Evaluating the quality of approximations to the non-dominated set

Michael Pilegaard Hansen
Institute of Mathematical Modelling
Technical University of Denmark
Building 321
2800 Lyngby
Denmark
Email: mph@imm.dtu.dk

and

Andrzej Jaszkiewicz
Institute of Computing Science
Poznan University of Technology
ul. Piotrowo 3a, 60-965 Poznan
Poland
Email: Jaszkiewicz@cs.put.poznan.pl

March 22, 1998

IMM Technical Report IMM-REP-1998-7

Evaluating the quality of approximations to the non-dominated set

Michael Pilegaard Hansen

Institute of Mathematical Modelling, Technical University of Denmark, 2800 Lyngby, Denmark, email: mph@imm.dtu.dk

Andrzej Jaszkiewicz

Institute of Computing Science, Poznan University of Technology, ul. Piotrowo 3a, 60-965 Poznan, Poland, email: Jaszkiewicz@cs.put.poznan.pl

Abstract: The growing interest in hard multiple objective combinatorial and non-linear problems resulted in a significant number of heuristic methods aiming at generating sets of feasible solutions as approximations to the set of non-dominated solutions. The issue of evaluating these approximations is addressed. Such evaluations are useful when performing experimental comparisons of different multiple objective heuristic algorithms, when defining stopping rules of multiple objective heuristic algorithms, and when adjusting parameters of heuristic algorithms to a given problem. A family of outperformance relations that can be used to compare approximations under very weak assumptions about a decision-maker's preferences is introduced. These outperformance relations define incomplete orders in the set of all approximations. It is shown that in order to compare approximations, which are incomparable according to the outperformance relations, much stronger assumptions about the decision-maker's preferences are necessary. A general framework that can be used to compare and evaluate approximations under the presence of various types of additional information is proposed. Some particular comparison and evaluation methods based on this framework are suggested. The proposed framework is also used to characterize some previously proposed evaluation methods.

Keywords: Multiple objective optimization; Heuristics; Evaluation

Introduction

Research on multiple objective optimization was traditionally concentrated on linear programming problems (see e.g. Steuer, 1986 and White, 1990). Efficient solutions of such problems are in most cases generated by solving some single objective optimization problems, e.g. by optimizing a scalarizing function on the original set of feasible solutions. As very effective methods and software for single objective linear optimization exist, most multiple objective linear programming problems can be effectively solved in this way, except of some very large scale problems.

In recent years, however, researchers have become increasingly interested in other classes of multiple objective problems, for instance multiple objective combinatorial problems (cf. Ulungu and Teghem, 1994). This interest is raised by practical applications, e.g. in project scheduling (see e.g. Slowinski, 1989), vehicle routing (see e.g. Assad, 1988) and engineering design (Dasgupta and Michalewicz 1997). For example, solutions to vehicle routing problems are usually evaluated by e.g. total cost, distance, travel time and the number of vehicles used. In practice, it can therefore be difficult to evaluate a solution to such a problem with only a single objective. The objectives, however, are traditionally used separately or they are combined into a single objective.

Multiple objective problems are often hard even in the single objective case. For example, most vehicle routing problems are extensions of travelling salesperson problem (TSP) which is already NP-hard. It is also worth mentioning that some combinatorial problems, which are easy in single objective case, turn hard when multiple objectives are considered. For example, the single objective shortest path problem is one of the simplest combinatorial problems while the corresponding multiple objective problem is NP-hard (Serafini, 1987).

For hard multiple objective problems it may be necessary to settle for approximations to the efficient solutions. As single objective metaheuristic procedures, e.g. simulated annealing (Cerny, 1982, Kirkpatrick et al., 1983 and Laarhoven and Aarts, 1987), tabu search (Glover, 1989) and genetic algorithms (Goldberg, 1988), often are successful in the single objective optimization problems it seems natural to use them in the case of multiple objective optimization.

Several authors have proposed multiple objective metaheuristic procedures that aim at the effective generation of approximations of the non-dominated set. The methods are based on ideas of genetic algorithms (Schaffer, 1985, Fonseca and Fleming, 1993, Horn, Nafpliotis and Goldberg, 1994, Srinivas and Deb, 1995; see also Fonseca and Fleming, 1995, for a review), simulated annealing (Serafini, 1994, Ulungu et al., 1994 and Czyzak and Jaszkiewicz, 1995) or tabu search (Gandibleux et. al., 1996, and Hansen, 1997). Authors of such methods usually state that the methods should generate “good approximations” of the non-dominated set. The term “good approximation”, however, is often only defined intuitively, as being close to the real non-dominated set and well-dispersed over this set.

The main purpose of this paper is a more rigid definition of the goal of multiple objective heuristic algorithms. More precisely, we propose some tools that may be used to evaluate and compare approximations and define what is understood by “a good approximation” to the set of non-dominated solutions.

Please note that the issue of measurement is simple if a single objective is considered. In that case, we evaluate and compare approximate solutions using the obvious quality measure, which is the value of the objective function. No such natural measure exists in multiple objective case.

Evaluations of approximation of the non-dominated set may be used to:

- support experimental evaluation of different multiple objective heuristic algorithms,
- adjust parameters of metaheuristic algorithms for a given problem and
- define stopping rules of metaheuristic algorithms.

The possible functions are shortly characterized below.

The number of different multiple objective metaheuristics proposed in recent years is significantly increasing. It is obvious that they are of different quality, i.e. some of them generate better approximations in a shorter time than others. A natural way to evaluate the methods is to perform comparative experiments. At present, however, different authors use different approaches to evaluate quality of obtained approximations. This makes it practically impossible to perform a fair comparison of different multiple objective heuristics.

Both single and multiple objective metaheuristic procedures involve many parameters that have to be adjusted for a given class of problems. Settings of the parameters may have crucial influence on the quality of the algorithm applied to a given class of problems. Although some general guidelines exist, adjustment of the parameters usually requires some experiments with different settings and evaluations of the results. In the multiple objective case, it requires evaluating and comparing approximations obtained with different settings of the parameters.

Stopping rules of single objective metaheuristic algorithms are often based on observations of the objective function value (see e.g. Laarhoven and Aarts, 1987, and Goldberg, 1988). A procedure may for example be stopped if in a given number of iterations the improvement of the objective function value is below a given threshold. In the multiple objective case, such stopping rules require an evaluation of the quality of the current approximation of the non-dominated set.

We believe also that clear understanding of the goal of multiple objective heuristic algorithms is a necessary preliminary step towards any kind of theoretical analysis of such algorithms.

The paper is organized in the following way. In the next Section some basic definitions are given. In the second Section, we introduce outperformance relations for comparing pairs of approximations under very weak assumptions about the decision-maker's (DM's) preferences. Quantitative comparison methods using probability distribution of the DM's possible preferences are presented in the third Section. In the fourth Section, we characterize types of additional information that may be necessary in quantitative comparison methods. Practical guidelines for constructing and computing quantitative measures are presented in the fifth Section. In the sixth Section we comment on some previously used approaches for evaluation and comparison of approximations to the non-dominated set. The issue of evaluating approximations under presence of more precise preference information is discussed in the seventh Section. In the eighth Section, we outline the possibilities of using the preference information within multiple objective heuristics. Conclusions and directions for further research are summarized in the last Section.

1. Basic definitions

The general multiple objective optimization (MOO) problem is formulated as:

$$\text{maximize}\{f_1(\mathbf{x}) = z_1, \ldots, f_J(\mathbf{x}) = z_J\} \qquad \text{(P1)}$$

$$\text{s.t.} \quad \mathbf{x} \in D,$$

where: *solution* $\mathbf{x} = [x_1, \ldots, x_I]$ is a vector of *decision variables* and D is the set of feasible solutions. The type of the variables, constraints and objective functions may then describe different classes of problems. If the decision variables are continues, one may have a multiobjective linear program (MOLP) or a multiobjective non-linear program (MONLP) depending on the linearity of the constrains and the objective functions. If the variables are integers, (P1) turns into a multiobjective integer program (MOIP), which again may be both linear and non-linear. Multiobjective combinatorial optimization (MOCO) problems can often be formulated as linear multiobjective integer programs.

The image of a solution $\mathbf{x}$ in the objective space is a *point* $\mathbf{z}^{\mathbf{x}} = [z_1^{\mathbf{x}}, \ldots, z_J^{\mathbf{x}}]$, such that $z_j^{\mathbf{x}} = f_j(\mathbf{x})$, $j=1,..,J$. The image of the set D in the criterion space is a set Z composed of *attainable* points, i.e. points being images of feasible solutions.

Problem (P1) can also be formulated more succinctly as:

$$\text{maximize}\{\mathbf{z}\} \qquad \text{(P2)}$$

$$\text{s.t.} \quad \mathbf{z} \in Z$$

where $\mathbf{z} = [z_1, \ldots, z_J]$ is a vector of objective functions $z_j = f_j(\mathbf{x})$, $j=1,..,J$.

Point $\mathbf{z} \in Z$ *dominates* $\mathbf{z}' \in Z$, $\mathbf{z} \succ \mathbf{z}'$, if $z_j \geq z_j'$ $\forall j$, and $z_i > z_i'$ for at least one i.

Point $\mathbf{z}' \in Z$ is *non-dominated* if there is no $\mathbf{z} \in Z$ that dominates $\mathbf{z}'$. The set N of all non-dominated points is called the *non-dominated set*. A solution $\mathbf{x}$ is efficient if its image in the objective space is non-dominated.

By $ND(S)$, where S is a set of (feasible or infeasible) points in the objective space, we will denote a *set of points non-dominated within the set S*.

The point $\mathbf{z}^*$ composed of the best attainable objective function values is called the *ideal point*:

$$z_j^* = \max\{z_j \mid \mathbf{z} \in Z\} \qquad j = 1, \ldots, J.$$

The point $\mathbf{z}_{**}$ composed of the worst attainable objective function values is called the *anti-ideal point*:

$$z_{**j} = \min\{z_j \mid \mathbf{z} \in Z\} \qquad j = 1, \ldots, J.$$

The point $\mathbf{z}_*$ composed of the worst objective function values in the non-dominated set is called the *nadir point*.

