

# Effects of Population Size on Selection and Scalability in Evolutionary Many-objective Optimization

Hernán Aguirre<sup>1</sup>, Arnaud Liefooghe<sup>2,4</sup>, Sébastien Verel<sup>3,4</sup>, and Kiyoshi Tanaka<sup>1</sup>

<sup>1</sup> Shinshu University Japan

<sup>2</sup> Université Lille 1, LIFL, UMR CNRS 8022, France

<sup>3</sup> Université Nice Sophia-Antipolis, France

<sup>4</sup> Inria Lille-Nord Europe, France

ahernan@shinshu-u.ac.jp

arnaud.liefooghe@lifl.fr

verel@i3s.unice.fr

ktanaka@shinshu-u.ac.jp

**Abstract.** In this work we study population size as a fraction of the true Pareto optimal set and analyze its effects on selection and performance scalability of a conventional multi-objective evolutionary algorithm applied to many-objective optimization of small MNK-landscapes.

## 1 Introduction

Conventional multi-objective evolutionary algorithms (MOEAs) [1] are known to scale up poorly to high dimensional objective spaces [2], particularly dominance-based algorithms. This lack of scalability has been attributed mainly to inappropriate operators for selection and variation. The population size greatly influences the dynamics of the algorithm. However, its effects on large dimensional objectives spaces are not well understood. In this work we set population size as a fraction of the true Pareto optimal set and analyze its effects on selection and performance scalability of a conventional MOEA applied to many-objective optimization. In our study we enumerate small MNK-landscapes with 3 – 6 objectives, 20 bits, and observe the number of Pareto optimal solutions that the algorithm is able to find for various population sizes.

## 2 Methodology

In our study we use four MNK-landscapes [3] randomly generated with  $m = 3, 4, 5$  and 6 objectives,  $n = 20$  bits, and  $k = 1$  epistatic bit. For each landscape we enumerate all its solutions and classify them in non-dominated fronts. The exact number of true Pareto optimal solutions  $POS^T$  found by enumeration are  $|POS^T| = 152, 1554, 6265,$  and  $16845$  for  $m = 3, 4, 5,$  and  $6$  objectives, respectively. Similarly, the exact number of non-dominated fronts of the landscapes are 258, 76, 29, and 22, respectively.

We run a conventional MOEA for a fixed number of generations. The algorithm uses a population  $P$  from which it creates an offspring population  $Q$

by recombination and mutation. The population  $P$  for the next generation is obtained from the joined population  $P \cup Q$  by survival selection. In this work we use NSGA-II as the evolutionary multi-objective optimizer, set with two point crossover with rate  $pc = 1.0$ , and bit flip mutation with rate  $pm = 1/n$ .

Once evolution is over, we compare the set of  $POS^T$  with the sets of unique non-dominated solutions obtained at each generation after survival selection to determine which are true Pareto optimal solutions, count their number at each generation, and their accumulated number found during evolution.

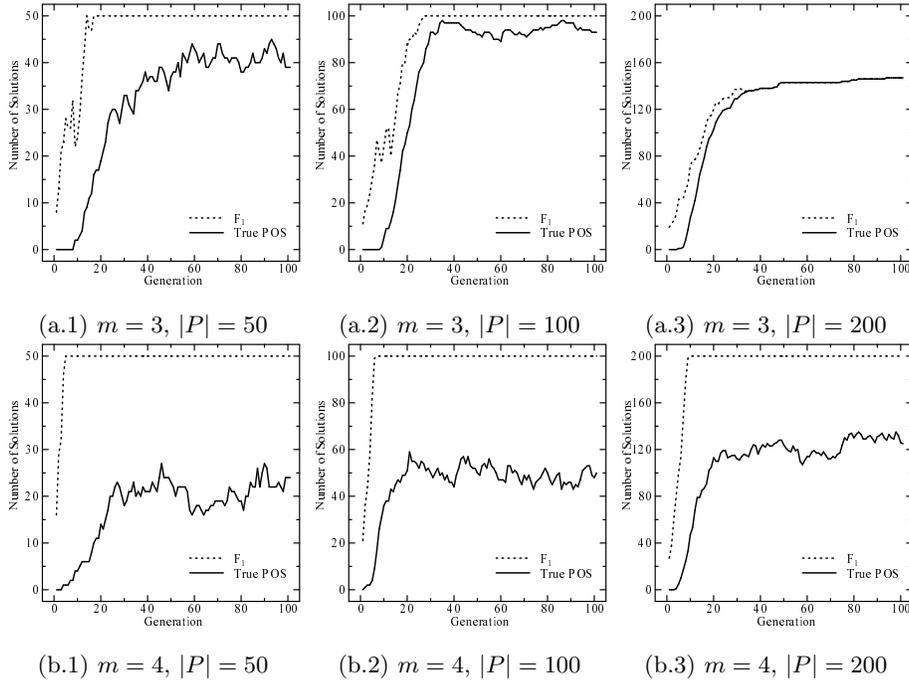
### 3 Experimental Results and Discussion

Let us denote by  $F_1$  the set of non-dominated solutions in population  $P$ , and  $F_1^T$  the set of solutions in  $F_1$  that are true Pareto optimal solutions. Fig.1 shows the number of solutions in  $F_1$  and  $F_1^T$  over the generations for  $m = 3$  and 4 objectives, running the algorithm for 100 generations with three different population sizes  $|P| = 50, 100$  and 200.

