On the Impact of Scalarizing Functions on Evolutionary Multiobjective Optimization

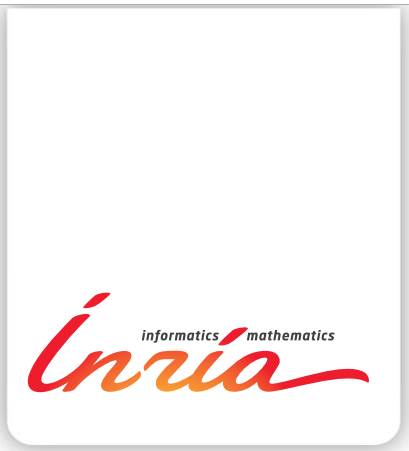
Bilel Derbel, Dimo Brockhoff, Arnaud Liefooghe, Sébastien Verel

On the Impact of Scalarizing Functions on Evolutionary Multiobjective Optimization

Bilel Derbel, Dimo Brockhoff, Arnaud Liefooghe, Sébastien Verel

RESEARCH REPORT
N° 8512
March 2014
Project-Teams Dolphin

ISSN 0249-6399 ISRN INRIA/RR--8512--FR+ENG



On the Impact of Scalarizing Functions on Evolutionary Multiobjective Optimization

Bilel Derbel*, Dimo Brockhoff†, Arnaud Liefooghe‡, Sébastien Verel§

Project-Teams Dolphin

Research Report n° 8512 — March 2014 — 18 pages

Abstract: Recently, there has been a renewed interest in decomposition-based approaches for evolutionary multiobjective optimization. However, the impact of the choice of the underlying scalarizing function(s) is still far from being well understood. In this paper, we investigate more carefully the effect of different scalarizing functions (weighted Tchebycheff, weighted sum, and combinations thereof) and their parameters on the search performance. We thereby abstract firstly from any specific algorithm and only consider the difficulty of the single scalarized problems in terms of the search ability of a $(1+\lambda)$-EA on biobjective NK-landscapes. Secondly, combining the outcomes of independent single-objective runs allows for more general statements on the impact of the scalarizing function parameters on set-based performance measures such as hypervolume and ε-indicator. Finally, we investigate the correlation between the opening angle of the scalarizing function's underlying contour lines and the position of the final solution found by the $(1+\lambda)$-EA in the objective space. Our analysis is of fundamental nature and sheds more light on the optimization of scalarizing functions within decomposition-based algorithms and also on how these algorithms might be improved.

Key-words: Evolutionary Multiobjective Optimization, scalarizing functions

* Inria Lille - Nord Europe, DOLPHIN team, France and Université Lille 1, LIFL, UMR CNRS 8022, France
† Inria Lille - Nord Europe, DOLPHIN team, France
‡ Inria Lille - Nord Europe, DOLPHIN team, France and Université Lille 1, LIFL, UMR CNRS 8022, France
§ Université du Littoral Côte d'Opale, LISIC, France

RESEARCH CENTRE
LILLE – NORD EUROPE

Parc scientifique de la Haute-Borne
40 avenue Halley - Bât A - Park Plaza
59650 Villeneuve d'Ascq

1 Introduction

Multiobjective optimization problems occur frequently in practice and evolutionary multiobjective optimization (EMO) algorithms have been shown to be well-applicable for them—especially if the problem under study is nonlinear and/or derivatives of the objective functions are not available or meaningless. Besides the broad class of Pareto-dominance based algorithms such as NSGA-II or SPEA2, a recent interest in the so-called *decomposition-based algorithms* can be observed. Those decompose the multiobjective problem into a set of single-objective, 'scalarized' optimization problems. The number of *scalarizing* subproblems usually indicate the final cardinality of the Pareto set approximation, identified by those methods. Examples of such algorithms include MSOPS [4], MOEA/D [11], and their many variants. We refer to [3] for a recent overview on the topic. The main idea behind those algorithms is to define a set of (desired) search directions in objective space and to specify the scalarizing functions corresponding to these directions. The scalarizing functions can then be solved independently (such as in the case of MSOPS), or in a dependent manner (like in MOEA/D where the recombination and selection operators are allowed to use information from the solutions maintained in neighboring search directions).

Many different scalarizing functions have been proposed in the literature, see e.g. [9] for an overview. Well-known examples are the weighted sum and the (augmented) weighted Tchebycheff functions, where the latter has an inherent parameter that controls the shape of the lines of equal function values in objective space. Especially with respect to decomposition-based EMO algorithms, it has been reported that the choice of the scalarizing function and their parameters has an impact on the search process [3]. Moreover, it has been noted that adapting the scalarizing function's parameters during the search can allow improvement over having a constant set of scalarizing functions [5]. Although several studies on the impact of the scalarizing function have been conducted in recent years, e.g. [6], to the best of our knowledge, all of them investigate it on a concrete EMO algorithm and on the quality of the resulting solution *sets* such as the hypervolume indicator when more than one scalarizing function is optimized (typically as mentioned above, in a dependent manner). Thereby, the focus is not in understanding why those performance differences occur but rather in observing them and trying to improve the global algorithm. However, we believe that it is more important to first understand thoroughly the impact of the choice of the scalarizing function for a *single* search direction before analyzing more complicated algorithms such as MOEA/D-like approaches with specific neighboring structures, recombination, and selection operators. In this paper, we fundamentally investigate the impact of the choice of the scalarizing functions and their parameters on the search performance, independently of any known EMO algorithm. Instead, we consider one of the most simple single-objective scalarizing search algorithm, i.e., a $(1+\lambda)$-EA with standard bit mutation, for optimizing a single scalarizing function, corresponding to a single search direction in the objective space. Experiments are conducted on well-understood bi-objective ρMNK-landscapes.

More concretely, we look experimentally at the impact of the parameters of a generalized scalarizing function (which covers the special cases of the weighted sum and augmented Tchebycheff scalarizing functions) in terms of the position (angle/direction) reached by the final points, as well as their quality with respect to the Tchebycheff function. We then consider how the opening of the cones that describe the lines of equal scalarizing function values can provide a theoretical explanation for the impact of the final position of the obtained solutions in objective space. We also investigate the resulting *set quality* in terms of hypervolume and ε-indicator if several scalarizing $(1+\lambda)$-EAs are run independently for different search directions in the objective space. Finally, we conclude our findings with a comprehensive discussion of promising research lines.

2 Scalarizing Functions

We consider the maximization of two objectives f_1, f_2 that map search points $x \in X$ from a search space X to an objective vector $f(x) = (f_1(x), f_2(x)) = (z_1, z_2)$ in the so-called objective space $f(X)$. A solution x is called *dominated* by another solution y if $f_1(y) \geq f_1(x)$, $f_2(y) \geq f_2(x)$, and for at least one i, $f_i(y) > f_i(x)$ holds. The set of all search solutions, not dominated by any other, is called Pareto set and its image Pareto front. To increase readability, we might refer to an objective vector as a solution and consider objective vectors only in the remainder of the paper.

Many ways of *decomposing* a multiobjective optimization problem into a (set of) single-objective *scalarizing functions* exist, including the prominent examples of *weighted sum* (WS), *weighted Tchebycheff* (T), or *augmented weighted Tchebycheff* (S_{aug}) [9]. For most of them, theoretical results, especially about which Pareto-optimal solutions are attainable, exist [9, 7] but they are typically of too general nature to allow for statements on the actual search performance of (stochastic) optimization algorithms. Instead, we are here not interested in any particular scalarizing function, but rather in understanding which general properties of them influence the search behavior of EMO algorithms. We argue by means of experimental investigations that it is not the actual choice of the scalarizing function or their parameters that makes the difference in terms of performance, but rather the general properties of the resulting lines of equal function values. To this end, we consider the minimization of the following general scalarizing function that covers the special cases of WS[1], T, and S_{aug} functions:

$$\mathsf{S}_{\text{gen}}(z) = \alpha \cdot \max\{\lambda_1 \cdot |\bar{z}_1 - z_1|, \lambda_2 \cdot |\bar{z}_2 - z_2|\} + \varepsilon\,(w_1 \cdot |\bar{z}_1 - z_1| + w_2 \cdot |\bar{z}_2 - z_2|)$$

where $z = (z_1, z_2)$ is the objective vector of a feasible solution, $\bar{z} = (\bar{z}_1, \bar{z}_2)$ a utopian point, $\lambda_1, \lambda_2, w_1$, and $w_2 > 0$ scalar weighting coefficients indicating a search direction in objective space, and $\alpha \geq 0$ and $\varepsilon \geq 0$ parameters to be fixed. For more details about the mentioned scalarizing functions and their relationship, we refer to Table 1.

