

Supplementary File of “R-Metric: Evaluating the Performance of Preference-Based Evolutionary Multi-Objective Optimization Using Reference Points”

Ke Li, *Member, IEEE*, Kalyanmoy Deb, *Fellow, IEEE*, and Xin Yao, *Fellow, IEEE*

I. MAJOR DIFFERENCES WITH [1]

Our proposed R-metric is a comprehensive extension and generalization of the metric developed by the second author and his collaborators [1]. The major differences are summarized as follows:

- [1] merely adapts the regular HV to evaluate the performance of reference-based EMO algorithms, while the proposed R-metric provides a general framework to adapt the existing performance metrics (we use the IGD and HV as the baseline) to assess and compare preference-based EMO algorithms using reference points.
- [1] does not have the prescreening procedure. Therefore, dominated solutions are able to participate in the metric calculation. This may lead to some noise in performance assessment.
- [1] uses a circle/sphere (or hyper-sphere) to trim overly spread solutions, while the proposed R-metric uses a rectangle/cube (or hyper-cube) to serve this purpose.
- [1] uses the centroid as the pivot point, while the proposed R-metric chooses the solution having the best ASF value as the pivot point. This is because the centroid may cause ambiguities in comparison. Let us consider the example discussed in Section IV-B3 and Fig. 5 of our manuscript again. As shown in Fig. 1, now we use the centroid as the pivot point, and the trimming is replaced by a circle. In addition, we change the reference point to the solution in S^1 that has the lowest f_2 value. Obviously, S^1 is preferable than S^2 as it indeed find the solution most preferred by the decision maker, whereas all solutions of S^2 are away from the region of interest. But the R-metric will conclude S^2 is better if we use the centroid as the pivot point.

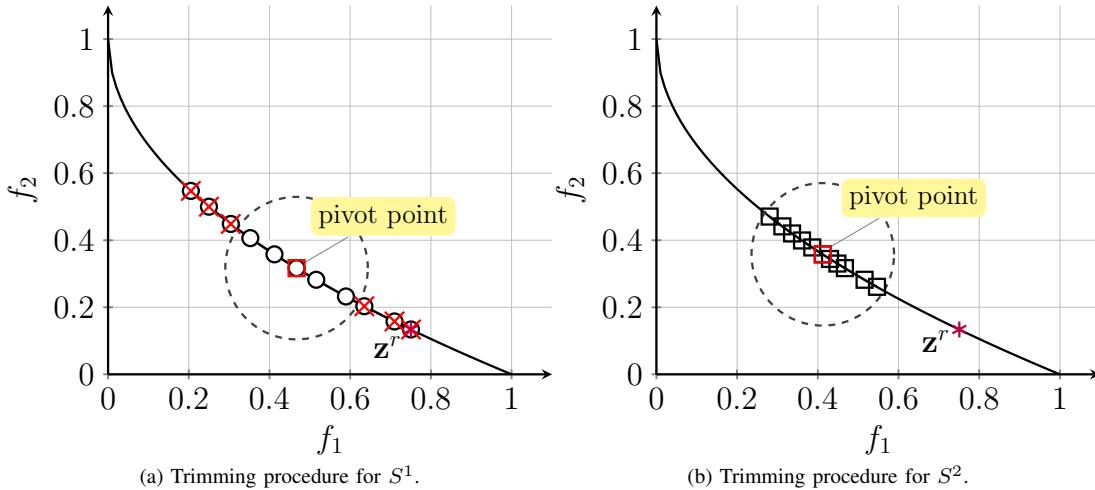


Fig. 1: Ambiguity caused by using the centroid as the pivot point.

- [1] does not provide methods for handling disconnected Pareto front and multiple reference points.

K. Li is with the College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, EX4 4QF, UK (e-mail: k.li@exeter.ac.uk)
K. Deb is with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824, USA (e-mail: kdeb@egr.msu.edu)
X. Yao is with the School of Computer Science, University of Birmingham, Birmingham, B15 2TT, UK (e-mail: x.yao@cs.bham.ac.uk)

II. INVESTIGATIONS OF IMPACTS OF DIFFERENT \mathbf{w} SETTINGS

To investigate the impacts of different \mathbf{w} settings on the R-metric assessment, we design three different solution sets (i.e., S^1 , S^2 and S^3) along an example PF (i.e., $f_2 = 1 - f_1$) as shown in Fig. 2. In particular, all solution sets have the same cardinality and spread. Their differences lie in their positions with respect to the reference point $\mathbf{z}^r = (0.45, 0.45)^T$. The corresponding R-HV values are given in Table I. Specifically, all objectives are equally important when \mathbf{w} is set to $(0.5, 0.5)^T$. In this case, S^1 shall be the best solution set as \mathbf{z}^r lies in the middle part of S^1 while the other two solution sets are deviated from \mathbf{z}^r . When \mathbf{w} is set to $(0.1, 0.9)^T$, the first objective is assumed to be more important than the second one. In this case, S^3 , which lies above \mathbf{z}^r , is the best solution set since its solutions have better values on the first objective. In contrast, the second objective is assumed to be more important when \mathbf{w} is set to $(0.9, 0.1)^T$. Accordingly, S^2 , which lies on the lower part of \mathbf{z}^r , becomes the best one as the second objective is better optimized. All in all, \mathbf{w} controls the preference of different objectives when setting \mathbf{z}^w , thus it influences the R-HV calculation.

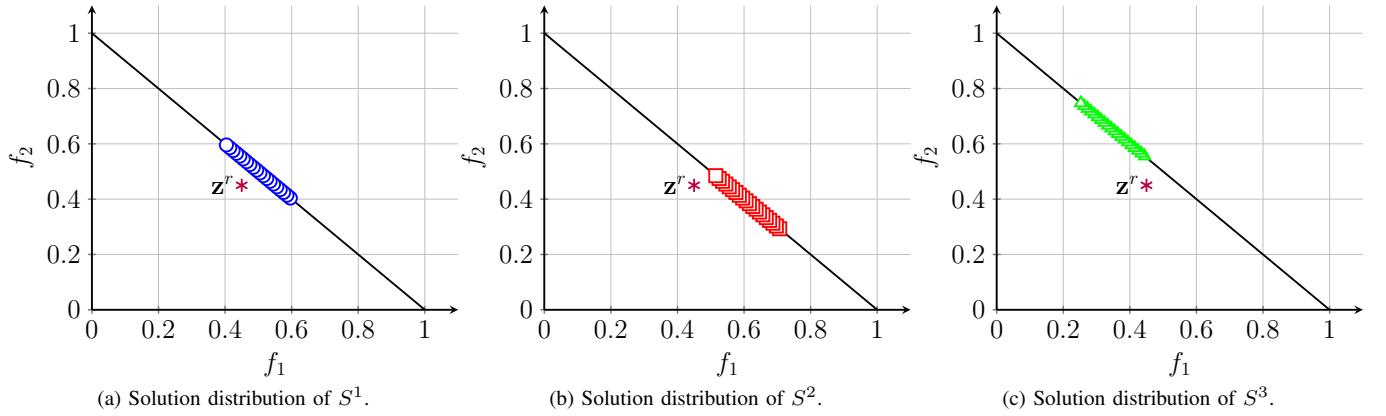


Fig. 2: Three different solution sets with respect to the same reference point.

