A Novel DOE-Based Selection Operator for NSGA-II Algorithm

Homa Amirian
Industrial Engineering Department of Shahed University
h.amirian@shahed.ac.ir

Mahdi Bashiri
Industrial Engineering Department of Shahed University
bashiri@shahed.ac.ir

Rashed Sahraeian
Industrial Engineering Department of Shahed University
sahraeian@shahed.ac.ir

Abstract

In the present paper, a modified variant of Non-dominated Sorting Genetic Algorithm (NSGA-II) is introduced. The proposed algorithm uses Design of Experiments (DOE) regression model to insert ideal points to the population in each generation. The performance of the proposed algorithm is investigated on five bi-objective benchmark problems and the results are compared with classic NSGA-II. The empirical comparison of the results show the efficiency of the Modified NSGA-II in finding non-dominated points much faster and often better than the classic version.

Keywords; Design of Experiments; Regression model; Non-dominated sorting genetic algorithm (NSGA-II); Selection Operator

1- Introduction

NSGA-II (Deb et al., 2002) is one of the most important and commonly used multi-objective optimization algorithms. This could be due to the three special characteristics of NSGA-II, i.e. fast non-dominated sorting approach, crowing distance estimation to preserve diversity and the double criteria selection operator. Many researches have studied the extensions of NSGA-II to make it more suitable in solving multi-objective optimization problems. Some modifications are based on various sorting methods e.g., Maocai et al. (2010) used a partial order relation as a new sorting method for non-dominated individuals. They also introduced a novel encoding schemes. Jensen (2003) proposed Pareto-based fitness sorting to reduce the overall run-time complexity of NSGA-II to O ($G*N \log^{M-1}N$), making the algorithm much faster than the O ($G*M*N^2$) complexity published in (Deb et al., 2002); where G is the number of generations, M is the number of objectives, and N is the population size. Some aimed to improve the mutation and/or crossover operators (Bandyopadhyay & Bhattacharya, 2013; Dhanalakshmi et al., 2011; Etghania et al., 2013). Bandyopadhyay and Bhattacharya (2011) proposed modifications to these operators in a fuzzy environment. Ramesh et al. (2012) proposed a modified version of NSGA-II for multi-objective Reactive Power Planning (RPP) problem. They used a Dynamic Crowding Distance procedure for better diversity. Additionally, they used TOPSIS to find the best compromise solution from the set of Pareto-solutions obtained from their modified NSGA-II. Others focused on fitness evaluation techniques e.g. Ishibuchi et al. (2009) used weighted sum fitness functions in NSGA-II. Liu et al. (2005) modified NSGA-II by a nearest neighbor (1-NN) classifier for numerical model calibration. Similarly, Pires et al. (2012) improved NSGA-II by adding a local search approach to the algorithm. They show that the modification results in better convergence towards the non-dominated front and ensures that the solutions attained are well spread over it. Pindoriya & Srinivasan (2010) proposed heuristic methods to seed the initial random population with a Priority list based solution for better convergence. They also studied a penalty-parameter-less constrained binary tournament method as the selection operator to handle the problem constraints efficiently. Over the years, design of experiments (DOE) has been vastly used in regression analysis. Cali et al. (2007) used a factorial response surface analysis for fitting regression models in a tubular SOFC generator problem. Li and Hickernell (2013) used DOE for linear regression models; fitted using both function and gradient data. Dette et al. (2006) have discussed the problem of designing experiments for exponential regression models on the basis that an appropriate choice of the experimental conditions can improve the quality of statistical inference substantially. A summary of the modifications applied to NSGA-II can be found at Table 1. Similar to the modifications mentioned in table 1, we introduce a DOE-based method to improve the overall solution estimation of NSGA-II. Since the algorithm is classified as a population-based evolutionary algorithm, preserving a better population in each iteration increases the number of solutions with low ranks (i.e. true Pareto front) and thus aims to accelerate the convergence. Simply put, we use DOE for estimation of the parameters in linear regression models. In the main loop of the proposed algorithm, a regression DOE model is used to insert optimized points concerning each objectives. The addition of these ideal points in each generation improves the general performance of the algorithm. The remainder of the paper is organized as follows. Section 2 describes the proposed modified NSGA-II. Experimental settings, test problems and performance metrics are given in Section 3. Results obtained from the tests are discussed in Section 4. Finally, the conclusion of the research can be found in Section 5.

