Performance of DEMO on New Test Problems: A Comparison Study

Tea Robič
Department of Intelligent Systems
Jožef Stefan Institute
Jamova 39, SI-1000 Ljubljana, Slovenia
tea.robic@ijs.si

Abstract

DEMO (Differential Evolution for Multiobjective Optimization) is a new algorithm for solving multiobjective problems. This paper presents a comparison between DEMO and three state-of-the-art algorithms for multiobjective optimization using a new comparison methodology and new test problems. Results show that DEMO outperforms the other algorithms used in comparison in four out of nine test problems. On only one test problem the other algorithms achieve significantly better results than DEMO, while in the remaining four test problems DEMO is the second-best algorithm.

1 Introduction

In the last decade, the ongoing research on evolutionary multiobjective optimization produced several successful multiobjective evolutionary algorithms (MOEAs). With growing number of new algorithms, the need to properly evaluate and compare their performance yielded many test problems and quality indicators. Unfortunately, the first proposed test problems were often poorly designed and did not test a wide range of characteristics. Similarly, many frequently used quality indicators were incapable of indicating whether a nondominated set of solution is better than another and were consecutively found to be unsuitable for evaluating the performance of MOEAs [10].

Only recently, researchers presented explicit directions for performance assessment of MOEAs [4]. At the same time, a new scalable test problem toolkit was constructed, which allows the user to choose arbitrary levels of complexity [6].

DEMO – a new MOEA based on differential evolution – was introduced in [7], where it outperformed state-of-the-art algorithms on five test problems. In the comparison, 'old' test problems and quality indicators were used. Following the suggestions from [4] and using the test problems from [6] this paper presents a new, thorough comparison between DEMO and state-of-the-art MOEAs.

2 DEMO

DEMO is a steady-state algorithm for multiobjective optimization, which uses differential evolution for constructing new individuals. Therefore it can only be used on problems, where an individual is encoded as a real vector. When a new individual, also called a candidate solution, is constructed, it is compared to its parent. If the candidate solution dominates the parent, it replaces the parent in the current population. If the parent dominates the candidate, the candidate is discarded. Otherwise, if the candidate and its parent are incomparable, the candidate is added to the population. After constructing candidates for each parent individual in the population, the population has possibly increased. In this case, it is truncated to the original size using nondominated sorting and crowding distance metric (as in NSGA-II). This steps are repeated until a stopping criterion is met.

DEMO is a very simple algorithm and was presented in [7] in three different variants. Throughout this paper, only the elementary variant, called DEMO/parent, is used.

3 Comparison Methodology

This paper follows the comparison methodology suggested in [4], which consists of the following steps:

1. Run each algorithm several times.
2. Compare the attainment surfaces of the applied algorithms.
3. Evaluate the algorithms using unary indicators.

3.1 Attainment Surfaces

An attainment surface is 'the family of tightest goals known to be attainable as a result of the opti-

problem	separability	modality	bias	geometry
WFG1	separable	uni	polynomial, flat	convex, mixed
WFG2	non-separable	f_1 uni, f_2 multi	no bias	convex, disconnected
WFG3	non-separable	uni	no bias	linear, degenerate
WFG4	separable	multi	no bias	concave
WFG5	separable	deceptive	no bias	concave
WFG6	non-separable	uni	no bias	concave
WFG7	separable	uni	parameter dependent	concave
WFG8	non-separable	uni	parameter dependent	concave
WFG9	non-separable	multi, deceptive	parameter dependent	concave

Table 1: Properties of the applied WFG test problems [6].

mization run' [3]. When performing multiple runs, a summary attainment surface is usually drawn. This is a visual way of summarizing a number of runs of a multiobjective algorithm. The interpretation of the 50% attainment surface is that, for every point on it, a point dominating this was obtained in at least 50% of the runs.

3.2 Quality Indicators

We use three indicators, which evaluate the quality of the given set of nondominated solutions: ϵ-indicator [10], hypervolume indicator [11] and R_R indicator [5]. Smaller values of the ϵ and R_R indicator and bigger values of the hypervolume indicator denote better sets.

ϵ-indicator. The ϵ-indicator gives the factor by which a nondominated set is worse than the Pareto-optimal front with respect to all objectives. The indicator is calculated as:

$$I_\epsilon(A) = \inf_{\epsilon \in \mathbf{R}} \{\forall w \in P \; \exists z \in A : z \succeq_\epsilon w\},$$

where $\succeq_\epsilon$ denotes the ϵ dominance relation and P is the Pareto-optimal front. If P is not known, a reference set R can be used instead.

