On the Performance Metrics of Multiobjective Optimization

Conference Paper *in* Lecture Notes in Computer Science · July 2012

DOI: 10.1007/978-3-642-30976-2_61

3 authors:

Shi Cheng
Shaanxi Normal University

Yuhui Shi
Southern University of Science and Technology

Quande Qin
Shenzhen University

On the Performance Metrics of Multiobjective Optimization*

Shi Cheng[1,2], Yuhui Shi[2], and Quande Qin[3]

[1] Department of Electrical Engineering and Electronics,
University of Liverpool, Liverpool, UK
[2] Department of Electrical & Electronic Engineering,
Xi'an Jiaotong-Liverpool University, Suzhou, China
shi.cheng@liverpool.ac.uk, yuhui.shi@xjtlu.edu.cn
[3] College of Management, Shenzhen University, Shenzhen, China
qinquande@gmail.com

Abstract. Multiobjective Optimization (MOO) refers to optimization problems that involve two or more objectives. Unlike in the single objective optimization, a set of solutions representing the tradeoff among the different objects rather than an unique optimal solution is sought in MOO. How to measure the goodness of solutions and the performance of algorithms is important in MOO. In this paper, we first review the performance metrics of multiobjective optimization and then classify variants of performance metrics into three categories: set based metrics, reference point based metrics, and the true Pareto front/set based metrics. The properties and drawbacks of different metrics are discussed and analyzed. From the analysis of different metrics, an algorithm's properties can be revealed and more effective algorithms can be designed to solve MOO problems.

Keywords: Multiobjective Optimization, Performance Metrics, Pareto Front/Set, Reference Point.

1 Introduction

An optimization problem in $\mathbb{R}^n$, or simply an optimization problem, is a mapping $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$, where $\mathbb{R}^n$ is termed as decision space [2] (or parameter space, problem space), and $\mathbb{R}^k$ is termed as objective space. Optimization problems can be divided into two categories based on the value of k. If $k = 1$, the problems are called Single Objective Problems (SOPs); if $k > 1$, problems are called Multiobjective Problems (MOPs), and specially, problems are called Many Objective Problems when k is large than 2 or 3 [1].

One of the main differences between SOPs and MOPs is that MOPs constitute a multidimensional objective space. In addition, a set of solutions representing the tradeoff among the different objectives rather than an unique optimal

* The authors' work is partially supported by National Natural Science Foundation of China under grant No.60975080.

solution is sought in Multiobjective optimization (MOO). How to measure the goodness of solutions and the performance of algorithms is important in MOO.

Although many articles have discussed metrics on multiobjective optimization [6, 9, 15], there is no one metric that can overwhelm others. It is necessary to have more analyses and discussions. In this paper, we classify variants of performance metrics into three categories: set based metrics, reference point based metrics, and the true Pareto front/set based metrics. The properties and drawbacks of different metrics are discussed and analyzed. From the analyses of the metrics, more effective algorithms can be designed to solve multiobjective problems.

This paper is organized as follows. Section 2 reviews the basic definitions of multiobjective optimization. Section 3 introduces set based metrics, which include outperformance relations, $\mathcal{C}$ measure, and $\mathcal{M}_3$ measure. Section 4 introduces a reference point based metrics, which include $\mathcal{S}$ measure (hypervolume) and $\mathcal{D}$ measure. Section 5 introduces the true Pareto front/set based metrics, which include inverted generational distance metric, hypervolume difference metric and spacing Δ metric. Section 6 concludes with some remarks and future research directions.

2 Multiobjective Optimization

A general *multiobjective optimization problem* can be described as a vector function $\mathbf{f}$ that maps a tuple of n parameters (decision variables) to a tuple of m objectives. Without loss of generality, minimization is assumed throughout this paper.

$$\begin{aligned} \text{minimize} \quad & \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \cdots, f_m(\mathbf{x})) \\ \text{subject to} \quad & \mathbf{x} = (x_1, x_2, \cdots, x_n) \in \mathbf{X} \\ & \mathbf{y} = (y_1, y_2, \cdots, y_m) \in \mathbf{Y} \end{aligned}$$

where $\mathbf{x}$ is called the *decision vector*, $\mathbf{X}$ is the *decision space*, $\mathbf{y}$ is the *objective vector*, and $\mathbf{Y}$ is the *objective space*, and $\mathbf{f} : \mathbf{X} \to \mathbf{Y}$ consists of m real-valued objective functions.

