Impact of selection methods on the diversity of many-objective Pareto set approximations

Article *in* Procedia Computer Science · September 2017

DOI: 10.1016/j.procs.2017.08.077

4 authors:

Luis Martí
Universidade Federal Fluminense

Eduardo Segredo
Edinburgh Napier University

Nayat Sánchez-Pi
Rio de Janeiro State University

Emma Hart
Edinburgh Napier University

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 112 (2017) 844–853

Procedia Computer Science

www.elsevier.com/locate/procedia

21st International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France

Impact of selection methods on the diversity of many-objective Pareto set approximations

Luis Martí[a,b,*], Eduardo Segredo[c,d], Nayat Sánchez-Pi[e], Emma Hart[c]

[a] *Institute of Computing, Universidade Federal Fluminense, Niterói (RJ) Brazil*
[b] *TAU team, CNRS/INRIA/LRI, Université Paris-Saclay, Paris, France*
[c] *School of Computing, Edinburgh Napier University, Edinburgh, Scotland, UK*
[d] *Dpto. Ingeniería Informática y de Sistemas, Universidad de La Laguna, San Cristóbal de La Laguna, Spain*
[e] *DICC/IME, Rio de Janeiro State University, Rio de Janeiro (RJ) Brazil*

Abstract

Selection methods are a key component of all multi-objective and, consequently, many-objective optimisation evolutionary algorithms. They must perform two main tasks simultaneously. First of all, they must select individuals that are as close as possible to the Pareto optimal front (convergence). Second, but not less important, they must help the evolutionary approach to provide a diverse population. In this paper, we carry out a comprehensive analysis of state-of-the-art selection methods with different features aimed to determine the impact that this component has on the diversity preserved by well-known multi-objective optimisers when dealing with many-objective problems. The algorithms considered herein, which incorporate Pareto-based and indicator-based selection schemes, are analysed through their application to the Walking Fish Group (WFG) test suite taking into account an increasing number of objective functions. Algorithmic approaches are assessed via a set of performance indicators specifically proposed for measuring the diversity of a solution set, such as the Diversity Measure and the Diversity Comparison Indicator. Hypervolume, which measures convergence in addition to diversity, is also used for comparison purposes. The experimental evaluation points out that the reference-point-based selection scheme of the Non-dominated Sorting Genetic Algorithm III (NSGA-III) and a modified version of the Non-dominated Sorting Genetic Algorithm II (NSGA-II), where the crowding distance is replaced by the Euclidean distance, yield the best results.

Keywords: multi-objective evolutionary algorithm; many-objective optimisation; selection mechanism; diversity preservation

1. Introduction

Multi-objective Optimisation Problems (MOPs) are those problems where several conflicting objective functions must be optimised simultaneously. MOPs with more than three objective functions are usually known as *Many-objective Optimisation Problems* (MaOPs) in the related literature [1]. The solution to MOPs and, consequently, MaOPs

* Corresponding author.
E-mail address: lmarti@ic.uff.br

1877-0509 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of KES International
10.1016/j.procs.2017.08.077

is a set of trade-off points referred to as the *Pareto optimal front* or *Pareto optimal set*. If a particular MOP satisfies a set of requirements, e. g., linearity or convexity of the objective functions or convexity of the feasible set, then the Pareto optimal set can be determined by mathematical programming approaches [2]. In the general case however, finding the solution of a MOP is an *NP*–complete problem [3]. As a result, heuristic or meta-heuristic methods, such as *Multi-objective Evolutionary Algorithms* (MOEAs) [4], have arisen as one of the most popular techniques to successfully address MOPs.

One key component of MOEAs is the selection scheme, i.e., the mechanism used to select individuals that will survive, and which is responsible for both convergence and diversity of the solution set provided. A significant number of MOEAs incorporates a Pareto-based selection scheme, which usually considers two separate selection criteria [5]. First, individuals are ranked by applying Pareto optimality, thus giving preference to those individuals that are non-dominated. Second, a diversity-based selection criterion is also applied to distinguish individuals belonging to the same rank. Although Pareto-based MOEAs have proven to be successful optimisers for a wide variety of MOPs, it has been recently demonstrated that Pareto-based selection is not suitable for MaOPs. One of the main reasons is that the number of non-dominated individuals exponentially increases with the number of objective functions. In this scenario, the selection scheme becomes inaccurate when ranking individuals and the selection pressure of the whole MOEA diminishes. Hence, individuals selected to survive may not be close enough to the Pareto optimal front due to the lack of convergence of the approach.

Two main paths have been explored in order to improve the performance of Pareto-based MOEAs when tackling MaOPs [5]. The first one is to propose novel definitions of dominance, such as dominance area control [6] and L-optimality [7], which allow the selection pressure of the MOEA to be increased. The second option focuses on improving or replacing the diversity-based selection criterion. In addition to the aforementioned options, different types of MOEAs have been proposed for solving MaOPs as an alternative to Pareto-based MOEAs [5]: *decomposition-based* MOEAs, *grid-based* MOEAs and *indicator-based* MOEAs.

Bearing all the above in mind, the main contribution of the current work is a comprehensive study about the impact that selection mechanisms have on the diversity preserved by MOEAs when dealing with MaOPs. For doing that, well-known MOEAs, which incorporate selection mechanisms with different features, are applied to MaOPs with a scalable number of objectives belonging to the *Walking Fish Group* (WFG) test suite [8]. The way in which the performance of MOEAs can be measured has arisen as an important research area [9]. As a result, a considerable number of quality indicators, like the *hypervolume indicator* [10] or the *Diversity Comparison Indicator* [9], has been proposed for measuring either convergence or diversity or both of them. In this work, we will focus on quality indicators specifically designed for measuring the diversity of a solution set.

