Online Supplementary Material

Bach Hoai Nguyen, *Member, IEEE*, Bing Xue, *Member, IEEE*, Peter Andreae,
Hisao Ishibuchi, *Fellow, IEEE*, and Mengjie Zhang, *Fellow, IEEE*

I. A SENSITIVITY STUDY ON THE NUMBER OF MOVING REFERENCE POINTS

In this section, a set of experiments are conducted to examine effect of different numbers of moving reference points M. Specifically, given a fixed total number of reference points (R), we examine five different values of the ratio M/R: 10%, 20%, 30%, 40%, and 50%. Note that a ratio of 0% is equivalent to the static version which has been already shown to be worse than the dynamic algorithm. Tables I and II shows the hypervolume values on the training set and the test set obtained by the five different ratios. As can be seen from the tables, there is very little difference between the hypervolume values of the five ratios. We also use the Wilcoxon signed-rank test at the 5% significance level to compare results obtained by 50% and the other four settings. The comparisons show that the results are not significantly different on all datasets, which illustrates that the proposed algorithm is not sensitive to the number of moving reference points.

II. FURTHER COMPARISONS WITH STATE-OF-THE-ART MANY-OBJECTIVE EMO ALGORITHMS

In this section, MOEA/D-STAT is compared with three state-of-the-art EMO algorithms: NSGA-III [1], MOEA/DD [2], θ-DEA [3]. All three benchmark algorithms use multiple reference points to generate reference lines, and then each solution is assigned to its closest reference line. Instead of using a crowding distance as in NSGA-II, NSGA-III uses the number of solutions associated with each reference line, known as a niche count, to select solutions from the boundary non-domination level. MOEA/DD is based on MOEA/D, but has a different update procedure. After generating offspring solutions, MOEA/DD sequentially adds one offspring solution to the current population and then removes one solution from the obtained population. The removed solution is selected according to its non-dominated rank, its PBI value, and the number of solutions assigned to its sub-region. θ-DEA divides the population into clusters corresponding to the reference lines. The solutions are then ranked by θ dominance. Basically, $\mathbf{x}$ is said to θ-dominate $\mathbf{y}$ if they are from the same cluster and the PBI value of $\mathbf{x}$ is smaller than that of $\mathbf{y}$.

Tables III and IV show the hypervolume values obtained by MOEA/D-STAT and the three many-objective algorithms on training and test sets, respectively. It can be seen that MOEA/D-DYN achieves similar or significantly better hypervolume values than the three benchmark algorithms. NSGA-III has the worst performance, mainly because, like NSGA-II, it relies heavily on the non-dominated sorting. Figs. 1 and 2 show the median fronts (i.e., the front corresponding to the median hypervolume value) obtained by the four algorithms. It can be seen that NSGA-III's front has the smallest number of solutions, and its solutions are generally dominated by solutions of the other algorithms. θ-DEA evolves more diverse fronts than NSGA-III. In comparison between θ-DEA and MOEA/DD, the latter usually has more diverse fronts. However, given the same number of features, feature subsets evolved by θ-DEA usually have smaller error rates. MOEA/D-DYN has much more diverse fronts than either MOEA/DD or θ-DEA, and it can evolve feature subsets with small error rates that cannot be achieved by either MOEA/DD or θ-DEA. The experimental results show that in feature selection, allocating reference points along the $fRatio$ axis results in a more diverse solution set than allocating reference points evenly on the line $w_f + w_e = 1$ where w_f and w_e are the weights corresponding to $fRatio$ and $eRate$, respectively.

III. A FURTHER COMPARISON WITH MOEA/D-BASED FILTER FEATURE SELECTION

In this section, MOEA/D-DYN is compared with the first MOEA/D-based feature selection algorithm from Paul and Das [4]. Their algorithm (which we refer to as MOEA/D-F) is a filter-based feature selection approach which has two objectives: minimizing the intra-class distance and maximizing the inter-class distance. The number of selected features is added to each distance as a penalty to increase the affinity towards selecting fewer features. Table V shows the comparison between MOEA/D-DYN and MOEA/D-F on the test sets. It can be seen that on all datasets MOEA/D-DYN achieves significantly better hypervolume than MOEA/D-F.

Fig. 3 shows the median fronts obtained by the two algorithms. Although MOEA/D-F considers the number of features during its evolutionary process, it tends to select a large number of features. On most datasets, the largest feature subset selected by MOEA/D-DYN is still smaller than the smallest feature subset selected by MOEA/D-F. The reason is that MOEA/D-DYN directly considers the number of selected features as an objective to optimize. On the other hand, MOEA/D-F uses the number of selected features as a penalty, which reduces the pressure to select fewer features. Further, combining the inter/intra-class distances and the number of selected features is problematic since they are different kinds of measures with different ranges. Simply adding them (even with weights) does not give good results. Regarding the classification accuracy, given the same number of features, MOEA/D-DYN evolves better feature subsets with lower classification error rates than MOEA/D-F. This is mainly because MOEA/D-DYN uses a classification algorithm during its training process, which considers the interaction between the selected features and the wrapped classification algorithm.