In the following, we also consider a case of S_{gen} that combines WS and T with a single parameter ε: the normalized $\mathsf{S}_{\text{norm}}(z) = (1-\varepsilon)\mathsf{T}(z) + \varepsilon\mathsf{WS}(z)$ where $\alpha = 1 - \varepsilon$ and $\varepsilon \in [0, 1]$. For optimizing in a given search direction (d_1, d_2) in objective space, we follow [4, 1] and set $\lambda_i = 1/d_i$.[2] In addition, we refer to the direction angle as $\delta = \arctan(d_1/d_2)$. For the case of S_{norm}, we furthermore choose $w_1 = \cos(\delta)$ and $w_2 = \sin(\delta)$ (thus, $w_1^2 + w_2^2 = 1$) for the weighted sum part in order to normalize the search directions in objective space uniformly w.r.t. their *angles*. Though, in many textbooks you can find statements like "ε has to be chosen small (enough)", we do not make such an assumption but want to understand which influence ε has on the finally obtained solutions and how it introduces a trade-off between the Tchebycheff approach and a weighted sum. For the question of how small ε should be chosen to find all Pareto-optimal solutions in exact biobjective discrete optimization, we refer to [2].

As mentioned above, one important property of a scalarizing function turns out to be the shape of its sets of equal function values, which are known for the WS, T, and S_{aug} functions [9]. However, no description of the equi-function-value lines for the general scalarizing function S_{gen} has been given so far. We think that it is necessary to state those opening angles explicitly in order to gain a deeper intuitive understanding of the above scalarizing approaches and related concepts such as the R2 indicator [1] or more complicated scalarizing algorithms such as MOEA/D [11]. Moreover, it allows us to investigate how a linear combination of weighted sum and Tchebycheff functions affect the search behavior of decomposition-based algorithms. The following theorem

[1] Contrary to the standard literature, our formalization assumes minimization of all scalarizing functions and we therefore have included the utopian point $\bar{z}$ that is typically assumed to be $\bar{z} = (0, 0)$ for minimization.

[2] The pathologic cases of directions parallel to the coordinates are left out to increase readability.

Table 1: Overview of the considered scalarizing functions, and the corresponding angles of the lines of equal function values with the standard Pareto dominance cone., and how to obtain the definition from the general scalarizing function S_{gen}.

scalar function	parameters in S_{gen}	opening angles	reference
$\mathsf{WS}(z) = w_1\lvert\bar{z}_1 - z_1\rvert + w_2\lvert\bar{z}_2 - z_2\rvert$	$\alpha = 0$, $\varepsilon = 1$	$\theta_1 = \arctan\left(-\frac{w_1}{w_2}\right)$ $\theta_2 = \frac{\pi}{2} + \arctan\left(\frac{w_1}{w_2}\right)$	[9, Eq. 3.1.1]
$\mathsf{T}(z) = \max\{\lambda_1\lvert\bar{z}_1 - z_1\rvert, \lambda_2\lvert\bar{z}_2 - z_2\rvert\}$	$\alpha = 1$, $\varepsilon = 0$	$\theta_1 = 0$ $\theta_2 = \pi/2$	[9, Eq. 3.4.2]
$\mathsf{S}_{\text{aug}}(z) = \mathsf{T}(z) + \varepsilon\left(\lvert\bar{z}_1 - z_1\rvert + \lvert\bar{z}_2 - z_2\rvert\right)$	$\alpha = 1$, $w_1 = w_2 = 1$	$\theta_1 = \arctan\left(-\frac{\varepsilon}{\lambda_1+\varepsilon}\right)$ $\theta_2 = \frac{\pi}{2} + \arctan\left(\frac{\varepsilon}{\lambda_2+\varepsilon}\right)$	[9, Eq. 3.4.5]
$\mathsf{S}_{\text{norm}}(z) = (1-\varepsilon)\mathsf{T}(z) + \varepsilon\mathsf{WS}(z)$	$\alpha = 1 - \varepsilon$, $w_i = 1/\lambda_i$	$\theta_1 = \arctan(-\frac{\varepsilon w_1}{(1-\varepsilon)\lambda_2+\varepsilon w_2})$ $\theta_2 = \frac{\pi}{2} + \arctan(\frac{\varepsilon w_2}{(1-\varepsilon)\lambda_1+\varepsilon w_1})$	here

Table 2: Parameter setting.

scalarizing functions	ρMNK-landscapes	$(1+\lambda)-$EA
$\bar{z} = (1,1)$	$\rho \in \{-0.9, -0.8, \ldots, 0.0, \ldots, 0.9\}$	$\lambda = n$
$\delta = j \cdot 10^{-2} \cdot \frac{\pi}{2}$, $j \in [\![1, 99]\!]$	$m = 2$	bit-flip rate $= 1/n$
S_{norm}: $\varepsilon = \ell \cdot 10^{-2}$; $\ell \in [\![0, 100]\!]$	$n = 128$	stopped after
S_{aug}: $\varepsilon = \ell \cdot 10^{-k}$; $\ell \in [\![0, 10]\!]; k \in [\![-1, 2]\!]$	$k = 4$	n iterations

states these opening angles θ_i between the cones of equal function values and the f_1-axis, see also Fig. 4 for some examples.

Theorem 1 *Let $\bar{z}$ be a utopian point, $\lambda_1, \lambda_2, w_1$, and $w_2 > 0$ scalar weighting coefficients, $\alpha \geq 0$ and $\varepsilon \geq 0$, where at least one of the latter two is positive. Then, the polar angles between the equi-utility lines of S_{gen} and the f_1-axis are $\theta_1 = \arctan(-\frac{\varepsilon w_1}{\alpha\lambda_2+\varepsilon w_2})$ and $\theta_2 = \frac{\pi}{2} + \arctan(\frac{\varepsilon w_2}{\alpha\lambda_1+\varepsilon w_1})$.*

Theorem 1 is only a corollary from the more general Theorem 2 that gives the concrete lines of equal function values and which, for readability reasons, is proven in the appendix.

3 Experimental Design

This section presents the experimental setting allowing us to analyze the scalarizing approaches introduced above on bi-objective ρMNK-landscapes. The family of ρMNK-landscapes constitutes a problem-independent model used for constructing multiobjective multimodal landscapes with objective correlation [10]. A bi-objective ρMNK-landscape aims at maximizing an objective function vector $f : \{0,1\}^n \to [0,1]^2$. Feasible solutions are binary strings of size n. The parameter k is the number of variables that influence a particular position from the bit-string (the epistatic interactions). By increasing the number of variable interactions k from 0 to $(n-1)$, landscapes can be gradually tuned from smooth to rugged. The parameter m defines the number of objective functions (here $m = 2$). At last, a correlation parameter ρ defines the degree of conflict between the objectives.