The rest of this paper is organised as follows. Section 2 is devoted to describe all the foundations related to the work carried out herein, including the formal definition of a MOP, the particular MOEAs we have considered for our study, as well as their selection mechanisms. The quality indicators selected to evaluate the diversity of the solution sets provided by those MOEAs are also introduced. Then, the computational experiments performed, as well as the results obtained, are shared in Section 3. Finally, Section 4 gives some conclusions and lines of future work.

2. Foundations

A *Multi-objective Optimisation Problem* (MOP) can be defined as the problem in which a set of *objective functions* $f_1(\boldsymbol{x}), \ldots, f_M(\boldsymbol{x})$ should be jointly optimised;

$$\min \; \boldsymbol{F}(\boldsymbol{x}) = \langle f_1(\boldsymbol{x}), \ldots, f_M(\boldsymbol{x}) \rangle \, ; \; \boldsymbol{x} \in \mathcal{S} \, ; \tag{1}$$

where $\mathcal{S} \subseteq \mathbb{R}^n$ is known as the *feasible set* and can be expressed as a set of restrictions over the *decision set*, in our case, $\mathbb{R}^n$. The *image set* of $\mathcal{S}$ produced by function vector $\boldsymbol{F}(\cdot)$, i.e., $\mathcal{O} \subseteq \mathbb{R}^M$ is called the *feasible objective set* or *criterion set*. The solution to these types of problems is a set of trade-off points. The optimality of a given solution can be defined in terms of the Pareto dominance relation.

Definition 1 (Pareto dominance relation)**.** *For the optimisation problem specified in (1) and having $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$, $\boldsymbol{x}$ is said to dominate $\boldsymbol{y}$ (expressed as $\boldsymbol{x} \prec \boldsymbol{y}$) iff $\forall f_j$, $f_j(\boldsymbol{x}) \leq f_j(\boldsymbol{y})$ and $\exists f_i$ such that $f_i(\boldsymbol{x}) < f_i(\boldsymbol{y})$.*

Definition 2 (Non-dominated subset)**.** *In problem (1) and having the set $\mathcal{A} \subseteq \mathcal{S}$, $\hat{\mathcal{A}}$, the non-dominated subset of $\mathcal{A}$, is defined as*

$$\widehat{\mathcal{A}} = \left\{ \boldsymbol{x} \in \mathcal{A} \middle| \nexists \boldsymbol{x}' \in \mathcal{A} : \boldsymbol{x}' \prec \boldsymbol{x} \right\}.$$

The solution of (1) is $\hat{\mathcal{S}}$, the *non-dominated subset* of $\mathcal{S}$. $\hat{\mathcal{S}}$ is known as the *efficient set* or *Pareto-optimal set*[2]. The *Pareto-optimal front*, $\hat{\mathcal{O}}$, is the image of $\hat{\mathcal{S}}$ in the feasible objective set.

2.1. Assessing diversity of a Pareto set approximation

Several are the quality indicators that have been proposed for measuring different aspects related to the shape of a Pareto set approximation, mainly, convergence, spread and uniformity*. From the above three aspects, we note that both spread and uniformity, which are closely related, determine the diversity of a Pareto set approximation. This section is devoted to briefly describe some well-known quality indicators that we apply herein as performance metrics. We have selected the *Diversity Measure* (Section 2.1.1) and the *Diversity Comparison Indicator* (Section 2.1.2) for our analyses because they focus on diversity according to the taxonomy proposed by Li et al.[9]. The *hypervolume* (Section 2.1.3) has also been chosen, since it not only focuses on diversity but also on convergence and is one of the most frequently used indicators to assess the performance of MOEAs.

2.1.1. Diversity Measure (DM)

The *Diversity Measure* (DM) was proposed to calculate the amount of diversity of a Pareto set approximation[11]. For doing that, it considers a reference set. The solutions belonging to both the reference set and the approximation are assigned to different grids or divisions of a hyperplane with $M - 1$ dimensions, with M being the number of objectives of the problem at hand. Those divisions are called *hyper-boxes*. The number of solutions assigned to a particular hyper-box, as well as the number of solutions assigned to its neighbour hyper-boxes, are used to evaluate that hyper-box. The larger the number of hyper-boxes containing solutions that belong to both the reference set and the approximation, the larger the indicator value, which is in the range $[0, 1]$. Bearing the above in mind, the value one indicates that maximum diversity has been reached. In this case, every solution of the approximation has been assigned to a hyper-box containing solutions of the reference set. At the same time, a DM value equal to zero means that the Pareto front approximation is not diverse at all, since no point belonging to the former has been assigned to a hyper-box which contains solutions of the reference set. According to Li et al.[9], DM has several disadvantages when dealing with MaOPs:

1. **A reference set with solutions uniformly distributed in the Pareto front is required**. Providing a reference set is an arduous task, even more in the case of many-objective optimisation. Furthermore, the number of solutions in the said reference set should be approximately equal to the number of solutions in the Pareto set approximation.
2. **A distribution estimation has to be calculated for each hyper-box**. There exist r^{M-1} hyper-boxes, with r being the number of divisions in every dimension.
3. **A value function has to be computed for each neighbour hyper-box** when estimating the distribution of a given hyper-box. A particular hyper-box has $3^M - 1$ neighbours. As a result, computing that value function may be challenging in the case of dealing with MaOPs.
4. **An inaccurate value of the approximation diversity could be provided**, since the Manhattan distance is taken into account instead of the Euclidean distance.