Range equalization factors (Steuer, 1986, sec. 8.4.2) are defined in the following way:

$$\pi_j = \frac{1}{R_j}, j=1, \dots, J \tag{1}$$

where R_j is the range of objective j in the set N or D. Objective function values multiplied by range equalization factors are called *normalized objective function values*.

A *utility function* $u:\Re^J \to \Re$, is a model of the DM's preferences that maps each point in the objective space into a value of utility. It is assumed that the goal of the DM is to maximize the utility.

A utility function u is *compatible with the dominance relation* if and only if $\forall \mathbf{z}^1, \mathbf{z}^2 \in \Re^J \quad \mathbf{z}^1 \succ \mathbf{z}^2 \Rightarrow u(\mathbf{z}^1) \geq u(\mathbf{z}^2)$. The set of all utility functions that are compatible with the dominance relation is denoted by U_c.

A utility function U is *strictly compatible with the dominance relation* if and only if $\forall \mathbf{z}^1, \mathbf{z}^2 \in \Re^J \quad \mathbf{z}^1 \succ \mathbf{z}^2 \Rightarrow u(\mathbf{z}^1) > u(\mathbf{z}^2)$. The set of all utility functions that are strictly compatible with the dominance relation is denoted by U_{sc}.

A convenient way to define a set of utility functions is by the use of parametric utility functions $u(\mathbf{z}, \mathbf{r})$, $\mathbf{r} \in D(\mathbf{r}) \subseteq \Re^n$, where $\mathbf{r}$ is a vector of parameters and $D(\mathbf{r})$ is domain of the parameter vector. A *parametric set of utility functions* is then defined in the following way:

$$U(\mathbf{r}) = \{u(\mathbf{z}, \mathbf{r}) \mid \mathbf{r} \in D(\mathbf{r})\}.$$

Weighted L_p norms are defined as:

$$L_p(\mathbf{z}^1, \mathbf{z}^2, \Lambda) = \left(\sum_{j=1}^{J} \lambda_i \left| z_j^1 - z_j^2 \right|^p \right)^{1/p}, \; p \in \{1, 2, \ldots\} + \{\infty\},$$

where $\Lambda = [\lambda_1, \ldots, \lambda_J]$, $\lambda_j \geq 0$, is a weight vector.

A set U_p of *utility functions based on weighted L_p norms* is a parametric set composed of functions of the following form:

$$u_p(\mathbf{z}, \mathbf{z}^*, \Lambda, p) = -\left(\sum_{j=1}^{J} \lambda_i \left(z_j^* - z_j\right)^p \right)^{1/p}, \; p \in \{1, 2, \ldots\} + \{\infty\}.$$

For $p=\infty$ one obtains a parametric set U_∞ of *weighted Tchebycheff utility functions*:

$$u_\infty(\mathbf{z}, \mathbf{z}^*, \Lambda) = -\max_j \{\lambda_i (z_j^* - z_j)\}. \tag{2}$$

For $p=1$ one obtains a parametric set U_1 of *weighted linear utility functions*:

$$u_p(\mathbf{z}, \mathbf{z}^*, \Lambda) = -\left(\sum_{j=1}^{J} \lambda_i \left(z_j^* - z_j\right) \right) = -\sum_{j=1}^{J} \lambda_i z_j^* + \sum_{j=1}^{J} \lambda_i z_j = \mathrm{const} + \sum_{j=1}^{J} \lambda_i z_j. \tag{3}$$

A finite set $A \subseteq Z$ is called an *approximation* of set N if $\forall \mathbf{z}^1, \mathbf{z}^2 \in A \quad \mathbf{z}^1 \not\succ \mathbf{z}^2 \wedge \mathbf{z}^2 \not\succ \mathbf{z}^1$, i.e. if it is composed of mutually non-dominated attainable points. The set of all approximations of the set N is denoted by Ω.

The maximum value reached by utility function u on an approximation A is denoted by $u^*(A) = \max_{\mathbf{z} \in A} \{u(\mathbf{z})\}$.

2. Outperformance relations

In the case of a MOO problem, the overall goal of the decision-maker (DM) is to select the single solution, which is most consistent with his or her preferences, the so-called best compromise. Generating an approximation to the non-dominated set is only a first phase in solving the problem. In the second phase, the DM selects the best compromise solution from the approximation, possibly supported by an interactive procedure. Therefore, the DM may consider approximation A as being better than approximation B if he or she can find a better compromise solution in A than in B.

We assume, however, that the DM's preferences are not known a priori. In fact, the use of heuristics generation of approximations to the full non-dominated set is justified only in this case. Nevertheless, one may be able to make some general assumptions about possible DM's preferences.

Using assumptions, we can state that an approximation A outperforms (is better than) B if, for some possible preferences held by the DM, the DM may find a better compromise solution in A than may be found in B and for other possible preferences, the solution found in A will be not worse than those found in B. Specifically, we will assume that all possible preferences of the DM may be modeled by utility functions belonging to a set U.

Let A and B be two approximations. Let $U(A > B) \subseteq U$ denote a subset of utility functions for which approximation A is better that B, i.e. $U(A > B) = \{u \in U \mid u^*(A) > u^*(B)\}$. Then, the following relation may be defined.

Definition 1. (Outperformance relation subject to a set of utility functions)

Approximation A *outperforms* B *subject to a set* U *of utility functions*, i.e. $A\ O_{/U}\ B$, if $U(A > B) \neq \emptyset$ and $U(B > A) = \emptyset$, i.e. there exist some utility functions in set U that achieve better values in A than in B, while the opposite is not true.

Obviously, the weaker the assumptions about DM's preferences, the more general the outperformance relation. The weakest, generally accepted assumption about the DM's preferences is that his/her utility function is compatible with the dominance relation (Rosenthal, 1985). In other words, the DM never prefers a solution that is dominated. This assumption means that the DM can limit the search for the best compromise solution to the set of efficient solutions. So, when two approximations A and B are known and from which the DM must select a compromise solution, the DM can limit the search for the best compromise to the set $ND(A \cup B)$, i.e. the set of points (and corresponding solutions) non-dominated within $A \cup B$. This allows us to define the following three dominance based outperformance relations.

Definition 2. (Weak outperformance)

Approximation A *weakly outperforms* B, and we write $A\ O_W\ B$, if $A \neq B$ and if $ND(A \cup B) = A$, i.e. if for each point $\mathbf{z}^2 \in B$ there exists a point $\mathbf{z}^1 \in A$ that is equal to or dominates $\mathbf{z}^2$ and at least one point $\mathbf{z}^1 \in A$ is not contained in B.

An approximation that weakly outperforms B can be obtained by adding to B a new point that is non-dominated with respect to all points in B as illustrated in Figure 1

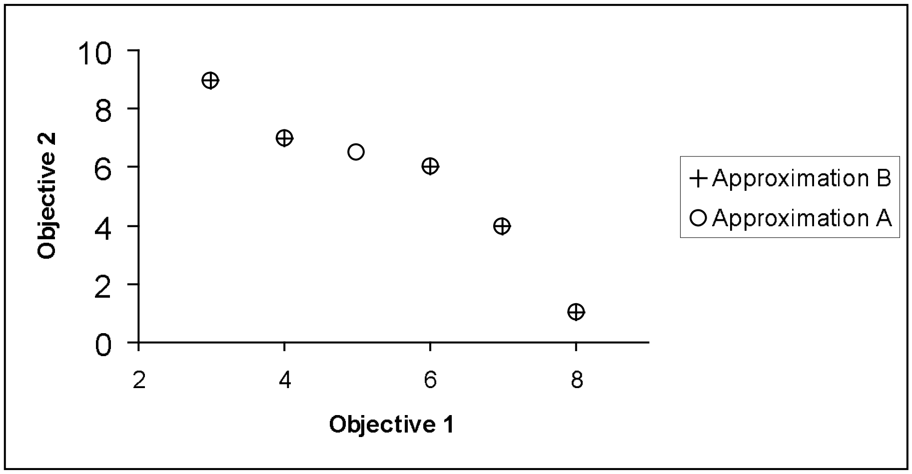

Figure 1. Example of weak outperformance – $A\ O_W\ B$**.**

Note that Definition 2 is equivalent to outperformance relation subject to the set U_C of utility functions compatible with the dominance relation. It is also equivalent to the outperformance relation subject to the set U_∞ of weighted Tchebycheff utility functions.

Definition 3. (Strong outperformance)

Approximation A *strongly outperforms* B, and we write $A\ O_S\ B$, if $ND(A \cup B) = A$ and $B \backslash ND(A \cup B) \neq \emptyset$, i.e. if for each point $\mathbf{z}^2 \in B$ there exists a point $\mathbf{z}^1 \in A$ that is equal to or dominates $\mathbf{z}^2$ and at least one point $\mathbf{z}^2 \in B$ is dominated by a point $\mathbf{z}^1 \in A$.

An approximation that strongly outperforms B can be obtained by adding to B a new point that dominates at least one point in B as illustrated in Figure 2.

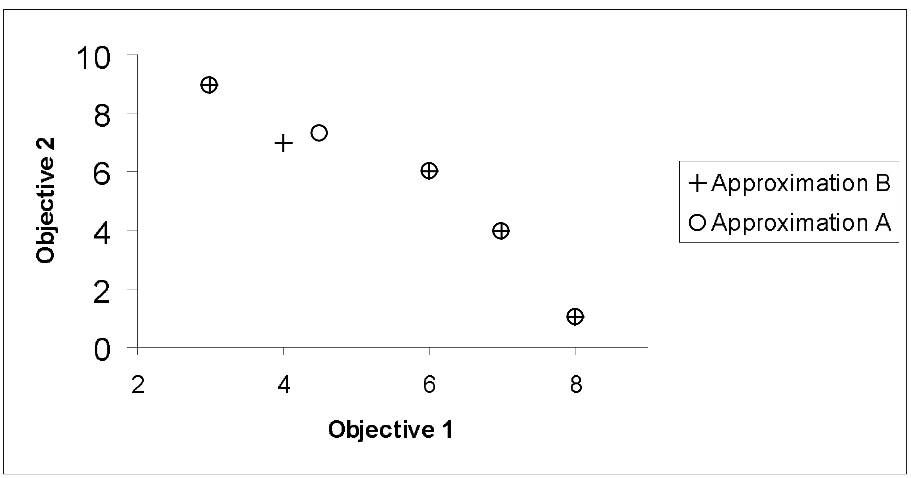

Figure 2. Example of strong outperformance – $A\ O_S\ B$**.**

Definition 4. (Complete outperformance)

Approximation A *completely outperforms* B, and we write $A\ O_C\ B$, if $ND(A \cup B) = A$ and $B \cap ND(A \cup B) = \emptyset$, i.e. if each point $\mathbf{z}^2 \in B$ is dominated by a point $\mathbf{z}^1 \in A$.

