A CLUSTERING-BASED NICHING FRAMEWORK FOR THE APPROXIMATION OF EQUIVALENT PARETO-SUBSETS

OLIVER KRAMER
Carl von Ossietzky University Oldenburg, Department of Computer Science
Uhlhornsweg 84, 26111 Oldenburg

HOLGER DANIELSIEK
TU Dortmund, Department of Computer Science
Otto-Hahn-Str. 14, 44221 Dortmund

Received (received date)
Revised (revised date)

In many optimization problems in practice, multiple objectives have to be optimized at the same time. Some multi-objective problems are characterized by multiple connected Pareto-sets at different parts in decision space – also called equivalent Pareto-subsets. We assume that the practitioner wants to approximate all Pareto-subsets to be able to choose among various solutions with different characteristics. In this work we propose a clustering-based niching framework for multi-objective population-based approaches that allows to approximate equivalent Pareto-subsets. Iteratively, the clustering process assigns the population to niches, and the multi-objective optimization process concentrates on each niche independently. Two exemplary hybridizations, i.e., rake selection and DBSCAN, as well as SMS-EMOA and kernel density clustering demonstrate that the niching framework allows enough diversity to detect and approximate equivalent Pareto-subsets.

Keywords: evolution strategies; evolutionary algorithms; multi-objective optimization; equivalent Pareto-subsets; niching; density-based clustering; rake selection

1. Introduction

1.1. Motivation

Many design problems involve multiple conflictive objectives. The designer often does not know in advance, how to balance the design objectives, and which solutions are feasible or satisfying. The approximation of a Pareto-set allows the designer to evaluate and compare various designs, and to choose an optimal solution a posteriori. Stochastic optimization methods turn out to be successful in this context. Although some evolutionary multi-objective approaches take into account diversity in decision space, few are designed to detect equivalent Pareto-subsets. Equivalent Pareto-subsets are subsets of Pareto-optimal solutions in different parts of design space. Some of these subsets may even cover the whole Pareto-front on their own. In turn, this means that for the same locations on the Pareto-front, multiple alternative
solutions in design space exist. As many multi-objective approaches do not consider, from where a Pareto-optimal solution stems in design space these approaches lack of flexibility in generating a preferably diverse set of solutions for a posteriori selection.

In this context, a main challenge is to generate alternative solutions in design space with equivalent balance of objectives. If solutions become infeasible in any phase of the design process, the designer can choose among equivalent alternatives. Without diversity enhancing mechanisms that support the optimization process, standard evolutionary multi-objective optimization techniques might fail to detect all Pareto-subsets reliably. Preuss et al. presented a real-world problem that shows such properties and turns out to be difficult to solve.

The hybridization between clustering of the design space and independent optimization of each cluster of solutions is a suitable approach for the detection of equivalent Pareto-subsets. We have presented preliminary work as a proof-of-concept, concentrating on rake selection combined with K-means, as well as DBSCAN. In this article we enhance the work to a general niching framework applicable for population-based derivative-free optimizers, and arbitrary clustering approaches. Two experimental case-studies of rake selection with DBSCAN, and SMS-EMOA with kernel density clustering demonstrate the capabilities of the clustering-based niching framework. Furthermore, we extend the approach by two indicators that automatize the niching process.

1.2. Multi-Objective Optimization

Many planning or decision making problems involve multiple objectives that have to be optimized simultaneously, in operations research also known as multiple criteria decision making. Formally, we assume $m$ conflictive objectives $f(x) = (f_1(x), \ldots, f_m(x))^T$:

$$
\min_{x \in \mathbb{R}^N} f(x) = \min_{x \in \mathbb{R}^N} (f_1(x), f_2(x), \ldots, f_m(x))
$$

with $f_i(x) : \mathbb{R}^N \rightarrow \mathbb{R}, i = 1, \ldots, m$. In most cases the decision maker is interested in a set of Pareto-optimal solutions, i.e., a set of solutions that are better or equal than other solutions w.r.t. all objectives, but that are strictly better in at least one objective. Formally, we seek for a Pareto-set:

$$
P = \{ x^* \in \mathbb{R}^q | \exists x \in \mathbb{R}^q : x \prec x^* \}.
$$

In objective space the solutions are called Pareto-front. Pareto-optimal solutions are also known as non-dominated or non-inferior solutions. After a Pareto-set has been generated, the decision maker can select the solutions that fit best to his preferences.

1.3. Overview

Section 2 gives an overview of related work on diversity in objective space, and niching approaches. In Section 3 we introduce the niching framework for the detection of equivalent Pareto-subsets. We introduce two instances of clustering-based
niching approaches complemented by an experimental analysis. In Section 4 we introduce two adaptive mechanisms that trigger the niching process with the help of cluster indicators, i.e., a parameterizable indicator, and a parameter-free one. Section 5 closes with a discussion of the results, and an outlook to future research possibilities. The parameterized test function that is basis of our experimental part is introduced in Appendix A.