Hypervolume indicator. The hypervolume indicator $I_H(A)$ gives the hypervolume of that portion of the objective space that is weakly dominated by the nondomitated set A. The computation of this indicator demands a reference point, which should be the worst point in all objectives.

R_R indicator. There are three variants of the R_R indicator, namely $R1_R$, $R2_R$ and $R3_R$. In this paper the variant $R2_R$ is used, which is calculated as:

$$I_{R2_R}(A) = \frac{\sum_{\lambda \in \Lambda} (u^*(\lambda, R) - u^*(\lambda, A))}{|\Lambda|},$$

where R is a reference set, λ is a scalarizing vector and the utility $u^*(\lambda, A)$ is the minimum distance of a point in set A from the reference point.

4 Experimental Setup

The applied test problems were selected from the WFG test problem toolkit [6], which defines nine test problems WFG1 – WFG9. This problems comprise many different characteristics, which are summarized in Table 1. The WFG toolkit allows the user to choose the complexity of the problems. In this study, the problems have 2 objectives and a 10-dimensional decision space composed by 6 position-related and 4 distance-related parameters.

The performance of DEMO on problems WFG1-WFG9 was compared to the performance of three state-of-the-art MOEAs: IBEA [8], NSGA-II [2] and SPEA2 [9]. All individuals were coded as 10-dimensional real vectors. The algorithms had population size 100 and were run until 250 generations were reached. DEMO was implemented in its DEMO/parent variant with scaling factor 0.5 and crossover probability 0.3. All other algorithms used tournament selection with tournament size 2, real-parameter SBX crossover with probability 1 and $\eta_c = 10$, and variable-wise polynomial mutation with probability 0.1 and $\eta_m = 20$. The parameter settings of all four algorithms were taken from the literature and were not optimized for this problems. All experiments were repeated 35 times.

Experiments with IBEA, NSGA-II and SPEA2 were performed using PISA interface [1]. PISA was also used for the complete performance assessment of all four algorithms: construction of attainment surfaces and calculation of quality indicators and statistics.

5 Results and Discussion

Figure 1 presents for each problem its Pareto-optimal front and 50% attainment surfaces of DEMO and IBEA as the best algorithm among IBEA, NSGA-II and SPEA2. Additional information on the performance of all algorithms is supplied in Figure 2, where the results of the ϵ-indicator are shown in box plots. The hypervolume and $R2_R$ indicators were