An approximation that strongly outperforms B can be obtained by adding to B a new point(s) that dominate all points in B as illustrated in Figure 3.

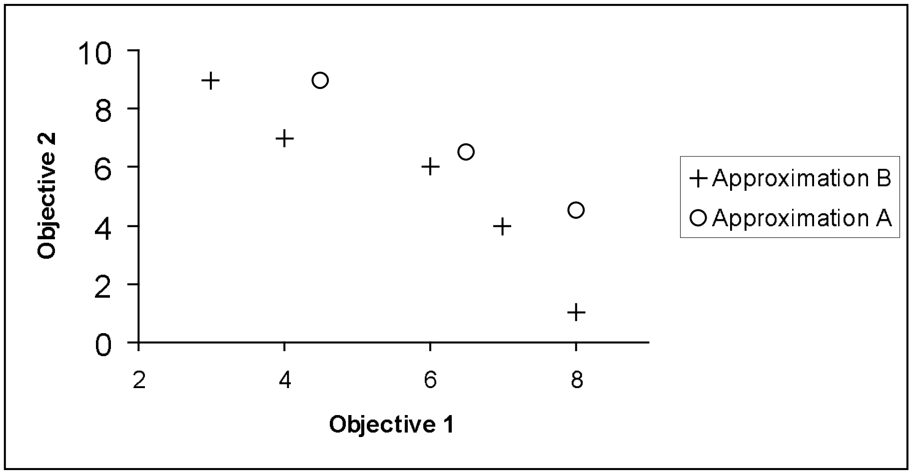

Figure 3. Example of complete outperfomance – $A\ O_C\ B$**.**

Note that $O_C \subset O_S \subset O_W$, i.e. complete outperformance is the strongest and the weak outperformance is the weakest of the outperformance relations. Each of the relations defines an incomplete ranking in the set Ω of all approximations.

If no additional assumptions about the DM's preferences (except of the compatibility with the dominance relation) are justified we propose to compare two or more approximations using the above three dominance based outperformance relations. They allow to assess that one approximation is better than another and also to grade the level of outperformance.

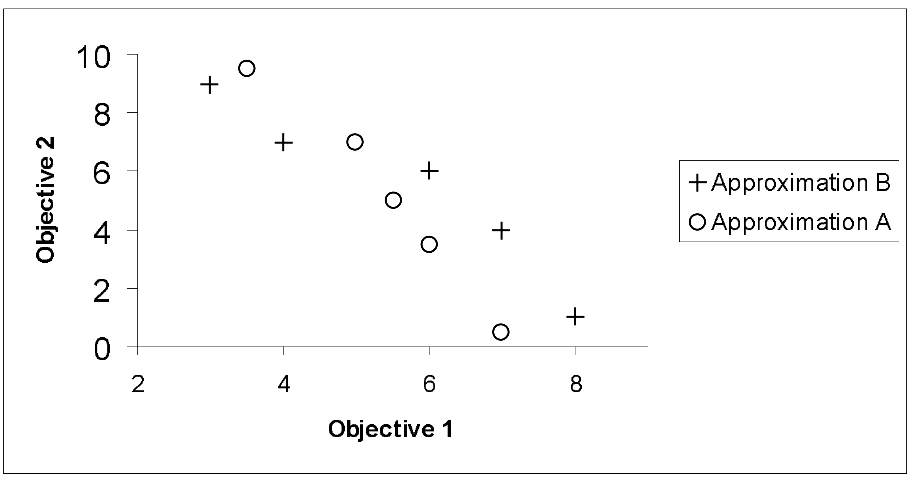

Figure 4. Approximations incomparable according to the weak outperformance.

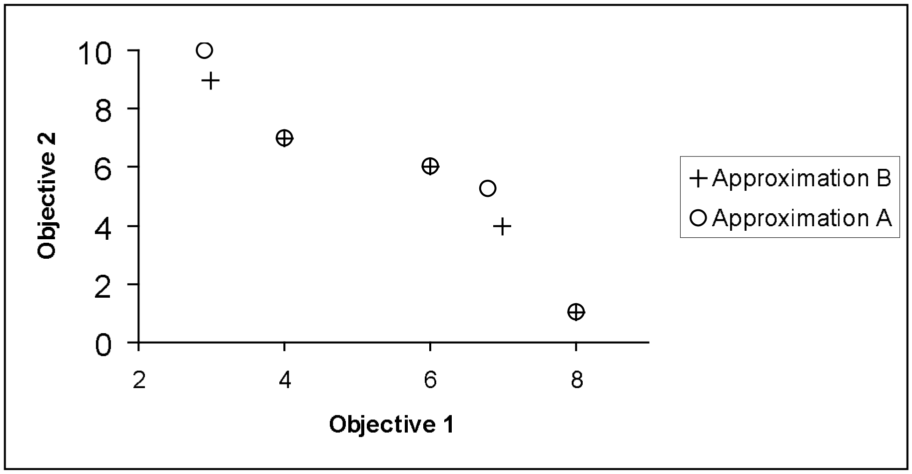

Figure 5. Approximations incomparable according to the weak outperformance but comparable according to outperformance subject to the set U_1 of weighted linear utility functions - $A \not{O}_W B$ **but** $A O_{/U_1} B$**.**

Obviously, many pairs of approximations will remain incomparable with respect to the weak outperformance (and so with respect to the stronger relations), i.e. none of the approximations will weakly outperform the other one, as is the case in Figure 4. In this case, one can introduce stronger assumptions about the DM's preferences, e.g. by defining another set U of possible utility functions, and test if one of the approximations will outperform the other one subject to this set. For example, the two approximations presented in Figure 5 are incomparable with respect to the weak outperformance, but approximation A outperforms B subject to the set U_1 of weighted linear utility functions.

Furthermore, the outperformance relations allow for qualitative comparison only. In many cases one will be interested in not only whether an approximation is better than another one but also how much better it is. For example both approximations B and C presented in figure 6 are completely outperformed by approximation A. Intuitively, however, the difference between A and B is much smaller than the differences between A and C and between B and C.

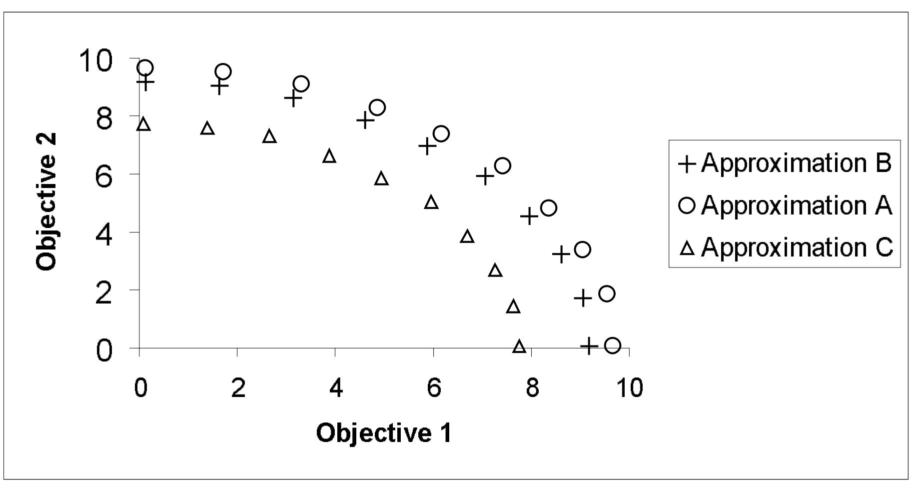

Figure 6. Different levels of differences between approximations.

3. Quantitative comparison methods

In this Section, we will describe some quantitative comparison methods that can be used to compare approximations that are incomparable according to the outperformance relations.

Quantitative comparison methods will also be based on some assumptions about the DM's preferences. It is natural to demand evaluations obtained by a quantitative comparison method to be concordant with outperformance relations based on these assumptions or on more general assumptions about the possible preferences. This demand is defined formally below.

Definition 5. (Weak compatibility with an outperformance relation)

A comparison method R is *weakly compatible* with an outperformance relation O, where O=$O_{/U}$, O_W, O_S or O_C, if for each pair of approximations A and B, such that $A\ O\ B$, R will evaluate approximation A as being not worse than B.

Definition 6. (Compatibility with an outperformance relation)

A comparison method R is *compatible* with an outperformance relation O, where O= $O_{/U}$, O_W, O_S or O_C, if for each pair of approximations A and B, such that $A\ O\ B$, R will evaluate approximation A as being better than B.

3.1 Comparison based on probability

Assume that each of the considered utility functions has a given probability of being the one held by the DM. When comparing two approximations A and B, it would then make sense to consider approximation A as being the better one, if there is a high probability for the utility functions in which approximation A presents a better solution than approximation B.

We express the probabilities of utility functions by an intensity function $p(u)$ and introduce an outcome function of the comparison between two approximations using a given utility function:

$$C(A,B,u)=\begin{cases}1 & if\ u^*(A)>u^*(B)\\ \tfrac{1}{2} & if\ u^*(A)=u^*(B)\\ 0 & if\ u^*(A)<u^*(B)\end{cases}$$

We then build the measure $R1$ to reflect the probability that approximation A is better than approximation B by integrating over all utility functions:

$$R1(A,B,U,p)=\int_{u\in U} C(A,B,u)\,p(u)\,du$$

According to this measure approximation A is better than B if $R2(A,B,U,p) > ½$ and A is not worse than B if $R2(A,B,U,p) \geq ½$. We notice, that $R1(A,B,U,p) = 1 - R1(B,A,U,p)$ and is therefore not possible, that one approximation can be better than another one while the opposite also is true.