2- Selection Phase of Modified NSGA-II

Selection provides the driving force in an evolutionary algorithm (EA) and the selection pressure (i.e. probability of the best individual selected) is a critical parameter. Too much, and the search will terminate prematurely, too little, and progress will be slower than necessary (Blickle & Thiele, 1995). Some of the selection methods, being stochastic, may lose the best value from the population. In tournament selection, for example, n individuals are chosen at random from the population, with the best being selected for reproduction. A fresh tournament is held for each member of the new population. Hence, the best member of the population may simply not be picked for any contests. Moreover, because each tournament is carried out individually, it suffers from sampling error. In Roulette Wheel selection scheme, each individual is given a chance to become a parent in proportion to its fitness. It is called roulette wheel selection as the chances of selecting a parent can be seen as spinning a roulette wheel with the size of the slot for each parent being proportional to its fitness. Obviously, those with the largest fitness (slot sizes) have more chance of being chosen. Hence, n trials have to be performed to obtain an entire population. As these trials are independent of each other, a relatively high variance in the outcome is observed. In Fitness Proportionate Selection (FPS), individuals are selected in proportion to their fitness on the evaluation function, relative to the average of the whole population. This scheme suffers from scaling problems, which are partially addressed by scaling; however, the selection pressure achieved is still dependent on the spread of fitness values in the population. Rank selection is an attempt to overcome the scaling problems of the direct fitness based approach. The population is ordered according to the measured fitness values. A new fitness value is then ascribed, inversely proportional to each individual's rank (Blickle & Thiele, 1995). In NSGA-II, individuals are selected according to rank and crowding distance, thus maintaining a balance between convergence and diversity respectively. It is obvious that this scheme does not have the problems mentioned above, such as sampling error, premature convergence or losing the best individual in a contest. In traditional NSGA-II, selection is carried out on the pool of elite individuals and the new individuals obtained from mutation and crossover. However, in our algorithm adding new points extracted by DOE analysis to the pool of individuals enhances the probability of finding better solutions for the next generation. These points are found by optimizing the linear regression model of a sample of individuals in respect to the separate objectives (i.e. one response per objective).

TABLE I. SUMMARY OF MODIFICATIONS TO NSGA-II

Modification	Result of Modification	Reference
ε-Elimination Algorithm as a diversity preserving mechanism	Better Diversity	Etghania et al.(2013)
New Mutation Algorithm	Better Mutated Individuals	Bandyopadhyay & Bhattacharya (2013)
Enhanced NSGA-II with local search	Better Convergence	Pires et al.(2012)
Dynamic Crowding Distance	Better Diversity	Ramesh et al.(2012)

Modification	Result of Modification	Reference
Dynamic Crowding Distance, Controlled Elitism	Uniform Diversity, Lateral diversity-preserving operator	Dhanalakshmi et al.(2011)
New Crossover & Mutation Operators for Fuzzy Values	Improvement in general Performance	Bandyopadhyay & Bhattacharya (2011)
Priority list based solution, Penalty-parameter-less Constrained Binary Tournament method	Better Convergence, Efficient selection	Pindoriya & Srinivasan(2010)
Partial Order relation, Crossover operater by Cauchy Distribution	Better Sorting, Enhancing the performance	Maocai et al. (2010)
Selection by Scalarizing Fitness functions	Additional Selection Pressure	Ishibuchi et al.(2009)
Numerical Calibration Method	Reduction in actual fitness evaluations	Liu et al.(2005)
Pareto-Based fitness	Reduction in the overall run-time complexity	Jensen (2003)

In our proposed operator, the variables are factors whose responses are evaluated by the current objectives and constraints. Since the model is available, there is no replication involved and the relation between the factors and responses is linear. The flowchart of modified NSGA-II is given in Fig. 1.a. A systematic outline of the selection phase of the algorithm can be found at Fig.1.b.

