Integrating Decision Space Diversity into Hypervolume-based Multiobjective Search

Tamara Ulrich[†] Johannes Bader[†] Eckart Zitzler[†]

[†]Computer Engineering and Networks Laboratory
ETH Zurich
8092 Zurich, Switzerland
firstname.lastname@tik.ee.ethz.ch

ABSTRACT

Multiobjective optimization in general aims at learning about the problem at hand. Usually the focus lies on objective space properties such as the front shape and the distribution of optimal solutions. However, structural characteristics in the decision space can also provide valuable insights. In certain applications, it may even be more important to find a structurally diverse set of close-to-optimal solutions than to identify a set of optimal but structurally similar solutions. Accordingly, multiobjective optimizers are required that are capable of considering both the objective space quality of a Pareto-set approximation and its diversity in the decision space.

Although NSGA, one of the first multiobjective evolutionary algorithms, explicitly considered decision space diversity, only a few other studies address that issue. It therefore is an open research question how modern multiobjective evolutionary algorithms can be adapted to search for structurally diverse high-quality Pareto-set approximations. To this end we propose an approach to integrate decision space diversity into hypervolume-based multiobjective search. We present a modified hypervolume indicator and integrate it into an evolutionary algorithm. The proof-of-principle results show the potential of the approach and indicate further research directions for structure-oriented multiobjective search.

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search

General Terms

Algorithms

Keywords

Multiobjective Optimization, Diversity in Decision Space, Hypervolume Indicator

GECCO'10, July 7–11, 2010, Portland, Oregon, USA.

1. INTRODUCTION

When approximating the Pareto-optimal set of a multiobjective optimization problem, the goal is to find a set of compromise solutions the quality of which is maximum. The quality of Pareto-set approximations can be evaluated in terms of set quality measures such as the hypervolume indicator [13] or – more general – set preference relations as discussed in [14]. Most of the time, the quality is determined only on the basis of the corresponding objective vectors, i.e. quality is measured in the objective space. However, in many applications not only the objective function values matter, but also the structural properties of the generated solutions. For instance, an engineer may be particularly interested in a range of structurally different solutions distributed over the Pareto-optimal front. In that case, the diversity within a Pareto-set approximation with regard to the decision space needs to be taken into account within the optimization model and the search algorithm. This study addresses that issue.

Interestingly, the idea to integrate decision space diversity into the optimization has been proposed as early as 1994 in the first NSGA paper [9]. NSGA uses fitness sharing on the decision vectors in combination with non-dominated sorting. After that, most algorithms concentrated on the objective space only. In recent years, however, a few studies have picked up on this idea and have proposed alternative approaches. In 2003, GDEA [10] integrated diversity into the search as an additional objective. In 2008, the Omni-Optimizer [5] was developed which extends the original idea of NSGA, but in contrast to NSGA, its diversity measure takes both the decision and objective space diversity into account. Finally in 2009, two further studies were proposed. [8] extended a CMA-ES niching framework to include decision space diversity. [12] on the other hand applies clustering in objective space and then builds a model from the solutions in these clusters. This model is then used during variation in order to generate new offspring.

Diversity is a set measure and therefore a separate goal to the optimization. The other goal – let's call it the objective space measure – is also a set measure which indicates how well the final population approximates the Pareto-optimal front. With two set measures the question arises how these two measures can be combined. NSGA and the Omni-Optimizer use a ranking of the two, where the objective space measure is always considered first, and only if there are ties using this measure, diversity is taken into consideration. The drawback of this approach is that the diversity plays an inferior role and there is no possibility to change the tradeoff between the two measures. Moreover, these kind of approaches suffer from cyclic behavior [14]. A second approach is considering the diversity as an additional objective. The problem here is that the diversity, which is defined on sets, is treated the same way as the original solution-oriented objectives. A second problem is that all tradeoffs between diversity and original objectives are explored concurrently, without any means to adjust the tradeoff. As the number of incomparable solutions increases, this may lead to an ineffective search.

In this paper we propose a method which combines two set measures, namely the hypervolume and the diversity, into one single set measure, where the tradeoff between the two measures is adjustable in a flexible manner. The focus lies on hypervolume-based search, a methodology that has gained popularity in recent years. Its advantage is that cyclic behavior can be prevented, and therefore convergence to the Pareto-optimal front can be proven [14]. To our best knowledge, the idea of incorporating diversity into hypervolume-based search has not yet been investigated. In the following, we propose a first step in this direction and present an integrated approach for hypervolume-based multiobjective search that combines decision space diversity and hypervolume indicator values. First, we introduce the basic idea in Section 2. In Section 3, we show how a diversity measure can be defined and and we introduce a modification of the hypervolume indicator that integrates this diversity measure. We then discuss in Section 4 how this concept can be implemented within a diversity integrating hypervolume-based search algorithm (DIVA) in the sense of a proof-of-principle, and finally in Section 5, we use practical considerations to adjust our algorithm.

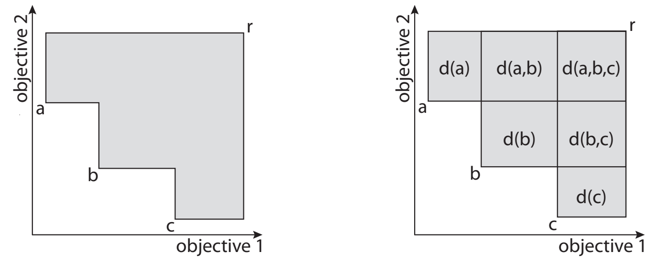

Figure 1: Original (left) and modified (right) hypervolume for a population of three solutions $A = \{a, b, c\}$ with reference set $R = \{r\}$. $d(a, b)$ for example is the diversity value of the subset $B = \{a, b\}$.

