2014 IEEE Congress on Evolutionary Computation (CEC)
July 6-11, 2014, Beijing, China

Evolutionary Many-objective Optimization by MO-NSGA-II with Enhanced Mating Selection

Shao-Wen Chen and Tsung-Che Chiang

***Abstract*—Many-objective optimization deals with problems with more than three objectives. The rapid growth of non-dominated solutions with the increase of the number of objectives weakens the search ability of Pareto-dominance-based multiobjective evolutionary algorithms. MO-NSGA-II strengthens its dominance-based predecessor, NSGA-II, by guiding the search process with reference points. In this paper, we further improve MO-NSGA-II by enhancing its mating selection mechanism with a hierarchical selection and a neighborhood concept based on the reference points. Experimental results confirm that the proposed ideas lead to better solution quality.**

I. Introduction

MULTIOBJECTIVE optimization deals with optimization problems with more than one objective. A simple formulation of a multiobjective minimization problem is as follows:

$$\text{Minimize } F(x) = \{f_1(x), f_2(x), \ldots f_M(x)\}\ x \in \Omega, \quad (1)$$

where Ω denotes the solution space, x denotes a solution, $f_i(x)$ denotes the i^{th} objective function, and M denotes the number of objectives. In most cases these objectives are conflicting, which means that improvement of one objective causes deterioration of some other objectives. The concept of Pareto optimality for multiobjective optimization is based on the Pareto dominance relationship. Assuming all objective functions are to be minimized, a solution x is said to *dominate* a solution y iff $\forall\ i \in \{1, \ldots, M\}, f_i(x) \le f_i(y)$ and $\exists\ i \in \{1, \ldots, M\}\ f_i(x) < f_i(y)$. A solution x^* is *Pareto optimal* iff it is not dominated by any solution in the solution space. We call the set of all Pareto optimal solutions the *Pareto* (*optimal*) *set* and the set of the corresponding objective vectors the *Pareto front*. The goal of solving multiobjective optimization problems (MOPs) to Pareto optimality is to find (or approximate) the Pareto optimal set.

Evolutionary algorithms (EAs) are a kind of approximation algorithms for optimization problems. It relies on a population to search for the optimal solution(s) through selection and variation. The *population*-based nature makes it fit the goal of finding the Pareto optimal *set*, and hence many multiobjective evolutionary algorithms (MOEA) such as NSGA-II [1], SPEA2 [2], IBEA [3], SMS-EMOA [4], and MOEA/D [5][6] have been proposed in the last decade. Evolutionary multiobjective optimization (EMO) has also been recognized as one of the three fastest growing fields of research among all computational intelligence topics in IEEE World Congress on Computational Intelligence (WCCI) in 2006 [7]. As more and more researchers join the field of EMO, design of MOEAs for two- and three-objective optimization problems has become well-established, which encourages researchers to tackle problems with many objectives.

Several pioneer studies [8][9] indicated that dealing with many-objective optimization problems ($M \ge 4$) poses additional challenges compared to problems with $M < 4$ objectives. Ishibuchi et al. [10] summarized three difficulties in many-objective optimization: (1) deterioration of the search ability of traditional MOEAs, (2) exponential growth of dimensionality of the Pareto set, and (3) visualization of the Pareto front. These difficulties have been tackled through several directions. The first direction is to change the problem or to make the goal clearer. Since the difficulty is caused by the large number of objectives, researchers have tried to reduce the number of objectives [11]–[14] or to incorporate user preferences to guide the search process [15]–[19]. The second direction is to improve the weak selective pressure of dominance-based selection when facing the large number of non-dominated solutions in the population. The techniques may modify Pareto dominance or assign different ranks to non-dominated solutions [20]–[26]. The third direction is to try selection mechanisms other than dominance-based ones, such as adopting performance indicators (e.g. hypervolume) [3][4][28]–[31] or scalarizing functions (e.g. weighted sum or Tchebycheff approach) [32]–[35].

NSGA-II played an important role in the field of EMO in the last decade. It served as the basis of many dominance-based MOEAs. However, researchers also noticed its weakness in solving many-objective problems. Recently the authors proposed a new version of the algorithm, called MO-NSGA-II [19]. It generates a set of reference points and associates individuals with the reference points. A reference point associated with fewer individuals is regarded as less-crowded. The selection mechanism favors the individuals that are associated with a less-crowded reference point and are closer to the associated reference point. In this paper, we consider the convergence degree of individuals toward the reference points and concentrate the search on the regions with less convergence. We also construct neighborhood between reference points and enforce individuals to mate with others inside the neighborhood to enhance the quality of the offspring. The rest of this paper is organized as follows. Section II will review the literature. Brief descriptions of MO-NSGA-II will be given in Section III. We will elaborate our algorithm in Section IV. Experiments and results are presented in Section V. Conclusions will be provided in Section VI.

S. W. Chen is with the Department of Computer Science and Information Engineering, National Taiwan Normal University, Taiwan (R.O.C.) (email: maxchen12@hotmail.com).

T. C. Chiang is with the Department of Computer Science and Information Engineering, National Taiwan Normal University, Taiwan (R.O.C.) (phone: +886-2-77346692; fax: +886-2-29322378; email: tcchiang@ieee.org).

II. Literature Review

Much research has been done for many-objective optimization. In this section we follow the categorization in Ishibuchi et al. [10] and report the recent progress.

A. Objective Reduction

An intuitive attempt to address the difficulties caused by the large number of objectives is to reduce the number of objectives. Brockhoff and Zitzler [11] proposed an error metric for a reduced set of objectives based on the concept of ε-dominance. Then, they defined two problems: the δ-MOSS problem aims at finding the minimum set of objectives within a given error, and the k-EMOSS problem aims at minimizing the error under a limited number of objectives. Jaimes et al. [12] proposed a reduction method based on a feature selection technique. They calculated the distance between objectives based on the correlation. Then, the objectives were clustered based on their distances. Following their previous research, Jaimes et al. [13] integrated their objective reduction method into an MOEA and proposed two algorithms. Saxena et al. [14] proposed a reduction method based on principal component analysis (PCA) and maximum variance unfolding (MVU).