Let $\mathbf{u} = (u_1, \cdots, u_m)$, $\mathbf{v} = (v_1, \cdots, v_m) \in \mathbf{Y}$, be two vectors, $\mathbf{u}$ is said to dominate $\mathbf{v}$ (denoted as $\mathbf{u} \preceq \mathbf{v}$), if $u_i \leq v_i, \forall i = 1, \cdots, m$, and $\mathbf{u} \neq \mathbf{v}$. A point $\mathbf{x}^* \in \mathbf{X}$ is called Pareto optimal if there is no $\mathbf{x} \in \mathbf{X}$ such that $\mathbf{f}(\mathbf{x})$ dominates $\mathbf{f}(\mathbf{x}^*)$. The set of all the Pareto optimal points is called the *Pareto set* (denoted as PS). The set of all the Pareto objective vectors, $PF = \{f(x) \in X | x \in PS\}$, is called the *Pareto front* (denoted as PF).

In a multiobjective optimization problem, we aim to find the set of optimal tradeoff solutions known as the Pareto optimal set. Pareto optimality is defined with respect to the concept of nondominated points in the objective space.

The optimization goal of an MOP consists of three objectives: (1) The distance of the resulting nondominated solutions to the true optimal Pareto front should be minimized; (2) A good (in most cases uniform) distribution of the obtained solutions is desirable; (3) The spread of the obtained nondominated solutions should be maximized, i.e., for each objective a wide range of values should be covered by the nondominated solutions.

3 Set based Metrics

In multiobjective optimization, a set of solutions representing the tradeoff among the different objectives rather than an unique optimal solution as sought. It's a straightforward way to measure solutions on set based metrics. These metrics are a kind of quality measures, which are difficult to measure the goodness of solutions.

3.1 Outperformance relations

Three kinds of outperformance relations are introduced in [5] to express the relations between two sets of internally nondominated objective vectors. The relations are as follow:

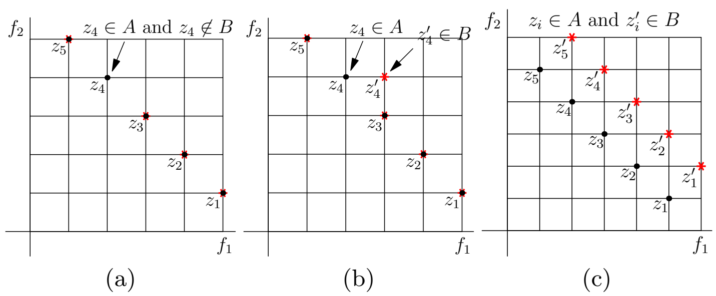

Fig. 1. Examples of outperformance relations, $\bullet \in \mathbf{A}$, and $* \in \mathbf{B}$: (a) **A** weak outperformance **B**, (b) **A** strong outperformance **B**, (c) **A** complete outperformance **B**.

- Weak outperformance: $\mathbf{ND}(\mathbf{A} \cup \mathbf{B}) = \mathbf{A}$ and $\mathbf{A} \neq \mathbf{B}$. **A** *weakly outperforms* **B** if all solutions in **B** are *contained* in **A** and there is at least one solution in **A** that is not contained in **B**, e.g. Fig. 1 (a).
- Strong outperformance: $\mathbf{ND}(\mathbf{A} \cup \mathbf{B}) = \mathbf{A}$ and $\mathbf{B} \backslash \mathbf{ND}(\mathbf{A} \cap \mathbf{B}) \neq \emptyset$. **A** *strongly outperforms* **B** if all solutions in **B** are equal to or dominated by solutions in **A** and there exists at least one solution in **B** that is dominated by solutions in **A**, e.g. Fig. 1 (b).
- Complete outperformance: $\mathbf{B} \cap \mathbf{ND}(\mathbf{A} \cup \mathbf{B}) = \emptyset$. **A** *completely outperforms* **B** if each solution in **B** is dominated by solutions in **A**, e.g. Fig. 1 (c).

where $\mathbf{ND}()$ denotes the set includes all nondominated solutions.

