Multiobjective Real-coded Bayesian Optimization Algorithm Revisited: Diversity Preservation

Chang Wook Ahn, R. S. Ramakrishna
Gwangju Institute of Science & Technology
Department of Information & Communications
Oryong-dong, Gwangju 500-712 Korea
{cwan,rsr}@gist.ac.kr

ABSTRACT

This paper provides empirical studies on MrBOA, which have been designed for strengthening diversity of nondominated solutions. The studies lead to modified sharing. A new selection scheme has been suggested for improving diversity performance. Empirical tests validate their effectiveness on uniformity and front-spread (i.e., diversity) of nondominated set. A diversity-preserving MrBOA (dp-MrBOA) has been designed by carefully combining all the promising components; i.e., modified sharing, dynamic crowding, and diversity-preserving selection. Experiments demonstrate that dp-MrBOA is able to significantly improve diversity performance (for the scaling problems), without weakening proximity of nondominated set.

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search—*Heuristic methods*

General Terms

Algorithms, Design, Experimentation

Keywords

Diversity preservation, Evolutionary algorithms, Multiobjective optimization, Probabilistic models, Scaling

1. INTRODUCTION

Real-world problems involve incommensurable and often competing objectives [1, 3–7]. They are known as *multiobjective* (or *multicriteria*) *optimization problems* (MOPs). In general, a real-valued MOP can be formulated as follows:

$$\begin{aligned} \min \quad & \vec{z} = \vec{f}(\vec{y}) = (f_1(\vec{y}), f_2(\vec{y}), \cdots, f_m(\vec{y})) \\ \text{s.t.} \quad & \vec{e}(\vec{y}) = (e_1(\vec{y}), e_2(\vec{y}), \cdots, e_c(\vec{y})) \geq 0 \end{aligned} \quad (1)$$

where $\vec{y} = (y_1, y_2, \cdots, y_n) \in \mathbb{R}^n$ is the *decision space*, and $\vec{z} = (z_1, z_2, \cdots, z_m) \in \mathbb{R}^m$ is the *objective space*. The set of decision vectors that satisfy the constraints is the *feasible set* $\mathcal{Y}$. The vector function $\vec{f} : \mathcal{Y} \rightarrow \mathbb{R}^m$ defines the feasible region $\mathcal{Z} = \{\vec{f}(\vec{y}) | \vec{y} \in \mathcal{Y}\}$. Due to the interdependence of the objectives, MOPs normally have a set of alternative solutions. These solutions are optimal in the sense that no solution is superior to them in an overall sense because no objective can be realized without losing on the others. The set of solutions is known as *Pareto optimal set* or *nondominated set* (of solutions). It is mathematically defined by

$$\mathcal{Q} = \{\vec{y}^0 \in \vec{\mathcal{Y}} | \nexists \vec{y}^1 \in \vec{\mathcal{Y}} : \vec{y}^1 \succ \vec{y}^0\} \quad (2)$$

where $\vec{y}^1 \succ \vec{y}^0$ indicates that the solution $\vec{y}^1$ *dominates* the solution $\vec{y}^0$; i.e., $\forall i : f_i(\vec{y}^0) \geq f_i(\vec{y}^1) \wedge \exists j : f_j(\vec{y}^0) > f_j(\vec{y}^1)$. The *Pareto* (*optimal*) *front* which is the image of the Pareto optimal set under the feasible objective space $\mathcal{Z}$ is given by,

$$\mathcal{F} = \{\vec{f}(\vec{y}^0) | \vec{y}^0 \in \mathcal{Q}\}. \quad (3)$$

Note that the goal of multiobjective optimization is to find the *global* Pareto optimal set $\mathcal{Q}^*$. It is identical to place nondominated solutions on the *true* (or *global*) Pareto front $\mathcal{F}^*$. However, achieving the goal is not practical since the number of Pareto optimal solutions is infinite. Thus, the down-to-earth goal is to discover representative nondominated solutions of the true Pareto front while maintaining a good spread of solutions over the front [1, 4–6].

Recently, multiobjective genetic and evolutionary algorithms (MGEAs) have attracted due attention due to their ability to search for multiple solutions in parallel as well as handle complex features such as discontinuity and multimodality [6, 9, 16]. The growing interest in difficult, higher dimensional problems has spurred the growth of MGEAs for over a decade. MGEAs can be broadly divided into two categories – *non-model-based* and *(probabilistic) model-based* approaches. The former attempts to improve proximity of the Pareto front by exploiting the domination information of individuals and preserve diversity of the solutions by employing a sharing strategy. Further, recent variants try to incorporate elitism in harmony and to take into account domination and density information at the same time, thereby improving both proximity and diversity. However, the approach may not be efficient for some complicated problems as it does not pay enough attention to linkage information[1] of the problems. *Nondominated sorting genetic algorithm* (NSGA) [13], *Multiobjective genetic algorithm* (MOGA) [7], *Strength Pareto evolutionary algorithm*

[1]It plays a key role in growing good partial solutions toward the global optima.

GECCO'07, July 7–11, 2007, London, England, United Kingdom.

(SPEA) [16], SPEA-II [15], NSGA-II [6], and *rank-density-based genetic algorithm* (RDGA) [9] are included in this category. On the other hand, the model-based approach concentrates on effectively combining the strengths of the state-of-the-art MGEAs with the EDAs' ability of automatic discovery and exploitation of problem regularities. The approach generally outperforms the other approaches with regard to both proximity and diversity, by virtue of its linkage-friendly evolution (of EDAs) and efficient fitness assignment policy (of competent MGEAs). In this sense, the approach is coming into limelight of late. *Bayesian mutiobjective optimization algorithm* (BMOA) [10], *multiobjective* (*hierarchical*) *Bayesian optimization algorithm* (m(h)BOA) [8, 12], *multiobjective mixture-based iterative density-estimation evolutionary algorithm* (MIDEA) [3, 4], and *multiobjective real-coded Bayesian optimization algorithm* (MrBOA) [1] are leading techniques.