B. Preference Points

In addition to reducing the number of objectives, another way to "change" the many-objective nature to make the problem easier is through the incorporation of user preferences. With user preferences, the MOEAs can focus on searching smaller and more specific region(s) to find solutions that users would be interested in. In R-NSGA-II [15], a preference operator replaced the crowding distance operator in the original NSGA-II. The preference distance of a solution was calculated by the minimum weighted normalized Euclidean distance to the reference points. Solutions with smaller preference distance were preferred during mating and environmental selection. Thiele et al. [16] proposed PBEA by introducing reference point(s) and achievement scalarization function (ASF) into IBEA. The fitness value in PBEA was that in IBEA divided by a normalized ASF value with respect to the reference point(s). The PIE algorithm by Sindhya et al. [17] also relied on reference points and ASF. It searches for the single optimal solution with respect to the ASF value rather than a set of solutions considering proximity and diversity simultaneously. Based on the concept of coevolving a family of decision-maker preferences together with a population of candidate solutions, Wang et al. [18] proposed the PICEA-g algorithm. PICEA-g does not ask users to provide reference points; instead, it evolves these reference points (called goals in the algorithm) by preferring those dominated by fewer solutions. On the other hand, solutions are regarded as better if they dominate more goals, especially when the goals are dominated by few solutions. MO-NSGA-II [19] generates structured set of reference points and aims at distributing solutions evenly to the reference points and moving solutions toward their corresponding reference points. In DI-EMOA [20], decision makers specified region of interests of the objectives and defined the desirability functions (DF) accordingly. A desirability index (DI) was then defined as a scalarization operator to map multiple DF values to a single value. The DI was used as the secondary criterion of the non-dominated sorting procedure. A DI-based archiving mechanism was also proposed.

C. Ranking and Modified Pareto Dominance

Noticing the low selective pressure of Pareto dominance in many-objective optimization, the other way is to improve the algorithm. Dreschsler et al. [21] proposed the favour relationship: an individual x is favored over an individual y iff the number of objectives in which x is better than y is larger than the number of objectives in which y is better than x. The favour relationship is identical to the dominance relationship when the number of objectives is two. In G-MOEA [22], users decided the linear trade-off between two objectives. The trade-off specified the minimal required improvement (resp. maximal allowed deterioration) of one objective for a unit of deterioration (resp. improvement) of the other objective. A modified dominance relationship can then be defined accordingly. Sato et al. [23] proposed a method to control the degree of expansion or contraction of the dominance area. Expanding the dominance area increases the number of fronts and the selective pressure. Their experiments showed that expansion of the dominance area improved algorithm performance when the problem dimension, search space, and problem difficulty increase. Corne and Knowles [24] tested seven ranking methods including average ranking [25] and favour relation [21] by two sets of combinatorial optimization problems. The experimental results showed that average ranking is a good method. Garza-Fabre et al. [26] proposed three ranking methods, two based on the difference of objective values between individuals and one based on the distance to the ideal point.

D. Indicator Function

In addition to algorithm design, performance assessment is also an important topic in the field of EMO [27]. Since we evaluate the performance of MOEAs by these indicators, it is natural to use them to select individuals. IBEA [3] was among the earliest studies on this direction. It evaluated individuals based on how they contribute to the indicator. While IBEA allowed the use of different indicators, SMS-EMOA [4] focused on the use of hypervolume. When all individuals are not dominated by each other, the worst individual is the one having the lowest contribution to the hypervolume. Wagner et al. [28] compared seven MOEAs belonging to Pareto-, aggregation-, and indicator-based methods, and the experimental results showed that SMS-EMOA had very good performance. Although SMS-EMOA performs well, expensive computational effort required by hypervolume leads to the use of other indicators. Díaz-Manríquez et al. [29] utilizes the contribution of individuals to the R2 indicator [30] to rank individuals. Phan and Suzuki [31] extended IBEA to R2-IBEA. They used a binary R2 indicator and also proposed an adaptive adjustment method for the reference point to correct the inherent bias of the R2 indicator toward the center of a Pareto front.

E. Scalarizing Function

The essential idea in the category of MOEAs based on scalarizing functions is to use multiple weight vectors and solve corresponding single-objective optimization sub-problems in parallel. In his MSOPS-II, Huges [32] proposed two scalarizing functions and a method to automatically generate weight vectors. Zhang and Li [5][6] developed MOEA/D and incorporated the concept of neighborhood of sub-problems to do mating restriction. Ishibuchi et al. [33] hybridized NSGA-II and linear weighted sum in the selection mechanism. Ishibuchi et al. [34] indicated that the large population made NSGA-II slow and inefficient but MOEA/D still worked very well. Ishibuchi et al. [35] compared linear weighted sum and weighted Tchebycheff. They found that linear weighted sum had faster convergence speed for high-dimensional problems but could not find good solutions for non-convex problems. Thus, they tried to detect the front shape (convex or non-convex) by the number of identical objective vectors in a neighborhood and chosen between the two scalarizing functions accordingly.

III. Review of MO-NSGA-II

In this paper we propose a many-objective evolutionary algorithm by improving the mating selection mechanism in MO-NSGA-II [19]. We will review MO-NSGA-II briefly in this section. Algorithm I presents the main flow of MO-NSGA-II. Comparing with the flow of classical EAs, there are three additional procedures: generation of hyper-plane, generation of reference points, and association of reference points. They will be explained in subsections A and B. After association of population members with reference points, a convergence measure and a diversity measure will be defined. Mating selection and environmental selection are carried out based on the non-domination level and these two new measures.