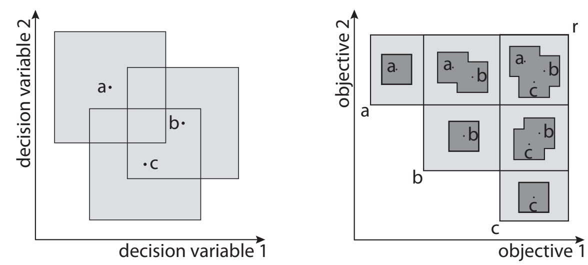

Figure 2: Symbolic representation of the modified hypervolume with one diversity value per hypervolume partition. The left figure shows the three solutions in the decision space. The right figure shows the hypervolume partitions, where in each partition the relevant diversity is shown qualitatively. For example in the partition dominated by solutions a and b, the used diversity value is the total area covered by the decision space boxes of solutions a and b.

2. BACKGROUND AND IDEA

In this study, we assume that two objective functions $f_i : X \to \mathbb{R}$, $i = \{1, 2\}$ are to be minimized. A solution x is mapped from the d-dimensional real-valued decision space $X \subseteq [0, 1]^d$ to its objective vector $f(x) = (f_1(x), f_2(x))$. Without loss of generality, we assume that all decision vectors lie in the interval $[0, 1]$. The underlying dominance relation is weak Pareto dominance, where a solution $a \in X$ weakly dominates another solution $b \in X$ if and only if the solution a is better or equal than b in all objectives, i.e. $a \preceq_{par} b$ iff $f(a) \leqslant f(b) := f_i(a) \leq f_i(b)$ for $i = \{1, 2\}$. We here consider the common optimization goal of finding a set A of μ solutions that maximizes the hypervolume of the set. The hypervolume with respect to a reference point set R is defined as the volume dominated by the solutions in A, but not dominated by the solutions in R.

We would now like to motivate the idea which is explained in detail in the next section. To this end we consider the example shown in the left part of Figure 1, with three solutions $A = \{a, b, c\}$ and one reference point $R = \{r\}$.

We assume that the hypervolume indicator is given, plus an additional diversity measure which returns the diversity of a subset $B \subseteq A$ of solutions. We would now like to integrate this diversity measure into the hypervolume indicator. But first, the question arises what diversity measure should be used. Intuitively, a measure similar to the hypervolume indicator, but defined in the decision space could be used, where the solutions cover some part of the decision space and the sum of the covered space indicates the diversity of the subset. In contrast to the objective space, the decision space has no clear direction of search. Therefore, we define a neighborhood around each solution, which indicates the part of the decision space covered by that solution. The diversity measure then is the union of these neighborhoods, see the left part of Figure 2 for an example.

There are different ways how such a neighborhood can be defined. The most intuitive approach would be a sphere with a given radius around the solution. A generalized version of the sphere would be a kernel function. In this study, we use a box around the solution, mainly for two reasons. First, it is much easier and faster to calculate if two boxes overlap than it is to calculate whether two spheres overlap. Second, the volume of a box with boxwidth 1 is always 1, whereas the volume of a sphere of radius 0.5 goes to zero for higher dimensions. Therefore, it is much easier to understand the influence and tune the boxwidth of a box than the radius of a sphere.

The next question which needs answering is how the hypervolume indicator and this diversity measure can be combined. The first idea which comes to mind is using a weighted sum. However, this approach comes with a serious drawback. Because only the non-dominated solutions have a contribution to the hypervolume, the dominated solutions are evaluated based on their contribution to diversity only. As solutions which are very diverse from non-dominating solutions usually also have very dissimilar objective values, this leads to populations where the non-dominated front optimizes the hypervolume and the dominated solutions optimize the diversity and are therefore randomly distributed instead of being close to the non-dominated solutions.

Therefore, our approach focuses on the hypervolume indicator. When looking at the hypervolume of a set of solutions, it can be seen that it is divided into partitions, where each partition is dominated by a specific subset of the whole population. In this study we propose to weight these partitions with the diversity of their dominating points before summing them up (see the right part of Figure 1 for an general example, and the right part of Figure 2 for an example with the proposed diversity measure). Note that in the original hypervolume indicator, the partitions are simply weighted with one.

This adaption has several nice properties. First, if a population is given, and the objective values of one solution improve, the modified hypervolume also improves. Second, if the diversity of a subset of the population improves (and the diversities of the remaining subsets remain the same), the modified hypervolume also improves. Third, if the diversity measure is chosen to be monotonically increasing with the number of solutions in the subset, adding a solution to the population cannot worsen the modified hypervolume. Fourth, it is more important that two solutions that are close in objective space are diverse than two solutions which are far apart in objective space. This is due to the fact that there are more subsets of the population that contain two close solutions than two far apart solutions.

3. DIVERSITY INTEGRATING HYPERVOLUME INDICATOR

In this section we provide a formal definition of the modified hypervolume indicator. First we discuss diversity measures and the properties they should have, then we show how such set measures in general can be integrated into the hypervolume indicator.

3.1 Diversity Measures

To calculate the decision space diversity of a given set of solutions, a so-called diversity function is used. Such a diversity function has to fulfill certain requirements such that the modified hypervolume indicator remains compliant with the underlying preference relation. First, the diversity of a set of solution must not decrease if a new solution is added to the set. Second, the diversity of a non-empty set of solution must be greater than zero, and the diversity of the empty set has to be zero. These properties are formally defined in Definition 1.

Definition 1 (Diversity Function). *Let Ψ denote the powerset of the decision space X. Then we call a function D from Ψ to $\mathbb{R}$* monotonic diversity function *if the following two properties hold:*

Monotonicity: *if $A, B \in \Psi$ are two sets of solutions for which $A \subseteq B$ holds, then $D(A) \le D(B)$.*

Positivity and null empty set: *for all $A \in \Psi \backslash \emptyset$ it holds $D(A) > 0$, while $D(\emptyset) = 0$.*

We here propose two different functions that are in accordance with Definition 1. The first function which is defined in Definition 2 has already been described in Section 2, where the diversity is equal to the volume of the union of neighborhoods around each solution.