TABLE I: Hypervolume obtained by different number of moving reference points on the training sets.

Dataset	10%	20%	30%	40%	50%
Wine	0.877±0.001 (○)	0.877±0.001 (○)	**0.878±0.001** (○)	**0.878±0.001** (○)	**0.878±0.001**
Australian	**0.795±0.000** (○)	**0.795±0.002** (○)	**0.795±0.000** (○)	**0.795±0.000** (○)	**0.795±0.000**
Vehicle	**0.802±0.001** (○)	**0.802±0.001** (○)	**0.802±0.001** (○)	**0.802±0.001** (○)	**0.802±0.001**
German	**0.720±0.003** (○)	**0.720±0.003** (○)	0.719±0.004 (○)	0.719±0.003 (○)	0.719±0.003
WBCD	**0.920±0.000** (○)	**0.920±0.000** (○)	**0.920±0.000** (○)	**0.920±0.000** (○)	**0.920±0.000**
Sonar	**0.894±0.008** (○)	0.885±0.007 (○)	0.890±0.007 (○)	0.891±0.009 (○)	0.886±0.009
Hillvalley	**0.625±0.003** (○)	0.624±0.003 (○)	**0.625±0.002** (○)	**0.625±0.003** (○)	**0.625±0.003**
Musk1	**0.935±0.004** (↑)	0.932±0.004 (○)	0.931±0.004 (○)	0.931±0.003 (○)	0.929±0.004
Arrhythmia	**0.958±0.001** (○)	**0.958±0.001** (○)	**0.958±0.001** (○)	0.957±0.001 (○)	0.957±0.001
Madelon	0.895±0.003 (○)	**0.897±0.003** (○)	0.896±0.003 (○)	0.895±0.003 (○)	0.896±0.003
Isolet5	**0.991±0.000** (○)	**0.991±0.000** (○)	**0.991±0.000** (○)	**0.991±0.000** (○)	**0.991±0.000**
MultipleFeatures	**0.994±0.000** (○)	**0.994±0.000** (○)	**0.994±0.000** (○)	**0.994±0.000** (○)	**0.994±0.000**

TABLE II: Hypervolume obtained by different number of moving reference points on the test sets.

Dataset	10%	20%	30%	40%	50%
Wine	**0.904±0.000** (○)	**0.904±0.000** (○)	**0.904±0.000** (○)	**0.904±0.000** (○)	**0.904±0.000**
Australian	**0.793±0.003** (○)	0.787±0.007 (○)	0.791±0.002 (○)	0.790±0.005 (○)	0.790±0.005
Vehicle	0.796±0.002 (○)	0.798±0.002 (○)	0.798±0.004 (○)	**0.799±0.004** (○)	0.798±0.002
German	0.681±0.006 (○)	**0.683±0.006** (○)	0.681±0.006 (○)	**0.683±0.007** (○)	0.680±0.008
WBCD	**0.914±0.000** (○)	**0.914±0.000** (○)	**0.914±0.000** (○)	**0.914±0.001** (○)	**0.914±0.000**
Sonar	**0.811±0.013** (↑)	0.800±0.016 (○)	0.805±0.015 (○)	0.795±0.020 (○)	0.797±0.013
Hillvalley	0.593±0.007 (○)	0.590±0.009 (○)	0.593±0.010 (○)	**0.599±0.011** (○)	0.596±0.009
Musk1	0.876±0.007 (○)	**0.878±0.013** (○)	0.869±0.005 (○)	0.875±0.009 (○)	0.871±0.006
Arrhythmia	0.952±0.002 (○)	**0.953±0.002** (○)	0.952±0.003 (○)	0.952±0.002 (○)	0.952±0.002
Madelon	0.886±0.005 (○)	0.885±0.003 (○)	**0.887±0.003** (○)	0.886±0.004 (○)	0.886±0.005
Isolet5	**0.989±0.000** (○)	**0.989±0.001** (○)	**0.988±0.001** (○)	**0.989±0.001** (○)	**0.989±0.001**
MultipleFeatures	**0.990±0.001** (○)	**0.990±0.000** (○)	**0.990±0.001** (○)	**0.990±0.001** (○)	**0.990±0.001**

TABLE III: Comparisons (hypervolume) between MOEA/D-DYN and many-objective algorithms on the training sets.

