Supplementary Document for "The Importance of Diversity in the Variable Space in the Design of Multi-objective Evolutionary Algorithms"

Carlos Segura[a], Joel Chacón Castillo[a], Oliver Schütze[b]

[a] *Center for Research in Mathematics (CIMAT), Computer Science Department, Callejón Jalisco s/n, Mineral de Valenciana, Guanajuato, Guanajuato 36240, Mexico*
[b] *Department of Computer Science, CINVESTAV-IPN, Mexico City, Mexico*

Abstract

Keywords: Diversity, Decomposition, Multi-objective Optimization, Evolutionary Algorithms.

1. Comparison against State-of-the-art MOEAs in long-term executions

One of the aims behind the design of AVSD-MOEA/D is to profit from long-term executions. Therefore, in this section we present the results attained by the different algorithms when setting the stopping criterion to 2.5×10^7 function evaluations. Table **??** shows the HV ratios obtained for the benchmark functions with two objectives. Note that the same results can be drawn with the IGD+ metric [1] and can be inspected in the supplementary material. For each method and problem, the best, mean and standard deviation of

Email addresses: carlos.segura@cimat.mx (Carlos Segura), joel.chacon@cimat.mx (Joel Chacón Castillo), shuetze@cs.cinvestav.mx (Oliver Schütze)

Table 1: Summary of the IGD+ attained for problems with two objectives

	AVSD-MOEA/D			**MOEA/D-DE**			**NSGA-II**			**NSGA-III**			**R2-EMOA**		
	Best	**Mean**	**Std**	**Best**	**Mean**	**Std**	**Best**	**Mean**	**Std**	**Best**	**Mean**	**Std**	**Best**	**Mean**	**Std**
WFG1	0.006	0.020	0.024	0.048	0.195	0.076	0.008	0.039	0.030	0.009	**0.014**	0.012	0.009	0.094	0.049
WFG2	0.003	**0.003**	0.000	0.008	0.008	0.000	0.004	0.004	0.000	0.007	0.045	0.066	0.004	0.006	0.001
WFG3	0.008	**0.008**	0.000	0.010	0.010	0.000	0.021	0.022	0.001	0.010	0.010	0.000	0.010	0.011	0.000
WFG4	0.006	**0.006**	0.000	0.009	0.009	0.000	0.014	0.016	0.001	0.009	0.010	0.001	0.008	0.014	0.003
WFG5	0.037	**0.056**	0.005	0.065	0.067	0.001	0.071	0.072	0.001	0.060	0.065	0.002	0.065	0.067	0.001
WFG6	0.024	0.047	0.013	0.009	0.022	0.011	0.014	0.016	0.001	0.026	0.039	0.008	0.007	**0.007**	0.000
WFG7	0.006	**0.006**	0.000	0.009	0.009	0.000	0.013	0.015	0.001	0.009	0.009	0.000	0.007	0.007	0.000
WFG8	0.034	**0.048**	0.005	0.110	0.115	0.002	0.119	0.124	0.002	0.116	0.117	0.001	0.116	0.118	0.001
WFG9	0.009	**0.011**	0.001	0.012	0.028	0.025	0.031	0.077	0.046	0.123	0.126	0.001	0.011	0.035	0.033
DTLZ1	0.001	**0.001**	0.000	0.001	**0.001**	0.000	0.002	0.002	0.000	0.001	**0.001**	0.000	0.002	0.002	0.000
DTLZ2	0.002	**0.002**	0.000	0.003	0.003	0.000	0.003	0.004	0.000	0.003	0.003	0.000	0.002	**0.002**	0.000
DTLZ3	0.002	**0.002**	0.000	0.003	0.003	0.000	0.003	0.003	0.000	0.003	0.003	0.000	0.002	**0.002**	0.000
DTLZ4	0.002	**0.002**	0.000	0.003	0.003	0.000	0.003	0.058	0.150	0.003	0.003	0.000	0.002	0.124	0.207
DTLZ5	0.002	**0.002**	0.000	0.003	0.003	0.000	0.003	0.004	0.000	0.003	0.003	0.000	0.002	**0.002**	0.000
DTLZ6	0.002	**0.002**	0.000	0.003	0.004	0.003	0.003	0.003	0.000	0.003	0.003	0.000	0.002	0.004	0.005
DTLZ7	0.002	**0.002**	0.000	0.004	0.004	0.000	0.003	0.003	0.000	0.004	0.004	0.000	0.002	**0.002**	0.000
UF1	0.003	**0.003**	0.000	0.007	0.007	0.000	0.005	0.006	0.000	0.004	0.004	0.000	0.004	0.004	0.000
UF2	0.003	**0.003**	0.000	0.006	0.007	0.001	0.009	0.011	0.001	0.009	0.014	0.003	0.008	0.009	0.001
UF3	0.030	0.042	0.007	0.004	**0.005**	0.000	0.006	0.008	0.002	0.008	0.022	0.018	0.005	0.010	0.004
UF4	0.007	**0.007**	0.000	0.027	0.031	0.002	0.034	0.037	0.001	0.038	0.039	0.001	0.030	0.033	0.001
UF5	0.016	**0.026**	0.006	0.140	0.251	0.063	0.094	0.129	0.032	0.103	0.146	0.022	0.094	0.135	0.067
UF6	0.017	**0.024**	0.012	0.036	0.225	0.151	0.078	0.202	0.060	0.078	0.130	0.074	0.081	0.220	0.103
UF7	0.003	**0.003**	0.000	0.004	0.004	0.000	0.008	0.010	0.001	0.010	0.016	0.002	0.004	0.012	0.005
Mean	0.010	**0.014**	0.003	0.023	**0.044**	0.015	0.024	**0.038**	0.014	0.028	**0.036**	0.009	0.021	**0.040**	0.021