In this paper, we investigate the two scalarizing functions S_{norm} and S_{aug} of Table 1 with different parameter settings for the weighting coefficient vector and the ε parameter, as reported in Table 2. In particular, the WS (resp. T) function corresponds to S_{norm} with $\varepsilon = 1$ (resp. $\varepsilon = 0$). The set of weighting coefficient *direction angles* δ_j with respect to the f_1-axis ($j \in \{1, \ldots, 99\}$)

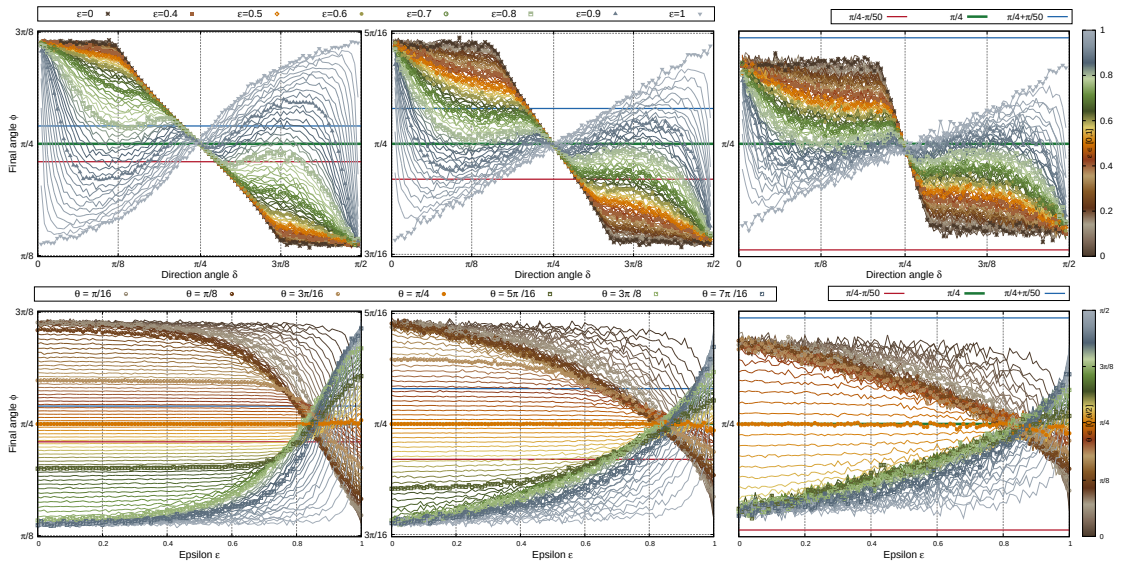


Figure 1: Angles of final solutions for S_{norm} as a function of δ (first row) and as a function of ε (second row). Columns show, from left to right, results for $\rho \in \{-0.7, 0.0, 0.7\}$.

are uniformly defined with equal distances in the angle space such that $\delta_j = j \cdot 10^{-2} \cdot \frac{\pi}{2}$. For both functions, we set $\lambda_1 = 1/\cos(\delta_j)$, and $\lambda_2 = 1/\sin(\delta_j)$. We recall that for S_{norm}, $w_i = 1/\lambda_i$, and for S_{aug}, $w_i = 1$. To evaluate the relative and the joint performance of the considered scalarizing functions, we investigate the dynamics and the performance of a randomized local search, a simple $(1+\lambda)-$EA. The search process is initialized with a random solution. At each iteration, λ offspring solutions are generated by means of an independent bit-flip mutation, where each bit of the parent solution is independently flipped with a rate $1/n$. The solution with the best (minimum) scalarizing function value among the parent and the offspring is chosen for the next iteration. For each configuration, 30 independent executions are performed.

4 Single Search Behavior

This section is devoted to the study of the optimization paths followed by *single* independent $(1+\lambda)$-EA runs for each direction angle δ and parameter ε of a scalarized problem. In particular, we study the final solution sets reached by the $(1+\lambda)$-EA in terms of diversity and convergence and give a sound explanation on how the search behaviour is related to the lines of equal function values of the scalarizing functions.

4.1 Diversity: Final Angle

In Fig. 1, we examine the average *angle* of the *final* solution reached by the algorithm with respect to the f_1-axis using S_{norm}. The final angle of solution x is defined as $\phi(x) = \arctan(f_2(x)/f_1(x))$. It informs about the actual *direction* followed by the search process. Firstly, we can observe that for the extreme values $\varepsilon = 0$ and $\varepsilon = 1$ corresponding to T and WS, the final solutions are in symmetric positions with respect to the direction angle $\pi/4$. This is coherent with the symmetric nature of ρMNK-landscapes [10]. Secondly, for WS ($\varepsilon = 1$), every single direction angle infers a

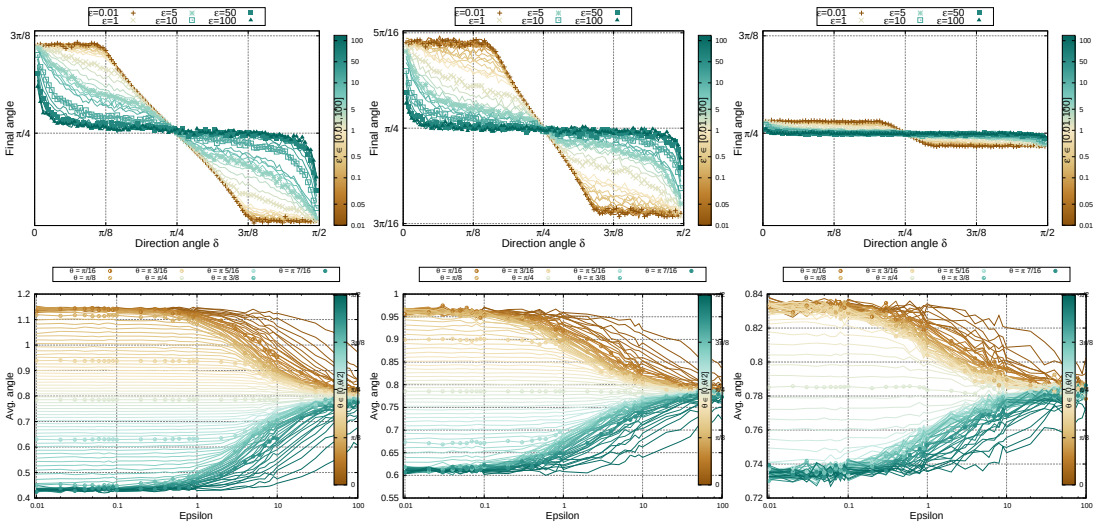


Figure 2: Angles of final solutions versus direction angle δ (first row) and ε (second row) for the S_{aug} function and a ρMNK-landscape with, from left to right, $\rho \in \{-0.7, 0.0, 0.7\}$.

different final angle whereas for T ($\varepsilon = 0$), the extreme direction angles end up reaching 'similar' regions of the objective space. These regions actually correspond to the lexicographically optimal points of the Pareto front[3], which is because of the choice of the utopian point that lies beyond them. Furthermore, when varying ε for a fixed δ, we can see that the search process is able to span a whole range of positions that are achieved by either T or WS but for variable δ values. Actually, when considering the direction angle being in the middle (i.e. $\delta \approx \pi/4$), the choice of ε does not substantially impact the search direction—because T and WS do allow to move to similar regions in this case. However, as the direction angle goes away from the middle, the influence of ε grows significantly; and the search direction is drifting in a whole range of values. This indicates that the choice of δ is not the only feature that determines the final angle but also the choice of ε highly matters: For some specific ε-values, the direction angles allow to distribute final angles fairly between the two lexicographically optimal points of the Pareto front—in the sense that each direction angle is inferring a different final angle, just like what we observe for WS. For some other ε-values, however, it may happen that the final angles are similar for two different direction angles. In particular, this is the case for large ε-values in S_{norm}, for which WS has more impact than T. Note that the same ε-value for instances with different objective correlation does not exactly lead to the same behavior, due to the different topologies of the Pareto front; see [10].

The previous discussion was for scalarizing function S_{norm}. Similar observations can also be made when considering scalarizing function S_{aug}, as illustrated in Fig. 2. When increasing the ε-value in S_{aug}, the T part of the function contributes only to a small extent. As a consequence, all final angles converge to the middle of the Pareto front, which is due to the WS part constituting S_{aug}, i.e., the search process degenerates to a pure WS search in the direction angle of $\pi/4$ given by the weights $w_1 = w_2 = 1$. As for S_{norm}, remark that there exists some ε-value for which every single direction angle is inferring a different region of the Pareto front.

The distribution of final directions is tightly related to the diversity of solutions computed by different independent single $(1+\lambda)$-EAs. As it will be discussed later, this is of crucial importance

[3] Points that reach the maximal possible value for at least one objective.

from a multiobjective standpoint, since diversity in the objective space is crucial to approach different parts of the Pareto front. However, the final angles depicted in Fig. 1 and Fig. 2 do not fully inform about how close every solution computed with a given parameter setting is to the Pareto front. This will be analyzed in the next subsection.

