Available online at www.sciencedirect.com

ScienceDirect

Procedia Technology 16 (2014) 532 – 541

Procedia Technology

CENTERIS 2014 - Conference on ENTERprise Information Systems / ProjMAN 2014 - International Conference on Project MANagement / HCIST 2014 - International Conference on Health and Social Care Information Systems and Technologies

On reference point free weighted hypervolume indicators based on desirability functions and their probabilistic interpretation

Michael T. M. Emmerich[a], André H. Deutz[a], and Iryna Yevseyeva[b]

[a]*LIACS, Leiden University, Niels Bohrweg 1, 2333-CA Leiden, The Netherlands*
[b]*Centre for Cybercrime and Computer Security, School of Computing Science, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK*

Abstract

The so-called a posteriori approach to optimization with multiple conflicting objective functions is to compute or approximate a Pareto front of solutions. In case of continuous objective functions a finite approximation to this set can be computed. Indicator-based multiobjective optimization algorithms compute solution sets that are optimal with respect to some quality measure on sets, such as the commonly used hypervolume indicator (HI). The HI measures the size of the space that is dominated by a given set of solutions. It has many favorable monotonicity properties but it requires a reference point the choice of which is often done ad-hoc. In this study the concept of set monotonic functions for dominated subsets is introduced. Moreover, this work presents a reference point free hypervolume indicator that uses a density that is derived from the user's preferences expressed as desirability functions. This approach will bias the distribution of the approximation set towards a set that more densely samples highly desirable solutions of the objective space. We show that the Harrington type and the Derringer-Suich type of desirability functions yield definite integrals and that the Harrington type has also the favorable property to provide a set-monotonic function over the set of dominated subspaces. It is shown that for a product type of aggregation the weighted hypervolume indicator is mathematically equivalent with an approach that computes the standard hypervolume indicator after transformation of the axes. In addition a probabilistic interpretation of desirability functions is discussed and how a correlation parameter can be introduced in order to change the aggregation type. Finally, practical guidelines for using the discussed set indicator in multiobjective search, for instance when searching for interesting subsets from a database, are provided.

Peer-review under responsibility of the Organizing Committee of CENTERIS 2014.

Keywords: Multicriteria Decision Analysis, Multiobjective Optimization, Weighted Hypervolume Indicator, Desirability Functions, Set Performance Indicators.

Peer-review under responsibility of the Organizing Committee of CENTERIS 2014.
doi:10.1016/j.protcy.2014.10.001

1. Introduction

This work is intended to extend the discussion of weighted hypervolume indicators that measure the performance of sets of alternative solutions in multiobjective optimization. We will discuss reference point free versions of the weighted hypervolume indicators that use probability distribution functions to model the decision maker's preferences. It offers a way to formulate user preferences for situations where a set of a small number of solutions is computed with the aim to find at least one solution that satisfies the decision maker. The use of desirability information steers the search to more relevant parts of the Pareto front. Moreover, we will motivate the concept of the *desirability of a set* based on the desirability of its elements. Desirability function here is understood as a special form of a utility function related to preferences (goals) of a decision maker (DM) rather than utility related to usefulness [9].

The new approach is similar to a recently proposed method by Wagner and Trautmann [1] that uses desirability functions in the hypervolume-based multiobjective optimization algorithm DF-SMS-EMOA. In DF-SMS-EMOA the coordinate axes are transformed using desirability functions and the standard hypervolume indicator is computed based on the transformed coordinates. Our approach will be different and based on the weighted hypervolume indicator using (probability) densities derived from the desirability functions. For the special case of using a product type of aggregation function, it will be shown that the new approach and the approach in [1] are equivalent. This work will therefore provide a view that reconciles the weighted hypervolume approach and desirability function approach in [1], by viewing the second as a special case of the first.

2. Set monotonic functions of dominated subspaces

In this work we consider the *multiobjective optimization problem* (MOP) where the objective functions $f_1: S \to \mathbb{R}, \ldots, f_m: S \to \mathbb{R}$ are to be maximized for some search space S. It is said that $x \in S$ dominates y ∈ S, if and only if x is better or equal in all objective function values and strictly better in at least one objective function value. The efficient set of an MOP is given by the set of non-dominated solutions in S, i.e. $X_E = \{x \in S |$ not exists $u \in S$ with u dominates $x\}$. The Pareto front is its image under $f = (f_1, \ldots, f_m)$ and is a subset of $\mathbb{R}^m$ of dimension at most m. In multiobjective optimization it is often of interest to find the Pareto front or an approximation of it [3]. Moreover, Pareto dominance in the objective space $\mathbb{R}^m$ is defined as: $y^{(1)} \in \mathbb{R}^m$ dominates $y^{(2)} \in \mathbb{R}^m$ if $y^{(1)} \neq y^{(2)}$ and $y^{(1)}$ is better or equal in all coordinates than $y^{(2)}$.