the HV ratio values are reported. Furthermore, in order to summarize the results attained by each method, the last row shows the mean for the whole set of problems. For each test problem, the method that yielded the largest mean and those that were not statistically inferior to the best are shown in **boldface**. Similarly, the method that yielded the best HV value among all the runs is underlined. From here on, the methods shown in **boldface** for a given problem are referred to as the winning methods. AVSD-MOEA/D, R2-EMOA, MOEA/D-DE, NSGA-III and NSGA-II belonged to the winning methods

Table 2: Statistical Tests and Deterioration Level of the IGD+ for problems with two objectives

	↑	↓	↔	**Score**	**Deterioration**
AVSD-MOEA/D	78	13	1	65	0.160
MOEA/D-DE	41	50	1	-9	1.181
NSGA-II	21	66	5	-45	1.057
NSGA-III	35	52	5	-17	1.119
R2-EMOA	47	41	4	6	1.066

in 17, 6, 2, 2 and 0 problems, respectively. The superiority of AVSD-MOEA/D is clear both in terms of this metric and in terms of the mean HV. Particularly, AVSD-MOEA/D attained a value equal to 0.976, while all the remaining methods attained values between 0.931 and 0.937. A careful inspection of the data shows that in those cases where AVSD-MOEA/D loses, the difference with respect to the best method is low. In fact, the difference between the mean HV ratio attained by the best method and by AVSD-MOEA/D is never greater than 0.1. However, in all the other methods, there were several problems where the distance with respect to the best approach was greater than 0.1. Specifically, it happened in 4, 4, 4 and 5 problems for R2-EMOA, MOEA/D-DE, NSGA-II and NSGA-III, respectively. This means that AVSD-MOEA/D wins in most cases and that when it loses, the difference is always small. Note also that in terms of standard deviation, AVSD-MOEA/D yields much lower values than all the other algorithms, meaning it is quite robust.

In order to better clarify these findings, pair-wise statistical tests were applied between each method tested in each test problem. For the two-objective cases, Table **??** shows the number of times that each method statistically won (column ↑), lost (column ↓) or tied (column ↔). The **Score** column shows