4.2 Convergence: Relative Deviation to Best

In the following, we examine the impact of the scalarizing function parameters on the performance of the $(1+\lambda)$-EA in terms of convergence to the Pareto front. For that purpose, we compute, for every direction angle δ, the best-found objective vector $z^{\star}_{\delta,\mathsf{T}}$ corresponding to the best (minimum) fitness value with respect to T, over all experimented parameter combinations and over all simulations we investigated. For both functions S_{norm} and S_{aug}, we consider the final objective vector z obtained for every direction angle δ and every ε-value. We then compute the relative deviation of z with respect to $z^{\star}_{\delta,\mathsf{T}}$, which we define as follows: $\Delta(z) = (\mathsf{T}(z) - \mathsf{T}(z^{\star}_{\delta,\mathsf{T}}))/\mathsf{T}(z^{\star}_{\delta,\mathsf{T}})$. Notice that this relative deviation factor is computed with respect to the T function, which is to be viewed as a reference measure of solution quality. This value actually informs about the performance of the $(1+\lambda)$-EA for a fixed direction angle, but variable ε-values.

In Fig. 3 (Left), we show the average relative deviation to best as a function of direction angles (δ) for different ε-values and the ρMNK-landscapes with an objective correlation of $\rho = -0.7$. To understand the obtained results, one has to keep in mind the results discussed in the previous section concerning the final angles inferred by a given parameter setting. In particular, since WS and T do not infer similar final angles, the final computed solutions lay in different regions of the objective space. Also, for the extreme direction angles, different ranges of ε imply different final angles. Thus, it is with no surprise that the average relative deviation to best can be substantial in such settings. However, the situation is different when considering direction angles in the middle ($\delta \approx \pi/4$). In fact, we observed that for such a configuration, the ε-value does *not* have a substantial effect on final angles, i.e., final solutions lie in similar regions of the objective space. Hence, one may expect that the search process has also the same performance in terms of average deviation to best. This is actually *not* the case since we can observe that the value of ε has a significant impact on the relative deviation for the non-extreme direction angles. To better illustrate this observation, we show, in the right-hand column of Fig. 3, the ε-value providing the minimum average relative deviation to best as a function of every direction angle. We clearly see that the best performance of the $(1+\lambda)$-EA for different direction angles are not obtained with the same ε-value.

4.3 Understanding the Impact of the Opening Angle

In this section, we argue that the dynamics of the search process observed previously is rather independent of the scalarizing function under consideration or its parameters. Instead, we show that the search process is guided by the positioning of the lines of equal function values in the objective space—described by the opening angle, i.e., the angle between the line of equal function values and the f_1-axis (cf. Theorem 1).

Fig. 4 shows four typical exemplary executions of the $(1+\lambda)$-EA in the objective space for different parameter settings. The set of brown squares is the best-known Pareto front approximation for the instance under consideration, the red line is the direction defined by the direction angle δ, and the semi-lines are the cones of equal function values. The offspring population is drawn at different iterations, together with the path followed by the current solution selected by the algorithm (in light blue).

The typical initial solution maps around the point $z = (0.5, 0.5)$ in the objective space, which

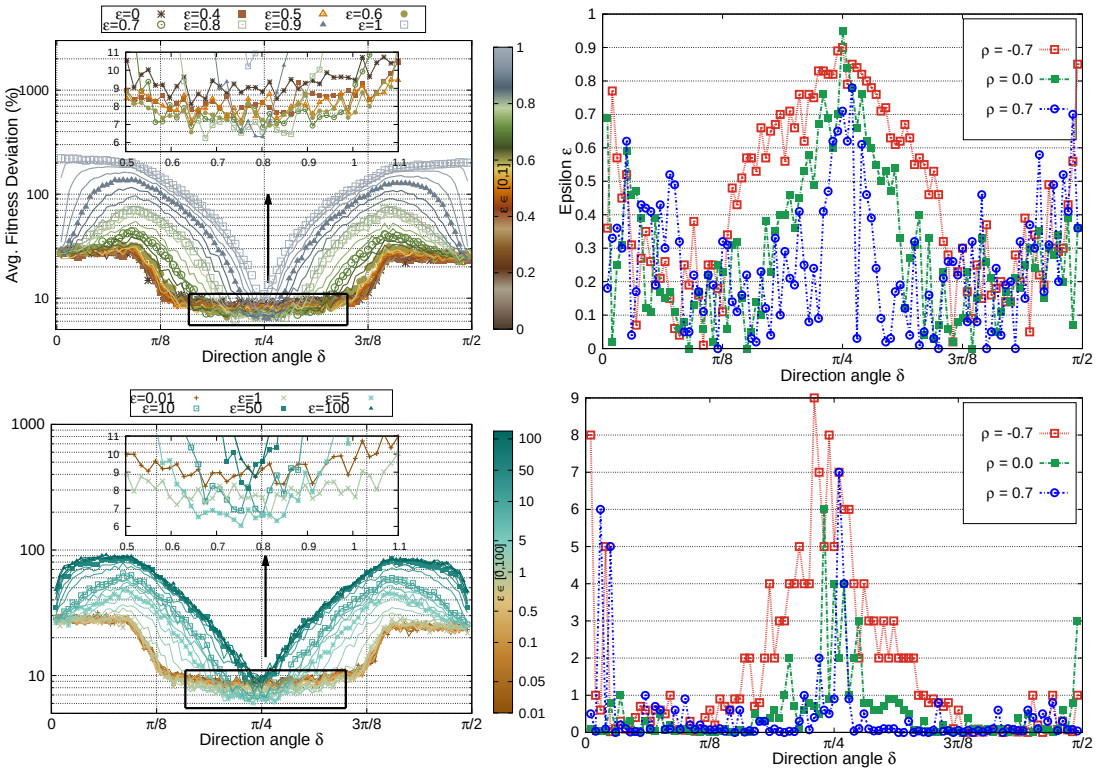


Figure 3: Left: Average deviation to best as a function of weight vector for S_{norm} (top) and for S_{aug} (bottom) and an objective correlation of $\rho = -0.7$ within the ρMNK-landscape instance. Right: ε-values providing the smallest deviation for a fixed direction (again for S_{norm} (top) and for S_{aug} (bottom)).

