Non-monotonicity of Obtained Hypervolume in 1-greedy S-Metric Selection

Leonard Judt*, Olaf Mersmann*, and Boris Naujoks*

*Faculty of Statistics, TU Dortmund University, 44221 Dortmund, Germany
*Institute of Computer Science, Cologne University of Applied Sciences, 51643 Gummersbach, Germany

ABSTRACT
The progression of the dominated hypervolume in the course of the optimization process, with respect to a global reference point, is thought to be monotonically increasing. This intuition is based on the observation that in each iteration the solution that contributes the least to the dominated hypervolume is eliminated. Derived from results of multiple SMS-EMOA runs with incorporated reference point adaptation, we show that this does not always hold for 2- and 3-dimensional objective spaces. For the 2-dimensional case, this is because the two boundary solutions are always retained in the population regardless of their hypervolume contribution. For the 3-dimensional case we are able to show that the cause of the drop in dominated hypervolume is the continuous adaption of the reference point.

KEY WORDS: Hypervolume Selection, Decreases in Hypervolume Progression, SMS-EMOA, Evolutionary Multiobjective Optimization

Contents

1 Introduction 2
2 Hypervolume & SMS-EMOA 3
3 Experiments 4
   3.1 Experiment 1: 2D-dim. case / boundary solution anomaly . . . . . . . . . . . . . . 4
   3.2 Experiment 2: 3-dim. case / Differences fixed and adaptive reference point . . 12
4 Conclusion & Outlook 15
1 Introduction

In recent years, Evolutionary Multi-objective Optimization (EMO, cf. Deb (2001); Coello Coello et al. (2007)) became a well renowned research field. Combining the two fields of Multi-Criteria Decision Making (MCDM, cf. Miettinen (1999); Ehrigott (2005)) and Evolutionary Computation (EA/EC, cf. Eiben and Smith (2003); DeJong (2006)), it became a flourishing combination of both fields of research. Originating in EMO, essential results and interesting new search direction have been carried over to MCDM as well as EA research.

Since comparing results in MCDM is not as easy as it is in single-objective optimization, performance indicators deserve special interest in MCDM and EMO. In the latter, such performance indicators can also be used as selection criteria by the optimization procedure. One example of this is the dominated hypervolume (cf. Zitzler and Thiele (1998); Zitzler (1999)), which has become one of the standard performance indicators in the past years.

In EMOA, the dominated hypervolume is at the core of the selection process of many modern algorithms, e.g. the IBEA (cf. Zitzler and Künzli (2004)), SMS-EMOA (cf. Beume et al. (2007)), MO-CMAES (cf. Igel et al. (2007)), or HypE (cf. Bader and Zitzler (2011)). In the SMS-EMOA, i.e. S-Metric Selection – EMO Algorithm (cf. Emmerich et al. (2005); Naujoks et al. (2005)), the individual contributing least to the dominated hypervolume of the set of current solutions, the population, is replaced in the next generation. Due to the exponentially increasing runtime complexity of the maximal-hypervolume subset selection w.r.t. the problem dimensionality, a \((\mu + 1)\) or 1-greedy selection mechanism (cf. Zitzler et al. (2008)) is preferred.

By employing a 1-greedy selection scheme as detailed above, it is thought that the overall measure of hypervolume w.r.t a global reference point is impossible to decrease. However, the work at hand shows that this is not the case on basis of multiple 2-dimensional (2-dim.) and 3-dimensional (3-dim.) SMS-EMOA runs. We present the main observations of these runs and state why decreases in hypervolume are possible.

Hypervolume selection has also been been extended to serve more complex selection tasks. For example, weights can be incorporated in the hypervolume calculation to implement user preferences in the selection process (Zitzler et al. (2007)). Moreover, the normal hypervolume calculation has been adopted to consider decision space diversity in addition to objective space diversity (Ulrich et al. (2010)). The integration of only objective space diversity in EMO is a major contrast to single-objective EA. In the latter, a lot of effort has been spend on decision space diversity preservation, which was almost neglected in EMO research in the last years.