2.1.2. Diversity Comparison Indicator (DCI)

The great majority of those indicators aimed to measure the amount of diversity of a Pareto set approximation are not suitable for problems with a large number of objective functions[9]. The *Diversity Comparison Indicator* (DCI) was specifically proposed to measure the diversity of many-objective Pareto front approximations[9]. It tries to solve the aforementioned drawbacks that arise when metrics, such as DM, deal with MaOPs. DCI takes different Pareto front

* Since the shape of Pareto set approximations are taken into consideration, quality indicators are usually defined by considering the objective function space rather than the decision variable space.

approximations and assesses their relative contribution to diversity instead of calculating the absolute contribution of a unique Pareto set approximation.

It considers a grid environment, which consists of a set of hyper-boxes, where the solutions belonging to the approximations are distributed. Only nonempty hyper-boxes, i.e., hyper-boxes where one or more non-dominated solutions belonging to the mixed set of approximations have been assigned, are taken into account by DCI to calculate the contribution of each approximation. Hence, given a particular approximation, if its solutions are assigned or are close to the majority of the nonempty hyper-boxes, then its contribution to diversity will be significant in comparison to the contribution of the other Pareto set approximations. If its solutions are not assigned or are far away from most of those nonempty hyper-boxes however, then its contribution to diversity will be poor. The contribution of each Pareto set approximation to each nonempty hyper-box has thus to be calculated. That contribution is measured in terms of the grid distance between the Pareto set approximation and the hyper-box considered.

The grid distance GD between two hyper-boxes h_1 and h_2 in the grid is computed as it is shown by (2), with h_1^k and h_2^k giving the coordinates of h_1 and h_2 in the k-th objective, respectively. It can be observed that the Euclidean distance is considered by DCI.

$$GD(h_1, h_2) = \sqrt{\sum_{k=1}^{M} (h_1^k - h_2^k)^2} \tag{2}$$

The grid distance D between an approximation P and a hyper-box h is the minimum grid distance between h and any other hyper-box, referred to as $G(p)$, containing at least one solution p belonging to P:

$$D(P, h) = \min_{p \in P}\{GD(h, G(p))\} \tag{3}$$

Therefore, the contribution CD of an approximation P to a hyper-box h can be computed as follows:

$$CD(P, h) = \begin{cases} \frac{1 - D(P,h)^2}{M+1}, & D(P, h) < \sqrt{M+1} \\ 0, & D(P, h) \geq \sqrt{M+1} \end{cases} \tag{4}$$

Finally, considering a Pareto front approximation P, its DCI value can be calculated as it is shown by (5), where the number of nonempty hyper-boxes is given by S.

$$DCI(P) = \frac{1}{S} \sum_{i=1}^{S} CD(P, h_i) \tag{5}$$

The main advantages of this indicator are the following ones:

1. **It does not require a reference set**, in opposition to other indicators, such as DM.
2. **It cannot only be applied to compare two Pareto front approximations, but several of them**.
3. **The execution time of DCI, which belongs to $O(M(LN)^2)$, is independent of the number of hyper-boxes**. L is the number of Pareto set approximations and N is the number of solutions in those approximations.

2.1.3. Hypervolume indicator

The hypervolume indicator, $I_{\text{hyp}}(\mathcal{A})$, [10,12–14] computes the volume of the region H delimited by a given set of points, $\mathcal{A}$, and a set of reference points, $\mathcal{N}$, as it is shown by (6). Therefore, larger values of the indicator will correspond to better solutions.

$$I_{\text{hyp}}(\mathcal{A}) = \text{volume}\left(\bigcup_{\forall \boldsymbol{a}\in\mathcal{A};\forall \boldsymbol{n}\in\mathcal{N}} \text{hypercube(a, n)}\right). \tag{6}$$

The hypervolume indicator is also known as the $\mathcal{S}$ metric or the Lebesgue measure. It has many attractive features that have favoured its application and popularity. In particular, it is the only indicator that has the properties of a metric and the only one to be strictly Pareto monotonic [15,16]. Because of these properties this indicator has been used not only for performance assessment but also as part of some MOEAs (see Section 2.4 for details).

To measure the absolute performance of an algorithm the reference points should be *nadir points* ideally. These points are the worst elements of O or, in other words, the elements of O that do not dominate any other element. To contrast the relative performance of two sets of solutions, though, one can be used as the reference set. These matters are further detailed in[14,17].

Having $\mathcal{N}$, the computation of the indicator is a non-trivial problem. Indeed, its determination is known to be computationally intensive, thus rendering it unsuitable for MaOPs.

A lot of research has focused on improving the computational complexity of this indicator[18–21]. The exact computation of the algorithm has been shown to be #*P*-hard[22] in the number of objectives. These types of problems are the analogous of *NP* for counting problems[23]. Therefore, all algorithms calculating the hypervolume must have an exponential run-time with regard to the number of objectives if $P \neq NP$, something that seems to be true[24].

According to the most recent results, the indicator is currently known to be $O(n \log n + n^{M/2})$[21] for more than three objectives ($M > 3$); $O(n \log n)$ for $M = 2, 3$[20]. One alternative to circumvent the complexity hurdle is to apply estimation algorithms capable of yielding an approximation of the indicator at a more convenient temporal cost. *Monte Carlo sampling*[25] and *k-greedy strategy*[26] have been applied with success.