3.2 $\mathcal{C}$ Measure

The $\mathcal{C}$ measure indicates the coverage of two sets [14]. This measure compares two sets of solutions and calculates the proportion of solutions in the second set for which there are solutions at least as good in every objective in the first set.

The definition of $\mathcal{C}$ measure is as follows: Let $\mathbf{A}, \mathbf{B} \subseteq \mathbf{X}$ be two sets of decision vectors. The function $\mathcal{C}$ maps the ordered pair $(\mathbf{A}, \mathbf{B})$ to the interval $[0, 1]$:

$$\mathcal{C}(\mathbf{A}, \mathbf{B}) := |\{\mathbf{b} \in \mathbf{B} \mid \exists \mathbf{a} \in \mathbf{A} : \mathbf{a} \preceq \mathbf{b}\}|/|B|$$

The value $\mathcal{C}(\mathbf{A}, \mathbf{B}) = 1$ means that all decision vectors in $\mathbf{B}$ are at least weakly dominated by $\mathbf{A}$. The opposite, $\mathcal{C}(\mathbf{A}, \mathbf{B}) = 0$, represents the situation when none of the points in $\mathbf{B}$ are weakly dominated by $\mathbf{A}$. Note that always both directions have to be considered, since $\mathcal{C}(\mathbf{A}, \mathbf{B})$ is not necessarily equal to $1 - \mathcal{C}(\mathbf{B}, \mathbf{A})$.

The $\mathcal{C}$ measure has some drawbacks: (1) It cannot measure the *subset* relation. In Fig. 2. (a), Set $\mathbf{A}$ *includes* Set $\mathbf{B}$, however, values of $\mathcal{C}$ measure are both zero; (2) If solutions in set $\mathbf{A}$ not dominated by solutions in set $\mathbf{B}$, while vice versa, the value of $\mathcal{C}$ measure is zero, e.g., Fig. 2. (b); (3) The magnitude of solutions is not considered. In Fig. 2. (c), the result of $\mathcal{C}$ measure is not obeyed the intuition.

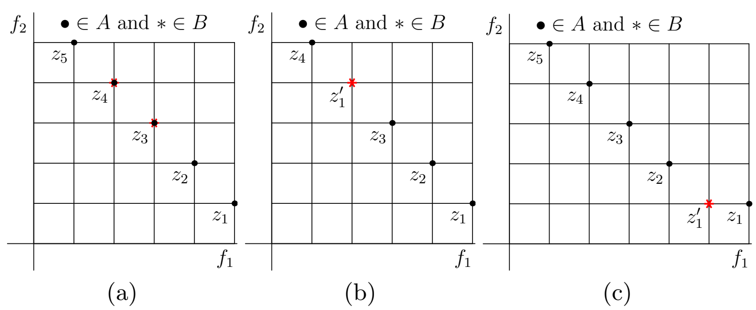

Fig. 2. Drawbacks of $\mathcal{C}$ measure: (a) Set $\mathbf{B}$ is a subset of set $\mathbf{A}$: $\mathcal{C}(\mathbf{A}, \mathbf{B}) = 0$, $\mathcal{C}(\mathbf{B}, \mathbf{A}) = 0$; (b) $\mathbf{A} \not\prec \mathbf{B}$ and $\mathbf{B} \not\prec \mathbf{A}$: $\mathcal{C}(\mathbf{A}, \mathbf{B}) = 0$, $\mathcal{C}(\mathbf{B}, \mathbf{A}) = 0$; (c) Different number of element in each set: $\mathcal{C}(\mathbf{A}, \mathbf{B}) = 0$, $\mathcal{C}(\mathbf{B}, \mathbf{A}) = 1/5$.