In detail, the article is organized as follows: The following section will first introduce the basic indicator used in our investigation and how it is integrated into the EMO algorithm under test. Thereafter, section 3 presents the results for the 2-dim. test cases starting with initial observation, frequencies of the observed phenomenon and a possible explanation (section 3.1). The same analysis is performed in section 3.2 for the 3-dim. test cases to illustrate the findings for higher dimensions. Although the reason is thought to be different, a comparable phenomenon is observed.

The paper is summarized in section 4 where an outlook with ideas for promising future investigations is also given. These would provide more insight into how and why this phenomenon affects the optimization process.
2 Hypervolume & SMS-EMOA

The hypervolume measure or \( S \)-Metric is a unique performance measure for EMOA performance introduced by Zitzler (cf. Zitzler and Thiele (1998); Zitzler (1999)). It measures the hypervolume that is spanned by a Pareto-front, mostly received as a result of an EMOA run, with respect to a pre-defined reference point. To this end, the hypervolume measures the convergence of a set of non-dominated solutions as well as their spread rewarding a better convergence and a uniform spread. The reference point is required for the hypervolume calculation and even early publications hint to the huge impact a proper choice of it might have.

The definition of the hypervolume goes back to the Lebesque measure of the union of rectangles \( a_i \) spanned by the reference point and one non-dominated point \( m_i \) of the considered set \( M \). It thus reads

\[
\Lambda \left( \bigcup \{a_i | m_i \in M \} \right)
\]

with \( \Lambda \) being the Lebesque measure of the given set.

One rather popular example considering the hypervolume as a selection criterion in an EMOA is the SMS-EMOA. Since the hypervolume was a very popular and turned out to be a very good performance indicator, the idea to use it for selection in an EMOA was near-by. The SMS-EMOA invokes the hypervolume in an easy and straight forward way adjusting the selection scheme to a greedy \((\mu + 1)\) scheme (Beume et al. (2007)).

This was new since single objective algorithms featuring a \((\mu + \lambda)\) selection scheme normally implement an offspring surplus, i.e. \( \lambda > \mu \). In EMO algorithms, \( \mu = \lambda \) was chosen in most prominent instances like NSGA-II (cf. Deb et al. (2000, 2002a)) or SPEA2 (cf. Zitzler et al. (2001, 2002)).

Choosing \( \lambda = 1 \) was due to the computational effort for the hypervolume calculation. This calculation was proven to be \( \#P \)-hard (cf. Bringmann and Friedrich (2008)), i.e. with increasing objective space dimension, the computational effort to calculated hypervolumes of sets grows rather rapidly. Nevertheless, quite efficient algorithms for the 2-dim.- and 3-dim. cases exist (cf. Emmerich and Fonseca (2011)).

The pseudocode of the SMS-EMOA is depicted in algorithm 1. In short, one new solution is generated by some arbitrary variation operator and the solution with the least hypervolume contribution (from the last Pareto front from the non-dominated sorting procedure) is truncated from the set of individuals. The latter is performed by the algorithm \( \text{Reduce} \), depicted in algorithm 2. This way, some monotonicity of the sequence of hypervolume values during on optimization run is expected. Due to the advantageous properties of the hypervolume, it is desired as well.

Since the handling of boundary solutions at the edge of Pareto fronts plays a major role in the forthcoming experiments, this should be explained in more detail. In the 2-dim. case, such solution can be easily be identified by an extreme value in one of the objectives. In the original SMS-EMOA (cf. Emmerich et al. (2005)), the solutions are always kept in the population. This is equivalent to assigning an infinite fitness function value to these solutions.

