

# The Multi-Objective Real-Valued Gene-Pool Optimal Mixing Evolutionary Algorithm

Anton Bouter  
Centrum Wiskunde & Informatica  
Amsterdam, The Netherlands  
Anton.Bouter@cwi.nl

Ngoc Hoang Luong  
Centrum Wiskunde & Informatica  
Amsterdam, The Netherlands  
Hoang.Luong@cwi.nl

Cees Witteveen  
Delft University of Technology  
Delft, The Netherlands  
C.Witteveen@tudelft.nl

Tanja Alderliesten  
Academic Medical Center  
Amsterdam, The Netherlands  
T.Alderliesten@amc.uva.nl

Peter A.N. Bosman  
Centrum Wiskunde & Informatica  
Amsterdam, The Netherlands  
Peter.Bosman@cwi.nl

## ABSTRACT

The recently introduced Multi-Objective Gene-pool Optimal Mixing Evolutionary Algorithm (MO-GOMEA) exhibits excellent scalability in solving a wide range of challenging discrete multi-objective optimization problems. In this paper, we address scalability issues in solving multi-objective optimization problems with continuous variables by introducing the Multi-Objective Real-Valued GOMEA (MO-RV-GOMEA), which combines MO-GOMEA with aspects of the multi-objective estimation-of-distribution algorithm known as MAMaLGaM. MO-RV-GOMEA exploits linkage structure in optimization problems by performing distribution estimation, adaptation, and sampling as well as solution mixing based on an explicitly-defined linkage model. Such a linkage model can be defined a priori when some problem-specific knowledge is available, or it can be learned from the population. The scalability of MO-RV-GOMEA using different linkage models is compared to the state-of-the-art multi-objective evolutionary algorithms NSGA-II and MAMaLGaM on a wide range of benchmark problems. MO-RV-GOMEA is found to retain the excellent scalability of MO-GOMEA through the successful exploitation of linkage structure, scaling substantially better than NSGA-II and MAMaLGaM. This scalability is even further improved when partial evaluations are possible, achieving strongly sub-linear scalability in terms of the number of evaluations.

## CCS CONCEPTS

•Mathematics of computing → Evolutionary algorithms;

## KEYWORDS

GOMEA, linkage learning, multi-objective optimization, real-valued optimization, scalability

## ACM Reference format:

Anton Bouter, Ngoc Hoang Luong, Cees Witteveen, Tanja Alderliesten, and Peter A.N. Bosman. 2017. The Multi-Objective Real-Valued Gene-Pool

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

GECCO '17, Berlin, Germany

© 2017 ACM. 978-1-4503-4920-8/17/07...\$15.00

DOI: <http://dx.doi.org/10.1145/3071178.3071274>

Optimal Mixing Evolutionary Algorithm. In *Proceedings of GECCO '17, Berlin, Germany, July 15-19, 2017*, 8 pages.

DOI: <http://dx.doi.org/10.1145/3071178.3071274>

## 1 INTRODUCTION

Oftentimes in evolutionary computation, only a Black-Box Optimization (BBO) setting is considered. This means that virtually no knowledge of the optimization problem is available to the algorithm. For multi-objective optimization the well-known NSGA-II [5] was introduced with univariate variation operators that vary each variable independently. As a consequence, NSGA-II is unable to efficiently solve problems with strongly dependent variables [10]. Given the nature of multi-objective optimization with multiple, conflicting objectives that could all be a source of dependencies between variables, having the ability to efficiently and effectively exploit such dependencies may often be important. Previously, the discrete Multi-Objective Gene-pool Optimal Mixing Evolutionary Algorithm (MO-GOMEA) [10] was found to be successful at exploiting linkage structure by using an explicit linkage model, consisting of linkage sets that each describe a set of dependent variables and by separately applying variation jointly to variables in each of the linkage sets. Knowledge of such dependencies, also called linkage in EAs, must be discovered during optimization in a BBO setting, but may also be (roughly) available a priori, in case more is known and understood about the problem at hand. We refer to such situations as a Gray-Box Optimization (GBO) setting. Specifically, we define the GBO setting as one where efficient partial evaluations are possible, enabling the efficient re-evaluation of the objective values of a solution after only a subset of its variables is modified. This leads to a further substantial improvement of the performance of MO-GOMEA, because its variation operation modifies many subsets of variables of solutions, after which the objective values of these solutions can thus be updated efficiently. Note that the possibility of efficient partial evaluations does however not mean that a problem is easy or even decomposable.

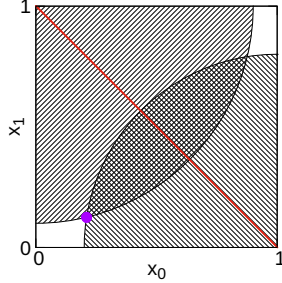
It is such properties that are key strengths of algorithms in the GOMEA family, including MO-GOMEA. Previously however, all members of the GOMEA family addressed problems with discrete variables only. Here, we set out to expand this research line to include real-valued variables. The specific goal being able to efficiently deal with typical aspects of real-valued problem landscapes while

maintaining the strengths of MO-GOMEA. To do so, we introduce the Multi-Objective Real-Valued GOMEA (MO-RV-GOMEA), which combines the strengths of MO-GOMEA and the Multi-Objective Adapted Maximum-Likelihood Gaussian Model Iterated Density-Estimation Evolutionary Algorithm (MAMaLGA) [1].

## 2 DEPENDENCIES INTRODUCED BY MULTI-OBJECTIVE OPTIMIZATION

Even in the simplest, completely dimension-wise decomposable multi-objective problems, dependencies between problem variables can arise. Due to the conflicting nature of objectives, the directions of improvement between objectives are very likely conflicting. Because of this, performing one-dimensional movements of solutions will rarely lead to solutions that dominate previously existing ones. Joint multi-dimensional movements of solutions through the parameter space are therefore required to find dominating solutions.

Such movements are therefore beneficial to efficiency of obtaining a Pareto front with a good coverage of the optimal Pareto front, but they require a more complex model than a univariately factorized one.



**Figure 1: Arbitrary solution (in purple) to the 2-dimensional genMED problem, with areas of improvement shaded in black, and the optimal Pareto set marked by a red line.**

We illustrate the above in Figure 1, which displays an instance of the convex genMED problem with two problem variables, i.e.,  $\ell = 2$ , defined as follows:

$$f_1^{genMED}(\mathbf{x}) = \frac{1}{2} \left( (x_0 - 1)^2 + \sum_{i=1}^{\ell-1} x_i^2 \right)$$

$$f_2^{genMED}(\mathbf{x}) = \frac{1}{2} \left( x_0^2 + (x_1 - 1)^2 + \sum_{i=2}^{\ell-1} x_i^2 \right)$$