Definition 2 (Coverage Diversity Function). *Given a set of solutions $A \subseteq X$ and a boxwidth b. The* coverage diversity function *$D_c(A)$ is then calculated as follows:*

$$D_c(A) = \frac{1}{b^d} \int_{\mathbb{R}^d} c_A^b(z) dz$$

where

$$c_A^b(z) = \begin{cases} 1 & \text{if } \exists x \in A : \forall 1 \le i \le d : |z_i - x_i| \le \frac{b}{2} \\ 0 & \text{else} \end{cases}$$

z_i and x_i is the i-th decision variable value of solution z and x, respectively.

The division by b^d makes the measure independent of the chosen neighborhood boxwidth b such that if the set consists of only one solution, this set has diversity one. This function fulfills the requirements of a monotonic diversity function. Adding solutions cannot decrease the diversity. A non-empty set has at least diversity 1, and the empty set has diversity zero.

The second diversity function is based on the distance between the solutions of a set A. More precisely, the diversity is the sum of distances of all solutions $x \in A$ to the median m_A of that set.

Definition 3 (Distance Diversity Function). *Given a set of solutions $A \subseteq X$ and a given distance measure $d(x, y)$, $x, y \in X$. Then the* distance diversity function *$D_d(A)$ is defined as:*

$$D_d(A) = \begin{cases} 0 & \text{if } A = \emptyset \\ 1 + \sum_{x \in A} d(x, m_A) & \text{else} \end{cases}$$

where $m(A) = (m_A^1, ..., m_A^d)$ is the median of the set A:

$$m_A^i = \frac{1}{|A|} \sum_{x \in A} x_i$$

Theorem 1 states the monotonicity of that function, i.e. that adding a solution to a given set cannot decrease the diversity value of that set. Positivity and null empty set are obvious.

Theorem 1. *$D_d(A \cup a) \ge D_d(A)$ holds for any set of solutions $A \subseteq X$ and any solution $a \in X$.*

Proof. The proof is by induction over the number of solutions $n = |A|$. $n = 1$ is obvious. For the transition of $n \to n+1$ We have to show that $1 + \sum_{x \in A \cup a} d(x, m_{A \cup a})$ $\ge 1 + \sum_{x \in A} d(x, m_A)$. Using the triangle inequality, we get $\sum_{x \in A} d(x, m_{A \cup a}) + d(a, m_{A \cup a}) \ge \sum_{x \in A} d(x, m_{A \cup a}) +$ $\sum_{x \in A} d(m_{A \cup a}, m_A) \ge \sum_{x \in A} d(x, m_A)$. We therefore have to show that $d(a, m_{A \cup a}) \ge \sum_{x \in A} d(m_{A \cup a}, m_A)$. Writing out the median sum we get $d(a, \frac{1}{n+1} \sum_{x \in A} x + \frac{1}{n+1} a) \ge n \cdot$ $d(\frac{1}{n+1} \sum_{x \in A} x + \frac{1}{n+1} a, \frac{1}{n} \sum_{x \in A} x)$
$= n \cdot d(\frac{1}{n+1} a, \frac{1}{n(n+1)} \sum_{x \in A} x)$. Multiplying by $(n+1)$ we get $d(na + a, \sum_{x \in A} x + a) = d(na, \sum_{x \in A} x)$ which completes the proof. □

3.2 Modified Hypervolume

We now explain how any such diversity function – or any set-based function which fulfills the above properties – can be integrated into the hypervolume indicator. As motivated in Section 2, we look at the hypervolume as a set of partitions which are dominated by a subset of the population. We call the solutions in A that dominate a certain point z the dominating points of z:

Definition 4 (Dominating Points). *Given a point $z \in \mathbb{R}^d$, and $A \subseteq X$ a set of solutions. We call the set of solutions $\mathrm{dom}_A(z) := \{x \,|\, x \in A \wedge f(x) \leqslant z\}$ the subset of A dominating the objective vector z.*

We can say that if one set $A \subseteq X$ has a better or equal diversity in all hypervolume partitions than another set $B \subseteq X$, the set A is weakly preferred to set B:

Definition 5 (Diversity Preference Relation). *Let $A, B \subseteq X$ be two sets of solutions and D a diversity function. A is weakly diversity preferred to B, denoted $A \preccurlyeq_D B$ iff $\forall z \in \mathbb{R}^d : \mathrm{D}(\mathrm{dom}_A(z)) \geq \mathrm{D}(\mathrm{dom}_B(z))$*

This preference relation has the property that if A is weakly diversity preferred to B, it is also weakly preferred to B according to Pareto dominance:

Theorem 2. *Given two sets $A, B \subseteq X$ and a diversity measure D, then $A \preccurlyeq_D B \Rightarrow A \preccurlyeq_{par} B$ holds, where $\preccurlyeq_{par}$ is the extension of Pareto dominance to sets, i.e. $A \preccurlyeq_{par} B$ holds iff $\forall y \in B : \exists x \in A : f(x) \leqslant f(y)$.*

Proof. As $B \subseteq \mathbb{R}^d$ it holds that $A \preccurlyeq_D B \Rightarrow \forall y \in B : \mathrm{D}(\mathrm{dom}_A(b)) \geq \mathrm{D}(\mathrm{dom}_B(b))$. We want to proof that this means that also $A \preccurlyeq_{par} B$ as defined in this theorem holds. Assume the contrary, i.e. $\exists y \in B : \nexists x \in A : f(x) \leqslant f(y)$. In this case, $\mathrm{dom}_A(b) = \emptyset \Rightarrow \mathrm{D}(\mathrm{dom}_A(b)) = 0$. At the same time $\mathrm{D}(b) > 0$ and with monotonicity $\mathrm{D}(\mathrm{dom}_B(b)) > 0$. Therefore $\mathrm{D}(\mathrm{dom}_B(b)) > \mathrm{D}(\mathrm{dom}_A(b))$ which is a contradiction. $\square$