$R1$ is weakly compatible with the outperformance relation for any set of utility functions $U \subseteq U_c$. $R1$ is compatible with outperformance relation subject to a U set of utility functions if the same set U is used in $R1$ and if the probability of selecting a utility function $u\in U(A > B)$ is always greater than zero whenever $U(A > B) \neq \emptyset$. The latter condition is, for example, assured if U is a parametric set and the probability intensity is a continuous and positive function of the parameter vector.

As an example, Figure 7 shows the u* values of two approximations A={[3,10],[5,7],[9,7]} and B={[2,9],[5,6],[10,6]} using the Chebycheff utility function with $\mathbf{z}^*$ = [10,10] and $\Lambda = [t, 1 - t]$, $t \in \langle 0,1 \rangle$:

$$U(t) = \left\{ u_\infty\left(\mathbf{z}, \mathbf{z}^*, \Lambda(t)\right) = -\max_j \left\{ t(10 - z_1), (1-t)(10 - z_2) \right\}, t \in \langle 0,1 \rangle \right\}$$

Only when $t > 0.8$ will approximation B contain the best alternative. If we presume that the utility functions are distributed as follows by letting t belong to the continuous uniform distribution, $t \in U\langle 0,1 \rangle$, we get $R1(A,B) = 0.8$.

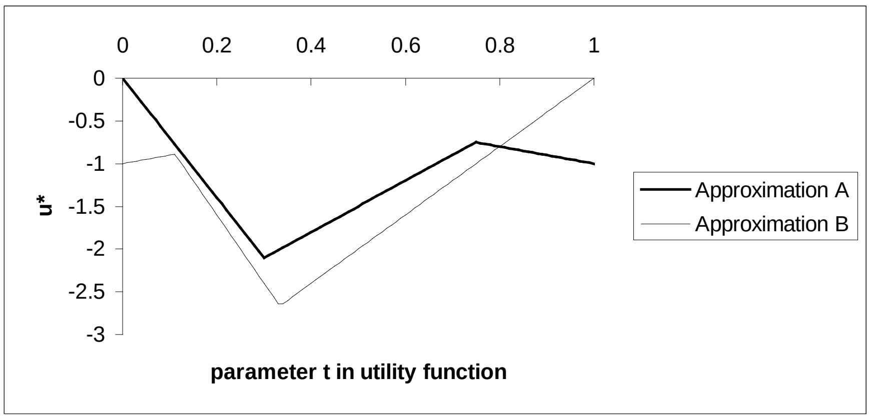

Figure 7. u* values of approximations A and B.

When using $R1$ to rank more than 2 approximations it is possible to obtain cycles where $R1(A,B,U,p) < ½$, where $R1(B,C,U,p) < ½$ and where $R1(B,C,U,p) < ½$, indicating that B is better than A, that C is better than B and that A is better than C. Figure 8 outlines such an example.

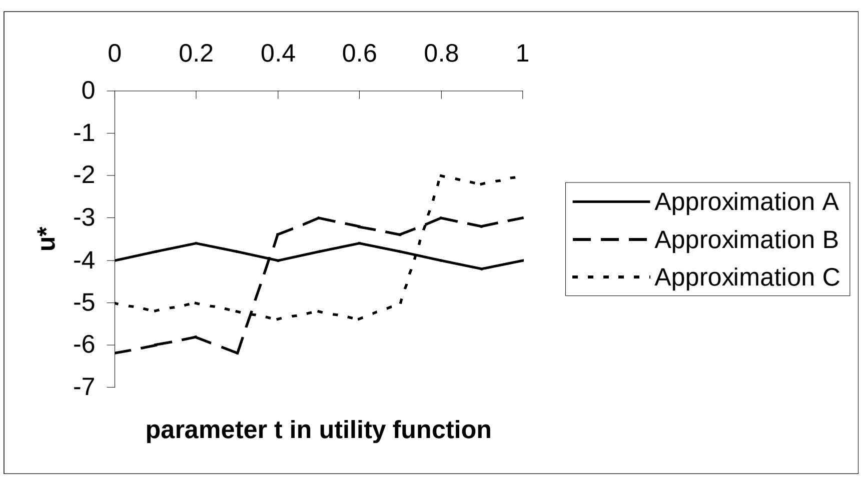

Figure 8. Example of cycles in ranking.

A different approach to ranking a number of approximation is to compare the approximations individually with a fixed reference set, R:

$$R1_R(A,U,p) = R1(A,R,U,p) = \int_{u \in U} C(A,R,u)\, p(u)\, du$$

This measure induces a complete ordering of approximations and cycles can therefore not occur. The measure is weakly compatible with the outperformance relation for any set of utility functions $U \subseteq U_c$.

However, it is not compatible even with the complete outperformance relation, if e.g. $A\ O_C\ B$ but $R\ O_C\ A$, then $R1_R(A,U,p) = R1_R(B,U,p) = 0$. This also shows that for this measure to be useful, the reference set R should reflect an *attainable quality* over the utility functions.

It is also worth mentioning that the resulting ranking of approximations will depend on the reference set used. With two different reference sets, R and P, we can be in a situation where $R1_R(A,U,p) > R1_R(B,U,p)$ and $R1_P(A,U,p) < R1_P(B,U,p)$, even if $R\ O_C\ P$. Still, we consider it beneficial to use more than one reference set, especially if these are on significantly different quality levels.

Notice, that the $C(A,B,u)$ function merely performs a ranking of the two approximations to see which contains the best point with respect to the utility function. This approach can easily be generalized to encompass more than two approximations, A_1, A_2, …, A_K by considering the rank of an approximation in $u^*(A_1)$, $u^*(A_2)$, …, $u^*(A_K)$:

$$C(A_i; A_1, A_2, ..., A_K, u) = \frac{rank(u^*(A_i); \{u^*(A_1), u^*(A_2), ..., u^*(A_K)\}) - 1}{K - 1},$$

where rank(x; X) is the rank of element x in the set X giving the value 1 is $x = \min\{X\}$ and the value $|K|$ if $x = \max\{X\}$. The measure from integrating over all u in U then yields the *expected relative rank* of an approximation among the K approximations.

A similar approach is to use a group of reference sets, R_1, R_2, …, R_K and calculate the expected relative rank of an approximation with respect to this set. These two generalizations are inspired by work of Fonseca and Fleming (1996).

3.2 Comparison based on expected values

Consider the following two approximations A={[1,10],[10,1.8]} and B={[2.2,10],[7,-1]}. Values of u^* are presented in Figure 9, where u is a Chebycheff utility function (2) defined in the same way as in Section 3.1. When t < 0.5125, approximation B contains the best alternative. So, applying measure $R1$ we obtain $R1(B) = 0.5215 > R1(A) = 0.4875$. However, in the region where approximation B gives better utility approximation A gives only slightly worse results, while in the other region the utility given by approximation A is significantly better. So, intuitively we might evaluate A better that B.

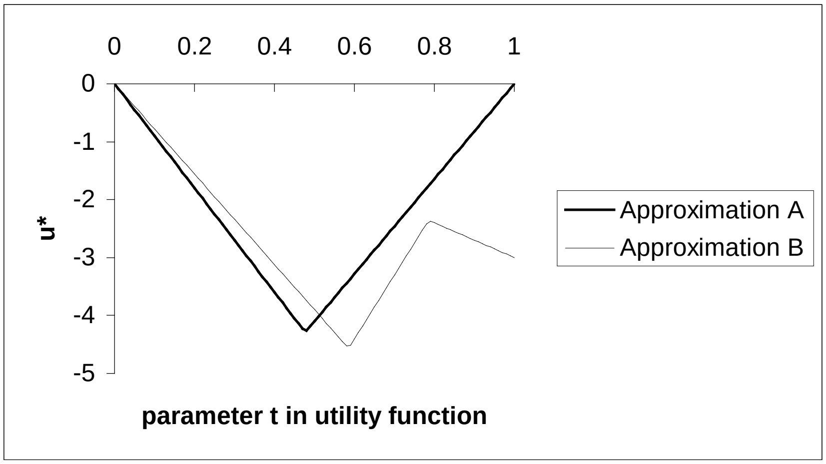

Figure 9. u* values of approximations A and B.

We propose to use the following measure that takes into account expected values of the utility:

$$R2(A,B,U,p) = E\left(u^*(A)\right) - E\left(u^*(B)\right) = \int_{u \in U} u^*(A) p(u) du - \int_{u \in U} u^*(B) p(u) du = \int_{u \in U} \left(u^*(A) - u^*(B)\right) p(u) du$$

According to this measure approximation A is better than B if $R2(A,B,U,p) > 0$ and A is not worse than B if $R2(A,B,U,p) \geq 0$. Obviously, $R2(A,B,U,p) = -R2(B,A,U,p)$.

In the above example we obtain $E(u^*(A)) = -2.1239$, while $E(u^*(B)) = -2.5924$, and therefore $R2(A,B,U,p) = 0.4684 > 0$. So, approximation A is evaluated better than B.

$R2$ is weakly compatible with the outperformance relation for any set of utility functions $U \subseteq U_c$. $R2$ is compatible with outperformance relation subject to a U set of utility functions under the same conditions as $R1$.

$R2$ induces complete ranking in the set of all approximations.

Measure $R2$ is based on an assumption that we are allowed to add values of different utility functions. Therefore, it is also dependent on the scaling of the different utility functions. Assume for example that in the above example each Chebycheff function is multiplied by $(1-t)^2$ (see Figure 10). In this case we obtain $E(u^*(A)) = -0.6361$, while $E(u^*(B)) = -0.59996$ and then $R3(A,B,U,p) = -0.03614 < 0$. So, approximation B is evaluated better than A.