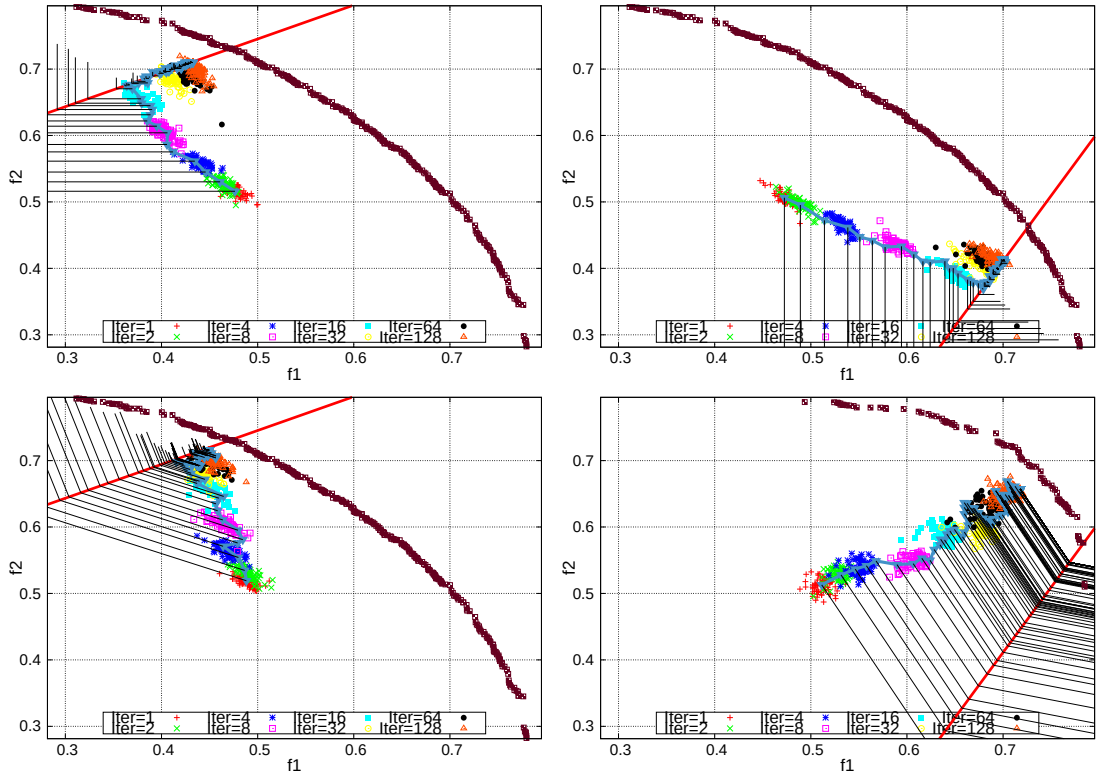


Figure 4: Exemplary runs of the $(1+\lambda)$-EA for different direction angles δ (red line) and different ε-values (S_{norm}, $\rho = -0.7$). Shown are the best known Pareto front approximation, the offspring at some selected generations, the evolution of the parent, and the lines of equal function values. Top left: $\varepsilon = 0$, $\delta = \frac{3}{10} \cdot \frac{\pi}{2}$. Top right: $\varepsilon = 1$, $\delta = \frac{3}{10} \cdot \frac{\pi}{2}$. Bottom left: $\varepsilon = 0.6$, $\delta = \frac{7}{10} \cdot \frac{\pi}{2}$. Bottom right: $\varepsilon = 0.6$, $\delta = \frac{3}{10} \cdot \frac{\pi}{2}$. All plots are for ρMNK-landscapes with a correlation of $\rho = -0.7$ except for the bottom right plot where $\rho = 0.0$.

is the average objective vector for a random solution of ρMNK-landscapes. The evolution of the current solution can be explained by the combination of two effects. The first one is given by the independent bit-flip mutation operator, that produces more offspring in a particular direction compared to the other ones, due to the underlying characteristics of the ρMNK-landscape under consideration. The second one is given by the lines of equal function values, i.e., the current solution moves perpendicular to the iso-fitness lines, following the gradient direction in the objective space. We can remark that the search process is mainly guided by the lower part of the cones of equal function values when the direction is above the initial solution, and *vice versa*: When the direction angle δ is smaller (resp. larger) than $\pi/4$, the dynamics of the search process is better captured by the opening angle θ_1 (resp. θ_2), defined between the equi-fitness lines and the f_1-axis. Geometrically, the optimal solution with respect to a scalarizing function should correspond to the intersection of one of the 'highest' lines of equal fitness values in the gradient direction and the feasible region of the objective space. Although the above description is mainly intuitive, a deeper and more fundamental analysis can support this general idea.

Let us focus, without loss of generality, on the influence of the opening angle θ_1 when the

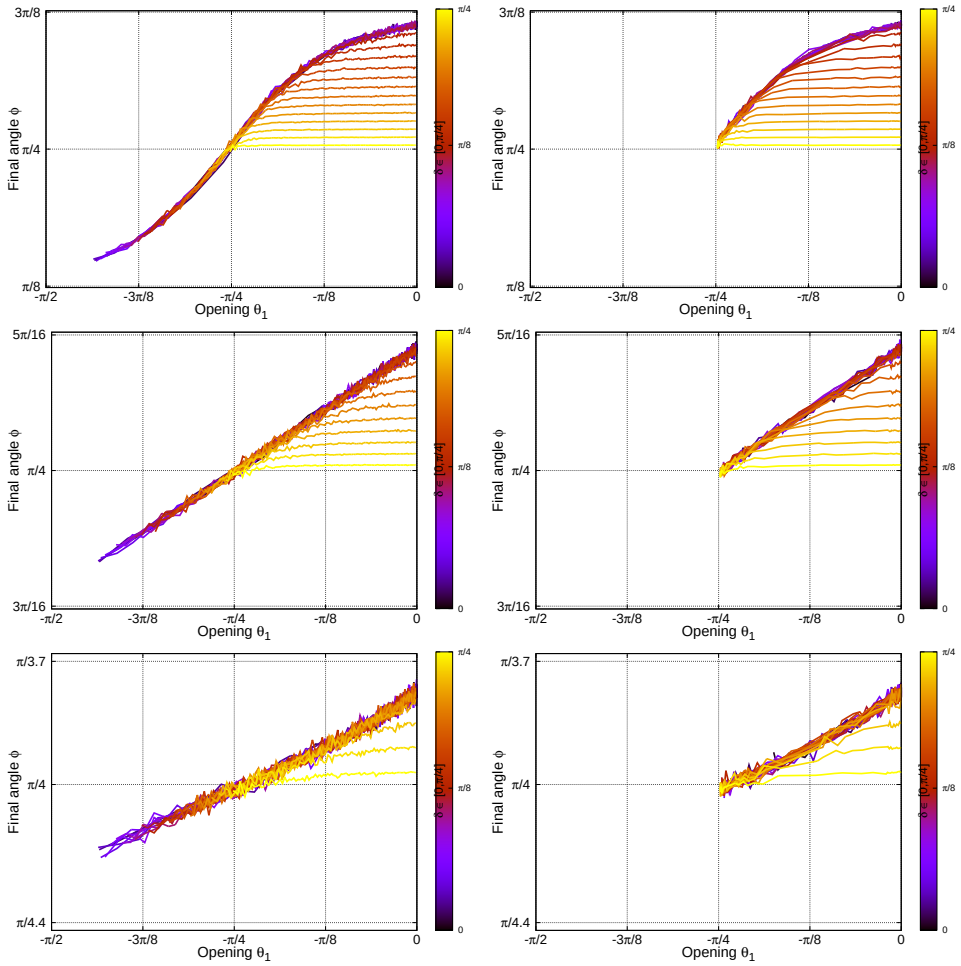


Figure 5: Scatter plots showing final angle $\phi(\varepsilon)$ and opening $\theta_1(\varepsilon)$ for $\rho = -0.7$ (top) $\rho = 0.0$ (middle) and $\rho = 0.7$ (bottom) and S_{norm} (left column) and S_{aug} (right column). Every color is for a fixed δ and variable ε.

direction angle δ is smaller than $\pi/4$ here. Similar results hold for $\delta > \pi/4$ and θ_2. Fig. 5 shows the scatter plots of the final angle ϕ as a function of the opening angle θ_1, for different objective correlation coefficients ρ and for different direction angles $\delta \in [0, \pi/4]$. A scatter plot gives a set of values $(\theta_1(\varepsilon), \phi(\varepsilon))$ for the ε-values under study. From Theorem 1, for a given direction angle δ, the opening angle θ_1 belongs to the interval $[\delta - \pi/2, 0]$ for S_{norm}, and to the interval $[-\pi/4, 0]$ for S_{aug}. Independently of the scalarizing function, when the direction angle is between 0 and around $3\pi/16$ (blue color), the value of ϕ is highly correlated with the opening angle θ_1. For such directions, a simple linear regression confirms this observation and allows us to explain the relation between the opening angle and the final angle by means of the following approximate equation: $\phi \approx (c + \pi/4) + c \cdot \theta_1$, such that c is respectively equal to 0.05, 0.2, and 0.4 for $\rho = -0.7$, 0, and 0.7. We emphasize that this is independent of the definition of the scalarizing function, and depends mainly on the property of the lines of equal function values. The previous equation tells us that the lines of equal fitness values are guiding the search process following the gradient direction given by the opening angle in the objective space. Figure 6 shows that the obtained

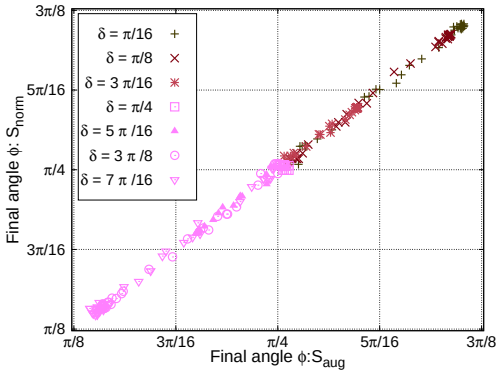


Figure 6: Scatter plot showing final angles obtained with S_{aug} (x-axis) against S_{norm} (y-axis) for some fixed direction angle δ and different ε providing the closest openings for the two scalarizing functions. Results are for $\rho = -0.7$.

final angles are equivalent when the opening angle is the same, even for different direction angles and/or scalarizing functions. In fact, we observe that the final angles obtained are very similar for the scalarizing functions S_{norm} and S_{aug} if δ is the same for both functions and the ε-values are chosen in order to have matching opening angles. Whatever the δ- and ε-values, the points in the plot of Fig. 6 are close to the line $y = x$, which shows that independently of the scalarizing function, the final angle is strongly correlated to the opening angle, and not to a particular scalarizing function. Also, the opening of the lines of equal function values have more impact on the dynamics of the search process than the direction angle alone. In this respect, the opening angle should be considered as a key feature to describe and understand the behavior of scalarizing search algorithms.

