drive their search for good solutions. The immediate effect when applying the criterion of Pareto optimality to high dimensional problems is that a large number of solutions within any generation are non-dominated and, therefore, *qualitatively* indistinguishable from each other. Hence, a criterion more stringent than Pareto non-dominance, yet general, would be highly desirable. On these premise, we propose the use of Preference Ordering [4] instead of Pareto non-domination as a selection criterion for high dimensional problems.

Definition (Efficiency of order). A point $\boldsymbol{x}^* \in \Omega$ is considered efficient of order k if $\boldsymbol{f}(\boldsymbol{x}^*)$ is not dominated by any member of $\mathcal{PF}^*$ for any of the k-element subsets of the objectives. In words, a point is efficient of order k if it is Pareto optimal in all the $\binom{m}{k}$ subspaces of Ω obtained considering only k objectives at a time. It is clear that the efficiency of order m for an MOP with exactly m objectives simply enforces the Pareto optimality.

Figure 2 illustrates the efficiency of order for a set of points in a three-dimensional objective space q : amidst this set, the point efficient of order 2 is the only point Pareto optimal for all the 3 two-dimensional objective spaces obtained by considering all combinations of the three objectives of q , taken two at a time.

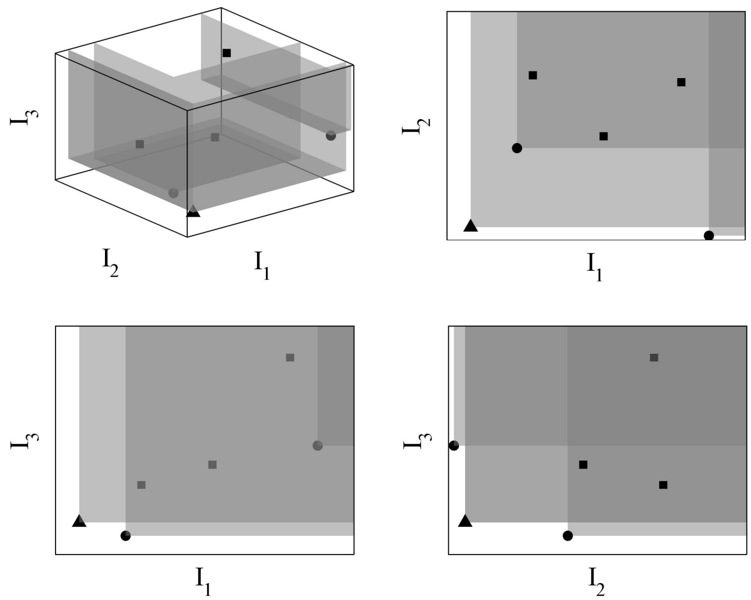


Figure 2: Pictorial representation of a set of points in a three-dimensional objective space and in its three two-dimensional projections onto the main planes. Squares and circles represent dominated and non dominated points respectively; the triangle represents the Pareto optimal point efficient of order 2.

Claim (*C1*): In a three-dimensional space there can be no more than one point efficient of order 2. This is a rather counterintuitive and interesting feature of Preference Ordering.

Proof: Suppose that A and B are two points in a three-dimensional objective space. Let us suppose that A and B are both efficient of order 2 and that $f_1(\mathrm{A}) < f_1(\mathrm{B})$. It follows that for the objective space $\{\mathrm{I}_1,\mathrm{I}_2\}$, the following relations must hold:

$$f_1(\mathrm{A}) < f_1(\mathrm{B}) \;\wedge\; f_2(\mathrm{B}) < f_2(\mathrm{A}) \tag{3}$$

From the first relation of (3), it follows that for the objective space $\{\mathrm{I}_1,\mathrm{I}_3\}$ the following must hold:

$$f_1(\mathrm{A}) < f_1(\mathrm{B}) \;\wedge\; f_3(\mathrm{B}) < f_3(\mathrm{A}) \tag{4}$$

and from the second relation (3) follows that for $\{\mathrm{I}_2,\mathrm{I}_3\}$ it must be:

$$f_2(\mathrm{B}) < f_2(\mathrm{A}) \;\wedge\; f_3(\mathrm{A}) < f_3(\mathrm{B}) \tag{5}$$

The second relations of (4) and (5) cannot both be contemporarily satisfied, unless $f_3(\mathrm{A}) = f_3(\mathrm{B})$. But according to relations (4) and (5), this implies that neither A nor B are efficient of order 2. In a similar way, it can be verified that the proof holds for $f_1(\mathrm{A}) > f_1(\mathrm{B})$.

A graphical explanation of *C1* is presented in Figure 3. As introduced above, given a point A in an objective space q , Pareto optimality delimits Ω^+_A and Ω^-_A. We denote ${}^{\{\mathrm{I}_i,\mathrm{I}_j\}}\Omega$, ${}^{\{\mathrm{I}_i,\mathrm{I}_j\}}\Omega^+_\mathrm{A}$ and ${}^{\{\mathrm{I}_i,\mathrm{I}_j\}}\Omega^-_\mathrm{A}$

their projections onto the two-dimensional spaces for every i,j = 1,2,3 with *i* Í *j,* respectively. Let us assume that A is efficient of order 2. For any point B to be efficient of order 2, its projections onto the plane I_iI_j should lie in $({}^{\{I_i,I_j\}}\Omega\ -\ {}^{\{I_i,I_j\}}\Omega_A^+\ -\ {}^{\{I_i,I_j\}}\Omega_A^-)$ for every i,j = 1,2,3 with *i* Í *j*. It can be seen that A is the only such point.

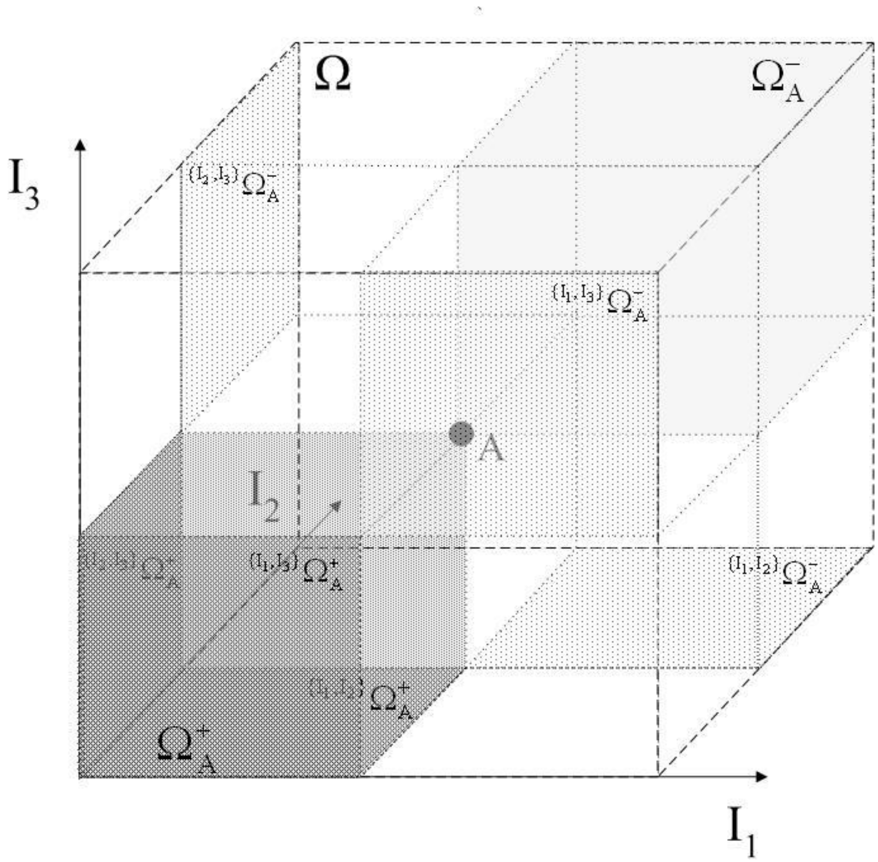


Figure 3: Graphical interpretation of Claim *C1*.

The condition of efficiency of order can thus be used to help reduce the number of points in a set by retaining only those that are regarded as ñbest compromisesò. In fact, it is intuitive that the less extreme components a point has, the more likely it is to be efficient of order. When the number of points selected is still considerably large, a more stringent criterion is required to sort out ñbetterò solutions:

Definition (Efficiency of order k with degree z): A point $\boldsymbol{x}^*$ is said to be efficient of order k with degree z if not dominated by any member of $\mathcal{P}^*$ for exactly z out of the possible $\binom{m}{k}$ k-element subsets.

As opposed to the condition of efficiency of order, the condition of efficiency with degree favors solutions that have extreme components and; therefore one should carefully orchestrate their cooperative usage. The next section describes a proposed ranking procedure based on Preference Ordering and suggests two alternative strategies that utilize these conditions of efficiency.

PREFERENCE ORDER RANKING PROCEDURE

In order to investigate on the effectiveness of Preference Ordering (PO) as an optimality criterion to be exploited in ranking procedures, NSGA-II [7] was selected as the framework evolutionary optimization algorithm. The choice of NSGA-II is motivated by two reasons: its poor scalability performance [16] as compared to other equally popular MOEA (PESA [2] and SPEA2 [25]) and its relatively simple mechanism for preserving diversity among the solutions. The former serves the main purpose of this paper, that is to say, to assess the usefulness of the best-rank solutions identified by Preference Ordering in the context of an evolutionary search. The latter challenges the ability of the algorithm to maintain a diversified genetic pool.

At every generation t of an NSGA-II run, the population P_t and the child population ${}^{C}P_t$ are combined into the compound population C_t and the ranking procedure can be summarized as follows:

(I) Identify the Pareto non-dominated individuals of C_t and group them into the subset R^I which is given rank 1;

(II) Identify the Pareto non-dominated individuals of C_t:$\{C_t \setminus R^I\}$ and group them into the subset R^{II} which is given rank 2;

(III) Iterate (I) and (III) until $C_t \setminus R^W = \emptyset$ where R^W is the subset that contains the worst individuals.

The proposed Preference Order ranking scheme that builds on the aforesaid procedure, but relies on the definitions of efficiency (*order* and *degree*) to classify individuals within the first subset identified (R^I). Consequently step (I) is revised and it can be stated as:

(I.I) Identify the Pareto non-dominated individuals of C_t and group them into R^I;

(I.II) Assign to the individuals of R^I a rank according to some strategy based on PO;

Once all the individuals of the first non dominated front are classified according to (I.II), then the NSGA-II ranking procedure resumes from step (II) and continues unaltered. Needless to say, after (I) not all the individuals in R^I have necessarily rank 1: the individuals classified as best according to strategy implemented in (I.II), referred to as R^{I*}, have rank 1 and the others will follow, the worst being given rank W. As a consequence, after (I.II), individuals in rank R^J will be given rank W+J.

So far, it has only been suggested that some strategy based on PO is used in the step (I.II) of the ranking procedure proposed to classify the first subset (R^I) of Pareto efficient individuals identified in (I.I). The next sub-section proposes two such alternative strategies, namely PO_k and $PO_{k,z}$, that were embodied in the ranking procedure detailed above and used for the experimental work of this study. It is clear that now we have at disposal three algorithms that differ altogether in their ranking procedure: the standard NSGA-II and NSGA-II geared with the ranking procedure based on PO and equipped with the strategies PO_k and $PO_{k,z}$. In the remainder of the paper, the last two algorithms will be referred to as $POGA_k$ and $POGA_{k,z}$ respectively.

Classification Strategies: PO_k and $PO_{k,z}$

The two definitions of efficiency, that of *order* and *degree*, provided by Preference Ordering offer two criteria that can help to progressively distinguish individuals in a set of Pareto non-dominated ones. However, as already pointed out, they act as to identify solutions that lay in different portions of a Pareto set $\mathcal{P}^*$, the former favoring the most balanced individuals (those with the fewer extreme components) and the latter rewarding those edging the set. Let us suppose that the MOP at stake has m objectives. If the Pareto set $\mathcal{P}^*$ identified is too numerous, one could resort to the condition of efficiency of order to identify the subset $\mathcal{P}^*_k$ that consists of solutions in $\mathcal{P}^*$ that are efficient of order

k. If $\mathcal{P}^*_k$ is still too numerous, one could retain only those solutions that are efficient with the highest degree, i.e. solutions in $\mathcal{P}^*_k$ that are not dominated by any point in $\mathcal{P}^*$ for the most of the possible $\binom{m}{k-1}$ (k-1)-element subsets. However, if $k = m$, i.e. if there are no solutions in $\mathcal{P}^*$ that are efficient of order m-1, applying the condition of efficiency with degree would result in the identification of the solutions that lay at the edge of the Pareto set $\mathcal{P}^*$.

In light of these considerations, two different strategies are proposed to rank the Pareto efficient set R^I (step (I.II) of the ranking scheme described in the previous sub section), and they are presented in Table I.

Table I: Preference Ordering based ranking strategies adopted in this study.