2. Related Work

2.1. Multi-Objective Optimization

Research on multi-objective optimization has a long tradition, beginning in the 19th century with the work of Edgeworth, Kuhn and Tucker, and Pareto. Evolutionary multi-objective optimization algorithms (EMOAs) have shown outstanding success in the last decade. Algorithms like NSGA-II by Deb et al., SPEA by Zitzler and Thiele, and the SMS-EMOA by Emmerich et al. are able to generate Pareto-sets of solutions in non-linear and multimodal scenarios. Most EMOAs generate a population of non-dominated solutions to approximate the Pareto-set. A comprehensive introduction to evolutionary multi-objective optimization is given in the book of Coello Coello et al.

Pareto-sampling techniques belong to the most successful methods. They stochastically generate a population of solutions that can be partially sorted by their rank, i.e., the number of population members that dominate the solution. Goldberg was the first who introduced domination as selection objective. To maintain diversity he introduced a niching-based approach. Also Horn et al., as well as Fonseca et al., proposed niching approaches. Many Pareto-sampling techniques have been introduced: MOGA by Fonseca and Fleming, MOMGA, the multi-objective messy genetic algorithm by Veldhuizen and Lamont, MOMGA II by Zydallis et al. and SPEA by Zitzler and Thiele, as well as its successor SPEA2 by Zitzler et al. One of the most famous approaches in this line of research is the Non-Dominated Sorting Genetic Algorithm NSGA by Srinivas and Deb, and its successor NSGA-II by Deb et al. The idea of non-dominated sorting is to rank individuals of the population according to their non-dominance level. All individuals that are not dominated are assigned to the first rank. In a next step they are removed from the population, and the non-dominance check is conducted again. These steps are repeated until all individuals have been assigned to a rank according to their non-dominance level. Afterwards, an indicator in objective space is the basis for a second-level selection of NSGA-II: best ranked solutions with maximum crowding distance values are added to the population of the next generation. The crowding distance selection is a diversity preserving mechanism in objective space. It is the average Manhattan distance to neighbored non-dominated solutions.

The $S$-metric is an indicator for the approximation of the Pareto-front by com-
puting the dominated hypervolume of a population. The metric has been introduced as basis of the SMS-EMOA. For an introduction we refer to Emmerich et al.\textsuperscript{13}, and Beume et al.\textsuperscript{2}. We will use the $S$-metric in the experimental analyses of this work as indicator for the ability to approximate the Pareto-front. We will also use an approach based on reference lines in objective space that has been proposed by Kramer and Koch\textsuperscript{21}, and that will shortly been introduced in Section 3.2.1.

2.2. Diversity in Objective and Decision Space

The goal of niching in multi-objective optimization is to maintain spatial diversity in solution space\textsuperscript{7,9,18,23}. A survey of niching methods gives Mahfoud\textsuperscript{24}. Niching in evolution strategies (ES) is also described by Shir\textsuperscript{30}. In this section we concentrate on techniques in the context of diversity in multi-objective optimization.

Rudolph et al.\textsuperscript{28} argue that it is important that the user is capable of obtaining distinct Pareto-subsets of similar quality. They investigate the success of standard EMOAs w.r.t. these characteristics, and come to the conclusion that none is able to reliably detect and preserve all equivalent Pareto-subsets, as they have not been designed for this purpose. In their work they have introduced the test problems our experimental analysis will concentrate on. Furthermore, Rudolph et al. introduced a multi-start approach that allows an approximation of $SYM-PART$, see Appendix A.

A combined consideration of objective and decision space has been introduced by Chan and Ray\textsuperscript{5}. Their approach considers the location of solutions in decision space, and makes use of two selection operators, one for diversity in decision, the other for diversity in objective space. Deb and Tiwari\textsuperscript{11} introduce the crowding distance for decision space, similar to the crowding distance of the NSGA variants. Their EMOA is also based on non-dominated sorting like our approach. The implementation is called omni-optimizer and aims at solving a wide variety of single- and multi-objective problems.

