Performance metrics for multiobjective optimization evolutionary algorithms

Crina Grosan
cgrosan@cs.ubbcluj.ro

Abstract. In the recent past many algorithms for multiobjective optimization have been proposed. To evaluate performances of these algorithms some measures of performances are needed. Many metrics of algorithms performances have been proposed. Existing performance metrics are briefly reviewed. Two metrics computing the convergence towards the Pareto front and the solution diversity on the Pareto front are proposed.

Keywords. Evolutionary Algorithms, Multiobjective Optimization, Pareto Front, Performance Metrics

1. Introduction

In the last years many multiobjective optimizations algorithms (MOA) have been proposed. To compute the performances of these algorithms some measures of performance were also introduced. Most of them are applied to the final nondominated set. It is now established that more than one metrics are necessary to evaluate the performances of the multiobjective evolutionary algorithms. Zitzler ([10]) has recently shown that for an *M*-objective optimization problem, at least *M* performance metrics must be used.
According to Deb ([1]) the existing performance metrics can be classified into three classes: metrics for convergence, metrics for diversity and metrics for both convergence and diversity.
Knowles and Corne ([6]) compared most of these metrics based on their compatibility with outperformance relations between two sets of solutions and their monotonicity properties. The metrics are classified based on strong, weak and complete outperformance of one set to another.
Veldhuizen ([8]) reports a number of performance metrics for multiobjective optimization.
Some of more recent and important metrics of performance are reviewed in the next section. Taking into account the suggestion made by Deb in ([2]) three metrics are important: one metric for convergence, one metric for diversity and one metric to analyze the running time. Two new metrics one for convergence and one for diversity are introduced in Section.

2. Performance metrics: a review

Jaszkiewicz ([5]) suggested that for evaluating the algorithm performances for multiobjective optimization two main criteria have to be taken into account
:

- computational requirement;
- the quality of the result.

The *quality measures* may be classified as either cardinal or geometrical. The idea of the cardinal measures is to count the number or ratio of points that meet some requirements. To define the requirements *equivalence* and/or *dominance* relation is used. Geometrical measures take into account geometrical position of the nondominated points in the objective space.
Zitzler et al. ([11]) suggest three goals of Pareto multiobjective search that can be identified and measured:

- distance of the resulting nondominated set to the true Pareto front should be minimized;
- a good distribution of the obtained solutions;
- the size of the obtained nondominated front should be maximized (i.e. for each objective, a wide range of values should be covered by the nondominated solutions).

The final goal of multiobjective optimization is to find a unique solution giving the best compromise between multiple objectives. But usually there is more than one solution. Selection of the best compromise solution requires taking into account Decision maker (DM) preferences. If decision making corresponds to a utility function, the best compromise is the solution that maximizes the utility. The solution selected by the DM from the approximation supplied by a MOA is generally worse than the best solution contained in the actual nondominated set. The DM may consider approximation A to be better than approximation B if it can find a better compromise solution in A than in B ([10]).

Hansen and Jaszkiewicz ([4]) defined a number of outperformance relations that express the relationship between two sets of (internally) nondominated objective vectors A and B.

To check if a final set of solutions is better than the other Jaszkiewicz ([5]) uses a set of utility functions. A utility function is defined as follows:

Definition 1.
A utility function $u : X \subset \mathbf{R}$ **->** $\mathbf{R}$ is a model of the Decision Maker's (DM) preferences that maps each point in the objective space into a value of utility.

Remarks
(i) It is assumed that the goal of the DM is to minimize the utility.
(ii) A utility function u is compatible with the dominance relation if and only if:

$$\forall \quad x_1, x_2 \in X \; x_1 \succ x_2 \Rightarrow u(x_1) \geq u(x_2).$$

Each utility function of this type has at least one global optimum belonging to the set of Pareto optimal solutions. For each Pareto optimal solution x there exist a utility function u compatible with the dominance relation such that x is a global optimum of u.

Let U be a set of utility functions. For an approximation set A the followings relations are assumed:

$$u^*(A) = \max_{z \in A}\{u(z)\};$$
$$U(A > B) = \{u \in U\text{: } u^*(A) >. u^*(B)\}.$$

Definition 2. (Outperformance relation subject to a set of utility functions)
Approximation A *outperforms* B subject to a set U of utility functions if $U(A > B) \neq \emptyset$ and $U(B > A) = \emptyset$, i.e. if there exist some utility functions in set U that achieve better values in A than in B, while the opposite is not true.