In the 3-dim. case, boundary solutions cannot be identified that easily. Consequently, there was no special treatment of these implemented in the original 3-dim. SMS-EMOA (cf. Naujoks et al. (2005)). Additionally, the definition of a fixed reference point led to problems in the 3-dim. case due to numerical instabilities for very small hypervolume contribution and a far away
Algorithm 1: SMS-EMOA

\[ P_0 \leftarrow \text{init}() \]

// Initialize random start population of \( \mu \) individuals:
\[ t \leftarrow 0 \]

repeat

// Generate one offspring by variation operators:
\[ q_{t+1} \leftarrow \text{generate}(P_t) \]

// Select \( \mu \) individuals for the new population:
\[ P_{t+1} \leftarrow \text{Reduce}(P_t \cup \{q_{t+1}\}) \]
\[ t \leftarrow t + 1 \]

until stop criterium reached

Algorithm 2: Reduce\((Q)\)

// All I non-dominated fronts of \( Q \):
\[ \{R_1, \ldots, R_I\} \leftarrow \text{fast-nondominated-sort}(Q) \]

// Find element of \( R_I \) with lowest \( \Delta_S(s, R_I) \):
\[ r \leftarrow \arg \min_{s \in R_I}[\Delta_S(s, R_I)] \]

// Eliminate \( r \) from \( Q \):
\[ Q' \leftarrow Q \setminus \{r\} \]

return \( Q' \)

reference point. Due to this, an adaptation scheme for the reference point was developed that allows for a control of the fraction of boundary points kept by the algorithm (cf. Naujoks et al. (2005)).

At first in this adaptation scheme, the points yielding an extreme value regarding one objective are identified in each generation. These extremal values in each objective are assembled and a predefined vector is added to guarantee a positive hypervolume value for the solutions yielding these extremal values. By default, this predefined vector is \(1^n\) with \(n\) being the objective space dimension.

3 Experiments

The results below are reported according to the scheme suggested by Preuss (2007). It forces the authors to concentrate on one aspect of an investigation at a time. These aspects are called experiments and the proposed scheme urges the authors to describe the setup in detail and to distinguish between objective observations and more subjective interpretations. Since we report on phenomena that are not fully explained and understood, our contribution places a higher emphasis on the visualization part of the scheme than the discussion part.

3.1 Experiment 1: 2D-dim. case / boundary solution anomaly

Pre-experimental planning: All SMS-EMOA runs reported here stem from a large study on the influence of the number of offspring on the performance of the SMS-EMOA (Judt
Figure 1: Course of hypervolume values received from an SMS-EMOA run on ZDT1 with a population size of $\mu = 100$. Only the interval between 16 800 and 18 000 fitness function evaluations is depicted to emphasize the hypervolume decrease after about 17 800 fitness function evaluations.

(2011)). During this study, an enormous number of runs were conducted for different values of $\lambda$ within a greedy $(\mu + \lambda)$ selection scheme. Indeed, first drops of hypervolume values were observed for $\lambda > 1$ and the questions whether this phenomenon also arises for $\lambda = 1$ led to the following results.

**Task:** Detect drops in the the dominated hypervolume of successive generations over the course of a SMS-EMOA run when the hypervolume is calculated with respect to a fixed global reference point. This is to be carried out for multiple test functions with 2-dim. objective space. Record the number of drops and in which generation they occur.

**Setup:** The SMS-EMOA with a $(\mu + 1)$-selection scheme is analyzed on the four bi-objective test functions ZDT1 – ZDT4 (cf. Zitzler et al. (2000)) with 30 decision variables each. Each run is terminated after 100 000 function evaluations. Different parameter sets are tested as follows: three population sizes $\mu \in \{10, 20, 100\}$ are examined and a Latin Hypercube Sample (Santner et al. (2003)) is used to choose 25 different combinations for the variation operator. Simulated Binary Crossover and Polynomial Mutation, cf. Deb (2001) for a description, have been invoked. The corresponding intervals for the different parameters have been set to $\eta_c \in [0, 40]$, $p_c \in [0, 1]$, $\eta_m \in [0, 40]$ and $p_m \in [0, 0.3]$.