A further niching-based approach has been introduced by Shir et al.\textsuperscript{31}. Their Niching CMA-ES combines a multi-objective variant of the CMA-ES with a special selection operator. The work extends an existing CMA-ES niching framework that has been designed for multi-objective problems. An experimental analysis shows that the niching approach maintains diversity without a significant loss in comparison to the convergence abilities of the original CMA-ES. Zhou et al.\textsuperscript{36} consider a class of multi-objective problems with different dimensionalities of the Pareto-set and Pareto-front manifolds. They propose a probabilistic model-based EMOA that clusters the sample points into subpopulations based on their distribution in objective space. A principal component analysis is used to estimate the Pareto-set manifold dimensionality, and a probabilistic model of the distribution in decision space is computed. Ulrich et al.\textsuperscript{34} proposed an approach to integrate decision space diversity enhancing the hypervolume indicator of the SMS-EMOA.
A Clustering-Based Niching Framework for the Approximation of Equivalent Pareto-Subsets

3. Clustering-Based Niching

In this section we introduce the clustering-based niching framework for the detection and approximation of equivalent Pareto-subsets. Afterwards, we demonstrate the behavior of concrete hybridizations and analyze them experimentally.

3.1. Framework

Algorithm 1 shows the clustering-based niching framework to detect equivalent Pareto-subsets. Until a termination condition the main iteration cycle is based on successive clustering and multi-objective optimization in niches. In the clustering step, the population of candidate solutions \( P_t = \{x_1, \ldots, x_\mu\} \) with \( x_j \in \mathbb{R}^N \) has to be assigned to niches. The task is to learn a meaningful assignment \( P_t \rightarrow P_i^t \) with \( 1 \leq i \leq K \) into \( K \) niches, based on the structure and distribution of the given population. An optimal niche\(^a\) assignment can be characterized as follows: (1) homogeneity among individuals in the same niche, and (2) heterogeneity of individuals in different niches.

Homogeneity and heterogeneity can be defined by similarity measures in data space. In general, it is unknown, how many niches exist, and which properties they have. Some clustering techniques require the number of clusters at the beginning, e.g., K-means, which is one of the most famous clustering method. If the clustering implementation allows the detection of new clusters, i.e., without the specification of a preliminary number at the beginning, the approach is able to discover new clusters in the population automatically. For this reason we recommend to employ density-based approaches.

Algorithm 1 Clustering-based niching framework

1: \( t := 0 \)
2: while termination condition not reached do
3: \( t = t + 1 \)
4: cluster population \( P_t \rightarrow P_i^t \)
5: for all niches \( P_i^t \) do
6: optimize \( P_i^t \) until termination condition
7: end for
8: merge niches \( P_i^t \rightarrow P_{t+1} \)
9: end while
10: post-optimization of \( P_{t+1} \)

After the individuals have been assigned to their niches, the optimization step has the task to improve the candidate solution within the niches, i.e., the subpopulation \( P_i^t \). An independent multi-objective optimization process within each

\(^a\)The terms \textit{cluster} and \textit{niche} are treated as synonyms in the following.
niche allows the approximation of the Pareto-front for each independent subset. The niching-based optimization process takes place until a termination condition is reached. This can be a fixed number of iterations – denoted as isolation time – or an indicator, see Section 4. Many population-based multi-objective algorithms can be integrated into this clustering framework, as each niche can be treated as independent multi-objective problem. The exploration capabilities of the EMOA will determine the capabilities to detect new niches, i.e., to leave the subspace of the solution domain that has been defined as a niche by the clustering process. After optimization within the niches, they are merged to one population $P_t \rightarrow P_{t+1}$, which becomes the basis of the following generation.

3.2. **Case-Study 1: Rake Selection and DBSCAN**

In the following, we will introduce two clustering-based niching instances, each based on different EMOAs and clustering techniques, and their experimental analyses. In case-study One we employ rake selection by Kramer and Koch \(^{21}\), and DBSCAN by Ester \textit{et al.} \(^{14}\).

3.2.1. **Rake Selection**

If the decision maker wants to select solutions a posteriori, i.e., after the Pareto-set has been generated, the question arises how the Pareto-front should look like. Rake selection \(^{21}\) is an approach that allows the arbitrary distribution of references lines (rakes) that guide the search in objective space. In particular, for two objectives ($m = 2$) it is reasonable to distribute the rakes in parallel, and equidistantly in objective space yielding approximately uniformly distributed solution on the Pareto-front.\(^b\)

The rakes are arranged orthogonally and equidistantly on the surface defined by the Pareto-optimal corner points $c_i$, $1 \leq i \leq m$ on a $m - 1$-dimensional hyperplane $\mathbf{h}$. The idea of adaptive rake corner points is to use the $m$ optimal solutions $x_i$, $1 \leq i \leq m$ of population $P_t$ at generation $t$ w.r.t. each objective $f_i$ as corner points for the rake base, i.e., $c_i = (f_1(x_i), \ldots, f_m(x_i))^T$. Various experiments have shown that the exploration capabilities can be enhanced by shifting the corner rakes to the outside, i.e., $c_i' = c_i + \gamma$, with a problem-specific parameter $\gamma$. Otherwise, the explorative character of the approach is limited, as the rakes are only distributed between the corner points, and not the whole Pareto-front is guaranteed to be found. The outer rakes in the neighborhood of the corner points have an explorative character, while the inner points exploit the knowledge about the location of the already explored Pareto-front. Rake selection uses a $(\mu + \lambda)$-ES with self-adaptive step sizes \(^3\).