Next we will introduce the notion of *dominated subspaces* on which the notion of *free hypervolume indicator* is based on. For each goal vector $y \in \mathbb{R}^m$ let us define $DomSet(\{y\}) := \{u \in \mathbb{R}^m \mid y \; dominates \; u\}$. For a finite multi-set (or population) of points $P = \{ y^{(1)}, \ldots, y^{(n)}\}$ of vectors in $\mathbb{R}^m$, define $DomSet(P) = DomSet(\{ y^{(1)}\}) \cup \ldots \cup DomSet(\{ y^{(n)}\})$. Furthermore, with $DomSets_m$ we denote the set of all subspaces A of $\mathbb{R}^m$ such that $A = DomSet(P)$ for some finite population $P \subset \mathbb{R}^m$.

By a set indicator we mean here a real valued function that measures how well a set of solution performs, for instance, as an approximation to the Pareto front. We are interested in set indicators that do not require the true Pareto front as a reference and thus can be used to guide the search towards the unknown Pareto front. A common set indicator of this kind is the hypervolume indicator: $HI_r(P) = Vol_m(DomSet(P) \cap [r, \infty])$ where Vol_m denotes the Lebesgue measure in dimension m (in 1-D length, in 2-D area, and in 3-D volume) and $r \in \mathbb{R}^m$ a user defined reference point.

Instead of taking the classical view that an indicator is an approximation measure to the Pareto front, we take the more general view that it measures how well the relevant parts of the subspace dominated by the Pareto front are covered by the space that is dominated by the given set of points (cf. [20]). In order to do so, set monotonic functions on the set of dominated subspaces will be introduced next. A *set monotonic function* $f: \mathcal{M} \to \mathbb{R}$ over a family of sets $\mathcal{M}$ is defined as a function that has the property: $S_1 \in \mathcal{M}$ and $S_2 \in \mathcal{M}$ and $S_1 \subset S_2$ implies $f(S_1) < f(S_2)$. The hypervolume indicator is for instance a set monotonic function over $\{D \cap R | D \in DomSets_m\}$, and R is the set of all points that dominate the reference point, but not for $DomSets_m$. A *weakly set monotonic function* $f: \mathcal{M} \to \mathbb{R}$ over a family of sets $\mathcal{M}$ is a function that has the property: $S_1 \in \mathcal{M}$ and $S_2 \in \mathcal{M}$ and $S_1 \subset S_2$ implies $f(S_1) \leq f(S_2)$.

A general way to construct (weakly) set monotonic functions over $DomSets_m$ is by using a density function $K: \mathbb{R}^m \to \mathbb{R}$. The density based hypervolume indicator function $DHI: DomSets_m \to \mathbb{R}$ is here introduced as:

$$DHI(S) := \int_{y \in S} K(y)\, dy.$$

The definition resembles the definition of the weighted hypervolume by Zitzler [12], but does not require a reference point. $K(y)$ must be chosen in such a way that the integral always exists. This can be guaranteed for density functions with the bounded improper integral (BI) property $\int_{x \in [-\infty, y]} K(x) dx < \infty$ for all $y \in \mathbb{R}^m$ (Here maximization is considered, for minimization use $\int_{x \in [y, \infty]} K(x)\, dx < \infty$ instead). If also the positive density (PD) property, $K(x) > 0$ for all $x \in \mathbb{R}^m$, holds* then the function is set monotonic over $DomSets_m$. This is the case, because firstly, the integral exists for all elements of $DomSets_m$ due to the BI property. Secondly, for $S_1 \in DomSets_m$ and $S_2 \in DomSets_m$ with $S_2 \subset S_1$the increment $DHI(S_1) - DHI(S_2)$ is simply given by the integral of the positive density K (see PD property) over the measurable set $S_1 - S_2$, which is strictly bigger than zero. If only the weaker property $K(x) \geq 0$ holds, then the function is still weakly set monotonic over $DomSets_m$. Note, that the standard hypervolume indicator is given by the special case of $K(x) = I(x\ dominates\ r)$ with I denoting the indicator function with $I(true) = 1$ and $I(false) = 0$. The hypervolume indicator is only a weakly set monotonic function over $DomSets_m$, as it is, e.g., assigning the same value of zero to all sets in $DomSets_m$ with elements dominated by the reference point.

In the following density functions based on desirability functions will be discussed. They accurately take into account the scaling of the objective functions according to the user's preferences. The density based approach requires no definition of a reference point, which is necessary for the standard hypervolume indicator. Moreover, it will be shown that it is a generalization of the concept of the desirability index from singletons to sets of alternative solutions, and for density functions in a product form offers a simple and fast computation scheme.

3. Construction of a density based on desirability functions

We consider the class of desirability functions which look like cumulative distribution functions (CDFs). The associated density functions are used in computing the weighted hypervolume. Of course, we can compute the weighted hypervolume for any density function but the above approach is very meaningful since we get an indicator which is compatible with certain decision theoretic interpretations of desirability functions.

Desirability functions, here for maximization, are functions $D_i: \mathbb{R} \to [0,1]$, $i = 1, \ldots, m$ that map objective function values to the interval [0,1] where the value of 0 indicates that the solution is not acceptable and a value of 1 indicates that the attained objective function value fully satisfies the decision maker and s/he would be indifferent to a further increase of the objective function value. There are two common types of desirability functions in the literature, the Harrington type [5,8], which was first proposed, and the Derringer-Suich type[6].