ALGORITHM I: MO-NSGA-II

$P_1 \leftarrow$ **GenerateInitialPopulation**()
$t \leftarrow 1$
$R \leftarrow$ **GenerateReferencePoints**() // Section III-B
while *stopping criterion is not satisfied*
  $H \leftarrow$ **GenerateHyperplane**(P_t) // III-A
  AssociateIndividualsWithReferencePoints(P_t, H, R) // III-B
  $Q_t \leftarrow \varnothing$
  NonDominatedSorting(P_t) // [1]
  for $i = 1$ to $|P_t|/2$
    $p_1 \leftarrow$ **MatingSelection**(P_t, R) // III-C
    $p_2 \leftarrow$ **MatingSelection**(P_t, R) // III-C
    $\{c_1, c_2\} \leftarrow$ **CrossoverAndMutation**(p_1, p_2)
    $Q_t \leftarrow Q_t \cup \{c_1, c_2\}$
  end for
  $P_{t+1} \leftarrow$ **EnvironmentalSelection**(P_t, Q_t, H, R) // III-D
  $t \leftarrow t + 1$
end while

A. Generation of Hyper-plane

The key innovation in MO-NSGA-II is that it uses reference points on a hyper-plane to guide the search process and maintain population diversity. In this subsection we describe how MO-NSGA-II generates the hyper-plane.

Given a set P_t of individuals, MO-NSGA-II first identifies the minimum value of each of the M objective functions by $f_i^* = \min \{f_i(x) \mid x \in P_t\}$, $i = 1 \ldots M$. The ideal point z^* is defined by $z^* = (f_1^*, f_2^*, \ldots, f_M^*)$. Then, for each individual x in P_t, it calculates the translated objective $f_i'(x) = f_i(x) - f_i^*$. Now, the ideal point becomes the zero point in the translated coordinate system with M axes (objectives).

In the coordinate system, the extreme point along each objective axis is identified by minimizing the achievement scalarizing function (ASF), defined by

$$ASF(x, w) = \max_{j=1\ldots M} f_i'(x) / w_i \text{, for } x \in P_t. \quad (1)$$

The weight vector w is the axis direction, and in the implementation a small positive value 10^{-6} replaces the zero value of w_i. When $M = 3$, for example, the weight vector w for axis f_1 is $(1, 10^{-6}, 10^{-6})$. Finally, M extreme points are found and are used to create a hyper-plane. The hyper-plane will be extended to reach the translated objective axes.

B. Reference Points and Clustering

MO-NSGA-II follows Das and Dennis' method [36] to generate a set of reference points. For each element r_i ($i = 1 \ldots M$) of a reference point r, $r_i \in \{0, 1/p, 2/p, \ldots, 1\}$ and the sum of r_i is 1. The total number H of reference points is determined by M and p:

$$H = \binom{M + p - 1}{p}. \quad (2)$$

In MO-NSGA-II, the population size is set roughly equal to the number of reference points. Users can set the value of p based on how many non-dominated solutions they want to obtain.

MO-NSGA-II uses reference points to assess the distribution of individuals in the objective space and to guide the search process. All individuals being considered are projected onto the constructed hyper-plane. Then, each projected counterpart is associated with the closest reference point. In other words, each reference point has a cluster of associated solutions. With H reference points, ideally each reference point should have $\rho_{\text{ideal}} = N/H$ solutions in its cluster for even distribution, where N is the population size. For each reference point r, we denote its associated cluster of solutions by $C(r)$, and its deficient count $d(r)$ is defined by

$$d(r) = \rho_{\text{ideal}} - |C(r)|. \quad (3)$$

If $d(r)$ is positive, it means that the number of individuals around r is not enough. (It also means that too many individuals are around some other reference points.) The selection mechanism will favor individuals associated with the reference points having large $d(r)$ to distribute individuals evenly.

C. Mating Selection

Mating selection chooses the individuals from the current population as the parents for producing the offspring. MO-NSGA-II uses binary tournament to select the parents: two individuals are picked up randomly, and the better one is selected as a parent.

In MO-NSGA-II, two individuals are compared by three criteria hierarchically. First, an individual is better if it has a smaller non-domination level. If they have the same level, an individual x is better if its associated reference point $r(x)$ has a larger deficient count $d(r(x))$. Finally, if the above two criteria cannot determine the winner, the individual x with a smaller $ASF(x, r(x))$ value is selected.

D. Environmental Selection

After generating offspring, environmental selection determines which individuals will survive to the next generation. Like NSGA-II, MO-NSGA-II first puts the individuals in the first non-domination level F_1 into the survival population, then F_2, F_3, until adding the individuals in F_l exceeds the population size. Individuals in F_1 to $F_{(l-1)}$ survive, and the deficiency count of their reference points are calculated accordingly. Among the reference points, the one r^* with the largest deficient count is identified. Then, the individual x in F_l that has the minimal $ASF(x, r^*)$ is selected to survive. The deficient count of r^* is reduced by one. The above process is repeated until the size of the survival population reaches N.

IV. EMS-MO-NSGA-II

We enhance mating selection of MO-NSGA-II in two ways: one uses a reference point-based selection procedure to put more focus on the regions with fewer individuals and lower convergence; the other applies the neighborhood concept to encourage mating of individuals with similar objective tendency. The proposed algorithm is named EMS-MO-NSGA-II, which stands for Enhanced-Mating-Selection-MO-NSGA-II. Details will be given in the following two sub-sections, respectively.

A. Reference Point-based Selection

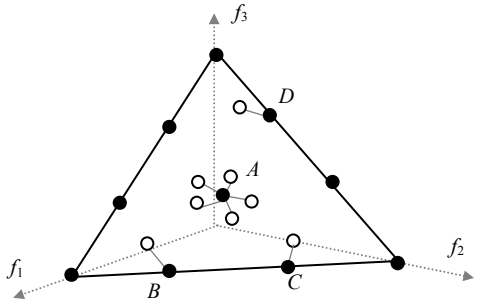

Fig. 1. Motivation of reference point-based selection

The binary tournament selection in MO-NSGA-II picks up two individuals randomly and then selects the one in a less-crowded region (i.e. the one whose reference point has a larger deficient count) when the two individuals have the same non-domination level. However, a region with more individuals is tried more frequently, and this may contradict the intention of preferring searching the less-crowded region. Take Fig. 1 as an example. Black circles and white circles refer to reference points and projected solutions, respectively. The cluster size of the central reference point A is five, and the size of the clusters of other three references points B, C, and D, is one. The binary tournament will select individuals in the cluster of A as parents in probability $(5/8)^2 = 39\%$, but will select individuals in each of the other three clusters in probability at most $1 - (7/8)^2 = 23\%$.

