Minimal Sets of Quality Metrics

A. Farhang-Mehr, and S. Azarm

Department of Mechanical Engineering, University of Maryland
College Park, Maryland 20742, U.S.A.
{mehr@wam, azarm@eng}.umd.edu

Abstract. Numerous quality assessment metrics have been developed by researchers to compare the performance of different multi-objective evolutionary algorithms. These metrics show different properties and address various aspects of solution set quality. In this paper, we propose a conceptual framework for selection of a handful of these metrics such that all desired aspects of quality are addressed with minimum or no redundancy. Indeed, we prove that such sets of metrics, referred to as 'minimal sets', must be constructed based on a one-to-one correspondence with those aspects of quality that are desirable to a decision-maker.

1 Introduction

There are various multi-objective heuristic search techniques in the literature among which Multi-Objective Evolutionary Algorithms (MOEAs) have received significant attention [1][2][3][4]. These techniques usually generate a finite set of solutions to approximate the Pareto frontier of a multi-objective optimization problem. However, obtaining the 'best possible' set to represent the entire Pareto frontier is not always a trivial (or even an objectively-defined) task. Indeed, researchers have developed a myriad of techniques over the last few years to improve the quality of such solution sets in one way or another. Naturally, performance assessment and comparison study of such techniques have also gained much attention. One obvious way to compare MOEAs is to simply visualize the final sets of solutions and rely on intuitive judgments to decide on superiority of one technique to another. However, as discussed by Van Valdhuizen and Lamont [2], intuitive and visual assessment is not a reliable tool for comparison of different multi-objective optimization techniques. Especially, for problems with more than three-dimensions, visual judgment is either impossible or quite misleading yet it is the prevailing tool used by the researchers in the field.

More recently, there has been an emerging theme in the literature to quantitatively assess and compare the quality of non-dominated solution sets via *quality* (or *performance*) *metrics.* (A Non-Dominated Set, abbreviated as NDS, is defined as a set of mutually non-dominated solutions obtained from a MOEA to approximate the Pareto frontier.) These quality metrics generally assign an absolute or relative value (or a set of values) to a NDS to determine whether it is a 'good' representation of the actual Pareto frontier. For instance, Zitzler and Thiele [5] performed a comparative study of several multi-objective optimization methods using two metrics: "size of the dominated space" and "fraction of solutions dominated by the other set". Van Veldhuizen [6] introduced several quality metrics, such as: 'error ratio', 'generational distance', 'maximum Pareto frontier error' and 'overall non-dominated vector generation ratio' to assess different aspects of a solution set quality. Sayin [7] defined metrics for coverage, uniformity and cardinality to determine how 'good' a set of discrete solution points represents the true Pareto frontier. (For a recent review of the quality metrics and their shortcomings, see [8].)

Having many different quality metrics in the literature poses a new question to researchers: which metric or host of metrics must be used for an exhaustive (but not redundant) comparison study of different MOEAs? In fact, many of these metrics are coupled in the sense that they address a common aspect of quality. For example, there are several metrics in the literature that are claimed to assess 'diversity of solutions' in one way or another, including: 'spacing metric' [9]; 'overall non-dominated vector generation', 'overall non-dominated vector generation ratio' [6]; 'coverage', 'uniformity', 'cardinality' [7]; 'number of distinct choices', 'Pareto spread', and 'cluster' [10]. In a similar fashion, researchers developed numerous metrics to assess the closeness of solution sets to the Pareto frontier (see [4], for examples of these metrics). Obviously, many of these overlapping metrics are correlated, introducing redundancy in the comparison study of MOEAs. On the other hand, selecting too few of these quality metrics does not guarantee an exhaustive comparison with respect to all aspects of quality. Indeed, a desirable collection of quality metrics must be *minimal*, in the sense that: 1) There is at least one metric for every aspect of the solution set quality to guarantee an exhaustive performance assessment; 2) There exists minimum (or no) correlation among quality metrics to avoid redundancies (see Section 4 for a formal definition). Due to the subjective nature of quality metrics, and quality of a set of solutions in general, the above-mentioned properties may not be noticeable from the formulation of these metrics and require a more thorough investigation of the underlying concepts.