In essence, MrBOA extends up-to-date real-coded BOA (rBOA) with a view to dealing with MOPs without in any way diluting its unique features; i.e., problem decomposition and probabilistic building-block crossover in the continuous space. In this regard, the rBOA is combined with the selection mechanism of NSGA-II. In order to further promote diversity performance, the selection scheme is enhanced by assigning fitness values by harmonizing domination ranks with sharing intensity measures (computed by adaptive sharing scheme) and modified crowding distances (measured by dynamic crowding technique). However, experiments in [1] have only focused on verifying the capability of dealing with problem regularities (in the continuous space), which denote statistical dependencies of variables (i.e., variable linkages).[2] In other words, a test on the effect of such operators proposed for diversity preservation has been disregarded. With this in view, this paper minutely examines, to begin with, what contributions the operators make to diversity promotion. On the basis of the investigations, a new sharing scheme that can help in maintaining diversity is suggested. A new selection method that significantly improves diversity performance has been developed. Finally, diversity-preserving MrBOA (dp-MrBOA) is investigated by a proper combination of such operators.

The rest of the paper is organized as follows. Section 2 briefly describes the central concepts of MrBOA. Section 3 takes a close look at empirical results provided by MrBOA's operators. Also, some modifications are incorporated for diversity preservation. Section 4 presents the diversity-preserving methodology. The paper concludes with a summary in Sect. 5.

2. A REVIEW OF MRBOA

This section briefly reviews the MrBOA developed for treating real-valued MOPs (RMOPs).

2.1 MrBOA Outline

The MrBOA extends the rBOA [2] into the realm of multi-objective optimization without blurring its unique features. Note that the rBOA has been developed in an effort to bring all the inherent strengths of (discrete) BOA into the continuous world. Thus, the rBOA performs proper (problem) decomposition by means of a Bayesian factorization and probabilistic building-block crossover by employing mixture models at the level of subproblems. A more detailed information on rBOA can be found in [1, 2]. Drawing on the procedures of rBOA, an outline of MrBOA is provided by the following pseudo-code.

STEP 1. INITIALIZATION
  Randomly generate initial population $\vec{\mathcal{P}}$
STEP 2. SELECTION
  Select a set of promising candidates $\vec{\mathcal{S}}$ from $\vec{\mathcal{P}}$
  2.1. PARETO RANKING
    Ranks $\vec{\mathcal{R}}$ are found by the nondominated sorting
  2.2. ADAPTIVE SHARING
    Sharing intensity $\vec{\mathcal{I}}$ is computed by adaptive sharing
  2.3. DYNAMIC CROWDING
    Crowding distances $\vec{\mathcal{D}}$ by dynamic crowding
  2.4. FITNESS ASSIGNMENT
    Fitness is assigned from $\vec{\mathcal{R}}$, $\vec{\mathcal{I}}$, and $\vec{\mathcal{D}}$
  2.5. ELITISM
    The elitist solutions are selected
STEP 3. LEARNING
  Learn a probabilistic model $\mathcal{M}$ from $\vec{\mathcal{S}}$ using a metric
STEP 4. SAMPLING
  Generate a set of offspring $\vec{\mathcal{O}}$ from the learned model
STEP 5. REPLACEMENT
  Create a new $\vec{\mathcal{P}}$ by replacing some of $\vec{\mathcal{P}}$ with $\vec{\mathcal{O}}$
STEP 6. TERMINATION
  If the termination criteria are not met, go to STEP 2

All the procedures except for the selection in STEP 2 are the same as those of rBOA. It implies that the selection procedure imparts to rBOA the capability to handle multiple objectives. Moreover, model learning and sampling methods (of rBOA) provide the MrBOA with necessary tools for discovering problem regularities and achieving the maximum BB-wise mixing rate in the real-valued multiobjective optimization. Thus, it is very important to develop an efficient selection scheme in such competent EDAs.

2.2 Details of Selection

The selection of MrBOA leads current solutions (i.e., individuals) towards the set of nondominated solutions.

2.2.1 Pareto Ranking

In MGEAs, ranking is of fundamental importance since it is closely related to fitness assignment to individuals. Many ranking schemes have been developed for achieving close convergence and uniform spread to the true (Pareto) front $\mathcal{F}^*$ [6, 7, 9, 15]. The MrBOA has employed the nondominated sorting method (i.e., pure Pareto ranking scheme) of NSGA-II, due to its simplicity and effectiveness. It decides domination ranks $\vec{\mathcal{R}}$ of individuals in such a manner that all the nondominated individuals in the population are assigned rank 1 (known as the *first* (Pareto), i.e., nondominated, front $\mathcal{F}_1$) and removed from temporary assertion, then a new set of nondominated individuals is assigned rank 2 (viz., the *second* front $\mathcal{F}_2$), and so forth. A solution with a lower rank is always preferred. An example of the Pareto ranking can be found in Fig. 3. Along with the ranks of individuals, it is necessary to discriminate among solutions with equal domination ranks. In the NSGA-II, a simple (but efficient) crowding method is applied to the individuals belonging to

[2]Recently, there has been another interesting attempt which extracts and exploits the regularity patterns of the Pareto set (in the decision space) using local principal component analysis [14].

identical fronts. The solutions with identical ranking are distinguished by comparing crowding distances (as a density estimate). However, its effect on the selection of individuals is definitely secondary to that of their domination ranks. Thus, some information that can facilitate diversity preservation must be more actively brought in for selecting promising individuals. In this regard, *adaptive sharing* and *dynamic crowding* described below have been developed [1].