TABLE I: Comparisons of R-HV values obtained on three different \mathbf{w} settings.

Weight	S^1	S^2	S^3
$\mathbf{w}^1 = (0.5, 0.5)^T$	1.0641	1.0340	0.9523
$\mathbf{w}^2 = (0.1, 0.9)^T$	0.4235	0.1721	0.6248
$\mathbf{w}^3 = (0.9, 0.1)^T$	0.4235	0.5896	0.0843

III. DISCUSSION OF SCALABILITY OF R-METRIC

In this section, we design two experiments to investigate the “scalability” of R-metric. Note that the “scalability” we studied here is about the number of reference points supplied by the DM.

- First, we increase the number of reference points to five, as shown in Fig. 3(a), where $\mathbf{z}^{r_1} = (0.1, 0.6)^T$, $\mathbf{z}^{r_2} = (0.2, 0.5)^T$, $\mathbf{z}^{r_3} = (0.45, 0.3)^T$, $\mathbf{z}^{r_4} = (0.6, 0.3)^T$ and $\mathbf{z}^{r_5} = (0.75, 0.1)^T$. Five point set combinations, i.e., $(S^2, S^3, S^5, S^6, S^8)$, $(S^1, S^2, S^3, S^4, S^5)$, $(S^4, S^5, S^6, S^7, S^8)$, $(S^3, S^4, S^8, S^9, S^{10})$ and $(S^1, S^4, S^6, S^7, S^{10})$ are chosen as the candidates for performance assessment. From the results shown in Fig. 3(b), we find that $(S^2, S^3, S^5, S^6, S^8)$ obtains the best R-IGD and R-HV values. From Fig. 3(a), we can see that S^2, S^3, S^5, S^6 and S^8 are in the corresponding ROI of each reference point, respectively. In contrast, since the other point set combinations are more or less away from the corresponding reference points, their R-IGD and R-HV values are worse than $(S^2, S^3, S^5, S^6, S^8)$. Note that the differences of R-metric values obtained by different point set combinations are not as significant as the example discussed in our manuscript where we only consider two reference points. The possible reason is the increase of the number of reference points which mitigates the difficulty of finding the ROIs.

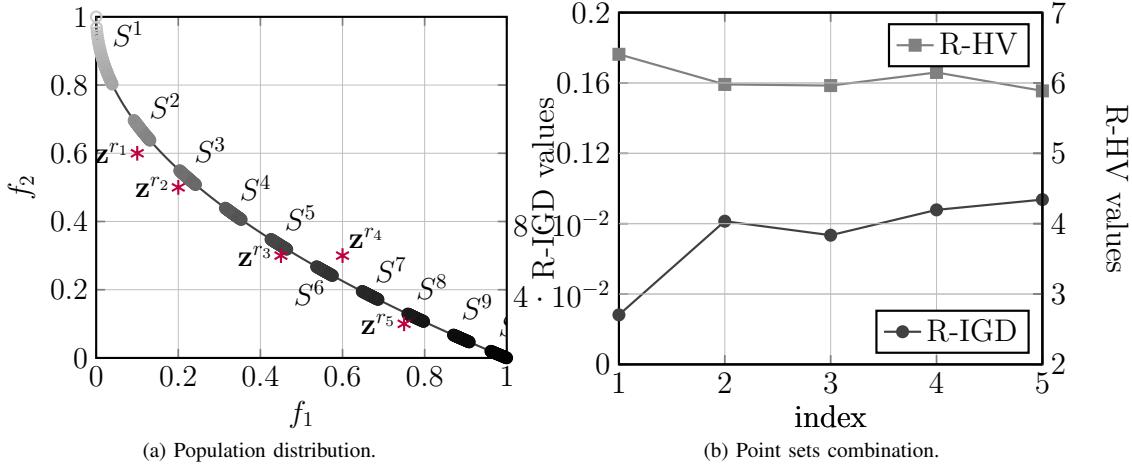


Fig. 3: Variations of R-IGD and R-HV with respect to five different reference points.

- As for the second experiment, we set the number of reference points to be the same as the population size. In particular, the reference points are the same as the weight vectors used in a decomposition-based EMO algorithm like MOEA/D. Here we use standard MOEA/D and NSGA-II as the baseline algorithms. For proof of principle purpose, we just choose ZDT1 as the test problem. Table II shows the comparisons of the mean and standard deviation of HV, R-HV, IGD and R-IGD obtained by MOEA/D and NSGA-II across 31 runs. From these results, we can see that R-HV and R-IGD have the same conclusion as the IGD and HV (the performance of MOEA/D is better than NSGA-II). In principle, we can treat the DM supplied reference points as a general approximation of the PF. Then, the R-metric can be used to assess the performance of the solution set obtained by an EMO algorithm for approximating each ROI.

TABLE II: Comparisons of IGD, HV, R-IGD and R-HV obtained by MOEA/D and NSGA-II on ZDT1.

Metric	MOEA/D	NSGA-II
IGD	4.325E-3(2.79E-5)	4.671E-3(4.11E-5)
HV	3.6601(3.35E-3)	3.6521(2.98E-3)
R-IGD	6.691E-3(5.14E-5)	6.722E-3(6.01E-5)
R-HV	8.6530(4.02E-3)	8.6402(3.41E-3)

IV. EMPIRICAL STUDIES ON ARTIFICIAL SCENARIOS

This section presents the empirical results on two artificial scenarios which use ZDT2 [2] and DTLZ2 [3] as the benchmark. The observations are similar to the ZDT1 and DTLZ1 cases discussed in the main manuscript. In particular, the point set closest to the ROI, i.e., DM supplied reference point, obtains the best R-metric values; while the R-metric values depend on the distance to the ROI.