We modify the (individual-based) selection in MO-NSGA-II to the reference point-based selection, which chooses a reference point r first and then chooses an individual in its cluster $C(r)$. A similar concept was mentioned in PESA2 [37]. According to how we choose the reference point, we have three strategies, RP-U, RP-P, and RP-W.

1) Uniform selection (*RP-U*): This strategy chooses a reference point r with the same probability, regardless of how many individuals are associated with it. Then, an individual in its cluster $C(r)$ is chosen randomly as a candidate in tournament.

2) Proportionate selection (*RP-P*): This strategy chooses a reference point r with a probability in proportion to $P(r)$:

$$P(r) = \begin{cases} MCS - |C(r)| + 1, & \text{if } |C(r)| > 0 \\ 0, & \text{if } |C(r)| = 0 \end{cases}, \quad (4)$$

where MCS is the maximum of $|C(r)|$ among all reference points. RP-U gives the equal chance to clusters with different sizes, and RP-P gives more chance to clusters with smaller sizes. An individual in the cluster of the chosen reference point is chosen randomly as a candidate.

The following is an example of the probability of selecting an individual as a candidate in binary tournament. Assume that there are three reference points r_1, r_2, and r_3. The population size is 30, and the cluster sizes are 15, 10, and 5, respectively. The probability of selecting an individual in each cluster as a candidate in tournament is given in Table I. Since MO-NSGA-II selects each individual randomly, the probability is always 1/30. RP-U selects each cluster in equal probability and then selects an individual randomly, and thus the probability is $(1/3)\times(1/|C(r)|)$. In RP-P, $P(r_i) = 1$, 6, and 11, respectively, for $i = 1$, 2, and 3. We can see that individuals in the cluster with a smaller size are selected as candidates in higher probability. Note that although RP-U selects each cluster in equal probability, the "reference-point-first-individual-second" selection mechanism already increases the probability of selecting each individual in r_3 as a candidate in tournament from 1/30 to 1/15.

TABLE I
EXAMPLE OF SELECTION PROBABILITY OF INDIVIDUALS IN CLUSTERS AS CANDIDATES IN TOURNAMENT

	r_1 ($C(r_1) = 15$)	r_2 ($C(r_2) = 10$)	r_3 ($C(r_3) = 5$)
MO-NSGA-II	1/30	1/30	1/30
RP-U	(1/3)×(1/15)	(1/3)×(1/10)	(1/3)×(1/5)
RP-P	(1/18)×(1/15)	(6/18)×(1/10)	(11/18)×(1/5)

3) Weak region selection (RP-W): When the cluster sizes of reference points are close, RP-U and RP-P turn to be the same as the selection mechanism in MO-NSGA-II. In addition to the cluster size, we can assess the convergence by the ASF values of individuals. We calculate $ASF(r)$ for each

reference point r by

$$ASF(r) = \begin{cases} \min\{ASF(x,r) \mid x \in C(r)\}, & \text{if } |C(r)| > 0 \\ 0, & \text{if } |C(r)| = 0 \end{cases} \quad (5)$$

Then, we calculate the average value $\overline{ASF}$ over all $ASF(r)$. We think that individuals in the clusters of reference points r with $ASF(r) > \overline{ASF}$ are not positioned along the direction well. Therefore, we give mating chances only to these individuals. The RP-W strategy randomly chooses a reference point r from those with $ASF(r) > \overline{ASF}$, and then randomly chooses an individual in $C(r)$ as a candidate in tournament.

We test three variants of this strategy. They are different in the timing of applying RP-W.

RP-W-a: It applies RP-W during the whole evolution.
RP-W-d: It starts to apply RP-W after each reference point is associated with at least one individual.
RP-W-g: It starts to apply RP-W after g% of generations.

B. Neighborhood-based Selection (NB)

The second way of enhancing mating selection in MO-NSGA-II is to introduce the neighborhood concept. It has been observed that mating of individuals that are close in the objective space helps to produce better offspring effectively [5][39]. In our EMS-MO-NSGA-II, the first parent p_1 can be chosen by the original MO-NSGA-II method or any of our three reference point-based methods. Then, the second parent p_2 must be chosen from the clusters of n_T reference points closest to $r(p_1)$ in terms of the Euclidean distance in the objective space. If all the n_T neighbor reference points have an empty cluster, p_2 is chosen in the same way in which p_1 is chosen. The value of parameter n_T will be investigated in Section V-E. Algorithm II summarizes the flow of EMS-MO-NSGA-II.

ALGORITHM II: EMS-MO-NSGA-II

```
W: a set of weight vectors

P_1 ← GenerateInitialPopulation()
t ← 1
R ← GenerateReferencePoints()     // Section III-B
while stopping criterion is not satisfied
  H ← GenerateHyperplane(P_t)     // III-A
  AssociateIndividualsWithReferencePoints(P_t, H, R)   // III-B
  Q_t ← ∅
  NonDominatedSorting(P_t)   // [1]
  for i = 1 to |P_t|/2
    p_1 ← MatingSelection(P_t, R)   // IV-A
    if neighborhood-based selection is activated
      B ← CollectNeighborIndividuals(P_t, R, r(p_1), n_T)   // IV-B
    else
      B ← P_t
    p_2 ← MatingSelection(B, R)   // IV-A
    end if
    {c_1, c_2} ← CrossoverAndMutation(p_1, p_2)
    Q_t ← Q_t ∪ {c_1, c_2}
  end for
  P_{t+1} ← EnvironmentalSelection(P_t, Q_t, H, R)   // III-D
  t ← t + 1
end while
```

V. EXPERIMENTS AND RESULTS

A. Benchmarks and Performance Metric

To verify the performance of the proposed EMS-MO-NSGA-II, we compared it with MO-NSGA-II on four well-known problems DTLZ1-4 [40]. Definitions of these problems are given in Table II. In these problems, function $g(X_M)$ requires $|X_M| = k$ variables. The total number of decision variables n is $k + (M - 1)$. We set k by 5 for DTLZ1 and DTLZ3 and by 10 for DTLZ2 and DTLZ 4. We followed the experiments in [19] and tested $M = \{3, 5, 8, 10\}$ for DTLZ1-3 and $M = \{3, 5, 8\}$ for DTLZ4. DTLZ4 with 10 objectives was not considered in [19] and is not included here. In total, there were 15 problem instances.