To modify the hypervolume indicator, we weight each hypervolume partition with the diversity of its dominating points. Because the diversity is a monotonically increasing function with the number of considered points, partitions which are dominated by only a few solutions are weighted with lower values than partitions which are dominated by a large number of solutions. In order to attenuate this effect, a so called desirability function is used. To that end, the hypervolume partitions are weighted with this desirability value of the partition, which in turn is derived from the diversity of the solutions dominating the partition:

Definition 6 (Desirability of Diversity). *Let $A \in \Psi \backslash \emptyset$ be a set of solutions with corresponding diversity $D(A)$. Then the desirability function $\xi : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{\geq 0}$, $D(A) \mapsto \mathbb{R}_{\geq 0}$ assigns a non-negative value to the diversity representing its desirability to the decision maker. The function ξ is assumed to be monotonically increasing and non-negative, i.e., $\forall x, y \in \mathbb{R}_{>0} : x \leq y \Rightarrow \xi(x) \leq \xi(y)$, and $\xi(x) \geq 0$.*

Many functions fulfill the required properties. We here propose the following class of functions which is tunable by a shape parameter s:

Definition 7. (Logarithmically Increasing Desirability) *Let $s \in \mathbb{R}_{\geq 0}$ be a shape parameter, then we define the logarithmically increasing desirability function as*

$$\xi_{\log}(D(A)) := \frac{\log(1 + s \cdot D(A))}{\log(1 + s)} \qquad (1)$$

Note that $\lim_{s \to 0} \xi_{\log}(D(A)) = D(A)$
and $\lim_{s \to \infty} \xi_{\log}(D(A)) = 1$

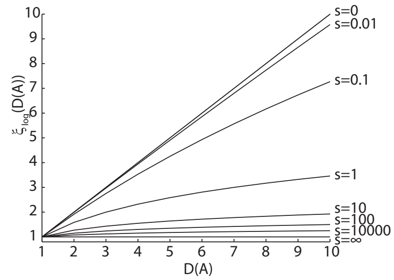

Figure 3: Shape of the desirability function $\xi_{\log}(\mathrm{D}(A))$ for $\mathrm{D}(A) \in [1, 10]$ and for different shape parameter values s.

The shape of this desirability function using different shape parameters is shown in Figure 3.

Now we are able to formally define the diversity integrating hypervolume. The objective space is divided into hypervolume partitions. Each partition is dominated by a specific subset of the population. The partitions weight is equal to the desirability value derived from the diversity of that subset of solutions. To achieve the total diversity integrating hypervolume, the partitions size multiplied with its weight is summed up.

Definition 8 (Diversity Integrating Hypervolume). *Let $A \in \Psi$ denote a set of solutions, let $w : \mathbb{R}^k \rightarrow \mathbb{R}_{>0}$ be a strictly positive and integrable weight function. Furthermore, let $D : \Psi \rightarrow \mathbb{R}_{\geq 0}$ be a monotonic diversity measure according to Def. 1, let ξ denote a desirability function of the diversity according to Def. 6, and let $\mathrm{dom}_A(z)$ according to Def. 4 give the subset of A dominating the objective vector z. Then the (weighted) diversity integrating hypervolume indicator $I_H^{w,\mathrm{D}}(A, R)$ corresponds to a weighted Lebesgue measure of the set of objective vectors weakly dominated by the solutions in A but not by a so-called reference set $R \in \mathbb{Z}$:*

$$I_H^{w,D}(A, R) = \int_{\mathbb{R}^d} \alpha_A(z) w(z) \xi(D(\mathrm{dom}_A(z))) dz \qquad (2)$$

with $\alpha_A(z) = \mathbf{1}_{H(A,R)}(z)$ where

$$H(A, R) = \{z \mid \exists a \in A \, \exists r \in R \; : \; f(a) \leqslant z \leqslant r\} \qquad (3)$$

and $\mathbf{1}_{H(A,R)}(z)$ being the characteristic function of $H(A, R)$ that equals 1 iff $z \in H(A, R)$ and 0 otherwise. $w(z)$ is a weight function which indicates how important it is to dominate z. In our case, $w(z) \equiv 1$.

This indicator is a weak refinement of the diversity preference relation defined in Definition 5:

Theorem 3. *If a set $A \subseteq X$ is weakly diversity preferred to another set $B \subseteq X$, the modified hypervolume of set A is larger or equal the one of B, i.e. $A \preccurlyeq_D B \Rightarrow I_H^{w,\mathrm{D}}(A, R) \geq I_H^{w,\mathrm{D}}(B, R)$.*

Proof. We know that $A \preccurlyeq_D B \Rightarrow A \preccurlyeq_{par} B$ (Theorem 2), therefore, $\{z : \alpha_B(z) \neq 0\} \subseteq \{z : \alpha_A(z) \neq 0\}$. From $\mathrm{D}(\mathrm{dom}_A(z)) \geq \mathrm{D}(\mathrm{dom}_B(z))$ it follows with a monotonic desirability function that $\xi(\mathrm{D}(\mathrm{dom}_A(z))) \geq \xi(\mathrm{D}(\mathrm{dom}_B(z)))$. Therefore, $I_H^{w,\mathrm{D}}(A, R) \geq I_H^{w,\mathrm{D}}(B, R)$. $\square$

Algorithm 1 Calculation of the modified hypervolume indicator. Takes a population $A \subseteq X$ and a boxwidth b and returns the indicator value. d is the number of decision variables.

```
function modHyp(A, b)
    h = 0 /* the indicator value */
    /* For all non empty hypervolume partitions */
    for all B ⊆ A\∅, ∃z ∈ R^d : dom_B(z) ≠ ∅ do
        /* Calculate the partition's volume */
        v ← ∏_{1≤i≤d} [max_{z:dom_B(z)≠∅} f_i(z) − min_{z:dom_B(z)≠∅} f_i(z)]
        /* Calculate D_c(B) */
        d ← 0 /* D_c(B), diversity of B */
        for x ∈ B do /* For all points */
            for 1, ..., m, m: Number of samples do
                /* Sample in the box around x */
                Sample s = (s^1, ..., s^d), s^i ∈ [x^i − b/2, x^i + b/2]
                /* Calculate the solutions of B which are within the box of s */
                F ← {y ∈ B : y^i ∈ [s^i − b/2, s^i + b/2]}
                d ← d + 1/(m·|F|) /* Increment diversity */
            end for
        end for
        h ← h + v · ξ(d) /* Increment indicator */
    end for
    Return h
end function
```