Table 3: Summary of the IGD+ attained for problems with three objectives

	AVSD-MOEA/D			**MOEA/D-DE**			**NSGA-II**			**NSGA-III**			**R2-EMOA**		
	Best	**Mean**	**Std**	**Best**	**Mean**	**Std**	**Best**	**Mean**	**Std**	**Best**	**Mean**	**Std**	**Best**	**Mean**	**Std**
WFG1	0.073	**0.085**	0.010	0.083	0.136	0.043	0.108	0.129	0.012	0.092	0.096	0.010	0.079	0.104	0.023
WFG2	0.055	**0.057**	0.001	0.062	0.069	0.004	0.096	0.135	0.021	0.097	0.113	0.018	0.119	0.120	0.000
WFG3	0.026	**0.027**	0.000	0.032	0.032	0.000	0.047	0.095	0.030	0.084	0.098	0.012	0.033	0.034	0.000
WFG4	0.088	**0.092**	0.001	0.133	0.133	0.000	0.132	0.142	0.009	0.133	0.133	0.000	0.119	0.124	0.002
WFG5	0.122	**0.137**	0.006	0.185	0.185	0.000	0.181	0.192	0.008	0.182	0.185	0.001	0.166	0.169	0.002
WFG6	0.112	0.130	0.009	0.140	0.158	0.009	0.159	0.183	0.012	0.145	0.162	0.009	0.120	**0.128**	0.005
WFG7	0.089	**0.091**	0.001	0.133	0.133	0.000	0.130	0.158	0.010	0.133	0.133	0.000	0.117	0.119	0.001
WFG8	0.122	**0.128**	0.003	0.191	0.193	0.001	0.242	0.254	0.006	0.194	0.198	0.002	0.174	0.178	0.002
WFG9	0.100	**0.103**	0.001	0.135	0.138	0.001	0.178	0.252	0.017	0.149	0.237	0.022	0.129	0.133	0.002
DTLZ1	0.015	0.015	0.000	0.014	**0.014**	0.000	0.019	0.021	0.001	0.014	**0.014**	0.000	0.015	0.015	0.000
DTLZ2	0.023	**0.024**	0.000	0.029	0.029	0.000	0.033	0.037	0.002	0.029	0.029	0.000	0.026	0.027	0.000
DTLZ3	0.023	**0.023**	0.000	0.029	0.029	0.000	0.035	0.039	0.002	0.029	0.029	0.000	0.026	0.027	0.000
DTLZ4	0.023	**0.023**	0.000	0.029	0.029	0.000	0.032	0.107	0.200	0.029	0.042	0.075	0.026	0.045	0.106
DTLZ5	0.004	0.004	0.000	0.005	0.005	0.000	0.003	**0.003**	0.000	0.008	0.010	0.002	0.003	**0.003**	0.000
DTLZ6	0.004	0.004	0.000	0.005	0.009	0.007	0.003	0.010	0.029	0.010	0.013	0.002	0.003	**0.003**	0.001
DTLZ7	0.033	**0.033**	0.000	0.059	0.059	0.000	0.040	0.060	0.056	0.050	0.061	0.005	0.075	0.113	0.047
UF8	0.030	**0.032**	0.001	0.040	0.054	0.016	0.089	0.111	0.026	0.040	0.075	0.066	0.042	0.050	0.008
UF9	0.029	**0.031**	0.001	0.038	0.169	0.071	0.103	0.164	0.058	0.032	0.046	0.041	0.034	0.110	0.085
UF10	0.060	**0.072**	0.010	0.105	0.309	0.091	0.229	0.273	0.043	0.154	0.276	0.055	0.254	0.261	0.017
Mean	0.054	**0.058**	0.002	0.076	**0.099**	0.013	0.098	**0.124**	0.029	0.084	**0.103**	0.017	0.082	**0.093**	0.016

the difference between the number of times that each method won and the number of times that each method lost. Additionally, for each method M, we calculated the sum of the differences between the mean HV ratio attained by the best method (the ones with the highest mean) and method M, for each problem where M was not in the group of winning methods. This value is shown in the *Deterioration* column. The data confirm that although AVSD-MOEA/D loses in some pair-wise tests, the overall numbers of wins and losses clearly favor AVSD-MOEA/D. More importantly, the total deterioration is much lower in the case of AVSD-MOEA/D, confirming that when AVSD-MOEA/D loses, the differences are low.