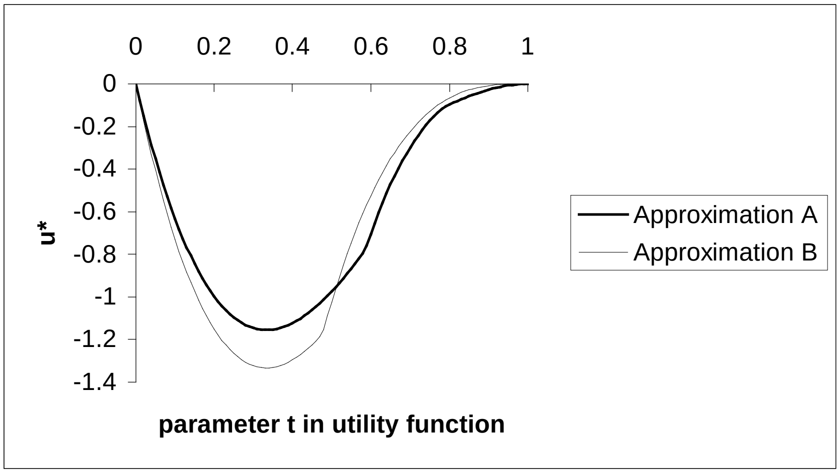

Figure 10. u* values multiplied by $(1-t)^2$ of approximations A and B.

The same method can be used to construct a quality measure that evaluates a single approximation A with respect to a fixed reference set R:

$$R2_R(A,U,p) = R2(R,A,U,p) = E\left(u^*(R)\right) - E\left(u^*(A)\right) = \int_{u \in U} u^*(R)p(u)du - \int_{u \in U} u^*(A)p(u)du = \int_{u \in U} \left(u^*(R) - u^*(A)\right)p(u)du$$

In this case, the lower the value of the measure is, the higher is the evaluation of A.

In some cases, ratios of best utility values may be more meaningful than their differences. Therefore, we propose also the following ratio measure:

$$R3(A,B,U,p) = E\left(\frac{u^*(B) - u^*(A)}{u^*(B)}\right) = \int_{u \in U} \frac{u^*(B) - u^*(A)}{u^*(B)} p(u)du\ .$$

When a reference set is known we may use the idea of $R3$ to construct the following measure:

$$R3_R(A,U,p) = R3(A,R,U,p) = E\left(\frac{u^*(R) - u^*(A)}{u^*(R)}\right) = \int_{u \in U} \frac{u^*(R) - u^*(A)}{u^*(R)} p(u)du$$

The idea of $R3_R$ follows an approach often used in single objective optimization where approximate solutions obtained by single objective heuristics are evaluated by the ratio of the deviation from a fixed bound, for instance the optimal value.

In general, also the values of the ratio measures are dependent on the scaling of the objective functions.

4. Reference information

The dominance-based outperformance relations introduced in Section 2 do not require information except the points of the two approximations being compared. However, in the case of outperformance relation subject to a set of utility functions and the quantitative comparison methods proposed in Section 3 there is a need for further information. In this Section, we characterize some sources of additional information about a given MOO problem (beyond those characterizing DM's preferences) that could be useful while comparing approximations of the non-dominated set.

4.1 Reference levels

In Section 3, we have used a reference set, R, with respect to which we can compare an approximation. Here we will discuss some of the requirements we have to such a set. First please notice, that we only use the reference set in order to obtain values of $u^*(R)$ for each given utility function. We may extend this by describing a *reference level* to each utility function. This is a generalization, since there might not exist a realization of a reference set, R, which can give the same values of $u^*(R)$ for each of the utility functions.

The reference level may then be described by, for instance, a relaxed correspondent to the original problem, relaxing the integer/binary constraints on decision variables or relaxing other constraints. This may result in upper bounds with known maximal (or consistent) deviation from optimum, insuring some degree of homogeneous quality over the non-dominated frontier.

Lower bounds for the exact reference levels can be found from e.g. approximation algorithms such as heuristics. Using these, one may be in a situation where an approximation performs better than the reference level for some utility functions. However, this is no different from the single-objective case. Again, the most important issue is to insure a homogeneous quality level over the non-dominated frontier.

Publication of such reference levels can be given as the relaxed model used, the approximation algorithm used, etc., but it can be useful for other researcher if also some of the actual utility function values are available. For measure $R2_R$ one can publish the integral value (perhaps with different numerical integration procedures and different sampling intervals) of $\int_{u \in U} u^*(R)\, p(u)\, du$.

It will often be useful to provide more than one reference level for a problem instance. As stated in Silver et. al. (1980), the purpose of heuristic methods is not only to be close to the optimal solution, but also to be far away from poor solutions. Reference levels describing poor solutions should therefore also be provided, for instance as generated by random solutions. In between, one may provide mediocre reference levels such as those obtained by generating local optima from e.g. simple steepest ascent heuristics.

4.2 Reference sets

Nevertheless, it can be more convenient to use an actual set of points, a reference set, to describe the reference levels since it may allow for easy interchange among researchers. Also, using an actual reference set may allow for fast calculation of some of the measures using efficient data-structures, such as quad-trees (see Habenicht, 1982, and Sun and Steuer, 1996).

The reference set should ideally be the full (optimal) non-dominated set but for computational reasons, one may have to settle for or a subset of this, for instance a well dispersed subset, all or some of the supported non-dominated points, or the like.

Still some problems may be so difficult to solve, that one cannot expect to find any (guaranteed) non-dominated points at all, to use in the reference set. A second approach is then to use a set of potentially non-dominated points as generated by approximation algorithms.

If the reference set does not solely consist of non-dominated points, one must pay special attention to insure a homogeneous quality of the potentially non-dominated points all-over the frontier. This must be argued for by the constructor of the reference set. Two main arguments will here be the deviation from optimality/non-dominance (which again can be measured in different ways) and the computational/algorithmic effort that has been used to obtain the points.

When the reference set contains non-dominated points, but only a subset of these, it will also be necessary to argue for homogeneous quality over the non-dominated frontier; here in terms of how well dispersed the subset is. The issue of well dispersed-ness is often relevant, since non-dominated sets can be very large indeed. Filtering techniques may here be useful (Steuer, 1986).

4.3 The ideal point

In the utility functions themselves, we will also often use additional information. Consider for example the utility functions within U_p. These use a weight vector and the ideal point, $\mathbf{z}^*$, as an anchoring point. We will first discuss the ideal point and then the weight vector.

The most used anchoring points are the ideal point and the nadir point. When we prefer using the ideal point, it is due to the fact, that the nadir point is notoriously difficult to determine for problems with more than two objectives. One risk to estimate the nadir point terribly wrong even for otherwise simple problems (Weistroffer, 1985). Even in the relatively simple case of multiple objective linear programming, the estimation is difficult, especially as the problems grow in size (Korhonen et. al., 1997).

Hence, we prefer to use the ideal point in the scalarizing functions defining the utility functions. However, determining the exact ideal point may also cause difficulties, since this imply solving each of the single objective programs to optimality. This may be possible in some cases whereas for others one has to estimate the ideal point. In the latter case it is important not to underestimate the ideal point since this implies that approximation points non-dominated with respect to the estimated ideal point normally will not contribute fully to the measure, and in some cases, will not contribute at all.

If one has to estimate the Ideal point, or some of its components, it is therefore important that the estimate is an overestimation. This may be found as upper bounds to the problem, perhaps from a relaxation. If some of the components of the Ideal point could be the optimal values, one should in order for the measures to function properly for points attaining the optimal level on one or more objectives add an ε-value to each component of the Ideal point (Steuer, 1986, Section 14.1). The ε-value can be set as a very small positive constant multiplied with the range equalization factors.

The utility function values will obviously depend on the coordinates of the ideal point and so will the measures. One must therefore always accompany the measuring results with the coordinates of the used ideal point including the ε-values. For the utility functions of family U_1 however, the utility function values only differ by a constant.

4.4 Range scaling

The purpose of the weight vector Λ is in fact twofold. One is to prescribe the importance of each objective, the higher the weight, the higher the importance of that objective. The second is to scale the ranges of the objectives, so those objectives with large ranges do not dominate the objectives with smaller ranges. These two are normally separated into two multiplicative factors, the importance weight and a range equalization factor. In Section 5 we suggest ways of using different importance weights in the measures and will for now only consider the range scaling in terms of the range equalization factors, as defined in formula (1).

The range equalization vector holds as the j'th element 1 divided by the range on objective j. Most naturally would be to use the ideal and nadir point for calculating the ranges, and this we suggest to do for bi-objective problems. With more than two objectives, the nadir point is as mentioned previously often so difficult to find, that alternative approaches can be desirable. Some possibilities will be discussed in the remainder of this Section.

The *anti-ideal point* (or an approximation for this) can be used in replacement for the nadir point. This makes most sense if the objectives are not correlated and if it is likely that the nadir point is placed close to a line between the ideal point and the anti-ideal point.

A *probabilistic approach* can be used if the objectives are independent and thereby not correlated. In this case, the levels on each objective will be independent of the level on other objectives and a fixed (low) percentile on each objective from e.g. a sample of random solutions can be used to estimate the levels of the nadir point. One must be careful when using this approach on problems where the objectives only are non-correlated but not independent.

One can *omit range* scaling in cases where it can be assumed, that the ranges are more or less equal. This may e.g. be the case for problems where all objective functions are of the same type and the parameters defining the corresponding single objective instances are generated independently and "in the same way".

The *pay-off table*, i.e. the matrix formed by the points which define the ideal point, can be used to obtain a setting for the nadir point as the lowest observed level on each objective. In cases of multiple optima for the single-objective problem, effort can be put into locating all optima. This approach may be best for problems with strongly correlated objectives and is often seen used in practice.

An *empirical approach* is to build an over-all best approximation from all approximations obtained in the experiments on the problem instance. From this set, an empirical nadir point can be found and used for the final calculations of measures. This may be the best option in cases with strong non-linearity in objectives, when the solution space is restricted in special ways or when there is some correlation among objectives.