2.2.2 Adaptive Sharing

Adaptive sharing is a way of estimating density information of individuals. It is based on the *domination count* [3,7]. The domination count of an individual is defined by the number of individuals in the current population by which it is dominated. The goal of the sharing scheme is to boost the solutions that are less dominated since they generally stand for their dominated solutions. In a similar way, an individual can also incorporate the information about the number of individuals which are dominated by it[3] (a solution that dominates more individuals is preferred) [5]. However, it risks being a primary factor. The reason can be found in [1]. Thus, the domination count is sufficient to distinguish representative individuals of a population. Helpful density information can be adaptively computed by employing the domination count alone. As a measure of density, *sharing intensity* is defined as follows:

$$\mathcal{I}(i) = 1 - \frac{1}{1 + N_{dom}(i)} \tag{4}$$

where $\mathcal{I}(i)$ is the sharing intensity of individual i, $N_{dom}(i)$ is the domination count of the individual i. A lower value is assigned to an individual that serves as a representative of its objective space (covered by the solution) regardless of the number of dominated individuals in that region. An example of sharing intensity of individuals can be found in Fig. 2. It is seen that lower values are assigned to individuals closer to nondominated solutions having smaller number of neighbors (i.e., less crowded). In other words, the individuals that can promote diversity preservation (and improve proximity as a bonus) are preferred.

2.2.3 Dynamic Crowding

It has been noted that the crowding method[4] of NSGA-II is quite effective in stimulating a diverse representation of the nondominated solutions (in $\mathcal{F}_1$). In this regard, the dynamic crowding has been established on the basis of the crowding scheme of NSGA-II. After sorting the individuals on objective function values, the crowding method is applied to the first front $\mathcal{F}_1$ in order to discriminate between the nondominated solutions (that have the same domination rank '1' and sharing intensity '0'). As a density measure, *crowding distance* is computed by

$$\mathcal{D}(\vec{\omega}_i) = \sum_{k=1}^{v} \frac{\prod_{l=0}^{1} \{f_k(\vec{\omega}_{i+l}) - f_k(\vec{\omega}_{i-1})\}}{(f_k^{max} - f_k^{min})^2} \tag{5}$$

where $\vec{\omega}_i$ is the ith individual in the sorted set of nondominated solutions, $\mathcal{D}(j)$ is the crowding distance of the individual j, $f_k(j)$ is the kth objective function value of the individual j. In addition, f_k^{min} (f_k^{max}) is the minimum (maximum)

[3]In this study, this is referred to as 'nondomination count.'
[4]In this paper, it is called the 'simple crowding' method.

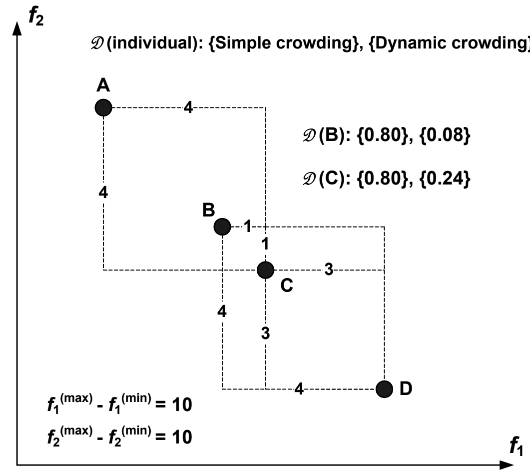


Figure 1: Example of crowding distance.

value of the kth objective function in the current nondominated set.[5] For each objective, as a matter of course, the first and last individuals are assigned an infinite distance to give an absolute preference to boundary solutions. An individual with a larger value is always preferred because it is regarded as a less crowded (i.e, representative) individual that can well approximate the nondominated solutions. On the other hand, its benefit over the simple crowding of [6] is illustrated in Fig. 1. It is seen that the existing crowding assigns to individuals B and C the same values so that both individuals are equally preferred. But, it is sufficient if one of them survives for a good approximation of the solutions in $\mathcal{F}_1$ because the two solutions are quite close (i.e., crowded). The dynamic crowding assigns to the individual C a crowding distance that is larger than that assigned to the individual B. That is, C is preferred to B. Therefore, the dynamic crowding method provides an efficient framework for promoting diversity of the nondominated solutions.

2.2.4 Fitness Assignment and Elitism

It is important to note that domination rank $\vec{\mathcal{R}}$ is a primary component of fitness assignment while sharing intensity $\vec{\mathcal{I}}$ and crowding distance $\vec{\mathcal{D}}$ prevent thickly crowded and dominated individuals from surviving the selection process. Taking into account their effect on preference of individuals, a fitness function is defined as follows:[6]

$$f(i) = \begin{cases} \mathcal{R}(i)\frac{1}{1+\mathcal{D}(i)}, & \text{if } i \in \mathcal{F}_1 \\ \mathcal{R}(i) + \gamma\mathcal{I}(i), & \text{otherwise.} \end{cases} \tag{6}$$

Here, $f(i)$ denotes the fitness value of individual i, $\mathcal{R}(i)$ is the domination rank of individual i, γ is a regularization parameter of penalty, and $\mathcal{I}(i)$ and $\mathcal{D}(i)$ are the sharing intensity and the crowding distance of individual i, respectively. An individual with a lower fitness value is always

[5]This is the normalization technique based on each local front suggested in [11]. It effectively mitigates barriers when encountering problems with badly scaled objective functions.
[6]In the original study, the fitness rule for $\mathcal{F}_1$ is that $f(i) = \mathcal{R}(i)\left(1 + \frac{1}{1+\mathcal{D}(i)}\right)$. However, they have no difference in performing individual selection.

preferable. As the value of parameter γ increases, sharing effect on the fitness grows so that diversity can be stimulated. In the reverse case, proximity can be emphasized due to the growth of the effect of domination rank on the fitness. A proper setting of γ may be crucial. The parameter value should not be too large.