Table 4: Statistical Tests and Deterioration Level of the IGD+ for problems with three objectives

	↑	↓	↔	**Score**	**Deterioration**
AVSD-MOEA/D	69	5	2	64	0.005
MOEA/D-DE	35	34	7	1	0.774
NSGA-II	6	65	5	-59	1.260
NSGA-III	22	48	6	-26	0.844
R2-EMOA	46	26	4	20	0.656

Tables **??** and **??** shows the same information for the problems with three objectives. In this case, the number of times that each method belonged to the winning groups were 17, 2, 0, 0 and 0 for AVSD-MOEA/D, R2-EMOA, MOEA/D-DE, NSGA-III and NSGA-II, respectively. Thus, AVSD-MOEA/D yielded quite superior results. Considering the whole set of problems, AVSD-MOEA/D obtained a much larger mean HV ratio than the other ones. Moreover, the difference between the mean HV ratio obtained by the best method and by AVSD-MOEA/D was never greater than 0.1. However, all the other methods exhibited a deterioration in excess of 0.1 in several cases. In particular, this happened in 2, 2, 2 and 6 problems for MOEA/D-DE, R2-EMOA, NSGA-III and NSGA-II respectively. Remarkably, AVSD-MOEA/D is quite superior in both the total deterioration and in the score generated from the pair-wise statistical tests. In fact, its deterioration for the entire problem set is just 0.006. Beating all the state-of-the-art algorithms in such a large number of problem benchmarks is a quite significant achievement, and shows the robustness of AVSD-MOEA/D. Our results show that the superiority of AVSD-MOEA/D persists, and even increases, when problems with three objective functions are considered. For a better comprehension of the strenghts and weakness of the algorithms, in the Figure **??** is shown the 50% attainment surfaces for WFG8 and UF5. An attainment surface approximation can be interpreted as the spatial region that is statistically attained among all the runs that were carried out by an algorithm [2, 3]. In other words, it can be understood as *the spatial region that is achieved by the k% among all the runs by one algorithm*. The most challenging characteristic of these problems are that WFG8 has strong dependencies among all the parameters, and UF5 is a multi-modal biased problem whose Pareto optimal front is discrete and consists of 21 points. In both problems AVSD-MOEA/D was the only one that converged adequately to the Pareto front at least 50% among all the runs. Even more, given that the standard deviation is too low it can be though that all the runs converged similarly well.

We can better understand the reasons behind the benefits of AVSD-MOEA/D against the state-of-the-art MOEAs by analyzing the evolution of the HV values and the diversity. Note that in some MOPs, variables can be classified into two types: distance variables and position variables. A variable x_i is a distance variable when for all x, modifying x_i results in a new solution that dominates x, is equivalent to x, or is dominated by x. Differently, if x_i is a position variable, modifying x_i in x always results in a vector that is incomparable or equivalent to x [4]. This is important because in some cases, MOEAs do not maintain a large enough diversity in the distance variables [5], so analyzing the diversity trend for these kinds of variables provides an useful insight into the dynamics of the population.

In order to show the behavior of the different schemes, we selected WFG5 and UF5. They are complementary in the sense that in WFG5, all the Pareto solutions exhibit constant values for the distant variables, which is not the case in UF5. Moreover, in UF5, the optimal regions are isolated in the variable space, meaning that more diversity is required. For each algorithm, the diversity is calculated as the average Euclidean distance between individuals (ADI) in the population by considering only the distance variables. Figures **??** and **??** show the evolution of the ADI (top) and the mean of HV (bottom) for WFG5 and UF5, respectively. In the WFG5 problem, the distance variables quickly converged to a small region in state-of-the-art MOEAs. Thus, the differential evolution operator loses it exploring power and as a result, those MOEAs were unable to significantly improve the quality of the approximations as the evolution progresses. By contrast, in the case of AVSD-MOEA/D, the decrease in ADI is quite linear until the midpoint of the execution, and the increase in HV is gradual. The final HV attained by AVSD-MOEA/D is the largest one, which shows the important benefit of gradually decreasing the diversity.