For each combination of test function, population size and variation operator set, 50 independent runs are conducted. In total, 15 000 runs are performed in this experiment. The dominated hypervolume with regard to a fixed reference point is calculated and stored for each generation. The number of times the hypervolume drops is counted. To calculate the hypervolume values, $[11, 11]$ is chosen to be the global reference point for ZDT1 – ZDT3 and $[11, 700]$ is separately used for the ZDT4 test function in order for all feasible solutions in the optimization process to be dominated by the reference point.
Table 1: Estimated probability (frequency) of a drop in dominated hypervolume w.r.t. a fixed reference point at any point in time of the optimization process. Results are shown separately for each 2-dim. test function and population size \( \mu \). The estimates are based on 20,000 fitness function evaluations of 1,250 runs on each test function.

<table>
<thead>
<tr>
<th>Test Function</th>
<th>( \mu = 10 )</th>
<th>( \mu = 20 )</th>
<th>( \mu = 100 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZDT1</td>
<td>0.00083</td>
<td>0.00093</td>
<td>0.00343</td>
</tr>
<tr>
<td>ZDT2</td>
<td>0.00291</td>
<td>0.00346</td>
<td>0.00549</td>
</tr>
<tr>
<td>ZDT3</td>
<td>0.00086</td>
<td>0.00096</td>
<td>0.00416</td>
</tr>
<tr>
<td>ZDT4</td>
<td>0.00195</td>
<td>0.00288</td>
<td>0.00789</td>
</tr>
</tbody>
</table>

Figure 2: Histogram of 1,250 SMS-EMOA runs showing the number of hypervolume drops vs. the generation when these take place. Here, the results for ZDT1, \( \mu = 100 \) and different parameter settings are plotted. Note that the \( y \)-axis is in logarithmic scale.

**Experimentation/Visualization:** Figure 1 depicts the course of hypervolume values received for ZDT1 and a population size of \( \mu = 100 \) between 16,800 and 18,000 fitness function evaluations.

In table 1 the frequencies of hypervolume drops is listed. These are broken down by the test functions and the different population sizes \( \mu \). Each frequency represents the number of generations with an observed drop in hypervolume in relation to the entire number of generations computed averaged over the 1,250 runs.

Figure 2 illustrates in which generations the hypervolume decreases occurred for the ZDT1 and \( \mu = 100 \) runs. Each bar represents the number of drops in an interval of generations. Complementing the overview of the obtained results, figure 6 shows the same information as figure 2 for all 2-dim. test functions considered and all population sizes \( \mu \). Figure 2 is therefore equivalent to the upper right corner of figure 6.

Figure 3 shows the magnitude of the drops in hypervolume for ZDT1, \( \mu = 100 \) after 20,000 fitness function evaluations. Each bar represents the number of drops that have similar amounts of decrease. Both axes are scaled logarithmically in order to better present the wide range of results. Figure 5 summarizes the same results of all test functions. To this end, figure 5 is
Hypervolume Non-monotonicity

Figure 3: Histogram of 1 250 SMS-EMOA runs showing the number of hypervolume drops vs. the magnitude of the drop, i.e. the difference in dominated hypervolume of the population before the drop and the following one. Here, the results for ZDT1, $\mu = 100$ and different parameter settings after 20 000 fitness function evaluations are plotted. Note that both axes have a logarithmic scale.

Observations: Figure 1 exemplifies that the progression of the dominated hypervolume with respect to a global reference point is not necessarily monotonically increasing. After about 17 800 function evaluations a decrease in hypervolume can be noticed. This decrease is significantly larger than the single increases previous to and following on the event. It can obviously be detected in the figure and it took about 150 generations to recover the amount of hypervolume from before the drop.

It can be seen from table 1 that decreases in hypervolume are more likely to appear with a higher population size $\mu$. Also, these drops were more frequently observed at runs on the test function ZDT2 with a concave Pareto front and on ZDT4, which has multiple local Pareto fronts that act as obstacles in the optimization process.

As presented in figures 2 and 6, the decreases happen more often during the start of the optimization process. In fact, most decreases are recorded in the first 2 000 generations. Nevertheless, decreases do take place during all phases of the optimization run. Additional runs of 100 000 generations, shown in figure 7, indicate that even at the end of such long optimization runs decreases do occur regularly. Note that due to the log axis in all these plots, the area of the bars do not directly reflect the frequency of a drop occurring.