The first objective of this problem is a solution's distance to the point (1, 0), and the second objective is the distance to the point (0, 1). An arbitrary solution is denoted by a purple circle, and its areas of improvement in either objective are marked in a black pattern, with the area of domination being the overlapping area of the two shaded areas. The optimal Pareto front is denoted by a red line. Note that the individual objectives are completely dimension-wise decomposable and improvements may be found by changing either variable independently. We see that domination of the point marked in purple requires the simultaneous adaptation of both  $x_0$  and  $x_1$ , because the directions of improvement of both objectives are naturally conflicting. Note that the optimal Pareto front can still be reached by univariate variation (i.e., move left or up), however, especially considering that in an EA there are multiple such

```

MO-RV-GOMEA //population  $\mathcal{P}$  size  $n$ ; number of clusters  $q$ 
1 NISPop  $\leftarrow t \leftarrow 0$ 
2 for  $i \in \{0, 1, \dots, n-1\}$  do
3    $\mathbf{x} \leftarrow \text{INITIALIZEANDEVALUATESOLUTION}()$ 
4    $\mathcal{P} \leftarrow \mathcal{P} \cup \{\mathbf{x}\}$ 
5    $\mathcal{A} \leftarrow \text{UPDATEELITISTARCHIVE}(\mathcal{A}, \mathbf{x})$ 
6   while  $\neg \text{TERMINATIONCRITERIASATISFIED}$  do
7      $t \leftarrow t + 1$ 
8     for  $\mathbf{x} \in \mathcal{P}$  do
9       improved $[\mathbf{x}] \leftarrow \text{False}$ 
10       $\mathcal{S} \leftarrow \text{NONDOMINATIONTRUNCATIONSELECTION}(\mathcal{P}, \lfloor \tau n \rfloor)$ 
11       $\{C_0, C_1, \dots, C_{q-1}\} \leftarrow \text{PERFORMCLUSTERING}(\mathcal{S}, n_c)$ 
12       $\{C_0^{\mathcal{P}}, C_1^{\mathcal{P}}, \dots, C_{q-1}^{\mathcal{P}}\} \leftarrow \text{ASSIGNCLUSTERINDEXFORPOPULATION}(\mathcal{P}, C)$ 
13      for  $C_k \in C$  do
14         $\mathcal{F}_{C_k} = \{F^0, F^1, \dots, F^{m-1}\} \leftarrow \text{LEARNLINKAGE}(C_k)$ 
15         $N_{C_k} = \{N_{F^0}, N_{F^1}, \dots, N_{F^{m-1}}\} \leftarrow \text{ESTIMATEPARAMETERS}(\mathcal{F}_{C_k}, C_k)$ 
16        COPYELITISTSOLUTIONSTOPOPULATION( $\mathcal{A}, C_k^{\mathcal{P}}, \lfloor \tau |C_k^{\mathcal{P}}| \rfloor$ )
17      for  $F^j \in \mathcal{F}_{C_k}$  do
18        for  $\mathbf{x} \in C_k^{\mathcal{P}}$  do
19           $\mathbf{x} \leftarrow \text{PERFORMPARTIALVARIATION}(\mathbf{x}, F^j, N_{F^j})$ 
20           $c_{F^j, C_k}^{\text{multiplier}} \leftarrow \text{ADAPTDISTRIBUTIONMULTIPLIER}(c_{F^j, C_k}^{\text{multiplier}})$ 
21          UPDATEELITISTARCHIVE( $\mathcal{A}, \mathbf{x}$ )
22        for  $\mathbf{x} \in \{\mathbf{x}_i \mid \mathbf{x}_i \in C_k^{\mathcal{P}}, i = 0, 1, \dots, \lfloor \frac{1}{2} \tau |C_k^{\mathcal{P}}| \rfloor\}$  do
23           $\mathbf{x} \leftarrow \text{APPLYANTICIPATEDMEANSHIFT}(\mathbf{x})$ 
24      for  $\mathbf{x} \in \mathcal{P}$  do
25        if improved $[\mathbf{x}]$  then
26          NIS $[\mathbf{x}] \leftarrow 0$ 
27        else
28          NIS $[\mathbf{x}] \leftarrow \text{NIS}[\mathbf{x}] + 1$ 
29        if NIS $[\mathbf{x}] > \text{NIS}^{\text{max}}$  then
30           $\mathbf{x} \leftarrow \text{PERFORMFORCEDIMPROVEMENT}(\mathbf{x})$ 
    
```

**Figure 2: Pseudo-code for MO-RV-GOMEA.**

solutions progressing toward the Pareto front, being able to jointly vary  $x_0$  and  $x_1$  and thus exploit this dependency, could therefore be beneficial to the optimization of this problem.

## 3 REAL-VALUED MO-GOMEA DESIGN

MO-RV-GOMEA maintains a population  $\mathcal{P}$  of  $n$  potential solutions, and an elitist archive  $\mathcal{A}$  according to the implementation in [9] to keep track of non-dominated solutions, because elitism is known to be beneficial to convergence [7]. The initial population consists of  $n$  randomly generated solutions (in the BBO case). The following steps are iterated until any termination criterion is satisfied (e.g., computing budget is spent or a Pareto front of sufficient quality is found). First, truncation selection is performed to select a set  $\mathcal{S}$  of  $\lfloor \tau n \rfloor$ ,  $\tau \in [1/n, 1]$ , solutions from the current population  $\mathcal{P}$  that are ranked the best after non-dominated sorting [5]. Second, the selection set  $\mathcal{S}$  is then clustered into  $q$  equal-sized clusters  $C_k$ 's. Third, a linkage model is learned for each cluster. Normal distribution parameters are then estimated for each linkage set of each cluster, which are used by the main variation operator to sample new partial solutions. The outline of MO-RV-GOMEA is displayed in pseudo-code in Figure 2 and the details of each step are presented in the following.

### 3.1 Clustering

At every generation, clustering is performed, because this is necessary to find a good spread of solutions across the entire optimal Pareto front [11]. The selection set  $\mathcal{S}$  is partitioned into  $q$  equal-sized clusters  $C_k$ 's by an efficient leader-based clustering procedure

as follows. All distances considered are in Euclidean objective space, where each objective value is scaled by dividing by the range between the minimum and maximum value of the objective in the population. First, for each of  $m$  objectives, a so-called single-objective cluster is created that contains the  $n_c = \frac{2}{q}|\mathcal{S}|$  solutions having the best objective value in this objective. Single-objective clusters are aimed at directing the far ends of the Pareto front outward. Second, to construct the  $q - m$  remaining clusters, leader solutions that are far apart from each other in the objective space are heuristically selected from  $\mathcal{S}$  as in [1]. Next, these leaders become the initial means of the  $q - m$  clusters, from which they are expanded to include other nearby solutions in  $\mathcal{S}$  until each cluster contains exactly  $n_c$  solutions, resulting in a set of equal-sized clusters where neighboring clusters can overlap. Such overlapping reduces the chance that some solutions are not included in any cluster, thereby decreasing the probability of finding a Pareto-front consisting of disconnected chunks, because the search effort is focused into separate clusters.

Cluster registration [1] is then applied to re-assign the cluster indices such that the cluster indexed  $k$  is the cluster in generation  $t$  that is closest to the cluster indexed  $k$  in generation  $t - 1$ . Such explicit registration is required to properly compute the direction of improvement for each cluster, in the form of the Anticipated Mean Shift (AMS) that accelerates search toward the regions where Pareto-optimal solutions are potentially located (see Section 3.5).