The above metrics are a kind of quality measure. It shows the relations of two sets, however, in most cases, two solutions both have part of non-dominated solutions. The set based metrics are difficult to utilize in that situation.

3.3 function $\mathcal{M}_3$

The function $\mathcal{M}_3$ is a spread metric, which measures the spread of the solutions set $\mathbf{A}$ in decision space or the spread of the obtained nondominated solutions $\mathbf{U}$

in objective space [14].

$$\mathcal{M}_3(\mathbf{A}) = \sqrt{\sum\nolimits_{i=1}^{n} \max\{\|a_i - b_i\| \;|\mathbf{a}, \mathbf{b} \in \mathbf{A}\}}$$
$$\mathcal{M}_3^*(\mathbf{U}) = \sqrt{\sum\nolimits_{i=1}^{n} \max\{\|u_i - v_i\| \;|\mathbf{u}, \mathbf{v} \in \mathbf{U}\}}$$

The function $\mathcal{M}_3$ ignores the magnitude of solutions.

4 Reference Point Based Metrics

The reference points based metrics are mostly used in multiobjective optimization. Through these metrics, the goodness of solutions is measured by a single scalar. These metrics are easy in concept and efficient in calculation, however, these metrics are sensitive to the choice of the reference point, and a solution in different part of Pareto front plays different role in the scalar calculation.

4.1 $\mathcal{S}$ Measure (hypervolume)

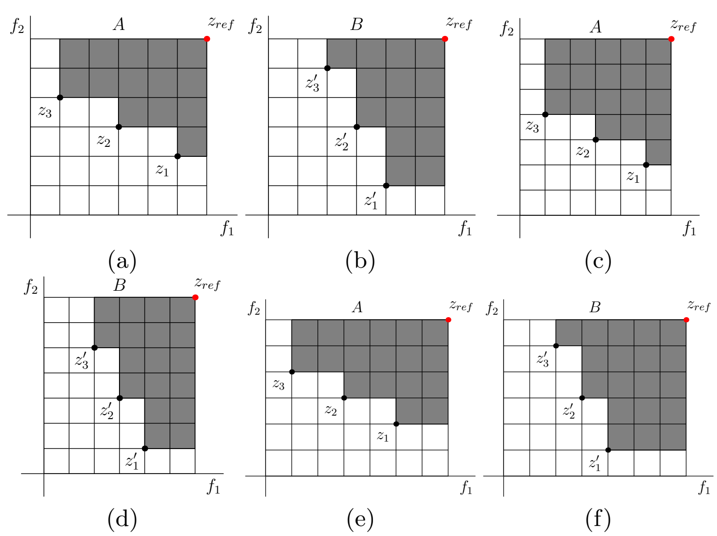

Fig. 3. The relative value of the $\mathcal{S}$ metric depends upon an arbitrary choice of reference point z_{ref}. Two nondominated sets are shown, A and B, in Fig. (a) and (b) $\mathcal{S}(A) = \mathcal{S}(B)$, in Fig. (c) and (d) $\mathcal{S}(A) > \mathcal{S}(B)$, and in Fig. (e) and (f) $\mathcal{S}(A) < \mathcal{S}(B)$. The same sets have a different ordering in $\mathcal{S}$ caused by a different choice of z^{ref}.

A favored metric is *hypervolume*, also known as the $\mathcal{S}$ measure [14] or Lebesgue measure. The hypervolume is a measure of how much of the objective space is weakly dominated by a given nondominated set. i.e., it measures the size of the portion of objective space that is dominated by these solutions collectively.

Generally, hypervolume is favored because it captures in a single scalar both the closeness of the solutions to the optimal set and, to some extent, the spread of the solutions across objective space. Hypervolume also has nicer mathematical properties than many other metrics; although it is difficult to calculate the accurate value of hypervolume, many fast algorithms are proposed to get an approximate scalar [11,12]. Also, it has been proved that hypervolume is maximized if and only if the set of solutions contains only Pareto optima.