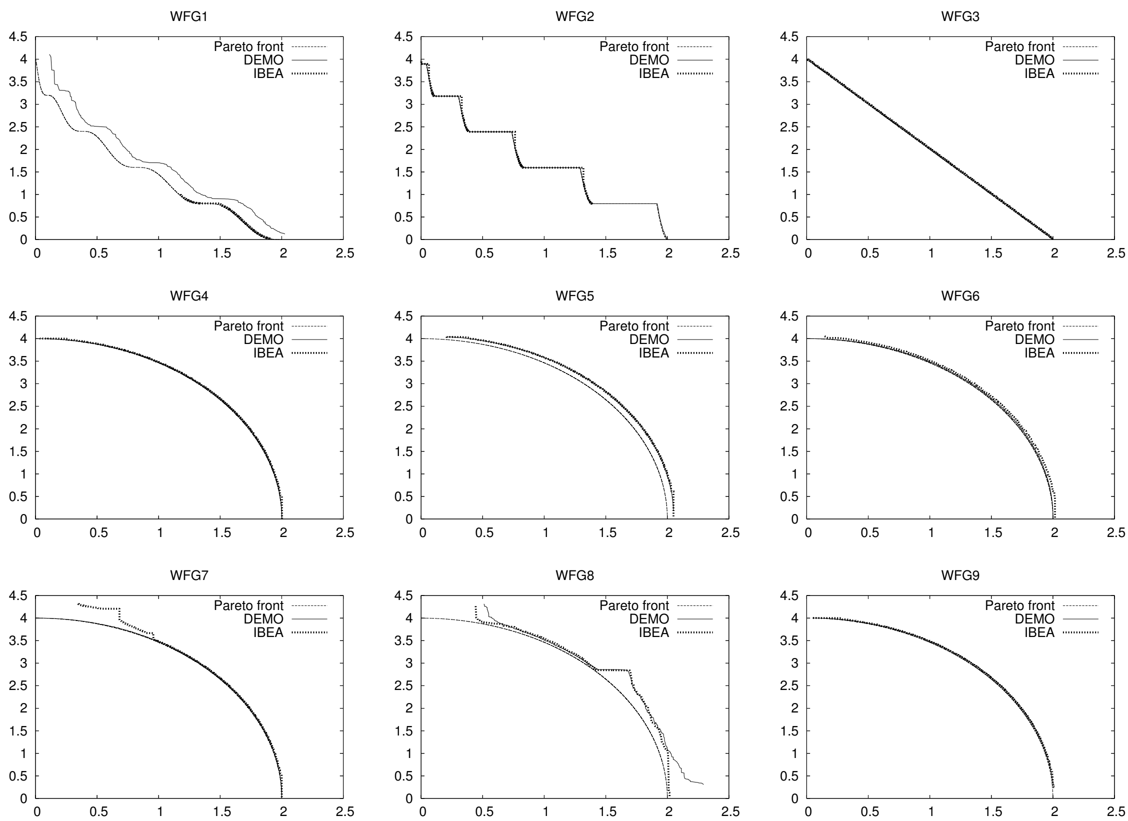

Figure 1: Pareto-optimal fronts and 50% attainment surfaces of DEMO and IBEA on problems WFG1-WFG9.

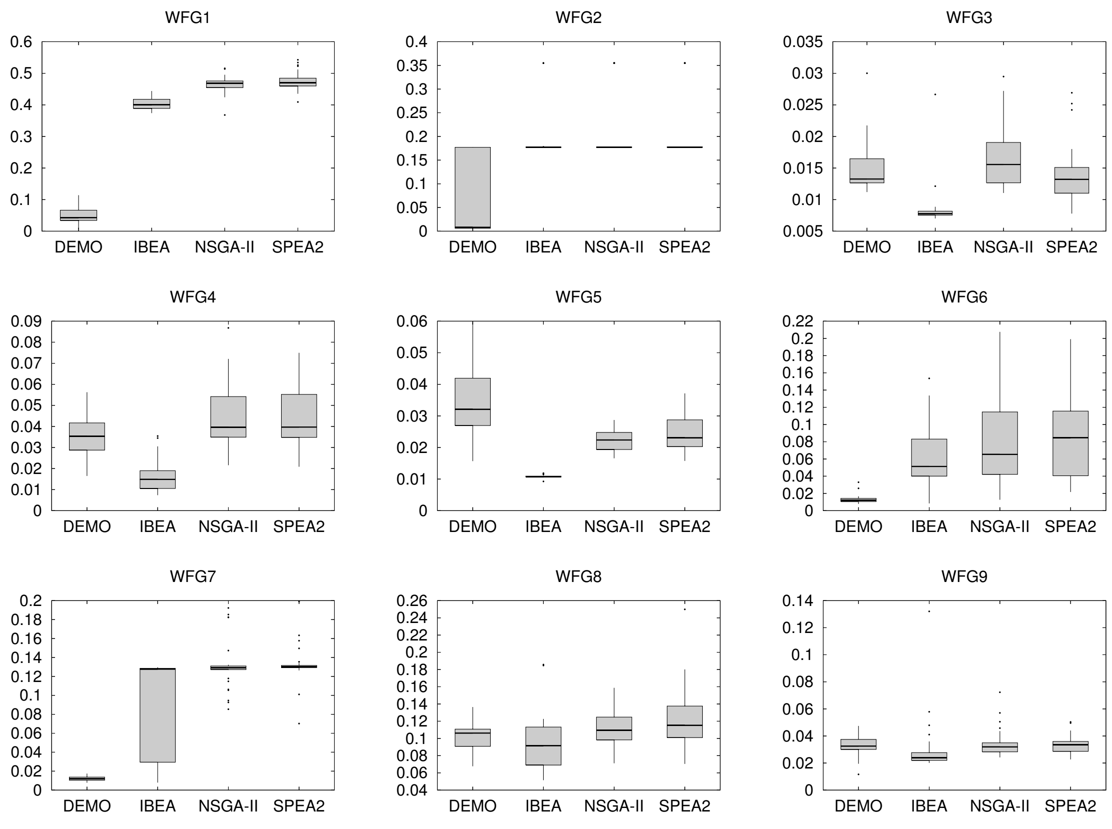

Figure 2: ϵ-indicator box plots for all algorithms on problems WFG1-WFG9.