The Harrington type of desirability functions is given by the function: $D_i(y_i) = \exp(-\exp(-(b_{0i} + b_{1i} y_i)))$ for some constants b_0 (absolutely satisfying level) and b_1 (marginally infeasible level). Harrington (1965) proposed the geometric mean $D(y) = (\prod_{i=1}^{m} D_i(y_i))^{1/m}$ for computing the desirability index that is an aggregate of m desirability functions. Here we will use the m-th power of the geometric mean instead, because this gives rise to a more simple probabilistic interpretation. The product form of a desirability index reads: $D(y) = \prod_{i=1}^{m} D_i(y_i)$. The two aggregations are rank invariant as the m-th root function is strict monotonic. Both forms have the property that a single desirability function value of zero causes the aggregate to be zero as well. However, the Harrington desirability functions obtain values in the open interval (0,1) and therefore this property will only be relevant in the discussion of limit properties. See Figure 1(a) for a visualization of shapes for different settings of b_0 and b_1.

The alternative choice of D_i by Derringer and Suich can well reach the boundaries of the interval [0,1]. For this approach we have:

---

* It is also true if the condition on x holds almost everywhere.

$$D_i = \begin{cases} 0, if \ \ y_i < L_i \\ ((y_i - L_i)/(U_i - L_i))^{l_i} \quad if \ L_i \leq y_i < U_i \\ 1, if \ y_i \geq U_i \end{cases}$$

The highest value where the i-th desirability function has a value of 0 is named L_i and the lowest value where the i-th desirability function has reached the value of 1 is named U_i. Moreover, l_i determines the steepness of the change from low to high values. See Figure 1 (b) for a visualization of the different parameter settings.

Our research question is now whether we can construct the kernel K for a DHI_K indicator that is (weakly) set monotonic over $DomSets_m$ and is also highly compatible with what is expressed by certain desirability functions. A straightforward proposal might be to just use the desirability index $D_m((y_1, \dots, y_m)^T)$ as a density function for the weighted hypervolume. However, it will be motivated next that it is a better choice to use the derivative of the desirability index instead. To be compatible with the above properties of desirability functions the kernel K should give rise to the DHI indicator which should at least have the following invariant properties:

1. For a singleton the weighted hypervolume indicator should obtain the value of the original desirability index.
2. For points in P where the desirability is zero for at least one function the contribution of the point to the weighted hypervolume indicator should always be zero.
3. For points in P where the desirability is 1 for the i-th function value, a further increment of this function value should not yield a further increment of the weighted hypervolume indicator.

Assuming, d_i being the derivative of the i-th desirability function D_i, which should be defined almost everywhere, a weighted hypervolume indicator that has these properties can be constructed by:

$$DHI(P) = \int_{y \in DomSet(P)} K(y)\, dy, \qquad K(y) \coloneqq \frac{\partial^m D}{\partial y_1 \dots \partial y_m}(y)$$

Here the integrand can be interpreted as a multivariate probability density function (see [13], page150), because desirability functions, for maximization, have mathematical properties of cumulative distribution functions (CDFs), albeit their standard interpretation is different.

When using the Harrington type desirability functions, then DHI will be set monotonic over $DomSets_m$, whereas for the Derringer-Suich type desirability functions this indicator is only weakly set monotonic.

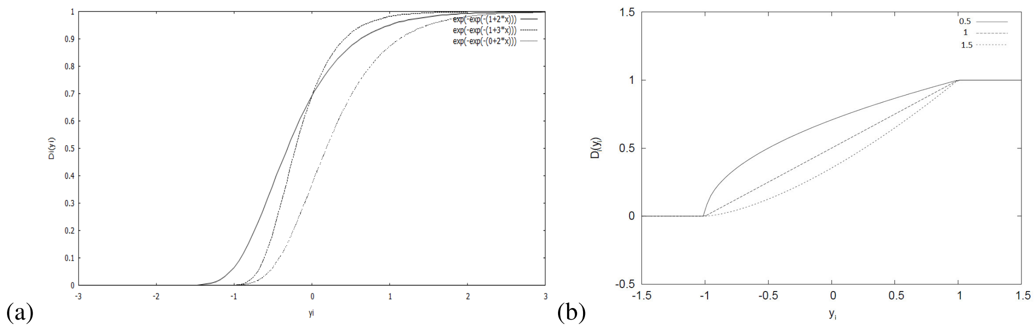

Figure 1 (a) Harrington type of desirability functions for different parameters of b_{0i} and b_{1i}. (b) Derringer-Suich desirability functions for maximization. L_i is set to -1 and U_i is set to 1. l_i is set to 0.5, 1 and 1.