Algorithm 2 Environmental Selection. Takes a Population A, $|A| \geq \alpha$, and the number α of selected individuals. R is the reference set.

```
function envSel(A, α)
    while |A| > α do
        /* Simulate removing each solution from the population. Remove the solution that induces the smallest indicator loss. */
        A ← A \ max_{I_H^{w,D_c}(A\x,R), x∈A} x
    end while
end function
```

4. PROOF OF PRINCIPLE

In this section we evaluate if our approach makes sense. We use the diversity measure D_c since it fits our intuitive notion of coverage in the decision space. The question is how the modified hypervolume indicator can be used in the evolutionary algorithm. Its calculation is complex, as the diversity of each partition has to be calculated separately. We elaborate on how we tackle that problem in the following.

4.1 Search Algorithm Design

First, we need to decide how the modified hypervolume indicator can be calculated. As we optimize a problem with two objective functions, the intuitive idea would be to use the hypervolume by slicing objectives algorithm [11] to calculate the hypervolume partitions, and then calculate the diversity of each partition.

The biggest problem with that approach is that calculating the diversity according to the D_c measure is #P-hard [2]. The fastest known algorithm [7] is of complexity $\mathcal{O}(n^{d/2} \log n)$, which makes it unusable for general optimization problems, as e.g. our testproblems have $d = 24$. We therefore suggest to use sampling in order to estimate the diversity of a set. With an increasing number of decision variables, the volume of the neighborhood of a solution (if the boxwidth is smaller than 1) converges to zero. Therefore sampling the whole decision space would require an infinite number of samples, as the probability of a sample falling into the neighborhood of a solution also converges to zero. We therefore suggest to only sample in the neighborhoods around each solution in the set, and to increase the diversity of the set depending on the number of solutions in the set that have the sample within their neighborhood (intersection size). Once the diversities of all hypervolume partitions are calculated, the desirability value of these diversities can be calculated, and then the hypervolume partition volumes are weighted with the corresponding desirability value and summed up to yield the diversity integrating hypervolume. This procedure is described in Algorithm 1.

Next, we need a fitness assignment strategy. We here propose to set the fitness of a solution equal to the loss in the modified hypervolume if that solution is removed from the population. Finally, for the environmental selection strategy we propose to use a popular greedy environmental selection scheme as described in Algorithm 2. In this greedy strategy, the solution with lowest fitness is removed until the population is of size α. As soon as one solution is removed, the fitnesses of the remaining solutions are reevaluated.

The whole environmental selection algorithm including the indicator calculation is of complexity $\mathcal{O}(n \cdot (n^2 + n \cdot m \cdot (n \cdot d + n^2 \cdot n \cdot n) + n \cdot n^2))$, where $n = |A|$ is the number of solutions, m is the number of samples and d is the decision space dimension. Assuming that $n^3 \geq d$ and $m \geq 1$ this simplifies to $\mathcal{O}(n \cdot n \cdot n^2 \cdot n \cdot n) = \mathcal{O}(n^6)$. The first n comes from the number of greedy steps in the environmental selection. The second n comes from the fact that the sampling is done for each point. The n^2 is the number of hypervolume partitions[1]. The fourth n stems from the fact that the diversity has to be calculated for each solution, simulating the diversity change if that solution is removed. The last n comes from the calculation of the intersection size. Due to its high combinatorial complexity, this algorithm can only be applied to very small population sizes, e.g. $|A| \leq 10$.

Now that we have designed an environmental selection strategy, we can integrate it in our diversity integrating hypervolume-based search algorithm (DIVA). DIVA additionally uses random mating selection which is based on pairwise tournament selection. As variation operators the simulated binary crossover (SBX) and the polynomial mutation operator [4] are used.

4.2 Results

DIVA was run on the WFG testsuite [6], with 4 position and 20 distance related parameters (in total 24 decision variables). The population size as well as the number of offspring was set to 10 and the algorithm was run for 300 generations. The number of samples per neighborhood box was set to 500. Different shape parameter values s for the desirability function were tested and it has been found that $s = 100$ is a value that yields good results. To test the influence of the neighborhood boxwidth, widths of 0.4 and 1.8 were used. For each test problem (WFG1 - WFG9) 11 runs were done. As a reference algorithm, we used HypE [1]. HypE is among the newest hypervolume optimizing algorithms and has been shown to be highly competitive with respect to other state-of-the-art multiobjective optimizers. HypE was used without sampling, and its mating selection was set to random tournament selection, like in DIVA.

The results are shown in Table 1 for the boxwidth of 0.4

[1]Note that this only holds for two dimensional objective spaces.

	Hypervolume		Diversity	
	DIVA	HypE	DIVA	HypE
WFG1	88.4 ± 1.36	89.0 ± 1.37	$10.0 \pm 0.03^*$	8.7 ± 0.67
WFG2	99.0 ± 3.35	99.7 ± 4.01	$10.0 \pm 0.00^*$	7.9 ± 0.86
WFG4	107.6 ± 0.69	$109.3 \pm 0.70^*$	$10.0 \pm 0.00^*$	8.8 ± 0.59
WFG5	107.9 ± 0.36	$110.8 \pm 0.70^*$	$10.0 \pm 0.00^*$	7.4 ± 0.76
WFG6	104.4 ± 1.65	$107.6 \pm 1.67^*$	$10.0 \pm 0.01^*$	8.7 ± 0.46
WFG7	107.8 ± 0.46	$110.0 \pm 0.56^*$	$10.0 \pm 0.02^*$	9.7 ± 0.20
WFG8	105.9 ± 1.03	$108.1 \pm 0.62^*$	$10.0 \pm 0.00^*$	9.9 ± 0.09
WFG9	107.2 ± 1.88	107.9 ± 2.70	$10.0 \pm 0.04^*$	7.9 ± 1.05

Table 1: Mean values and standard deviation of the hypervolume and diversity for DIVA and HypE. Significantly better results are highlighted with an asterisk. The used neighborhood boxwidth is 0.4.