The hypervolume can also be used to measure the progress of an algorithm as the evolution proceeds. For doing that, the relative formulation of the binary hypervolume indicator[27] is usually considered:

$$I_{\text{hyp}}(\mathcal{A}, \mathcal{B}) = I_{\text{hyp}}(\mathcal{A}) - I_{\text{hyp}}(\mathcal{B}) \ . \tag{7}$$

Substituting $\mathcal{A}$ and $\mathcal{B}$ by the non-dominated elements of the current and the previous iteration, $\mathcal{PF}^*_t$ and $\mathcal{PF}^*_{t-1}$, respectively, the indicator can be expressed as:

$$I_{\text{hyp}}(t) = I_{\text{hyp}}(\mathcal{PF}^*_t) - I_{\text{hyp}}\left(\mathcal{PF}^*_{t-1}\right) \ . \tag{8}$$

2.2. MOEA selection

A wide variety of algorithms, like the *Non-dominated Sorting Genetic Algorithm II*[28] (NSGA-II), which is described at Section 2.3, or the *improved Strength Pareto Evolutionary Algorithm*[29] (SPEA2), have been designed by incorporating Pareto optimality as the main selection criterion. Nevertheless, Pareto-based selection is not suitable for many-objective optimisation. One of the main options to increase the performance of Pareto-based MOEAs when dealing with MaOPs is to modify or replace the diversity-based criterion present at their selection scheme. This is the choice addressed, for instance, by the *Non-dominated Sorting Genetic Algorithm III*[30] (NSGA-III).

A different class of MOEAs are those incorporating a quality or performance indicator, like the *hypervolume*[10] or the *R2 indicator*[31], into the selection mechanism. These quality indicators are usually designed for assessing either convergence or diversity or both of them. Individuals are thus selected depending on their contribution to convergence and/or diversity of the solution set they belong to. That contribution is measured by the particular quality indicator applied. The indicator-based MOEAs we have selected for our analyses will be depicted at Section 2.4.

2.3. Pareto-based selection: NSGA-II

NSGA-II is an improvement over the original *Non-dominated Sorting Genetic Algorithm*[32] (NSGA). NSGA-II incorporates two key operations: fast non-dominated sorting of the population and crowding distance computation with the aim of promoting diversity in the population.

The crowding distance considers the size of the largest cuboid enclosing each individual without including any other member of the population. This feature is used to keep diversity in the population, where points belonging to the same rank and with a higher crowding distance are assigned a better fitness than those with a lower crowding distance, avoiding the use of the fitness sharing factor.

2.4. Indicator-based selection

This section is devoted to describe the particular indicator-based MOEAs we have selected in order to carry out our study. Particularly, we have selected two algorithms that incorporate two well-known quality indicators. The first one makes use of the hypervolume indicator, while the second one applies the R2 indicator.

2.4.1. SMS-EMOA

The $\mathcal{S}$*-metric Selection Evolutionary Multi-objective Optimisation Algorithm* (SMS-EMOA), which was proposed by Beume et al.[33], is a steady-state algorithm, i.e., only one offspring is produced and only one individual has to be removed from the population at every generation. The hypervolume is not computed exactly. Instead, the k-greedy strategy is employed. These decisions were made in the hope of tackling the high computational demands of computing the hypervolume.

The key element of SMS-EMOA is the method for determining which element of the population will be replaced by the offspring. This is done by applying a non-domination ranking. From the individuals that are dominated by the rest of the population, one individual is selected such that it makes the minimum contribution to the hypervolume considering the whole set.

2.4.2. R2-EMOA

The R2-EMOA algorithm was originally proposed by Trautmann et al.[34] and was analysed in depth by Brockhoff et al.[35]. As in the case of SMS-EMOA, R2-EMOA is a steady-state approach, but it incorporates the R2 indicator[31] as the secondary criterion for guiding the selection. The individual $\boldsymbol{a}^*$ belonging to the worst rank R_h and allowing the smallest R2 indicator value associated to the remaining individuals of that worst rank to be obtained, is selected to be replaced by the offspring. The aforementioned procedure is depicted by (9). Parameters $\boldsymbol{r}^*$ and Λ are the utopian point and the set of weight vectors, respectively, which allow preferences of a decision maker to be incorporated into R2-EMOA[34].

$$\boldsymbol{a}^* = \arg\min\{r(\boldsymbol{a}) : \boldsymbol{a} \in R_h\};\ \forall \boldsymbol{a} \in R_h : r(\boldsymbol{a}) = R2\left(R_h \setminus \{\boldsymbol{a}\}; \Lambda; \boldsymbol{r}^*\right) \tag{9}$$

We note that the R2 indicator, as the hypervolume, assess the main three features that a Pareto front approximation should fulfil: convergence, spread and uniformity. The R2 indicator, however, presents two main differences with respect to the hypervolume. First, the computation of the hypervolume, which takes exponential time in the number of objectives[†] is avoided. Second, it may avoid the biased behaviour of the hypervolume indicator regarding the solution sets provided, which are usually focused on the knee area of the Pareto front. The above is due to the potential incorporation of preferences of a decision maker.