Elitism allows the best solutions of the current generation to be copied into the next generation. In MGEAs, it plays an important role in further advancing the nondominated solutions due to the availability of (equally preferable) multiple solutions [4, 6, 10, 16]. In other words, elitism directly contributes to *exploration* because it preserves superior individuals which are difficult to generate again (when getting lost in a certain generation) [4]. Moreover, it can also help in *exploitation* as it determines the individuals which are selected to survive a generation [4]. Note that an easy (but efficient) approach for incorporating elitism into MGEAs is to employ *truncation selection*.

3. EXPERIMENTS AND MODIFICATION

This section takes a close look at key components of MrBOA through experiments. Some modifications are considered for performance improvement.

3.1 Test Problem

In [1], important features of MrBOA such as linkage-friendly evolution in MOPs have been tested on various decomposable problems. Experimental studies demonstrated its ability to compete with up-to-date MGEAs such as NSGA-II and MIDEA. However, the benefit seems to come, not from the operators suggested in [1], but from the inherent characteristics of (single-objective) rBOA. It is only natural to examine the contribution of the operators towards higher performance. As described in Sect. 2, the operators correspond to adaptive sharing and dynamic crowding. Since they have been developed for increasing diversity performance, the test can performed by applying them to problems with badly-scaled objective functions. This is because, due to the badly-scaling property, any algorithm hardly breaks free from poor diversity for the problems.

A new test problem can be designed from [11]. It has two important issues that are not encountered in usual benchmark MOPs; *scaling* (of objectives) and *dependency* (of variables). The problem is defined as follows:

$$\begin{aligned} \text{minimize } & f_{SCALE}(\vec{y}) = (f_1(\vec{y}), f_2(\vec{y})), \\ \text{where } & f_1(\vec{y}) = y_1^{\beta_1} + g(\vec{y}), \\ & f_2(\vec{y}) = y_1^{-\beta_2} + g(\vec{y}), \\ & g(\vec{y}) = \sum_{j=2}^{n} \left| y_j + \sum_{k=1}^{j-1} y_k \right| \end{aligned} \tag{7}$$

where $y_1 \in [0.1, 10]$ and $y_j \in [-100, 100]$ for $2 \le j \le n$, and n is the problem size. The true front is given as

$$f_2 = f_1^{-\rho} \text{ and } g(\vec{y}) = 0$$

where ρ is defined by β_2/β_1. As in [11], ρ adjusts the degree of scaling between objectives, and f_2 is much larger scaling than f_1 as ρ increases from the value of 1.0. To make matters worse, the function $g(\vec{y})$ involves (multivariate) interactions between variables. It means that discovering the optimal point (of $g(\vec{y})$) is very difficult without incorporating the knowledge on the dependencies. Due to the mixed effect of the (badly) scaling and the variable interaction, no algorithm finds it easy to achieve a good spread of solutions along the true front.

The barrier regarding dependency can be overcome by the linkage-guided evolution (as in MrBOA), but the scaling is still problematic. Thus, every MGEA suffers from discovering a set of solutions that uniformly sample the overall true front, even if final solutions reach a part of the front. As the number of variables grows, the dependencies disable the algorithms from breaking down the barrier erected by the scaling property.

3.2 Performance Measures

The aim of multiobjective optimization is to achieve higher proximity of the nondominated solutions while preserving better diversity. In this regard, several metrics have been suggested for properly measuring performances of MGEAs, but none of thme is universal. Thus, suitable metrics must be chosen/employed in order to correctly evaluate accepted performance. With this in view, we consider the following three metrics.

The first is *proximity* metric introduced in [4]. It measures the extent of convergence (of the nondominated set) to the true front. The proximity metric is given by,

$$\Upsilon = \frac{1}{|\mathcal{F}_1|} \sum_{\vec{z}^0 \in \mathcal{F}_1} \min_{\vec{z}^1 \in \mathcal{F}^*} \{d_E(\vec{z}^0, \vec{z}^1)\} \tag{8}$$

where $d_E(\vec{z}^0, \vec{z}^1)$ is the Euclidean distance between objective values and it is given by

$$d_E(\vec{z}^0, \vec{z}^1) = \sqrt{\sum_{k=1}^{M} (f_k(\vec{y}^1) - f_k(\vec{y}^0))^2} \tag{9}$$

where M is the number of objectives. A smaller value denotes a higher proximity of the nondominated solutions.

The second is *uniformity* metric given in [1].[7] It measures diversity in the sense of indicating the extent of uniformly spread of the nondominated solutions. It is defined by

$$\Delta = \frac{d_f + d_l + \sum_{i=1}^{|\mathcal{F}_1|-1} |d_i - \mu_d|}{(|\mathcal{F}_1| - 1)\mu_d} \tag{10}$$

where d_i and μ_d are the Euclidean distance between consecutive solutions in the set $\mathcal{F}_1$ and the average of these distances respectively, and d_f and d_l are the Euclidean distances between the extreme solutions and the boundary solutions of the computed nondominated set. The concrete methodology of computing d_f and d_l can be found in [6]. The metric is essentially identical to that in NSGA-II, but the difference lies in the consistency in indicating good/bad distribution of the nondominated solutions. Thus, the metric always has a smaller value for well distributed nondominated set.