Ranking Strategy Name	Description
PO_k	1. Compute the order of efficiency of every solution in R^I; 2. Assign to every individual i in R^I a rank equal to k_i ï K + 1;
$PO_{k,z}$	1. Perform steps 1 and 2 of the strategy PO_k; 2. If there is more than one solution that share the same best order of efficiency k^{I*} Í m, then compute the degree of efficiency of order k^{I*} -1 for this subset of solutions R^{I*}; 3. Assign to every individual i in R^{I*} a rank equal to Z ï z_i + 1; 4. Assign to every individual i in { $R^I \setminus R^{I*}$} a rank equal to Z + k_i;
Symbols	**Description**
m	number of objectives
k	order of efficiency
z_i	degree of efficiency
K	$\min_{i \in R^I} (k_i)$
Z	$\max_{i \in P^{I*}} (z_i)$

Since $R^{I*} \subseteq R^I$, it is clear that $PO_{k,z}$ engenders a higher selective pressure than PO_k. Moreover, the former strategy degenerates to the latter when $k^{I*} = m$, i.e. when none of the individuals in R^I is at least efficient of order m-1.

Algorithm

Prior to describing the algorithm that it is proposed to implement the strategy PO_k, let us recall that [4] claims and proves that:

Claim 2 (C2). If $\boldsymbol{x}^*$ is efficient of order k, then it is efficient of order k+1.

In addition, we submit that:

Claim 3 (C3). If $\boldsymbol{x}^*$ is not efficient of order k, then it is not efficient of order k-1.

The proof immediately follows from C2. By induction, it is clear that if $\boldsymbol{x}^*$ is not efficient of order k, then it is not efficient of order j, for any $j < k$. The algorithm that we propose to compute the efficiency of order of a set $\mathcal{P}^*$ of N Pareto efficient solutions to an MOP is based on *C3* and it is described by the following pseudo code:

Algorithm: EFFICIENCY_OF_ORDER

```
Input: ℘* ⊆ ℝ^m , I ≡ {1,2,…,m}
rank_i ← 0  ∀ x^i ∈ ℘*
isefficient_i ← 1  ∀ x^i ∈ ℘*
j ← 0
  for c = 1 to (m choose m - j)
    for i = 1 to N
      if isefficient_i == 0
        break
      elseif ∃x^g ∈ ℘*, x^g ≠ x^i | x^g ≺^I_{(m-j),c} x^i
        isefficient_i ← 0
      else
        rank_i ← rank_i + 1
      end
    end
  j ← j +1
end
Output: rank_i  ∀ x^i ∈ ℘*
```

$$x^g \underset{(m-j),c}{\overset{\mathrm{I}}{\prec}} x^i \;\mathit{iff}\; \begin{cases} x_u^g \le x_u^i \;\forall u \in \mathrm{I}_c^{(m-j)} \\ x_u^g < x_u^i \text{ for at least one } u \in \mathrm{I}_c^{(m-j)} \end{cases}$$

$$\mathrm{I}_c^{(m-j)} : \left\{ \mathrm{I}_h^{(m-j)} \mid \mathrm{I}_h^{(m-j)} \subseteq \mathrm{P}(\mathrm{I}), \left|\mathrm{I}_h^{(m-j)}\right| = (m-j), h = 1,\ldots,\binom{m}{m-j}, h = c \right\}$$

$\mathrm{P}(\mathrm{I})$: Power Set of I, i.e. the set of all subsets of I

$|\mathrm{I}|$: cardinality of the set I

Figure 4 shows the experimental computation time required by the algorithm suggested to evaluate the degree of efficiency (ranking strategy PO_k) of a set $\mathcal{P}^*$ of Pareto efficient points for a variable cardinality of $\mathcal{P}^*$ (N) and number of objectives (m). Five values of m (4, 5, 6, 7 and 8) and eleven of N (between 20 and 80 with interval of 6) were considered altogether. The following steps were performed to evaluate a single point of the graphs shown, i.e. a triplet (Cp time, m_i, N_j):

1. Generate a random set of points $\mathcal{P}$ within the unitary m_i-dimensional hypercube;
2. Retain only the Pareto efficient individuals of $\mathcal{P}$ and insert them into $\mathcal{P}^*$.
3. If $|\mathcal{P}^*|$Ó N_i, then rank the first N_i, individuals of $\mathcal{P}^*$ according to PO_k and record the computation time.
4. Iterate steps 1 to 3 until a thousand computation time samples have been recorded and average them.

It has been shown experimentally ([6], pages 404-405) that the proportion of non-dominated individuals decreases with an increasing population size and the rate is higher for low-dimensional objective spaces than for higher. This suggested limiting the maximum population size of the set $\mathcal{P}^*$ to 80, so that an appropriate size of $\mathcal{P}$ was still affordable even for the worst case, i.e. m = 4 and N = 80.

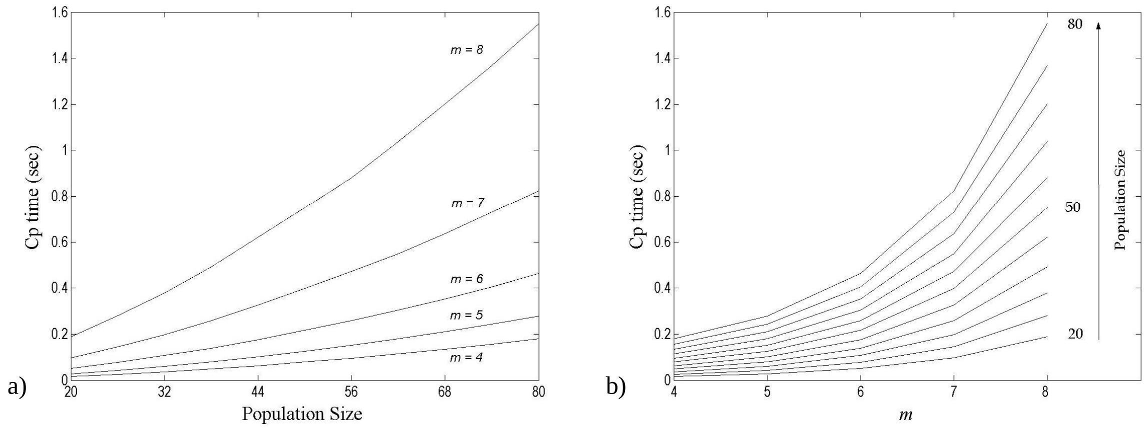


Figure 4: Experimental PO_k computation time versus: a) population size for different number of objectives, i.e. m = 4, 5, 6, 7 an 8 b) number of objectives for increasing population size.

The implementation of the strategy $PO_{k,z}$ is a simple extension of the algorithm presented above and it is omitted.

TEST FUNCTIONS

In order to compare the scalability behavior of the three algorithms NSGA-II, $POGA_k$ and $POGA_{k,z}$ four well known test functions were chosen, namely DTLZ1, DTLZ2, DTLZ3 and DTLZ5. They were selected for this study because they all share these important features: a) the relatively small implementation effort (Bottom-up approach [10]), b) can be scaled to any number of objectives and decision variables, c) the global Pareto front is known analytically, d) convergence and diversity difficulties can be easily controlled. More specifically, DTLZ1-3 challenge MOEA ability to converge to the optimal Pareto Front since they introduce a huge (controlled) number of local optima; DTLZ5 is peculiar in that its Pareto front is a degenerated hyper-surface and therefore it represents an interesting problem to test the proposed procedure.

All these test problems require the specification of a different functional $\boldsymbol{g}(\boldsymbol{x}_M)$ where $\boldsymbol{x}_M$ represents the last $M = n\text{-}m+1$; n and m are the number of decision variables and objective functions respectively and M is also referred to as the difficulty factor. Table II presents the mathematical formulation of the test problems used in this study, along with the decision space and the optimal solutions.

Since, according to *Claim C1*, the difference between Pareto dominance and Preference Ordering ranking procedures is almost negligible for a three-dimensional MOP, it was decided to focus the analysis on four, five, six, seven and eight objectives, i.e. m = 4, 5, 6, 7 and 8 respectively.

Table II: Description of the test problems used in this study.

Name	Problem Description	Optimal Solutions
DTLZ1	$\min_{\boldsymbol{x}\in\Omega} \boldsymbol{f}(\boldsymbol{x})$ $\Omega=\{\boldsymbol{x} \mid 0\le x_i \le 1\,\forall i=1,\ldots,n\}$ $f_1(\boldsymbol{x})=(1+\boldsymbol{g}(\boldsymbol{x}_M))\cdot 1/2\cdot x_1x_2\cdots x_{m-1}$ $f_2(\boldsymbol{x})=(1+\boldsymbol{g}(\boldsymbol{x}_M))\cdot 1/2\cdot x_1x_2\cdots(1-x_{m-1})$ $\vdots$ $f_m(\boldsymbol{x})=(1+\boldsymbol{g}(\boldsymbol{x}_M))\cdot 1/2\cdot(1-x_1)$ $\boldsymbol{g}(\boldsymbol{x}_M)=100\cdot\left[\lvert\boldsymbol{x}_M\rvert+\sum_{x_i\in\boldsymbol{x}_M}(x_i-0.5)^2-\cos(20\pi\cdot(x_i-0.5))\right]$	$x_i=0.5\ \ \forall i\in\boldsymbol{x}_M$
DTLZ2	$\min_{\boldsymbol{x}\in\Omega} \boldsymbol{f}(\boldsymbol{x})$ $\Omega=\{\boldsymbol{x} \mid 0\le x_i \le 1\,\forall i=1,\ldots,n\}$ $f_1(\boldsymbol{x})=(1+\boldsymbol{g}(\boldsymbol{x}_M))\cdot\cos(x_1\cdot\pi/2)\cdots\cos(x_{m-2}\cdot\pi/2)\cdot\cos(x_{m-1}\cdot\pi/2)$ $f_2(\boldsymbol{x})=(1+\boldsymbol{g}(\boldsymbol{x}_M))\cdot\cos(x_1\cdot\pi/2)\cdots\cos(x_{m-2}\cdot\pi/2)\cdot\sin(x_{m-1}\cdot\pi/2)$ $\vdots$ $f_m(\boldsymbol{x})=(1+\boldsymbol{g}(\boldsymbol{x}_M))\cdot\sin(x_1\cdot\pi/2)$ $\boldsymbol{g}(\boldsymbol{x}_M)=\sum_{x_i\in\boldsymbol{x}_M}(x_i-0.5)^2$	$x_i=0.5\ \ \forall i\in\boldsymbol{x}_M$
DTLZ3	$\min_{\boldsymbol{x}\in\Omega} \boldsymbol{f}(\boldsymbol{x})$ $\Omega=\{\boldsymbol{x} \mid 0\le x_i \le 1\,\forall i=1,\ldots,n\}$ $f_1(\boldsymbol{x})=(1+\boldsymbol{g}(\boldsymbol{x}_M))\cdot\cos(x_1\cdot\pi/2)\cdots\cos(x_{m-2}\cdot\pi/2)\cdot\cos(x_{m-1}\cdot\pi/2)$ $f_2(\boldsymbol{x})=(1+\boldsymbol{g}(\boldsymbol{x}_M))\cdot\cos(x_1\cdot\pi/2)\cdots\cos(x_{m-2}\cdot\pi/2)\cdot\sin(x_{m-1}\cdot\pi/2)$ $\vdots$ $f_m(\boldsymbol{x})=(1+\boldsymbol{g}(\boldsymbol{x}_M))\cdot\sin(x_1\cdot\pi/2)$ $\boldsymbol{g}(\boldsymbol{x}_M)=100\cdot\left[\lvert\boldsymbol{x}_M\rvert+\sum_{x_i\in\boldsymbol{x}_M}(x_i-0.5)^2-\cos(20\pi\cdot(x_i-0.5))\right]$	$x_i=0.5\ \ \forall i\in\boldsymbol{x}_M$
DTLZ5	$\min \boldsymbol{f}(\boldsymbol{x})$ $\Omega=\{\boldsymbol{x} \mid 0\le x_i \le 1\,\forall i=1,\ldots,n\}$ $f_1(\boldsymbol{x})=(1+\boldsymbol{g}(\boldsymbol{x}_M))\cdot\cos(\theta_1\cdot\pi/2)\cdots\cos(\theta_{m-2}\cdot\pi/2)\cdot\cos(\theta_{m-1}\cdot\pi/2)$ $f_2(\boldsymbol{x})=(1+\boldsymbol{g}(\boldsymbol{x}_M))\cdot\cos(\theta_1\cdot\pi/2)\cdots\cos(\theta_{m-2}\cdot\pi/2)\cdot\sin(\theta_{m-1}\cdot\pi/2)$ $\vdots$ $f_m(\boldsymbol{x})=(1+\boldsymbol{g}(\boldsymbol{x}_M))\cdot\sin(\theta_1\cdot\pi/2)$ $\boldsymbol{g}(\boldsymbol{x}_M)=\sum_{x_i\in\boldsymbol{x}_M}x_i^{0.1}$ $\theta_i=\frac{1}{2\cdot(1+\boldsymbol{g}(\boldsymbol{x}_M))}\left(1+2\cdot\boldsymbol{g}(\boldsymbol{x}_M)\cdot x_i\right)\qquad\forall i=2,\ldots,(m-1)$	$x_i=0\ \ \forall i\in\boldsymbol{x}_M$

Symbol	Description
n	No. of decision variables
m	No. of objective functions
$\boldsymbol{x}_M$	$\{x_i\forall\ i=n-M+1,n-M,\ldots,n\}$
M	n-m+1

PERFORMANCE METRICS

A set of five *functionally independent* [9] metrics commonly used in the literature was adopted to compare the performance of NSGA-II, $POGA_k$ and $POGA_{k,z}$. They are: generational distance (*GD*), normalized hyper volume (*Hn*), two set coverage (*CS*), diversity metric 1 (*DM1*) and diversity metric 2 (*DM2*).