Dataset	NSGA-III	MOEA/DD	θ-DEA	MOEA/D-DYN
Wine	0.869±0.012 (↓)	0.870±0.004 (↓)	0.863±0.015 (↓)	**0.877±0.001**
Australian	0.779±0.014 (↓)	0.788±0.007 (↓)	0.777±0.014 (↓)	**0.795±0.000**
Vehicle	0.787±0.021 (↓)	0.792±0.005 (↓)	0.787±0.013 (↓)	**0.802±0.001**
German	0.698±0.017 (↓)	0.714±0.006 (↓)	0.711±0.007 (↓)	**0.719±0.003**
WBCD	0.902±0.026 (↓)	0.919±0.001 (↓)	0.916±0.006 (↓)	**0.920±0.000**
Sonar	0.796±0.024 (↓)	0.874±0.011 (↓)	0.871±0.010 (↓)	**0.889±0.009**
Hillvalley	0.510±0.024 (↓)	0.617±0.004 (↓)	0.619±0.003 (↓)	**0.625±0.003**
Musk1	0.745±0.028 (↓)	0.925±0.005 (↓)	0.928±0.005 (↓)	**0.931±0.004**
Arrhythmia	0.749±0.023 (↓)	0.951±0.004 (↓)	0.950±0.004 (↓)	**0.957±0.001**
Madelon	0.618±0.021 (↓)	0.886±0.008 (↓)	0.894±0.005 (○)	**0.896±0.003**
Isolet5	0.721±0.019 (↓)	0.944±0.014 (↓)	0.964±0.005 (↓)	**0.991±0.000**
MultipleFeatures	0.744±0.017 (↓)	0.969±0.010 (↓)	0.973±0.003 (↓)	**0.994±0.000**

TABLE IV: Comparisons (hypervolume) between MOEA/D-DYN and many-objective algorithms on the test sets.

Dataset	NSGA-III	MOEA/DD	θ-DEA	MOEA/D-DYN
Wine	0.883±0.024 (↓)	0.895±0.021 (↓)	0.884±0.028 (↓)	**0.904±0.000**
Australian	0.747±0.060 (↓)	0.743±0.063 (↓)	0.742±0.059 (↓)	**0.790±0.005**
Vehicle	0.784±0.020 (↓)	0.790±0.004 (↓)	0.783±0.015 (↓)	**0.798±0.003**
German	0.655±0.023 (↓)	0.678±0.011 (○)	0.675±0.015 (○)	**0.680±0.006**
WBCD	0.887±0.029 (↓)	**0.914±0.000** (○)	0.911±0.007 (○)	**0.914±0.000**
Sonar	0.741±0.033 (↓)	0.784±0.021 (○)	0.780±0.024 (○)	**0.791±0.020**
Hillvalley	0.500±0.023 (↓)	0.596±0.008 (○)	**0.602±0.010** (○)	0.598±0.012
Musk1	0.701±0.035 (↓)	0.842±0.019 (↓)	0.847±0.021 (↓)	**0.874±0.008**
Arrhythmia	0.748±0.023 (↓)	0.946±0.004 (↓)	0.943±0.004 (↓)	**0.953±0.002**
Madelon	0.612±0.022 (↓)	0.874±0.009 (↓)	0.879±0.005 (↓)	**0.886±0.004**
Isolet5	0.719±0.019 (↓)	0.941±0.014 (↓)	0.962±0.005 (↓)	**0.989±0.001**
MultipleFeatures	0.742±0.017 (↓)	0.965±0.010 (↓)	0.968±0.003 (↓)	**0.990±0.001**

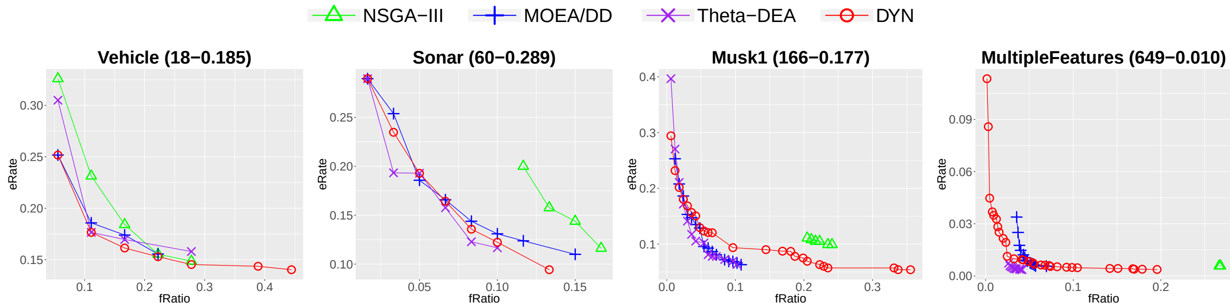

Fig. 1: Median fronts obtained by many-objective algorithms and MOEA/D-DYN on the training sets.