The fact that desirability functions for maximization objectives can be seen as cumulative distribution functions raises the question on whether there could be a probabilistic interpretation for them. Although the standard interpretation of desirability function is not probabilistic, a probabilistic interpretation of the desirability function could be as follows: The desirability function $D_i(f_i(x))$ could be interpreted as the probability that the decision maker will accept a solution x if s/he would only look at the objective function f_i in isolation. The desirability index $D(f(x)) = \prod_{i=1}^{m} D_i(f_i(x))$ is the probability that a decision maker accepts all objective function values of x, assuming the probability variables for the acceptance of the single objective function values are independent of each other. Extending the probabilistic interpretation of desirability to sets, it can be said that $DHI_K(P)$ is the probability that the decision maker accepts at least one solution in the set. Hence, within this interpretation context, it is justified to interpret the weighted hypervolume indicator of a set as the *acceptance probability of that set*.

4. Analysis

Next we will show how the integral can be computed, first for the 2-D case and then for the general case. A schematic picture (for the Harrington desirability function) is drawn in Figure 2. In the following we will denote the value of f_1 with y_1 and the value of f_2 with y_2. Starting from desirability functions which have the form of cumulative probability functions, we will take the product of the associated density functions as the density for computing the weighted hypervolume: Let $P = \{(y_1^{(1)}, y_2^{(1)}), (y_1^{(2)}, y_2^{(2)}), \dots, (y_1^{(n)}, y_2^{(n)})\}$ be a non-dominated subset of $\mathbb{R}^2$ which is sorted in the first coordinate in ascending order (this entails that the second coordinate is sorted in descending order). Let the desirability functions be $D_i: \mathbb{R} \to \mathbb{R}, i = 1,2$ and the derivatives of these functions are denoted by $d_i, i = 1,2$ (assume the derivatives of the D_is exist almost everywhere). We assume (for $i = 1,2$) that $\lim_{u\to\infty} D_i(u) = 1$, $\lim_{u\to-\infty} D_i(u) = 0$ and $1 \geq D_i(u) \geq 0$, for all u (these are necessary conditions for a function to be a CDF). The density in a point (s_1, s_2) of $\mathbb{R}^2$ is defined as follows:

$$K(s_1, s_2) = d_1(s_1) d_2(s_2),$$

and in the points where the d_1 or d_2 is not defined the density is defined to be 0.

For a box of $\mathbb{R}^2$ which is denoted by the left lower corner (a_1, a_2) and the right upper corner (b_1, b_2) the contribution to the weighted hypervolume is as follows:

$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} d_1(s)\, d_2(u)\, ds\, du.$$

Using the fact that we know the anti-derivatives of $d_i, i = 1,2$ are D_1, D_2 we get:

$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} d_1(s)\, d_2(u)\, ds\, du = (D_1(b_1) - D_1(a_1))(D_2(b_2) - D_2(a_2)),$$

where a_1 or a_2 could possibly be $-\infty$. The DHI for a given P and desirability functions D_1, D_2 is easy to compute:

$$DHI = DHI(P, D_1, D_2) = \sum_{i=1}^{n-1} (D_1(y_1^{(i)}) - D_1(-\infty))\,(D_2(y_2^{(i-1)}) - D_2(y_2^{(i)})) +$$
$$\left(D_1(y_1^{(n)}) - D_1(-\infty)\right)\left(D_2(y_2^{(n)}) - D_2(-\infty)\right).$$

Or unfolding the products in the above formula we can get a slightly simpler expression for the weighted hypervolume:

$$DHI(P, D_1, D_2) = D_1(y_1^{(1)}) D_2(y_2^{(1)}) - D_1(y_1^{(1)}) D_2(y_2^{(2)}) + D_1(-\infty) D_2(y_2^{(1)})$$
$$\sum_{i=2}^{n-1} D_1(y_1^{(i)}) D_2(y_2^{(i)}) - D_1(y_1^{(i)}) D_2(y_2^{(i+1)}) +$$
$$D_1(y_1^{(n)}) D_2(y_2^{(n)}) - D_1(y_1^{(n)}) D_2(-\infty) + D_1(-\infty) D_2(-\infty).$$

For the Harrington and Derringer-Suich desirability functions the limit values $D_1(-\infty)$ and $D_2(-\infty)$ both are 0. Therefore the computation is equivalent to computing for each point $x \in P$ the transformed objective function vector $(D_1(f_1(x)), D_2(f_2(x)))$ and then compute the standard hypervolume indicator for the transformed population of

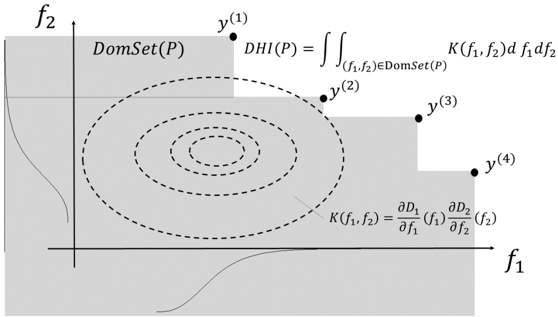

Figure 2 Schematic drawing of the Integral of the desirability function based hypervolume DHI_D(P) for some population P. Below the f_1 axis and to the left of the f_2 axis are the desirability functions D_1 and D_2.

objective function vectors and a reference point of 0. This corresponds to the approach suggested by Wagner and Trautmann [1] to transform a MOP using desirability functions before computing the hypervolume indicator. The DHI is therefore not new for the product type of desirability indices. Rather our paper motivates this indicator in an alternative way. This can provide a new insight on the relationship between the weighted hypervolume indicator and the desirability function based transformation of the coordinate axis used in [1], identifying the second as a special case of the first.