In this paper, we assume that the decision-maker's specific preferences are not known a priori, while the general aspects of interest are known. In other words, we would like to determine a minimal set of quality metrics that exhaustively addresses all desired aspects of quality without redundancy. Hansen and Jaszkiewicz [11] performed a similar study in which a family of outperformance relations was defined to compare Pareto-optimality of non-dominated solutions sets (see Section 2 for more on this). These relations account for pairs of NDSs where one solution set is objectively better (based only on the notion of dominance) than the other set and thus, they establish *strict partial orders* among NDSs. (A strict partial order in the set of all possible NDSs is defined as an irreflexive, antisymmetric, and transitive relation that compares some and not every given pair of NDSs. For details see almost any book on set theory, e.g. [12].) In other words, not all solution sets are comparable in this way, but at least one can verify the validity of a quality metric by examining its compatibility with these relations. Based on this idea, Zitzler et al. [13] proposed a theoretical framework to investigate the compatibility and completeness of different comparison methods and derived a set of theoretical restrictions for the existence of compatible and complete *unary* quality metrics. (A unary quality metric measures the *absolute* goodness of a NDS.) In fact, they prove that a finite combination of unary metrics that is compatible and complete at the same time does not exist in general. They also mention that this limitation does not apply to *binary* quality metrics (i.e., those that quantify only the *relative* goodness of two NDSs), wherein one can construct compatible and complete metrics with respect to any dominance relations.

The focus of this paper is on *binary* quality metrics and their correspondence with outperformance relations. More specifically, the contributions of this paper are as follows.

- While outperformance relations in the previous works address only one aspect of the quality, namely closeness to the Pareto frontier, in this paper, we propose the more general notion of *Excellence Relations*. These relations establish strict partial orders in the set of all NDSs with respect to different aspects of quality.

- Unlike previous studies by Hansen and Jaszkiewicz [11] and Zitzler et al. [13] who considered several outperformance relations (e.g., weak/strong/ outperformance) to address the closeness of a NDS to the Pareto frontier; here we assume that a decision-maker provides a *combination* of excellence relations, each addressing a *distinct* aspect of quality (e.g., a dominance relation for closeness to the Pareto frontier, a coverage relation for the coverage of the Pareto frontier, and so on). Then we define minimal sets of binary quality metrics, and investigate their ability to address all of these excellence relations at the same time (e.g., whether a set of quality metrics can address both closeness and coverage of the Pareto frontier.)

- We extend the definition of compatibility with outperformance relations [11][13] to excellence relations and show that: Given two uncorrelated binary quality metrics, they cannot be both compatible with the same excellence relation.

- Finally, one may find several minimal combinations of quality metrics, however, all of them possess the following important property (referred to as Minimality Lemma in Section 4): if one assumes n excellence relations with certain properties, a minimal set of binary quality metrics contains n and only n metrics, each compatible with exactly one excellence relation, i.e., there is a one-to-one correspondence. This property significantly narrows the search for minimal sets.

Although there are relatively fewer binary quality metrics in the literature as compared to unary metrics [13], the above property makes them very attractive for a comparison study of multiobjective optimization algorithms, mainly because one could select a minimal set of binary metrics to address all desired aspects of quality exhaustively and distinctly. (This is not possible in general with unary metrics, see [13].)

2 Excellence Relations

As mentioned in Section 1, Hansen and Jaszkiewicz [11] defined the outperformance relations to establish a strict partial order among NDSs, where some pairs of solution sets are objectively comparable in terms of Pareto optimality (or dominance). Here, the definition of a strong outperformance relation is given for the completeness of this paper. (In addition to the strong relation, Hansen and Jaszkiewicz [11] also defined weak and complete outperformance relations.)

Definition 1. A non-dominated set A strongly outperforms a non-dominated set B, denoted by AR_OB, iff (i.e., if and only if) $A \neq B$ and (in the objective space) for each $y \in B$, there exists $x \in A$ such that $x \succeq y$.

The notation '$\succeq$' in $x \succeq y$ indicates that a point x is either equal to or dominates a point y with respect to all objectives. Fig. 1 demonstrates two non-dominated solution sets generated for a 2-objective maximization problem. According to Definition 1, we observe that: AR_OB.

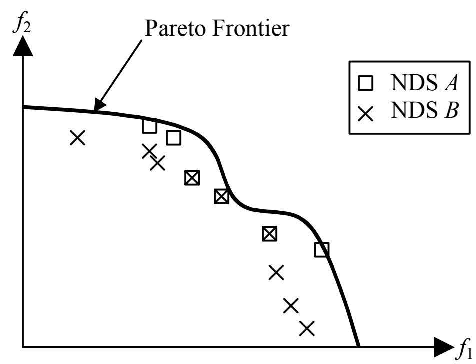

Fig. 1. AR_OB: Set A strongly outperforms Set B.