A *common-sense* addition to the other approaches can be desirable. For instance, an objective may obtain extremely poor values, as can be the case in e.g. bi-criteria project scheduling with minimization of costs and project completion time. Here, the schedules of low cost may be so lengthy, that they are without practical interest. A lowest-value-of-interest on some objectives may therefore be relevant in calculating the range equalization factors. This is obviously also necessary if the problem has unbounded objectives as can be the case in multiobjective programs other than MOCO problems.

Again, the range equalization factors should always be given with the measuring results. This allows other researchers to calculate directly comparable measures (using measures $R2$ and $R3$) for their approximations.

Also, if the nadir point is estimated through feasible solutions (as is the case with the pay-off table and the empirical approach) the points defining the nadir point should be reported so as to possibly improve the estimation of the nadir point for other researchers.

5. Guidelines for constructing measures

The quality measures proposed in Section 3 include parameters which may have crucial influence on the results of the evaluation. If the measures are used in experiments which results are published, the authors should precisely specify what was the setting of the parameters, namely:

- what set of utility functions was used,
- what was the probability distribution of the utility functions and
- how the objective functions were scaled (in the case of measures $R2$ and $R3$).

In this Section we give some general guidelines for setting these parameters and suggest "best practice" for a number of cases.

5.1 The choice of the set of utility function

If no additional information about the possible preferences of the DM is available, one should choose a set of utility functions that does not disregard any non-dominated points. Obviously, this is assured by the set U_c of utility functions compatible with the dominance relation but this set cannot be used in practice. We propose to use the set U_∞ of weighted Tchebycheff utility functions in this case.

In other cases, additional assumptions about the DM's preferences may be justified. For example, one may know that the DM's preferences can be expressed by a weighted sum of objectives, but where the weights are otherwise unknown. In this case, the set U_1 of weighted linear utility functions should be used.

5.2 The choice of probability distribution of the utility functions

Assume that two approximation A, B are incomparable according to outperformance relation subject to the selected set of utility functions U. It is then always possible to choose a probability distribution such that one of the approximations will be evaluated higher than the other one by the quality measures. This can be done by defining intensity function $p(u)$ that take on high values in regions of set U where one approximation gives higher utility and low values in the other regions. Of course, some of the distributions will be more "artificial" that the others.

An obvious requirement is that the intensity function $p(u)$ should be greater than 0 for all $u \in U$. Otherwise, functions with intensity function value equal to zero can be excluded from U. The probability distributions also should not "favor" any subregion of U.

In the case of weighted Tchebycheff or weighted linear utility functions, the utility functions are defined by the selected weight vector Λ (see formulas (2) and (3)). Therefore, the distribution of the utility functions can be defined via the distribution of weights. For practical reasons, we propose to use normalized weight vectors, i.e. vectors belonging to the following set:

$$\Psi = \left\{ \Lambda \in \Re^J \mid \sum_{j=1}^{J} \lambda_j = 1 \text{ and } \lambda_j > 0, j = 1,.., J \right\},$$

We then propose to use a uniform distribution of weights, i.e. a distribution for which:

$$\forall \Psi' \subseteq \Psi \quad \int_{\Lambda \in \Psi'} p(\Lambda) d\Lambda \Big/ \int_{\Lambda \in \Psi} p(\Lambda) d\Lambda = V(\Psi') / V(\Psi)$$

where $V(\Psi)$ and $V(\Psi')$ are Euclidean hyper-volumes of Ψ and Ψ', respectively. In other words, the probability of a weight vector belonging to Ψ' should be proportional to the hyper-volume of Ψ'. The normalized weight vectors should be applied to normalized objective values, i.e. original objective values multiplied by range equalization factors (1).

Of course, if there are reasons to assume that some utility functions are more probable, one can modify the probability distribution to reflect this knowledge, as we will discuss in Section 7.

5.3 Utility functions' scaling

Utility functions' scaling is crucial for measures $R2$ and $R3$ based on expected values. The measures require that similar changes of the best values of different utility functions reflect similar changes of the quality of the best solutions, i.e. the utility functions should be linear with respect to the intuitive interpretation of the quality of solutions.

Having a set U of utility functions, we propose to scale them to the range 0-1, where 0 is the worst and 1 is the best value. The extreme values should be achieved by points of some intuitively understood quality, e.g. the quality of a reference set.

If a good reference set R is known, the utility functions should be scaled such that their maximum in this set is equal to 1. Otherwise, one can use points obtained by solving a relaxed correspondent to the original problem, if a tight relaxation can be found (cf. Section 4.1). If no such information is known, we propose to use ideal point or its approximation to scale the utility functions, so that they achieve the value 1 in the (approximation of) ideal point.

The value of 0 should be associated with some "poor" points, preferably worse than all the approximations evaluated. One can for instance use (the approximations of) the anti-ideal or the nadir point. If no approximation of the nadir point is known but range equalization factors π_j are defined through other methods, one can associate value of 0 with the following point:

$$\mathbf{z}^- = [z_1^-, ..., z_J^-], \qquad z_j^- = z_j^* - \beta \, 1/\pi_j, \; j = 1, ..., J,$$

where $\beta > 0$ and large enough for $\mathbf{z}^-$ to be dominated by all the points in evaluated approximations.

5.4 Computational issues

Measures defined in Section 3 are generally difficult to calculate analytically since they require integration of non-smooth functions. Daniels (1992) proposed a method that can be used to calculate $R3$ in the case of linear utility functions.

Of course, the values of all the measures can be found by numerical integration. This requires sampling the set U of utility functions according to its distribution. In the case of weighted Tchebycheff or weighted linear utility functions this turns into sampling weight vectors Λ. In Section 5.2 we proposed to use set Ψ of normalized weight vectors. We propose to sample this set by generating all weight vectors in which each individual weight takes on one of the following values: $\left\{ l/k,\ l = 0,...,k \right\}$, where k is a sampling parameter defining the number of weight levels. With a combinatorial argument, we notice that this produces $\binom{k+J-1}{J-1}$ weight vectors.

For example, for k=3 and J=3, we obtain the following set of 10 vectors: {[0,0,1], [0,1/3,2/3], [0,2/3,1/3], [0,1,0], [1/3,0,2/3], [1/3,1/3,1/3], [1/3,2/3,0], [2/3,0,1/3], [2/3,1/3,0], [1,0,0]}. Again, the weight vectors should be applied to normalized objective values.

If the results of numerical integration are reported in the description of an experiment the authors should precise all the necessary details of the algorithm.

6. Comments on other approaches

In this Section we comment on some quality measures used in published experiments with multiple objective heuristics. We characterize the quality measures using our framework, i.e. we take into account compatibility with the outperformance relations and (implicit) assumptions about the DM's preferences.

6.1 Cardinal measures

If a reference set composed of all the efficient solutions is known, it may seem that the most natural quality measure is the ratio of the reference points found. The measure may be defined in the following way:

$$C1_R(A) = \frac{|A \cap R|}{|R|}.$$

This measure was used by e.g. Ulungu (1993) and by Morita and Katoh (to appear).

If the reference set does not contain all non-dominated points then the points from A, which are non-dominated by points contained in R, may actually belong to the non-dominated set. In this case, it may be more reasonable to use the following measure, which is defined as the ratio of approximation points non-dominated by R:

$$C2_R(A) = \frac{|\{\mathbf{u} \in A \mid \nexists \mathbf{r} \in R \ \mathbf{r} \succ \mathbf{u}\}|}{|A|}.$$

This measure was used by Tuyttens et al. (to appear).

The cardinal measures have, however, some significant disadvantages. They are very insensitive to the improvements of the approximation. Consider for example the two approximations of a reference set presented in Figure 11 obtained for a bi-objective knapsack problem. Clearly, approximation 1 is much better than approximation 2. All points in approximation 1 are very close to the reference set and they cover most regions of the reference set. Both approximations will have, however, the same worst values of the measures $C1$ and $C2$. In fact, $C1$ is only weakly compatible with relations O_C, O_S and O_W. Measure $C2$ is weakly compatible with O_C relation only.

The use of cardinal measures seems to be reasonable only if there is a high probability that a method is able to find a significant percentage of non-dominated solutions. In the case of larger problems, finding non-dominated points may be very difficult (compare results of experiment described in Czyzak and Jaszkiewicz, 1998). Please note, that single objective heuristic normally not are treated as global optimization tools. They should rather generate, in relatively short time, solutions close to the optimal one. Analogously, multiple objective metaheuristic should, within a realistic computational time, give a good approximation to the whole non-dominated set.

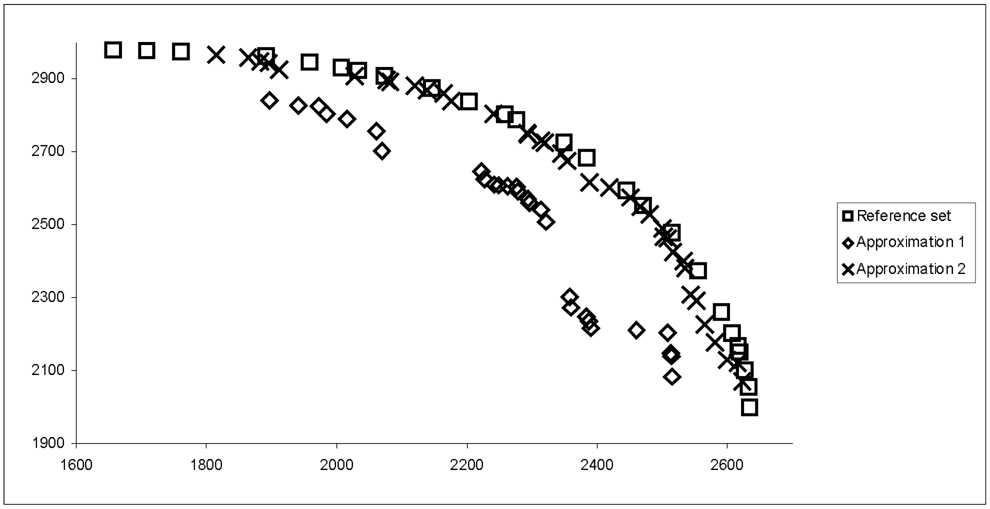

Figure 11. Reference set composed of all supported solutions and two approximations obtained for a two objective knapsack problem (100 elements).