Generational Distance (GD). The Pareto surfaces of the 4 test functions herein considered are known analytically. Therefore, instead of computing this metric as suggested in [1], as the expected value of the minimum distance between a point and a reference set, with respect to the entire population, i.e.

$$GD = \frac{\left(\sum_{i=1}^{N} d_i^2\right)^{1/2}}{N}$$

it is possible to evaluate this metric as the expected value of the distance between each individual in a population and the Pareto surface.

Normalized Hyper Area (Hn). This metric relates to both the sparseness of a set of points and its convergence to an ideal set and is a variation of the one used by [25]. It measures the fraction of the space between the axes origin and a reference point that is dominated by the members of the set. It is defined as

$$Hn = \frac{\bigcup_{i=1}^{N^*} hv_i}{hv_O}$$

where N^* is the number of non-dominated points in the set, hv_i is the *hyper*-volume enclosed between the point *i* and the Reference Point (RP) and hv_O is the *hyper*-volume enclosed between the origin and the reference point.

Two set Coverage (CS) [27]. This metric relates to the relative coverage of two sets. Consider $X', X'' \subseteq X'$ as two sets of points in the objective space of an MOP. CS is defined as

$$CS(X', X'') = \frac{|\{\boldsymbol{x''} \in X''; \exists\, \boldsymbol{x'} \in X' : \boldsymbol{x'} \preceq \boldsymbol{x''}\}|}{|X''|}$$

If all the points in X' dominate or are equal to all points in X'', then $CS(X', X'') = 1$; if all points dominate in X'' dominate the points in X', $CS(X', X'') = 0$. Since, in general $CS(X', X'') \neq 1 - CS(X'', X')$, both values should be considered.

Diversity Metric (DM) 1 & 2. These metrics [9] were introduced to measure the diversity amongst a set of points P, with respect to a Reference Set RS. In words, a set of *m*-dimensional points is projected into a (*m*-1)-dimensional hyper plane, which is discretized in a number of grids. The more grids that contain a point of RS also contain a point in P, the higher the diversity metric value. *DM1* and *DM2* differ in that the former entails the discretization into grids of the RS set, which is to be representative of the global Pareto front of an MOP; the latter, instead, requires the discretization of the (*m*-1)-dimensional objective space represented by P. It is easy to understand that *DM2* enables the assessment of the diversity of P even when its convergence to RS is poor.

To compute these metrics:

1. For each grid indexed by $i, j, .., m$-1 the following two arrays must be computed:

$$H(i,j,\ldots,m-1)=\begin{cases}1 & \text{if the grid has a representative point in RS}\\ 0 & \text{otherwise}\end{cases}$$

$$h\ (i,j,\ldots,m-1)=\begin{cases}1 & \text{if } H(i,j,\ldots,m-1)=1 \text{ and the grid has a representative point in RS}\\ 0 & \text{otherwise}\end{cases}$$

2. Compute $m(H(i,j,\ldots,m\text{-}1))$ and $m(h(i,j,\ldots,m\text{-}1))$ according to a scheme that takes into account the neighbor grids.
3. Compute the diversity metric for the set P with respect to RS as

$$DM(P)=\frac{\sum\limits_{\{i,j,\ldots m-1 \mid H(i,j,\ldots m-1)\neq 0\}} m(h(i,j,\ldots m-1))}{\sum\limits_{\{i,j,\ldots m-1 \mid H(i,j,\ldots m-1)\neq 0\}} m(H(i,j,\ldots m-1))}$$

For boundary grids, a neighboring imaginary grid with unitary H and h is always assumed. Consequently, to avoid the artificial effects introduced by the imaginary grids, the formula above is modified to

$$DM(P)=\frac{\sum\limits_{\{i,j,\ldots m-1 \mid H(i,j,\ldots m-1)\neq 0\}} m(h(i\ ,j,\ldots m-1))-\sum\limits_{\{i,j,\ldots m-1 \mid H(i,j,\ldots m-1)\neq 0\}} m(\boldsymbol{0})}{\sum\limits_{\{i,j,\ldots m-1 \mid H(i,j,\ldots m-1)\neq 0\}} m(H(i,j,\ldots m-1))-\sum\limits_{\{i,j,\ldots m-1 \mid H(i,j,\ldots m-1)\neq 0\}} m(\boldsymbol{0})}$$

Deb [9] and Khare [16] suggested a neighboring scheme to compute the values of $m(h)$ and $m(H)$; the same scheme is used for the present study and it is presented in Table III. They also suggested to compute dimension wise values of $m(h)$ and $m(H)$ for 2 or more-dimensional hyper planes: herein, an average value over all principle axes is used. It is important to point out that h and H could be evaluated directly over the lesser-dimensional hyper plane were the P and RS are projected onto, or over the segments used to compute the dimension-wise $m(h)$ and $m(H)$. Since the former entails a much finer discretization of the objective space as compared to the latter for a given dimension wise grid size, it is also likely to yield values of *DM1* almost nil when the P has not quite converged to RS thus reducing the significance of any comparison between two ore more set of points. Consequently, the second method to evaluate h and H was chosen.

Furthermore, when computing *DM2*, instead of assuming that each grid of the discretized P set be represented by a point in the RS, $H(i,j,\ldots,m\text{-}1)$ is computed on the discretized RS objective space and its values are used to evaluate $h((i,j,\ldots,m\text{-}1)$ as from the formula presented above over the discretized objective space of the set P.

Table III: neighboring scheme to evaluate m(h) and m(H) for a 2-dimensional surface.

***h*(…,*j*-1,…)**	***h*(…,*j*,…)**	***h*(…,*j*+1,…)**	***m*(*h*(…,*j*,…))**
0	0	0	0
0	0	1	0.5
1	0	0	0.5
0	1	1	0.67
1	1	0	0.67
1	0	1	0.75
0	1	1	0.75
1	1	1	1

EXPERIMENTAL STUDY

The main focus of this study is to assess if the ranking procedures based on Preference Ordering favor individuals that carry good genetic materials so that the evolutionary search for better solutions is guided towards promising regions compared to that of a Pareto non-dominance based scheme. The second aim is to examine the scalability of POGA compared to the standard NSGA-II. In order to investigate these two issues with the greatest clarity, it was decided to trim off all the elements that might blemish the outcomes of the experiments to be conducted. Consequently, a canonical form of MOEA was used and a fixed population size was chosen instead of what was suggested in [16;17]. This canonical MOEA allows the assessment of the difference in selective pressure exerted by the ranking procedures over objective spaces of different dimensionality. Consequently, a compensative effect provided by a (not linearly) increasing population size (as suggested in [16;17]) was disregarded for the purpose of this study. Similarly, the binary coding of the decision variables was preferred to the real one. The reason for this choice requires further explanation. First, all the experiments with the four test functions were conducted with a constant difficulty factor M. Recalling that $M = m\text{-}n\text{+}1$ it is easy to appreciate that, when the number of objectives (m) increases, so must the number of decision variables (n) in order to maintain M at a constant value. If the real coding had been used, the mutation probability should have been set to some function of the number of decision variables $G(n)$, as commonly suggested in the literature [21;22]. Again, this would have introduced an unnecessary complication. Hence, a binary coding, with a mutation probability as a function of the (constant) population size was used [5]. Second, it was shown [20, pg. 148] that the adaptive nature of the crossover operator featured by the real-coded NSGA-II, i.e. the Simulated Binary Crossover (SBX), was the major reason for the poor scalability performance of NSGA-II on one of the test problems considered in this study.

A proper calibration of the MOEAs parameters for each test functions and each number of objectives (m) was considered unnecessarily time consuming and not completely aligned to the aim of the study. A more simplistic parameter tuning approach was then preferred: the 6-dimensional (m =6) DTLZ2 was selected for this procedure and the best set of parameter values found for this test were applied to all the other experiments performed hereafter. Table IV summarizes the genetic operators and the parameterization used for NSGA-II, POGA_k and $\text{POGA}_{k,z}$ to solve the test MOPs considered.

Table IV: MOEA settings adopted for this study.

Operator type	Operator Name	Parameter Name	Parameter Value
Cross over	single-point	probability	0.6
Mutation	uniform	probability	$1/N$
		Population size (N)	100

Each experiment consists of 10 runs of a MOEA over the same MOP, which is completely identified by the pair (Test_Function_Name-m). This would allow the consideration of the effect of different random number generator seeds on the result of the optimizations. The maximum number of generations for each run varies and was set according to [16]. Table V summarizes the difficulty factor values, the number of runs and the maximum number of generations for each MOP.

Table V: Difficulty factor (M), number of runs and number of generations for each test problem considered.

Test Problem	M	Number of runs	Number of generations	
			m={4,5,6}	m={7,8}
DTLZ1	5	10	300	600
DTLZ2	10	10	300	600
DTLZ3 & DTLZ5	10	10	500	1000

In order to compute the metrics *Hn*, *DM1* and *DM2* a few parameters need to be set. More specifically, *Hn* requires the Reference Point to be specified. A careful choice of this parameter turned out being essential: too high a vector-value would have avoided the NSGA-II *Hn* metric to fall to 0, but would have also shrunk the difference between the metric values of POGA_k and $\text{POGA}_{k,z}$ with a consequent loss of useful information. The converse effect induced by too close an RP to the origin is rather intuitive and likewise to be avoided. The best choice for the RP was then carried out for each test problem and for each number of objectives. In order to compute the diversity metric *DM1* and *DM2*, the number of points in the reference set, the lesser-dimensional reference plane and the number of grids used to discretize each principal axis were to be specified. Table VI summarizes all the parameter values used for this study.

Table VI: Parameter values used to compute metric Hn, DM1 and DM2 for the experiments performed in this study.

Parameter	m	***Hn***			
		DTLZ1	DTLZ2	DTLZ3	DTLZ5
	4	[5,5,5,5]	[1,1,1,1]	[5,5,5,5]	[5,5,5,5]
	5	[5,5,5,5,5]	[1,1,1,1,1]	[5,5,5,5,5]	[5,5,5,5,5]
RP	6	[50,50,50,50,50,50]	[1,1,1,1,1,1]	[50,50,50,50,50,50]	[5,5,5,5,5,5]
	7	[50,50,50,50,50,50,50]	[1,1,1,1,1,1,1]	[50,50,50,50,50,50,50]	[5,5,5,5,5,5,5]
	8	[50,50,50,50,50,50,50,50]	[1,1,1,1,1,1,1,1]	[50,50,50,50,50,50,50,50]	[5,5,5,5,5,5,5,5]

Parameter	***DM1* & *DM2***
|RS|	100
Reference Plane	$f_m = 0$
# of grids	100

RESULTS

This section details the results of the experiments conducted, organized by group of metrics: convergence metrics (*GD* and *Hn*), diversity metrics (*DM1* and *DM2*) and *CS* values are presented respectively.

Convergence Metrics: *GD* and *Hn*.

Figure 5 shows the trend of the mean values of *GD* and *Hn* versus the number of objective functions (*m*) for the three algorithms NSGA-II, POGA_k and $\text{POGA}_{k,z}$ for the test problems considered. The numerical values of mean and standard deviation are shown in Table A - I in the Appendix section, along with the formulae to compute the *GD* for each test function.

Overall, all the experiments suggest a good convergence to the true Pareto front. In particular, the parameterization chosen enabled NSGA-II to attain better performances than ever reported in the literature so far [16] on test problems DTLZ1 and DTLZ2. It can also be appreciated that POGA_k and $\text{POGA}_{k,z}$ performed similarly both in terms of *GD* and *Hn* values; consequently, in this sub-section emphasis is given to the difference between the Preference Ordering based algorithm, referred as to POGA and NSGA-II.