As mentioned before, not every two sets are comparable in this way. If we denote the set of all possible non-dominated sets by U, then the above relation constructs a strict partial order in U. Here we define the *partially ordered domain* of this comparison in UxU, shown by Λ_{Ro}, as the set of all 2-tuples of NDSs that are comparable via R_O, i.e.,

$$\Lambda_{Ro} = \{(A,B) \in U \mathrm{x} U \mid \text{either } AR_OB \text{ or } BR_OA\}$$

This relation by itself does not provide a tool to compare any given pair of NDSs, however, it can be used to verify the validity of quality metrics. In fact, this relation is based only on the concept of dominance (or closeness to the Pareto frontier), and therefore, if a quality metric aims at comparing NDSs in terms of dominance, it must be compatible with this relation in the first place. (The formal definition of compatibility is given later in the paper.)

We mentioned that the outperformance relation accounts for the Pareto optimality of solution sets. However, there are other aspects of quality that are especially important in the assessment of solution sets obtained from MOEAs, e.g., diversity of the solution sets, extent of the Pareto frontier that is covered by the solutions. Very similar to the outperformance relation, one can collect all 2-tuples of NDSs that are objectively comparable with respect to any aspect of quality and construct a strict partial order accordingly. This prompts for the definition of a more general concept, i.e., *excellence relations*, as proposed in this paper.

Definition 2. An excellence relation, denoted by R, is defined as a strict partial order in U that relates all non-dominated sets that are objectively comparable with respect to a common aspect of quality. The partially ordered domain of R in $U \times U$, denoted by Λ_R, is defined as: $\Lambda_R = \{(A,B) \in U \times U \mid \text{either } ARB \text{ or } BRA\}$.

For example, an outperformance relation is an excellence relation with respect to dominance. As another example, in the following we define a new excellence relation (i.e., *coverage relation*) to address a different aspect of quality: *coverage* (i.e., the span of the solution set over the Pareto frontier). In this example, it is assumed that all objective functions are positive.

Definition 3. A non-dominated set, B, is strictly superior to another non-dominated set, A, in terms of *coverage*, denoted by BR_CA, iff all solution points of Set A are contained in a convex cone generated by Set B, while there exists at least one solution in Set B that is not contained in a convex cone generated by Set A.

Here the convex cone generated by a solution set $A=\{\mathbf{a}_1, \mathbf{a}_2, \ldots\}$ is defined as all nonnegative linear combinations of $\mathbf{a}_i$'s, i.e., $\{\mathbf{v} \mid \mathbf{v} = \Sigma w_i \mathbf{a}_i\ ;\ \ w_i \geq 0\}$ [14]. The shaded area in Fig. 2(b) demonstrates the convex cone generated by Set B. This cone clearly contains all solution points of Set A. In contrast, the convex cone of Set A (Fig. 2(a)) does not include all solution points of Set B. Thus, according to Definition 3, we have: BR_CA. Finally, note that in this definition it is assumed that all objectives are to be maximized. Also, the nadir point (i.e., the lower bound of the Pareto frontier) is assumed to be located at the origin of the Cartesian objective space. If these assumptions do not hold for a given problem, one could always transform the objectives to meet these assumptions.

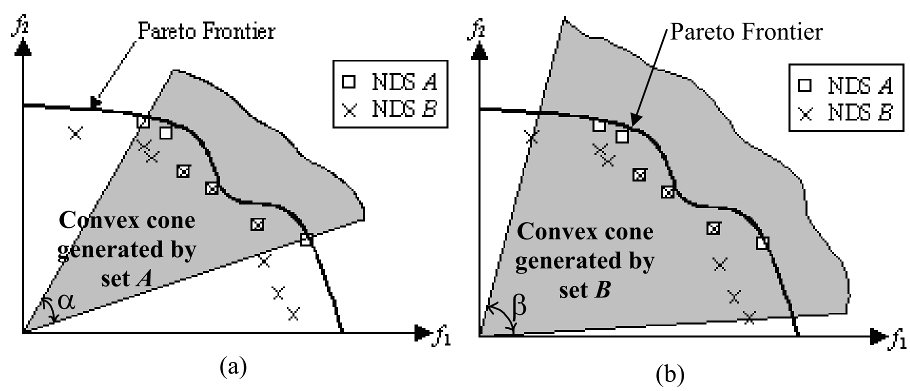

Fig. 2. Convex cones generated by: (a) NDS A; and (b) NDS B.
According to Definition 3: BR_CA.

As expected, the excellence relation for this example establishes a strict partial order in U, i.e., objectively compares a pair of NDSs in terms of coverage.