Because MO-RV-GOMEA does not create an entire offspring solution each time, but adapts each existing solution in the population through a series of partial variations, each population solution ultimately needs to be assigned to exactly one appropriate cluster so that the linkage model of that cluster can be used for performing variation. Thus,  $\mathcal{P}$  must be partitioned into  $q$  non-overlapping clusters  $\mathcal{C}_k^{\mathcal{P}}$ 's. First, to assure that each population cluster  $\mathcal{C}_k^{\mathcal{P}}$  has a minimum size of  $n_c$  solutions,  $n_c$  rounds are performed in which for each cluster, considered in a random order, the solution that is closest to this cluster's mean and is yet unassigned, is assigned to this cluster. For the remaining population solutions that do not belong to any  $\mathcal{C}_k^{\mathcal{P}}$ , each solution is assigned to the cluster  $\mathcal{C}_k^{\mathcal{P}}$  that has the shortest distance to the cluster mean.

Having several of the current best non-dominated solutions in the population can benefit performance [1]. To this end, each elitist archive solution is associated with its nearest cluster  $\mathcal{C}_k^{\mathcal{P}}$  based to the cluster means. For each cluster  $\mathcal{C}_k^{\mathcal{P}}$ , at most  $n_k = \lfloor \tau |\mathcal{C}_k^{\mathcal{P}}| \rfloor$  of the most-dominated solutions are replaced by elitist solutions associated with  $\mathcal{C}_k^{\mathcal{P}}$ . If the number of associated elitist solutions is more than  $n_k$ , then  $n_k$  elitist solutions that are far apart from each other can be chosen by the same leader-selecting heuristic as used above, from [1].

### 3.2 Distribution and Linkage Learning

Similar to the discrete (MO-)GOMEA [3, 10], the MO-RV-GOMEA employs the general Family Of Subset (FOS) linkage model to describe the linkage structure of a problem. Let  $L = \{0, 1, \dots, \ell - 1\}$  be the set of all  $\ell$  problem variable indices. A FOS  $\mathcal{F}$  is written as  $\{\mathbf{F}^0, \mathbf{F}^1, \dots, \mathbf{F}^{|\mathcal{F}|-1}\}$  where  $\mathbf{F}^j \subseteq \{0, 1, \dots, \ell - 1\}$ ,  $j \in \{0, 1, \dots, |\mathcal{F}|-1\}$ . A FOS  $\mathcal{F}$  is thus a set of subsets of  $L$ . Each subset  $\mathbf{F}^j$  represents a linkage set of variables that are considered to exhibit some degree of dependency. Variables in such a linkage set should thus be

jointly considered when performing variation to generate offspring solutions.

If all problem variables are independent from each other, the univariate FOS can be used, in which all linkage sets contain a single variable index, i.e.,  $\mathcal{F} = \{\mathbf{F}^j = \{j\} \mid j = 0, 1, \dots, \ell - 1\}$ . The most commonly used form of FOS in the discrete (MO-)GOMEA is the linkage tree (LT) [3, 10]. An LT FOS  $\mathcal{F}$  has  $2\ell - 1$  linkage sets, in which  $\ell$  linkage sets are singleton sets like in the univariate FOS. The remaining linkage sets are constructed in such a way that any multivariate linkage set  $\mathbf{F}^i$  is the result of merging two other linkage sets  $\mathbf{F}^j$  and  $\mathbf{F}^k$ ,  $\mathbf{F}^j \cap \mathbf{F}^k = \emptyset$ ,  $|\mathbf{F}^j| < |\mathbf{F}^i|$ ,  $|\mathbf{F}^k| < |\mathbf{F}^i|$ , and  $\mathbf{F}^j \cup \mathbf{F}^k = \mathbf{F}^i$ . The two constituent linkage sets  $\mathbf{F}^j$  and  $\mathbf{F}^k$  will then not be considered for further merging. This merging procedure stops when a linkage set containing all variables, i.e., the set  $L$  itself, is created. The LT FOS thus allows hierarchical dependencies to be modeled, ranging from univariate all-independence to full all-dependence. A similarity/distance metric is required here to give priority to pairs of linkage sets that are strongly related to each other. Problem-specific information or expert knowledge, if available as in GBO, can be exploited to develop such a metric and the LT FOS can be learned once, offline, before (MO-RV-)GOMEA is run. For BBO, a viable option to measure similarity between linkage sets of (continuous) variables is based on Mutual Information (MI). For real-valued variables and using the normal distribution, MI can be derived from the sample Pearson correlation coefficient  $r$  [8], which can be estimated from the population's candidate solutions in each generation. Learning LTs can be efficiently done by the Unweighted Pair Grouping Method with Arithmetic-mean (UPGMA) for which an optimal implementation exists that has  $O(\ell^2)$  time complexity, given a pairwise distance matrix [6]. To compute this matrix based on MI from the population costs  $O(n\ell^2)$ , however.

For multi-objective optimization, solutions that are far from each other on the Pareto-optimal front may bear little resemblance as they excel in different objectives. Similarly, the landscape in these regions may have very different dependency structures. Therefore, if problem-specific knowledge is not available, a separate FOS is learned from the solutions in each cluster.

After learning FOS models, values of solution variables in each linkage set  $\mathbf{F}^j$  of each cluster's FOS  $\mathcal{F}$  are then modeled by a (multivariate) Gaussian distribution  $N_{\mathbf{F}^j}$ , the parameters of which can be estimated with Maximum Likelihood (ML) [13] over the cluster's solutions, as displayed in Figure 3. More specifically, let the cluster  $\mathcal{C}_k$  be the sample data, then the mean vector  $\mu_{\mathbf{F}^j, \mathcal{C}_k}$  and the covariance matrix  $\Sigma_{\mathbf{F}^j, \mathcal{C}_k}$ , i.e., the parameters of  $N_{\mathbf{F}^j}$ , are estimated by the sample average and sample covariance matrix, respectively.

### 3.3 Performing Partial Variation

Similar to continuous Estimation of Distribution Algorithms (EDAs) such as AmALGaM and CMA-ES, MO-RV-GOMEA samples the estimated distributions to create offspring solutions. However, instead of generating an entire solution each time, we here integrate the mechanism of partial variation from the Gene-pool Optimal Mixing (GOM) operator of the discrete (MO-)GOMEA [3, 10]. Specifically, in each cluster  $\mathcal{C}_k^{\mathcal{P}}$ , linkage sets  $\mathbf{F}_{\mathcal{C}_k}^j$  are iteratively considered in a random order. For each linkage set, the corresponding multivariate Gaussian distribution  $N_{\mathbf{F}^j}$  is sampled  $|\mathcal{C}_k^{\mathcal{P}}|$  times to alter all existing

```

ESTIMATEPARAMETERS( $\mathcal{F}_{C_k}, C_k$ )
1  $N \leftarrow \{\}$ 
2 for  $F^j \in \mathcal{F}_{C_k}$  do
3    $\hat{\mu} \leftarrow \frac{1}{|C_k|} \sum_{i=0}^{|C_k|-1} (C_{k(i)})_{F^j}$ 
4    $\hat{\Sigma} \leftarrow \frac{1}{|C_k|} \sum_{i=0}^{|C_k|-1} ((C_{k(i)})_{F^j} - \hat{\mu})((C_{k(i)})_{F^j} - \hat{\mu})^T$ 
5    $\hat{\mu}_{(F^j, C_k)}^{shift}(t) \leftarrow \hat{\mu}$ 
6    $\hat{\mu}_{(F^j, C_k)}^{shift}(t) \leftarrow \hat{\mu}_{(F^j, C_k)}(t) - \hat{\mu}_{(F^j, C_k)}(t-1)$ 
7    $\hat{\Sigma}_{(F^j, C_k)}^{multiplier}(t) \leftarrow c_{(F^j, C_k)}^{multiplier} \hat{\Sigma}$ 
8    $L_{(F^j, C_k)}^* \leftarrow \text{CHOLESKYDECOMPOSITION}(\hat{\Sigma}_{(F^j, C_k)}^{multiplier}(t))$ 
9    $N \leftarrow N \cup \{N_{F^j}(\hat{\mu}_{(F^j, C_k)}^{shift}(t), \hat{\Sigma}_{(F^j, C_k)}^{multiplier}(t))\}$ 
10 return( $N$ )
    
```