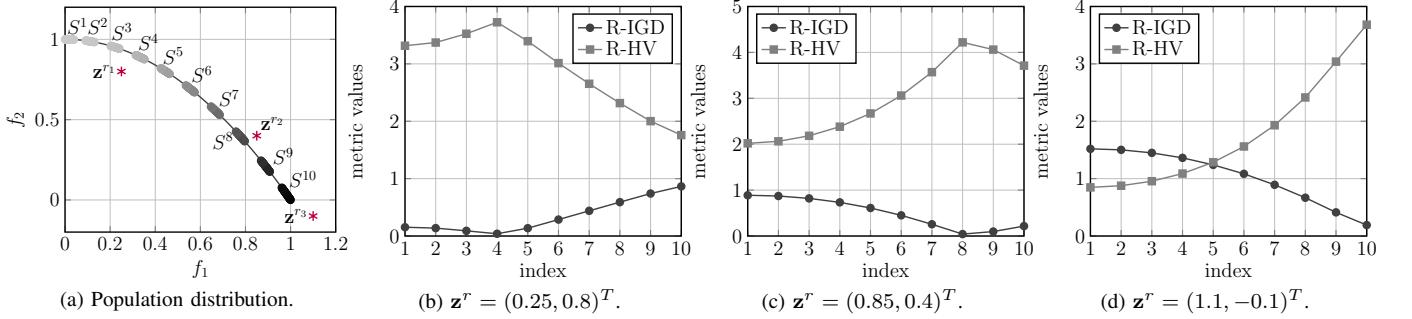


Fig. 4: Variations of R-IGD and R-HV with respect to an unattainable reference point $\mathbf{z}^r = (0.25, 0.8)^T$, an attainable reference point $\mathbf{z}^r = (0.85, 0.4)^T$ and a outside reference point $\mathbf{z}^r = (1.1, -0.1)^T$ on ZDT2 problem.

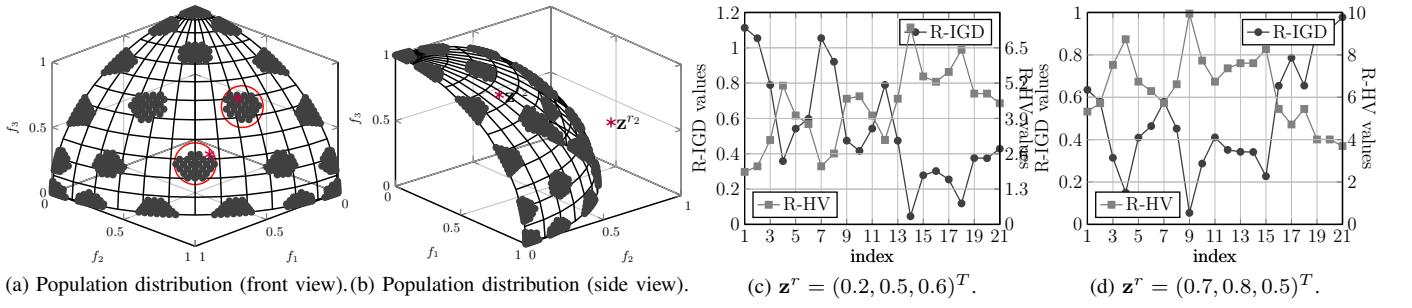


Fig. 5: Variations of R-IGD and R-HV with respect to an unattainable reference point $\mathbf{z}^r = (0.2, 0.5, 0.6)^T$, an attainable reference point $\mathbf{z}^r = (0.7, 0.8, 0.5)^T$ on DTLZ2 problem.

V. PARAMETER SETTINGS

All multi-objective optimizers use simulated binary crossover (SBX) [4] and polynomial mutation [5] as the reproduction operators. For SBX, the crossover probability is set as $p_c = 1.0$ and distribution index is set as $\eta_c = 10$; for polynomial mutation, the mutation probability is set as $p_m = \frac{1}{n}$ and distribution index is set as $\eta_m = 20$. The population size is set to 100 for all algorithms, and the stopping criterion is the number of function evaluations. Specifically, it is set as 40,000 for two-objective problems, and 50,000 for three-objective ones except DTLZ3, where 100,000 function evaluations are used due to its difficulties. Table III gives the general parameter settings for the four preference-based EMO algorithms according to their original references [6]–[9]. Table IV gives the reference point settings for different test problems.

TABLE III: Parameter Settings of EMO Algorithms.

Algorithm	Parameter Settings
r-MOEA/D-STM	neighborhood size $T = 20$, probability to select in neighborhood $\delta = 0.9$
R-NSGA-II	clearing parameter $\epsilon = 0.001$ for bi-objective problems and $\epsilon = 0.002$ for tri-objective cases
r-NSGA-II	non-r-domination parameter $\delta = 0.2$

TABLE IV: The Settings of Reference Points.

Problem	Unattainable Reference Point \mathbf{z}^{r_1}	Attainable Reference Point \mathbf{z}^{r_2}
ZDT1 and ZDT4	$(0.3, 0.4)^T$	$(0.65, 0.3)^T$
ZDT2	$(0.2, 0.8)^T$	$(0.9, 0.4)^T$
ZDT3	$(0.15, 0.4)^T$	$(0.4, 0.0)^T$
ZDT6	$(0.9, 0.3)^T$	$(0.5, 0.7)^T$
DTLZ1	$(0.05, 0.05, 0.2)^T$	$(0.3, 0.3, 0.2)^T$
DTLZ2 to DTLZ 4	$(0.2, 0.5, 0.6)^T$	$(0.7, 0.8, 0.5)^T$
DTLZ5 to DTLZ 6	$(0.1, 0.3, 0.5)^T$	$(0.6, 0.7, 0.5)^T$
DTLZ7	$(0.165, 0.71, 4.678)^T$	$(0.75, 0.15, 6.0)^T$

VI. PLOTS OF FINAL POPULATIONS ON ZDT AND DTLZ PROBLEMS

This section presents the plots of the final populations with the lowest R-IGD metric values found by four preference-based EMO algorithms on all ZDT and DTLZ problems. In particular, the blue circle indicates the solutions for the unattainable reference point \mathbf{z}^{r_1} ; while the red circle represents the solutions for the attainable reference point \mathbf{z}^{r_2} .