Note that the first step in a comparison study of NDSs is to determine different aspects of quality that are of interest to the decision maker (e.g., Pareto optimality, coverage, diversity, and so on). Then, we collect all pairs of NDSs that are objectively comparable with respect to any of these quality aspects. These collections establish strict partial orders in U that we referred to as excellence relations. However, not all solution sets are comparable using these relations, and one must formulate quality metrics that enable an exhaustive comparison of all non-dominated solution sets. Each binary quality metric constructs a *total order* that quantitatively compares all NDSs pairs in U. (A total order in the set of all possible NDSs is defined as an irreflexive, antisymmetric, and transitive relation that compares all pairs of NDSs in the set U.) The correspondence of these quality metrics (i.e., total order) with excellence relations (i.e., partial orders) is the subject of the next section. We prove two key theorems that will be used later in Section 4 to investigate the properties of a minimal set of quality metrics.

3 Compatibility, Concordance, and Correlation

As briefly explained in Section 1, a binary quality metric compares the quality of two NDSs and returns a relative value, while a unary metric computes an absolute value for quality. The focus of this paper is only on binary metrics, i.e., if A and B are two NDSs, then $Q(A,B)$ returns a scalar that reflects how much better set A is when compared to set B. (The arguments of this paper also applies to other metrics that can be transformed into the above binary format.) Moreover, Q is assumed to be symmetric and homogeneous, meaning that: $Q(B,A) = -Q(A,B)$. Note that a symmetric metric as defined by Knowles and Corne [4] is: $Q'(B,A) = C - Q'(A,B)$. This latter format can be easily transformed into a homogenous symmetric metric by assuming: $Q(A,B) = C/2 - Q'(A,B)$. (See also [13] for a discussion on the properties of such metrics.) Without loss of generality, it is assumed that $Q(A,B) > 0$ iff A is strictly better than B. A binary quality metric (hereafter, 'metric' implies 'binary metric' unless otherwise stated) as defined above constructs a total order in U and compares any two non-dominated sets on a quantitative basis.

Now, assume a symmetric and homogeneous quality metric, Q, that compares any two given NDSs in terms of a certain aspect of quality. If R is an excellence relation that addresses the same aspect of quality, it is natural to expect Q to be compatible with R, as defined formally in the following. (This is very similar to the definition of compatibility with outperformance relation given in [11] and [13]; tailored for symmetric metrics, and generalized for excellence relations.)

Definition 4. A symmetric homogeneous binary metric, Q, is *compatible* with an excellence relation, R, iff: for any given pair of non-dominated sets A and B such that ARB, we also have $Q(A,B) > 0$, which implies set A has a better quality than set B. (The compatibility of metric Q with relation R is denoted as $Q \sim R$ in this paper.)

Knowles [8] studied the compatibility of several unary and binary quality metrics with respect to an outperformance relation. The same study can be carried out to determine the compatibility of those metrics with respect to any other excellence relations, such as the coverage relation, R_C. Obviously, if a metric is not intended to compare two NDSs in terms of a certain aspect of quality (e.g., a diversity assessment metric is not intended to account for closeness to the Pareto frontier) it does not need to be compatible with that excellence relation. In fact, we will prove that each quality metric in a minimal set must be compatible with one and only one excellence relation. Nevertheless, this compatibility is dependant on the definition of the excellence relation itself. Before formally defining minimal sets of quality metrics and their correspondence with excellence relations, in the following we introduce the notion of *concordance* among excellence relations.

Definition 5. Two excellence relations R and R' are *concordant* iff for each $(A,B) \in \Lambda_R \cap \Lambda_{R'}$ such that ARB, we also have $AR'B$.

Put another way, R and R' are concordant iff there do not exist two non-dominated sets A and B such that: ARB and $BR'A$. Concordance basically implies that the two excellence relations *cannot* work against each other. If two excellence relations are referring to different aspects of quality (e.g., diversity and Pareto optimality), there are always examples of NDSs that are better with respect to one aspect of quality and worse with respect to another, and therefore, those relations are not concordant (or are *non-concordant*). In contrast, if being better in terms of one relation always implies better with respect to another, then it implies that the two relations have essentially the same nature (i.e., refer to the same aspect or notion of quality) and thus are concordant. The two excellence relations of Section 2, i.e., R_C and R_O, are non-concordant because set A in Figures 1 and 2 is better than set B in terms of the outperformance relation (AR_OB), but worse in terms of coverage (BR_CA). Moreover, from the above definition two relations R and R', such that $R \subset R'$, are always concordant. Therefore, the family of outperformance relations defined by Hansen and Jaszkiewicz [11] are concordant because: complete outperfromance is a subset of strong outperformance, which in turn is a subset of weak outperformance. Therefore, although these relations are different, they are concordant according to Definition 5.