5 Global Search Behavior

In the previous section, we considered every single $(1+\lambda)$-EA separately. However, the goal of a general-purpose decomposition-based algorithm is to compute a set of solutions approximating the whole Pareto front. In this section, we study the quality of the set obtained when combining the solutions computed by different configurations of the scalarizing functions. A natural way to do so, is to use the same ε-value for all direction angles. Fig. 7 illustrates the relative performance, in terms of hypervolume difference and multiplicative epsilon indicators [12], when considering such a setting.

The hypervolume difference indicator gives the difference between the portion of the objective space that is dominated by the approximation found by the global decomposition-based approach and some reference set. The reference point is set to the origin, and the reference set is the best-known approximation for the instance under consideration. The epsilon indicator gives the minimum multiplicative factor by which the approximation found by the global scalarizing algorithm has to be translated in the objective space to weakly dominate the reference set. Notice that both indicator-values are to be minimized.

To understand the results of Fig. 7, we once again should refer to Fig. 1 eliciting the actual search direction of every $(1+\lambda)$-EA separately. We remark that the ε-values minimizing both indicator-values correspond to those that allow to distribute the final angles fairly among direction angles. In fact, the quality of the obtained approximation set depends on the diversity of

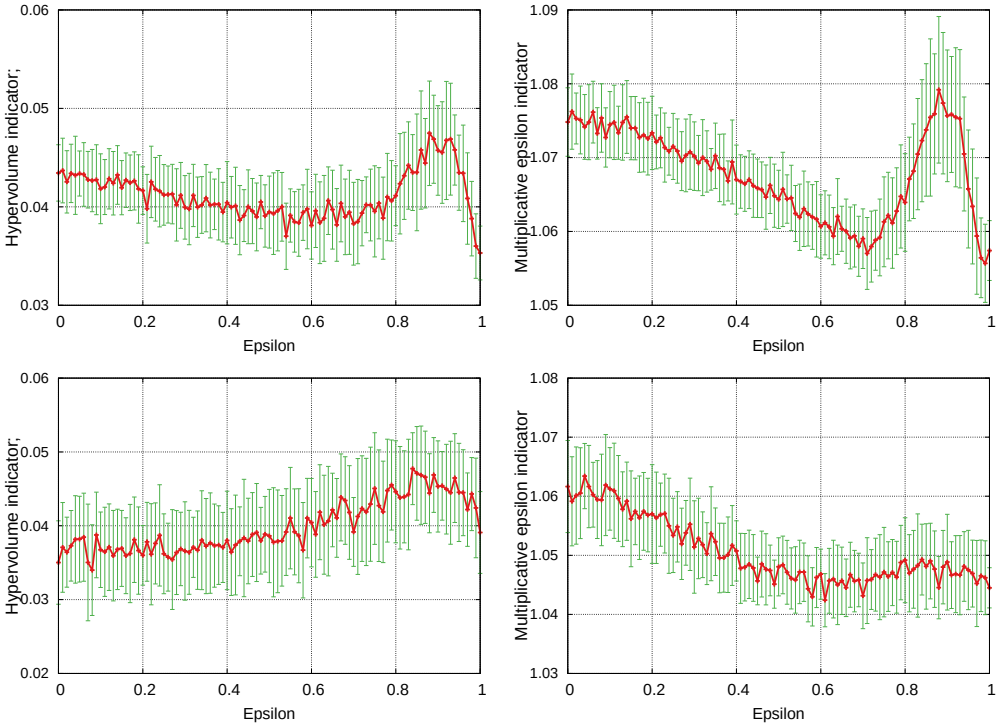


Figure 7: Column 1 and 3 (resp. 2 and 4) depict the hypervolume (resp. epsilon) indicators for scalar function S_{norm}. Top (resp. bottom): objective correlations $\rho = -0.7$ (resp. $\rho = 0.7$).

combined solutions; and we can consistently see that the better the distribution of final angles, the better the performance of the global algorithm. In fact, when the ε-value does not produce diverse solutions, the multiobjective search is only focused on restricted number of regions of the objective space. For those regions, the aggregated set is more likely to contain more solutions which approach better the front since several scalarizing functions are actually optimizing solutions in the same direction. The simplest example to illustrate this is when we compare $\varepsilon = 0$ (T) to $\varepsilon = 1$ (WS). From Fig. 1, there are substantially more (resp. less) direction angles leading to the extremes (resp. middle) of the Pareto front when using T compared to WS. Thus, it is more likely to better approach the lexicographically optimal regions, rather than the middle, of the Pareto front when using T; and *vice-versa*. For landscapes with positive objective correlation, this effect gets more pronounced because the Pareto front has a smaller extend. This also explains why in Fig. 7, we see that for positive objective correlation, the performance of T compared to WS is conflicting depending on whether the hypervolume and the epsilon indicator is considered.

To illustrate this further, we show in Fig. 8 the empirical attainment function differences (EAFs, [8])[4] comparing pure WS and pure T for a ρMNK-landscape with positive objective

[4]The EAF provides the empirical probability that an arbitrary objective vector is dominated by, or equivalent to, a solution obtained by a single run of the algorithm. The difference between the EAFs for two different algorithms enables to identify the regions of the objective space where one algorithm performs better than another. The magnitude of the difference in favor of one algorithm is plotted within a gray-colored graduation. Plots are generated by the R package of Carlos Fonseca et al. which can be downloaded from

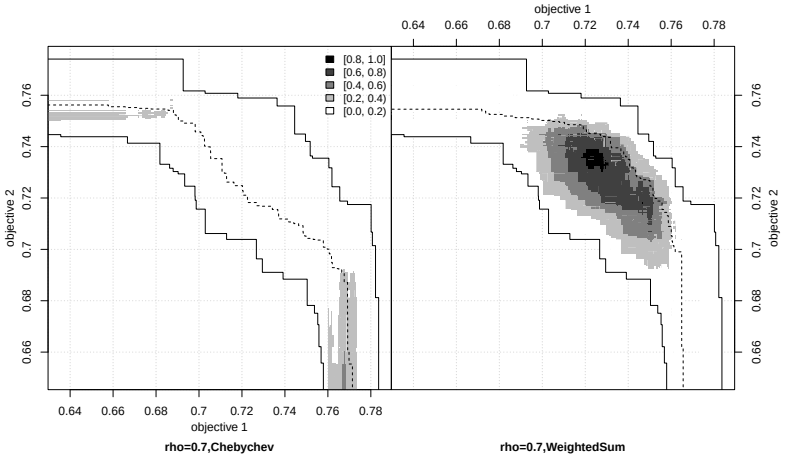


Figure 8: T vs. WS, positive objective correlation 0.7.

correlation. Notice from Fig. 1, and because the Pareto front's extent is small for a positive objective correlation, there are substantially more (resp. less) direction angles leading to the extremes (resp. middle) of the Pareto front when using T compared to WS. Thus, it is more likely to better approach the lexicographically optimal regions, rather than the middle, of the Pareto front when using T; and *vice-versa*.

Moreover, WS is found to be in general competitive with respect to other fixed ε-values. This observation might suggest that WS is the best-performing parameter setting, since every different direction angle leads to a different final angle. Nevertheless, the diversity of final angles is not the only criterion that can explain quality. The efficiency of the $(1+\lambda)$-EA with respect to the single-objective problem implied by the scalarizing function is also crucial. In Fig. 3 (right), we observed that the ε-value exhibiting the minimal average deviation to best is not necessarily the same for every direction. We also observed that for direction angles in the middle of the weight space, the final angles obtained for different ε-values can end up being very similar. Thus, it might be possible that, by choosing different ε-values for different directions, one can find a configuration for which final solutions are diverse, but also closest to the Pareto front. Indeed, we can observe a significant difference between the non-uniform case where the scalarizing function S_{norm} (or S_{aug}) is configured with an ε providing the best deviation to best for every direction, and the situation where ε is the same for all directions. As shown in Table 3, such non-uniform configurations are both substantially better than T and also competitive compared to WS.