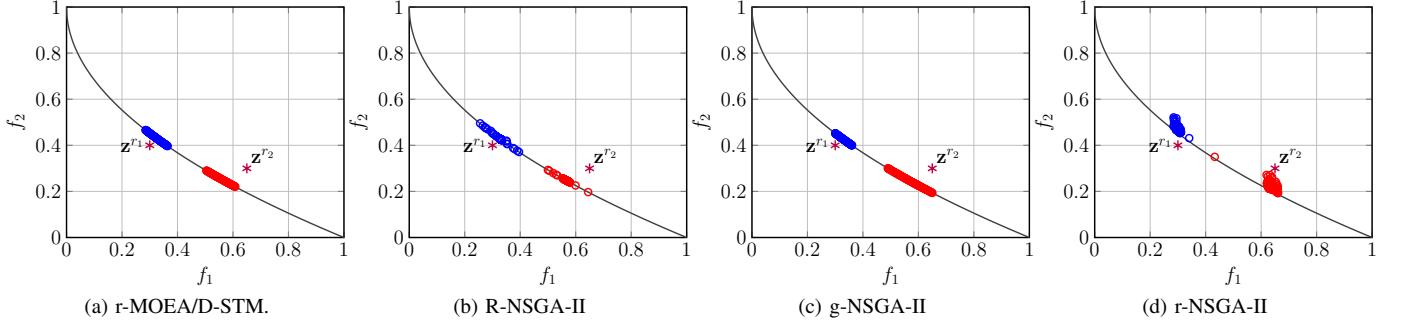


Fig. 6: Plots of the final populations with the lowest R-IGD metric values found by four preference-based EMO algorithms on ZDT1.

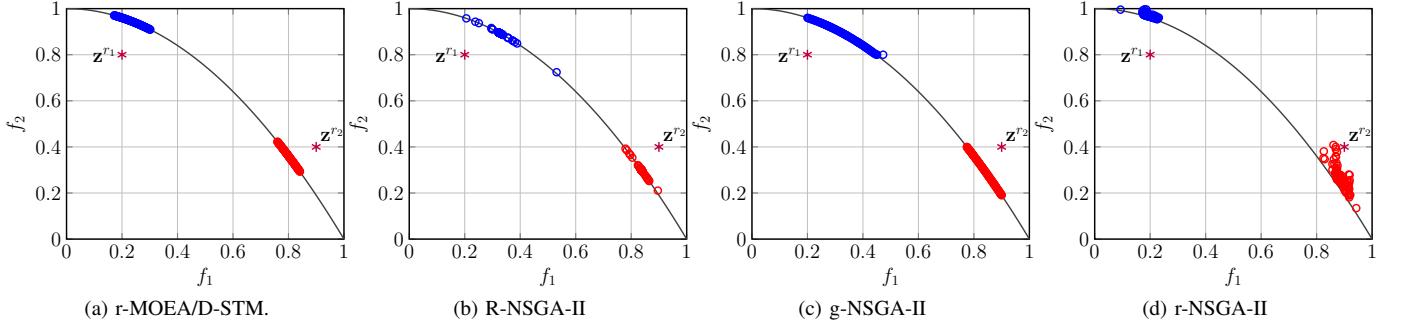


Fig. 7: Plots of the final populations with the lowest R-IGD metric values found by four preference-based EMO algorithms on ZDT2.

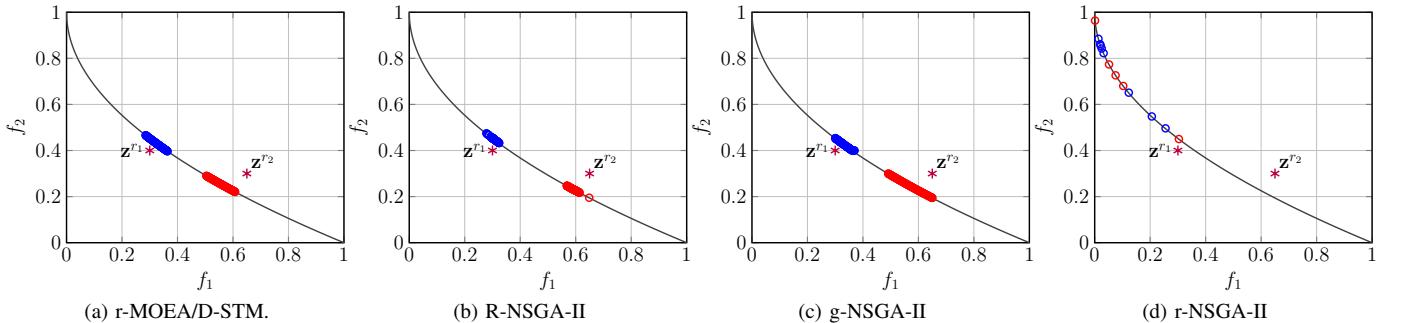


Fig. 8: Plots of the final populations with the lowest R-IGD metric values found by four preference-based EMO algorithms on ZDT4.

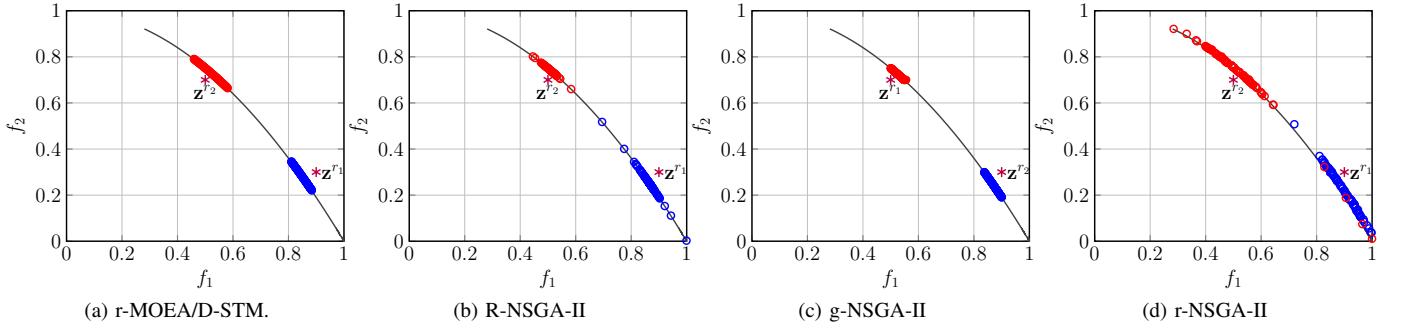


Fig. 9: Plots of the final populations with the lowest R-IGD metric values found by four preference-based EMO algorithms on ZDT6.