Figure 3: Estimating parameters of normal distributions.

```

PERFORMPARTIALVARIATION( $x, F^j, C_k$ )
1  $b \leftarrow x$ 
2  $y_{F^j} \leftarrow \text{SAMPLEDISTRIBUTION}(N_{F^j}(\hat{\mu}_{(F^j, C_k)}^{shift}, \hat{\Sigma}_{(F^j, C_k)}^{multiplier}))$ 
3 if  $x \in \{x_i \mid x_i \in C_k, i = 0, 1, \dots, \lfloor \frac{1}{2} \tau |C_k| \rfloor\}$  do
4    $y_{F^j} \leftarrow y_{F^j} + c_{(F^j, C_k)}^{multiplier} \delta_{AMS}^{shift} \hat{\mu}_{(F^j, C_k)}^{shift}(t)$ 
5    $x_{F^j} \leftarrow y_{F^j}$ 
6 EVALUATESOLUTION( $x$ )
7 if  $x \not\prec b$  and ISDOMINATEDBYARCHIVE( $\mathcal{A}, x$ )
8    $x_{F^j} \leftarrow b_{F^j}$ 
9 else
10   improved[ $x$ ]  $\leftarrow$  True
11 return( $x$ )
    
```

Figure 4: Performing linkage set-based variation.

solutions in the cluster at the indices indicated by  $F^j$ . Let  $v = F^j$ , a  $|v|$ -dimensional vector can be randomly drawn from a multivariate Gaussian distribution  $N(\mu_v, \Sigma_v)$  as  $\mu_v + L_v(z_0, z_1, \dots, z_{|v|-1})$ , where  $z_i$  is a single-dimensional random variable following standard normal distribution  $N(0, 1)$  and  $L_v$  is the lower triangular matrix resulting from the Cholesky decomposition of the covariance matrix  $\Sigma_v$ , i.e.,  $L_v L_v^* = \Sigma_v$ . For each partially-altered solution  $x$ , the changes are accepted if such changes result in a new solution that dominates the old one or if the new solution is a non-dominated solution (compared to the elitist archive). Otherwise,  $x$  reverts to its previous state. If  $x$  is part of a single-objective cluster, the changes are accepted if they lead to an improved objective value for the objective that this cluster is focused on, and reverted otherwise. Before evaluating objective values, a random subset of size  $\lfloor \frac{1}{2} \tau |C_k^P| \rfloor$ , of these partially-altered solutions are further modified by applying AMS to bias the search effort toward the direction where Pareto improvements happen (see Section 3.5). Figure 4 shows pseudo-code for performing linkage set-based partial variation.

While the discrete (MO-)GOMEA evolves each existing solution  $x$  through a series of partial variations to iteratively construct a single offspring solution at a time, each partial variation (associated with a linkage set  $F^j$ ) in MO-RV-GOMEA is performed on all solutions in the cluster before moving to the next linkage set. Interchanging these loops is required to quantify the improvements that are brought about in the cluster by performing variation for linkage set  $F^j$ . These improvements are used to adjust the corresponding multiplier that adaptively controls the range of exploration (see Sections 3.2 and 3.4) and the AMS movement in the problem dimensions indicated by  $F^j$  (see Section 3.5).

The capability of performing partial variation matches with the possibility of doing partial evaluations. Overall efficiency can be

```

ADAPTDISTRIBUTIONMULTIPLIER( $c_{(F^j, C_k)}^{multiplier}$ )
1 if ELITISTARCHIVECANBEIMPROVED( $\mathcal{A}, C_k^P$ ) then
2    $NIS^{pop} \leftarrow 0$ 
3   if  $c_{(F^j, C_k)}^{multiplier} < 1.0$  then  $c_{(F^j, C_k)}^{multiplier} \leftarrow 1.0$ 
4    $SDR \leftarrow \text{COMPUTE STANDARD DEVIATION RATIO}(F^j, C_k^P)$ 
5   if  $SDR > \theta^{SDR}$  then  $c_{(F^j, C_k)}^{multiplier} \leftarrow \eta^{inc} c_{(F^j, C_k)}^{multiplier}$ 
6 else
7   if  $c_{(F^j, C_k)}^{multiplier} > 1.0$  or  $NIS^{pop} \geq NIS^{max}$  then
8      $c_{(F^j, C_k)}^{multiplier} \leftarrow \eta^{dec} c_{(F^j, C_k)}^{multiplier}$ 
9   if  $c_{(F^j, C_k)}^{multiplier} < 1.0$  and  $NIS^{pop} < NIS^{max}$  then  $c_{(F^j, C_k)}^{multiplier} \leftarrow 1.0$ 
10 return( $c_{(F^j, C_k)}^{multiplier}$ )

COMPUTE STANDARD DEVIATION RATIO( $F^j, C_k^P$ )
1  $n^{imp} \leftarrow 0$ ;  $SDR \leftarrow 0.0$ 
2 for  $x \in C_k^P$  do
3   if ELITISTARCHIVECANBEIMPROVED( $\mathcal{A}, x$ ) then
4      $n^{imp} \leftarrow n^{imp} + 1$ 
5     if  $|F^j| = 1$  then
6        $SDR \leftarrow SDR + (x_{F^j} - \hat{\mu}_{(F^j, C_k)}) / \hat{\sigma}_{(F^j, C_k)}$ 
7     else
8        $SDR \leftarrow SDR + \max_{0 \leq i < |F^j|} \{|(L_{(F^j, C_k)}^{-1})(x_{F^j} - \hat{\mu}_{(F^j, C_k)})_i|\}$ 
9   if  $n^{imp} > 0$  then  $SDR \leftarrow SDR / n^{imp}$ 
10 return( $SDR$ )
    
```

Figure 5: Adapting distribution multipliers.

substantially improved if the objective (and/or constraint) functions allow such local changes to be evaluated efficiently.