Hypervolume has some nonideal properties:

- It is sensitive to the choice of reference point. Fig. 3 displays that the same sets have a different ordering in $\mathcal{S}$ caused by a different choice of z^{ref} [6,7].
- Extreme points play an important role than points in the middle of the Pareto front. For example, in Fig. 3 (a), z_3 is more important than z_2, and in Fig. 3 (b). z'_1 is worth more than z'_2.
- Hypervolume is expensive to calculate, an approach needs to be designed to approximate it within a reasonable error [12].

4.2 $\mathcal{D}$ Measure

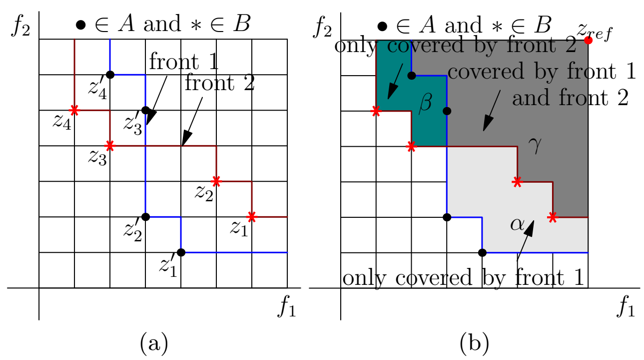

Fig. 4. The comparison between $\mathcal{C}$ measure and $\mathcal{D}$ measure. (a) the $\mathcal{C}$ measure: $\mathcal{C}(\mathbf{A}, \mathbf{B}) = \mathcal{C}(\mathbf{B}, \mathbf{A}) = 1/2$; (b) The $\mathcal{D}$ measure: $\mathcal{D}(\mathbf{A}, \mathbf{B}) = \alpha$, $\mathcal{D}(\mathbf{B}, \mathbf{A}) = \beta$, and $\mathcal{D}(\mathbf{A}, \mathbf{B}) > \mathcal{D}(\mathbf{B}, \mathbf{A})$.

The $\mathcal{D}$ measure indicates the coverage difference of two sets [14]. This measure combines the C measure and hypervolume measure.

The definition of $\mathcal{D}$ measure is as follows: Let $\mathbf{A}, \mathbf{B} \subseteq \mathbf{X}$ be two sets of decision vectors. The function $\mathcal{D}$ is defined by

$$\mathcal{D}(\mathbf{A}, \mathbf{B}) := \mathcal{S}(\mathbf{A} + \mathbf{B}) - \mathcal{S}(\mathbf{B}) \tag{1}$$

and gives the size of the space weakly dominated by **A** but not weakly dominated by **B** (regarding the objective space).

As shown in Fig. 4, (a) is for $\mathcal{C}$ measure, (b) is for $\mathcal{D}$ measure. There is the area of size α that is covered by front 1 but not by front 2; and area of size β that is covered by front 2 but not by front 1. The dark-shaded area (of size γ) is covered by both front in common. It holds that $\mathcal{D}(\mathbf{A},\mathbf{B}) = \alpha$, and $\mathcal{D}(\mathbf{B},\mathbf{A}) = \beta$.

In this example, $\mathcal{D}(\mathbf{B},\mathbf{A}) > \mathcal{D}(\mathbf{A},\mathbf{B})$ which reflects the quality difference between the two fronts in contrast to the $\mathcal{C}$ metric. In addition, the $\mathcal{D}$ measure gives information about whether either set entirely dominates the other set, e.g., $\mathcal{D}(\mathbf{A},\mathbf{B}) = 0$ and $\mathcal{D}(\mathbf{B},\mathbf{A}) > 0$ means that **A** is dominated by **B**.

The $\mathcal{D}$ measure is based on the hypervolume calculation. It is sensitive to the choice of reference point. Fig. 5 displays that the same sets have a different ordering in $\mathcal{D}$ caused by a different choice of z^{ref}.