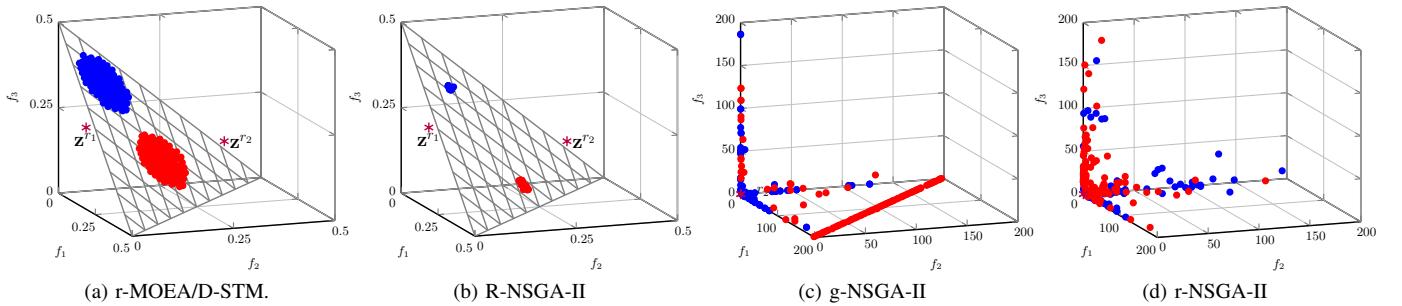


Fig. 10: Plots of the final populations with the lowest R-IGD metric values found by four preference-based EMO algorithms on DTLZ1.

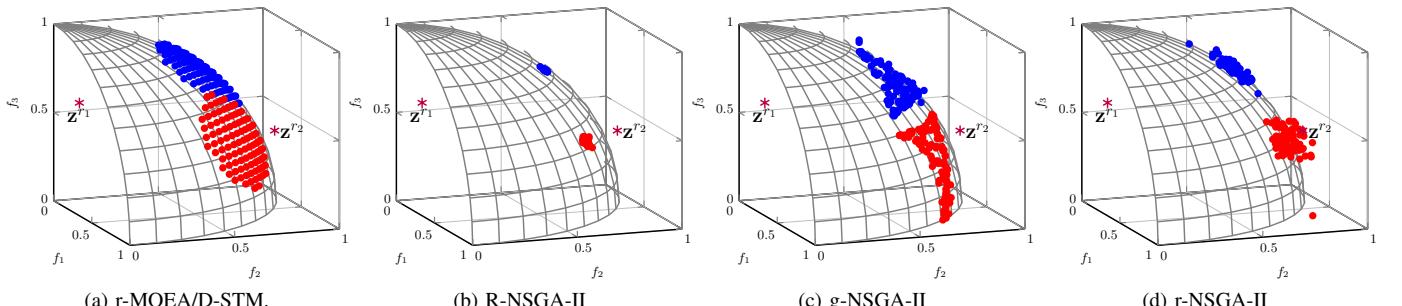


Fig. 11: Plots of the final populations with the lowest R-IGD metric values found by four preference-based EMO algorithms on DTLZ2.

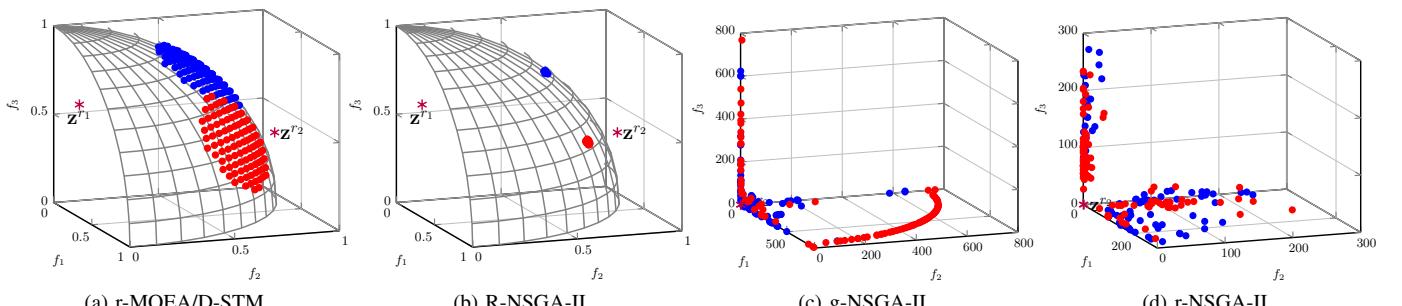


Fig. 12: Plots of the final populations with the lowest R-IGD metric values found by four preference-based EMO algorithms on DTLZ3.

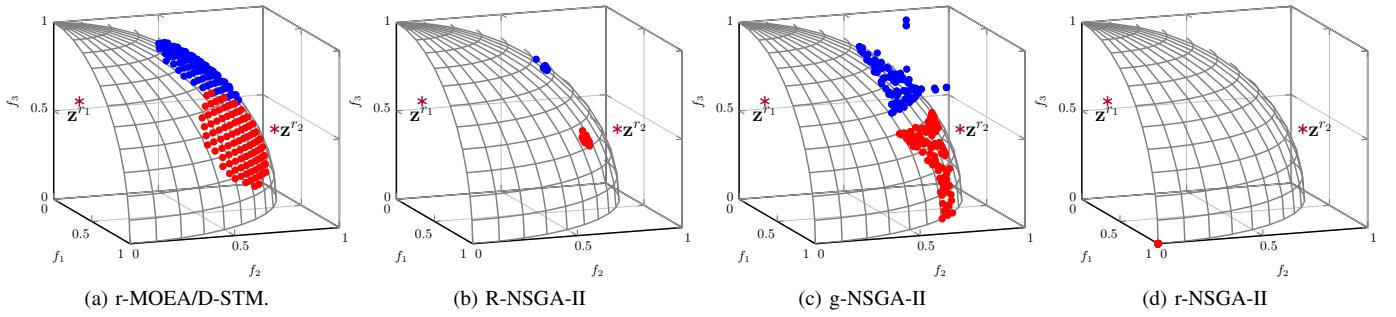


Fig. 13: Plots of the final populations with the lowest R-IGD metric values found by four preference-based EMO algorithms on DTLZ4.