DM never prefers a dominated solution. In this way DM can limit the search to the set of Pareto optimal solutions only. When two approximations A and B from which DM must select a compromise solution are known DM can limit the search to the set of nondominated solutions from $A \cup B$. The following definitions are needed (see [5]).

Let us denote $ND(A)$ the set of nondominated solutions from the set A.

Definition 3. (Weak outperformance)
Approximation A *weakly outperforms* approximation B (and we write $A\ O_w\ B$) if the following conditions are fulfilled:

(i) $A \neq B$;
(ii) $ND(A \cup B) = A$.

That means that for each point $y \in B$ there exists a point $x \in A$ that is equal to or dominates y and at least one point $x \in A$ is not contained in B.

Definition 4. (Strong outperformance)
Approximation A *strongly outperforms* B (and we write $A\ O_s\ B$) if the following conditions are fulfilled:

(i) $ND(A \cup B) = A$;
(ii) $B / ND(A \cup B) \neq \emptyset$.

That means that for each point $y \in B$ there exists a point $x \in A$ that is equal to or dominates y and at least one point $y \in B$ is dominated by a point $x \in A$.

Definition 5. (Complete outperformance)
Approximation A *completely outperforms* B (and we write $A\ O_c\ B$) if the following conditions are fulfilled:
$ND(A \cup B) = A$;
$B \cap ND(A \cup B) = \emptyset$.
That means that each point $y \in B$ is dominated by a point $x \in A$.

Remark
$O_c \subset O_s \subset O_w$.

We will classify here measures for evolutionary algorithms performances in two major classes:

1. convergence metrics – evaluate how far from the true Pareto front solutions obtained in final population are;
2. diversity metrics – evaluate scatter of solutions in the final population on the Pareto front;
3. direct comparison metrics – evaluate solution sets using utility functions.

Each of them is described in detail bellow.

2.1 Metrics for convergence

Many metrics for measuring the convergence of a set of nondominated solutions towards the Pareto front have been proposed. Almost all of these metrics were constructed in order to directly compare two sets of nondominated solutions. There are also approaches which compare a set of nondominated solutions with a set of Pareto optimal solutions if the true Pareto front is known.
In what follows we review some existing metrics for convergence.

2.1.1 Metric *S*

The S metric has been introduced by Zitzler in [9] and improved in [10]. The S metric measures how much of the objective space is dominated by a given nondominated set A.

Definition 6. (Size of the dominated space)

Let X be set of decision vectors for the considered problem and $A = \{x_1, x_2,\ldots,x_t\} \subseteq X$ a set of t decision vectors. The function $S(A)$ gives the volume enclosed by the union of the polytopes p_1, p_2,…, p_t, where each p_i is formed by the intersection of the following hyperplanes arising out of x_i, along with the axes: for each axis in the objective space there exist a hyperplane perpendicular to the axis and passing through the point $(f_1(x_i), f_2(x_i),\ldots,f_k(x_i))$.

Example 1

In the two-dimensional case, each p_i represents a rectangle defined by the points (0,0) and $(f_1(x_i), f_2(x_i))$. An example for two-dimensional case is presented in Figure 1.

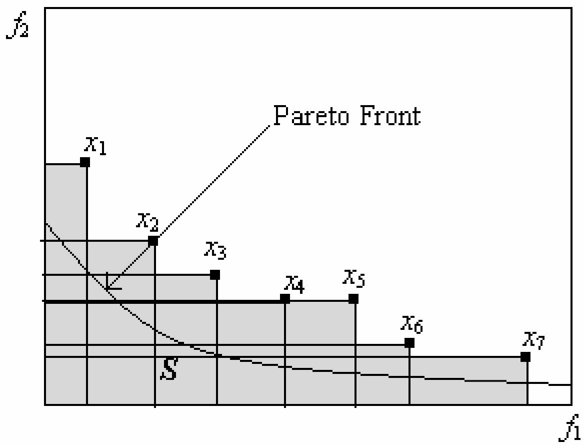

Figure 1. The metric *S* for the case of two objective functions and 7 decision vectors (x_1, x_2, …, x_7) for a minimization problem.