\(^b\)We have to point out that rake selection is in particular tailored for $m = 2$, but can also be used for more than two objectives.
3.2.2. DBSCAN

DBSCAN is based on the density of data samples, and has two essential advantages that are useful in the context of the niching procedure. Density-based clustering methods are based on the assumption that a certain least number of data samples have to lie in the radius $\epsilon$ of a reference data sample. The density of data samples has to exceed a given threshold. Hence, it allows to cluster data samples of arbitrary shape, e.g., non-convex and intertwined sets. A further advantage of DBSCAN is the ability to adapt the number of clusters automatically in the course of the clustering process.

DBSCAN starts at an arbitrary point $x$ in data space, and computes the $\epsilon$-neighborhood $N_\epsilon(x)$. If $x$ is a core point, for each point $x'$ of set $N_\epsilon(x)$ the set of neighbored points is computed. If $|N_\epsilon(x')|$ is smaller than $\nu$, $x'$ is classified as border point of the found clusters, and does not have to be considered anymore. Otherwise, i.e., for $|N_\epsilon(x')| \geq \nu$, all points in $N_\epsilon(x')$ are assigned to the current cluster number $\rho$, if this has not been done yet. Points that have already been classified as noise are assigned to the clusters. Before a new iteration starts, the current point $x$ is deleted from set $N_\epsilon(x)$. If $N_\epsilon(x)$ is empty, the method terminates indicating that the cluster has been identified completely. Thom and Kramer have presented a method to accelerate DBSCAN based on the reduction of neighborhood evaluations. In the following, we are going to explore, if DBSCAN is an appropriate niching mechanism to detect equivalent Pareto-subsets.

3.2.3. Experiments

Table 1 shows the experimental analysis of the niching approach on problems SYM-PART2 and SYM-PART3. The figures shows the total $S$-metric, an indicator for the approximation of the Pareto-front of all niches at once, and the average $S$-metric, an indicator for the average Pareto-front covering of the niches.

<table>
<thead>
<tr>
<th>problem</th>
<th>runtime in sec</th>
<th>ffe</th>
<th>total $S$</th>
<th>$S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYM-PART2</td>
<td>361.54</td>
<td>80,000</td>
<td>99.33</td>
<td>99.30</td>
</tr>
<tr>
<td>SYM-PART3 $[-15,15]$</td>
<td>1807.89</td>
<td>382,000</td>
<td>99.32</td>
<td>99.22</td>
</tr>
<tr>
<td>SYM-PART3 $[-20,20]$</td>
<td>742.72</td>
<td>162,800</td>
<td>99.32</td>
<td>99.28</td>
</tr>
<tr>
<td>SYM-PART3 $[-30,30]$</td>
<td>645.54</td>
<td>144,000</td>
<td>99.33</td>
<td>99.29</td>
</tr>
</tbody>
</table>

An analysis of the Pareto-sets revealed that all nine Pareto-set clusters of SYM-PART have been found, and in each niche the solutions are smoothly distributed in decision space. The solutions are also smoothly distributed on the Pareto-front, a result that is confirmed by the high $S$-metric values. Figure 1 shows a section of
the Pareto-front in detail. The Pareto-front consists of solutions from nine different
niches, marked with different symbols. Table 1 also shows the experiments on the
more complex problem $SYM-PART3$ with different parameterizations. Problem
$SYM-PART3$ is hard to solve for parameterizations like $L = -15$ and $U = 15$. In
this case, the decision space is bulged, the area around $(0, -15)$ is restricted, and
cannot be found in every run. In three of the runs only eight clusters have been
detected by the $(100 + 200)$-ES.

![Detailed section of the Pareto-front of an experiment with the Rake-DBSCAN-hybrid on $SYM-PART2$. Different symbols represent different niches, i.e., solutions from different Pareto-subsets. The lines in decision space show the rakes that are basis of the selection process.](Fig. 1)

3.3. Case-Study 2: SMS-EMOA and Kernel Density Clustering

The second analysis is based on the SMS-EMOA by Emmerich et al. as multi-
objective approach, and kernel density-based clustering that has been proposed,
e.g., by Schnell.