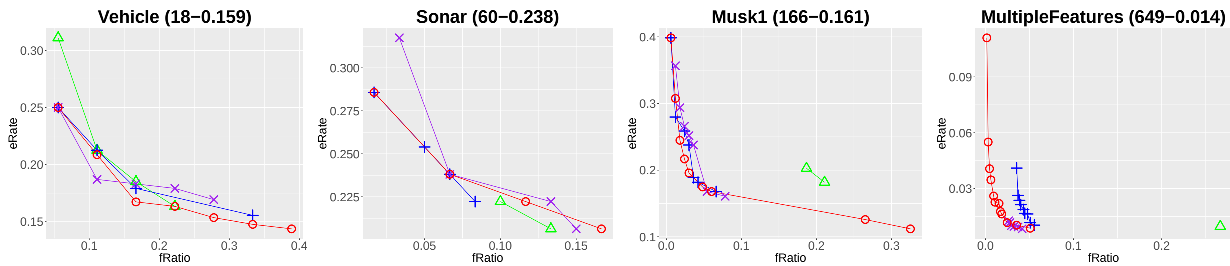

Fig. 2: Median fronts obtained by many-objective algorithms and MOEA/D-DYN on the test sets.

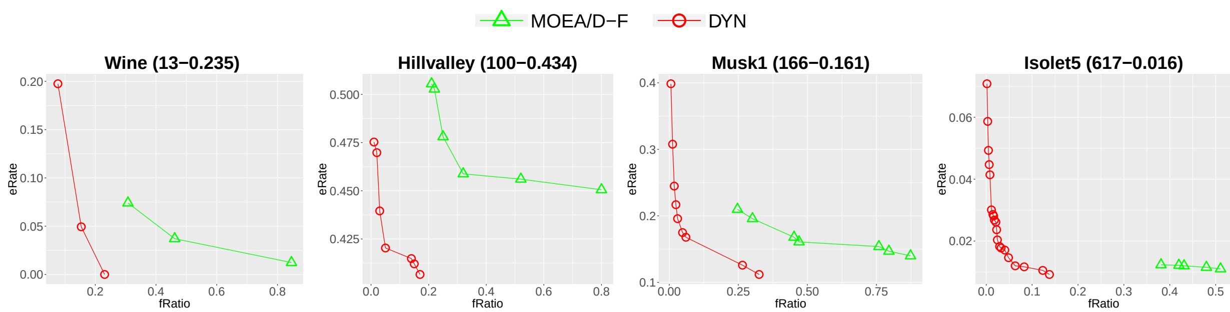

Fig. 3: Median fronts obtained by MOEA/D-DYN (DYN) and MOEA/D-F.

TABLE V: Comparison between MOEA/D-DYN and MOEA/D-F in terms of hypervolume.

Dataset	MOEA/D-F	MOEA/D-DYN
Wine	0.737±0.067 (↓)	**0.904±0.000**
Australian	0.665±0.067 (↓)	**0.790±0.005**
Vehicle	0.689±0.059 (↓)	**0.798±0.003**
German	0.587±0.048 (↓)	**0.680±0.006**
WBCD	0.861±0.035 (↓)	**0.914±0.000**
Sonar	0.694±0.038 (↓)	**0.791±0.020**
Hillvalley	0.454±0.026 (↓)	**0.598±0.012**
Musk1	0.657±0.028 (↓)	**0.874±0.008**
Arrhythmia	0.614±0.030 (↓)	**0.953±0.002**
Madelon	0.385±0.019 (↓)	**0.886±0.004**
Isolet5	0.641±0.032 (↓)	**0.989±0.001**
MultipleFeatures	0.649±0.029 (↓)	**0.990±0.001**

On the other hand, MOEA/D-F uses inter/intra-class distances as an estimation of the classification accuracy, which results in larger classification errors than MOEA/D-DYN.

References

[1] K. Deb and H. Jain, "An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: Solving problems with box constraints," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 4, pp. 577–601, 2014.

[2] K. Li, K. Deb, Q. Zhang, and S. Kwong, "An evolutionary many-objective optimization algorithm based on dominance and decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 5, pp. 694–716, 2015.

[3] Y. Yuan, H. Xu, B. Wang, and X. Yao, "A new dominance relation-based evolutionary algorithm for many-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 1, pp. 16–37, 2016.

[4] S. Paul and S. Das, "Simultaneous feature selection and weighting - an evolutionary multi-objective optimization approach," *Pattern Recognition Letters*, vol. 65, pp. 51–59, 2015.