Concordance is a very strong assumption in the sense that if $(A,B) \in U \times U$ is comparable via two given concordant excellence relations, the outcome of the comparison from the first relation is always the same as that of the second one. On the other hand, non-concordance is a weak assumption in the sense that: two excellence relations are non-concordant even if there exists only one pair of non-dominated solution sets, (A,B), such that A is better than B with respect to one relation and worse with respect to another. Finally, the following theorem demonstrates an important property of non-concordant relations.

Theorem 1. There does not exist a symmetric and homogeneous quality metric that is compatible with two (or more) non-concordant excellence relations.

Proof. For the sake of contradiction, suppose there exists a quality metric, Q, which is compatible with two non-concordant excellence relations, namely R and R'. Since R and R' are non-concordant, there exists a pair of non-dominated sets, namely $A,B \in U$, $(A \neq B)$, such that ARB and $BR'A$. But since Q is compatible with R, from ARB we conclude $Q(A,B) > 0$. Similarly, Q is compatible with R' and from $BR'A$ we have $Q(B,A) > 0$, which is a contradiction because $Q(A,B) = -Q(A,B)$.

In fact, this theorem is somewhat intuitive from the definition of concordance and compatibility: a metric cannot simultaneously be compatible with two excellence relations that work against each other. For example a symmetric and homogeneous coverage metric, which is compatible with R_C, is necessarily incompatible with R_O (i.e., recall that R_O and R_C are non-concordant according to Definition 5 and Figures 1 and 2). The above theorem indicates that there must be at least one separate metric in a minimal set to individually address each aspect of quality, e.g., at least one metric compatible with diversity, another one compatible with Pareto optimality, and so on. Moreover, in the following we show that any two given metrics that address the same aspect of quality are necessarily correlated.

Theorem 2. If two symmetric homogeneous metrics are both compatible with an excellence relation, R, they are positively correlated within Λ_R.

Proof. Assume two symmetric and homogeneous metrics, Q and Q', are both compatible with R. Then the covariance of Q and Q' within Λ_R can be written as:

$$\mathrm{Cov}\,[Q(A,B), Q'(A,B)] = \langle Q(A,B)Q'(A,B)\rangle - \langle Q(A,B)\rangle \langle Q'(A,B)\rangle \quad ; (A,B) \in \Lambda_R$$

where the expected value of Q within Λ_R, i.e., $\langle Q(A,B)\rangle$, is zero, because Q is symmetric, and therefore, for each $(A,B) \in \Lambda_R$, we also have $(B,A) \in \Lambda_R$, and $Q(A,B) = -Q(B,A)$. Similarly: $\langle Q'(A,B)\rangle = 0$. On the other hand, Q and Q' are both compatible with R, and therefore, for each $(A,B) \in \Lambda_R$, $Q(A,B)$ and $Q'(A,B)$ have the same sign (both negative or both positive). Thus, $\langle Q(A,B)Q'(A,B)\rangle$ is strictly positive, and the theorem follows.

Note that being 'positively correlated' is a necessary and not sufficient condition for 'compatibility with the same relation', i.e., two metrics that are not compatible with the same relation are not guaranteed to be uncorrelated. In the following section, we take advantage of the above theorems to investigate the properties of minimal sets of quality metrics.

4 Minimal Sets of Quality Metrics

In the following, we formally state the minimality conditions for a set of quality metrics.

Definition 6. A set of quality metrics, namely Γ, is said to be *minimal* with respect to a given set of non-concordant excellence relations, Φ, iff:

1- Each quality metric, $Q \in \Gamma$, is compatible with at least one excellence relation in Φ. Formally, $\forall Q \in \Gamma : \exists R \in \Phi$ such that *Q~R*.

2- For each excellence relation in Φ, there is at least one compatible quality metric in Γ. Formally, $\forall R \in \Phi : \exists Q \in \Gamma$ such that *Q~R*.

3- There is minimum (or no) correlation among quality metrics of Γ within the partially ordered domain of excellence relations.

The first property rejects unnecessary metrics that are not compatible with any of the excellence relations. The second property guarantees that Γ is exhaustive, in the sense that it addresses all aspects of quality that are of any interest to the decision-maker (i.e., expressed via excellence relations in Φ). The last property eliminates or minimizes the redundancy within the set, i.e., the selected quality metrics should have minimum or no correlation. From this definition and Theorems 1 and 2 we observe the following.