### 3.4 Adapting Distribution Multipliers

Similar to MAMaLGaM [1], MO-RV-GOMEA employs distribution multipliers to scale the variance of Gaussian distributions adaptively according to optimization progress (i.e., Adaptive Variance Scaling (AVS) according to the Standard Deviation Ratio (SDR)). MO-RV-GOMEA estimates a separate Gaussian distribution  $N_{F^j}$  for each linkage set  $F^j$  in the linkage model  $\mathcal{F}_{C_k}$  of each cluster  $C_k$ . To counter the variance-diminishing effect of selection when needed, the covariance matrix is scaled by a distribution multiplier  $c_{(F^j, C_k)}^{multiplier}$ , which is adapted according to improvements achieved by mixing with  $F^j$  in  $C_k$ . Pseudocode for this procedure is displayed in Figure 5. If improvements, on average, happen far away from the cluster mean at a distance of more than one standard deviation, i.e.,  $SDR > \theta^{SDR} = 1$ , the variance should thus be increased by scaling the multiplier with some  $\eta^{inc} > 1$ . Further variance enlargement is not required if most improvements happen near the cluster mean.

If no improvements are found, the exploration range might be too large to be effective and should thus be narrowed by decreasing the distribution multiplier by a factor  $\eta^{dec} \in (0, 1)$ . If the distribution multiplier is at its neutral value of 1, no further decrease is allowed, i.e., the search distribution then operates at the normal modus of a maximum-likelihood Gaussian EDA. However, if a sufficiently large number of generations (at least the maximum No-Improvement Stretch  $NIS^{max}$ ) has passed without improvements, a distribution multiplier with a value lower than 1 is allowed, enabling the distribution to converge to a fully contracted state. Here, MO-RV-GOMEA adopts the same SDR parameter values as in [1], which gave good results, i.e.,  $\eta^{dec} = 0.9$ ,  $\eta^{inc} = 1/\eta^{dec}$ . The threshold  $NIS^{max}$  is empirically determined as  $2 + (25 + \ell)/(m + 1)$ , where  $\ell$  is the number of variables and  $m$  is the number of objectives.

```

APPLYANTICIPATEDMEANSHIFT(x)
1  $\mathbf{b} \leftarrow \mathbf{x}$ 
2  $\delta^{\text{AMS}} \leftarrow 2.0$ 
3  $\mathbf{x} \leftarrow \mathbf{x} + \delta^{\text{AMS}} \cdot \hat{\boldsymbol{\mu}}_{C_k}^{\text{shift}}(t)$ 
4 EVALUATESOLUTION(x)
5 if  $\mathbf{x} > \mathbf{b}$  or  $\neg \text{IsDOMINATEDBYARCHIVE}(\mathcal{A}, \mathbf{x})$  then
6    $\text{improved}[\mathbf{x}] \leftarrow \text{True}$ 
7 else  $\mathbf{x} \leftarrow \mathbf{b}$ 
8 return(x)

```

Figure 6: Applying AMS.

### 3.5 Applying AMS

Similar to MAMaLGaM [1], MO-RV-GOMEA employs AMS to increase search effort in the direction of improvement (i.e., toward the Pareto-optimal front). In each cluster, the AMS is computed as the difference between the cluster mean of this generation and that of the previous generation, i.e.,  $\hat{\boldsymbol{\mu}}_{C_k}^{\text{shift}}(t) = \hat{\boldsymbol{\mu}}_{C_k}(t) - \hat{\boldsymbol{\mu}}_{C_k}(t-1)$ . Thanks to cluster registration, cluster indexed  $k$  in generation  $t$  is the cluster that is closest to cluster indexed  $k$  in generation  $t-1$ , thus allowing the AMS to be properly computed between pairs of corresponding clusters in subsequent generations.

During partial variation for all solutions in a cluster  $C_k^{\mathcal{P}}$  by using a linkage set  $F^j$  and its corresponding Gaussian distribution  $N_{F^j}$ , a number, specifically  $\lfloor \frac{1}{2} \tau |C_k^{\mathcal{P}}| \rfloor$ , of these newly (partially-) altered solutions are further moved in the direction of the AMS along the dimensions indicated by linkage set  $F^j$ , i.e.,  $\mathbf{x}_{F^j} \leftarrow \mathbf{x}_{F^j} + c_{F^j, C_k}^{\text{multiplier}} \delta^{\text{AMS}} \hat{\boldsymbol{\mu}}_{C_k}^{\text{shift}}(t)$ . The distribution multiplier  $c_{F^j, C_k}^{\text{multiplier}}$  is employed here to accelerate the movement of solutions in the direction of AMS because the size of the multiplier relates to whether performing variation at variables indicated by  $F^j$  results in improvements far away from the cluster mean. The value  $\delta^{\text{AMS}} = 2$  was previously shown to give good results [1].

It is possible that not all dependency structures are captured by the linkage model, potentially deteriorating the efficiency of the search. To address this problem, a second round of AMS is invoked after all sampling procedures are completed in each cluster. For each cluster,  $\lfloor \frac{1}{2} \tau |C_k^{\mathcal{P}}| \rfloor$  solutions are moved in the direction of the AMS along all dimensions (instead of only the ones associated with the variables in a certain linkage set), i.e.,  $\mathbf{x} \leftarrow \mathbf{x} + \delta^{\text{AMS}} \hat{\boldsymbol{\mu}}_{C_k}^{\text{shift}}(t)$ . Since all variables are considered at the same time, the distribution multipliers  $c_{F^j, C_k}^{\text{multiplier}}$ 's of linkage sets are not used here. Figure 4 shows how AMS is applied during partial variations and Figure 6 shows pseudo-code for the second round of AMS.

### 3.6 Performing Forced Improvements

MO-RV-GOMEA adapts the Forced Improvements (FI) procedure of discrete (MO-)GOMEA [3, 10] for the context of continuous optimization. If a solution  $\mathbf{x}$  has not yielded any Pareto improvement in a stretch of consecutive generations  $\text{NIS}[\mathbf{x}]$  that exceeds the threshold  $\text{NIS}^{\text{max}}$ , the FI procedure is invoked to move  $\mathbf{x}$  in the direction of the current best found solutions (i.e., elitist archive solutions). First, the cluster  $C_k^{\mathcal{P}}$  that  $\mathbf{x}$  belongs to is specified. Second, among the elitist archive solutions associated with cluster  $C_k^{\mathcal{P}}$  (see Section 3.1), the one that is nearest to the cluster mean is chosen as the donor solution  $\mathbf{d}$ . Then, the linkage sets in the cluster's linkage model  $\mathcal{F}_{C_k}$  are traversed. For each linkage set  $F^j$ ,  $\mathbf{x}$  is moved in the direction

```

PERFORMFORCEDIMPROVEMENT(x)
1  $\mathbf{b} \leftarrow \mathbf{x}; k \leftarrow \text{cluster}[\mathbf{x}]; \alpha \leftarrow 0.5$ 
2  $\mathbf{d} \leftarrow \text{FINDELITISTOLUTIONNEARESTTOCENTROIDOFCLUSTER}(C_k)$ 
3 while  $\alpha > 0.05$  do
4   for  $F^j \in \mathcal{F}_{C_k}$  do
5      $\mathbf{x}_{F^j} \leftarrow \alpha \mathbf{x}_{F^j} + (1 - \alpha) \mathbf{d}_{F^j}$ 
6     EVALUATESOLUTION(x)
7     if  $\mathbf{x} > \mathbf{b}$  then
8        $\text{improved}[\mathbf{x}] \leftarrow \text{True}$ 
9       break all loops
10    else  $\mathbf{x} \leftarrow \mathbf{b}$ 
11     $\alpha \leftarrow 0.5\alpha$ 
12 if  $\neg \text{improved}[\mathbf{x}]$  then  $\mathbf{x} \leftarrow \mathbf{d}$ 
13 return(x)