Remarks

(i) Veldhuizen ([8]) suggest that metric *S* can be misleading if the Pareto optimal front is non-convex.

(ii) An advantage of the metric *S* is that each algorithm can be assessed independently of the other algorithms. The *S* value of two sets *A* and *B* cannot be used to decide if one set entirely dominates the other.

2.1.2 Metric *C*

The metric *C*, like metric *S*, was introduced by Zitzler in [9] and improved in [10]. Using metric *C* two sets of nondominated solutions can be compared to each other.

Definition 7. (Coverage of two sets)

Let *X* be the set of decision vectors for the considered problem and $A, B \subseteq X$ two sets of decision vectors. The function *C* maps the ordered pair (*A*, *B*) into the interval [0,1]:

$$C(A,\ B) = \frac{\left|\{b \in B \,/\, \exists a \in A : a \succeq b\}\right|}{|B|}.$$

Remarks

(i) The value $C(A, B) = 1$ means that all decision vectors in *B* are dominated by *A*.

(ii) The value $C(A, B) = 0$ represent the situation when none of the points in *B* are dominated by *A*.

(iii) $C(A, B)$ is not necessary equal to $1 - C(B, A)$.

Example 2

There are situations when the metric *C* cannot decide if an obtained front is better than the other. Let us suppose that front1 correspond to a set A and front 2 to a set B.

In Figure 2, the surface covered by the front 1 is equal to the surface covered by the front 2 but front 2 is closer to the Pareto optimal front than front 1. In this situation (and in other situations similar with this) the *C* metric is not applicable. To eliminate this shortcoming a new metric – *D* metric – was proposed.

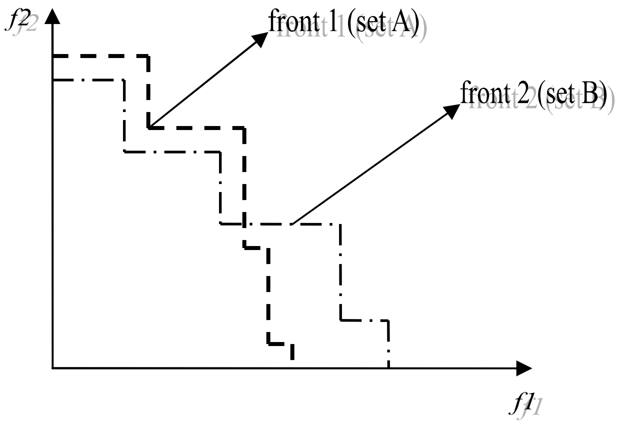

Figure 2. An example when metric *C* can not decide between front 1 and front 2 (the surface covered by front 1 is equal to the surface covered by front 2).

Definition 8. (Coverage difference of two sets)
Let $A, B \subseteq X$ be two sets of decision vectors. the size of the space dominated by A and not dominated by B (regarding the objective space) is denoted $D(A, B)$ and is defined as:

$$D(\mathrm{A}, \mathrm{B}) = S(\mathrm{A} + \mathrm{B}) - S(\mathrm{B}),$$

where $S(A)$ is defined above.

Example 3
Metric *D* can be used to solve the inconvenience of Example 2.Consider the notations from Figure 3.
By applying metric *D* the followings equalities are obtained:

$$S(A + B) = \alpha + \beta + \gamma,$$
$$S(A) = \alpha + \gamma,$$
$$S(B) = \alpha + \beta.$$

The metric *D* for this example is expressed below.

$$D(A, B) = \gamma,$$
$$D(B, A) = \beta.$$

From

$$D(A, B) < D(B, A)$$

it results that front 2 dominates front 1.