3.3.1. SMS-EMOA

The SMS-EMOA belongs to the family of indicator-based EMOAs using the S-
metric as selection criterion. The S-metric – also called hypervolume indicator –
measures the size of the space dominated by the set of non-dominated solutions.
Maximization of the S-metric means to push the population into the direction of the
Pareto-front. Selection of the non-dominated solutions that maximize the S-metric
is a combinatorial optimization problem. Hence, the SMS-EMOA applies a $(\mu + 1)$
(or steady-state) selection scheme in each generation discarding the individual that
contributes least to the S-metric value of the population. The invoked variation
operators are often polynomial mutation and simulated binary crossover with the same parameterization as for NSGA-II. Auger et al. analyze whether the hypervolume indicator is biased towards certain regions, and show that for linear fronts the distribution is uniform with constant distance between two consecutive points. They come to the conclusion that for general fronts it is presumably impossible to characterize exactly the distribution. Bringmann and Friedrich show that large populations sets that maximize the hypervolume indicator can quickly approach the optimal approximation ratio, while small populations achieve a near-optimal approximation ratio.

3.3.2. Kernel Density Clustering

The idea of kernel density clustering is to place cluster centers at maxima of distributions of sample points. The distributions are estimated via non-parametric estimates, i.e., with the Parzen-window estimator. The idea is that regions with high densities are potential cluster centers. Clustering algorithms that use the kernel density estimates have been introduced by Schnell. They are based on assigning each point to the closest local optimum of the kernel density estimate. Local kernel density maxima define clusters. The approach is based on computing the gradient of the kernel density function to find their local maxima. Hinneburg and Keim introduced an extension of Schnell’s approach. The approach computes the gradient of the kernel density function with a Gaussian kernel, employing hill climbing to find the maximum of the kernel density estimate.

An attractive heuristic to cluster with kernel density estimation is to find the points with the highest kernel density, and assign all points in the neighborhood to these cluster centers. To identify the corresponding cluster centers automatically, we apply the following steps. The regions with the highest kernel density and a least distance to neighbored cluster centers are potential cluster centers. Iteratively, in iteration $t$ the points with the highest relative kernel density

$$d(x_j) = \sum_{i=1,i\neq j}^{N} K_H(x_i, x_j) > nu$$

are identified. We consider a least distance $\rho \in \mathbb{R}^+$ to previously computed cluster centers (codebook vectors) in set $C$ with $d(x_j, c_k) > \rho$, $c_k \in C$, and a least kernel density $\nu \in \mathbb{R}^+$. The points that fulfill both conditions are added to set $C$ of cluster centers. Clusters arise by assigning each data point $x_j$ to its closest cluster center $c_k \in C$.

3.3.3. Experiments

The experimental analysis of the SMS-EMOA and kernel density clustering hybridization are presented in Table 2. For the SMS-EMOA we employ a $(\mu + 1)$-ES with polynomial mutation and simulated binary crossover. Furthermore, we use...
the kernel density clustering heuristic introduced in the last section with $\rho = 0.1$, and $\nu = 10^{-3}$. The experimental results show that the SMS-EMOA and kernel density clustering hybridization shows sufficient capabilities to approximate the Pareto-subsets. But in comparison to the rake selection and DBSCAN hybrid more optimization steps are necessary (both in terms of fitness function evaluations and runtime).

<table>
<thead>
<tr>
<th>problem</th>
<th>runtime in sec</th>
<th>ffe</th>
<th>total</th>
<th>$S$</th>
<th>$S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SYM\text{-PART2}$</td>
<td>940.10</td>
<td>102,800</td>
<td>99.33</td>
<td>99.31</td>
<td></td>
</tr>
<tr>
<td>$SYM\text{-PART3} [-15, 15]$</td>
<td>2002.21</td>
<td>421,200</td>
<td>99.32</td>
<td>99.27</td>
<td></td>
</tr>
<tr>
<td>$SYM\text{-PART3} [-20, 20]$</td>
<td>1322.79</td>
<td>156,600</td>
<td>99.33</td>
<td>99.30</td>
<td></td>
</tr>
<tr>
<td>$SYM\text{-PART3} [-30, 30]$</td>
<td>1022.56</td>
<td>128,400</td>
<td>99.33</td>
<td>99.31</td>
<td></td>
</tr>
</tbody>
</table>

4. Niching with Indicators

The question arises, if an analysis of the search can improve the optimization and clustering process. In this section we propose two clustering indicators that allow the recognition of new potential niches, and automatically trigger the clustering process. We show their behavior for the rake selection and DBSCAN hybridization introduced in Section 3.2.