The most interesting behavior can be observed on ZDT2 with $\mu \in \{10, 20\}$. Here, the amounts of drops decrease rather rapidly between 2 000 and 10 000 fitness function evaluations. Thereafter, the progression is rather constant on a level between $10^1$ and $10^{1.5}$ decreases per unit. Since all other subfigures depict a monotonically decreasing progression, the situation on ZDT2 and ZDT4 for $\mu = 100$ has to be mentioned. Here, the progression displays a plateau between 5 000 and 7 000 fitness function evaluations.
Figures 3 and 5 show that the drops observed have a wide range in magnitude. Most drops in hypervolume are of size between $10^{-4}$ and $10^{-2}$. Drops of higher magnitude are rarely observed, yet drops of smaller magnitude are likely to be seen. Decreases that are below a certain threshold are rounded up to that threshold due to computational reasons. This threshold depends on the scale of the chosen global reference point. For ZDT1 – ZDT3, the threshold lies in the region of $10^{-14}$. Therefore, the peak at the left of each histogram can be ignored.

Furthermore, the drops tend to be of lower magnitude when the population size $\mu$ is larger.

**Discussion:** Figure 4 displays the cause for the decrease in hypervolume at bi-objective optimization problems observed in this experiment. The figure shows three blue points that span a 2-dim. Pareto front in a minimization problem and one green point, representing the offspring in a generation. The new individual is also considered as a boundary point as it is best in one objective. Despite the fact that this offspring would virtually contribute the least in hypervolume, its hypervolume contribution is deliberately assigned an infinite value, so that all boundary points are being conserved in $S$-metric selection with two objectives. Due to this exceptional rule, the hypervolume attained after selection is then lower as the hypervolume reached before the offspring was generated. Thus, a decrease in hypervolume is possible in bi-objective optimization problems.

It has to be noted that these decreases appear more frequently at the start of the optimization process as it appears in figures 2 and 6, presumably because the points generated have not yet converged to the true Pareto front and are situated rather close to each other, resulting in a higher probability of new boundary solutions being generated. Later in the process, there is a higher spread of points along the Pareto front, causing a smaller likelihood of the existing boundary points being overtrumped by offspring.

As seen in figures 3 and 5, the size of the drops range from $10^{-14}$ (presumably even lower) and $10^{-2}$. At first, these amounts of decrease may appear insignificant. However, these drops are of similar scale to the increases attained, which continue to appear more frequently. The drops observed in runs with a higher population size $\mu$ tended to be of lower magnitude. This is probably due to the fact that the neighbors of the boundary points are located closer to the boundary points and therefore the hypervolume contributions gained are also smaller.
Figure 5: Histograms illustrating the magnitude of the drops in hypervolume. For each of the 2-dim. test cases and each $\mu$ a figure is plotted. It gives the number of hypervolume drops over all runs on this function-$\mu$ combination regardless of further parametrization. Both axes are log-scaled.
Figure 6: Histograms indicating when hypervolume drops take place in the progression of an optimization run. For each of the 2-dim. test cases and each $\mu$ a figure is plotted. It gives the number of hypervolume drops over all runs on this function-$\mu$ combination regardless of further parametrization.
Figure 7: Histograms indicating when hypervolume drops take place in the progression of an optimization run as shown in figure 6 but for runs with 100,000 fitness function evaluations.
3.2 Experiment 2: 3-dim. case / Differences fixed and adaptive reference point

Pre-experimental planning: Since the current experiment extends the results received for the 2-dim. test cases, these experiments on lower dimensional test cases build the groundwork for the following ones. Again, some of the results stem from the afore mentioned study for $\lambda \geq 1$ in $(\mu + \lambda)$ - SMS-EMOA (Judt (2011)). In addition, runs with the same settings have been conducted for further 3-dim. test cases.