```

Figure 7: Performing Forced Improvements.

resulting from a linear combination of  $\mathbf{x}$  and  $\mathbf{d}$  along the dimensions indicated by  $F^j$ , i.e.,  $\mathbf{x}_{F^j} \leftarrow \alpha \mathbf{x}_{F^j} + (1 - \alpha) \mathbf{d}_{F^j}$ , where  $\alpha$  is the combination coefficient. If this partially-altered  $\mathbf{x}$  (Pareto) dominates the original solution, the changes are accepted. Otherwise,  $\mathbf{x}$  reverts to its original state. At any moment, when acceptance occurs, FI will be terminated to prevent premature convergence due to excessive bias toward elitist archive solutions. Different values for  $\alpha$  are tried to move closer to  $\mathbf{d}$ . If FI cannot (Pareto) improve  $\mathbf{x}$ , then  $\mathbf{x}$  will be replaced by  $\mathbf{d}$ , i.e.,  $\mathbf{x} \leftarrow \mathbf{d}$ . The associated  $\text{NIS}[\mathbf{x}]$  is reset to 0, preventing FI from being applied to  $\mathbf{x}$  again in the next generation. Similarly, when elitist solutions are copied from the archive to replace some solutions in each cluster (see Section 3.1), the associated  $\text{NIS}$ 's of the replaced solutions should be reset to 0 as well. Note that MO-RV-GOMEA keeps track of an overall  $\text{NIS}^{\text{pop}}$  for the whole population to adapt the distribution multipliers and multiple individual  $\text{NIS}[\mathbf{x}]$ 's, one for each candidate solution  $\mathbf{x}$ , to trigger the FI procedure. Figure 7 shows pseudo-code for FI.

## 4 SCALABILITY BENCHMARKS

### 4.1 Benchmark Problems

A set of well-known benchmark problems is used, consisting of the ZDT1, ZDT3, ZDT6 problems [4], and the convex genMED problem. We also introduce the Sum of Rotated Ellipsoid Blocks (SoREB) problem, for which one objective has tight interdependencies within blocks of a constant number of consecutive variables. All tested benchmark problems consist of two objectives.

Both objectives of the genMED problem are entirely dimension-wise decomposable. This makes the problem very simple, but we use this problem to show that exploiting dependencies between problem variables can be helpful in multi-objective optimization, even when the individual objectives are dimension-wise decomposable. The genMED problem was formally defined in Section 2.

We select three of the well-known ZDT functions with different properties. The ZDT1 problem is the easiest of this selection, and its optimal Pareto front is convex. This problem is defined as follows:

$$\begin{aligned}
 f_0^{\text{ZDT1}}(\mathbf{x}) &= x_0 \\
 f_1^{\text{ZDT1}}(\mathbf{x}) &= g(\mathbf{x}) \cdot h(f_0^{\text{ZDT1}}(\mathbf{x}), g(\mathbf{x})) \\
 g(\mathbf{x}) &= 1 + \frac{9}{\ell - 1} \sum_{i=1}^{\ell-1} x_i \\
 h(f(\mathbf{x}), g(\mathbf{x})) &= 1 - \sqrt{\frac{f(\mathbf{x})}{g(\mathbf{x})}}
 \end{aligned}$$

The ZDT3 problem is characterized by its discontinuous optimal Pareto front, and is defined as follows:

$$\begin{aligned} f_0^{\text{ZDT3}}(\mathbf{x}) &= \mathbf{x}_0 \\ f_1^{\text{ZDT3}}(\mathbf{x}) &= g(\mathbf{x}) \cdot h(f_0^{\text{ZDT3}}(\mathbf{x}), g(\mathbf{x})) \\ g(\mathbf{x}) &= 1 + \frac{9}{\ell-1} \sum_{i=1}^{\ell-1} \mathbf{x}_i \\ h(f(\mathbf{x}), g(\mathbf{x})) &= 1 - \sqrt{\frac{f(\mathbf{x})}{g(\mathbf{x})} - \frac{f(\mathbf{x})}{g(\mathbf{x})} \sin(10\pi f(\mathbf{x}))} \end{aligned}$$

The last of our selection is the ZDT6 problem, which has a concave optimal Pareto front. The ZDT6 problem is defined as follows:

$$\begin{aligned} f_0^{\text{ZDT6}}(\mathbf{x}) &= 1 - \exp(-4\mathbf{x}_0) \sin(6\pi \mathbf{x}_0)^6 \\ f_1^{\text{ZDT6}}(\mathbf{x}) &= g(\mathbf{x}) \cdot h(f_0^{\text{ZDT6}}(\mathbf{x}), g(\mathbf{x})) \\ g(\mathbf{x}) &= 1 + 9 \left( \sum_{i=1}^{\ell-1} \frac{\mathbf{x}_i}{\ell-1} \right)^{\frac{1}{4}} \\ h(f(\mathbf{x}), g(\mathbf{x})) &= 1 - \left( \frac{f(\mathbf{x})}{g(\mathbf{x})} \right)^2 \end{aligned}$$

Finally, we introduce the SoREB problem. The first objective is simply the first dimension, and the second objective is the sum of rotated ellipsoid functions, where each rotated ellipsoid function is applied to a block of  $k$  consecutive variables. The rotation function  $R_\theta(\mathbf{x})$  applies a counter-clockwise rotation around the origin with an angle  $\theta$  to  $\mathbf{x}$ . When a non-trivial rotation is applied to each ellipsoid block, strong interdependencies exist within each block of  $k$  variables, but no dependencies exist between variables in different blocks. Specifically, we use blocks of size  $k = 5$  and a rotation angle  $\theta = 45^\circ$ .

$$\begin{aligned} f_1^{\text{SoREB}}(\mathbf{x}) &= \mathbf{x}_0 \\ f_2^{\text{SoREB}}(\mathbf{x}, k) &= \sum_{i=0}^{(\ell-1)/k} [f_{\text{ellipsoid}}(R_\theta([\mathbf{x}_{ki+1}, \dots, \mathbf{x}_{k(i+1)}]), k)] \\ f_{\text{ellipsoid}}(\mathbf{x}) &= \sum_{i=0}^{\ell-1} \left[ 10^{\frac{6i}{\ell-1}} \mathbf{x}_i^2 \right] \end{aligned}$$

## 4.2 Setup of Experiments

For each problem and each problem size, 30 independent runs of each algorithm are performed. All algorithms are implemented in C, and the source code of MO-RV-GOMEA is available on the homepage of the last author. All runs are performed using an 'AMD Opteron Processor 6386 SE' CPU.