4.1. Parameterized Cluster Indicator

The cluster indicator $\phi$, considers the genotypic and phenotypic distances between two individuals. A threshold $\theta$ for $\phi$ indicates, whether two solutions lie in the same or in different niches. The cluster indicator $\phi$ measures the relation between the distances of two solutions in decision space, and in objective space:

$$
\phi(x_1, x_2) = \frac{\|x_1 - x_2\|_2}{\|f(x_1) - f(x_2)\|_2}
$$

employing the Euclidean distance. The idea of the cluster indicator is based on the assumption that two neighbored individuals of one niche in objective space are also neighbored in decision space. Hence, if distance $d_1 = \|f(x_1) - f(x_2)\|_2$ of two neighbored solutions $x_1$ and $x_2$ in objective space, e.g., of neighbored rakes, is higher in decision space than the distance $d_2$ of two solutions that are more unlike in objective space, a novel niche has been found. Rake selection selects the closest non-dominated solution for each rake. Hence, the comparison for the cluster indicator can be restricted to solutions of neighbored rakes. Each new potential
niche has to be checked, if it is new or already known. If Equation 3 is fulfilled, DBSCAN is restarted, and the population is assigned to corresponding niches.

Important for the success of the indicator approach is the choice of threshold parameter $\theta$. If $\theta$ is chosen too small, too many cluster processes are started, and the indicator is not advantageous. If $\theta$ is chosen too large, the algorithm may not be able to find solutions from other niches, and computes a necessary cluster assignment too late, leading to an unnecessarily high number of fitness function evaluations.

Table 3. Results of a (100 + 200)-ES Rake-DBSCAN-hybrid with cluster indicator and various setting for $\theta$ on $SYM$-$PART2$ (upper part), and $SYM$-$PART3$ (lower part). The results also show the average number of iterations in each niche, and the average number of clustering calls.

<table>
<thead>
<tr>
<th>problem</th>
<th>$\theta$</th>
<th>runtime in sec</th>
<th>niching iter.</th>
<th>clust. calls</th>
<th>ffe</th>
<th>total</th>
<th>$S$</th>
<th>$S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SYM$-$PART2$</td>
<td>2</td>
<td>346.67</td>
<td>6.7</td>
<td>6.7</td>
<td>76,000</td>
<td>99.33</td>
<td>99.3</td>
<td></td>
</tr>
<tr>
<td>$SYM$-$PART2$</td>
<td>4</td>
<td>342.77</td>
<td>6.3</td>
<td>6.1</td>
<td>74,600</td>
<td>99.33</td>
<td>99.3</td>
<td></td>
</tr>
<tr>
<td>$SYM$-$PART2$</td>
<td>10</td>
<td>363.97</td>
<td>5.8</td>
<td>5.4</td>
<td>64,000</td>
<td>99.33</td>
<td>99.3</td>
<td></td>
</tr>
<tr>
<td>$SYM$-$PART3$ $[-15,15]$</td>
<td>2</td>
<td>2,267.60</td>
<td>34.6</td>
<td>33.9</td>
<td>479,600</td>
<td>99.32</td>
<td>99.17</td>
<td></td>
</tr>
<tr>
<td>$SYM$-$PART3$ $[-15,15]$</td>
<td>4</td>
<td>3,127.56</td>
<td>45.9</td>
<td>43.8</td>
<td>655,200</td>
<td>99.32</td>
<td>99.23</td>
<td></td>
</tr>
<tr>
<td>$SYM$-$PART3$ $[-15,15]$</td>
<td>10</td>
<td>3,376.61</td>
<td>41.3</td>
<td>35.4</td>
<td>582,200</td>
<td>99.32</td>
<td>99.28</td>
<td></td>
</tr>
<tr>
<td>$SYM$-$PART3$ $[-30,30]$</td>
<td>2</td>
<td>581.54</td>
<td>11.9</td>
<td>11.5</td>
<td>130,200</td>
<td>99.32</td>
<td>99.29</td>
<td></td>
</tr>
<tr>
<td>$SYM$-$PART3$ $[-30,30]$</td>
<td>4</td>
<td>419.79</td>
<td>9.2</td>
<td>8.1</td>
<td>94,600</td>
<td>99.33</td>
<td>99.29</td>
<td></td>
</tr>
<tr>
<td>$SYM$-$PART3$ $[-30,30]$</td>
<td>10</td>
<td>476.26</td>
<td>10.8</td>
<td>6.7</td>
<td>98,400</td>
<td>99.32</td>
<td>99.29</td>
<td></td>
</tr>
</tbody>
</table>

The upper part of Table 3 shows a parameter study of rake selection with DBSCAN, and cluster indicator on problem $SYM$-$PART2$. In average, the setting $\theta = 10$ achieved the best results in comparison to $\theta = 2$ or $\theta = 4$ in terms of fitness function evaluations, while achieving the same $S$-metric values. In comparison to the approach without cluster indicator, see Table 1 significant performance wins can be observed for all parameter combinations. The lower part of Table 3 shows the experimental results on $SYM$-$PART3$ for a (100 + 200)-ES. Qualitatively, the results are mostly the same like for $SYM$-$PART2$. An explanation for the difficulties of parameterization $L = -15$ and $U = 15$ are the small distances between the three Pareto-sets with negative second function value. It is a difficult task to differentiate between the niches. Again, a reduction of population sizes leads to a significant speedup. Significant savings of clustering activities (i.e., not clustering calls, but runtime of clustering processes) are the main reason for the improvements. To summarize, the introduction of the cluster indicator leads to a significant performance win.