In Fig. 9, we furthermore illustrate the empirical attainment function (EAF) [8] differences. obtained when comparing WS and T, to a non-uniform configuration where the ε-value for each direction is not the same anymore, but extracted from Fig. 3 (Right). We can see that this non-uniform configuration is both substantially better than T and also competitive compared to WS. Let us remark that we only show the performance of the above non-uniform configuration in order to illustrate how choosing different ε-values can improve the quality of the resulting approximation set. However, this particular non-uniform configuration might not be 'optimal'. In fact, in order to obtain the results of Fig. 3 (Right), we fixed every weight vector and computed the ε-value that provides the best relative deviation. It might be the case that a different weight vector provides even 'better' solutions in the same direction, since we showed that different

`http://cran.r-project.org/web/packages/eaf/`.

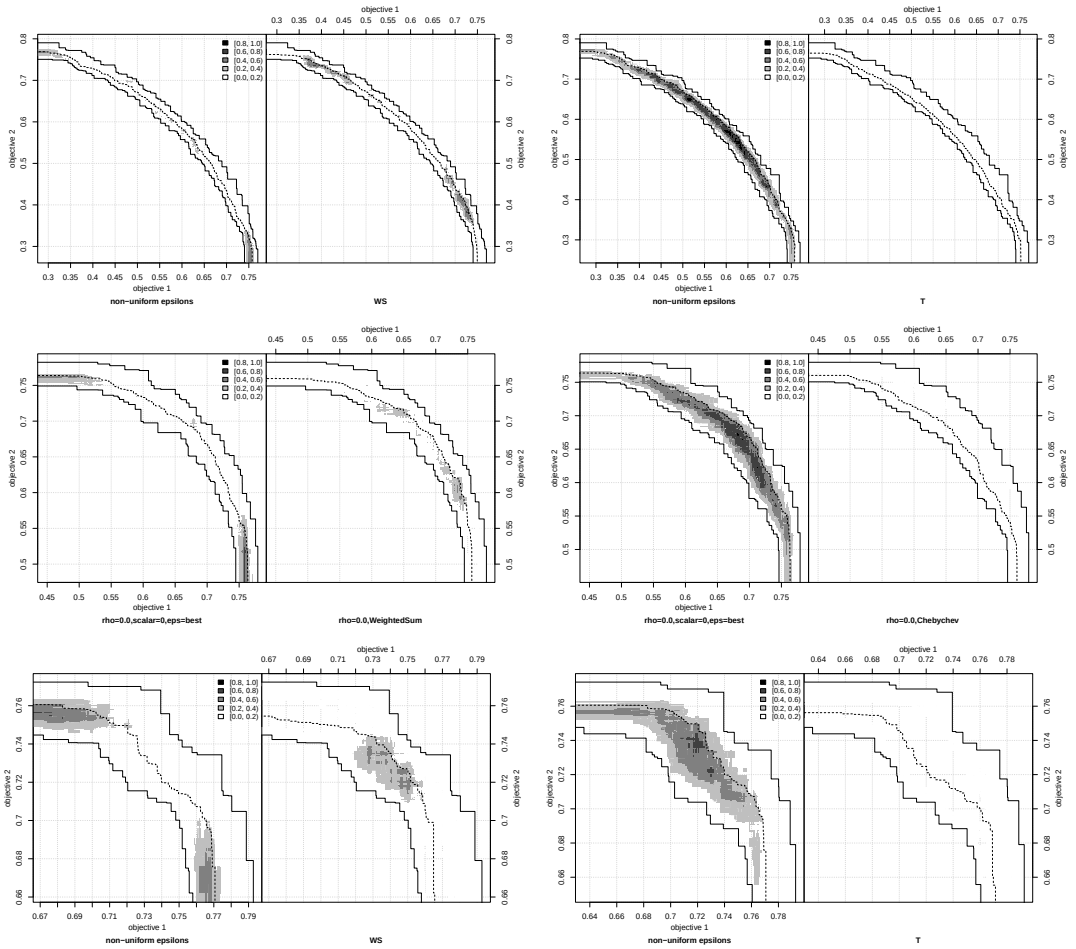


Figure 9: EAF differences: mixed configuration of Fig. 3 *vs.* WS (left) and T (right). From top to bottom: $\rho \in \{-0.7, 0, 0.7\}$.

directions with different ε-values can induce a similar distribution of final angles. In other words, finding the 'best' parameter configuration in a setting where μ independent single $(1+\lambda)$-EAs are considered, can be itself formulated as an optimization problem with variables ε and δ; such that direction angles in the optimal configuration might not be necessarily pair-wisely different.

6 Open(ing) (Re)search Lines

We presented an extensive empirical study that sheds more light on the impact of scalarizing functions within decomposition-based evolutionary multiobjective optimization. Our results showed that, given a weighting coefficient vector and a relative importance of the weighted sum and the Tchebycheff term in the function, it is fundamentally the opening of the lines of equal function values that explicitly guides the search towards a specific region of the objective space. When combining multiple scalarizing search processes to compute a whole approximation set,

Table 3: Comparison of WS, T, and non-uniform $\mathsf{S}^{\varepsilon^\star}_{\text{norm}}$ and $\mathsf{S}^{\varepsilon^\star}_{\text{aug}}$ configured with ε-values giving the best deviation w.r.t every direction. The number in braces shows the number of other algorithms that statistically outperform the algorithm under consideration w.r.t. a given indicator and a Mann-Whitney signed-rank statistical test with a p-value of 0.05 (the lower, the better).

	Avg. hypervolume difference ($\times 10^{-1}$)				Avg. multiplicative epsilon			
ρ	WS	T	$\mathsf{S}^{\varepsilon^\star}_{\text{norm}}$	$\mathsf{S}^{\varepsilon^\star}_{\text{aug}}$	WS	T	$\mathsf{S}^{\varepsilon^\star}_{\text{norm}}$	$\mathsf{S}^{\varepsilon^\star}_{\text{aug}}$
-0.7	0.353 (2)	0.434 (3)	0.324 (**0**)	0.307 (**0**)	1.057 (**0**)	1.075 (3)	1.059 (**0**)	1.057 (**0**)
0.0	0.418 (2)	0.458 (3)	0.357 (1)	0.322 (**0**)	1.056 (**0**)	1.084 (3)	1.062 (1)	1.064 (1)
0.7	0.391 (3)	0.350 (2)	0.303 (**0**)	0.292 (**0**)	1.044 (**0**)	1.062 (3)	1.047 (1)	1.047 (1)

these lines play a crucial role to achieve diversity. While our results are with respect to a rather simple setting where multiple scalarizing search procedures are run independently, they make a fundamental step towards strengthening the understanding of the properties and dynamics of more complex algorithmic settings. It is our hope that the lessons learnt from our study can highly serve to better tackle the challenges of decomposition-based approaches. They also rise new interesting issues that were hidden by the complex design of well-established algorithms. In the following, we identify a non-exhaustive number of promising research directions that relate directly to our findings.

❶ **Improving existing algorithms.** Eliciting the best configuration to tackle a multiobjective optimization problem by decomposition can highly improve search performance. As we demonstrated, similar regions can be achieved using different parameter settings, and the performance could be enhanced by adopting non-uniform configurations. One research direction would be to investigate how such *non-uniform* configurations perform when plugged into existing approaches. To our best knowledge, there exists no attempt in this direction, and previous investigations did only consider uniform parameters, which do not necessarily guarantee to reach an optimal performance.

❷ **Tuning the opening angles.** Generally speaking, the parameters of existing scalarizing functions can simply be viewed as one specific tool to set up the openings of the lines of equal function values. In this respect, one can consider other types of opening angles, without necessarily using a particular scalarizing function. This could lead to more flexibility when tuning decomposition-based algorithms, by offering new configuration possibilities, e.g., defining the opening angles under and below the line crossing the utopian point independently, without being bound to a closed-form relationship, but with respect to the current search state. Notice that tuning the opening angles is a challenging issue which was not studied *per-se*. We believe that classical paradigms for on-line and off-line parameter tuning are worth to be investigated in order to accurately tackle this issue, because of the number and the complexity of possible configurations.