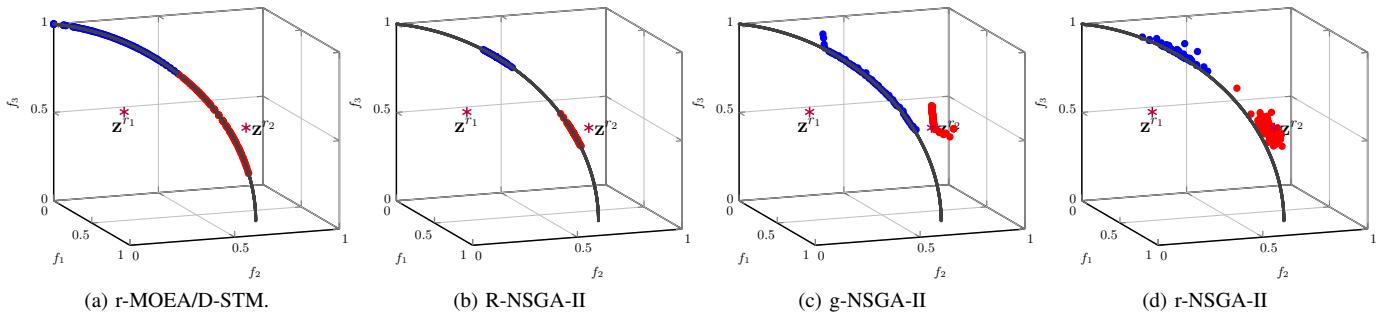


Fig. 14: Plots of the final populations with the lowest R-IGD metric values found by four preference-based EMO algorithms on DTLZ5.

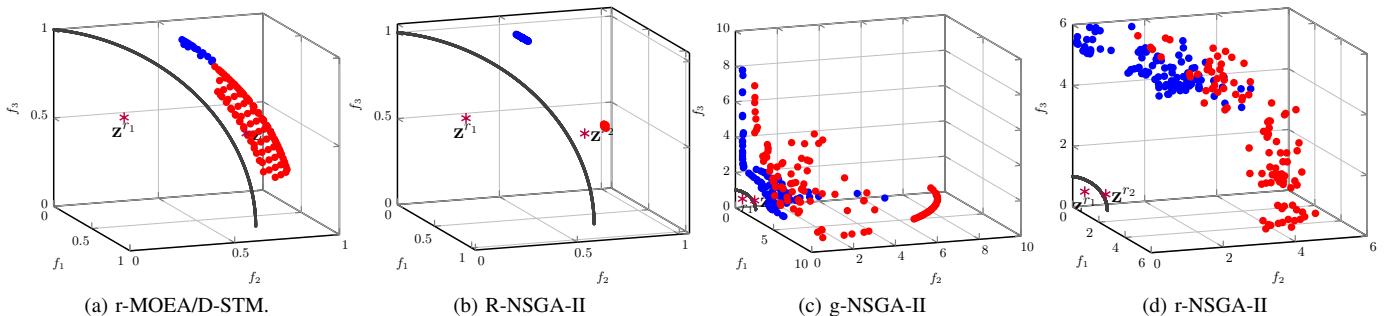


Fig. 15: Plots of the final populations with the lowest R-IGD metric values found by four preference-based EMO algorithms on DTLZ6.

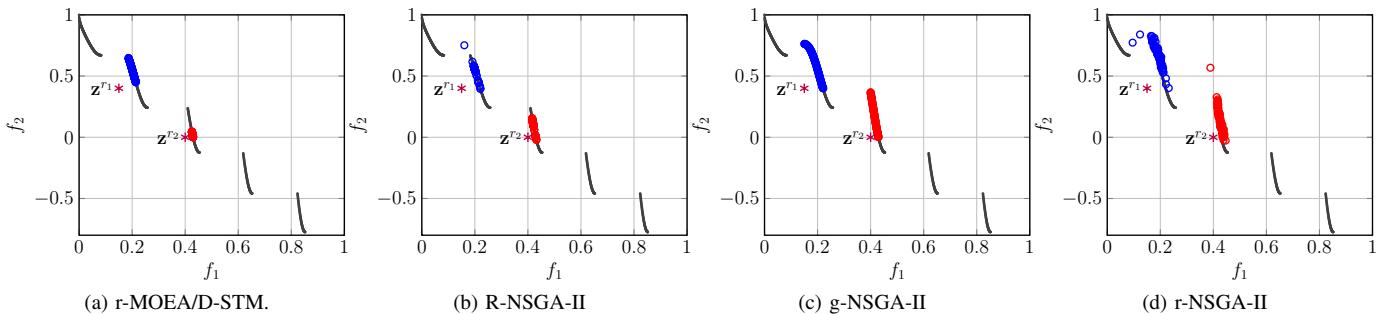


Fig. 16: Plots of the final populations with the lowest R-IGD metric values found by four preference-based EMO algorithms on ZDT3.

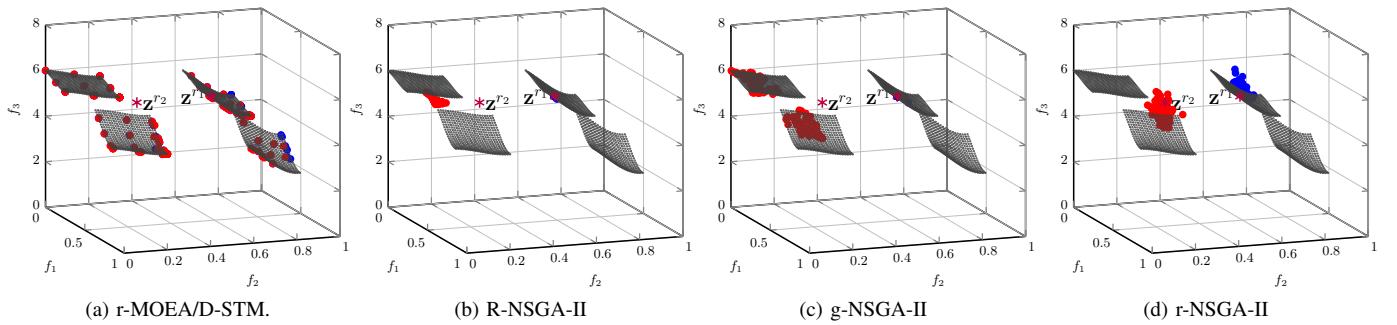


Fig. 17: Plots of the final populations with the lowest R-IGD metric values found by four preference-based EMO algorithms on DTLZ7.

VII. PLOTS FOR MANY-OBJECTIVE PROBLEMS

This section presents the plots of the final populations with the lowest R-IGD metric values found by four preference-based EMO algorithms on 5- and 10-objective DTLZ2. The reference point is highlighted as the red dotted line.