Minimality Lemma. Given a set of *n* non-concordant excellence relations, Φ, a corresponding minimal set of symmetric and homogeneous metrics, Γ, contains *n* and only *n* quality metrics. (Also, there is a one-to-one correspondence between Γ and Φ.)

Proof. From Theorem 1, a metric in Γ cannot be compatible with more than one excellence relations in Φ (because the excellence relations in Φ are non-concordant). Therefore, following the first property of Definition 6, each metric is compatible with exactly one excellence relation. Also, Theorem 2 indicates that two uncorrelated metrics cannot be compatible with the same excellence relation (because otherwise they would be positively correlated according to Theorem 2). Therefore, following the second property in Definition 6, there is a one-to-one correspondence between Γ and Φ.

Minimality Lemma suggests a recipe with a set of steps to be followed for the selection of a minimal set of quality metrics, as given next.

Step 1. The general aspects of performance that are of interest to the decision-maker are determined or presumed (e.g., closeness to the Pareto frontier, coverage, diversity, etc.)

Step 2. For a given aspect of quality, a strict partial order is established in *U*. In other words, an excellence relation is constructed that accounts for of all pairs of NDSs that are objectively comparable with respect to that given aspect of performance. For instance, outperformance relation addresses the closeness to the

Pareto frontier; coverage relation of Definition 3 addresses coverage of the set, and so on. If the aspects of quality are defined properly in Step 1, these excellence relations are non-concordant (because if they are indeed referring to different aspects of quality, there exists a non-dominated sets that is better than another set with respect to one excellence relation and worse with respect to another). These non-concordant excellence relations constitute Φ.

Step 3. Suppose Φ consists of n excellence relations. To construct a minimal set, we select one and only one quality metric, compatible with each excellence relation in Φ (recall Minimality Lemma). Note that since Theorem 2 provides only a necessary condition for being uncorrelated, establishing a one-to-one compatibility correspondence does not guarantee a minimal set. However, it rules out many of non-minimal collections of metrics and significantly narrows the search for minimum correlations. The result is a set of size n of performance assessment metrics, i.e., Γ.

Γ constructs exactly n total orders in U, and can be used to compare any two given NDSs in terms of the quality aspects expressed in Step 1, and formulated as excellence relations (i.e., strict partial orders) in Step 2. It exhaustively and distinctly covers all desired aspects of quality, without unnecessary correlation among metrics. According to the Minimality Lemma, collections of more than n metrics are necessarily correlated, while less than n metrics cannot distinctly address all desired aspects of quality.

As an example, if the decision-maker desires only two aspects of quality: 1) Pareto optimality, and 2) Coverage, Φ consists of exactly two relations: $\Phi = \{R_O, R_C\}$ (recall that these two relations are non-concordant). A minimal set of size two of quality metrics should then be selected from the pool of existing metrics, $\Gamma=\{Q_1, Q_2\}$, such that $Q_1 \sim R_O$, and $Q_2 \sim R_C$. No other combination can address both of these quality aspects without redundancy. According to the Minimality Lemma, the same argument holds for any number of non-concordant excellence relations in Φ. Finally, the above guideline is only an abstraction of the notion of quality metrics and their desired properties, and therefore, it does not define or formulate new metrics by itself.

5 Practicality of the Minimality Lemma

Although the theoretical framework in this paper provides an approach for an objective selection of minimal sets of binary quality metrics, its real-world application may be hindered by several factors:

- A decision-maker may not be able to state his/her idea of 'quality of a solution set' explicitly in the form of excellence relations.

- Even if the decision-maker is able to state a set of excellence relations, Φ, as a basis for quality assessment, it is not always possible to find a corresponding set of minimal binary and symmetric quality metrics, Γ, such that each metric in Γ is compatible with an excellence relation in Φ.

For example consider a case where Φ consists of the two non-concordant excellence relations of Section 2, i.e., $\Phi = \{R_O, R_C\}$. Table 1 shows the compatibility of several quality metrics in the literature with these relations.