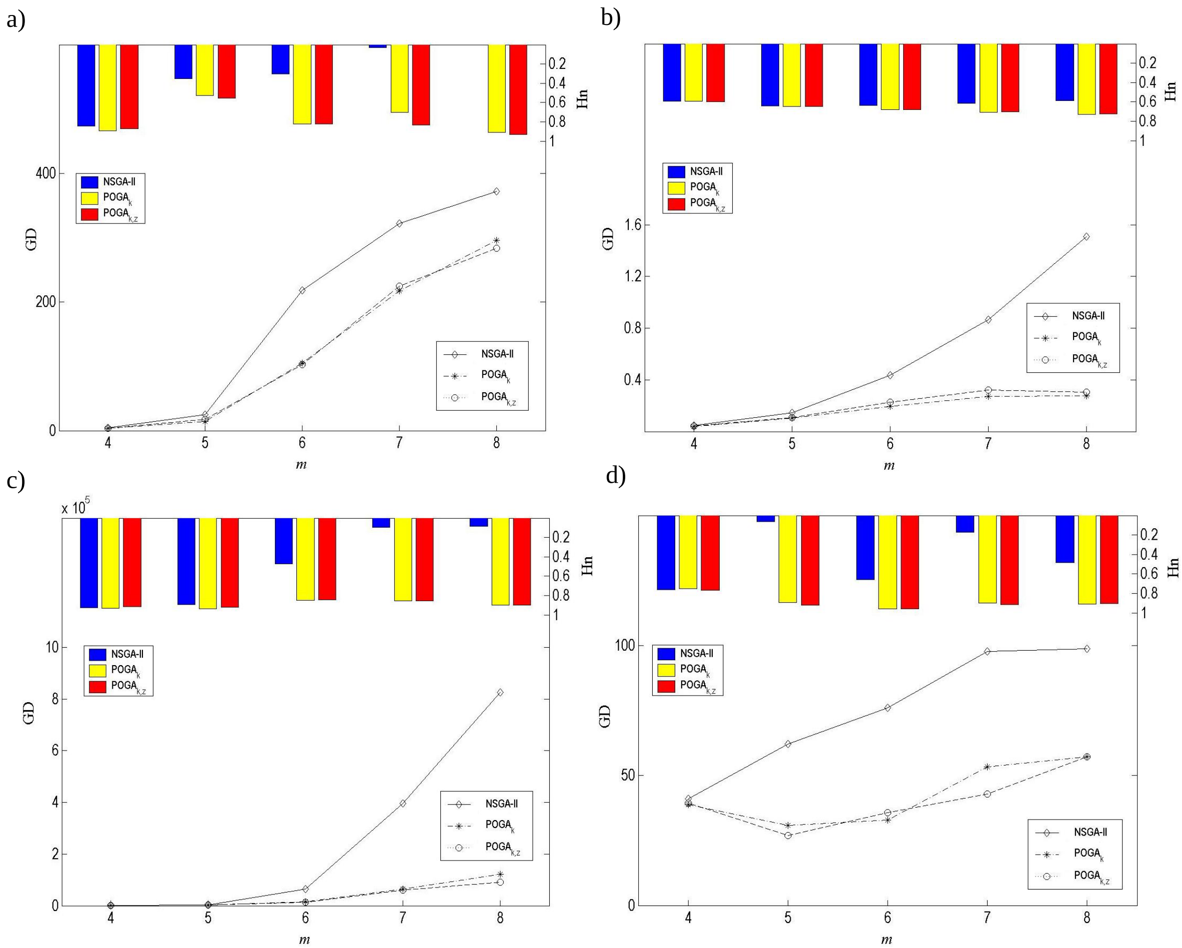


Figure 5: *GD* and *Hn* metrics for the test problems: a) DTLZ1; b) DTLZ2; c) DTLZ3; d) DTLZ5.

DTLZ1. POGA$_k$ and POGA$_{k,z}$ show a generational distance comparable to that attained by NSGA-II when the number of objectives considered are four and five. When the test problem has 6 objectives, then there is a clear advantage of the POGA algorithms over NSGA-II, and this performance gap holds for higher number of objectives (m = 7, 8). The Hyper-volume metric (*Hn*) shows a somewhat different behavior. In fact, POGA algorithms and NSGA-II attain similar values for m = 4, but the former show a progressively better performance as the number of objectives considered increases.

DTLZ2. Both *Hn* and *GD* metric values show a progressively better performance of POGA algorithms over NSGA-II as the number of objectives increases. This is result is particularly significant since the tuning of the optimization parameters of NSGA-II was performed on this MOP. Therefore, it testifies that the better scalable properties achieved by POGA in terms of convergence properties are in fact attributable to the lower effectiveness of Pareto Efficiency as compared to Preference Ordering as optimality criterion to rank individuals in a high-dimensional objective space.

DTLZ3. The results show a pattern similar to that observed on the test problem DTLZ2. In fact, POGA algorithms perform progressively better than NSGA-II, but the performance gap shown by the first two algorithms on test problem DTLZ3 grows more dramatically than that observed on DTLZ2 as the number of objectives increases. Since DTLZ2 and DTLZ3 have the same global Pareto Surface, but the latter introduces many more local Pareto front than the former, this results enforces that the ranking procedure of POGA are more suitable than that of NSGA-II when the number of objectives of the MOP is high.

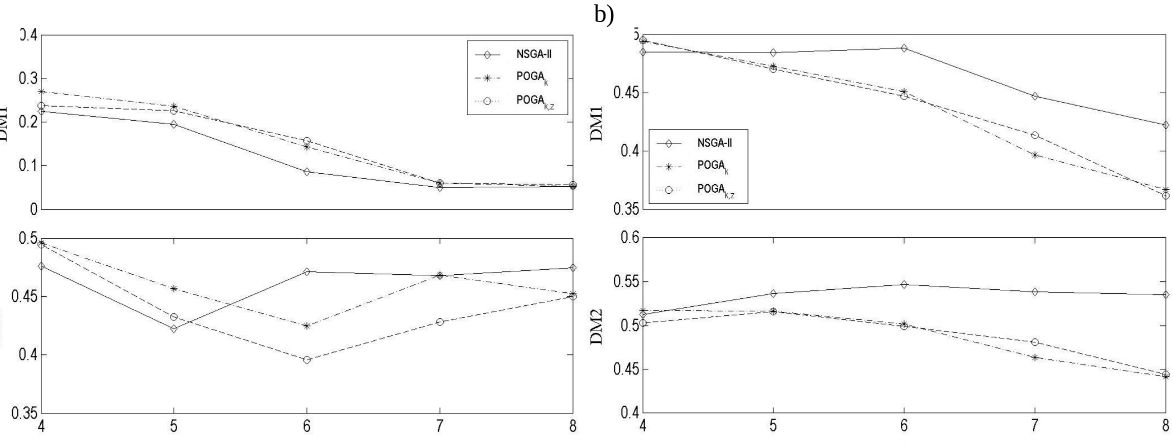
DTLZ5. POGA algorithms show a remarkably better convergence than NSGA-II over all the number of objectives considered, except for $m = 4$, where the three algorithms perform almost identically. For this test problem, the Hyper-volume metric was found particularly sensitive to the Reference Point chosen. As a consequence, a pattern arising from the observation of the *Hn* values over the five values of *m* considered cannot be safely deduced. Instead, it can be clearly stated that POGA attained better values than those generated by NSGA-II, for all the numbers of objective functions taken into account.

On the basis of the results just presented, it is argued that Preference Ordering rewards individuals that constructively contribute to build solutions of better quality. It also implicitly suggests that the diversity preserving operator based of the NSGA-II ranking procedure is not as effective as Preference Ordering to resolve the tie between equally rank individuals from the compound population competing to be inserted into the next generation population. In fact, on average, after only 15 to 30 generations for all experiments conducted ï depending on the number of objectives - all individuals of the compound population are Pareto efficient, thus equally ranked. Consequently, for the majority of the search duration, NSGA-II purely relies on crowding distance to identify those that will survive to the next generation, whereas POGA benefits from extra efficiency-based information.

Diversity Metrics: *DM1* and *DM2*

Figure 6 shows the trend of the mean values of *DM1* and *DM2* versus the number of objective functions (*m*) for the three algorithms NSGA-II, $POGA_k$ and $POGA_{k,z}$ for the test problems considered. Numerical values are shown Table A - II in the Appendix section. Overall, all the experiments show that POGA algorithms are less effective than NSGA-II in maintaining a good spread of solutions over the Pareto surface. This is in fact quite expected, since, according to what was pointed out in the first section, the two levels of Preference Ordering (order and degree of efficiency) favor sub-portions of a Pareto efficient set.

a) b)



c) d)

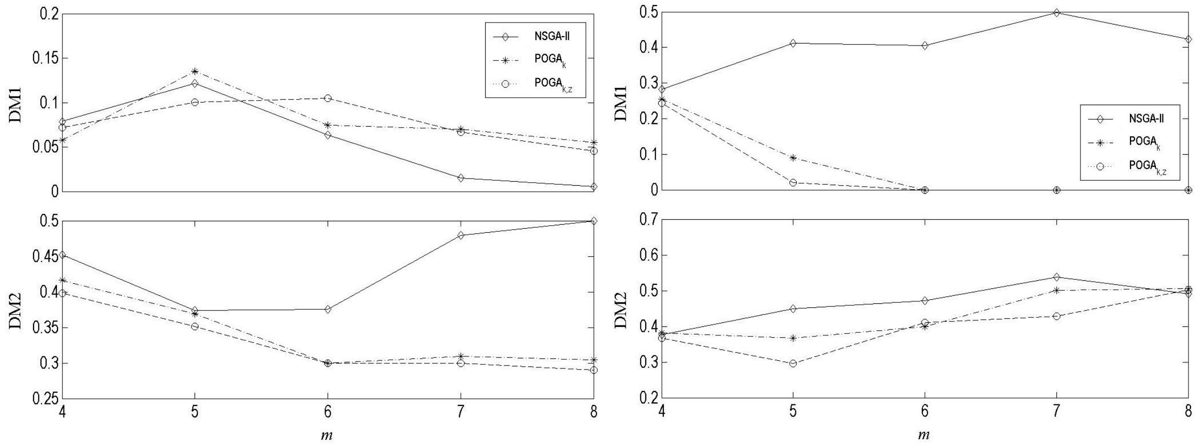


Figure 6: *DM1* and *DM2* metrics for the test problems: a) DTLZ1; b) DTLZ2; c) DTLZ3; d) DTLZ5.

DTLZ1. The first diversity metric indicates that NSGA-II generated a worse spread of solutions as compared to Preference Orderingï based algorithms. However, since NSGA-II, as evidenced in the previous sub-section, did not converge to the extent as POGA algorithms did, it is more appropriate to use the metric *DM2* as the basis for a comparison. Even so, a clear pattern can not be drawn for all the number of objectives considered, despite NSGA-II showing an overall better diversity.

DTLZ2. NSGA-II is clearly more effective than POGA algorithms in maintaining a good spread of solutions over the Pareto surface. This is shown for both metrics *DM1* and *DM2*, whose values also suggest that the advantage of the former algorithm over the latter ones grows with the dimension of the MOP tackled.

DTLZ3. The observation made about the test problem DTLZ1 holds for DTLZ3: since NSGA-II did not converge as quite well as POGA to the true Pareto Surface, in order to compare the algorithms the values of the second diversity metric *DM2* are taken into account. These indicate that, while the diversities of the Pareto sets found by NSGA-II and POGA are comparable when the number of objective functions (*m*) is low, NSGA-II outperforms the POGA for higher dimension spaces of the test problem.

DTLZ5. NSGA-II clearly attains a spread amongst the Pareto Set solutions unpaired by POGA algorithms over all cases considered. As it can be appreciated by looking at Figure 6, the *DM2* values of the latter algorithms become nil when the test function has 5 or more objectives. It is worth pointing out that this does not signify that all the solutions found by POGA are identical. Instead, they cover a relatively small portion of the true Pareto surface that is not represented (sampled) by the metric. In fact, the grid size and the dimension of the Reference Set chosen to evaluate this metric (Table VI) conjunctively generate a discretization of the objective space that is too fine for the test problem DTLZ5, particularly when the number of objectives is high, the reason being the degenerated nature of its Pareto Surface.

The analysis of the diversity metrics values suggests that while NSGA-II is more effective than POGA at maintaining a set of well spread solutions over the Pareto Surface, the two algorithms based on Preference Ordering, POGA$_k$ and POGA$_{k,z}$ are comparable in terms of performance, with a slight advantage of the former shown for the test problems DTLZ1 and DTLZ3.

Coverage Set Metric: *CS*

Figure 7 shows a pictorial representation of the *CS* metric values for the three algorithms on each test problem. It consists of four matrix-like graphs; each graph is a collection of box plots of the *CS* values computed on the last generation solutions of each run for m = 4, 5, 6, 7 and 8 respectively. The graph (i,j) relates to the values of the metric $CS(i,j)$, that is the proportion of solutions generated by the algorithm j that are dominated by those generated by i; i and j identifies the row and column algorithms respectively. For instance, the graph (2,3) of Figure 7a) represents the metric CS(POGA$_k$, NSGA-II) for the test problem DTLZ1, while the graph (3,1) represents the metric CS(NSGA-II, POGA$_{k,z}$).

DTLZ1. As it can be appreciated from Figure 7a), when the number of objectives is low (m = 4 and 5), POGA$_k$ is the algorithm that performs best, since it generates the highest proportion of solutions that are non-dominated neither by POGA$_k$ nor NSGA-II. However, it is worth pointing out that both algorithms based on Preference Ordering show quite an irregular behaviour, as graphs (1,3), (2,1) and (2,3) suggest, where the standard deviation of the metric *CS* is fairly high for low values of m. As the number of objectives grows, the performance gap experienced by NSGA-II holds almost unchanged (as already noticed for the convergence metrics), while the algorithms POGA obtain comparable *CS* values.

DTLZ2. Figure 7b) indicates a clear pattern: while for m = 4 the three algorithms perform comparably, as the number of objectives increases, an increasing proportion of solutions generated by NSGA-II are dominated by those generated by POGA algorithms, the latter yielding *CS* values almost identical. This behavior is also supported by the analysis of convergence and diversity metrics presented in the previous sub-sections.

DTLZ3. Similarly to what was observed for the previous test problem, the performance of NSGA-II progressively worsens compared to PO algorithms, but the gap in the *CS* values is significantly higher than that reported on DTLZ3. Both POGA$_{k,}$ and POGA$_{k,z}$ *CS* values are characterized by a high standard deviation when the number of objective functions considered is low (m = 4, 5), the behaviour already evidenced for the test problem DTLZ1. However, on average, it is safe to state that the performances of the two POGA algorithms do not differ appreciably over the entire range of objective numbers considered. This is in accordance with the results of both convergence and diversity metric values presented in the previous sub-sections.