2.5. Reference-point-based selection: NSGA-III

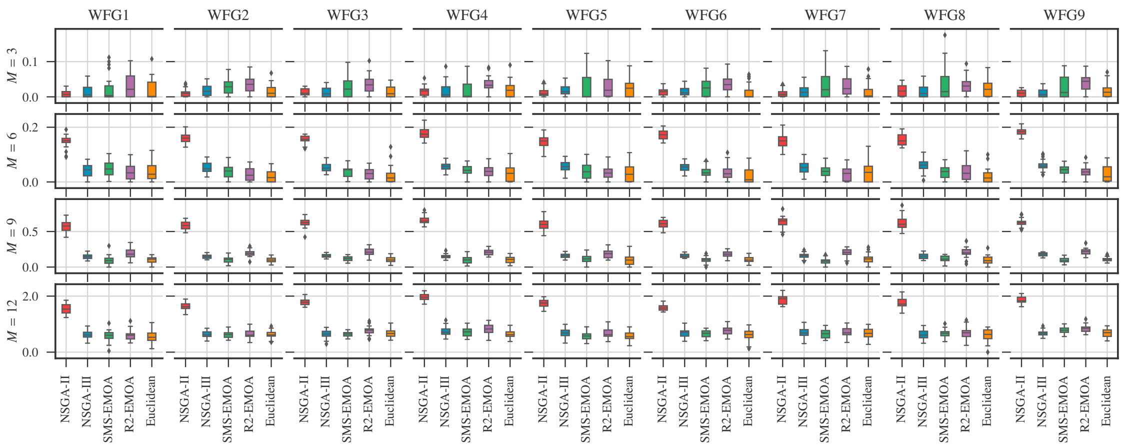
Another promising line for tackling MaOPs comes from the reference-point-based many-objective version of the NSGA-II, referred to as NSGA-III. Similarly to NSGA-II, NSGA-III employs the Pareto non-dominated sorting to partition the population into a number of fronts. In the last front however, rather than using the crowding distance to determine the surviving individuals, a novel niche-preservation operator is applied.

This niche-preservation operator relies on reference points organised in a hyper-plane in order to promote a diverse population. As a result, solutions associated with a smaller number of crowded reference points are more likely to be selected. Finally, we note that a sophisticated normalisation scheme is incorporated into the NSGA-III, which is aimed to effectively handle objective functions of different scales.

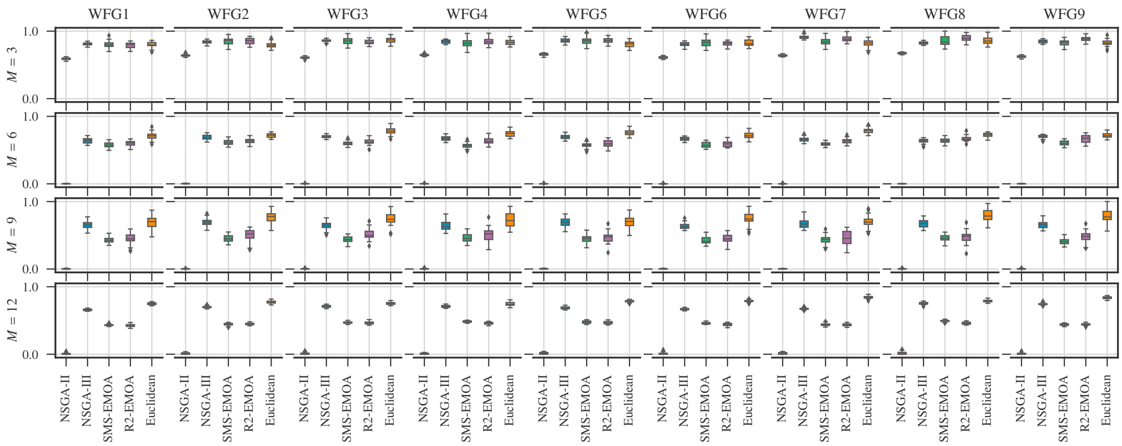
3. Experimental results

The *leitmotiv* of this work is to study the impact that different selection methods have on population diversity when dealing with MaOPs. A shared MOEA framework was used to analyse the selection methods considered under the same experimental conditions. The said framework provided a testing ground common to all approaches, and as a result, we were able to solely focus on the topic of interest. The shared MOEA is similar to other previously proposed algorithms and falls into the $(\mu + \lambda)$ evolutionary strategy scheme. The algorithm is summarised in Fig. 2 as the procedure `shared_moea()`. It starts with an initial random population, $\mathcal{P}_0$, of μ individuals. After that, at every iteration, t, an offspring population with λ individuals, $\mathcal{P}_{\text{off}}$, is created by applying the variation operators

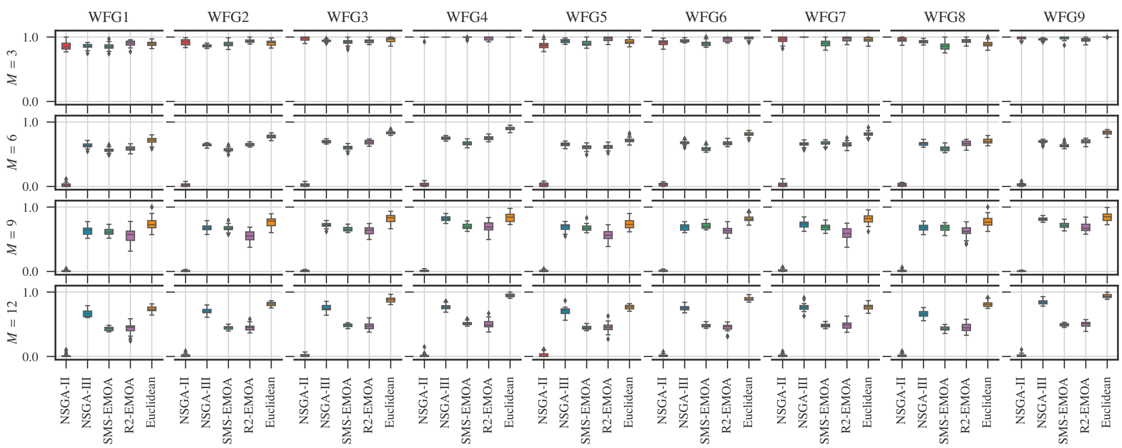
[†] In the case of dealing with two or three objectives efficient multi-objective optimisers based on the hypervolume indicator have been proposed.