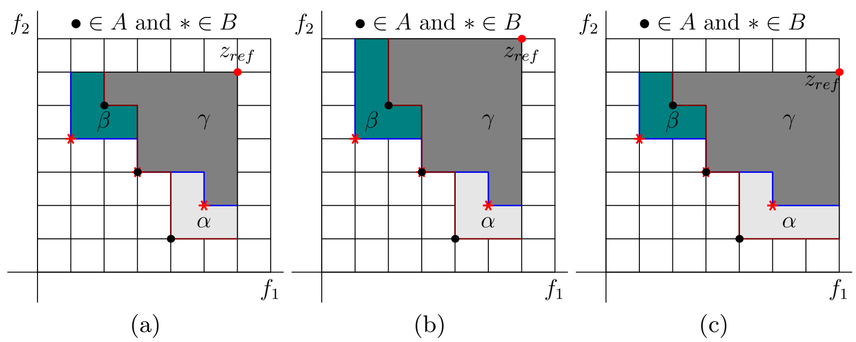

Fig. 5. The relative value of the $\mathcal{D}$ metric depends upon an arbitrary choice of reference point z^{ref}. In Fig.(a) $\mathcal{D}(A,B) = \mathcal{D}(B,A)$, (b) $\mathcal{D}(A,B) < \mathcal{D}(B,A)$, (c) $\mathcal{D}(A,B) > \mathcal{D}(B,A)$.

5 True Pareto Front/Set Based Metrics

True Pareto front based metrics compares the distribution of Pareto front found by the search algorithm and the true Pareto front. This kind of metrics is only utilized on benchmark problems, because the true Pareto front is unknown for real-world problems. However, utilizing these metrics, the search efficiency of different algorithms can be compared.

5.1 Inverted Generational Distance (IGD)

One frequently used metric is Inverted generational distance (IGD) [8, 13] also known as reverse proximity indicator (RPI) [3, 10], or the convergence metric γ [4]. It measures the extent of convergence to a known set of Pareto-optimal solutions.

The definition of this metric is as follows: Let P^* be a set of uniformly distributed Pareto-optimal points in the PF (or PS). Let $\mathbf{P}$ be an approximation to the PF (or the PS). The IGD metric is defined as follows:

$$IGD(\mathbf{P}^*, \mathbf{P}) = \sum\nolimits_{v \in \mathbf{P}^*} d(v, \mathbf{P}) / |\mathbf{P}^*|$$

where $d(v, \mathbf{P})$ is the minimum Euclidean distance between v and all of the points in the set $\mathbf{P}$; and $|\mathbf{P}^*|$ is the cardinality of $\mathbf{P}^*$. In this metric, the number of solutions in $\mathbf{P}$ should be large enough to obtain an accurate result.

The IGD metric can be utilized both in solution space and objective space. In objective space, $\mathbf{P}^*$ is a set of points in the PF and $d(v, \mathbf{P})$ is the minimum distance between fitness values of solutions and the Pareto front. While in decision space, $\mathbf{P}^*$ is a set of points in the PS and $d(v, \mathbf{P})$ is the minimum distance between solutions and the Pareto set.

5.2 Hypervolume Difference (I_H^-) Metric

The Hypervolume difference I_H^- metric is defined as

$$I_H^-(\mathbf{P}^*, \mathbf{P}) = I_H(\mathbf{P}^*) - I_H(\mathbf{P})$$

where $I_H(\mathbf{P}^*)$ is the hypervolume between the true Pareto front $\mathbf{P}^*$ and a reference point, and $I_H(\mathbf{P})$ is the hypervolume between the obtained Pareto front $\mathbf{P}$ and the same reference point. The hypervolume difference measure is also based on the hypervolume calculation. The result may be different by the choice of reference point.

Both the IGD metric and the I_H^- metric measure convergence and diversity. To have low IGD and I_H^- values, $\mathbf{P}$ must be close to the PF (or PS) and cannot miss any part of the whole PF (or PS) [13].

5.3 Spacing Δ Metric

The spacing Δ metric measures the extent of spread achieved among the obtained solutions [4]. The following metrics is utilized to calculate the nonuniformity in the distribution:

$$\Delta = \frac{d_f + d_l + \sum_{i=1}^{|\mathbf{P}|-1} |d_i - \bar{d}|}{d_f + d_l + (|\mathbf{P}| - 1)\bar{d}}$$

where d_i is the Euclidean distance between consecutive solutions in the obtained nondominated set of solutions $\mathbf{P}$, d_f and d_l are the distances between the extreme solutions in true Pareto front and the boundary solutions of $\mathbf{P}$. $\bar{d}$ is the average of all distance d_i, $i \in [1, |\mathbf{P}| - 1]$.