We use the following linkage models in MO-RV-GOMEA: Univariate (Uni), Linkage Tree (LT), and Bounded Fixed Linkage Tree (BFLT). In the BFLT model, the maximum size of each linkage set was bounded to the non-trivial size of 100, and the LT was learned based on a distance metric. A completely random distance metric was used for the genMED and ZDT problems, while for the SoREB problem a random low distance was assigned to all pairs of variables within the same block, and a random high distance was assigned to all pairs of variables in different blocks. Due to the distance metric requiring problem-specific knowledge, the BFLT model can only fairly be used in GBO. Efficient partial evaluations on each GBO benchmark problem were used in MO-RV-GOMEA, with the partial evaluation of  $k$  modified variables being counted as  $k/\ell$  evaluations

on the genMED and all ZDT problems. On the SoREB problem, the modification of any variable requires the re-evaluation of its entire block, with the re-evaluation of any block of 5 being counted as  $5/\ell$  evaluations.

The scalability of MO-RV-GOMEA using all introduced linkage models is compared to NSGA-II, MAMaLgAM using the univariate (MAMaLgAM-Uni), and non-factorized (MAMaLgAM-Full) linkage models. The parameters of MO-RV-GOMEA and MAMaLgAM are set according to the guidelines discussed in this paper.

To avoid excessive tuning of the population size parameter, an interleaved multistart scheme, which we shall refer to as IMS, inspired by [12] is used. This scheme interleaves generations of multiple instances of an EA, performing 1 generation of an instance with population size  $2n$  for each  $c^{\text{IMS}}$  generations of an instance with population size  $n$ . For each increment of the population size by a factor 2, the number of clusters is incremented by 1. The initial instance of an EA starts with a population size  $n^{\text{base}}$  and  $q^{\text{base}}$  clusters, which are typically chosen to be small. Specifically, we use  $c^{\text{IMS}} = 8$ ,  $q^{\text{base}} = m + 1$ , and  $n^{\text{base}} = 10 \cdot q^{\text{base}}$ . Instances with small populations can be terminated when they are regarded to be of worse quality than instances with a larger population. We terminate an instance if its population, and each instance with a smaller population, each contribute less than 10% to the complete set of rank-0 solutions. This set of rank-0 solutions is a Pareto set that consists of all solutions in any non-terminated population that are not dominated by any solution in any other population.

For a notion of Pareto front quality, we use the  $D_{\mathcal{P}_F \rightarrow \mathcal{S}}$  metric [2]. This metric describes the mean distance from each point in  $\mathcal{P}_F$  to their nearest neighbor in a given approximation front  $\mathcal{S}$ . A run of any algorithm is considered successful when a  $D_{\mathcal{P}_F \rightarrow \mathcal{S}}$  from the elitist archive to a selection of 5000 points on the optimal Pareto front smaller than 0.001 is reached within the one hour time limit.

For the SoREB problem, the population is initialized uniformly in the range  $[-115, -110]^\ell$ . This range is far away from the global optimum, which makes the problem increasingly difficult to optimize if only a univariate operator is used. For all other problems, the population is initialized uniformly in the range  $[0, 1]^\ell$ . All ZDT problems are constrained to the range  $[0, 1]^\ell$ , and the first parameter of the SoREB problem is constrained to the range  $[0, 1]$ . The remaining problems are unconstrained. Boundary repair is used when a variable value is outside the feasible range of the problem, meaning that the value is set to its nearest feasible value. A maximum elitist archive size of 1000 was used for each algorithm.

## 4.3 Results of Experiments

The observed scalability of all considered algorithms in terms of time and evaluations required to solve the benchmark problems is shown in Figure 8. In the following we discuss our findings.

**4.3.1 GBO.** In GBO, MO-RV-GOMEA performs and scales better than both MAMaLgAM and NSGA-II in terms of time and required number of evaluations on all considered benchmark problems when an appropriate linkage model is used. On all ZDT problems, MO-RV-GOMEA with the BFLT model performed the best, but was only a constant factor better than MO-RV-GOMEA with the univariate model. This does indicate that, even though the ZDT problems are completely dimension-wise decomposable, it

is indeed beneficial to consider linkage structure, as discussed in Section 2. For the ZDT problems with low dimensionality, the LT model performed approximately as good or even better than the univariate and BFLT models. However, for larger dimensional problems, the LT scales clearly worse. This is partially caused by the elitist archive, which is updated after applying one linkage set to the entire population in GOM. Moreover, when using the full LT, full covariance matrices are estimated and decomposed for sampling, which is far more expensive in terms of computational overhead than is for instance the case for the univariate operators used by default in NSGA-II. Another issue is the required population size for the full LT to work well, which is quite a bit larger than the population size that is started from in the IMS. This is a direct consequence of using AMaLGaM mechanisms for each FOS element. As a consequence, scalability could be improved by starting the IMS from a larger initial population size for larger dimensional problems in case of using a full LT. We aim to explore this effect further in future research.

On the genMED problem, the BFLT performed best, but it was approximately as good as the univariate and LT models. Finally, NSGA-II and univariate MAMaLGaM were unable to solve the SoREB problem within the time limit for any problem dimensionality. This is because it is very difficult for EAs with univariate operators to optimize highly-dependent problems, especially when the parameters are initialized far away from their globally optimal value. The univariate MO-RV-GOMEA performed slightly better, and was able to solve the SoREB problem with low dimensionality, but poor scalability is clearly observed. Conversely, excellent scalability is achieved by MO-RV-GOMEA when the BFLT model is used, because this model can efficiently exploit the linkage structure of the highly-dependent blocks of variables of the SoREB problem, while avoiding the large overhead caused by using a non-factorized (full) linkage set. In terms of (discounted) evaluations, strongly sub-linear scalability is observed, even on problems where the individual objectives are not univariately decomposable, such as SoREB.

**4.3.2 BBO.** In this setting, MO-RV-GOMEA performs better than NSGA-II and MAMaLGaM on all benchmark problems in terms of time when an appropriate linkage model is used. The LT model performs very well, seeing as MO-RV-GOMEA with the LT model performs better than all other algorithms on the genMED, ZDT6, and SoREB problems in terms of evaluations. Also in BBO it is thus beneficial to consider linkage structure if a degree of decomposability exists. However, for similar reasons as mentioned for the GBO setting, for larger dimensionalities, the scalability of MO-RV-GOMEA when using the full LT starts to become worse, especially in terms of actual computation time. Finally, MO-RV-GOMEA performs slightly worse than NSGA-II on the ZDT3 problem, but scales better and is expected to perform better for higher-dimensional problems. Even though no efficient partial evaluations can be performed in BBO, MO-RV-GOMEA still performs very well, and the LT has been observed to be a very general and competent model.