The last metric that measures diversity is *front-spread* metric given in [15]. Unlike the uniformity metric, it measures the extent of covering the true front by the nondomi-

[7] In general, it is known as diversity metric [6]. In this study, the term of 'diversity' is used as a measure that how well and widely the nondominated solutions are distributed over the true front. Thus, the term of 'uniformity' is employed for avoiding some confusion.

nated solutions. The metric is given by

$$\Lambda = \frac{1}{\sqrt{m}} \max_{(z^0, z^1) \in \tilde{\mathcal{F}}^* \times \tilde{\mathcal{F}}^*} \{d_E(\vec{z}^0, \vec{z}^1)\} \qquad (11)$$

where $\tilde{\mathcal{F}}^*$ is defined by

$$\tilde{\mathcal{F}}^* = \{\vec{z}^0 | \exists \vec{z}^0 \in \mathcal{F}^1 \text{ s.t. } \min_{\vec{z}^1 \in \mathcal{F}^*} \{d_E(\vec{z}^0, \vec{z}^1)\} \le \epsilon\}.$$

Here, the set $\tilde{\mathcal{F}}^*$ is a collection of nondomiated solutions that are ϵ-close to the true front. It is called an approximate set of the true front. The metric is normalized to within [0,1]; hence, a closer value to 1.0 means a wider spread of nondominated solutions.

3.3 Empirical Modification

The experiment set-up exploited in the original study [1,2] to get better overall performance is employed. MrBOA uses normal mixture distributions obtained by clustering the selected individuals in order to learn probabilistic models. As a computationally efficient clustering way, k-means algorithm is employed for model selection and randomized leader algorithm (RLA) with a threshold value ξ of 0.3 is used for model fitting. Truncation selection with parameter $\tau = 0.5$ and the BIC with regularization parameter $\lambda = 0.5$ are invoked for learning a probabilistic model. The maximum number of parents, viz., $n-1$, is allowed. The test problem size n and its parameter ρ are set to 10 and 5, respectively. Moreover, the population size used is empirically obtained in order to get acceptable convergence (to the true front) with at most the maximum number of (multiobjective function) evaluations. In the problem under focus, it has been observed that the population size N is 1000 and the number of 10^5 evaluations is adequate for termination. All the results are collected/averaged over 30 runs.

3.3.1 Effect of Regularization Parameter

Table 1 presents all the results found by MrBOA employing the adaptive sharing with varying values of the regularization parameter. To discriminate between the solutions in the first front, the simple crowding scheme (of NSGA-II) is applied. Selection dynamics of MrBOA based on the fitness assigned by $\gamma = 1$ resembles that of NSGA-II [1].[8] Thus, the performance with $\gamma = 1$ can be regarded as a reference for NSGA-II. It is seen that all the results almost approach the true front. Even though a proper setting of γ exists for the best performance, no statistical significance appears. It may be concluded that adaptive sharing does not help much in promoting diversity of nondominated solutions. In the following, the MrBOA employing adaptive sharing with $\gamma = 20$, RLA of $\xi = 0.3$, simple crowding, and truncation selection with $\tau = 0.5$ is taken as a reference.[9] Its Pareto front can be found in Fig. 4(a).

3.3.2 Modified Adaptive Sharing

We now develop a modified (adaptive) sharing scheme that further improves the overall performance. In adaptive sharing, the fitness computed based on domination count only may not be very effective. In [1], the reason for employing domination count only has been given as follows: "A solution dominated by a smaller number of individuals

[8]The selection order of $\gamma = 1$ is not exactly identical to that of NSGA-II, but their performances are not much different.
[9]In this study, it is called 'simple MrBOA (s-MrBOA).'

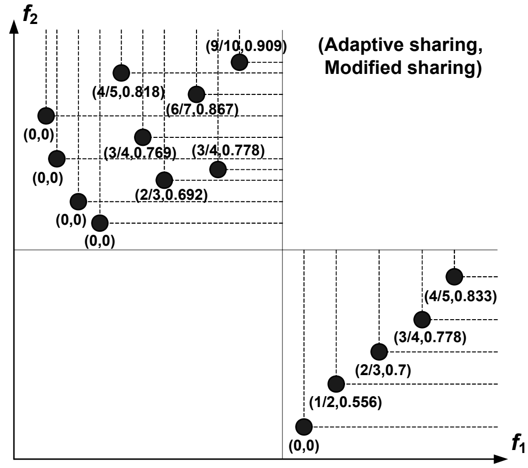


Figure 2: Adaptive sharing vs. modified sharing

is essentially representative; but a solution that dominates a smaller number of individuals is not necessarily unrepresentative." The assertion is obvious if the nondomination count only is of interest. However, it is doubtful when both domination count and nondomination count are taken into account.[10] This is because smaller/larger nondomination count is not necessarily but probabilistically more unrepresentative/representative. Thus, the nondomination count can be used as auxiliary information. A particular solution that has the smallest domination count and the largest nondomination count must be assigned the highest selection priority. As the domination count (of a solution) increases and the nondomination count decreases, the selection priority goes down. This property plays a role in encouraging more representative solutions to survive the course of selection. However, its sensitivity to the nondominated count must be minor. A way to incorporate such dynamics is to incrementally take the reverse nondomination count into account. Thus, a modified sharing intensity is suggested as follows:

$$\mathcal{I}_m(i) = 1 - \frac{1}{1 + N_{dom}(i) + \frac{1}{1+\overline{N}_{dom}(i)}} \qquad (12)$$

where $\overline{N}_{dom}(i)$ is the number of individuals which are dominated by the individuals i, i.e., nondomination count. Also, a lower value is preferable. An example can be found in Fig. 2. As compared with adaptive sharing, the modified sharing can further discriminate which one is better to survive even if some individuals are dominated by the same number of solutions (i.e., equal domination count). In [1], the regularization parameter γ was used for computing fitness values. But, it is undesirable to introduce any system parameter. Instead, the fitness can be directly computed by the modified sharing intensity without diluting its innate feature. That is, the fitness of individuals that do not belong to the first front $\mathcal{F}^1$ is given by $f(i) = \mathcal{R}(i)\{1+\mathcal{I}_m(i)\}$, where $i \notin \mathcal{F}^1$; otherwise, it follows that $f(i) = \mathcal{R}(i)\{1 + \mathcal{D}(i)\}^{-1}$, where $i \in \mathcal{F}^1$. Table 2 shows some positive effects

[10]SPEA-II [15] assigns the fitness of an individual by the sum of strength values of all the individuals it dominates. In other words, it also tires to take into account the information of both dominating and dominated individuals.