As expected, explicitly promoting diversity is also beneficial for problems with disconnected optimal regions. As the data in Figure **??** show, the advantage of promoting diversity in the UF5 test problem is clear. In this case, state-of-the-art algorithms maintain some degree of diversity in the distance variables for the entire search. However, a large degree of diversity is required to obtain the 21 optimal solutions, and these MOEAs do not maintain the required amount of diversity, and as a result, they miss many of the solutions. In the case of AVSD-MOEA/D, enforcing a large degree of diversity in the initial phases promotes more exploration, which makes it possible to find additional optimal regions. Once these regions are located, they are not discarded, meaning that a larger level of diversity is maintained throughout the execution. This way, AVSD-MOEA/D not only attained better HV values for the first 10% of the total function evaluations, but it also kept looking for promising regions. In fact, its HV values improved significantly until the midpoint of the execution period i.e., the final moment when diversity was explicitly promoted. Then, an additional increase was obtained due to intensification in the regions identified. This analysis shows that the dynamic of the population depends on the problem at hand. The behavior of AVSD-MOEA/D with all the problems tested was similar to those already presented. Scenaries where the optimal regions consists of constant values for the distance variables behave like WFG5, whereas the behavior in those cases where the optimal regions consist of non-constant values for the distance variables is more similar to the UF5 case. Note, however, that in these cases, different levels of diversity are required, so the behavior is not as homogeneous.

In order to better understand the importance of D_I, the entire set of benchmark problems was tested with different values of D_I. As in previous experiments, the stopping criterion was set to 2.5×10^7 function evaluations. Since normalized distances are used, the maximum attainable distance between pairs of individuals is 1.0. Also note that setting D_I to 0 implies not promoting diversity in the variable space. Thus, several values in this range were considered. Specifically, the values $D_I = \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$ were tested. Figure **??** shows the mean HV ratio obtained for both the two-objective and the three-objective case with the D_I values tested. The AVSD-MOEA/D performed worst when D_I was set to 0. The HV ratio quickly increased as higher D_I values up to 0.2 were used. Larger values yielded quite similar performances. Thus, a wide range of values (from 0.2 to 1.0) exhibited very good performance, meaning that the behavior of AVSD-MOEA/D is quite robust. Thus, properly setting this parameter is not a complex task.

In order to better understand the implications of D_I on the dynamics of the population, Figure **??** shows, for AVSD-MOEA/D, the evolution of diversity in the distance variables in the WFG9 case for three different values of D_I. When setting $D_I = 0$, the diversity is reduced quite quickly, which results in premature convergence. The result is a hypervolume that is not too high. However, when $D_I = 0.4$ and $D_I = 1$ are used, the loss of diversity is slowed down, and the resulting hypervolume is quite large. Note that setting $D_I = 1$ promotes greater diversity, so the hypervolume increases slower than when $D_I = 0.4$. However, the degree of exploration in both cases is enough to yield high-quality solutions. The behavior is quite similar in every problem, which explains the stability of the algorithms for different values of D_I. Note that for shorter periods, setting a proper D_I value is probably much more important. However, for long-term executions at least, practically any value higher than 0.2 yields similar solutions, which we regard as a highly positive feature.

2. On the Convergence of MOEAs in Test Problems with Bias Features

As pointed out in [6, 7, 4], the bias feature is one of the most challenging difficulties that MOEAs might face. Recently, the BTs test problems were proposed to facilitate the study of the ability of MOEAs for dealing with biases. In this context bias means that small variations in the decision space around

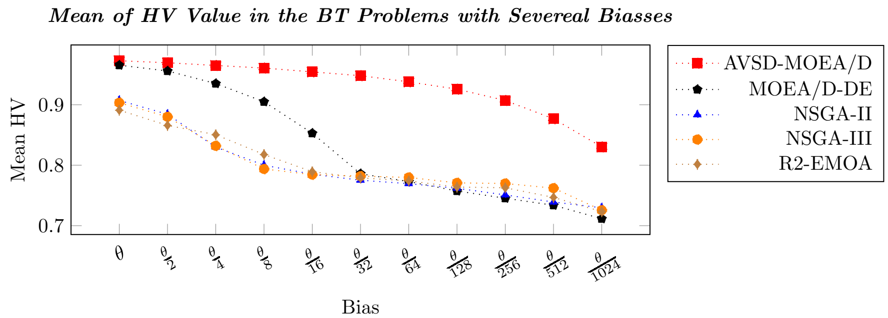

Figure 1: Mean of HV values for eight BTs problems (y-axis) against several biasses ratios (x-axis). The BT2 problem is not taken into consideration due that it suffers of numerical stability.

the Pareto set cause significant changes in vicinities of some Pareto front solutions [4]. Particularly, those problems are built with transformations that induce position-related bias and distance-related bias. While the former means that a small change on the position-related variables of one solution in the Pareto set projects a significant change along the Pareto front. The later imposes that a small variation on the distance-related variables of one solution in the Pareto set causes a significant deterioration on the convergence towards the Pareto front.