Figure 3 shows the behaviour for singletons. The aggregation used in Harrington type and Derringer-Suich type of desirability functions is the product. The figure shows that in case of singleton sets the values of the DHI coincide with the values of the product of desirability function values and satisfy the the requirements for the DHI stated earlier in Section 3.

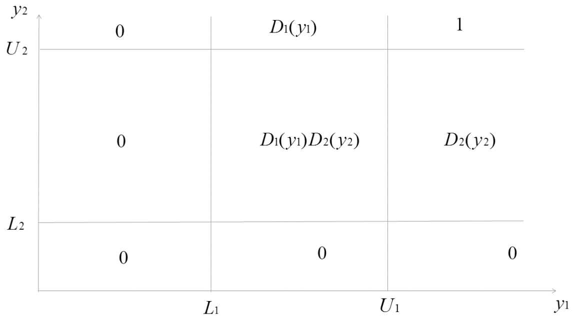

Figure 3 The desirability set-indicator DHI({y}) of a singleton set S={y}.

The discussion above is only for the two dimensional case. It is straightforward to extend the computation to higher dimensional cases. As long as the density function is in a product form, all that needs to be done is to partition the dominated subspace $DomSet(P)$ into boxes (see [7] and [14]) and compute for each one of the boxes the contribution to the integral using Fubini's theorem. To show this, let us denote a single box of the decomposed dominated hypervolume (bound by the reference point) with the orthogonal range $[(l_1, \dots l_m)^T, (u_1, \dots, u_m)^T]$. Then

$$\int_{y_1=l_1}^{u_1} \dots \int_{y_m=l_m}^{u_m} PDF(y_1, \dots, y_m)\, dy_1 \dots dy_m$$
$$= \int_{y_1=l_1}^{u_1} \dots \int_{y_m}^{u_m} \prod_{i=1}^{m} \frac{\partial D_i}{\partial y_i} dy_1 \dots dy_m$$
$$= \prod_{i=1}^{m} (D_i(u_i) - D_i(l_i)) = Vol([(D_1(l_1), \dots, D_m(l_m)), (D_1(u_1), \dots, D_m(u_m))]).$$

Again, given a product form, it becomes obvious that the DHI can be computed by a simple coordinate transformation: To compute the DHI, replace all points a in the given set P by $(D_1, \dots, D_m)(a)$ and compute the standard hypervolume indicator with reference point zero for this transformed set. This is equivalent to the axes transformation in [1]. We also see that, in case of a product form, we can compute the DHI with the same amount of computational resources as for the standard hypervolume indicator. For fast algorithms see [7] for the 2- respectively 3-dimensional cases and [18] for dimensions higher than 3. Once we deviate from the product type of aggregation we can no longer apply Fubini's theorem and have to compute the integrals over the boxes $[l, u]$, instead of using the coordinate transformation. Therefore, in terms of aggregation possibilities the weighted hypervolume approach is more general than the coordinate transformation approach. An example for a non-product aggregation is provided in the next section.

5. Numerical Study for Bivariate Gaussian Distribution

Instead of using the classical desirability functions, in the following study a desirability function that has the shape of a Gaussian cumulative distribution function will be discussed. The parameters mean value μ and standard deviation σ can be used to control its shape. The corresponding bivariate joint probability distribution has an additional correlation parameter ρ. The general shape for the bivariate PDF is given by:

$$PDF(y_1, y_2) = \frac{1}{2\,\pi\,\sigma_1\,\sigma_2\,\sqrt{(1-\rho^2)}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(y_1-\mu_1)}{\sigma_1}\cdot\frac{(y_2-\mu_2)}{\sigma_2} + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right)\right].$$

See Figure 4 for an example of a PDF for $\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 1$, and $\rho = +0.9$. Whenever $\rho \neq 0$ the PDF loses its product form, and the transformation of the coordinate axes cannot be applied anymore for the computation of DHI. Instead other forms of cumulative distribution functions arise that can be interpreted as different means of aggregation of desirability functions. In a decision theoretic setting the contours of the cumulative joint distribution functions are interpreted as indifference curves. Figure 5 depicts these curves for the cases $\rho = -0.9, \rho = 0,$ and $\rho = 0.9$. For the discussion we interpret $D_i(y_i) = \Phi\mu_i, \sigma_i(y_i),\ i = 1,2$ as the desirability of the i-th objective function value y_i. In the extreme case of a negative correlation $\rho \to -1$ the resulting indifference curves given by the contours of the bivariate distribution function resemble lines and this corresponds to a linear aggregation: $D(y_1, y_2) = D_1(y_1) + D_2(y_2)$. In case of $\rho = 0$ the aggregation is done in the standard way of a product, i.e. $D(y_1, y_2) = D_1(y_1) \cdot D_2(y_2)$. Then, in case of $\rho = 1$ the aggregation is of the form $max\{D_1(y_1), D_2(y_2)\}$. In the last case the best desirability function value of a solution determines the desirability index of the solution. All other settings of ρ fall into the 'gray area' between these pure forms of using aggregation to define a desirability index.