(a) hypervolume indicator



(b) Diversity Comparison Indicator (DCI)



(c) Diversity Measure (DM)

Fig. 1. Box plots of the hypervolumes (lower values are better), DCI and DM (higher values are better) obtained after executing the experiment runs on each problem and number of objectives.

function `shared_moea`(`sel_func`, μ, λ)
 ▷ `sel_func`, selection function to be used.
 ▷ μ and λ, population and offspring sizes.
 $t \leftarrow 0$; $\mathcal{P}_0 \leftarrow$ `random_population`(μ).
 while end criterion not met **do**
  $\mathcal{P}_{\text{off}} \leftarrow$ `apply_variation`($\mathcal{P}_t, \lambda$).
  $\mathcal{P}_{t+1} \leftarrow$ `sel_func`($\mathcal{P}_t \cup \mathcal{P}_{\text{off}}, \mu$).
  $t \leftarrow t + 1$.
 return `non_dom_set`($\mathcal{P}_{t+1}$), final non-dominated set.

Fig. 2. Algorithmic representation of the shared MOEA framework.

in `apply_variation()` on the current population, $\mathcal{P}_t$. Subsequently, the best μ individuals are kept for the next generation population $\mathcal{P}_{t+1}$ by applying a given selection function, `sel_func()`, which is specified as a parameter. Once the stopping criterion is met, the non-dominated subset of $\mathcal{P}_t$ is returned.

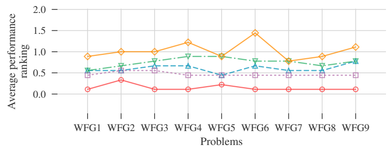
We note that for our analyses we considered the selection schemes of the following MOEAs, which were introduced at Section 2.2: NSGA-II, NSGA-III, SMS-EMOA, R2-EMOA and a modified version of NSGA-II, where the Euclidean distance substitutes the crowding distance in the selection mechanism.

We chose the WFG multi-objective problem toolkit[36] as the benchmark suite. It describes nine complex problems, referred to as WFG1–WFG9, that test whether the optimisation algorithms are capable of handling different challenges, like separability, multi-modality and deceptive local optima, among others. Each problem was addressed with $M = 3, 6, 9, 12$ objective functions. At the same time, the number of function evaluations was used as the stopping criterion. Particularly, executions were stopped once $10^{3+\frac{M}{3}}$ function evaluations were carried out, thus performing longer runs for those problems with a higher number of objective functions. The population size of the different approaches was fixed by considering the number of objective functions of the problem at hand as well, i.e. , $\mu = \lambda = 50 \times 10^{\frac{M}{3}}$. For all cases, we ran all experiment instances 50 times[‡]. The diversity of the resulting solution sets was computed using the hypervolume, DCI and DM indicators, which were described at Section 2.1. As we previously mentioned, the hypervolume indicator is able to capture both convergence of the approximation and its diversity, while both DCI and DM are meant for assessing diversity only.

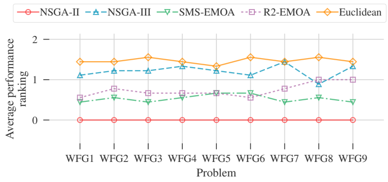
These results are summarised as box plots in Fig. 1. It is particularly interesting that the selection mechanism of NSGA-III, as well as the application of the Euclidean distance rather than the crowding distance by NSGA-II, consistently yielded better results as the number of objective functions grew. Although illustrative, box plots can not be used to reach a definitive conclusion. That is why statistical hypothesis tests are called for. In our case, for each problem/number of objective functions combination, we performed a Kruskal–Wallis test with the indicator values achieved by each algorithm. In this context, the null hypothesis of this test is that all algorithms are equally capable of solving the problem. If the null hypothesis was rejected, which was actually the case in all instances of the experiment, the Conover–Inman procedure was applied in a pairwise manner to determine whether a particular algorithm attained better results than another. A significance level $\alpha = 0.05$, corrected using the Dunn-Šidák correction, was applied.

To further simplify the understanding of the results, we decided to adopt a more integrative representation like the one proposed by Bader[37]. This representation groups, either by problem or by number of objectives, the results provided by the different algorithms. It does so by computing the number of times a given algorithm was statistically better than the others. Fig. 3 conveys these analyses. This summarised representation allows the results previously identified in Fig. 1 to be easily verified. It becomes evident that the NSGA-II applying the Euclidean distance yielded important and outstanding results. When doing the analysis grouping results by problems (Figs. 3a–3c), it is clear that the Euclidean distance-based approach managed to yield consistently better results, with a few exceptions where it was outperformed by the selection scheme of SMS-EMOA. At the same time, the analysis grouping results by number of objectives (Figs. 3d–3f) brings even more interesting results. In this case, the Euclidean distance-based approach provided the best results for numbers of objectives larger than three, with the only exception that considers

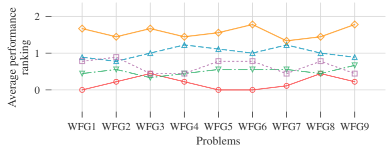
‡ Full experiment parameters, raw data, statistical hypothesis tests and source code will be available on-line upon publication.