In order, to analyze the capability of the MOEAs to deal with bias features the BTs problems are taken into account. Specifically, this section analyses the sensitivity of the algorithms imposing several levels of bias in the distance-related variables. Initially, for each problem the position-related bias and distance-related bias (θ) are kept exactly as the one proposed in the original work [6]. Then, for each problem its initial distance-related bias value (θ) is iterativelly decreased by a factor of two. Specifically, the distance-related bias taken into account are $\{\theta, \frac{\theta}{2}, \frac{\theta}{4}, \frac{\theta}{8}, \frac{\theta}{16}, \frac{\theta}{32}, \frac{\theta}{64}, \frac{\theta}{128}, \frac{\theta}{256}, \frac{\theta}{512}, \frac{\theta}{1028}\}$. Figure 1 shows the mean HV ratio obtained with several distance-related biasses. Also note that the BT2 problem is not taken into consideration due that increasing its bias values provokes numerical instability since that it incorporates a different bias transformation, nevertheless all the results can be consulted in the supplementary document. Taking exactly the original configuration (bias of θ) [6] AVSD-MOEA/D is sigthly better than MOEA/D-DE, but as soon as the bias is decreased to $\frac{\theta}{32}$ the performance of MOEA/D-DE decays aggressively. Furthermore, the performance of AVSD-MOEA/D is superior than 0.9 with biasses values upper or equal to $\frac{\theta}{256}$ which is quite superior than the state-of-the-art MOEAs whose values at that point are approximately of 0.75. Figure **??** shows the 50% of attainment surface of BT6, BT7 and BT8 with a bias of $\frac{\theta}{32}$. BT6 and BT8 have simple nolinear Pareto set while BT7 has a complicated nolinear Pareto set. BT8 is multimodal. Although that MOEA/D-DE converged to a region of the Pareto front with BT6 AVSD-MOEA/D covered a huge region of the Pareto front, in fact this shows that for this problem promoting diversity in the decision space results in diversity in the objective space. In addition, AVSD-MOEA/D converges quite well in complicates nonlinear Pareto sets shown in the 50% attained surface of BT7 (Figure **??**). Finally but not less important AVSD-MOEA/D shows a superior behaviour with biased and multimodal problems as is the case of BT8 whose attainment surfaces have converged much better to the Pareto front.

References

[1] H. Ishibuchi, H. Masuda, Y. Tanigaki, Y. Nojima, Modified distance calculation in generational distance and inverted generational distance,

in: International conference on evolutionary multi-criterion optimization, Springer, 2015, pp. 110–125.

[2] J. Knowles, A summary-attainment-surface plotting method for visualizing the performance of stochastic multiobjective optimizers, in: 5th International Conference on Intelligent Systems Design and Applications (ISDA'05), IEEE, 2005, pp. 552–557.

[3] C. M. Fonseca, P. J. Fleming, On the performance assessment and comparison of stochastic multiobjective optimizers, in: International Conference on Parallel Problem Solving from Nature, Springer, 1996, pp. 584–593.

[4] S. Huband, P. Hingston, L. Barone, L. While, A review of multiobjective test problems and a scalable test problem toolkit, IEEE Transactions on Evolutionary Computation 10 (2006) 477–506.

[5] J. C. Castillo, C. Segura, A. H. Aguirre, G. Miranda, C. León, A multi-objective decomposition-based evolutionary algorithm with enhanced variable space diversity control, in: Proceedings of the Genetic and Evolutionary Computation Conference Companion, ACM, 2017, pp. 1565–1571.

[6] H. Li, Q. Zhang, J. Deng, Biased multiobjective optimization and decomposition algorithm, IEEE transactions on cybernetics 47 (2016) 52–66.

[7] K. Deb, Multi-objective genetic algorithms: Problem difficulties and construction of test problems, Evolutionary computation 7 (1999) 205–230.