Task: Verify whether drops in the progression of the hypervolume do also occur for higher dimensional, here 3-dim., test cases. Observe how often and when such drops are recorded. Since boundary solutions within the SMS-EMOA as well as a reference point need to be invoked differently from the 2-dim. case, what are the reasons for such decreases here?

Setup: Except for the test functions, the setup from the previous 2-dim. experiment is not changed. The test functions examined in this experiment are DTLZ1, DTLZ2, and DTLZ3 (cf. Deb et al. (2002b)). The global reference points $[1000, 500, 500]$ for DTLZ1, $[11, 11, 11]$ for DTLZ2 and $[1000, 1000, 1000]$ for DTLZ3 are chosen in order to calculate the hypervolume values. The number of decision variables is reduced to 7 and 12 as suggested by Deb et al. (2002b), in order for each optimization run to converge. Thus, for each population size $\mu$ and test function 2 500 runs are conducted, making a total of 22 500 runs in this 3-dim. experiment.

Experimentation/Visualization: Comparable figures and tables to the ones from the previous experiment have been conducted again to demonstrate the proximity of results. To this end, figure 8 first provides an example for drops on the progression of the hypervolume for DTLZ2, $\mu = 100$ in the interval 1 000 to 5 000 fitness function evaluations. Like table 1, table 2 summarizes the frequencies of hypervolume drops for the three 3-dim. test functions and different $\mu$ values.

Illustrative subfigures are not highlighted this time, so figure 11 provides an overview on the amount of drops over the generation when it takes place for all 3-dim. test functions and all population sizes $\mu$. This is similar to figure 6 for the 2-dim. case.

In addition, this figure for the situation after 20 000 fitness function evaluations is again complemented by the figure showing the situation after 100 000 function evaluations (cf. figure 12).

Furthermore, a figure depicting the extent of hypervolume decreases is conducted on basis of the results of the 3-dim. test cases. This is figure 10, which is similar to the 2-dim. situation in figure 5. For every 3-dim. test function and every population size $\mu$ a histogram shows the magnitude of decrease observed during 20 000 fitness function evaluations.

A possible explanation for the decreases in the 3-dim. case is illustrated in figure 9. It provides a figure of a 3-dim. example taken from some arbitrary run. The lower row displays the whole dominated hypervolume (blue) with the contributions of two single solutions marked differently (w.l.o.g. solution 1: red, solution 2: green). In the upper row, the dominated hypervolume of other solutions is omitted (the blue part) and only the contributions of solutions 1 and 2 are depicted.
Figure 8: Course of hypervolume values received from an SMS-EMOA run on DTLZ2 with a population size of $\mu = 100$. Only the interval between 1 000 and 5 000 fitness function evaluations is depicted to emphasize the bouncing hypervolume values between 1 200 and 2 200 fitness function evaluations.

Table 2: Estimated probability (frequency) of a drop in dominated hypervolume w.r.t. a fixed reference point at any point in time of the optimization process. As in table 1, results are shown separately for each 3-dim. test function and $\mu$. The estimates are based on the first 20 000 FE of 1 250 runs on each test function.
While the left column of the figure depicts the situation during an algorithm run with respect to a dynamic reference point, the right part exhibits the situation with respect to a fixed reference point, \([11, 11, 11]\) for DTLZ2 here. This latter situation is akin to calculating hypervolume indicator values to compare multiple algorithm results. In this setting, a static reference point must obviously be chosen a-priori.

**Observations:** In contrast to the observations made in the 2-dim. experiment, the frequencies seen in table 2 show that decreases in hypervolume are less likely to appear with a higher population size \(\mu\). Much higher amounts of drops were observed at runs on the test functions DTLZ1 and DTLZ3. Both functions are characterized by multiple local Pareto fronts as opposed to DTLZ2.

When taking a look at figure 10, one sees that decreases with a higher magnitude are achieved when the population size \(\mu\) is lower. Also, bigger drops were observed at DTLZ1 and DTLZ3. The peak bar at the left side of each histogram should again be ignored. For decreases of the size below a certain threshold, the drop in hypervolume is rounded up to that threshold for computational reasons.