We measured performance by the inverted generational distance (IGD) because it is commonly adopted in the EMO literature [5][6][12][23][31] and can be computed efficiently. The definition is as follows:

$$\text{IGD} = \left(\sqrt{\sum_{i=1}^{|P^*|} d_i^2} \right) / |P^*|, \quad (6)$$

where P^* is a reference set of Pareto optimal points and d_i is the Euclidean distance between the i^{th} point in P^* and the nearest solution in the approximated set of solutions obtained by a tested algorithm. The Pareto front of each test problem is known. We connect the zero point and the reference points (Section III-B) to form lines. The intersection of these lines and the Pareto front form the reference set P^*. For example, the Pareto front of the 3-objective ($M = 3$) DTLZ1 problem lies on the linear hyper-plane $f_1 + f_2 + f_3 = 0.5$. If we set $p = 2$, we will generate $C(3+2-1, 2) = 6$ points in the reference set: {(0.5, 0, 0), (0.25, 0.25, 0), (0.25, 0, 0.25), …, (0, 0, 0.5)}.

TABLE II
PROBLEM INSTANCES: DTLZ1–4 [40]

Problem	Definition
DTLZ1	Minimize $f_1(\mathrm{X}) = \frac{1}{2} x_1 x_2 \ldots x_{M-1}(1+g(\mathrm{X}_M))$, Minimize $f_2(\mathrm{X}) = \frac{1}{2} x_1 x_2 \ldots (1-x_{M-1})(1+g(\mathrm{X}_M))$, … Minimize $f_{M-1}(\mathrm{X}) = \frac{1}{2} x_1 (1-x_2)(1+g(\mathrm{X}_M))$, Minimize $f_M(\mathrm{X}) = \frac{1}{2}(1-x_1)(1+g(\mathrm{X}_M))$, subject to $0 \le x_i \le 1$, for $i = 1, 2, \ldots, n$. $g(\mathrm{X}_M) = 100\left[|\mathrm{X}_M| + \sum_{x_i \in \mathrm{X}_M} (x_i - 0.5)^2 - \cos(20\pi(x_i - 0.5))\right]$.
DTLZ2	Minimize $f_1(\mathrm{X}) = (1+g(\mathrm{X}_M)) \cos(x_1\pi/2) \ldots \cos(x_{M-1}\pi/2)$, Minimize $f_2(\mathrm{X}) = (1+g(\mathrm{X}_M)) \cos(x_1\pi/2) \ldots \sin(x_{M-1}\pi/2)$, … Minimize $f_M(\mathrm{X}) = (1+g(\mathrm{X}_M)) \sin(x_1\pi/2)$, subject to $0 \le x_i \le 1$, for $i = 1, 2, \ldots, n$. $g(\mathrm{X}_M) = \sum_{x_i \in \mathrm{X}_M} (x_i - 0.5)^2$.
DTLZ3	Use the same definition as DTLZ2 but replace g function by that in DTLZ1.
DTLZ4	Use the same definition as DTLZ2 but replace x_i by x^α in f_k ($k = 1 \ldots M$). ($\alpha = 100$ in the experiments.)

B. Parameter Setting

To compare our EMS-MO-NSGA-II with MO-NSGA-II and focus on the mating selection mechanism, we used the same crossover and mutation operators as those in MO-NSGA-II [19]. We also followed the same parameter setting: crossover rate was set by 1.0, and mutation rate was $1/n$, where n is the number of decision variables. Values of parameters η_c in the simulated binary crossover and η_m in polynomial mutation were both set by 20. The population size (N) and the maximum number of generations (G) were set identically to the values used in [19]. Both algorithms were run to solve each problem instance for 20 times. We recorded the IGD values of 20 runs and carried out Mann Whitney U test with a significance level 0.05. To save space, we will not report the IGD values. Instead, we provide the number of instances that the compared algorithms have significantly difference performance. In Tables IV to VII, the notation x+y- means that EMS-MO-NSGA-II outperforms MO-NSGA-II significantly for x out of 15 problems and is outperformed for y problems.

TABLE III
PARAMETER SETTING OF MO-NSGA-II AND EMS-MO-NSGA-II

	Number of objectives (M)			
	3	5	8	10
number of divisions (p)	12	5	4	3
number of clusters	91	126	330	220
population size (N)	92	128	332	220
generation number (G) for DTLZ1	750	1500	2500	5000
generation number (G) for DTLZ2	750	1500	3500	4000
generation number (G) for DTLZ3	1000	3000	4500	6000
generation number (G) for DTLZ4	1500	3000	4000	

C. Experiment 1: RP-U/RP-P

In our first experiment, we checked the effects of the two selection strategies RP-U and RP-P. The difference between them and the mating selection mechanism in MO-NSGA-II is in that they select a reference point and then select an individual associated with that reference point. (Here the proposed NB strategy was not applied.) This aims at avoiding the high selection probability for a crowded region due to the original individual-based selection. Table IV shows the experimental results.