## 5 CONCLUSIONS

We have introduced MO-RV-GOMEA, which uses an explicitly defined linkage model to exploit linkage structure of multi-objective real-valued optimization problems. Aligning the linkage model

with the problem structure can be done through the application of linkage learning on the population, or by using problem-specific knowledge to define a linkage model a priori. When the linkage model is roughly aligned with the problem structure, substantial improvement gains are possible due to the variation operator of MO-RV-GOMEA, which incrementally improves parts of solutions that are considered dependent by the linkage model. We have compared the scalability of MO-RV-GOMEA to the state-of-the-art algorithms NSGA-II and MAMaLGaM on a set of well-known benchmark problems that are to some degree decomposable. In a BBO setting, MO-RV-GOMEA performs substantially better than NSGA-II and MAMaLGaM in terms of time, and is sometimes better and sometimes worse in terms of the required number of evaluations. When efficient partial evaluations are possible, MO-RV-GOMEA scales substantially better than MAMaLGaM and NSGA-II, achieving sub-linear scalability in terms of the number of evaluations on all tested benchmark problems. Moreover, NSGA-II has large difficulty solving optimization problems when the initialization range of the population is far away from the global optimum, even for very low-dimensional problems, whereas this is no problem for MO-RV-GOMEA. We conclude that MO-RV-GOMEA is a promising new EA for BBO when an appropriate linkage model is used, with the univariate model being the simplest model and the LT model being the most general. Moreover, considering optimization problems in the setting that we were at the outset primarily interested in, i.e., the GBO case where efficient partial evaluations are possible, MO-RV-GOMEA really excels, achieving even strongly sub-linear scalability in the required number of evaluations on a set of well-known benchmark problems.

## REFERENCES

- [1] P.A.N. Bosman. 2010. The anticipated mean shift and cluster registration in mixture-based EDAs for multi-objective optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2010)*. 351–358.
- [2] P.A.N. Bosman and D. Thierens. 2003. The balance between proximity and diversity in multiobjective evolutionary algorithms. *IEEE Transactions on Evolutionary Computation* 7, 2 (2003), 174–188.
- [3] P.A.N. Bosman and D. Thierens. 2013. More concise and robust linkage learning by filtering and combining linkage hierarchies. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2013)*. 359–366.
- [4] K. Deb. 1999. Multi-objective genetic algorithms: Problem difficulties and construction of test problems. *Evolutionary Computation* 7, 3 (1999), 205–230.
- [5] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE transactions on Evolutionary Computation* 6, 2 (2002), 182–197.
- [6] I. Gronau and S. Moran. 2007. Optimal implementations of UPGMA and other common clustering algorithms. *Information Processing Letters* 104, 6 (2007), 205–210.
- [7] J. Knowles and D. Corne. 2003. Properties of an adaptive archiving algorithm for storing nondominated vectors. *IEEE Transactions on Evolutionary Computation* 7, 2 (2003), 100–116.
- [8] A. Kraskov, H. Stögbauer, and P. Grassberger. 2004. Estimating mutual information. *Physical Review E* 69 (2004), 066138. Issue 6.
- [9] N.H. Luong and P.A.N. Bosman. 2012. Elitist archiving for multi-objective evolutionary algorithms: To adapt or not to adapt. In *International Conference on Parallel Problem Solving from Nature*. Springer, 72–81.
- [10] N.H. Luong, H. La Poutre, and P.A.N. Bosman. 2014. Multi-objective gene-pool optimal mixing evolutionary algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2014)*. 357–364.
- [11] M. Pelikan, K. Sastry, and D.E. Goldberg. 2005. Multiobjective hBOA, clustering, and scalability. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2005)*. ACM, 663–670.
- [12] J.C. Pereira and F.G. Lobo. 2015. A Java Implementation of Parameter-less Evolutionary Algorithms. *arXiv preprint arXiv:1506.08694* (2015).
- [13] V.N. Vapnik. 1998. *Statistical learning theory*. Vol. 1. Wiley New York.

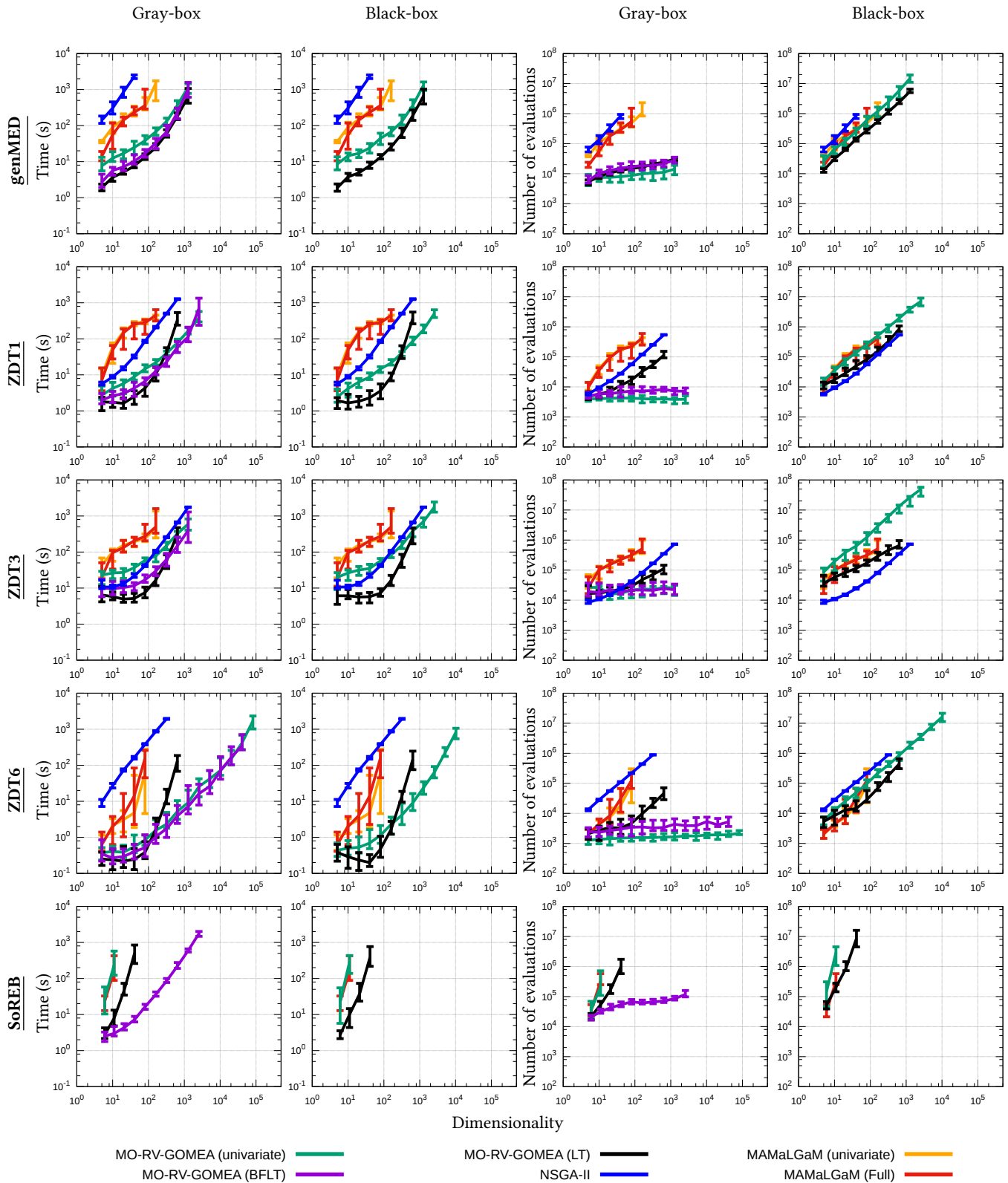


Figure 8: Medians and interdecile ranges of all experiments with each data point being the median of 30 successful runs.