	Hypervolume		Diversity	
	DIVA	HypE	DIVA	HypE
WFG1	87.7 ± 1.10	89.0 ± 1.37	$9.9 \pm 0.02^*$	4.8 ± 0.80
WFG2	100.9 ± 3.45	99.7 ± 4.01	$9.7 \pm 0.06^*$	4.6 ± 0.58
WFG4	107.4 ± 0.56	$109.3 \pm 0.70^*$	$9.8 \pm 0.04^*$	5.6 ± 0.34
WFG5	107.9 ± 0.56	$110.8 \pm 0.70^*$	$9.9 \pm 0.02^*$	3.5 ± 0.69
WFG6	104.9 ± 1.42	$107.6 \pm 1.67^*$	$9.9 \pm 0.03^*$	4.9 ± 0.60
WFG7	107.4 ± 0.23	$110.0 \pm 0.56^*$	$9.6 \pm 0.07^*$	6.1 ± 0.29
WFG8	106.1 ± 0.50	$108.1 \pm 0.62^*$	$9.7 \pm 0.05^*$	6.9 ± 0.49
WFG9	105.2 ± 1.60	$107.9 \pm 2.70^*$	$9.8 \pm 0.04^*$	3.8 ± 0.68

Table 2: Similar to Table 1, but with a neighborhood boxwidth of 1.8.

and Table 2 for the boxwidth of 1.8.[2] In order to test the results for significant differences between DIVA and HypE, a Kruskal-Wallis test as described in [3] is applied, using the Conover-Inman procedure, Fisher's least significant difference method performed on ranks and a significance level of 5%. It can be seen that DIVA produces populations with a significantly better diversity than HypE for all eight testproblems and for both boxwidth values. It also seems that the hypervolume of WFG4, WFG5, WFG6, WFG7 and WFG8 is better optimized by HypE than by DIVA. For WFG1 and WFG2, there is no significant difference whereas WFG9 is only better optimized by HypE if the larger boxwidth is used.

The results indicate that there is a tradeoff between hypervolume and diversity. An explanation of this behavior is that the Pareto-optimal front of the WFG testsuite has very similar solutions. In our case where there are 24 decision variables in total, the 20 distance related parameters have a fixed value for Pareto-optimal solutions. This behavior can be seen in Figure 4, where for example for WFG7 and WFG8 the HypE populations of the distance related parameters (decision variables 5 to 24) are more similar than those of the position related parameters (decision variables 1 to 4).

From Figure 4 it also becomes clear that the degree to which diversity is optimized depends on the chosen neighborhood size. As long as the box width is smaller than 1, two solutions whose distance in one decision space variable is larger or equal to this boxwidth do not overlap. With a boxwidth that is larger than 1, the neighborhoods of two solutions always overlap (hence the diversity value never reaches the optimal value of 10), and to minimize this overlap, the solutions distances in all decision variables must be as large as possible. Therefore if an optimization problem is used where the Pareto-optimal solutions are very similar, a boxwidth of 1.8 yields more diverse solutions than a boxwidth of 0.4. Interestingly enough, the hypervolume value is not impeded by using a larger boxwidth.[3]

[2] WFG3 is not shown because the algorithms failed to achieve an acceptable approximation of the Pareto-front.

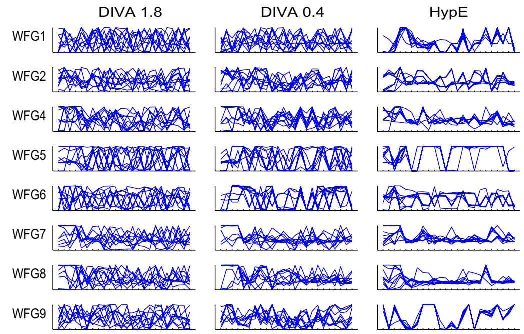

Figure 4: Parallel coordinates plot of the decision variable values of one population (ten individuals) for each WFG problem (top to bottom rows are WFG1 to WFG9). Results are shown for DIVA with a neighborhood boxwidth of 1.8 and 0.4 (left and middle column, respectively), as well as for HypE (right column).

When interpreting these results it has to be noted that the explanatory power of runs with population sizes of 10 is rather low. The goal of this section was to investigate the tradeoff between hypervolume and diversity, as well as to get an idea about the influence of the boxwidth. Also, the mean runtime for a population size of 10 is roughly 11 seconds per generation. As stated in Section 4, DIVA has a complexity of $\mathcal{O}(n^6)$, therefore making it unusable for larger population sizes. In order to alleviate this problem, we propose two adjustments of the original algorithm in the next section.

5. PRACTICAL CONSIDERATIONS

In the following we propose a modified environmental selection scheme and evaluate it on three testproblems.

5.1 Search Algorithm Design

There are two ways how the proposed environmental selection scheme can be modified for speed. The first way is to facilitate the calculation of the diversity and the second way is to reduce the number of considered hypervolume partitions.

Replacing Diversity Measure Out of the $\mathcal{O}(n^6)$ complexity, $\mathcal{O}(n^2)$ is due to the used diversity measure D_c, as n points have to be sampled with m samples and for each sample, the intersection size has to be calculated. The distance diversity measure D_d, on the other hand, only needs to go through the solutions twice (once for calculating the median and once for calculating the solutions distance to the median) and therefore has complexity $\mathcal{O}(n)$, so we propose to use that measure for larger population sizes.