6 Conclusions

Multiobjective Optimization refers to optimization problems that involve two or more objectives, and a set of solutions is obtained instead of one. How to measure the goodness of solutions and the performance of algorithms is important in multiobjective Optimization.

In this paper, we reviewed variants of performance metrics and classified them into three categories: set based metric, reference point based metric, and the true Pareto front/set based metric. The properties and drawbacks of different metrics are discussed and analyzed.

A proper metric should be chosen under different situations, and on the contrary, an algorithm's ability can be measured by different metrics. An algorithm's properties can be revealed through different metrics analysis on different problems, then different algorithms can be utilized in an appropriate situation. From the analysis of different metrics, an algorithm's properties can be revealed and more effective algorithms can be designed to solve MOO problems.

References

1. Adra, S.F., Fleming, P.J.: Diversity management in evolutionary many-objective optimization. IEEE Transactions on Evolutionary Computation 15(2), 183–195 (April 2011)
2. Adra, S.F., Dodd, T.J., Griffin, I.A., Fleming, P.J.: Convergence acceleration operator for multiobjective optimization. IEEE Transactions on Evolutionary Computation 12(4), 825–847 (August 2009)
3. Bosman, P.A.N., Thierens, D.: The balance between proximity and diversity in multiobjective evolutionary algorithms. IEEE Transactions on Evolutionary Computation 7(2), 174–188 (April 2003)
4. Deb, K., Agrawal, S., Pratap, A., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6(2), 182–197 (April 2002)
5. Hansen, M.P., Jaszkiewicz, A.: Evaluating the quality of approximations to the non-dominated set. Tech. Rep. MM-REP-1998-7, Technical University of Denmark (March 1998)
6. Knowles, J.D., Corne, D.: On metrics for comparing non-dominated sets. In: Proceedings of the 2002 Congress on Evolutionary Computation (CEC2002). pp. 711–716 (2002)
7. Knowles, J.D., Corne, D.: Properties of an adaptive archiving algorithm for storing nondominated vectors. IEEE Transactions on Evolutionary Computation 7(2), 100–116 (April 2003)
8. Li, H., Zhang, Q.: Multiobjective Optimization Problems With Complicated Pareto Sets, MOEA/D and NSGA-II. IEEE Transactions on Evolutionary Computation 13(2), 284–302 (April 2009)
9. Okabe, T., Jin, Y., Sendhoff, B.: A critical survey of performance indices for multiobjective optimisation. In: Proceedings of the 2003 Congress on Evolutionary Computation (CEC2003). vol. 2, pp. 878–885 (December 2003)
10. Soylu, B., Köksalan, M.: A favorable weight-based evolutionary algorithm for multiple criteria problems. IEEE Transactions on Evolutionary Computation 14(2), 191–205 (April 2010)

11. While, L., Bradstreet, L., Barone, L.: A fast way of calculating exact hypervolumes. IEEE Transactions on Evolutionary Computation 16(1), 86–95 (February 2012)
12. While, R.L., Hingston, P., Barone, L., Huband, S.: A faster algorithm for calculating hypervolume. IEEE Transactions on Evolutionary Computation 10(1), 29–38 (February 2006)
13. Zhou, A., Zhang, Q., Jin, Y.: Approximating the set of Pareto-optimal solutions in both the decision and objective spaces by an estimation of distribution algorithm. IEEE Transactions on Evolutionary Computation 13(5), 1167–1189 (October 2009)
14. Zitzler, E.: Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications. Ph.D. thesis, Swiss Federal Institute of Technology Zurich (ETH) (November 1999)
15. Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., da Fonseca, V.G.: Performance assessment of multiobjective optimizers: An analysis and review. IEEE Transactions on Evolutionary Computation 7(2), 117–132 (April 2003)