Figure 11 shows a very interesting phenomenon. Partly due to the logarithmic scale of the \(y\)-axis, the negative slope towards the end of the optimization runs are very small. Nevertheless, some of configurations do not show the expected negative slope towards this end. This expected behavior was also supported by the 2-dim. experiments and figure 6. Moreover, the count of decreases does not fall below \(10^2\) in the 3-dim. case.

In contrast to figure 6, figure 11 shows a more stable or continuous behavior. Although the expected decrease takes place for lower population sizes \((\mu \in \{10, 20\})\), in particular the results for \(\mu = 100\) depict a different characteristic. Here, the amounts first drop rather frequently but recover after approx. 10,000 fitness function evaluations. After crossing this valley, the progressions show a stable attitude. Surprisingly, a small positive slope can be recognized in the figures, which is even more interesting having the logarithmic scale of the axes in mind. Consequently, a real increase in the amounts of decreases takes place.

Figure 12 depicts the situation after 100,000 fitness function evaluations. A slight decrease can be observed over all generations in the majority of runs. The only configuration showing a real negative slope toward the end of this long optimization runs, after an increasing behavior over the first 20,000 fitness function evaluations, is \(\mu = 100\) for DTLZ2. On DTLZ1 and DTLZ3, \(\mu = 100\), the subfigures depict a rather continuous behavior after crossing the valley of the first 7,500 generations (cf. figure 11). However, this is still not in compliance with the expected behavior of decreasing counts of hypervolume drops.

Interestingly, the solutions providing the least hypervolume contribution change if a dynamic or a fixed reference point is considered. In the left part of figure 9 the contribution of solution 1 (red) is smaller than the contribution of solution 2 (green). In the right part considering a large, fixed reference point, the amount of the dominated hypervolume of solution 1 is increased due to the increased distance to the reference point. As a result, the hypervolume contribution of solution 1 is greater than the contribution of solution 2 in this situation.

**Discussion:** The reason for decreases in hypervolume in the 3-dim. case is somehow different from the reason for the effect in the 2-dim. case. Nevertheless, boundary solutions are
involved in both cases.

A possible explanation for decreases in the 3-dim. case is the algorithmic handling of the reference point during an optimization run. While internally a dynamic reference point is considered, the hypervolume of consecutive generations are calculated using a fixed reference point. As it is explained by figure 9 this can lead to different solutions providing the least hypervolume contribution. Thus, different hypervolume values for solutions exist, what leads to possible drops in generation’s hypervolume values depending on which reference points, hypervolume values respectively, are considered.

As seen in figure 10, drops in hypervolume with a higher magnitude are achieved when the population size $\mu$ is lower. This is most likely due to the fact, that the objective space is less crowded and therefore all hypervolume contributions are higher. Also, keeping in mind that all of the 3-dim. test functions use different global reference points and thus, have a different scale in hypervolume, the fact that drops of larger magnitude are to be seen with DTLZ1 and DTLZ3 is not surprising.

As mentioned previously, looking at when hypervolume decreases take place in the optimization process, figure 11 shows a more stable behavior in comparison to the 2-dim. case depicted in figure 6 and figure 7, even when not neglecting the logarithmic scale of the $y$-axis. Figure 11 also supports this observation. This demonstrates that hypervolume decreases in the 3-dim. case are relatively independent of convergence to the Pareto-front as opposed to the 2-dim. case.

4 Conclusion & Outlook

This article, to our knowledge, is the first publication emphasizing hypervolume decreases during hypervolume selection based EMOA. Due to the inherent greediness of the widely employed $(\mu+1)$ selection scheme used in most hypervolume selection EMOA, the progression of the hypervolume was expected to monotonically increase. That is not always the case, as was shown in this investigation.

Reproducible results are presented for different 2-dim. and 3-dim. test cases, namely different ZDT and DTLZ test functions. Moreover, different parameterizations were tested to exclude these as a possible cause for the observed effect.