In Figure 6 results are shown of approximations of the Pareto Fronts which arise in 2-dimensional problems by using the weighted hypervolume for a density with positive correlation and a density with correlation equal to 0. The Pareto Fronts considered are the Lamé curves for different values of the parameter γ (see [19]). Here, SMS-EMOA was used as a search algorithm, and the hypervolume contributions were computed according to the density given by a normal distribution with mean values 0 and standard deviations 1. Different values of the correlation parameter ρ were compared. The pictures give some evidence that a positive correlation gives rise to approximations which are more concentrated about the more desirable region around the knee point. Note, that this study only approximates the optimal distribution of points, as the algorithm that is used is based on heuristic search. We computed 10000 iterations (function evaluations) which for the given problems typically yields accurate approximations. Source code for reproduction of the experiments is available at http://natcomp.liacs.nl/index.php?page=code or on request by the authors. Analytical methods to determine the optimal distribution of points for a given density were described in [2]. In higher dimensions, the single parameter would have to be replaced by a covariance matrix with entries $\rho_{i,j}$ for the correlations between each two marginal probability distributions.

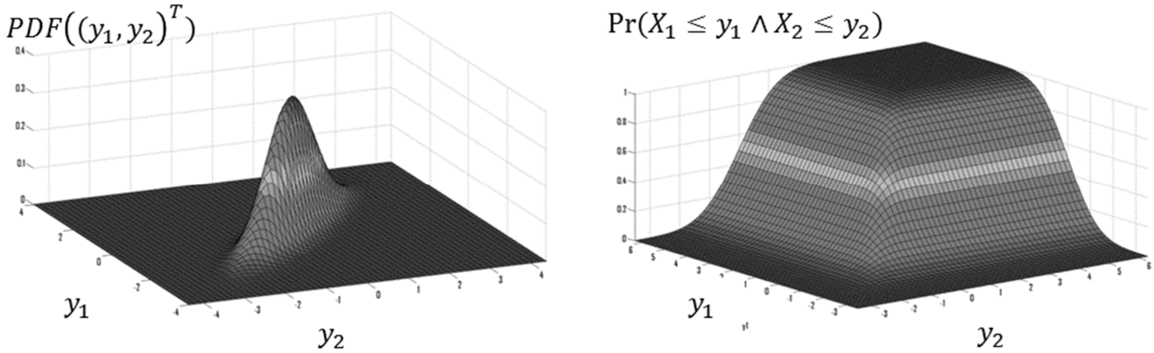

Figure 4 PDF (left) and CDF (right) of a correlated Gaussian joint probability distribution with correlation coefficient ρ=0.9.

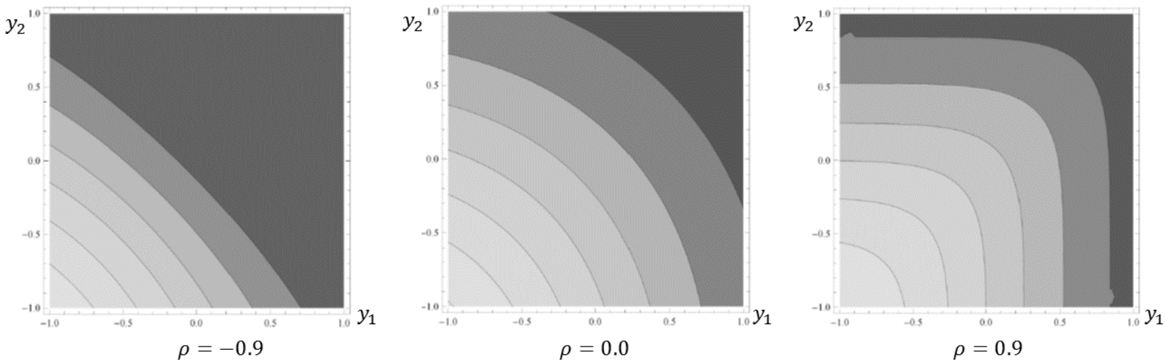

Figure 5 Bivariate Gaussian CDF for different values of the correlation parameter ρ.