TABLE IV
NUMBER OF INSTANCES WITH STATISTICALLY SIGNIFICANT DIFFERENCE BETWEEN MO-NSGA-II AND THE VARIANT USING RP-U AND RP-P

	MO-NSGA-II	RP-U	RP-P
MO-NSGA-II		0+0-	0+0-
RP-U	0+0-		0+0-
RP-P	0+0-	0+0-	

The results indicate that the RP-U and RP-P strategies do not have significant effect on the performance of MO-NSGA-II. After observing the search process, we found that MO-NSGA-II is able to distribute population members evenly to reference points. When setting the population size roughly equal to the number of reference points (please see Table III), we found that almost every reference point was ideally associated with one individual. Besides, this even distribution was achieved at very early stage (about 5%–10% of the maximum generation number). The even distribution makes individual-based selection and reference point-based selection behave identically, which explains the results in Table IV.

D. Experiment 2: RP-W

In the second experiment we analyzed the effect of the proposed RP-W strategy. (Again, the proposed NB strategy was not applied.) Like RP-U and RP-P, RP-W selects a reference point first and then an individual associated with the reference point. The difference is in that RP-U and RP-P aims at improving distribution by considering the number of individuals associated with the reference points but RP-W focuses on improving convergence by considering the *ASF* value. Only references points whose *ASF* values are higher than the average *ASF* can be selected in the RP-W strategy. We have three RP-W variants. They are different in the timing of activating the strategy. RP-W-a activates it at the first generation, RP-W-d activates it when each reference point is associated with at least one individual, and RP-W-g activates it when g% of maximum generations is reached. In the experiment, we tested three values of g: 50, 70, and 90. Table V shows the experimental results.

TABLE V
NUMBER OF INSTANCES WITH STATISTICALLY SIGNIFICANT DIFFERENCE BETWEEN MO-NSGA-II (MO-NS-II) AND THE VARIANT USING RP-W

	MO-NS -II	RP-W -a	RP-W-d	RP-W -50	RP-W-70	RP-W -90
MO-NS-II		13+1-	13+0-	7+2-	2+7-	0+11-
RP-W-a	1+13-		0+9-	2+12-	1+14-	1+14-
RP-W-d	0+13-	9+0-		1+13-	1+13-	1+13-
RP-W-50	2+7-	12+2-	13+1-		0+10-	1+9-
RP-W-70	7+2-	14+1-	13+1-	10+0-		1+1-
RP-W-90	11+0-	14+1-	13+1-	9+1-	1+1-	

On one hand, the results indicate that RP-W is able to improve the performance. RP-W-70 is significantly better than MO-NSGA-II for 7 problem instances and is worse for 2 instances. RP-W-90 is better for 11 instances and worse for none. On the other hand, RP-W-a and RP-W-d do not perform well. RP-W-a activated the strategy at the beginning. According to the finding in the first experiment, we can know that RP-W-d activated the strategy early when 5%-10% of maximum generations was reached. The result implies that RP-W should be activated at the later stage of the search process. One possible reason is that RP-W strategy puts focus on searching some certain regions. If we do it too early, the search direction may be biased and the approximation would also become bad.

E. Experiment 3: NB

In this experiment, we want to verify the effects of the NB strategy, which allows a selected individual x to mate only with individuals that are associated with one of the n_T reference points closet to $r(x)$. To focus on the NB strategy, we did not apply our reference point-based selection here. We tested four values of n_T: $N\cdot\{5\%, 10\%, 15\%, 20\%\}$, where N is the population size. Table VI presents the experimental results.

TABLE VI
NUMBER OF INSTANCES WITH STATISTICALLY SIGNIFICANT DIFFERENCE BETWEEN MO-NSGA-II AND THE VARIANT USING NB

	MO-NSGA-II	NB-5	NB-10	NB-15	NB-20
MO-NSGA-II		2+10-	0+13-	1+12-	1+13-
NB-5	10+2-		2+6-	3+6-	5+5-
NB-10	13+0-	6+2-		5+0-	6+0-
NB-15	12+1-	6+3-	0+5-		7+0-
NB-20	13+1-	5+5-	0+6-	0+7-	

The results show that the NB strategy can improve the performance of MO-NSGA-II and is not sensitive to the value of n_T. The NB strategy leads to significantly better performance for at least 10 out of 15 problem instances.

F. Experiment 4: MO-NSGA-II vs. EMS-MO-NSGA-II

In the last experiment, we integrated both RP-W and NB strategies into MO-NSGA-II, leading to EMS-MO-NSGA-II. We set the value of n_T in the NB strategy by N·10%. Then, we tested several values of g for the RP-W strategy and finally set it by 80. The experimental results are given in Table VII. We can see that EMS-MO-NSGA-II significantly outperforms MO-NSGA-II for 12 problem instances and is outperformed by 1 instance. Although EMS-MO-NSGA-II looks slightly worse than NB-10 when both are compared with MO-NSGA-II, EMS-MO-NSGA-II is significantly better than NB-10 for five instances but worse for none. It shows that both of the proposed RP-W and NB strategies help to improve algorithm performance.

TABLE VII
NUMBER OF INSTANCES WITH STATISTICALLY SIGNIFICANT DIFFERENCE BETWEEN MO-NSGA-II AND PROPOSED EMS-MO-NSGA-II

	MO-NSGA-II	NB-10	EMS-MO-NSGA-II
MO-NSGA-II		0+13-	1+12-
NB-10	13+0-		0+5-
EMS-MO-NSGA-II	12+1-	5+0-	

Table VIII gives the mean and standard deviation of the IGD and additive ε-indicator values [27] over 20 runs for MO-NSGA-II and EMS-MO-NSGA-II. The ε-indicator is calculated by

$$I_{\varepsilon+} = \inf_{\varepsilon}\{\forall y \in P^*, \exists x \in P : y_i \geq x_i - \varepsilon, i = 1,...,M\}, \quad (7)$$

where P and P^* denote the approximation set and the reference set. Better mean values between the two tested algorithms are marked in bold. We can see that EMS-MO-NSGA-II has better mean IGD and $I_{\varepsilon+}$ for 14 and 11 out of 15 problem instances, respectively.