4.2. Parameter-Free Indicator

In the previous section we have introduced the cluster indicator based on a threshold $\theta$ to trigger the clustering processes. In the following, we introduce a parameter-free
variant. Let \( x_{i+1} \) be the nearest, and \( x_{i+2} \) be the second nearest point of solution \( x_i \) in objective space. If
\[
||x_i - x_{i+2}||_2 < ||x_i - x_{i+1}||_2,
\]
i.e., \( x_{i+2} \) is closer to \( x_i \) in decision space than \( x_{i+1} \) to \( x_i \), then \( x_{i+1} \), and \( x_{i+2} \)

![Diagram](image)

**Fig. 2.** Illustration of parameter-free indicator: if the distance in design space between two neighbored solutions in decision space \( x_i \) and \( x_{i+1} \) is higher than the distance in design space to a third solution \( x_{i+2} \), it is probable that \( x_{i+1} \) belongs to another niche.

probably do not belong to the same cluster, but \( x_i \), and \( x_{i+2} \) do. If this condition is fulfilled, a new cluster process is reasonable.

In the case of rake selection this operator is easy to implement assuming \( m = 2 \) and \( k \) solutions, when \( k \) is also the number of rakes belonging to one niche: If the distance in decision space of a solution \( x_i \) selected by rake \( r_i \) to solution \( x_{i+1} \) belonging to the neighbored rake \( r_{i+1} \) is higher than the distance of solution \( x_i \) to solution \( x_{i+2} \) belonging to rake \( r_{i+2} \) that is neighbored to \( r_{i+1} \) (which is not \( r_i \) itself), a new cluster process is started, and the solutions are assigned to niches, see Figure 2. Hence, only \( k \) comparisons are necessary. The advantage of this heuristic is the independence of scalings in solution and objective space. Not the distance values, but the relative comparisons are basis of the parameter-free indicator.

Table 4 shows the experimental results of the niching technique with the parameter-free indicator, and adaptive corner points on problems \( SYM\text{-}PART2 \) and \( SYM\text{-}PART3 \) for a \((100 + 200)\)-ES. The approximation process is slower than the optimization with the parameterized cluster indicator of Section 4.1, but similar \( S \)-metric values were achieved. The higher number of fitness function evaluations can be explained as follows. The number of clustering calls show that fewer clustering processes are started. Although all niches have been found, less effort is spent
A Clustering-Based Niching Framework for the Approximation of Equivalent Pareto-Subsets

Table 4. Results of the niching approach with a $(100 + 200)$-ES with the parameter-free indicator and adaptive corner points on $SYM\text{-}PART2$ (upper part) and various parameterization of $SYM\text{-}PART3$ (lower part), averaged over 10 runs.

<table>
<thead>
<tr>
<th>problem</th>
<th>runtime in sec</th>
<th>ffe</th>
<th>total $S$</th>
<th>$\varnothing$</th>
<th>$S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SYM\text{-}PART2$</td>
<td>572.31</td>
<td>15.3</td>
<td>6.8</td>
<td>142,600</td>
<td>99.32</td>
</tr>
<tr>
<td>$SYM\text{-}PART3 \begin{bmatrix}-15,15\end{bmatrix}$</td>
<td>4.039.25</td>
<td>88.5</td>
<td>26.5</td>
<td>1,085,200</td>
<td>99.32</td>
</tr>
<tr>
<td>$SYM\text{-}PART3 \begin{bmatrix}-20,20\end{bmatrix}$</td>
<td>1.150.75</td>
<td>31.7</td>
<td>10.4</td>
<td>295,400</td>
<td>99.32</td>
</tr>
<tr>
<td>$SYM\text{-}PART3 \begin{bmatrix}-30,30\end{bmatrix}$</td>
<td>1.153.77</td>
<td>51.5</td>
<td>5.1</td>
<td>394,600</td>
<td>99.32</td>
</tr>
</tbody>
</table>

on optimization within these niches. Hence, if a reasonable estimation of the cluster indicator $\theta$ is possible, the niching approach should use this measure to control the balance between optimization and clustering. If $\theta$ cannot be estimated, the parameter-free cluster indicator guarantees a good $S$-metric result, which has to be paid with a slight loss in performance.