		IBEA			NSGA-II			SPEA2		
		I_ϵ	I_H	I_R	I_ϵ	I_H	I_R	I_ϵ	I_H	I_R
DEMO	WFG1	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$
	WFG2	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$
	WFG3	$\downarrow$	$\downarrow$	$\Uparrow$	$\approx$	$\Uparrow$	$\Uparrow$	$\approx$	$\Uparrow$	$\Uparrow$
	WFG4	$\downarrow$	$\downarrow$	$\downarrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$
	WFG5	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
	WFG6	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$
	WFG7	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$
	WFG8	$\approx$	$\downarrow$	$\downarrow$	$\approx$	$\approx$	$\Uparrow$	$\Uparrow$	$\approx$	$\Uparrow$
	WFG9	$\downarrow$	$\downarrow$	$\downarrow$	$\approx$	$\Uparrow$	$\Uparrow$	$\approx$	$\Uparrow$	$\Uparrow$

Table 2: Outcomes of the Mann-Whitney significance test for ϵ-indicator, hypervolume indicator and $R2_R$ indicator. With $\Uparrow$ ($\downarrow$) we denote that DEMO is significantly better (worse) than the algorithm in the column regarding the underlying indicator. The sign $\approx$ marks there is no significant difference between DEMO and the algorithm in the column.

also computed, but their results are not presented in box plots due to space limitations.

The values of all three indicators were further tested for significance. We used the Mann-Whitney test with 5% significance level to check if the results of DEMO were significantly better, worse or equivalent to the results of IBEA, NSGA-II and SPEA2. The outcome of this test is presented in Table 2.

The results show that DEMO outperforms the state-of-the-art algorithms on four problems. DEMO achieves the worst results only on the problem WFG5. On the remaining four problems, DEMO is worse than IBEA, but equivalent or significantly better than NSGA-II and SPEA2.

Note that DEMO often achieves a better spread of solutions than the other algorithms. This is manifested on problems WFG1, WFG2, WFG6 and WFG7 (see Figure 1), where IBEA covers only a portion of the Pareto-optimal front while DEMO covers the entire extent of the front.

6 Conclusion

Following the new comparison methodology and using new test problems DEMO again showed to be comparable to state-of-the-art algorithms for multiobjective optimization. Its main weakness remains the inability to solve combinatorial problems due to the use of differential evolution.

Since we only considered two-dimensional problems, additional experiments are needed for a proper comparison between DEMO and other MOEAs.

References

[1] S. Bleuler, M. Laumanns, L. Thiele, and E. Zitzler. PISA – a platform and programming language independent interface for search algorithms. In *Evolutionary Multi-Criterion Optimization (EMO 2003)*, pages 494–508, 2003. `http://www.tik.ee.ethz.ch/pisa/`.

[2] K. Deb, A. Pratap, S. Agrawal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA–II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

[3] C. M. Fonseca and P. J. Fleming. On the performance assessment and comparison of stochastic multiobjective optimizers. In *Parallel Problem Solving from Nature (PPSN IV)*, pages 584–593, 1996.

[4] C. M. Fonseca, J. D. Knowles, L. Thiele, and E. Zitzler. A tutorial on the performance assessment of stochastic multiobjective optimizers. Tutorial at the *Third International Conference on Evolutionary Multi-Criterion Optimization (EMO 2005)*. `http://dbk.ch.umist.ac.uk/knowles/emo-tutorial-2up.pdf`.

[5] M. P. Hansen and A. Jaszkiewicz. Evaluating the quality of approximations to the non-dominated set. Technical Report IMM-REP-1998-7, Technical University of Denmark, 1998.

[6] S. Huband, L. Barone, L. White, and P. Hingston. A scalable multi-objective test problem toolkit. In *Evolutionary Multi-Criterion Optimization (EMO 2005)*, pages 280–295, 2005. `http://www.wfg.csse.uwa.edu.au/datafiles.html`.

[7] T. Robič and B. Filipič. DEMO: Differential evolution for multiobjective optimization. In *Evolutionary Multi-Criterion Optimization (EMO 2005)*, pages 520–533, 2005.

[8] E. Zitzler and S. Künzli. Indicator-based selection in multiobjective search. In *Parallel Problem Solving from Nature (PPSN VIII)*, pages 832–842, 2004.

[9] E. Zitzler, M. Laumanns, and L. Thiele. SPEA2: Improving the strength pareto evolutionary algorithm. TIK report no. 103, TIK, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, 2001.

[10] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132, 2003.

[11] E. Zitzler and L. Thiele. Multiobjective optimization using evolutionary algorithms—A comparative study. In *Parallel Problem Solving from Nature (PPSN V)*, pages 292–301, 1998.

This paper is the corrected version of the paper *Performance of DEMO on New Test Problems: A Comparison Study* published in *Proceedings of the fourteenth International Electrotechnical and Computer Science Conference ERK 2005, Volume B.*

BibTex entry:

```
@inproceedings{Robic05Performance,
   author = {Tea Robič},
   title = {Performance of DEMO on New Test Problems: A Comparison Study},
   booktitle = {Proceedings of the Fourteenth International Electrotechnical and Computer
Science Conference ERK 2005},
   volume = {B},
   year = {2005},
   pages = {121--124}
}
```