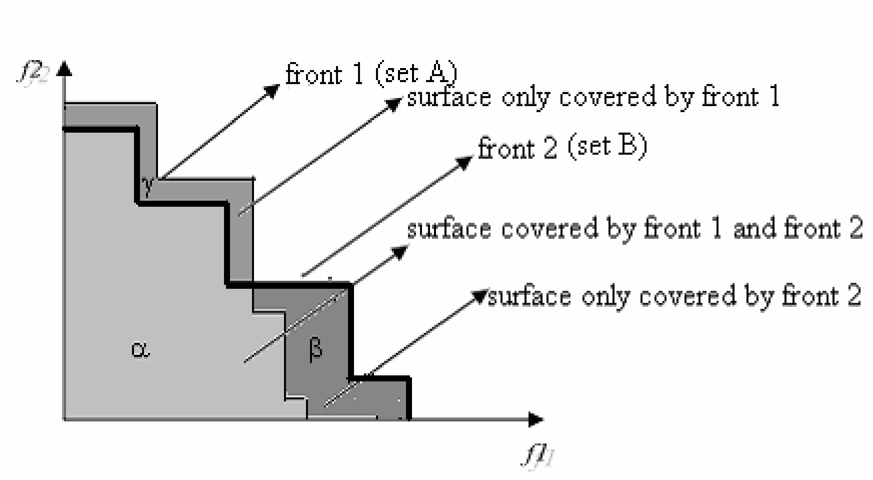

Figure 3. Example of difference between *C* metric and *D* metric for considered fronts front 1 and front 2.

Zitzler ([10]) suggest that (ideally) the *D* metric is used in combination with the *S* metric where the values may be normalized by a reference volume V, where (for a maximization problem) V is given by:

$$V = \prod_{i=1}^{k} (f_i^{\max} - f_i^{\min}) .$$

$f_i^{\max}$ and $f_i^{\min}$ represent the maximum respectively minimum value for the objective f_i. Thus, the value

$$D'(A,B) = \frac{D(A,B)}{V}$$

represents the relative size of the region (in the objective space) dominated by A and not dominated by B.

2.1.3 A convergence metric

Deb ([2]) proposed a convergence metric that evaluates the convergence towards a reference set.
A target set P^* is used. P^* can be either a set of Pareto optimal solutions (if known) or the nondominated set of points in a combined pool of all generations-wise populations obtained from a multiobjective evolutionary algorithm run.
For a population $P(t)$ of an evolutionary algorithm the convergence metric is computing following steps:

Step 1. Identify the nondominated set $F(t)$ of $P(t)$.
Step 2. From each solution i in $F(t)$ calculate the smallest normalized Euclidian d_i to P^* :

$$d_i = \min_{j=1}^{|P^*|} \sqrt{\sum_{k=1}^{M} \left(\frac{f_k(i) - f_k(j)}{f_k^{\max} - f_k^{\min}} \right)^2} \; ,$$

where M denotes the number of objectives and $f_k^{\max}$ and $f_k^{\min}$ are the maximum and the minimum function values of the k–th objective function in P^* respectively.

Step 3. Calculate the convergence metric value $C(P(t))$ for $P(t)$ by averaging the normalized distance for all points in $F(t)$:

$$C(P(t)) = \frac{\sum_{i=1}^{|F(t)|} d_i}{|F(t)|} .$$

Remark

In order to keep the convergence metric within [0, 1], once the above metric values are calculated for all generations, the value $C(P(t))$ is normalized by the maximum value (usually $C(P(0))$):

$$\overline{C}(P(t)) = \frac{C(P(t))}{C(P(0))} .$$

2.1.4 Error ratio

Let A = $\{e_1, e_2, \ldots, e_n\}$ an approximation set; $e_i = 0$ if vector i is in the true Pareto front and 1 otherwise.

Error ratio has been proposed by Veldhuizen ([8]) and measures the ratio, amongst the approximation set (Pareto front known in Veldhuizen's terminology), of those vectors that are in the true Pareto front, to those not in the true Pareto front. This metric uses the true Pareto front as reference set and is given by:

$$Err(A) = \frac{\sum_{i=1}^{n} e_i}{n} \; ,$$

where n is the number of vectors in the approximation set.

Lower values of the error ratio are preferable. A low value of $Err\,(A)$ means that many solutions in the set A lies on the true Pareto front.

2.2 Diversity metrics

In this section some metrics for diversity are outlined. Each of them is shortly described.

2.2.1 Spacing metric (1)

Spacing metric (Δ) has been proposed by Deb et al. in [3] and measures how evenly the points in the approximation set are distributed in the objective space.
This metric is given by:

$$\Delta = \sum_{i=1}^{|PF|} \frac{d_i - \overline{d}}{| PF |},$$

where PF represent the known Pareto front, d_i is the Euclidian distance between two consecutive vectors in the nondominated front of the approximation set and $\overline{d}$ is the average of these distances.