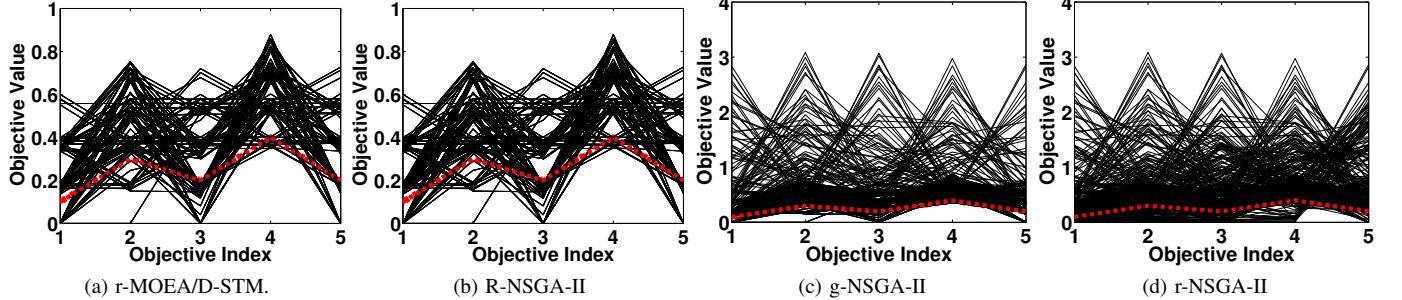


Fig. 18: Plots of the final populations with the lowest R-IGD metric values found by four preference-based EMO algorithms on five objective DTLZ2 with regard to the reference point $\mathbf{z}^r = (0.1, 0.3, 0.2, 0.4, 0.2)^T$.

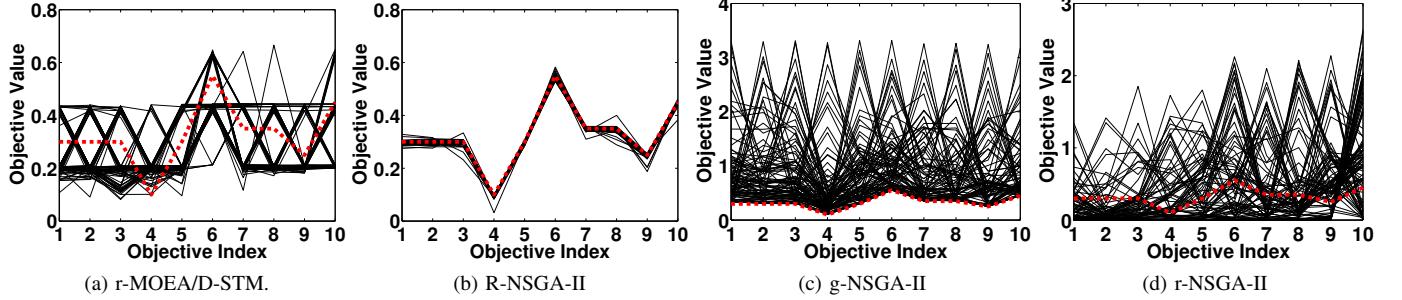


Fig. 19: Plots of the final populations with the lowest R-IGD metric values found by four preference-based EMO algorithms on five objective DTLZ2 with regard to the reference point $\mathbf{z}^r = (0.3, 0.3, 0.3, 0.1, 0.3, 0.55, 0.35, 0.35, 0.25, 0.45)^T$.

VIII. TRAJECTORIES OF R-METRIC VALUES

This section presents the trajectories of R-metric values over different generations. In addition, Table V gives the results of using R-metrics as a stopping criterion.

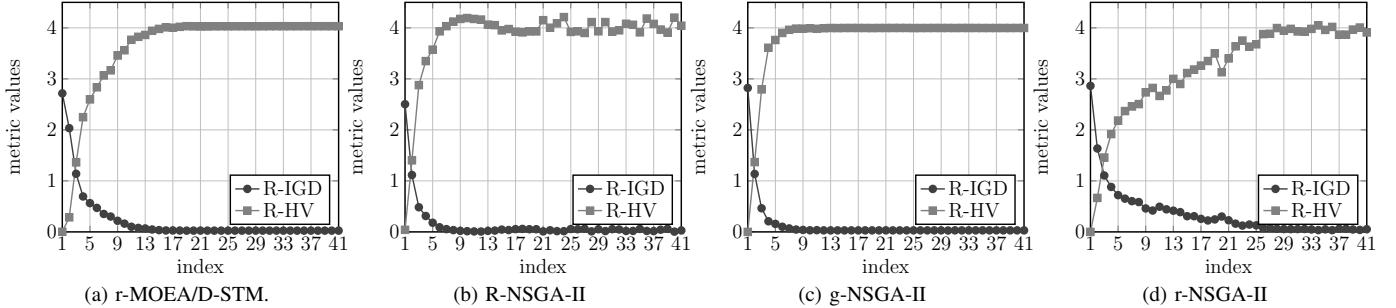


Fig. 20: Trajectories of R-IGD and R-HV values versus the number of generations on ZDT1 with reference point $\mathbf{z}^r = (0.3, 0.4)^T$.

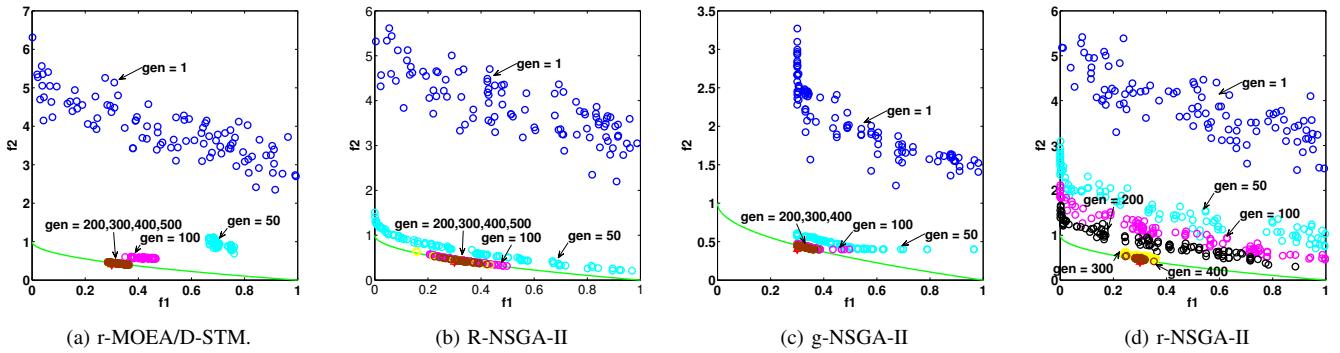


Fig. 21: Plots of all 40 populations on ZDT1 with reference point $\mathbf{z}^r = (0.3, 0.4)^T$.