Table 1: Comparison of Pareto fronts discovered by MrBOA with various γ.

Metric	$\gamma = 1$	$\gamma = 5$	$\gamma = 10$	$\gamma = 20$	$\gamma = 50$
μ_Υ	1.9723E-4	1.8616E-4	1.7437E-4	1.7450E-4	2.1526E-4
σ_Υ	1.9518E-4	1.0019E-4	6.9910E-5	6.9989E-5	2.0537E-4
μ_Δ	0.9073	0.9276	0.8998	0.8740	0.8939
σ_Δ	0.1486	0.2030	0.1795	0.1516	0.1500
μ_Λ	0.7902	0.7968	0.7973	0.8014	0.7831
σ_Λ	0.1100	0.1105	0.1195	0.1077	0.1394
Statistical t-test; (Υ, Δ, Λ)					
	$\gamma = 1$	$\gamma = 5$	$\gamma = 10$	$\gamma = 20$	$\gamma = 50$
$\gamma = 1$	–	(0.27, -0.43, -0.23)	(0.61, 0.16, -0.22)	(0.60, 0.84, -0.38)	(-0.34, 0.33, 0.21)
$\gamma = 5$	(-0.27, 0.43, 0.23)	–	(0.52, 0.51, -0.02)	(0.52, 1.01, -0.15)	(-0.69, 0.72, 0.42)
$\gamma = 10$	(-0.61, -0.16, 0.22)	(-0.52, -0.51, 0.02)	–	(-0.01, 0.78, -0.14)	(-1.04, 0.13, 0.41)
$\gamma = 20$	(-0.60, -0.84, 0.38)	(-0.52, -1.01, 0.15)	(0.01, -0.78, 0.14)	–	(-1.02, -0.48, 0.55)
$\gamma = 50$	(0.34, -0.33, -0.21)	(0.69, -0.72, -0.42)	(1.04, -0.13, -0.41)	(1.02, 0.48, -0.55)	–

There is no statistical significance by a paired, two-tailed test of $\alpha = 0.05$.

Table 2: Performance of modified sharing.

	μ_Υ (σ_Υ)	μ_Δ (σ_Δ)	μ_Λ (σ_Λ)
Modified sharing	1.6200E-4 (9.1551E-6)	0.8628 (0.0868)	0.8105 (0.0347)
t-test {Adaptive−Modified}; (Υ, Δ, Λ)			
t-value	(0.99, 0.33, -0.41)		

No significance by a paired, two-tailed test of $\alpha = 0.05$.

Table 3: Performance of dynamic crowding.

	μ_Υ (σ_Υ)	μ_Δ (σ_Δ)	μ_Λ (σ_Λ)
Dynamic crowding	3.5899E-4 (9.7356E-4)	0.8439 (0.2423)	0.8121 (0.1288)
t-test {Simple−Dynamic}; (Υ, Δ, Λ)			
t-value	(-1.03, 0.56, -0.36)		

No significance by a paired, two-tailed test of $\alpha = 0.05$.

of the modified sharing on proximity and diversity performances. Although there is no statistical significance, the involvement of nondomination count can further boost up proximity, uniformity and front-spread of nondominated set.

3.3.3 *Effect of Dynamic Crowding*

Table 3 compares overall performances of simple and dynamic crowding schemes. The results present how the dynamic crowding works in diversity preservation. The diversity achieved by dynamic crowding is better than that of simple crowding. No statistical significance is implied. It can be seen that dynamic crowding is somewhat helpful for achieving better diversity. If dynamic crowding is balanced with other operators such as modified sharing, the synergy arising from them might drastically improve the diversity performance.[11] Thus, it is promising to directly accept the dynamic crowding scheme in order for better diversity of nondominated solutions.

4. DIVERSITY PRESERVATION

This section presents diversity-preserving algorithms.

4.1 Diversity-Preserving Selection

From the investigation in Sect. 3, it is necessary to develop a more powerful mechanism for improving diversity performance. In general, most MGEAs employ truncation selection due to its ability to naturally incorporate elitism. That is, the best τ portion of the population is selected as parents for the next generation. It is helpful in advancing the equally preferable solutions towards the true front. However, it can also degenerate the diversity of solutions in some sense. Thus, it is desirable to incorporate elitism in a controlled manner. In the selection phase, the top $\tau\varrho$ portion (of the individuals) is chosen by the truncation selection. After that, the rest $\tau(1-\varrho)$ portion that is set aside for diversity promotion is filled with some solutions that are largely far from the worst $\tau(1-\varrho)N$ individuals in the objective space. (Here, N is the population size.) A set of such individuals is referred to as 'diversity-steering set' (DS). At first, a number of w individuals is randomly chosen (without redundancy) from the remaining $(1-\tau)N$ individuals[12]. One of the w individuals, which is mostly apart from the worst individual of DS, is selected. The same procedure follows for the second worst individual of DS. If the fittest individual has already been selected, the next promising candidate is taken. This continues until going over all the elements of DS. An example is illustrated in Fig. 3. Let us consider the selected half of the population. By truncation selection with $\tau = 0.5$, a set of individuals $\{A, B, C, D, E, F, G\}$ is selected. When employing the proposed selection with $\tau = 0.5$, $\varrho = (4/7)$, and $w = (1-\tau)N$,[13] the individuals

[11] The composite effect would be significant for high-dimensional and badly-scaling problems.