Remark
Metric Δ is only suitable for two-dimensional objective spaces because it is not clear how 'consecutive' would be defined in the case of more than two objectives.

2.2.2 Spacing metric (2)

Spacing metric Δ' has been introduced by Schott in [7]. The purpose of this metric is to gauge haw evenly the points in the approximation set are distributed in the objective space.
This metric is given by:

$$\Delta' = \sqrt{\frac{1}{n-1} \sum_{i+1} n(\overline{d} - d_i)^2},$$

where

$$d_i = \min_j (| f_1^i(\vec{x}) - f_1^j(\vec{x}) | + | f_2^i(\vec{x}) - f_2^j(\vec{x}) |, i, j = 1, 2, \ldots n, \overline{d}$$

is the mean of all d_i and n is the size of the known Pareto front.

If this metric is used in conjunction with other metrics it may provide information about the distribution of vectors obtained. It has low computational overhead. The metric can be generalized to more than two dimensions by extending the definition of distance d_i to cover more objectives.

2.2.3 A diversity metric

In this section we consider a metric for diversity proposed by Deb in [2]. The obtained nondominated points at each generation are projected on a suitable hiperplan. The plan is divided into a number of small grids ((M – 1) dimensional boxes, M being the number of objectives). The diversity metric is defined depending on whether each grid contains an obtained nondominated point or not. The best possible result is obtained if all grids are

represented with at least one point. If some grids are not represented by a nondominated point the diversity is poor.
Let us denote by P^* a set of optimal solutions, by $P(t)$ the population at the generation t, by $F(t)$ the set of nondominated solutions at the generation t.
The procedure for computing diversity metric is given in the next steps:

Step 1. From $P(t)$ determinate the set $F(t)$ which are nondominated to P^*.

Step 2. For each grid indexed by $(i, j, \ldots)$ calculate following two arrays:

$$H(i, j, \ldots) = \begin{cases} 1, \text{ if the grid has a representative point in } P^* \\ 0, \text{ otherwise} \end{cases},$$

$$h(i, j, \ldots) = \begin{cases} 1, \text{ if } H(i, j, \ldots) = 1 \text{ and the grid has a representative point in } F(t) \\ 0, \text{ otherwise} \end{cases}.$$

Step 3. Assign a value $m(h(i, j \ldots))$ to each grid depending on its and its neighbor's $h(i, j, \ldots)$. Similarity, calculate $m(H(i, j, \ldots))$ using $H(i, j, \ldots)$ for reference points.

Step 4. Calculate the diversity metric by averaging the individual $m(i, j, \ldots)$ values for $h(i, j, \ldots)$ with respect to that for $H(i, j, \ldots)$:

$$D(P(t)) = \frac{\sum_{H(i,j\ldots)\neq 0}^{i,j,\ldots} m(h(i, j, \ldots))}{\sum_{H(i,j\ldots)\neq 0}^{i,j,\ldots} m(H(i, j, \ldots))}.$$

Remark
For greater number of objectives the value function will be difficult to define.

2.3 Metrics for direct comparison of two sets

The metrics from this class directly compare two set of solutions using a set of utility functions.

2.3.1 *R*1 metric

Consider the metric $R1$ proposed by Hansen and Jaszkiewicz in [4]. $R1$ is based on calculating the probability that approximation A is better than approximation B over an entire set of utility functions.
Consider A and B are two approximation sets, U is a set of *utility functions*, $u : \mathbf{R}^K \to \mathbf{R}$ which maps each point in the objective space into a measure of utility. $p$is an *intensity function* expressing the probability density of each utility function $u \in U$.
Let u^* be the function defined as:

$$u^*(A) = \max_{z\in A}\{u(z)\}.$$

The $R1$ metric is defined as follows:

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

where

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

Remark

The metric $R1$ have a lower computational overhead than the S metric.

2.3.2 Metric *R*2

Metric $R2$ has been proposed by Hansen and Jaszkiewicz in [4]. As $R1$ just uses the function $C(A, B, u)$ to decide which of two approximations is better on utility function u, without measuring by how much, R2 take into account the expected values of the utility. R2 calculates the expected difference in the utility of an approximation A with another one B.

Consider A and B are two approximation sets, U is a set of *utility functions*, $u : \mathbf{R}^K \to \mathbf{R}$ which maps each point in the objective space into a measure of utility. $p$is an *intensity function* expressing the probability density of each utility function $u \in U$.