DTLZ5. Figure 7d) shows comparable *CS* values obtained by the three algorithms for m = 4. For higher number of objective functions, NSGA-II clearly performed worse, but a definite pattern can not be evinced; in fact, graphs (1,3) and (2,3) show oscillating mean and standard deviation values that do not exhibit any trend. The two POGA algorithms have explored altogether diverse portions of the Pareto surface since their convergence and diversity metrics values are different, even though not remarkably, and the *CS* metric is nil when more than four objectives are considered.

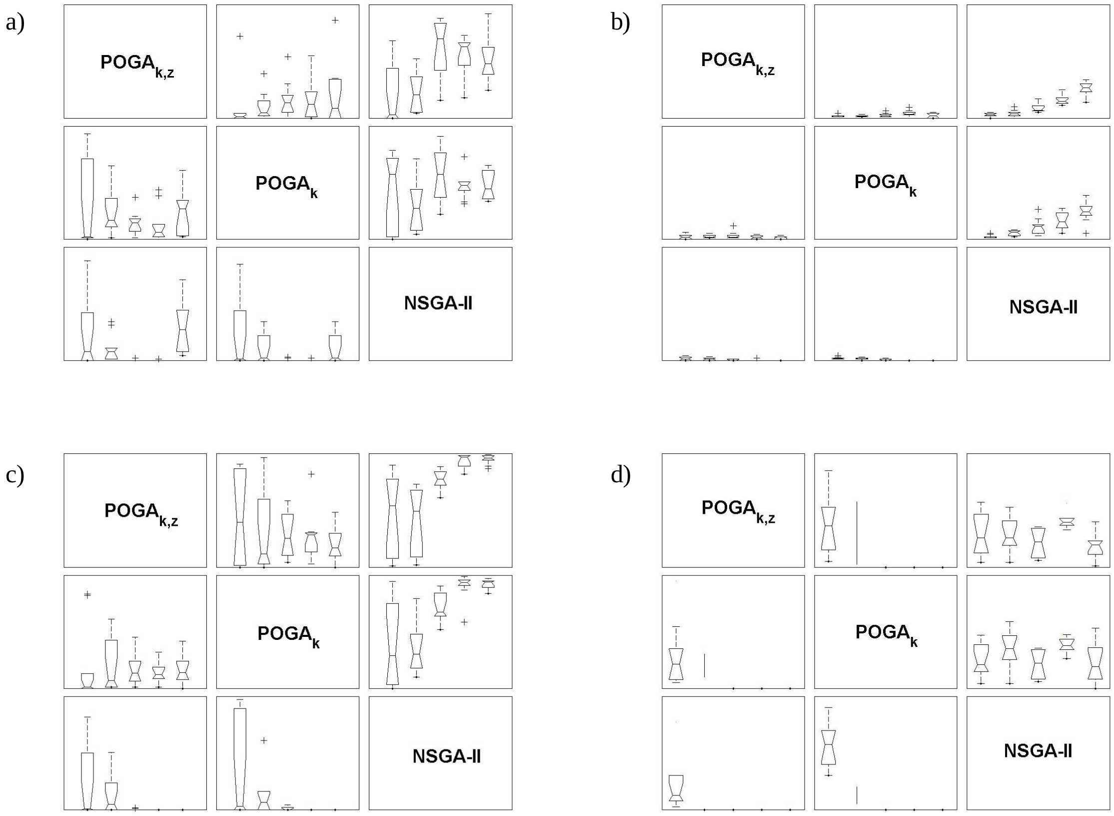


Figure 7: *CS* metrics for the test problems: a) DTLZ1; b) DTLZ2; c) DTLZ3; d) DTLZ5.

DISCUSSION

The results presented in the previous section bring empirical evidence to the claim that motivates this study: the proposed ranking procedure based on Preference Ordering considerably enhances the scalability properties of multi-objective genetic algorithms. This is testified through the results obtained by the two versions of the NSGA-II algorithm geared with the novel ranking procedure, namely POGA$_k$ and POGA$_{k,z}$. These two algorithms show superior convergence and *CS* metric values when compared to NSGA-II with the traditional ranking procedure based on Pareto efficiency. In order to further substantiate the assert, the Games-Howell [13] approximate test of equality of means was performed on the *GD*, *Hn* and *CS* metric values and the outcome (Table A - III and Table A - IV in the Appendix section) enforces the findings of the previous section: the means obtained by POGA algorithms are significantly better than those obtained by NSGA-II almost over the whole range of *m* (number of objective functions) investigated and for all test problems. Neither of the two POGA algorithms is found to perform distinctively and consistently better than the other. Further experiments are therefore necessitated to investigate how different types of strategies for orchestrating the two definitions of Preference Ordering relate to the structure (domain) of the MOP to be tackled. In the following subsections we explore in greater details two phenomena that emerged from the observation and the analysis of the values of *CS* and *DM* metrics.

Premature Convergence of POGA on DTLZ5

As far as the diversity properties are concerned, all the experiments show quite expectedly that POGA algorithms are less effective than NSGA-II at maintaining a well spread of solutions over the Pareto surface. In particular, test problem DTLZ5 has proven to greatly challenge the mechanism to preserve diversity of both POGA$_k$ and POGA$_{k,z}$. This is not enough to declare that Preference Ordering is mislead by the so called redundant solutions [10] occurring as a consequence of a degenerated Pareto Surface, and a comprehensive analysis on the matter is beyond the scope of this study. Nonetheless, the results on the test DTLZ5 overall suggest that POGA algorithms are affected to some extent by premature convergence. This, as it is vividly reasoned in [15], should not be attributed to the loss of diversity, which is usually a symptom of the aforementioned premature convergence, but rather to an earlier loss of explorative capabilities. On one hand, it remains unclear how to best detect and quantify this early loss of explorative capabilities. On the other hand, however, we argue that in the case of test problem DTLZ5, it is attributed to a concomitance of two reasons. However, prior to their detailed discussion, we note that it is convenient to think of the selective pressure generated by the PO-based ranking procedure of POGA as consisting of two components: one pointing towards region of low order of efficiency, and one pointing towards the global Pareto front. Hereafter, these components, which we name *lateral* and *frontal*, will be referred to as sp_l and sp_f respectively. We can now move onto discussing the two reasons we have just mentioned.

First, sp_l increases as the local Pareto fronts approach the global one. Let us, without loss of generality, focus on the three dimensional instance of test problem DTLZ5, and consider two local Pareto fronts and the global Pareto front, referred to as PF_{L2}, PF_{L1}, and PF_G respectively, as shown in Figure 8.

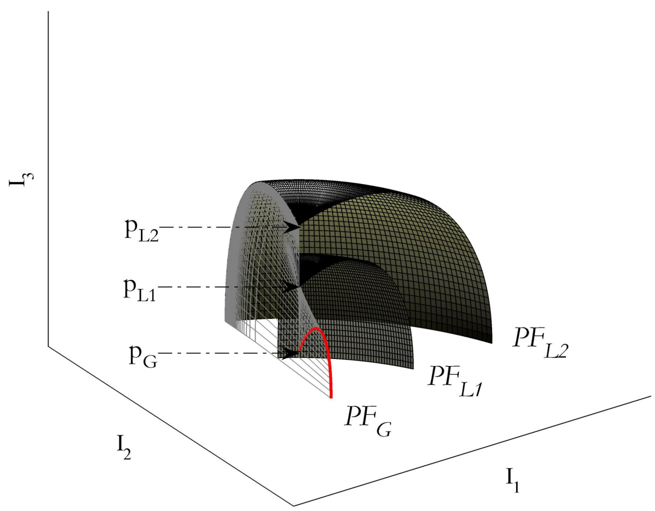


Figure 8: Increasing level of sp_l moving towards the global Pareto front (marked in red) of test problem DTLZ5.

It is easy to verify that the vectors p_{L2}, p_{L1} and p_G, are the only points efficient of order two for the Pareto fronts PF_{L2}, PF_{L1}, and PF_G respectively. More generally, for the test problem DTLZ5, the condition of efficiency of order two is an operator, which we refer to as $\mathbf{PO}_2$, that generates a mapping $\mathbf{PO}_2 : \mathrm{P} \subseteq \mathbb{R}^m \rightarrow S \subseteq \mathbb{R}^1$, where P is the collection of all local Pareto fronts of test problem DTLZ5 and S is the set of all feasible solutions that have the first m-1 components equal to 0, i.e. $S := \{\boldsymbol{x} \in \boldsymbol{f}(\Omega) \mid (\boldsymbol{x}_i = 0, \forall i = 1,\ldots,m-1) \wedge (\boldsymbol{x}_m \in \boldsymbol{f}_m(\Omega))\}$. However, if Preference Ordering is applied to determine the rank of any finite set of solutions PF_L^f drawn from a local Pareto front PF_L, three possible situations may occur: a) the set contains $\mathbf{PO}_2(PF_L)$, which is therefore assigned the best rank; b) the set does not contain $\mathbf{PO}_2(PF_L)$, but it contains another point $p \notin \mathbf{PO}_2(PF_L)$ that is efficient of order two, and therefore is assigned the best rank; c) the set does not contain $\mathbf{PO}_2(PF_L)$, and there may or may not be points in PF_L^f that are assigned better ranks than others, according to their relative position. It is also intuitive to understand that, the closer the local Pareto front PF_L to the global Pareto Front, the more likely the second situation. In fact, if Preference Ordering is applied to any finite set of solutions drawn from the global Pareto front, one point of the set is always assigned the best rank and this point is always the closest to S. This is in virtue of the fact that over the entire PF_G of DTLZ5, the first m-1 objectives are non-conflicting with each other. We can therefore conclude that, due to the sampling error generated by a finite population size, as the search progresses towards the global Pareto front, the lateral component of the selective pressure generated by POGA, i.e. sp_l, strengthens and progressively points towards S. We note that the reasoning above is based on the tacit assumption that POGA progresses through finite sets of points drawn from local Pareto fronts of DTLZ5. It is immediate to verify, however, that the conclusion holds for the more general situation.

Second, the diversity amongst the solutions is enforced within POGA through the Crowding Distance density-based estimator, which is biased in favor of boundary solutions. However as a consequence of the lateral component of the selective pressure of POGA, i.e. sp_l, steering the search towards S, the ranking procedure also favors solutions with extreme components. Thus, on test problem DTLZ5, the PO-based ranking procedure, coupled with diversity preserving mechanism based on Crowding Distance, triggers a positive feedback between an increasing sp_l and a decreasing diversity between the set of solutions generated.

From what it has been discussed thus far, one would expect POGA to discover at very early stages of the search solutions that are progressively closer to S and solutions that are increasingly closer to the global Pareto front alike. However, as the search progresses and sp_l tends to prevail, favored, in addition, by the diversity preserving mechanism (as it was previously argued), POGA should drive the entire population to converge towards S. At this stage, since sp_l is nil, POGA should move the population towards p_G along S. In order to ascertain what we have reasoned, it was decided to trace through an optimization run the values of the functional $\boldsymbol{g}(\boldsymbol{x}_M)$ and the distance d from the segment S (defined previously) as indirect measures of sp_f and sp_l respectively. Figure 9 shows the 25th, 50th and 75th percentile of $\boldsymbol{g}(\boldsymbol{x}_M)$ and d of a population of 100 solutions during an optimization run of POGA$_k$ on the test problem DTLZ5 with 5 objectives. As it is easy to appreciate, the pattern shown in Figure 9 is in complete accordance with the behavior that was previously predicted. At the early stage of the run, sp_f and sp_l equally drive the search and the population is uniformly converging towards the global Pareto front. It is interesting to note the increasing trend of d over the first few generations. This, as it was previously discussed, is the effect of the sampling error generated by the PO-based ranking scheme with a fixed population size. From about generation 100 until generation 170, POGA discovers an increasing number of solutions that lay on S (at generation 150 the 50th percentile of d drops to 0). From this stage until about generation 350, the frontal component of the selective pressure, i.e. sp_f, is weakened ($\boldsymbol{g}(\boldsymbol{x}_M)$ decreases at a lower rate), whereas sp_l, although decreasing, is still active (the 50th percentile of d often drops to 0). Around generation 350, sp_l is significantly reduced and the search almost progresses along S, towards p_G (the 25th percentile of $\boldsymbol{g}(\boldsymbol{x}_M)$ decreases) until generation 400, where the search stalls.