Another drawback of the cardinal measures is illustrated by the example presented in figure 12. The two approximations are composed of 5 non-dominated points, so, their cardinal measures are equally good. All points composing approximation 3 are, however, very close in the objective space, i.e. they represent the same region of the non-dominated frontier. The points of approximation 4, on the other hand, are dispersed over whole reference set. They carry much richer information, e.g. about the possible ranges of objectives. This example shows that for the cardinal measures, each point in the approximation has the same weight regardless of their proximity and information concerning the shape of the non-dominated set.

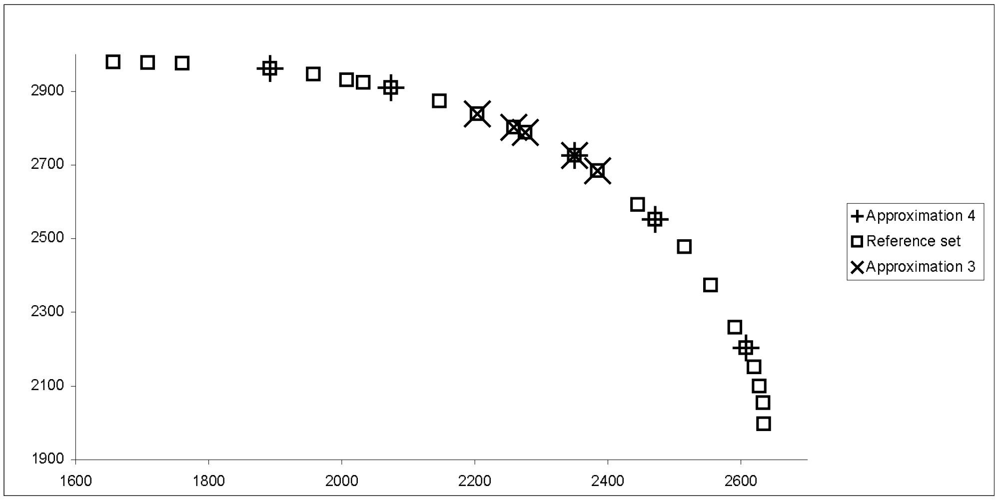

Figure 12. Reference set composed of all supported solutions and two other approximations obtained for a two objective knapsack problem (100 elements).

6.2 Distance measure

Czyzak and Jaszkiewicz (1998) proposed the following distance measure based on a reference set R:

$$D1_R(A,\Lambda)=\frac{1}{|R|}\sum_{\mathbf{r}\in R}\min_{\mathbf{z}\in A}\{d(\mathbf{r},\mathbf{z})\},$$

where $d(\mathbf{r},\mathbf{z})=\max_j\{\lambda_j(r_j-z_j)\}$ and $\Lambda=[\lambda_1,....,\lambda_J]$, $\lambda_j = 1/R_j$, $j=1,...,J$ with R_j being the range of objective j in set R.

The measure was also used by Tuyttens et al. (to appear) and by Viana (1997).

Obviously, $D1$ is weakly compatible with the O_C, O_S and O_W outperformance relations. It is not compatible, however, even with the strongest complete-outperformance relation O_C. Consider for example the bi-objective problem with the following reference set:

$R = \{\mathbf{r}^1 = [9, 4], \mathbf{r}^2 = [4, 9]\}$,

and the following two approximations

$A = \{\mathbf{u}^1 = [8, 2], \mathbf{u}^2 = [2, 8]\}$,

$B = \{\mathbf{u}^3 = [8.5, 2], \mathbf{u}^4 = [2, 8.5]\}$.

Obviously, $\mathbf{u}^3 \succ \mathbf{u}^1$ and $\mathbf{u}^4 \succ \mathbf{u}^2$, so, $B\ O_C\ A$, however, $D1_R(A) = D1_R(B) = 0.2$. This happened because points $\mathbf{u}^1$ and $\mathbf{u}^2$ are weakly non-dominated with respect to the points $\mathbf{u}^3$ and $\mathbf{u}^4$. In practice, however, it is rather unlikely that all points in an approximation, that is completely outperformed, are only weakly dominated by the points of the other approximation.

Consider now compatibility of the measure $D1$ with the strong outperformance relation. Assume that the reference set and the weighting vector are the same as used in previous example and the following two approximations are evaluated:

$A = \{\mathbf{u}^1 = [8, 2], \mathbf{u}^2 = [2, 8], \mathbf{u}^3 = [4, 4]\}$,

$B = \{\mathbf{u}^4 = [8.5, 2], \mathbf{u}^5 = [2, 8.5], \mathbf{u}^6 = [4.5, 4.5]\}$.

Clearly, B strongly outperforms A, however, $D1_R(A) = D1_R(B) = 0.2$.

The value of $D1$ may be interpreted as the average of maximum losses on all objectives when the approximation A is available in place of reference set R, where the losses on objectives are expressed as percentages of the (approximated) ranges of particular objectives in the non-dominated set.

One should be aware however, that this is in fact a weighted average, where the reference points have equal weight. In some cases, it may therefore give results against intuition. Consider for example the same bi-objective problem with the following reference set:

$R = \{\mathbf{r}^1 = [9, 4], \mathbf{r}^2 = [4, 9], \mathbf{r}^3 = [6, 6]\}$,

and the following two approximations

$A = \{\mathbf{u}^1 = [8, 2], \mathbf{u}^2 = [3, 8], \mathbf{u}^3 = [4, 4]\}$,

$B = \{\mathbf{u}^4 = [8.5, 3], \mathbf{u}^5 = [2, 7], \mathbf{u}^6 = [5, 5]\}$.

According to measure $D1$, approximation B is better than A. Assume now that the reference set is "improved" by adding two new points:

$R' = \{\mathbf{r}^1 = [9, 4], \mathbf{r}^2 = [4, 9], \mathbf{r}^3 = [6, 6], \mathbf{r}^4 = [3.99, 9.01], \mathbf{r}^5 = [4.01, 8.99]\}$,

Approximation A is now evaluated better than B, because points $\mathbf{r}^4$ and $\mathbf{r}^5$, although very close to $\mathbf{r}^2$ have the same importance.

According to our framework, functions $d(\mathbf{r}, \mathbf{z})$ may be interpreted as different utility functions and the value of $D1$ as the expected value ever these functions. Distribution of the functions depends on the distribution of points in the reference set that may be distributed very unevenly. Please notice that this very well may be the case even if the reference set is composed of all the non-dominated points.

There is no reason to assume that the regions of the non-dominated set with high density of points are more important for the DM. For example, the fact that in a bi-objective project scheduling problem most efficient schedules have long completion time does not mean that they are more important. Before using this measure one should therefore assure uniform distribution of points in the reference set or, alternatively, weight the points so as to give low importance (e.g. low weight) to points which are located close to other points in the objective space.

7. Evaluating an approximation under preference information

In practical applications, one may not seek an approximation to the entire non-dominated set, but only to parts hereof. This can be the case if the output of the heuristic is used in an interactive MCDM system where preference information is elicited and fed back to the heuristic. Indeed, an important reason for our choice of using probability of utility functions is that it in a practical and elegant way facilitates incorporation of certain types of elicited preference information into the measures.

The nature of preference information can be manifold, and the easiest to handle in our framework is an expression with respect to weights. One simply shapes the probabilities for the utility functions so as to reflect the preference information. Say for example that the DM has come to the realization that the preferred solution will have a weight (in the used scalarizing function) in the interval of $\langle 0.4, 0.6 \rangle$. Alternatively, that the weight is around 0.5 with a given certainty.

These types of preference information can be included in the measurement of an approximation by using probabilities for utility functions as illustrated in Figure 13. In the first case, we have limited the set *U* to contain weights in the interval of interest and here consider a constant probability function. In the second case, the set *U* is the entire set, but the probability function is changed to reflect the preference information.

Logically one should use probabilities, which integrate to one. This could also be omitted, however, since it only serves to scale the measure.

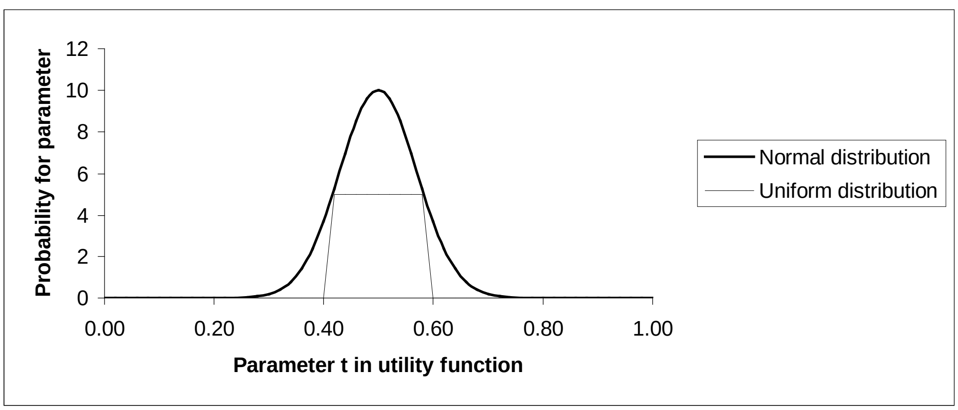

Figure 13. Weights with preference information

Another type of preference information is expressed by aspiration levels, which, for some or all objectives define a level of achievement on which the DM is satisfied. (The precise interpretation of such levels varies among interactive procedure.) This aspiration point (in which some levels may be aspiration levels and other may be ideal/optimal levels) gives the possibility of using a *displaced ideal* point instead of the Ideal point as the anchoring point in the utility functions of family u_∞. In this way, we do not measure improvements in the approximation beyond the aspiration levels.