5. Conclusions

Without diversity enhancing mechanisms that support the optimization process, standard evolutionary multi-objective optimization techniques may fail to detect all Pareto-subsets reliably. Furthermore, they fail to cover found subsets satisfactory, a necessity to allow the choice among solutions with equivalent balance of objectives. In our work we have proposed a flexible clustering-based framework, open for the integration of arbitrary clustering methods and EMOAs. In a case-study we have shown two instances of hybridizations that fit into the framework. Both approaches were successful in optimizing a test problem known in literature with a varying degree of difficulty. Furthermore, we could show how the performance can be increased with indicators that control the clustering process. The approach with cluster indicator $\phi$, and threshold $\theta$ turned out to be the most successful variant. If threshold $\theta$ is not easy to determine, the parameter-free indicator should be employed.

As not many test problems with equivalent Pareto-subsets exist, it must be future work to develop a test suite of multi-objective problems with equivalent Pareto-subsets that may also come from practical applications. Properties like non-convex and intertwined Pareto-sets will be tested as a next step in this line of research. As such data distributions belong to the strengths of density-based clustering techniques, we can expect that methods like DBSCAN and kernel density clustering will be capable to solve this problem class as well.

Appendix A. Definition of Test Problem $SYM\text{-}PART$

$SYM\text{-}PART$ introduced by Rudolph et al.\cite{Rudolph2001} exhibits a Pareto-front with disconnected Pareto-sets in decision space. Each set covers the whole Pareto-front. The
points of the Pareto-set of each section lie on a line. The complexity of SYM-PART can be controlled with the following manipulations:

1. Scaling of each section,
2. shift of each section,
3. rotation of each section by angle $\omega$, and
4. transformation, i.e., control of the individual size of each section.

A simple version of SYM-PART consists of two objective functions, and a two-dimensional decision space:

$$f_1(x_1, x_2) = (x_1 + a)^2 + x_2^2,$$

(A.1)

and

$$f_2(x_1, x_2) = (x_1 - a)^2 + x_2^2,$$

(A.2)

for arbitrary $a > 0$, and with Pareto-set:

$$P^* = \{x \in \mathbb{R}^2 : x = (x_1, 0)' \text{ with } x_1 \in [-a, a]\},$$

(A.3)

and Pareto-front:

$$F^* = \{z \in \mathbb{R}^2 : z = (4a^2\nu^2, 4a^2(1 - \nu)^2) \text{ with } \nu \in (0, 1)\}.$$

(A.4)

In this simple definition the Pareto-set covers a line of length $2 \cdot a$, the vertical distances between the sections are determined by parameter $b$, horizontal distances by parameter $c$. In the next step the function is distributed to $3 \times 3$-sections in decision space. The section size is determined by width $2 \cdot a + c$ and height $b$. The values are mapped to set $\{-1, 0, 1\}$ as follows:

$$t_i = \text{sgn}(\hat{t}_i) \times \min\{|\hat{t}_i|, 1\}.$$

(A.5)

The function is symmetric and the zero-point lies in the middle of the Pareto-set of section $(0, 0)$. Hence, the whole decision space lies between $((-3a - \frac{3}{2}c), -\frac{3}{2}b)$ and $((3a + \frac{3}{2}c), \frac{3}{2}b)$. In the following, we will use the fixed parameterization of $a = 1$, $b = 10$, $c = 8$, and a decision space between $(-15, -15)$ and $(15, 15)$. Based on the preliminary equation, function SYM-PART1 is defined by:

$$f_s^{(1)}(x_1, x_2) = f_s(x_1 - t_1(c + 2a), x_2 - t_2b)$$

(A.6)

To rotate the function, the following rotation matrix is used:

$$r(x) = \begin{pmatrix} \cos \omega & -\sin \omega \\ \sin \omega & \cos \omega \end{pmatrix} \cdot x$$

(A.7)

with rotation angle $\omega$, we use $\omega = 45^\circ$. The function SYM-PART2 sounds as follows:

$$f_s^{(2)}(x_1, x_2) = f_s^{(1)}(r_1(x), r_2(x)).$$

(A.8)

A further extension of SYM-PART2 makes the optimization problem even more difficult. The transformation

$$d(x_1, x_2) = x_1 \times \left(\frac{x_2 - L + \varepsilon}{U - L}\right)^{-1}$$

(A.9)
for small $\varepsilon > 0$, $U$ as upper bound and $L$ as lower bound is the basis of SYM-PART3, which is defined as follows:

$$f_{\ast}^{(3)}(x_1, x_2) = f_{\ast}^{(2)}(d(x_1, x_2), x_2).$$  \hspace{1cm} (A.10)

References


A Clustering-Based Niching Framework for the Approximation of Equivalent Pareto-Subsets