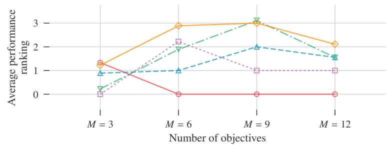
(a) Hypervolumes grouped by problem.



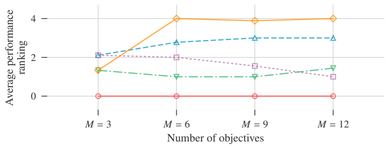
(b) DCIs grouped by problem.



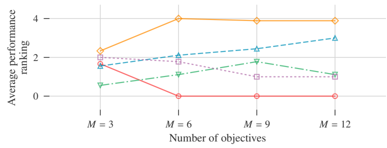
(c) DMs grouped by problem.



(d) Hypervolumes grouped by number of objectives.



(e) DCIs grouped by number of objectives.



(f) DMs grouped by number of objectives.

Fig. 3. Summarised representation of the statistical hypothesis tests. A higher position on the charts implies better results.

the hypervolume indicator and nine objectives, where it addressed results slightly worst than SMS-EMOA. Here it should be noted that the computational cost of the Euclidean distance-based approach is a fraction of that invested by SMS-EMOA.

4. Final remarks

Maintaining population diversity is a key issue of MOEAs. This is particularly important when dealing with many-objective test cases as it has been repeatedly reported that current approaches are not capable of sustaining diverse populations. In this paper we have studied the impact that different selection schemes belonging to well-known MOEAs have on population diversity when dealing with MaOPs. We have dealt with nine complex benchmark problems with an increasing number of objectives. The results point out the limitations of current selection schemes and the directions for future progress in this area. Particularly, they have shown how the reference-point-based selection approach incorporated into NSGA-III and the modified version of the NSGA-II, where the Euclidean distance replaces the crowding distance, are able to provide better performance, and not only in terms of diversity as DM and DCI indicates, but also in terms of convergence (hypervolume), specially in the case of the modified NSGA-II.

Similarly, there are some important results that should be explored more in depth: the use of the Euclidean distance as part of the selection scheme. Our experimental evaluation indicates this is a promising line of research and, to the best of our knowledge, it has not been investigated yet.

References

1. Ishibuchi, H., Tsukamoto, N., Nojima, Y.. Evolutionary many–objective optimization: A short review. In: *2008 IEEE Congress on Evolutionary Computation (CEC 2008)*. Piscataway, New Jersey: IEEE Press; 2008, p. 2419–2426.
2. Branke, J., Miettinen, K., Deb, K., Słowiński, R., editors. *Multiobjective Optimization*; vol. 5252 of *Lecture Notes in Computer Science*. Berlin/Heidelberg: Springer–Verlag; 2008.
3. Bäck, T.. *Evolutionary Algorithms in Theory and Practice*. New York — Oxford: Oxford University Press; 1996.
4. Coello Coello, C.A., Lamont, G.B., Van Veldhuizen, D.A.. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Genetic and Evolutionary Computation. New York: Springer; second ed.; 2007.
5. Liu, Y., Gong, D., Sun, J., Jin, Y.. A many-objective evolutionary algorithm using a one-by-one selection strategy. *IEEE Transactions on Cybernetics* 2017;**PP**(99):1–14.
6. Sato, H., Aguirre, H., Tanaka, K.. Improved S-CDAs using crossover controlling the number of crossed genes for many-objective optimization. In: *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*; GECCO '11. New York, NY, USA: ACM. ISBN 978-1-4503-0557-0; 2011, p. 753–760.
7. Zou, X., Chen, Y., Liu, M., Kang, L.. A new evolutionary algorithm for solving many-objective optimization problems. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 2008;**38**(5):1402–1412.