[12] They do not contain the individuals selected by truncation selection and those in DS

[13] Such w value that has been considered for simple illustration is undesirable in practice.

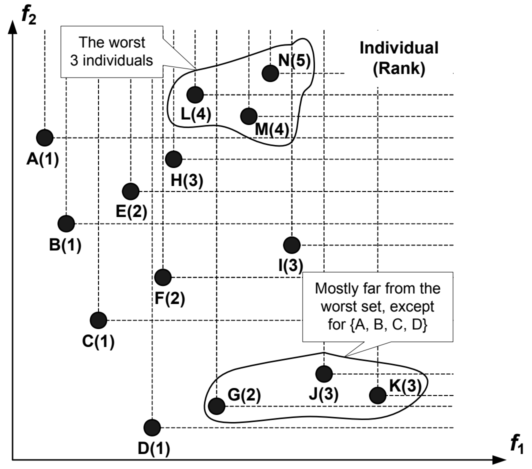


Figure 3: Example of diversity-preserving selection.

Table 4: Results of diversity-preserving selection.

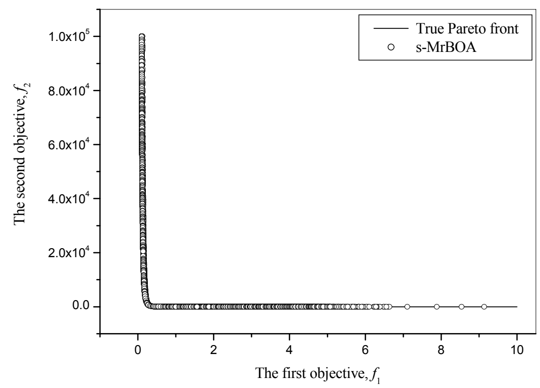
	μ_Υ (σ_Υ)	μ_Δ (σ_Δ)	μ_Λ (σ_Λ)
dp-selection	1.7575E-4 (1.7269E-5)	0.6177 (0.0855)	0.9789 (0.0273)
t-test {Truncation−dp-Selection}; (Υ, Δ, Λ)			
t-value	(-0.10, 8.12†, -8.63†)		

† t-value is *significant* at α=0.05 by a paired, two-tailed test.

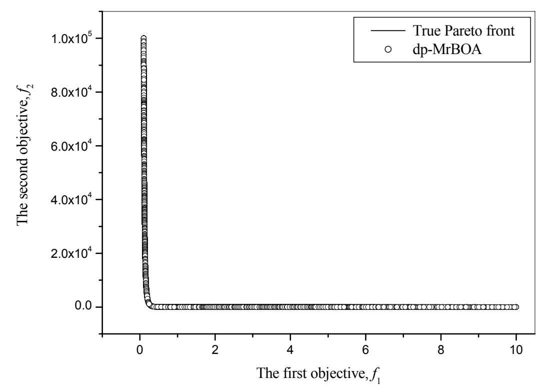
$\{A, B, C, D, G, J, K\}$ would be chosen. In other words, the solutions $\{J, K\}$ are preferable to the solutions $\{E, F\}$ even though the former has worse (Pareto) ranking values. It is clear that $\{J, K\}$ is less crowded than $\{E, F\}$, thereby making for better diversity of nondominated set. The strength of the diversity-preserving selection on the diversity performance is empirically validated in Table 4. The results have been obtained with $\tau = 0.5$, $\varrho = 0.8$, $w = 10$ and the same set-up used in Sect. 3.3, as applied to the scaling problem. The proposed selection encourages (statistically significant) uniformity and front-spread of nondominated solutions (compared with truncation selection) while maintaining comparable proximity performance.

4.2 Diversity Preserving MrBOA: dp-MrBOA

A diversity preserving MrBOA (dp-MrBOA) will be proposed in what follows. The idea is to incorporate modified sharing, dynamic crowding and diversity-preserving selection into the algorithm. This combination synergically accumulates their effects, thereby significantly enhancing both uniformity and front-spread (i.e., diversity performance) without compromising on the proximity of nondominated set. Figures 4(a) and (b) compare the Pareto fronts found by s-MrBOA and dp-MrBOA as applied to the badly-scaled problem, respectively. Recall that s-MrBOA consists of adaptive sharing with $\gamma = 20$, simple crowding, truncation selection with $\tau = 0.5$, and RLA with $\xi = 0.3$. It is seen that dp-MrBOA returns much better diversity of the nondominated solutions than does s-MrBOA. Table 5 shows the (statistically) significant improvement of both uniformity and front-spread without degrading proximity performance.



(a) Simple MrBOA.



(b) Diversity-preserving MrBOA.

Figure 4: Pareto fronts by s-MrBOA & dp-MrBOA.

To investigate the effect of the diversity-preserving mechanism on non-scaling problems (i.e., objective functions are not scaled), the dp-MrBOA is also tested on four well-known MOPs; MDP-II, MNSP, ZDT_4, ZDT_6. Their inherent characteristics can be found in [1]. Being quite different from the scaling problem[14], any algorithm finds it very hard to approach close to their true fronts; i.e., front-spread performance is often unmeasurable. Only the proximity and uniformity performances have been investigated here. The experimental set-up is the same as that employed in the original study [1]. The results are tabulated in Table 6. Depending on the problems' features, the dp-MrBOA seems to be superior or inferior to the s-MrBOA, but their performances have no statistical difference. Evidently, the dp-MrBOA has no undesirable effect on the non-scaling problems. In conclusion, it can be said that the dp-MrBOA achieves better diversity of nondominated solutions than does the s-MrBOA (according to the scaling degree of objectives), while maintaining comparable convergence to the true front.