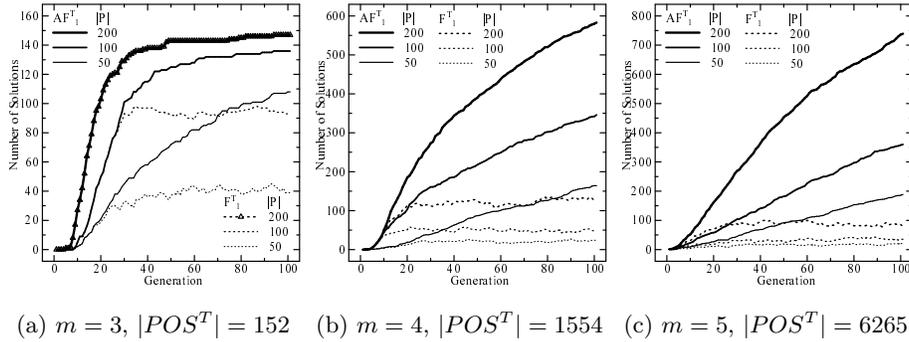
First we analyze results for  $m = 3$  objectives. When we set population size to  $|P| = 50$  or 100, a value smaller than the number of true Pareto optimal solutions  $|POS^T| = 152$ , it can be seen in Fig.1 (a.1) and (a.2) that after few generations all solutions in the population are non-dominated,  $|F_1| = |P|$ . However, not all solutions in  $F_1$  are true Pareto optimal solutions, i.e.  $|F_1^T| < |F_1| = |P|$ . Also, it is important to note that  $F_1^T$  fluctuates up and down after an initial increase. On the other hand, when we set the population size to a value larger than the number of true Pareto optimal solutions,  $|P| = 200 > |POS^T| = 152$ , it can be seen in Fig.1 (a.3) that the instantaneous non-dominated set is a subset of the population,  $F_1 \subset P$ . Also, note that from generation 35 onwards, all non-dominated solutions in the population are also true Pareto optimal,  $F_1 = F_1^T$ . In this case, the algorithm finds and keeps in  $P$  almost all true Pareto optimal solutions, 147 out of 152, during the latest stage of the search.

It is known that the number of true Pareto optimal solutions  $|POS^T|$  increases considerably with the number of objectives. However, this is often ignored and the algorithm is set with a very small population size compared to  $|POS^T|$ . To study these cases, Fig.1 (b.1)-(b.3) show results for  $m = 4$  objectives setting population size to the same values used for  $m = 3$  objectives, which are very small compared to  $|POS^T|$ , i.e.  $|P| \leq 200 < |POS^T| = 1554$ . Note that these settings of population size magnify the difficulties observed for  $m = 3$  with  $|P| = 50$  or  $|P| = 100$ . That is, fewer solutions are true Pareto optimal, although the set of non-dominated solutions of the population quickly contains mutually non-dominated solutions only. Also, larger fluctuations are observed in the number of true Pareto optimal solutions  $F_1^T$ .

In general, if  $|P|$  is set to a value smaller than  $|POS^T|$ , the algorithm cannot keep all true Pareto optimal solutions in the population. However, we would expect an ideal algorithm to keep as many true Pareto optimal solutions as the size of its population,  $|F_1^T| = |F_1| = |P| < |POS^T|$ . This is not what we observe in our results. To explain this behavior, Fig.2 shows the instantaneous number

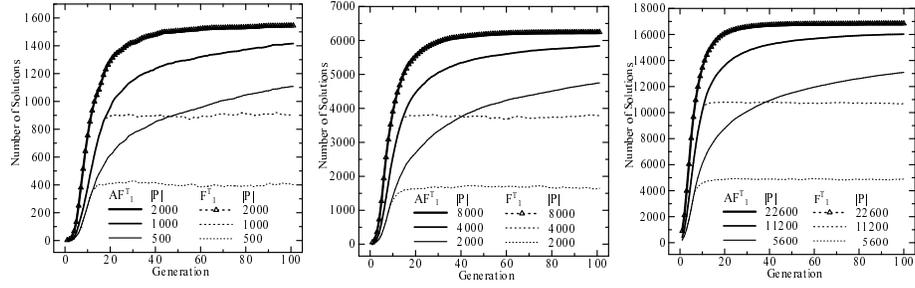


**Fig. 1.** Number of non-dominated  $F_1$  and actual number of true Pareto optimal solutions  $F_1^T$  in the population over the generations.  $|POS^T| = 152$ , and 1554 for  $m = 3$ , and 4 objectives, respectively. Population sizes  $|P| = 50, 100$ , and 200.