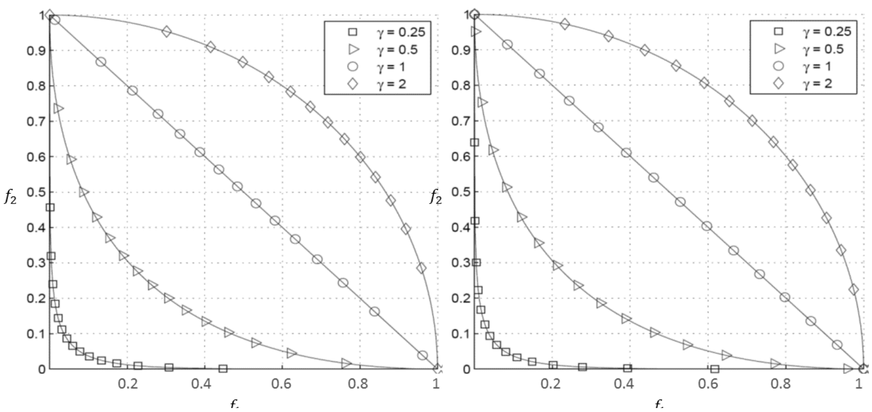

Figure 6 Sets with optimal DHI on the bivariate Gaussian density with $\mu = 0, \sigma = 1$ and for different Pareto fronts given by arcs of Lamé supercircles computed by SMS-EMOA with different values of correlation ρ. Left: $\rho = 0.9$. Right: $\rho = 0$.

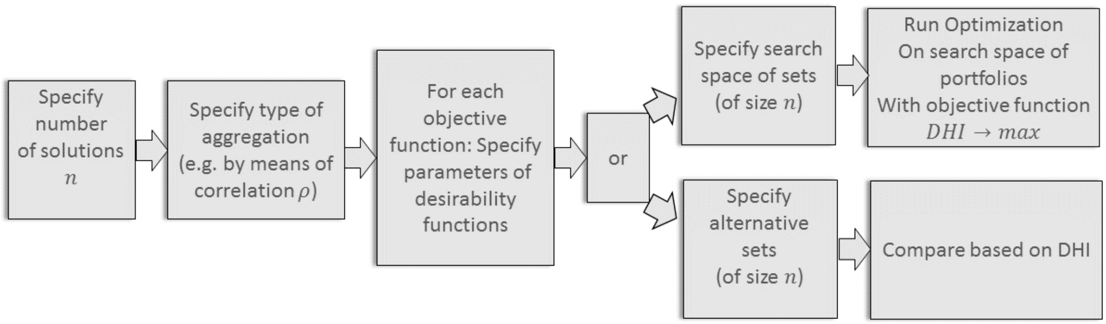

Figure 7 Workflow diagram for setting up and using the density based hypervolume DHI in set comparison and set oriented optimization.

6. Practical guidelines for using the DHI for finding optimal sets of alternative solutions

A typical example of the use of the proposed preference based set performance indicator DHI can be found, e.g., in search applications. For instance, we can consider a search by keywords in a search engine or a query in a database. The decision maker is interested in obtaining a subset of solutions, which are good with respect to several objectives, e.g., contain all the keywords in the resulting text and catch the diversity of the results in case of possible multiple meanings for the same combinations of the keywords.

In terms of multiobjective optimization, solving such problems corresponds to obtaining approximate Pareto front solutions considering that some Pareto front regions are more interesting for the decision maker when compared to the rest of the front. Desirability functions allow decision makers to specify such preferences in a form of parameters of desirability functions that are intuitive to the decision makers.

The proposed procedure to follow when using the proposed approach is as described below (see also Figure 7):

1. The decision maker specifies the initial parameters of the desirability functions for each objective.
    a. The decision maker has to specify the number of solutions s/he wishes to obtain as a result.
    b. The type of the desirability functions D_i and their aggregation function to form the index should also be selected (e.g. among arithmetic mean, geometric mean, max function or any other type), or in case of using the Gaussian CDF (Section 5) by providing the correlation parameter ρ.
    c. The decision maker should define the specific parameters of the desirability functions. For instance, for the Harrington type the absolutely satisfying b_0 and marginally infeasible b_1 values for each objective should be specified; and for Derringer-Suich type the highest value L_i of the objective function where the desirability is equal to 0, and the lowest value U_i of the objective function where the desirability is equal to 1 should be defined.
2. Running the comparison based on DHI or optimization methods with objective function DHI for a collection or a search space of sets of alternative solutions.
3. Analyzing the resulting best set of alternative solutions and selecting a single solution from it.

Under the given assumptions a set is obtained that maximizes the probability of containing an acceptable solution. The set can also be viewed as a Pareto front approximation that has an increased density in areas of high desirability. The weighted hypervolume indicator can be interpreted as the probability for the obtained set. Its value will depend on the available solutions in the search space and the size of the set.

7. Discussion

The family of weighted hypervolume indicators that has been obtained generalizes a desirability index for a single point to a desirability index for a set of points. All DHI indicators can be viewed as a special reference point free case of weighted hypervolume [2]. The new approach yields a weighted hypervolume indicator that for the Harrington and Derringer-Suich desirability functions, if aggregated by a product, is equivalent to the coordinate transformation approach used in the DF-SMS-EMOA [1]. From a modeling perspective the new approach, however, offers some interesting possibilities of extensions. It is now possible and straightforward to integrate the correlation of objective functions (by means of correlated joint probability distributions) and study density functions for set-indicators that are compatible with desirability function aggregations that are not in a product form. More importantly, the discussion on desirability functions points towards a widely applicable and very general technique to assign a meaningful density function in order to obtain a (weakly) set monotonic weighted hypervolume indicator for dominated sets. Also in the context of optimization with fuzzy constraints [4] this is of interest. The probability distribution models the probability of constraint violation and can be used in the DHI. The maximization of such indicators will yield solution sets that dominate a part of the search space that with high probability contains at least one feasible solution.