Bounding Recursion Depth The second adjustment is due to the fact that the number of hypervolume partitions is $\mathcal{O}(n^2)$. To reduce that number, we propose to

[3] Except for WFG7 and WFG9, where the hypervolume achieved with a boxwidth of 0.4 is significantly better than the hypervolume achieved with a boxwidth of 1.8.

Algorithm 3 Takes a population A and a cutoff value c and returns the modified hypervolume indicator value.

function MODHYPSPEEDUP(A, c)
  $h = 0$ /* *the indicator value* */
  /* *For all non empty hypervolume partitions dominated by less or equal c solutions* */
  for all $B \subseteq A \backslash \emptyset$, $\exists z \in \mathbb{R}^d : |\operatorname{dom}_B(z)| \in [1, c]$ **do**
    /* *Calculate the partitions volume* */
    $v \leftarrow \prod_{1 \leq i \leq d} [\max_{z:\operatorname{dom}_B(z) \neq \emptyset} f_i(z) - \min_{z:\operatorname{dom}_B(z) \neq \emptyset} f_i(z)]$
    $h \leftarrow h + v \cdot \xi(D_d(B))$ /* *Increment indicator* */
  end for
  Return h
end function

not use all partitions, but only those which are dominated by less than c solutions. c in this case serves as a cutoff value. The smaller c is, the fewer partitions have to be considered. This new partition number is of complexity $\mathcal{O}(c \cdot n)$. The new diversity integrating hypervolume with cutoff c is

$$I_H^{w,D}(A, R, c) = \int_{\mathbb{R}^d} \alpha_A(z) w(z) \xi(D^c(\operatorname{dom}_A(z))) dz$$

where

$$D^c(\operatorname{dom}_A(z)) = \begin{cases} D(\operatorname{dom}_A(z)) & \text{if } |\operatorname{dom}_A(z)| \leq c \\ D(A) & \text{else} \end{cases}$$

Note that $D(A)$ is a constant and therefore does not influence the ranking of the fitnesses of the solutions.

If these two modifications are used, the overall complexity of the algorithm reduces to $\mathcal{O}(c \cdot n^4)$, which makes the algorithm applicable to reasonable population sizes.

5.2 Search Algorithm Design

In order to integrate these two modifications into the search, Algorithm 1 needs to be adapted. The modified algorithm including the new diversity measure is described in Algorithm 3.

5.3 Results

In the following tests, the distance diversity measure D_d was used. The population size was chosen to be 100, with 100 offspring in each generation. The desirability function shape parameter was again set to $s = 100$ and on each selected WFG testproblem there were 11 runs with 300 generations each.

In addition to HypE, we used the Omni-Optimizer [5] as a second reference algorithm. It was implemented according to [5], with $\varepsilon = 0.001$. The only adaption with respect to [5] is that normal SBX (simulated binary crossover) on 50% of the individuals and the polynomial mutation operator [4] were used instead of the slightly adapted versions proposed in the Omni-Optimizer paper. That way, DIVA, HypE and the Omni-Optimizer all used the same variation operators, which makes them better comparable. However, note that while DIVA and HypE used random mating selection such that differing results are solely due to the different environmental selection strategies, the Omni-Optimizer implementation used the fitness-based mating selection proposed in its paper.

The results are shown graphically in Figures 5 and 6. For all three testproblems, the hypervolume as well as the diversity has a tendency to increase with an increasing cutoff value. As in Section 4.2, Kruskal-Wallis was used to

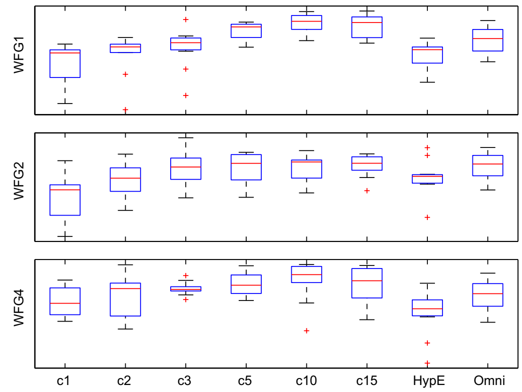

Figure 5: Hypervolume values (higher = better) for three different testproblems (WFG1, WFG2 and WFG4), for DIVA with different cutoff values ($c = \{1, 2, 3, 5, 10, 15\}$) as well as for HypE and the Omni-Optimizer (Omni).

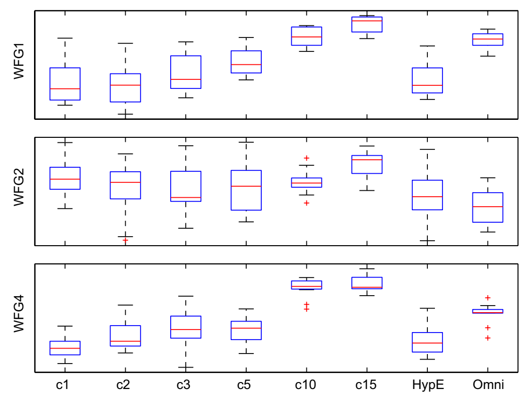

Figure 6: Distance diversity (D_d) values (higher = better) for the same runs as shown in Figure 5.

test whether DIVA's results are significantly better than HypE's. The results are shown in Table 3. Interestingly enough, DIVA seems to work better on the selected test problems when using the new measure. With a cutoff value of 15, DIVA's diversity is significantly better than the Omni-Optimizer's as well as HypE's. With a cutoff value of 3 of higher, DIVA's hypervolume is not significantly different and sometimes even significantly better than the Omni-Optimizer's and HypE's. This indicates that the optimization of the diversity also helps to explore the search space and therefore to achieve better objective values. One reason for this could be that more unusual solutions are produced which have quite a different decision space representation from the solutions which are usually found during optimization.

The decision space values of one population of DIVA and HypE and all three selected testproblems are shown in Figure 7. It can be seen that DIVA's distance diversity measure allows for much more subtle diversity changes which are difficult to make out at a glance, but still yield a mathematically significantly better diversity than HypE.