Possible explanations for the test functions yielding different objective space dimensions are presented. In the 2-dim. case, the handling of boundary (extremal) solutions is thought to be the root cause for the effect. In the 3-dim. situation, this is not the case. Instead, the reference point adaptation scheme, which is necessary to avoid numerical instabilities, is the likely culprit. However, it remains unclear, if all the observed decreases are due to these reasons.

At this point it should be noted that during the investigation the numerical stability of current implementations of hypervolume algorithms was at times questionable. Most of the spikes that can be observed in the histograms can be traced back to lack of numerical precision. This aspect, while not the focus of this article, should receive further treatment from the community in general and from algorithm designers in particular.

Based on the results of the 3-dim. case, a generalization to a higher number of objectives is apparent. Empirical results from hundreds of runs on different test problems show that this effect occurs rather frequently. Moreover, it might be hard for the algorithm to regain the
Figure 9: HV contributions on DTLZ2 ($\mu = 10$) with respect to an adaptive reference point as handles internally in the SMS-EMOA and hypervolume contributions with respect to a fixed reference point (11, 11, 11) like considered for performance comparison.
Figure 10: Histograms illustrating the magnitude of the drops in hypervolume observed during 20,000 fitness function evaluations. For each of the 3-dim. test cases and each $\mu$ a figure is plotted. It gives the count of hypervolume drops over all runs on this function-$\mu$ combination regardless of further parametrization. Both axes are log-scaled.
Figure 11: Histograms indicating when hypervolume drops take place in the progression of 2500 optimization runs, each consisting of 20000 fitness function evaluations. For each of the 3-dim. test cases and each $\mu$ a figure is plotted. It gives the count of hypervolume drops over all runs on this function-$\mu$ combination regardless of further parametrisations.
Figure 12: Histograms indicating when hypervolume drops take place in the progression of 2,500 optimization runs, each consisting of 100,000 fitness function evaluations. For each of the 3-dim. test cases and each $\mu$, a figure is plotted. It gives the count of hypervolume drops over all runs on this function-$\mu$ combination regardless of further parameterizations.
hypervolume loss due to the observed effects. As a consequence, the working principle of such algorithms, in particular the (adaptive) choice of the reference point, might need to be revised.

Constantly adapting the reference point implies that the EMOA is in fact solving a dynamic optimization problem. The final set of points returned by any run will only be “optimal” for the last reference point. Therefore, comparing algorithm runs solely based on their dominated hypervolume w.r.t a fixed reference point, as is done in benchmarking studies, might not be the best strategy to choose a good algorithm, especially when tuning one algorithm for a specific problem.

The implications and new research directions implied by our results are manifold. From a theoretical perspective, our results show that \((\mu + 1)\) selection algorithms are not strictly elitist w.r.t. to a fixed reference point. This implies that theoretical results like those by Bringmann and Friedrich (2011) are not directly transferable to these algorithms. This holds in particular if the discussed techniques for boundary solution handling and reference point adaptation are considered. Can such theoretical results be adapted to the situation in practice or vice versa?

The latter question leads to the discussion of new techniques for boundary solution handling and reference point adaptation and therefore to new algorithm designs. Possible routes to explore are

- Consider dividing population into “interior” and “boundary” solutions and evolve them separately.
- Find a sane way to specify trade-offs between boundary and interior solutions and derive an optimal reference point from this.

Continuing the work at hand, the influence of the parametrization will be examined. Results are already available but have to visualized, described, and analyzed in an appropriate way. To this end, the influence on frequencies and amounts of decreases have to be set in relation to all mentioned parameters.

Moreover, the effect on higher dimensional objective spaces need to be studied as well as the effect on real world applications. Since hypervolume approximation schemes have recently become available that allow practitioners to tackle high and very-high dimensional problems and these inherently also include a source of hypervolume decreases, namely the approximation error, further studies of the influence of these phenomena on the overall optimization process are necessary. Last but not least, a possible influence on final results in comparison to new algorithm designs, by possibly not accepting decreases, need to be investigated.

Despite all the results presented here, hypervolume selection still remains the most effective selection scheme for multiobjective optimization problems.
References