Table 1. Compatibility of quality metrics with $\Phi = \{R_O, R_C\}$.
('Y' indicates compatibility)

	Strong Outperformance Relation (R_O)	Coverage Relation (R_C, Definition 3)
$\mathcal{C}$ metric [4]	Y	N
Inferiority Index (*InfI*) [15]	Y	N
k-th Objective Pareto Spread (OS_k) [10]	N	N
Entropy [16]	N	N

(See [8] for a comprehensive compatibility analysis of common quality metrics with outperformance relations.) Note that Wu and Azarm's k-th Objective Pareto Spread (OS_k) is a unary metric [10]. So, in this paper, we revise this metric, i.e., compute the difference in the value of OS_k between two given NDSs to create a binary metric: $OS_k(A,B)=OS_k(A)-OS_k(B)$, which constructs a cardinal total order in U. Table 1 shows that only a portion of the examined metrics are compatible with R_O, while we were not able to find any quality metric in the literature to be compatible with R_c. Indeed, as shown by Knowles [8], a relatively small portion of the existing quality metrics is compatible with one of the outperformance relations introduced by Hansen and Jaszkiewicz [11]. Does this mean that we cannot obtain a set of quality metrics to address these two aspects of quality simultaneously? Zitzler et al. [13] show that this is in fact impossible in general with a finite number of unary quality metrics. As shown in this paper, however, this is in theory possible for the case of binary quality metrics, although it may not be practical because such metrics may not actually exist in the literature. In a case like this, one may go back to the initial set of the excellence relations, Φ, and try to compromise these relations such that a corresponding minimal set of quality metrics can be found. Alternatively, one may try to create new metrics to match the compatibility criterion with a given excellence relation. In the following, for example, we propose a new binary quality metric to quantify the difference between the spans of two non-dominated sets as a measure of extent of coverage. Later we show that this metric is compatible with R_c.

Definition 7. *Binary coverage metric*, denoted by $Q_c(A,B)$, is defined as:

$$Q_c(A,B)= \inf \{(\mathbf{b}_1.\mathbf{b}_2)/(\|\mathbf{b}_1\|\ \|\mathbf{b}_2\|) \text{ s.t. } \mathbf{b}_1,\mathbf{b}_2 \in \mathbb{B}\} - \inf \{(\mathbf{a}_1.\mathbf{a}_2)/(\|\mathbf{a}_1\|\ \|\mathbf{a}_2\|) \text{ s.t. } a_1,a_2 \in \mathbb{A}\}$$

where $\mathbb{A}$ and $\mathbb{B}$ are the convex cones generated by solution sets A and B. The term: $\inf \{(\mathbf{a}_1.\mathbf{a}_2)/(\|\mathbf{a}_1\|\ \|\mathbf{a}_2\|) \text{ s.t. } a_1,a_2 \in \mathbb{A}\}$ measures the cosine of the largest possible angle between two vectors in the convex cone generated by the solution set A. This for example corresponds to $cos(\alpha)$ in Fig. 2, and $Q_c(A,B)= cos(\beta) - cos(\alpha) < 0$. We use this as a measure of maximum span of the solution sets on the Pareto frontier. Q_c is compatible with R_c because: if BR_cA, we have $\mathbb{A} \subset \mathbb{B}$, and therefore, the second term in the above equation is greater than the first term. Therefore, $Q_c(A,B)<0$. Similarly, if AR_cB we obtain $Q_c(A,B)>0$, and compatibility follows.

From Table 1, Definition 7, and minimality lemma we observe that $\Gamma = \{InfI; Q_c\}$ is a candidate minimal set of binary quality metrics with respect to $\Phi = \{R_O, R_C\}$.

Other than defining new compatible metrics, one could also modify the definition of excellence relations in Φ to make the existing quality metrics compatible. If the decision-maker modifies the definition of the coverage excellence relation (Definition 3) for example, Γ may or may not remain minimal. In the following, for example, we introduce a modified definition for coverage excellence relation that makes Wu and Azarm's OS_k metric compatible.

Definition 8 (Modified Coverage Relation; Compare to Definition 3). In a normalized multi-objective maximization, a non-dominated set $B=\{\mathbf{b}_1, \mathbf{b}_2, \ldots\}$ is strictly superior to another non-dominated set $A=\{\mathbf{a}_1, \mathbf{a}_2, \ldots\}$ in terms of a *modified coverage*, denoted by $BR'_C A$, iff: $\max_{i,j}(b_i^k - b_j^k)$ is strictly greater than $\max_{i,j}(a_i^k - a_j^k)$ for all k's (where b_i^k refers to the k-th objective value of the i-th solution in B).

OS_k is compatible with R'_C, and therefore, a combination of this quality metric and Inferiority Index (*InfI*) [15], i.e., $\Gamma' = \{InfI, OS_k\}$, is a candidate minimal set with respect to $\Phi' = \{R_O, R'_C\}$. In contrast, $\Gamma = \{InfI, Q_c\}$ which is minimal with respect to $\Phi = \{R_O, R_C\}$, is not minimal with respect to Φ', because Q_c is not compatible with R'_C (See Definition 7 for Q_C). Note that the process of defining and redefining Φ becomes increasingly difficult as more excellence relations are included. Nonetheless, it provides a formal platform and an objective starting point for selection of binary quality metrics.