❸ **Variation operators and problem-specific issues.** Our results show that, not only the line of equal function values impact the search process, but also the properties of the mechanisms allowing to produce improving solutions in every step of the scalarizing search procedure. In our experimental study, we consider the independent one-bit-flip mutation operator and ρMNK-landscapes. Preliminary experiments, not shown in this paper for readability reasons, comfort us in the validity of our conclusions for a large spectrum of instances and operators. However, this issue should be throughly investigated in future work at the aim of gaining in generality and to better understand the impact of other search components or problem types in the performance of scalarizing search approaches. Equally important is to investigate problems with more than two objectives.

❹ **Theoretical modeling.** Last but not least, it would be nice to provide theoretical results that fit our empirical findings. The goal would be to provide a theoretical framework, abstracting away the problem-specific issues, and allowing us to reason about decomposition-based approaches in a purely theoretical manner. Such a framework would enable to better harness decomposition-based approaches by means of scalarizing functions, and to derive new methodological tools. Although this is a challenging research direction, it can be necessary in order to improve our practice of decomposition-based evolutionary multiobjective optimization approaches for real-world multiobjective optimization.

7 Acknowledgments

This report is based on investigations within the project "Global Research on the Framework of Evolutionary Solution Search to Accelerate Innovation" as part of the "Strategic Young Researcher Overseas Visits Program for Accelerating Brain Circulation" which is kindly supported by the Japan Society for the Promotion of Science (JSPS). In addition, the first three authors would like to acknowledge support by the French national research agency (ANR) within the Modèles Numérique project "NumBBO - Analysis, Improvement and Evaluation of Numerical Blackbox Optimizers" (ANR-12-MONU-0009-03).

References

[1] D. Brockhoff, T. Wagner, and H. Trautmann. On the Properties of the $R2$ Indicator. In *Genetic and Evolutionary Computation Conference (GECCO 2012)*, pages 465–472, 2012.

[2] K. Dächert, J. Gorski, and K. Klamroth. An Augmented Weighted Tchebycheff Method With Adaptively Chosen Parameters for Discrete Bicriteria Optimization Problems. *Computers & Operations Research*, 39(12):2929–2943, 2012.

[3] I. Giagkiozis, R. C. Purshouse, and P. J. Fleming. Generalized Decomposition. In *EMO*, pages 428–442, 2013.

[4] E. J. Hughes. Multiple Single Objective Pareto Sampling. In *CEC*, pages 2678–2684, 2003.

[5] H. Ishibuchi, Y. Sakane, N. Tsukamoto, and Y. Nojima. Adaptation of scalarizing functions in MOEA/D: An adaptive scalarizing function-based multiobjective evolutionary algorithm. In *EMO*, pages 438–452, 2009.

[6] Hisao Ishibuchi, Naoya Akedo, and Yusuke Nojima. A study on the specification of a scalarizing function in MOEA/D for many-objective knapsack problems. In *LION7*, pages 231–246, 2013.

[7] Ignacy Kaliszewski. Using trade-off information in decision-making algorithms. *Computers & Operations Research*, 27(2):161 – 182, 2000.

[8] M. López-Ibáñez, L. Paquete, and T. Stützle. Exploratory Analysis of Stochastic Local Search Algorithms in Biobjective Optimization. In *Experimental Methods for the Analysis of Optimization Algorithms*, pages 209–2033. 2010.

[9] K. Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer, Boston, MA, USA, 1999.

[10] Sébastien Verel, Arnaud Liefooghe, Laetitia Jourdan, and Clarisse Dhaenens. On the structure of multiobjective combinatorial search space: MNK-landscapes with correlated objectives. *Eur J Oper Res*, 227(2):331–342, 2013.

[11] Q. Zhang and H. Li. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE TEC*, 11(6):712–731, 2007.

[12] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. Grunert da Fonseca. Performance Assessment of Multiobjective Optimizers: An Analysis and Review. *IEEE TEC*, 7(2):117–132, 2003.

Appendix

Theorem 2 *Given the scalarizing function*

$$\mathsf{S}_{gen}(z) = \alpha \cdot \max\{\lambda_1|\bar{z}_1 - z_1|, \lambda_2|\bar{z}_2 - z_2|\} + \varepsilon\,(w_1|\bar{z}_1 - z_1| + w_2|\bar{z}_2 - z_2|)$$

of Eq. 2 where $\bar{z} = (\bar{z}_1, \bar{z}_2)$ *is a utopian vector,* $\lambda_1, \lambda_2, w_1,$ *and* $w_2 > 0$, $\alpha \geq 0$, $\varepsilon > 0$. *The lines of equal function values going through an objective vector* $z^1 = (z_1^1, z_2^1)$ *are given by*

$$y = -\frac{\varepsilon w_1}{\alpha\lambda_2 + \varepsilon w_2} \cdot x + \frac{\alpha\lambda_2 z_2^1 + \varepsilon\left(w_1 z_1^1 + w_2 z_2^1\right)}{\alpha\lambda_2 + \varepsilon w_2}$$

for $z = (x, y)$ below *the line* D *through* $\bar{z}$ *and in direction* $(1/\lambda_1, 1/\lambda_2)$ *and by*

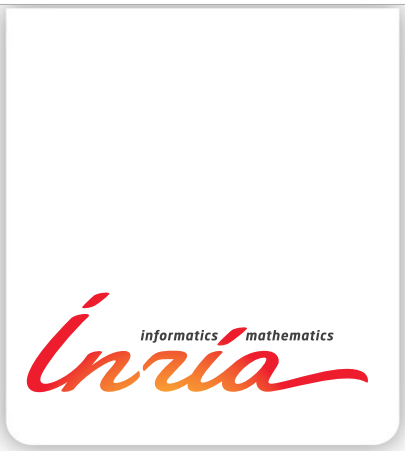
$$y = -\frac{\alpha\lambda_1 + \varepsilon w_1}{\varepsilon w_2} \cdot x + \frac{\alpha\left(\lambda_2 z_2^1 + \lambda_1\bar{z}_1 - \lambda_2\bar{z}_2\right) + \varepsilon\left(w_1 z_1^1 + w_2 z_2^1\right)}{\varepsilon w_2}$$

for $z = (x, y)$ above *the same line* D. *In case that* $\varepsilon = 0$, *the lines of equal function values are parallel to the coordinate axes of the objective space.*

Proof: Let us fix the point $z^1 = (z_1^1, z_2^1)$ and take a point $z = (x, y)$ such that z is below the line D. Then we can solve $\mathsf{S}_{\text{gen}}(z) = \mathsf{S}_{\text{gen}}(z^1)$ in order to get the lines of equal utility with respect to z^1. By the definition of $\mathsf{T}(z)$ and the fact that z lies below D, we get $\mathsf{T}(z) = \lambda_2 \cdot |\bar{z}_2 - y|$ and thus

$$\begin{aligned}
&\mathsf{S}_{\text{gen}}(z) = \mathsf{S}_{\text{gen}}(z^1)\\
\Leftrightarrow\ &\alpha\lambda_2\left(\bar{z}_2 - y\right) + \varepsilon\left(w_1\left(\bar{z}_1 - x\right) + w_2\left(\bar{z}_2 - y\right)\right) = \alpha\lambda_2\left(\bar{z}_2 - z_2^1\right) + \varepsilon\left(w_1\left(\bar{z}_1 - z_1^1\right) + w_2\left(\bar{z}_2 - z_2^1\right)\right)\\
&\Leftrightarrow y = -\frac{\varepsilon w_1}{\alpha\lambda_2 + \varepsilon w_2} \cdot x + \frac{\alpha\lambda_2 z_2^1 + \varepsilon\left(w_1 z_1^1 + w_2 z_2^1\right)}{\alpha\lambda_2 + \varepsilon w_2}
\end{aligned}$$

which proves the first part of the theorem. The second part follows analogously for a point $z = (x, y)$ above D. Similarly, the case where $\varepsilon = 0$ can be derived by standard calculus. Theorem 1 follows from the above by simple geometrical considerations and the slope of the lines. □



RESEARCH CENTRE
LILLE – NORD EUROPE

Parc scientifique de la Haute-Borne
40 avenue Halley - Bât A - Park Plaza
59650 Villeneuve d'Ascq

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399