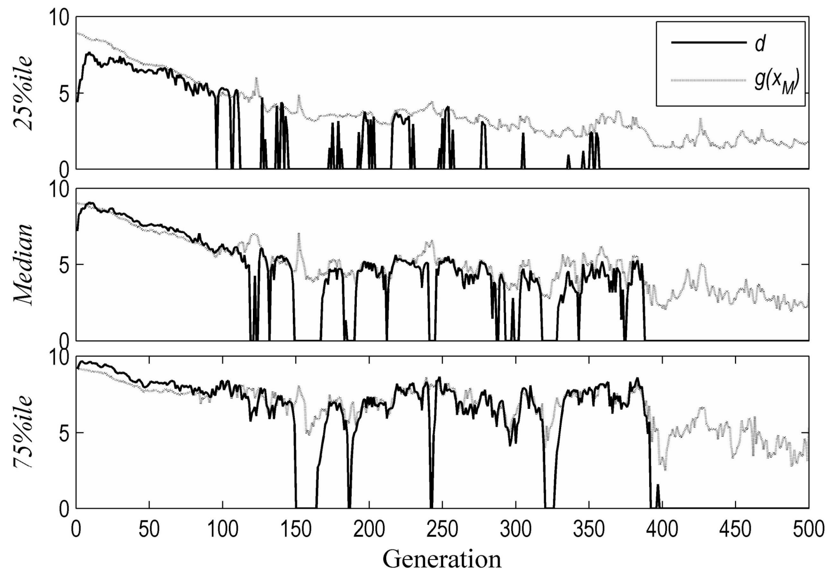


Figure 9: 25th, 50th and 75th percentile of the functional $\boldsymbol{g(x_M)}$ and the distance d from the segment S of a population of 100 solutions during an optimization run of POGA$_k$ on the test problem DTLZ5 with 5 objectives.

Figure 6 shows decreasing values of the diversity metric DM1 obtained by POGA on instances of DTLZ5 with increasing number of objectives. The reason for this behavior is that, as the number of objectives increases, the frontal component of the selective pressure, sp_f, decreases and the search becomes quickly driven solely by the lateral component, with a consequent loss of diversity and explorative capabilities.

Oscillating values of *CS* metric

The box plots of the Coverage Set metric (Figure 7) reveal that it is generally affected by an oscillatory behavior across the values of m (number of objective functions). This is the effect of a feature intrinsic to the definition of Preference Ordering, rather than the result of some random mechanism affecting the experiments. Figure 10 is a pictorial representation of the percentage of points in a m-dimensional Pareto efficient set that are efficient to the lowest order (the best points), for a variable number of objective spaces and cardinalities (N) of Pareto efficient sets (sets that consist of Pareto efficient points only). For these experiments, the same values for m (4, 5, 6, 7 and 8) and population size (20 to 80 with a step five of 6) that were used to evaluate the computation time of PO$_k$ (Figure 4) were chosen. Each box depicts median and inter-quartile range of the number of individuals of a Pareto efficient set that would be given the highest rank if classified by means of PO$_k$, that is according to their order of efficiency. For instance, if a set of 20 4-dimensional Pareto efficient points were classified according to PO$_k$, on average, only 2 points (10%) of the set would be given the highest rank. As it can be appreciated, a marked oscillatory behavior across different values of m is shown to affect median inter-quartile range. The latter, in particular, grows considerably when m and N approach the lowest and highest values, respectively, of the ranges considered. Roughly speaking, this signifies that for low-dimensional objective spaces (m = 4) the chances that none of the points of a Pareto efficient set is better than the others according to PO$_k$ is high. This confirms what shown in Figure 7, that is the high fluctuations of *CS* values for low number of objectives.

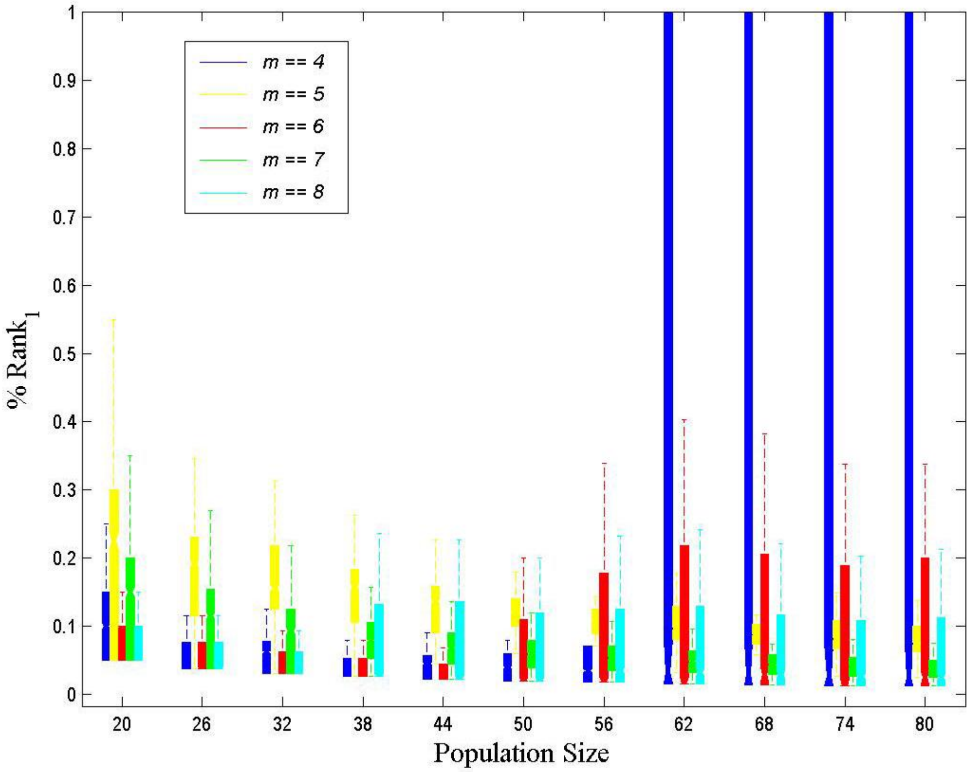


Figure 10: Percentage of points in an *m*-dimensional Pareto set that are efficient to the lowest order versus population size (*N*), for different dimensions of the objective space (*m*).

Figure 11 provides a different representation of the data presented in Figure 10 and it shows an interesting pattern: when the Pareto efficient set is even-dimensional, the inter-quartile range grows with the *N* (the cardinality of the set). Conversely, it shrinks when the Pareto efficient set is odd-dimensional. Instead, the median shows decreasing trend with *N*, irrespective of the number of dimensions of the objective space. Even though theoretical proves are necessitated to allow for the generalization over *N* and *m* values, it can be argued that the oscillating trend of inter-quartile ranges observed about the *CS* metric values across the number of objectives, can be partly attributed to this feature inherent in the definition of *efficiency of order* (PO_k).

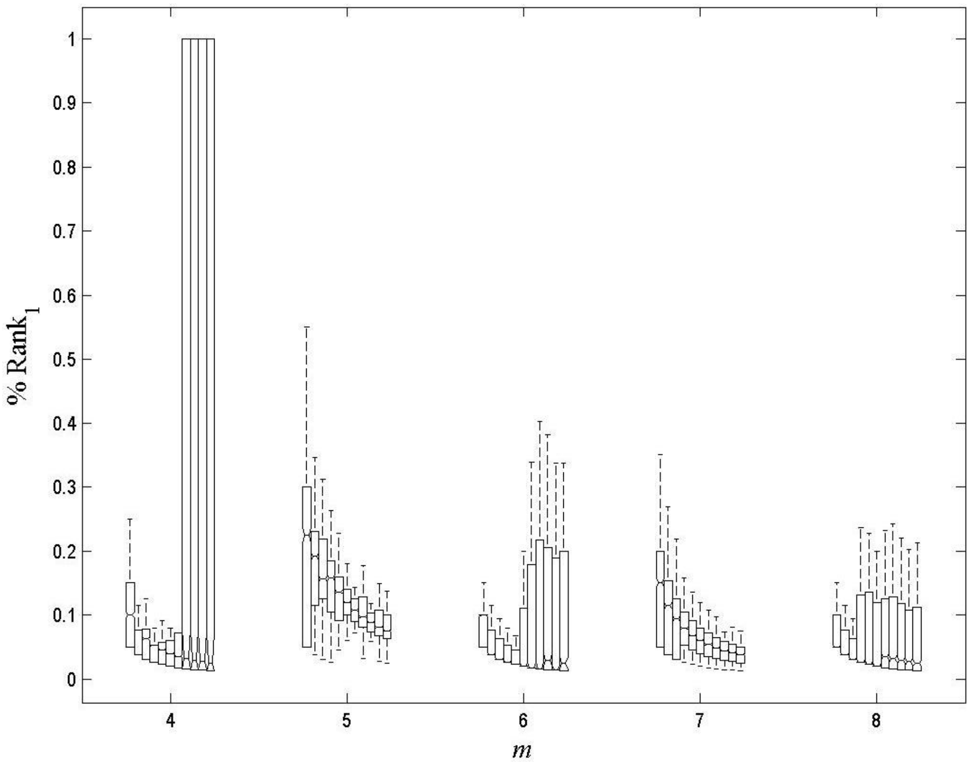


Figure 11: percentage of points in an m-dimensional Pareto efficient set that are efficient to the lowest order versus dimension of the objective space (m) for different population sizes (N).

CONCLUSIONS

This paper explored the potentials of Preference Ordering (PO), an optimality criterion more stringent than Pareto efficiency, as the basis for classifying solutions in the context of multi objective evolutionary search. It was submitted that, since it is less affected than Pareto efficiency by the progressive loss of effectiveness occurring as the dimension of the objective space of an MOP increases, it could greatly enhance the scalability properties of MOEA.

In order to substantiate this claim, two alternative ranking procedures that exploit the definition of PO were proposed, namely PO_k, and $PO_{k,z}$, and compared to a more traditional Pareto efficiency-based ranking scheme. To this end, a most acknowledged MOEA algorithm, NSGA-II was selected as optimization shell, geared with the three ranking schemes at stake and challenged in a series of experiments on four widely applied test functions (DTLZ1, DTLZ2, DTLZ3 and DTLZ5), for five different number of objectives, namely 4, 5, 6,7 and 8. Five performance metrics ï Generational Distance (*GD*), Hyper-volume (*Hn*), Diversity Metric 1 and 2 (*DM1* and *DM2*) and Coverage Set (*CS*) ï were used to highlight convergence and diversity preserving properties.

The results testify that the algorithms equipped with the novel ranking procedures attain better convergence to the true Pareto surface of all the test problems; furthermore, the performance gap shows a markedly increasing trend with the number of objectives considered. The analysis of the Coverage Set metric confirmed this pattern, which was also supported by the Games and Howell approximate test performed on *GD*, *Hn* and *CS* mean values.

Overall, the experiments also showed that PO_k, and $PO_{k,z}$ are less effective than the Pareto efficiency-based ranking scheme in maintaining a well diversified set of solutions over the Pareto surface. This is in fact quite expected, since the two levels of Preference Ordering that they exploit, favor sub-portions of a Pareto efficient set, and, for many engineering problems, it is not to be necessarily considered as a drawback [11]. Test problem DTLZ5, in particular, proved to greatly challenge the mechanism to preserve diversity of both $POGA_k$ and $POGA_{k,z}$. An in depth analysis revealed that the premature convergence occurring on this test problem was encouraged by the inherent inadequacy of the Crowding Distance density estimator in ensuring diversity amongst solutions in portions of the objective space favored by the PO-based ranking scheme. We therefore believe that MOEAs that have selective mechanisms conceptually different from that of NSGA-II, e.g. PESA and SPEA2, could benefit to a larger extent from the ranking procedure that we have proposed and this is a matter of our current research.

REFERENCES

[1] Coello Coello, C. A., Van Veldhuizen, D. A., and Lamont, G. B., (2002). *Evolutionary Algorithms for Solving Multi-Objective Problems*. New York: Kluwer Academic.

[2] Corne, D. W., Knowles, J. D., and Oates, M. J., (2000). "The Pareto Envelope-based Selection Algorithm for Multiobjective Optimization", in Proceedings of the Parallel Problem Solving from Nature {VI} Conference, pp. 839-848, M. Schoenauer and K. Deb and G. Rudolph and X. YAO and E. Lutton and J. J. Merelo and H. P. Schwefel Eds. Springer. Lecture Notes in Computer Science No. 1917.

[3] Das, I., (1999). "A Preference Ordering Among Various Pareto Optimal Alternatives", *Structural Optimization*, vol. 18(1):30-35.

[4] Das, I. and Dennis, J. E., (1998). "Normal Boundary Intersection: A New Method for generating the Pareto Surface in nonlinear multi objective optimization problems", *SIAM J. on Optimization*, vol. 8(3):631-657.

[5] de Jong, A. K., (1975). "An Analysis of the Behaviour of a Class of Genetic Adaptive Systems." Ph.D. thesis, University of Michigan.

[6] Deb, K., (2001). *Multi-Objective Optimization using Evolutionary Algorithms* John Wiley & Sons, Chichester, UK.

[7] Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T., (2002). "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II", *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION*, vol. 6(2):182-197.

[8] Deb, K., Agrawal, S., Pratap, A., and Meyarivan, T., (2000). "A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II", Technical Report, Kanpur Genetic Algorithm Laboratory (KanGAL), 200001.

[9] Deb, K. and Jain, S., (2002). "Running Performance Metrics for Multi-Objective Optimization", in Proceedings of the 4th Asia-Pacific Conference on Simulated Evolution

and Learning (SEAL '02), vol. 1, pp. 13-20, L. Wang and K. C. Tan and T. Furuhashi and J.-H. Kim and X. Yao Eds., Orchid Country Club, Singapore: Nanyang Technical University.

[10] Deb, K., Thiele, L., and Zitzler, E., (2002). "Scalable Multi-Objective Optimization Test Problems", in IEEE Congress on Evolutionary Computation, (CEC 2002), pp. 825-830, Piscataway (NJ): IEEE press.