Let u^* be the function defined as:

$$u^*(A) = \max_{z\in A}\{u(z)\}.$$

The metric R2 is defined as follows:

$$\begin{aligned} R2(A,B,U,p) &= E(u^*(A)) - E(u^*(B)) \\ &= \int_{u\in U} u^*(A)\,p(u)\,du - \int_{u\in U} u^*(B)\,p(u)\,du, \\ &= \int_{u\in U} (u^*(A) - u^*(B))\,p(u)\,du \end{aligned}$$

3. Two new metrics for convergence and diversity

In this section two metrics - one for evaluate the convergence to the Pareto set and the other to determinate the spread of the solutions on the Pareto set are proposed.

3.1 New convergence metric

Assume that the Pareto front is known. Let us denote by P a set of Pareto optimal solutions.
For each individual from the final population distance (Euclidian distance or other suitable distance) to the all points of P is computed. The minimum distance is kept for each individual. The average of these distances represents the measure of convergence to the Pareto front.

3.2. New diversity metric

For each individual from the final population we consider the point from the set of Pareto optimal points P situated at minimal distance. Several concepts of distance to a set may be considered. Here we consider $d(x, P)$ as being:

$$d(x, P) = \min_{y \in P} d(x, y) .$$

We called each such point from P a *marked* point. The total number of different marked points from P over the size of P represents the diversity metric.

Remarks
(i) These two metrics have a low computational cost.
(ii) These metrics can be applied for high dimensional space.

4. Conclusions

Many metrics have been proposed in the last years. Most of them calculate the convergence to an obtained set of solutions to the true Pareto front. The others measure the diversity of the obtained set of solutions on the Pareto front. We can not say that one metric is the best. Some of these metrics are preferred considering some aspects the others from the others aspects. Some of them are preferred to the others by considering the computation complexity. For different classes of problems different types of metrics can be preferred.
Two new metrics have been proposed. Results of these metrics are comparable with the other popular metrics but their complexity is lower.

References

1. Deb, K., *Multiobjective optimization using evolutionary algorithms*. Chichester, UK: Wiley, 2001.
2. Deb, K., Jain, S., Running performance metrics for Evolutionary Multiobjective Optimization, KanGAL Report No. 2002004, Kanpur Genetic Algorithms Laboratory, Indian Institute of Technology Kanpur, India, 2002.

3. Deb, K., Agrawal, S., Pratab, A., Meyarivan, T., A fast elitist non-dominated sorting genetic algorithms for multiobjective optimization: NSGA II. KanGAL report 200001, Indian Institute of Technology, Kanpur, India, 2000.
4. Hansen, M.P., Jaszkiewicz, A., Evaluating the quality of approximations to the nondominated set, Technical Report IMM-REP-1998-7, Institute of Mathematical Modeling, Technical University of Denmark, 1998.
5. Jaszkiewicz, A., Multiple objective metaheuristic algorithms for combinatorial optimization, Habilitation thesis, Poznan University of Technology, Poznan, 2001.
6. Knowles, J., Corne, D., One metrics for computing non-dominated set. Proceedings of the World Congress on Computational Intelligence, pp. 711-716, 2002.
7. Schott, J. R., Fault tolerant design using single and multicriteria genetic algorithms optimization. Master’s thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, 1995.
8. Van Veldhuizen, D.A., Multiobjective evolutionary algorithms: Classification, analyses and new innovations. Doctoral dissertation, Air Force Institute of Technology, Dayton, Ohio, 1999.
9. Zitzler, E., Thiele, L., Multiobjective optimization using evolutionary algorithms-a comparative case study. In A. E. Eiben, T. Back, M. Schoenauer and H. P. Schwefel (Eds.), Fifth International Conference on Parallel Problem Solving from Nature (PPSN-V), Berlin, Germany, pp. 292 – 301, 1998.
10. Zitzler, E., Evolutionary algorithms for multiobjective optimization: Methods and applications, PhD thesis, Swiss Federal Institute of Technology Zurich, 1999.
11. Zitzler, E., Deb, K., Thiele, L., Comparison of multiobjective evolutionary algorithms: Empirical results. Evolutionary Computation, 8(2): 173-195, 200.