**Fig. 2.** Accumulated and instantaneous number of true Pareto optimal solutions,  $AF_1^T$  and  $F_1^T$ ,  $m = 3, 4$ , and 5 objectives. Population sizes  $|P| = 50, 100$ , and 200.

of true Pareto optimal solutions in the population  $|F_1^T|$  and its accumulated number  $|AF_1^T|$  over the generations for population sizes  $|P| = 50, 100$ , and 200. Note that a large number of true Pareto optimal solutions are found by the algorithm. However, not all these solutions remain in the population (except in



(a)  $m = 4$ ,  $|POS^T| = 1554$  (b)  $m = 5$ ,  $|POS^T| = 6265$  (c)  $m = 6$ ,  $|POS^T| = 16845$

**Fig. 3.** Accumulated and instantaneous number of true Pareto optimal solutions,  $AF_1^T$  and  $F_1^T$ ,  $m = 4, 5$ , and  $6$  objectives. Population sizes  $1/3, 2/3$  and  $4/3$  of  $POS^T$ .

the case  $m = 3$   $|P| = 200$ ). Some of these solutions are lost from one generation to the next one during the survival selection step of the algorithm. At this step, the algorithm joins the population  $P$  with the offspring population  $Q$  and ranks individuals with respect to dominance-depth. The best rank is given to true Pareto optimal solutions and also to some others that are not true optimal but appear non-dominated in the combined population. Let us call the set of best ranked non-dominated solutions obtained from  $P \cup Q$  as  $F_1^{P \cup Q}$ . If this set  $F_1^{P \cup Q}$  is larger than the population  $P$ , a sample of them  $P = F_1^1 \subset F_1^{P \cup Q}$  is chosen based on crowding distance during the survival step. At this point, some true Pareto optimal solutions are dropped in favor of less crowded non-optimal solutions. Summarizing,  $P = F_1^1 \subset F_1^{P \cup Q}$  and therefore  $F_1^T \subset F_1$  is more likely to occur for population sizes smaller than the number of true Pareto optimal solutions  $|POS^T|$ .

Fig.2 (a) and Fig.3 (a)-(c) show results for  $m = 3, 4, 5$  and  $6$  objectives using population sizes that correspond approximately to  $1/3, 2/3$  and  $4/3$  of the set  $POS^T$ , respectively. From these figures note that increasing population size from  $1/3$  to  $4/3$  of  $POS^T$  translates into a striking performance scalability of the algorithm, measured on terms of the number of true Pareto optimal solutions found and kept in the population. For population size  $4/3$  of  $POS^T$  the number of  $AF_1^T = F_1^T \subset F_1$  and the algorithm can actually find and keep in the population 147 out of 152, 1545 out of 1554, 6248 out of 6265, and 16842 out of 16845 true Pareto optimal solutions for 3, 4, 5 and 6 objectives, respectively.

These results show that the effectiveness of the algorithm in many-objective landscapes depends strongly on the size of the population. However, it should be noted that larger populations demand more computational time and memory. Also, a relatively larger number of solutions need to be evaluated. For example, after 100 generations, using a population size  $4/3$  of  $POS^T$ , the conventional MOEA used in this study evaluates approximately a number of solutions equivalent to 2%, 19%, 76% and 215% of the size of the search space for  $m = 3, 4, 5$ , and  $6$  objectives, respectively. In the future, we would like to analyze the efficiency of MOEAs in many-objective landscapes.

## 4 Conclusions

In this work we analyzed the effects of population size on selection and scalability of a conventional dominance-based MOEA for many-objective optimization. We showed that the performance of a conventional MOEA can scale up fairly well to high dimensional objective spaces if a sufficiently large population size is used compared to the size of the true Pareto optimal set.

## References

1. K. Deb, *Multi-Objective Optimization using Evolutionary Algorithms*, John Wiley & Sons, Chichester, West Sussex, England, 2001.
2. H. Ishibuchi, N. Tsukamoto, and Y. Nojima, “Evolutionary Many-Objective Optimization: A Short Review”, *In Proc. IEEE Congress on Evolutionary Computation (CEC 2008)*, IEEE Press, pp.2424-2431, 2008.
3. H. Aguirre and K. Tanaka, “Insights on Properties of Multi-objective MNK-Landscapes”, *Proc. 2004 IEEE Congress on Evolutionary Computation*, IEEE Service Center, pp.196–203, 2004.