5. CONCLUSION

This paper has tried to take a close look at the main operators of MrBOA. A potential source of diversity preservation has been investigated. Adaptive sharing has been modified by involving both the domination and the nondom-

[14]Convergence to the true front is easy, but spreading individuals uniformly and widely over the front is quite difficult.

Table 6: Performance comparison of s-MrBOA and dp-MrBOA on non-scaling problems.

Algorithm	Metric	MDP-II (800, 11)	MNSP (400, 10)	ZDT_4 (800, 10)	ZDT_6 (400, 10)
s-MrBOA	μ_Υ	0.0956	0.1258	0.0534	0.0983
	σ_Υ	0.1240	0.1700	0.0926	0.1374
	μ_Δ	0.5124	0.4248	0.5009	1.2125
	σ_Δ	0.1742	0.0671	0.1422	0.6230
dp-MrBOA	μ_Υ	0.0774	0.1428	0.0423	0.1475
	σ_Υ	0.1130	0.2492	0.0598	0.2074
	μ_Δ	0.5177	0.4622	0.4594	1.0507
	σ_Δ	0.2259	0.0778	0.0767	0.6247

t-test {s-MrBOA−dp-MrBOA}; (Υ, Δ)				
	MDP-II	MNSP	ZDT_4	ZDT_6
t-value	(0.63, -0.11)	(-0.30, -1.82)	(0.52, 1.44)	(-1.01, 0.97)

No statistical significance appears from a paired, two-tailed test of $\alpha = 0.05$. The values of parenthesis stand for population size N and problem size n, respectively.

Table 5: Comparison of s-MrBOA and dp-MrBOA.

	μ_Υ (σ_Υ)	μ_Δ (σ_Δ)	μ_Λ (σ_Λ)
dp-MrBOA	1.7784E-4 (1.7886E-5)	0.5964 (0.0675)	0.9789 (0.0217)
t-test {s-MrBOA−dp-MrBOA}; (Υ, Δ, Λ)			
t-value	(-0.26, 8.93†, -8.84†)		

† t-value is *significant* at α=0.05 by a paired, two-tailed test.

ination counts. Experimental results have shown that this strategy can be effective in improving the diversity of the nondominated set. Further, a diversity-preserving selection has been developed by introducing some solutions largely apart from the worst group of individuals. It has shown that the proposed selection rather than the truncation selection is more beneficial to diversity promotion. These results lead to the dp-MrBOA which combines modified sharing, dynamic crowding, and diversity-preserving selection. Experimental results have shown that the concomitant synergy significantly improves uniformity and front-spread of nondominated solutions (i.e., diversity performance) while achieving appreciable proximity.

6. ACKNOWLEDGMENTS

This work was supported by the second stage of the Brain Korea 21 Project in 2007.

7. REFERENCES

[1] C. W. Ahn. *Advances in Evolutionary Algorithms: Theory, Design and Practice.* Springer-Verlag, New York, 2006.

[2] C. W. Ahn, D. E. Goldberg, and R. S. Ramakrishna. Real-coded bayesian optimization algorithm: Bringing the strength of BOA into the continuous world. *Lecture Notes in Computer Science*, 3102:840–851, 2004.

[3] P. A. N. Bosman and D. Thierens. Multiobjective optimization with diversity preserving mixture-based iterated density estimation evolutionary algorithm. *Int. J. Approx. Reasoning*, 31(3):259–289, 2002.

[4] P. A. N. Bosman and D. Thierens. The balance between proximity and diversity in multiobjective evolutionary algorithm. *IEEE Trans. Evol. Comput.*, 7(2):174–188, 2003.

[5] J. H. Chen. *Theory and Applications of Efficient Multiobjective Evolutionary Algorithms.* Ph.D. Dissertation, Feng Chia Univ., Taiwan, R.O.C., 2004.

[6] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.*, 6(2):182–197, 1995.

[7] C. M. Fonseca and P. J. Fleming. Multiobjective optimization and multiple constraint handling with evolutionary algorithms – partI: a unified formulation. *IEEE Tran. Syst., Man, Cybern. B*, 28(1):26–37, 1998.

[8] N. Khan. *Bayesian optimization algorithms for multiobjective and hierarchically difficult problems.* M.S. Thesis, Univ. Illinois at Urbana-Champaign, 2003.

[9] H. Lu and G. Yen. Rank-density-based multiobjective genetic algorithm and benchmark test function study. *IEEE Trans. Evol. Comput.*, 7(4):325–343, 2003.

[10] J. Ocenasek and J. Schwarz. Estimation of distribution algorithm for mixed continuous-discrete optimization problems. In *Proc. 2nd Int. Symp. Computational Intelligence*, pages 227–232, 2002.

[11] G. Pedersen and D. E. Goldberg. Dynamic uniform scaling for multiobjective genetic algorithms. *Lecture Notes in Computer Science*, 3103:11–23, 2004.

[12] M. Pelikan, K. Sastry, and D. E. Goldberg. Multiobjective hBOA, clustering, and scalability. In *Proc. GECCO'05*, pages 663–670. ACM Press, 2005.

[13] N. Srinivas and K. Deb. Multiobjective optimization using nondominated sorting in genetic algorithms. *Evol. Comput.*, 2(3):221–248, 1995.

[14] Z. A. Zhang, Q. and Y. Jin. RM-MEDA: A regualrity model based multiobjective estimation of distribution algorithm. *IEEE Trans. Evol. Comput.*, 2007 (to appear).

[15] E. Zitzler, M. Laumanns, and L. Thiele. SPEA2: Improving the strength pareto evolutionary algorithm. In *Proc. Evol. Methods for Design, Optimization, and Control*, pages 95–100, 2002.

[16] E. Zitzler and L. Thiele. Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach. *IEEE Trans. Evol. Comput.*, 3(4):257–271, 1999.