8. Huband, S., Barone, L., While, L., Hingston, P.. A scalable multi-objective test problem toolkit. In: Coello Coello, C.A., Hernández Aguirre, A., Zitzler, E., editors. *Evolutionary Multi-Criterion Optimization: Third International Conference, EMO 2005, Guanajuato, Mexico, March 9-11, 2005. Proceedings*. Springer Berlin Heidelberg. ISBN 978-3-540-31880-4; 2005, p. 280–295.
9. Li, M., Yang, S., Liu, X.. Diversity comparison of Pareto front approximations in many-objective optimization. *IEEE Transactions on Cybernetics* 2014;**44**(12):2568–2584.
10. Zitzler, E., Deb, K., Thiele, L.. Comparison of Multiobjective Evolutionary Algorithms on Test Functions of Different Difficulty. In: Wu, A.S., editor. *Proceedings of the 1999 Genetic and Evolutionary Computation Conference. Workshop Program*. Orlando, Florida; 1999, p. 121–122.
11. Deb, K., Jain, S.. Running performance metrics for evolutionary multi-objective optimization. Tech. Rep.; Kanpur Genetic Algorithms Laboratory, ITT Kanpur; 2002.
12. Zitzler, E., Brockhoff, D., Thiele, L.. The hypervolume indicator revisited: On the design of pareto-compliant indicators via weighted integration. In: Obayashi, S., et al., editors. *Conference on Evolutionary Multi–Criterion Optimization (EMO 2007)*; vol. 4403 of *LNCS*. Springer Berlin Heidelberg; 2007, p. 862–876.
13. Knowles, J.D.. *Local-Search and Hybrid Evolutionary Algorithms for Pareto Optimization*. Ph.D. thesis; The University of Reading, Department of Computer Science; Reading, UK; 2002.
14. Knowles, J., Thiele, L., Zitzler, E.. A tutorial on the performance assessment of stochastic multiobjective optimizers. TIK Report 214; Computer Engineering and Networks Laboratory (TIK), ETH Zurich; 2006.
15. Fleischer, M.. The measure of Pareto optima. Applications to multi-objective metaheuristics. In: Fonseca, C.M., Fleming, P.J., Zitzler, E., Deb, K., Thiele, L., editors. *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*. Faro, Portugal: Springer. Lecture Notes in Computer Science. Volume 2632; 2003, p. 519–533.
16. Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., Grunert da Fonseca, V.. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation* 2003;**7**(2):117–132.
17. Zitzler, E., Laumanns, M., Thiele, L., Fonseca, C.M., Grunert da Fonseca, V.. Why Quality Assessment of Multiobjective Optimizers Is Difficult. In: Langdon, W.B., Cantú-Paz, E., Mathias, K., Roy, R., Davis, D., Poli, R., et al., editors. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'2002)*. San Francisco, California: Morgan Kaufmann Publishers; 2002, p. 666–673.
18. While, L., Hingston, P., Barone, L., Huband, S.. A faster algorithm for calculating hypervolume. *IEEE Transactions on Evolutionary Computation* 2006;**10**(1):29–38.
19. While, L., Bradstreet, L., Barone, L., Hingston, P.. Heuristics for Optimising the Calculation of Hypervolume for Multi-Objective Optimization Problems. In: *2005 IEEE Congress on Evolutionary Computation (CEC'2005)*; vol. 3. Edinburgh, Scotland: IEEE Service Center; 2005, p. 2225–2232.
20. Fonseca, C.M., Paquete, L., López-Ibánez, M.. An improved dimension–sweep algorithm for the hypervolume indicator. In: *2006 IEEE Congress on Evolutionary Computation (CEC'2006)*. 2006, p. 1157–1163.
21. Beume, N.. S–metric calculation by considering dominated hypervolume as Klee's measure problem. *Evolutionary Computation* 2009; **17**(4):477–492.
22. Bringmann, K., Friedrich, T.. Approximating the volume of unions and intersections of high–dimensional geometric objects. *Computational Geometry* 2010;**43**(6-7):601–610.
23. Papadimitriou, C.M.. *Computational Complexity*. Reading, MA, USA: Addison–Wesley; 1994.
24. Deolalikar, V.. P≠NP. Tech. Rep.; Hewlett Packard Research Labs; Palo Alto, CA, USA; 2010.
25. Bader, J., Zitzler, E.. HypE: An Algorithm for Fast Hypervolume-Based Many-Objective Optimization. TIK Report 286; Computer Engineering and Networks Laboratory (TIK), ETH Zurich; 2008.
26. Zitzler, E., Thiele, L., Bader, J.. On set-based multiobjective optimization. *IEEE Transactions on Evolutionary Computation* 2010; **14**(1):58–79.
27. Knowles, J., Thiele, L., Zitzler, E.. A Tutorial on the Performance Assessment of Stochastic Multiobjective Optimizers. 214; Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Switzerland; 2006. Revised version.
28. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 2002;**6**(2):182–197.
29. Zitzler, E., Laumanns, M., Thiele, L.. SPEA2: Improving the strength pareto evolutionary algorithm. Tech. Rep. 103; Computer Engineering and Networks Laboratory (TIK), Department of Electrical Engineering, Swiss Federal Institute of Technology (ETH) Zurich; 2001.
30. Jain, H., Deb, K.. An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part II: Handling constraints and extending to an adaptive approach. *IEEE Transactions on Evolutionary Computation* 2014;**18**(4):602–622.
31. Hansen, M.P., Jaszkiewicz, A.. Evaluating the quality of approximations to the non-dominated set. Tech. Rep.; Department of Mathematical Modelling, Technical University of Denmark; 1998.
32. Srinivas, N., Deb, K.. Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation* 1994; **2**(3):221–248.
33. Beume, N., Naujoks, B., Emmerich, M.. SMS–EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research* 2007;**181**(3):1653–1669.
34. Trautmann, H., Wagner, T., Biermann, D., Weihs, C.. Indicator-based selection in evolutionary multiobjective optimization algorithms based on the desirability index. *Journal of Multi-Criteria Decision Analysis* 2013;**20**(5–6):319–337.
35. Brockhoff, D., Wagner, T., Trautmann, H.. R2 indicator-based multiobjective search. *Evolutionary Computation* 2015;**23**(3):369–395.
36. Huband, S., Hingston, P., Barone, L., While, L.. A review of multiobjective test problems and a scalable test problem toolkit. *IEEE Transactions on Evolutionary Computation* 2006;**10**(5):477–506.
37. Bader, J.. *Hypervolume-Based Search for Multiobjective Optimization: Theory and Methods*. Ph.D. thesis; ETH Zurich; Switzerland; 2010.