6 Concluding Remarks

In this paper, we presented a theoretical framework for selection of minimal sets of quality metrics that can be used for a comparison study of different MOEAs. These metrics exhaustively account for all desired aspects of quality in non-dominated solution sets (obtained by MOEAs) without redundancy. In this framework, once the decision-maker's desired aspects of performance are determined, it is necessary to find all pairs of non-dominated sets that are objectively comparable. This in turn constructs partial orders in the set of all possible non-dominated sets, referred to as excellence relations in this paper. We proved that there is a one-to-one compatibility correspondence between these excellence relations (partial orders) and a minimal set of quality metrics (total orders), i.e., for each excellence relation there is one and only one compatible quality metric in a minimal set. This important result (referred to as Minimality Lemma) helps the decision-maker select a minimal set among the existing quality metrics in the literature, and thus, enables a quantitative and objective comparison of the solution sets obtained from different MOEAs.

7 Acknowledgments

The work presented in this paper was supported in part by the U.S. National Science Foundation, Grant number 0112767. This support is gratefully acknowledged.

References

1. Zitzler, E.: 1999, Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications, Ph.D. Dissertation, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland (1999)

2. Van Valdhuizen, D. A., Lamont G. B.: Multiobjective Evolutionary Algorithms: Analyzing the State-of-the-Art," Evolutionary Computation, 8(2), (2000) 125-147

3. Deb, K.: Multi-objective Optimization Using Evolutionary Algorithms. John Wiley and Sons, Chichester, UK (2001)

4. Knowles, J. D., Corne, D. W.: On Metrics for Comparing Non-Dominated Sets, Proceedings of the 2002 Congress on Evolutionary Computation Conference, (2002) 711-716

5. Zitzler, E., Thiele, L., 1998, "Multiobjective Optimization Using Evolutionary Algorithms – A Comparative Study," Proc. 5th International Conference: Parallel Problem Solving from Nature – PPSN V, Springer, Amsterdam, The Netherlands, (1998) 292-301

6. Van Veldhuizen, D. A.: Multiobjective Evolutionary Algorithm: Classification, Analyses and New Innovations, Ph.D. Dissertation, Dept. of Electrical and Computer Engineering, Air Force Institute of Technology, Wright-Patterson AFB, OH (1999)

7. Sayin, S.: Measuring the Quality of Discrete Representations of Efficient Sets in Multiple Objective Mathematical Programming, Mathematical Programming, 87, (2000) 543-560

8. Knowles, J. D.: Local-Search and Hybrid Evolutionary Algorithms for Pareto Optimization, Ph.D. Dissertation, University of Reading, Department of Computer Science, Reading, U.K. (2002)

9. Schott, J. R.: Fault Tolerant Design Using Single and Multicriteria Genetic Algorithm Optimization, M.S. Thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA (1995)

10. Wu, J., Azarm, S.: Metrics for Quality Assessment of a Multiobjective Design Optimization Solution Set," Transactions of the ASME, Journal of Mechanical Design, 123, (2001) 18-25
11. Hansen, M. P., Jaszkiewicz, A.: Evaluating the Quality of Approximations to the Non-dominated Sets, Technical Report IMM-REP-1998-7, Technical University of Denmark, Lyngby, Denmark (1998)
12. Enderton, H. B.: Elements of Set Theory. Academic Press, Orlando, FL (1977)
13. Zitzler, E., Laumanns, M., Thiele, L., Fonseca, C. M., Grunert da Fonseca V.: Why Quality Assessment of Multiobjective Optimizers Is Difficult. Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2002) (2002) 666-674
14. Steuer, R. E.: Multiple Criteria Optimization: Theory, Computation, and Application, John Wiley & Sons, NY (1986)
15. Farhang-Mehr, A., Wu, J., Azarm, S.: Some Preliminary Results on the Development and Comparison of a New Multi-Objective Genetic Algorithm. Paper No. DETC2001/DAC-21021, CD-ROM Proceedings of the ASME Design Automation Conference, Pittsburgh, PA (2001)
16. Farhang-Mehr, A., Azarm, S.: Diversity Assessment of Pareto Optimal Solutions: An Entropy Approach. IEEE 2002 World Congress on Evolutionary Computation, Honolulu, Hawaii (2002)