[11] di Pierro, F., Djordjevic, S., Khu, S.-T., Savic, D., and Walters, G. A., (2004). "Automatic calibration of urban drainage model using a novel multi-objective genetic algorithm", in Urban Drainage Modelling (UDM'04), pp. 41-52, P. Krebs and L. Fuchs Eds.

[12] Farina, M. and Amato, P., (May2004). "A Fuzzy Definition of Optimality for Many-Criteria Optimization Problems", *IEEE Transactions on Systems, Man, and Cybernetics Part A---Systems and Humans*, vol. 34(3):315-326.

[13] Games, P. A. and Howell, J. F., (1976). "Pairwise multiple comparison procedures with unequal N's and/or variances: A Monte Carlo study", *Journal of Educational Statistics*, vol. 1:113-125.

[14] Horn, J., Nafpliotis, N., and Goldberg, D. E., (1994). "A Niched Pareto Genetic Algorithm for Multiobjective Optimization", in Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence, vol. 1, pp. 82-87, Piscataway, New Jersey: IEEE Service Center.

[15] Hu, J., Goodman, E. D., Seo, K., Fan, Z., and Rosenberg, R. C., (2003). "HFC: a Continuing EA Framework for Scalable Evolutionary Synthesis", in Computational Synthesis: From Basic Building Blocks to High Level Functionality, pp. 106-113.

[16] Khare, V., (2002). "PERFORMANCE SCALING OF MULTI-OBJECTIVE EVOLUTIONARY ALGORITHMS." MSc. School of Computer Science, The University of Birmingham, Birmingham , U.K.

[17] Khare, V., Yao, X., and Deb, K., (2003). "Performance Scaling of Multi-objective Evolutionary Algorithms", in Evolutionary Multi-Criterion Optimization.Second International Conference, EMO 2003, vol. 2632, pp. 376-390, C. M. Fonseca and P. J. Fleming and E. Zitzler and K. Deb and L. Thiele Eds. Springer.

[18] Khu, S. T. and Madsen, H., (2004). "Multi-objective calibration with Pareto preference ordering: An application to rainfall-runoff model calibration", *WATER RESOURCES RESEARCH*.

[19] Knowles, J. and Corne, D., (1999). "The Pareto Archived Evolution Strategy: A New Baseline Algorithm for Pareto Multiobjective Optimisation", in Proceedings of the 1999 Congress on Evolutionary Computation (CEC'99), vol. 1, pp. 98-105, P. J. Angeline and Z. Michalewicz and M. Schoenauer and X. Yao and A. Zalzala Eds. IEEE Press.

[20] Purshouse, R. C., (2003). "On the Evolutionary Optimisation of Many Objectives." Ph.D. thesis, Department of Automatic Control and Systems Engineering, The University of Sheffield.

[21] Reed, P., Minsker, B., and Goldberg, D. E., (2001). "The Practitioner's Role in Competent Search and Optimization Using Genetic Algorithms", in Bridging the Gap: Meeting the

World's Water and Environmental Resources Challenges, Don Phelps and Gerald Sehlke Ed., Washington, DC: American Society of Civil Engineers.

[22] Reed, P., Minsker, B., and Goldberg, D. E., (2003). "Simplifying multiobjective optimization: An automated design methodology for the nondominated sorted genetic algorithm-II", *WATER RESOURCES RESEARCH*, vol. 39(7).

[23] Srinivas, N. and Deb, K., (1995). "Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms", *Evolutionary Computation*, vol. 2(3):221-248.

[24] A. Trujillo-Ortiz and R. Hernandez-Walls, "GHtest: Games-Howell's approximate test of equality of means from normal population when variances are heterogeneous.," Mathworks, Matlab File Exchange, 2003.

[25] Zitzler, E., Laumanns, M., and Thiele, L., (2001). "SPEA2: Improving the Strength Pareto Evolutionary Algorithm", Swiss Federal Institute of Technology (ETH) Zurich, TIK, 103.

[26] Zitzler, E. and Thiele, L., (1999). "Multiobjective Evolutionary Algorithms: A Comparative Case study and the Strength Pareto Approach", *IEEE Transaction on Evolutionary Computation*, vol. 3(4):257-271.

[27] Zitzler, E., Thiele, L., and Deb, K., (2000). "Comparison of Multiobjective Evolutionary Algorithms: Empirical Results", *Evolutionary Computation*, vol. 8(2):173-195.

APPENDIX

Table A - I: Mean and standard deviation of the metrics *GD* and *Hn*. The formula to compute the *GD* metric for each test problem is also reported, with *N* and *m* being the population size and number of objectives respectively. Mean and standard deviation are evaluated over the last generation population of the ten runs performed for each problem instance.

Test	Metric	*m*	NSGA-II	$POGA_k$	$POGA_{k,z}$	NSGA-II	$POGA_k$	$POGA_{k,z}$
			mean			*std*		
DTLZ1	$GD = E_N\left[\sum_{i=1}^{m} f_i - 0.5\right]$	4	4.58E+00	3.37E+00	3.94E+00	2.63E+00	1.96E+00	1.16E+00
		5	2.49E+01	1.43E+01	1.73E+01	1.20E+01	7.80E+00	6.63E+00
		6	2.18E+02	1.05E+02	1.03E+02	2.34E+01	2.21E+01	3.46E+01
		7	3.22E+02	2.17E+02	2.25E+02	3.39E+01	1.99E+01	2.23E+01
		8	3.72E+02	2.96E+02	2.84E+02	1.68E+01	5.83E+01	7.83E+01
	Hn	4	8.44E-01	8.92E-01	8.70E-01	1.04E-01	8.22E-02	6.69E-02
		5	3.54E-01	5.29E-01	5.52E-01	2.97E-01	3.76E-01	2.42E-01
		6	3.01E-01	8.24E-01	8.21E-01	2.82E-01	6.94E-02	7.79E-02
		7	3.13E-02	7.03E-01	8.33E-01	9.90E-02	2.90E-01	1.29E-01
		8	0.00E+00	9.09E-01	9.29E-01	0.00E+00	7.54E-02	3.81E-02
DTLZ2	$GD = E_N\left[\sum_{i=1}^{m} f_i^2 - 1\right]$	4	4.69E-02	3.98E-02	4.40E-02	1.44E-02	1.04E-02	2.00E-02
		5	1.45E-01	1.03E-01	1.09E-01	3.81E-02	1.64E-02	3.27E-02
		6	4.35E-01	1.96E-01	2.29E-01	1.16E-01	4.11E-02	4.07E-02
		7	8.64E-01	2.71E-01	3.21E-01	1.78E-01	1.01E-01	9.54E-02
		8	1.51E+00	2.76E-01	3.06E-01	2.24E-01	7.06E-02	1.21E-01
	Hn	4	5.92E-01	5.94E-01	5.95E-01	7.18E-03	7.32E-03	8.60E-03

		5	6.40E-01	6.47E-01	6.48E-01	7.22E-03	1.30E-02	1.14E-02
		6	6.35E-01	6.82E-01	6.80E-01	1.91E-02	9.10E-03	5.98E-03
		7	6.12E-01	7.04E-01	7.03E-01	2.39E-02	3.12E-02	2.32E-02
		8	5.84E-01	7.27E-01	7.24E-01	1.86E-02	2.34E-02	1.92E-02
DTLZ3	$GD \triangleq \underset{N}{E}\left[\sum_{i=1}^{m} f_i^2 - 1\right]$	4	1.05E+03	7.95E+02	8.05E+02	1.65E+03	8.88E+02	7.62E+02
		5	3.47E+03	2.05E+03	2.34E+03	4.93E+03	1.78E+03	2.94E+03
		6	6.46E+04	1.55E+04	1.24E+04	2.51E+04	8.27E+03	6.19E+03
		7	3.96E+05	6.38E+04	6.02E+04	5.96E+04	3.35E+04	2.23E+04
		8	8.25E+05	1.22E+05	9.20E+04	1.19E+05	4.54E+04	3.39E+04
	Hn	4	9.24E-01	9.33E-01	9.16E-01	3.12E-02	3.92E-02	5.85E-02
		5	8.93E-01	9.36E-01	9.21E-01	7.08E-02	2.86E-02	4.40E-02
		6	4.72E-01	8.46E-01	8.42E-01	4.20E-01	6.28E-02	6.46E-02
		7	9.29E-02	8.55E-01	8.54E-01	2.94E-01	4.59E-02	5.43E-02
		8	8.65E-02	9.00E-01	8.95E-01	2.73E-01	3.84E-02	4.70E-02
DTLZ5	$GD \triangleq \underset{N}{E}\left[\sum_{i=1}^{m} f_i^2 - 1\right]$	4	4.10E+01	3.87E+01	3.93E+01	3.75E+00	4.06E+00	2.98E+00
		5	6.22E+01	3.07E+01	2.69E+01	3.09E+00	9.90E+00	9.20E+00
		6	7.61E+01	3.28E+01	3.57E+01	4.02E+00	6.28E+00	7.38E+00
		7	9.77E+01	5.33E+01	4.29E+01	3.67E+00	5.17E+00	8.39E+00
		8	9.88E+01	5.71E+01	5.72E+01	2.08E+00	5.65E+00	3.09E+00
	Hn	4	7.61E-01	7.52E-01	7.68E-01	5.84E-02	5.70E-02	6.30E-02
		5	6.22E+01	8.91E-01	9.20E-01	3.09E+00	9.02E-02	6.48E-02
		6	6.58E-01	9.60E-01	9.60E-01	3.85E-02	0.00E+00	0.00E+00
		7	1.70E-01	8.96E-01	9.16E-01	1.14E-01	3.16E-02	1.67E-02
		8	4.84E-01	9.09E-01	9.05E-01	1.97E-01	2.68E-02	2.24E-02

Table A - II: Mean and standard deviation of the diversity metrics *DM1* and *DM2*. Mean and standard deviation are evaluated over the last generation population of the ten runs performed for each problem instance.

Test	Metric	*m*	NSGA-II	$POGA_k$	$POGA_{k,z}$	NSGA-II	$POGA_k$	$POGA_{k,z}$
			mean			*std*		
DTLZ1	*DM1*	4	0.22480	0.26970	0.23709	0.03830	0.08526	0.05367
		5	0.19458	0.23685	0.22633	0.03455	0.05474	0.04175
		6	0.08563	0.14279	0.15752	0.02159	0.02480	0.03224
		7	0.04937	0.05960	0.05981	0.03290	0.02733	0.03721
		8	0.05233	0.05306	0.05681	0.01618	0.04746	0.05370
	DM2	4	0.47605	0.49587	0.49442	0.06376	0.10023	0.06793
		5	0.42246	0.45692	0.43257	0.05877	0.06899	0.05008
		6	0.47136	0.42479	0.39593	0.01633	0.05821	0.03931
		7	0.46801	0.46849	0.42807	0.02730	0.03237	0.02644
		8	0.47439	0.45257	0.45015	0.02601	0.04112	0.07814
DTLZ2	*DM1*	4	0.48457	0.49406	0.49505	0.03035	0.03071	0.04069
		5	0.48445	0.47251	0.46996	0.03900	0.02362	0.02012
		6	0.48821	0.45080	0.44697	0.02240	0.02956	0.04099
		7	0.44698	0.39623	0.41346	0.02904	0.03695	0.04737
		8	0.42217	0.36659	0.36162	0.02025	0.03583	0.03057
	DM2	4	0.51242	0.51737	0.50261	0.02441	0.02822	0.02902
		5	0.53638	0.51655	0.51611	0.02948	0.03116	0.01395
		6	0.54691	0.50193	0.49903	0.01914	0.02054	0.03333

		7	0.53857	0.46381	0.48106	0.01894	0.02180	0.03572
		8	0.53499	0.44189	0.44414	0.02447	0.02496	0.03590
DTLZ3	*DM1*	4	0.07869	0.05769	0.07177	0.07560	0.03112	0.03343
		5	0.12180	0.13492	0.10047	0.10513	0.03175	0.03075
		6	0.06378	0.07426	0.10495	0.08691	0.02028	0.02533
		7	0.01494	0.07029	0.06658	0.00931	0.02299	0.02674
		8	0.00520	0.05534	0.04578	0.00708	0.01908	0.01389
	DM2	4	0.45188	0.41613	0.39814	0.06918	0.09938	0.09018
		5	0.37356	0.36862	0.35106	0.04849	0.04581	0.07551
		6	0.37567	0.29961	0.29979	0.02506	0.03753	0.02192
		7	0.47977	0.30938	0.29921	0.01912	0.04132	0.04490
		8	0.49951	0.30462	0.28994	0.02344	0.03769	0.05148
DTLZ5	*DM1*	4	0.28187	0.25516	0.24322	0.03007	0.01675	0.02321
		5	0.41159	0.08981	0.02073	0.03722	0.14086	0.06554
		6	0.40493	0.00000	0.00000	0.02494	0.00000	0.00000
		7	0.49720	0.00000	0.00000	0.02462	0.00000	0.00000
		8	0.42254	0.00000	0.00000	0.01520	0.00000	0.00000
	DM2	4	0.37650	0.38078	0.36694	0.02100	0.03028	0.04180
		5	0.44844	0.36754	0.29556	0.03966	0.12403	0.16271
		6	0.47222	0.39953	0.40992	0.03078	0.03331	0.02112
		7	0.53761	0.50002	0.42802	0.01639	0.03891	0.03794
		8	0.49166	0.50633	0.50183	0.02361	0.03394	0.04873

Games and Howell approximate test

The Games and Howell approximate test is used assess whether the means of two or more distributions, from which a series of samples are drawn, are Significantly (S) or Non Significantly (NS) different when the variances are heterogeneous. It is a test associated to the Behrens-Fisher problem and it uses the Tukey's studentized range with specially weighted average degrees of freedom (df ') and a standard error based on the averages of the variances of the means. In this study, the Games and Howell approximate test is used to compare the mean values of the convergence metrics, namely *GD* and *Hn*, and Coverage Set (*CS*) metric attained by the algorithms NSGA-II, POGA$_k$ and POGA$_{k,z}$. The results presented in the following tables were generated by the open source Matlab file [24], for a significance level Ŭ= 0.05.