VI. CONCLUSIONS

MO-NSGA-II is a recently proposed EA for many-objective optimization. The core idea is to guide the search process by reference points on the hyper-plane constructed based on obtained solutions. In this paper, we enhanced the mating selection mechanism in MO-NSGA-II by turning the original individual-based selection to a hierarchical selection and incorporating the neighborhood concept. Experimental results showed that both ideas help to improve algorithm performance. The proposed EMS-MO-NSGA-II outperformed MO-NSGA-II in terms of IGD and additive ε-indicator for most of DTLZ1-4 instances with 3 to 10 objectives.

In our future work, an important topic is to study how to activate the proposed RP-W strategy automatically without setting a predefined parameter (g). In the experiments we found that MO-NSGA-II was able to distribute individuals to reference points evenly. We are also interested in investigating why the individual-based selection can achieve the even distribution. To have a deeper analysis of the performance of our algorithm, we will consider more problems such as the WFG problems [41]. Interaction effects between the reference point-based strategy and NB strategy are also worth more investigation.

TABLE VIII
MEAN AND STANDARD DEVIATION OF IGD AND ADDITIVE ε-INDICATOR OF MO-NSGA-II AND PROPOSED EMS-MO-NSGA-II ($\times 10^{-6}$)

		IGD		$I_{\varepsilon+}$	
	M	EMS-MO-NSGA-II	MO-NSGA-II	EMS-MO-NSGA-II	MO-NSGA-II
DTLZ1	3	**324** (274)	359 (223)	**841** (591)	1275 (824)
	5	**43** (40)	284 (149)	**132** (74)	1449 (892)
	8	**21** (13)	442 (64)	**413** (529)	4530 (4385)
	10	**13** (15)	116 (41)	**643** (935)	1365 (1093)
DTLZ2	3	**730** (208)	948 (179)	1660 (668)	**1431** (426)
	5	**369** (83)	1149 (233)	**2401** (2013)	6970 (4252)
	8	**87** (12)	868 (102)	**481** (245)	5972 (5993)
	10	**44** (8)	508 (80)	**336** (187)	2945 (2210)
DTLZ3	3	**871** (457)	1028 (626)	2388 (1550)	**1860** (729)
	5	**98** (28)	381 (50)	**486** (323)	2235 (1293)
	8	**160** (343)	653 (69)	5421 (10280)	**4923** (1622)
	10	248 (354)	**243** (52)	6312 (9434)	**2065** (1402)
DTLZ4	3	**53451** (193760)	53649 (163655)	**71423** (217860)	71736 (217882)
	5	**97** (26)	637 (774)	**604** (330)	5743 (6069)
	8	**28** (5)	390 (64)	**168** (76)	6365 (3218)

ACKNOWLEDGMENT

The authors are very grateful to the reviewers for their valuable and instructive comments. This research was supported by the Ministry of Science and Technology of Taiwan (R.O.C.) under Grant NSC101-2221-E-003-007.

REFERENCES

[1] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[2] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength pareto evolutionary algorithm for multiobjective optimization," in K.C. Giannakoglu et al. (eds.) *Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems*, 2002, pp. 95–100.

[3] E. Zitzler and S. Künzli, "Indicator-based selection in multiobjective search," in *Proc. International Conference on Parallel Problem Solving from Nature (PPSN VIII)*, pp. 832–842, 2004.

[4] N. Beume, B. Naujoks, and M. Emmerich, "SMS-EMOA: Multiobjective selection based on dominated hypervolume," *European Journal of Operational Research*, vol. 181, no. 3, pp. 1653–1669, 2007.

[5] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712–731, 2007.

[6] H. Li and Q. Zhang, "Multiobjective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 2, pp. 284–302, 2009.

[7] Deb, "Introduction to Evolutionary Multiobjective Optimization," in J. Branke et al. (Eds.) *Multiobjective Optimization* (LNCS 5252), 2008, pp. 59–96.

[8] R. C. Purshouse and P. J. Fleming, "Evolutionary many-objective optimization: an exploratory analysis," in *Proc. IEEE Congress on Evolutionary Computation (CEC)*, 2003, pp. 2066–2073.

[9] E. J. Hughes, "Evolutionary many-objective optimization: many once or one many," in *Proc. IEEE Congress on Evolutionary Computation (CEC)*, 2005, pp. 222–227.

[10] H. Ishibuchi, N. Tsukamoto, and Y. Nojima, "Evolutionary many-objective optimizations: a short review," in *Proc. IEEE Congress on Evolutionary Computation (CEC)*, 2008, pp. 2419–2426.

[11] D. Brockhoff and E. Zitzler, "Are all objectives necessary? On dimensionality reduction in evolutionary multiobjective optimization," in *Proc. Parallel Problem Solving from Nature (PPSN)*, 2006, pp. 533–542.

[12] A. L. Jaimes, C. A. C. Coello, and D. Chakraborty, "Objective reduction using a feature selection technique," in *Proc. Genetic and Evolutionary Computation Conference (GECCO)*, 2008, pp. 673–680

[13] A. L. Jaimes, C. A. C. Coello,.and J. E. U. Barrientos, "Online objective reduction to deal with many-objective problems," in *Proc. 5th International Conference on Evolutionary Multi-Criterion Optimization (EMO)*, 2009, pp. 423–437.

[14] D. K. Saxena, J. A. Duro, A. Tiwari, K. Deb, and Q. Zhang, "Objective reduction in many-objective optimization: linear and nonlinear algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 1, pp. 77–99.

[15] K. Deb, J. Sundar, N. Udar, and S. Chaudhuri, "Reference point based multi-objective optimization using evolutionary algorithms," *International Journal of Computational Intelligence Research*, vol. 2, no. 3, pp. 273–286, 2006.

[16] L. Thiele, K. Miettinen, P. J. Korhonen, and J. Molina, "A preference-based evolutionary algorithm for multiobjective optimization," *Evolutionary Computation*, vol. 17, no. 3, pp. 411–436, 2009.