8. The use of measures in multiobjective heuristics

In the beginning of this paper we stated, that the measures are useful information during execution of a heuristic procedure, namely as stopping rules and as a global objective function. However, we cannot expect to have reference information such as a reference set or the Ideal point available during the execution of a heuristic. This information is obviously part of the solution, which is sought and knowing it in advance would correspond to, in the case of single objective optimization, if we knew the optimal solution (or a good bound for it) before the heuristic procedure begins. Still, the heuristic may contain components aiming at acquiring such information. For example, the knowledge of the ranges of the objectives is often essential to a multi-objective heuristic aiming at generating an approximation which is equally good all over the non-dominated frontier. If reference information is used in heuristics for benchmark experiments, this must be explicitly stated along with the publicized results.

So, while a priori known reference information in general should not be used in the heuristic, it may to some extent be necessary if preference information is to be included. For instance, it does not make sense to define preference weights (such as a sub-space of the weight-space) for the Tchebycheff based utility function without prescribing an anchoring point. However, the anchoring point may be some point located arbitrarily far away from the ideal point on the diagonal of the contours defined by the scalarizing function used and going through the ideal point. In this way, the heuristic can not use the anchoring point for reference information.

Finally, the formula (but without a priori calculated parameters) by which the output of the heuristic is to be measured can be used in the heuristic to guide the search. The heuristic may then use the measuring formula with reference information as obtained during the calculation, to guide the search, for stopping criteria, etc. The computational effort of calculating the measure should always be included in the evaluation of the heuristic.

9. Conclusions and directions for further research

In this paper, we have introduced dominance-based outperformance relations between approximations that form a basis for other methods evaluation and comparison methods. The dominance-based outperformance relations are based on very weak and generally accepted assumptions about the DM's preferences and they do not require any additional information about the MOO problem.

The outperformance relations leave, however, many pairs of approximations incomparable and do not deliver quantitative information. Quantitative comparison of approximations requires stronger assumptions about the DM' preferences. It seems impossible however to propose general assumptions that could be accepted in all situations. This is not surprising, as different decision-makers (whose preferences are compatible with the dominance relation) among approximations, which are incomparable according to the outperformance relations, indeed may prefer different approximations.

We have therefore proposed a framework for quantitative measures of approximations. Specifically, three generic families of measures have been presented. These measures are simple in their conceptual basis, yet flexible enough for various types of needs. They are based on common-sense assumptions and are characterized by their concordance with the dominance-based outperformance relations. They form guidelines for construction of quantitative measures under different practical circumstances. However, other types of measures can also be developed, based on the presented framework.

We have also used our framework to characterize properties of some previously used quality measures. We have shown that these in some cases can be discordant with intuition and with the introduced dominance-based outperformance relations.

An important issue not fully discussed in this paper is the measurement of uniformness of the quality of an approximation. In general, we want the approximation to be equally good all over the non-dominated frontier. Expressed within the framework of the presented measuring approach, this refers to the approximation being equality good considering all utility functions. The consistency of quality of the approximation may be analyzed statistically and many types and sources for variation can occur. The most useful types of analysis is probably to investigate the quality at extreme utility functions (e.g. those generated with extreme weights) and in larger clusters of the generating weight space. This may give insightful information on the performance of the heuristics in terms of identifying its strengths and weaknesses.

We believe that the framework and the quality measures proposed in this paper are useful for experimental evaluations of multiple objective heuristic procedures and may contribute to further improvements of such procedures. We were of the opinion that the literature was in need for good and consistent quality measures within the field of heuristic multiobjective optimization and hope that the presented work is a step towards filling this gap.

Acknowledgements

The work of the second author was supported by State Committee for Scientific Research, KBN research grant no. 8T11C 013 13 and by CRIT 2 - Esprit Project no. 20288.

References

Andersen, K., A., Jörnsten, K., Lind, M. (1996). On bicriterion minimal spanning trees: An approximation, *Computers and Operations Research*, Vol. 23, No. 12, pp. 1171-1182.

Assad, A.A. (1988). Modelling and implementation issues in vehicle routing. In: B.L. Golden, A.A. Assad (eds.): *Vehicle routing: Methods and Studies*, North-Holland, Amsterdam, 7-46.

Cerny V. (1982). *A thermodynamical approach to the travelling salesman problem: An efficient simulation algorithm*, Report, Comenius University, Bratislava, Czechoslovakia.

Czyzak P., Jaszkiewicz A. (1996). A multiobjective metaheuristic approach to the localization of a chain of petrol stations by the capital budgeting model. *Control and Cybernetics*, **25**, 1, 177-187.

Czyzak P., Jaszkiewicz A. (1998). Pareto simulated annealing - a metaheuristic technique for multiple-objective combinatorial optimization. *Journal of Multi-Criteria Decision Analysis*, **7**, 34-47.

Daniels R.L. (1992). Analytical evaluation of multi-criteria heuristics. *Management Science*, **38**, 4, 501-513.

Dasgupta D., Michalewicz Z. (1997). *Evolutionary Algorithms in Engineering Applications*. Springer Verlag.

Fonseca C.M., Fleming P.J. (1993). Genetic algorithms for multi-objective optimization: Formulation, discussion and generalization. In: S. Forrest (ed.), *Genetic algorithms: Proceedings of the Fifth International Conference*, Morgan Kaufmann, San Mateo, CA, 141-153.

Fonseca C.M., Fleming P.J. (1995). An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary Computation*, **3**, 1, 1-16.

Fonseca C.M., Fleming P.J. (1996). On the performance assessment and comparison of stochastic multiobjective optimizers. *Lecture Notes in Computer Science*, **1141**, Springer Verlag, 584-593

Gandibleux, X., Mezdaoui N., Fréville A. (1996). A tabu search procedure to solve multiobjective combinatorial optimization problems, to appear in *Proceedings volume of MOPGP '96*, R. Caballero and R. Steuer (eds.), Springer-Verlag.

Glover F. (1989). Tabu Search - Part I, *ORSA J. Comput.* **1**, 190-206.

Goldberg D.E. (1988). *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, Massachusetts.

Habenicht, W. (1982). Quad trees, A data structure for discrete vector optimization problems, *Lecture notes in economics and mathematical systems*, **209**, 136-145.

Hansen, M (1997). *Tabu search for multiobjective optimization: MOTS*, presented at the 13th MCDM conference, Cape Town, South Africa, January 6-10

Horn. J., Nafpliotis N. (1994). A niched Pareto genetic algorithm for multiobjective optimization. *Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence*, vol. 1, IEEE, New York, 82-87.

Kirkpatrick S., Gelatt C.D., Vecchi M.P. (1983). Optimization by simulated annealing, *Science*, **220**, 671-680.

Korhonen P., Salo S., Steuer R.E. (1997). A heuristic for estimating nadir criterion values in multiple objective linear programming. *Operations Research*, **45**, 5, 751-757.

Laarhoven P.J.M., Aarts E.H.L. (1987). *Simulated Annealing: Theory and Applications*, Reidel, Dordrecht.

Landsowne Z. F. (1996). Ordinal ranking methods for multicriterion decision making. *Naval Research Logistics*, **43**, 613-627.

Morita H., Katoh N. (to appear). A genetic algorithm for bicriteria scheduling problems. *Journal of Heuristics*.

Rosenthal R.E. (1985), Principles of multiobjective optimization. *Decision Sciences*, **16**, 133-152.

Serafini P. (1987). Some considerations about computational complexity for multiobjective combinatorial problems, *LNEMS*, **294**, 222-232, Springer Verlag.

Serafini P. (1994). Simulated annealing for multiple objective optimization problems. In: Tzeng G.H., Wang H.F., Wen V.P., Yu P.L. (eds), *Multiple Criteria Decision Making. Expand and Enrich the Domains of Thinking and Application*, Springer Verlag, 283-292.

Silver, E.A., Vidal R.V.V., de Werra D. (1980). A tutorial on heuristic methods, *European Journal of Operation Research*, **5**, 153-162.

Sun, M., Steuer R.E. (1996). InterQuad: An interactive quad tree based procedure for solving the discrete alternative multiple criteria problem, *European Journal of Operational Research*, **89**, 462-472.

Steuer R.E. (1986). *Multiple Criteria Optimization - Theory, Computation and Application*, Wiley, New York.

Schaffer J.D. (1985). Multiple objective optimization with vector evaluated genetic algorithms. In: J.J. Grefenstette (ed.), *Genetic Algorithms and Their Applications: Proceedings of the Third International Conference on Genetic Algorithms*, Lawrence Erlbaum, Hillsdale, NJ, 93-100.

Slowinski R. (1989). Multiobjective project scheduling under multiple-category resource constraint. In: Slowinski R., Weglarz J. (eds.), *Advances in Project Scheduling*, Elsevier, Amsterdam, 151-167.

Srinivas N., Deb K. (1994). Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, **2**, 2, 221-248.

Tuyttens D., Teghem J., Fortemps Ph, Van Nieuwenhuyse K. (to appear). Performance of the MOSA method for the bicriteria assignment problem. *Journal of Heuristics*.

Ulungu E.L. (1993). *Optimisation combinatoire multicritère: détermination de l'ensemble des solutions efficaces et méthodes interactives*, Thèse de doctorat, Université de Mons-Hainaut (Belgim).

Ulungu E.L., Teghem J. (1994). Multi-ojective Combinatorial Optimization Problems: A Survey. *Journal of Multi-Criteria Decision Analysis* **3**, 83-101.

Ulungu E.L., Teghem J., Fortemps Ph. (1995). Heuristics for multi-objective combinatorial optimization by simulated annealing. *Multiple Criteria Decision Making: Theory and Applications. Proceedings of the 6th National Conference on Multiple Criteria Decision Making*, Beijing (China), 228-238.

Viana, A. (1997). *Applicação de Meta-heurísticas Multiobjectivo ao Problema de Sequenciamento de Actividades com Restrições de Recursos*, Master thesis at the Faculty of Engineering, University of Porto, Portugal (In Portuguese).

Weistroffer H.R. (1985). Careful usage of pessimistic values is needed in multiple objectives optimization. *Operations Research Letters*, **4**, 1, 23-25.

White D.J. (1990). A bibliography on the applications of mathematical programming multiple-objective methods. *J. Opl Res. Soc.*, **41**, 669-691.