The calculation time in relation to the chosen cutoff value is shown in Figure 8. It can be seen that the calculation time

	Hypervolume					
	c1	c2	c3	c5	c10	c15
WFG1	0	0	+	+	+	+
WFG2	-	0	0	0	0	0
WFG4	0	+	+	+	+	+

	Diversity					
	c1	c2	c3	c5	c10	c15
WFG1	0	0	0	+	+	+
WFG2	+	0	0	0	0	+
WFG4	0	0	0	0	+	+

Table 3: Significance of hypervolume (top) and diversity D_d (bottom) values of DIVA with different cutoff values with reference to HypE. For example, $c1$ means that the cutoff value was set to one. A $+/-$ indicates that DIVA's result for the given cutoff value and the given testproblem is significantly better/worse than the HypE values. A 0 indicates that there is no significant difference to HypE.

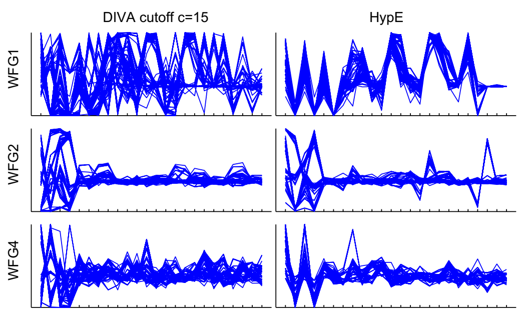

Figure 7: Parallel coordinates plot of the decision variable values of one population (100 individuals) for WFG problems 1, 2, and 4 (rows). Results are shown for DIVA with a cutoff value of $c = 15$ (left column), as well as HypE (right column).

achieved with the distance diversity measure D_d, a cutoff value of 10 and a population size of 100 is comparable to the calculation time when using the coverage diversity measure D_c, no cutoff and a population size of 10.

6. CONCLUSIONS AND OUTLOOK

This work has laid a foundation for diversity integrating hypervolume based search. We have introduced a method to integrate decision space diversity into the hypervolume indicator, such that these two set measures can be optimized simultaneously. We have proposed a proof of principle method which only works on small populations, and some modification of this method such that it is also applicable to reasonable population sizes. Using the latter method we could show that the diversity can be significantly increased without impeding the hypervolume. Interestingly enough, there are several problem settings where the hypervolume even increases if diversity is also optimized. Whether this finding can be generalized to other problems has to be the focus of further studies.

In the future it might be interesting to have an adaptive setting of the neighborhood size for the coverage based diversity measure, or to use different neighborhood shapes,

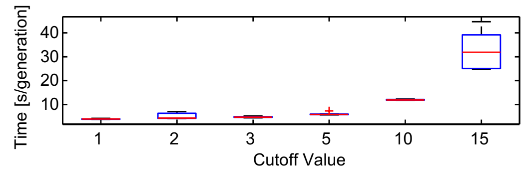

Figure 8: The calculation time for different cutoff values in seconds per generation. The population size and the number of offspring is 100. The used testproblems were WFG1, WFG2 and WFG4, with 11 runs for each problem.

and maybe even kernel functions. Furthermore, the algorithms should be adapted such that they are applicable to higher-dimensional objective spaces. Finally, the proposed diversity measures should be extended such that they can be used with other than real-valued decision spaces, such as integer or nominal spaces.

7. REFERENCES

[1] J. Bader and E. Zitzler. HypE: An Algorithm for Fast Hypervolume-Based Many-Objective Optimization. *Evolutionary Computation*, 2009. to appear.

[2] K. Bringmann and T. Friedrich. Approximating the Volume of Unions and Intersections of High-Dimensional Geometric Objects. In *International Symposium on Algorithms and Computation (ISAAC 2008)*, volume 5369 of *LNCS*, pages 436–447, 2008.

[3] W. J. Conover. *Practical Nonparametric Statistics*. John Wiley, 3rd edition, 1999.

[4] K. Deb. *Multi-Objective Optimization Using Evolutionary Algorithms*. Wiley, 2001.

[5] K. Deb and S. Tiwari. Omni-optimizer: A generic evolutionary algorithm for single and multi-objective optimization. *European Journal of Operational Research*, 185(3):1062–1087, 2008.

[6] S. Huband, P. Hingston, L. Barone, and L. While. A Review of Multiobjective Test Problems and a Scalable Test Problem Toolkit. *IEEE Transactions on Evolutionary Computation*, 10(5):477–506, 2006.

[7] M. H. Overmars and C. Yap. New upper bounds in klee's measure problem. *SIAM J. Comput.*, 20(6):1034–1045, 1991.

[8] O. M. Shir, M. Preuss, B. Naujoks, and M. Emmerich. Enhancing decision space diversity in evolutionary multiobjective algorithms. In *Conference on Evolutionary Multi-Criterion Optimization (EMO 2009)*, pages 95–109, 2009.

[9] N. Srinivas and K. Deb. Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms. *Evolutionary Computation*, 2(3):221–248, 1994.

[10] A. Toffolo and E. Benini. Genetic diversity as an objective in multi-objective evolutionary algorithms. *Evolutionary Computation*, 11(2):151–167, 2003.

[11] L. While, P. Hingston, L. Barone, and S. Huband. A Faster Algorithm for Calculating Hypervolume. *IEEE Transactions on Evolutionary Computation*, 10(1):29–38, 2006.

[12] A. Zhou, Q. Zhang, and Y. Jin. Approximating the set of pareto optimal solutions in both the decision and objective spaces by an estimation of distribution algorithm. *IEEE Transactions on Evolutionary Computation*, 2009. Accepted.

[13] E. Zitzler and L. Thiele. Multiobjective Optimization Using Evolutionary Algorithms - A Comparative Case Study. In *Conference on Parallel Problem Solving from Nature (PPSN V)*, pages 292–301, Amsterdam, 1998.

[14] E. Zitzler, L. Thiele, and J. Bader. On Set-Based Multiobjective Optimization (Revised Version). TIK Report 300, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, 2008.