[17] K. Sindhya, A. B. Ruiz, and K. Miettinen, "A preference based interactive evolutionary algorithm for multi-objective optimization: PIE," in *Proc. 6th International Conference on Evolutionary Multi-Criterion Optimization (EMO)*, 2011, pp. 212–225.

[18] R. Wang, R. C. Purshouse, and P. J. Fleming, "Preference-inspired coevolutionary algorithm for many-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 4, pp. 474–494, 2013.

[19] K. Deb and H. Jain, "Handling many-objective problems using an improved NSGA-II procedure," in *Proc. IEEE Congress on Evolutionary Computation (CEC)*, 2012.

[20] H. Trautmann, T. Wagner, D. Biermann, and C. Weihs, "Indicator-based selection in evolutionary multiobjective optimization algorithms based on the desirability index," *Journal of Multi-Criteria Decision Analysis*, vol. 20, no. 5-6, pp. 319–337, 2014.

[21] N. Drechsler, R. Drechsler, and B. Becker, "Multi-objective optimization based on favour," in *Proc. 1st International Conference on Evolutionary Multi-Criterion Optimization (EMO)*, 2001, pp. 154–166.

[22] J. Branke, T. Kaußler, and H. Schmeck, "Guidance in evolutionary multi-objective optimization," *Advances in Engineering Software*, vol. 32, no. 6, pp. 499–507, 2001.

[23] H. Sato, H. E. Aguirre, and K. Tanaka, "Controlling dominance area of solutions and its impact on the performance of MOEAs," in *Proc. 4th International Conference on Evolutionary Multi-Criterion Optimization (EMO)*, 2007, pp. 5–20.

[24] D. Corne and J. Knowles, "Techniques for highly multiobjective optimisation: Some nondominated points are better than others," in *Proc.Genetic and Evolutionary Computation Conference (GECCO)*, 2007, pp. 773–780.

[25] P. J. Bently and J. P. Wakefield, "Finding acceptable solutions in the Pareto-optimal range using multiobjective genetic algorithms," In: Chawdry et al., (eds.) *Soft Computing in Engineer Design and Manufacturing*, 1997, pp. 231-240.

[26] M. Garza-Fabre, G. T. Pulido, and C. A. C. Coello, "Ranking methods for many-objective optimization," in *Proc.8th Mexican International Conference on Artificial Intelligence (MICAI)*, 2009, pp. 633–645.

[27] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. Grunert da Fonseca, "Performance assessment of multiobjective optimizers: an analysis and review," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 2, pp. 117−132, 2003.

[28] T. Wagner, N. Beume, and B. Naujoks, "Pareto-aggregation-, and indicator-based methods in many-objective optimization," in *Proc. 4th International Conference on Evolutionary Multi-Criterion Optimization (EMO)*, 2007, pp. 742–756.

[29] A. Díaz-Manríquez, G. Toscano-Pulido, C. A. C. Coello, and R. Landa-Becerra, "A ranking method based on the R2 indicator for many-objective optimization," in *Proc. IEEE Congress on Evolutionary Computation (CEC)*, 2013, pp. 1523–1530.

[30] M. P. Hansen and A. Jaszkiewicz, "Evaluating the quality of approximations of the non-dominated set," Technical University of Denmark, Technical Report IMM-REP-1998-7, 1998.

[31] D. H. Phan and J. Suzuki, "R2-IBEA: R2 indicator based evolutionary algorithm for multiobjective optimization," in *Proc. IEEE Congress on Evolutionary Computation (CEC)*, 2013, pp. 1836–1845.

[32] E. J. Hughes, "MSOPS-II: a general-purpose many-objective optimizer," in *Proc. IEEE Congress on Evolutionary Computation (CEC)*, 2007, pp. 3944–3951.

[33] H. Ishibuchi, T. Doi, and Y. Nojima, "Incorporation of scalarizing fitness functions into evolutionary multiobjective optimization algorithms," in *Proc. Parallel Problem Solving from Nature (PPSN)*, 2006, pp. 493–502.

[34] H. Ishibuch, Y. Sakane, N. Tsukamoto, and Y. Nojima, "Evolutionary many-objective optimization by NSGA-II and MOEA/D with large populations," in *Proc. IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2009, pp. 1758–1763.

[35] H. Ishibuchi, Y. Sakane, N. Tsukamoto, and Y. Nojima, "Adaptation of scalarizing functions in MOEA/D: an adaptive scalarizing function-based multiobjective evolutionary algorithm," in *Proc. 5th International Conference on Evolutionary Multi-Criterion Optimization (EMO)*, 2009, pp. 438–452.

[36] I. Das and J. Dennis, "Normal-boundary intersection: a new method for generating the Pareto surface in nonlinear multicriteria optimization problems," *SIAM Journal of Optimization*, vol. 8, no. 3, pp. 631 – 657, 1998.

[37] D. W. Corne, N. R. Jerram, J. K. Knowles, and M. J. Oates, "PESA-II: region-based selection in evolutionary multiobjective optimization," in *Proc. Genetic and Evolutionary Computation Conference (GECCO)*, 2001, pp. 283–290.

[38] H. Jain and K. Deb, "An improved adaptive approach for elitist nondominated sorting genetic algorithm for many-objective optimization," in *Proc. 7th International Conference on Evolutionary Multi-Criterion Optimization (EMO)*, 2013, pp. 307–321.

[39] H. Ishibuchi, N. Akedo, and Y. Nojima, "Recombination of similar parents in SMS-EMOA on many-objective 0/1 knapsack problems," in *Proc. Parallel Problem Solving from Nature (PPSN)*, 2012, pp. 132–142.

[40] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable test problems for evolutionary multi-objective optimization," in A. Abraham et al. (Eds.) *Evolutionary Multi-objective Optimization*, pp. 105–145, 2005.

[41] S. Huband, L. Barone, L. While, and P. Hingston, "A scalable multiobjective test problem toolkit," in *Proc. 3rd International Conference on Evolutionary Multi-Criterion Optimization (EMO)*, 2005, pp. 280–294.