Table A - III: Games-Howell's approximate test of equality of means of convergence metrics, that is *GD* and *Hn*. Mean difference, (D), standard error (q), weighted degrees of freedom (df ') conclusion of the test (Significant and Non Significant) are shown respectively, for a significance Ŭ= 0.05 (5%).

Test Problem	*m*	Code Pair	Performance Metric							
			GD				*Hn*			
			D	q	df '	Conclusion	D	q	df '	Conclusion
DTLZ1	4	3-2	0.57	0.79	14	NS	0.02	0.66	17	NS
		3-1	0.64	0.70	12	NS	0.03	0.66	15	NS
		2-1	1.21	1.17	16	NS	0.05	1.15	17	NS
	5	3-2	3.00	0.93	17	NS	0.02	0.16	15	NS
		3-1	7.60	1.75	14	NS	0.20	1.63	17	NS
		2-1	10.60	2.34	15	NS	0.18	1.16	17	NS
	6	3-2	2.00	0.15	15	NS	0.00	0.09	17	NS
		3-1	115.00	8.71	15	S	0.52	5.62	10	S

		2-1	113.00	11.10	17	S	0.52	5.69	10	S
		3-2	8.00	0.85	17	NS	0.13	1.30	12	NS
	7	3-1	97.00	7.56	15	S	0.80	15.59	16	S
		2-1	105.00	8.45	14	S	0.67	6.93	11	S
		3-2	12.00	0.39	16	NS	0.02	0.75	13	NS
	8	3-1	88.00	3.47	9	NS	0.93	77.11	9	S
		2-1	76.00	3.96	10	S	0.91	38.12	9	S
		3-2	0.00	0.59	13	NS	0.00	0.28	17	NS
	4	3-1	0.00	0.37	16	NS	0.00	0.85	17	NS
		2-1	0.01	1.26	16	NS	0.00	0.62	17	NS
		3-2	0.01	0.52	13	NS	0.00	0.18	17	NS
	5	3-1	0.04	2.27	17	NS	0.01	1.87	15	NS
		2-1	0.04	3.20	12	NS	0.01	1.49	14	NS
		3-2	0.03	1.80	17	NS	0.00	0.58	15	NS
DTLZ2	6	3-1	0.21	5.30	11	S	0.05	7.11	10	S
		2-1	0.24	6.14	11	S	0.05	7.02	12	S
		3-2	0.05	1.14	17	NS	0.00	0.08	16	NS
	7	3-1	0.54	8.50	13	S	0.09	8.64	17	S
		2-1	0.59	9.16	14	S	0.09	7.40	16	S
		3-2	0.03	0.68	14	NS	0.00	0.31	17	NS
	8	3-1	1.20	14.96	13	S	0.14	16.56	17	S
		2-1	1.23	16.62	10	S	0.14	15.13	17	S
		3-2	10.00	0.03	17	NS	0.02	0.76	15	NS
	4	3-1	245.00	0.43	12	NS	0.01	0.38	13	NS
		2-1	255.00	0.43	13	NS	0.01	0.57	17	NS
		3-2	290.00	0.27	14	NS	0.02	0.90	15	NS
	5	3-1	1130.00	0.62	14	NS	0.03	1.06	15	NS
		2-1	1420.00	0.86	11	NS	0.04	1.78	11	NS
		3-2	3100.00	0.95	16	NS	0.00	0.14	17	NS
DTLZ3	6	3-1	52200.00	6.39	10	S	0.37	2.75	9	NS
		2-1	49100.00	5.88	10	S	0.37	2.79	9	NS
		3-2	3600.00	0.28	15	NS	0.00	0.04	17	NS
	7	3-1	335800.00	16.69	11	S	0.76	8.05	9	S
		2-1	332200.00	15.37	14	S	0.76	8.10	9	S
		3-2	30000.00	1.67	16	NS	0.01	0.26	17	NS
	8	3-1	733000.00	18.73	10	S	0.81	9.23	9	S
		2-1	703000.00	17.45	11	S	0.81	9.33	9	S
		3-2	0.60	0.38	16	NS	0.02	0.60	17	NS
	4	3-1	1.70	1.12	17	NS	0.01	0.26	17	NS
		2-1	2.30	1.32	17	NS	0.01	0.35	17	NS
		3-2	3.80	0.89	17	NS	0.03	0.83	16	NS
	5	3-1	35.30	11.50	11	S	0.86	0.88	9	NS
		2-1	31.50	9.60	10	S	0.83	0.85	9	NS
		3-2	2.90	0.95	17	NS	-[+]	-[+]	-[+]	-[+]
DTLZ5	6	3-1	40.40	15.20	13	S	0.30	24.81	9.00	S
		2-1	43.30	18.36	15	S	0.30	24.81	9.00	S
		3-2	10.40	3.34	14	S	0.02	1.77	13	NS
	7	3-1	54.80	18.92	12	S	0.75	20.48	9	S
		2-1	44.40	22.15	16	S	0.73	19.41	10	S
		3-2	0.10	0.05	13	NS	0.00	0.36	17	NS
	8	3-1	41.60	35.32	15	S	0.42	6.71	9	S
		2-1	41.70	21.90	11	S	0.43	6.76	9	S

Code	1	2	3
Algorithm	NSGAII	$POGA_k$	$POGA_{k,z}$
+	Test can not be performed because of nil variance		

Table A - IV: Games-Howell's approximate test of equality of means of Coverage Set metric (*CS*). Mean difference (D), standard error (q), weighted degrees of freedom (df ') and conclusion of the test (Significant and Non Significant) are shown respectively, for a significance Ŭ= 0.05 (5%).

m	**Code Pairs**	**Performance Metric**							
		CS							
		DTLZ1				DTLZ2			
		D	q	df '	Conclusion	D	q	df '	Conclusion
4	6-5	0.01	0.09	17	NS	0.00	0.29	17	NS
	6-4	0.16	1.03	17	NS	0.00	0.46	17	NS
	5-3	0.03	0.19	16	NS	0.00	0.50	16	NS
	5-2	0.05	0.35	17	NS	0.01	1.89	17	NS
	4-3	0.14	0.85	17	NS	0.00	0.08	16	NS
	4-2	0.22	1.54	17	NS	0.01	1.17	17	NS
	3-1	0.18	1.26	14	NS	0.00	0.26	14	NS
	2-1	0.11	0.92	16	NS	0.01	1.73	17	NS
5	6-5	0.01	0.20	17	NS	0.00	0.48	17	NS
	6-4	0.20	2.46	14	NS	0.04	5.40	11	S
	5-3	0.12	1.62	14	NS	0.01	1.77	16	NS
	5-2	0.13	1.82	15	NS	0.03	3.00	11	NS
	4-3	0.07	0.70	17	NS	0.03	3.74	16	NS
	4-2	0.06	0.66	17	NS	0.01	1.09	16	NS
	3-1	0.13	1.79	14	NS	0.01	0.92	14	NS
	2-1	0.14	1.99	15	NS	0.02	2.50	10	NS
6	6-5	0.00	0.76	14	NS	0.00	0.04	16	NS
	6-4	0.58	7.54	9	S	0.11	5.01	9	NS
	5-3	0.14	4.49	9	NS	0.02	2.18	9	NS
	5-2	0.62	7.55	9	S	0.08	6.60	9	S
	4-3	0.44	5.29	11	S	0.09	3.52	13	NS
	4-2	0.03	0.31	17	NS	0.03	1.02	14	NS
	3-1	0.02	0.39	15	NS	0.01	0.50	14	NS
	2-1	0.45	4.73	14	S	0.07	4.72	13	S
7	6-5	0.00	0.45	13	NS	0.00	1.00	9	NS
	6-4	0.47	12.94	9	S	0.16	6.61	9	S
	5-3	0.12	2.47	9	NS	0.01	2.65	12	NS
	5-2	0.56	10.46	9	S	0.16	10.50	9	S
	4-3	0.35	5.65	16	S	0.15	5.87	9	S
	4-2	0.09	1.43	15	NS	0.00	0.08	14	NS
	3-1	0.05	0.62	16	NS	0.02	2.47	14	NS
	2-1	0.39	4.62	17	S	0.12	7.16	13	S
8	6-5	0.00	1.00	9	NS	-[+]	-[+]	-[+]	-[+]
	6-4	0.48	11.21	9	S	0.25	8.09	9	S
	5-3	0.23	3.67	9	NS	0.02	4.03	9	NS
	5-2	0.51	8.33	9	S	0.26	13.19	9	S
	4-3	0.25	3.26	15	NS	0.23	7.51	9	S
	4-2	0.03	0.36	16	NS	0.01	0.25	15	NS
	3-1	0.04	0.40	16	NS	0.01	0.96	14	NS
	2-1	0.32	3.01	16	NS	0.23	11.38	10	S

		DTLZ3				DTLZ5			
		D	q	df '	Conclusion	D	q	df '	Conclusion
4	6-5	0.14	0.98	15	NS	0.04	1.15	12	NS
	6-4	0.10	0.61	17	NS	0.13	2.82	10	NS
	5-3	0.03	0.23	16	NS	0.04	0.78	17	NS
	5-2	0.31	2.29	16	NS	0.19	2.93	12	NS
	4-3	0.21	1.25	17	NS	0.13	2.32	16	NS
	4-2	0.07	0.41	17	NS	0.02	0.26	17	NS
	3-1	0.16	1.09	17	NS	0.03	0.67	17	NS
	2-1	0.12	0.78	17	NS	0.12	1.84	13	NS
5	6-5	0.01	0.19	17	NS	0.01	1.25	9	NS
	6-4	0.26	3.03	15	NS	0.34	6.86	9	S
	5-3	0.09	1.09	15	NS	0.02	1.31	9	NS
	5-2	0.32	3.20	13	NS	0.28	6.07	9	S
	4-3	0.15	1.51	17	NS	0.33	6.47	10	S
	4-2	0.07	0.59	17	NS	0.06	0.94	17	NS
	3-1	0.05	0.39	17	NS	0.01	0.31	13	NS
	2-1	0.18	1.40	17	NS	0.26	4.98	13	S
6	6-5	0.01	1.47	11	NS	-[+]	-[+]	-[+]	-[+]
	6-4	0.71	17.47	9	S	0.21	5.87	9	S
	5-3	0.17	3.73	9	NS	-[+]	-[+]	-[+]	-[+]
	5-2	0.77	30.02	9	S	0.21	5.87	9	S
	4-3	0.55	8.79	17	S	0.21	5.87	9	S
	4-2	0.05	1.06	15	NS	-[+]	-[+]	-[+]	-[+]
	3-1	0.05	0.80	17	NS	-[+]	-[+]	-[+]	-[+]
	2-1	0.54	9.92	13	S	0.21	5.87	9	S
7	6-5	-[+]	-[+]	-[+]	-[+]	-[+]	-[+]	-[+]	-[+]
	6-4	0.90	24.74	9	S	0.38	17.82	9	S
	5-3	0.14	5.29	9	S	-[+]	-[+]	-[+]	-[+]
	5-2	0.94	48.88	9	S	0.42	16.72	9	S
	4-3	0.76	17.01	16	S	0.38	17.82	9	S
	4-2	0.03	0.79	13	NS	0.04	1.25	17	NS
	3-1	0.08	1.34	12	NS	-[+]	-[+]	-[+]	-[+]
	2-1	0.72	12.44	11	S	0.42	16.72	9	S
8	6-5	-[+]	-[+]	-[+]	-[+]	-[+]	-[+]	-[+]	-[+]
	6-4	0.92	67.06	9	S	0.22	4.17	9	NS
	5-3	0.16	4.04	9	NS	-[+]	-[+]	-[+]	-[+]
	5-2	0.95	72.14	9	S	0.18	5.39	9	S
	4-3	0.77	18.56	11	S	0.22	4.17	9	NS
	4-2	0.02	1.26	17	NS	0.04	0.57	15	NS
	3-1	0.00	0.09	17	NS	-[+]	-[+]	-[+]	-[+]
	2-1	0.79	18.81	10	S	0.18	5.39	9	S

Code	1	2	3	4	5	6
Metric	$POGA_{k,z}$-$POGA_k$	$POGA_{k,z}$-NSGAII	$POGA_k$-$POGA_{k,z}$	$POGA_k$-NSGAII	NSGAII-$POGA_{k,z}$	NSGAII-$POGA_k$

[+] Test can not be performed because of nil variance

View publication stats