There are many directions to extend the approaches discussed in this paper in the future. The integral computation techniques discussed in this work also offer a simple and fast computation of the multiobjective probability of improvement discussed in [14, 15, 16]. For this the same transformation can be applied as discussed in Section 4 (using the i.i.d. Gaussian distribution function instead of a desirability index), yielding a linear time reduction to the problem of computing the standard hypervolume indicator for which efficient algorithms are known. It would also be interesting to consider instead of desirability functions other utility functions and theories on multicriteria decision analysis in the future. For instance, one can take into account different grades of desirability, differentiating between the decision maker's attitude towards risk, e.g., similar to methods in multi-attribute utility or value theory [10]. Also the possibility of relative evaluation of solutions with respect to the current solution needs to be studied in order to capture real world decision biases as, e.g., in prospect theory it was found that in order to capture subjective preferences realistically the possibility of biasing from standard attitudes towards gaining and losing is crucial [11].

References

[1] Wagner, T., & Trautmann, H. (2010). Integration of preferences in hypervolume-based multiobjective evolutionary algorithms by means of desirability functions. Evolutionary Computation, IEEE Transactions on, 14(5), 688-701.

[2] Auger, A., Bader, J., Brockhoff, D., & Zitzler, E. (2009, July). Investigating and exploiting the bias of the weighted hypervolume to articulate user preferences. In Proceedings of the 11th Annual conference on Genetic and evolutionary computation (pp. 563-570). ACM.

[3] Miettinen, K. (1999). Nonlinear multiobjective optimization. Springer.

[4] Kruisselbrink, J. W., Emmerich, M. T., Bäck, T., Bender, A., IJzerman, A. P., & van der Horst, E. (2009, January). Combining aggregation with Pareto optimization: a case study in evolutionary molecular design. In Evolutionary Multi-Criterion Optimization (pp. 453-467), Springer.

[5] Suich, R., & Derringer, G. C. (1977). Is the Regression Equation Adequate?: One Criterion. Technometrics, 213-216.

[6] Harrington, E.C. (1965), The desirability function, Industrial Quality Control, 21, 494-498.

[7] Guerreiro, A. P., Fonseca, C. M., & Emmerich, M. T. (2012). A Fast Dimension-Sweep Algorithm for the Hypervolume Indicator in Four Dimensions. In CCCG (pp. 77-82).

[8] Trautmann, H., & Weihs, C. (2006) On the distribution of the desirability index using Harrington's desirability function. Metrika, 63(2), 207-213.

[9] Bermúdez JL (2009). Decision Theory and Rationality, Oxford University Press Inc., New York.

[10] Keeney, R. & Raiffa H. (1976). Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Wiley, New York.

[11] Kahneman, D. & Tversky A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2): 263-291.

[12] Zitzler, E., Brockhoff, D., & Thiele, L. (2007). The hypervolume indicator revisited: On the design of Pareto-compliant indicators via weighted integration. In Evolutionary Multi-Criterion Optimization (pp. 862-876). Springer.

[13] A. Rényi: Wahrscheinlichkeitsrechnung, VEB Deutscher Verlag der Wissenschaften, Berlin, 1971

[14] Emmerich, M. (2005). Single-and multi-objective evolutionary design optimization assisted by gaussian random field metamodels. Dissertation, LS11, FB Informatik, Universität Dortmund, Germany.

[15] Keane, A. J. (2006). Statistical improvement criteria for use in multiobjective design optimization. AIAA journal, 44(4), 879-891.

[16] Couckuyt, I., Deschrijver, D., & Dhaene, T. (2013, October). Fast calculation of multiobjective probability of improvement and expected improvement criteria for Pareto optimization. Journal of Global Optimization, 1-20. DOI: 10.1007/s10898-013-0118-2

[17] Beume, N., Fonseca, C. M., López-Ibáñez, M., Paquete, L., & Vahrenhold, J. (2009). On the complexity of computing the hypervolume indicator. Evolutionary Computation, IEEE Transactions on, 13(5), 1075-1082.

[18] Chan, T. M. (2013, October). Klee's Measure Problem Made Easy. In Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on (pp. 410-419). IEEE.

[19] Emmerich, M.T.M. & Deutz A.H. (2007), Test Problems Based on Lamé Superspheres. In Evolutionary Multi-Criterion Optimization (pp. 922-936), Springer.

[20] Grunert da Fonseca, V. & Fonseca C. M. (2011). The Relationship between the Covered Fraction, Completeness and Hypervolume Indicators. Artificial Evolution (pp. 25-36), Springer.