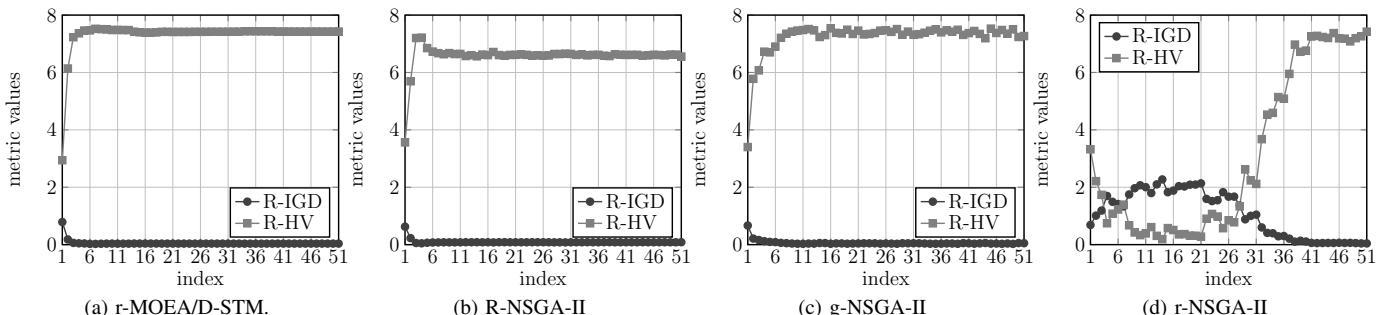


Fig. 22: Trajectories of R-IGD and R-HV values versus the number of generations on DTLZ2 with reference point $\mathbf{z}^r = (0.2, 0.5, 0.6)^T$.

REFERENCES

- [1] K. Deb, F. Siegmund, and A. H. C. Ng, "R-HV: A metric for computing hyper-volume for reference point based EMOs," in *SEMCOCO'14: Proc. of the 5th International Conference on Swarm, Evolutionary, and Memetic Computing*, 2014, pp. 98–110.
- [2] E. Zitzler, K. Deb, and L. Thiele, "Comparison of multiobjective evolutionary algorithms: Empirical results," *Evolutionary Computation*, vol. 8, no. 2, pp. 173–195, 2000.
- [3] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable test problems for evolutionary multiobjective optimization," in *Evolutionary Multiobjective Optimization*, ser. Advanced Information and Knowledge Processing, A. Abraham, L. Jain, and R. Goldberg, Eds. Springer London, 2005, pp. 105–145.
- [4] K. Deb and R. B. Agrawal, "Simulated binary crossover for continuous search space," *Complex Systems*, vol. 9, pp. 1–34, 1994.
- [5] K. Deb and M. Goyal, "A combined genetic adaptive search (GeneAS) for engineering design," *Computer Science and Informatics*, vol. 26, pp. 30–45, 1996.

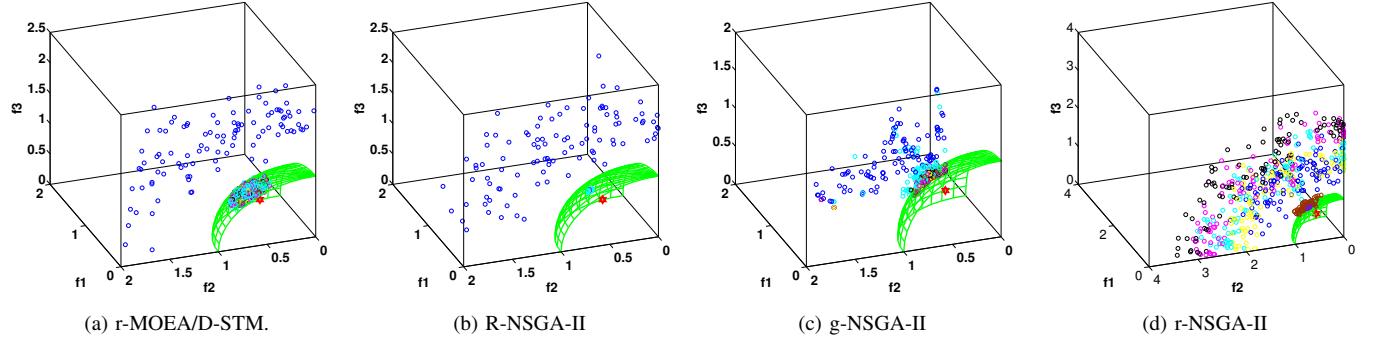
Fig. 23: Plots of all 50 populations on DTLZ2 with reference point $\mathbf{z}^r = (0.2, 0.5, 0.6)^T$.

TABLE V: Standard Deviations of R-metrics for Every 10 and 25 Consecutive Generations.

Gap = 10		r-MOEA/D-STM		R-NSGA-II		g-NSGA-II		r-NSGA-II	
Test Instance	R-metric	< 0.1	< 0.01	< 0.1	< 0.01	< 0.1	< 0.01	< 0.1	< 0.01
ZDT1	R-IGD	35	126	39	73	31	89	301	345
	R-HV	54	174	56	—	44	92	322	341
DTLZ2	R-IGD	19	54	17	364	17	91	470	476
	R-HV	37	90	—	—	—	—	489	—
Gap = 25		r-MOEA/D-STM		R-NSGA-II		g-NSGA-II		r-NSGA-II	
Test Instance	R-metric	< 0.1	< 0.01	< 0.1	< 0.01	< 0.1	< 0.01	< 0.1	< 0.01
ZDT1	R-IGD	63	157	60	89	48	105	348	352
	R-HV	128	188	71	—	76	116	—	—
DTLZ2	R-IGD	33	64	44	—	44	—	—	—
	R-HV	64	104	—	—	100	—	—	—

— indicates that the stopping criterion is not satisfied until the maximum number of generations.

- [6] K. Li, Q. Zhang, S. Kwong, M. Li, and R. Wang, "Stable matching based selection in evolutionary multiobjective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 6, pp. 909–923, 2014.
- [7] K. Deb, J. Sundar, U. Bhaskara, and S. Chaudhuri, "Reference point based multiobjective optimization using evolutionary algorithms," *International Journal of Computational Intelligence Research*, vol. 2, no. 3, pp. 273–286, 2006.
- [8] J. Molina, L. V. Santana, A. G. Hernández-Díaz, C. A. C. Coello, and R. Caballero, "g-dominance: Reference point based dominance for multiobjective metaheuristics," *European Journal of Operational Research*, vol. 197, pp. 685–692, 2009.
- [9] L. B. Said, S. Bechikh, and K. Ghédira, "The r-dominance: A new dominance relation for interactive evolutionary multicriteria decision making," *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 5, pp. 801–818, 2010.