3- Experimental Settings

3-1- Test Problems

The performance of the proposed algorithm is tested on a set of five unconstrained benchmark problems generally used to validate the performance of different Evolutionary Algorithms (EAs). All these problems are taken from Deb et al. (2002). These are Fonseca and Fleming's first (FON) and second (MOP2) functions, Kursawe's function (MOP4), Laumanns' function (LAU) and Lis' function (LIS). The initial population is set to 100; the mutation (mu), crossover (Cr) and DOE probability are taken as 0.5, 0.3, and 0.5 respectively. The problem is compiled in MATLAB R2012a and is executed on Intel(R) Corel(TM)2 Duo 2 GHz PC with 1 GB RAM. In each case, a run is terminated when the number of generations reaches to 250.

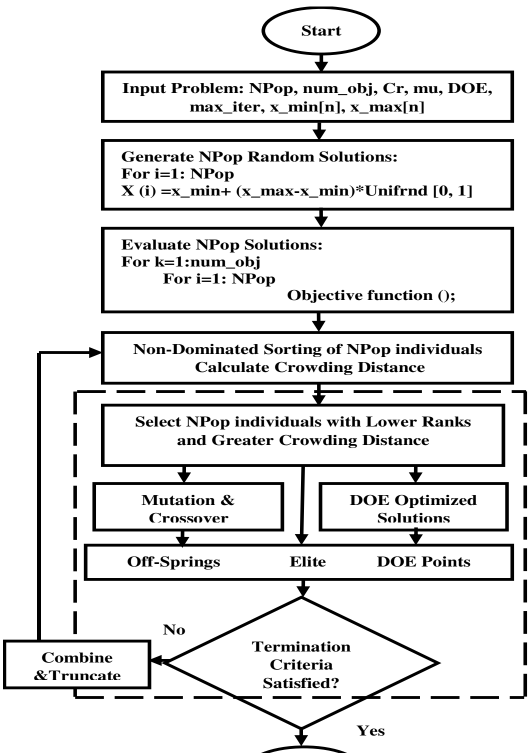

Fig 1.a. Flow chart of Modified NSGA-II

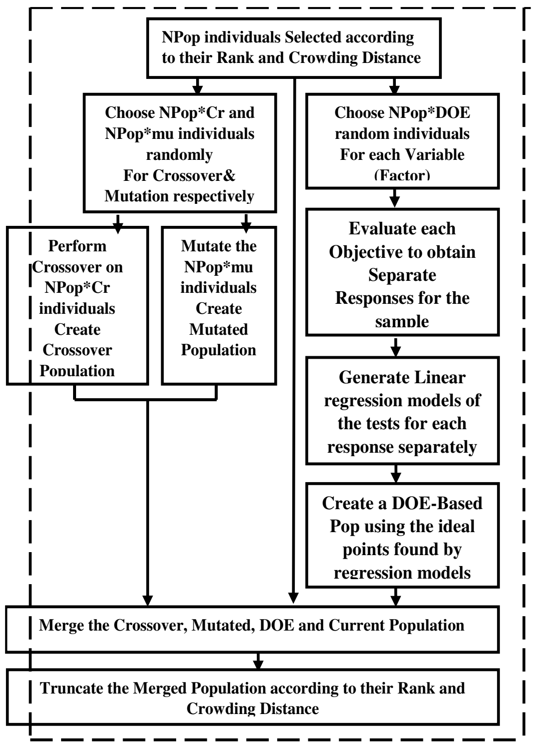

Fig 1.b. Selection Phase of Modified NSGA-II

3-2- Performance Metric

To validate the proposed algorithm, we use two commonly adopted performance metrics in evolutionary multi-objective optimization literature. Since the key to any good evolutionary algorithm is maintaining the balance between precision and dispersion of estimated solutions, convergence and diversity metrics are used respectively. A brief introduction of these metrics is given here:

3-2-1 Convergence Metric

Deb et al. (2002) proposed this metric to evaluate the convergence towards a reference set (P*). P* can be either a set of Pareto optimal solutions (if known) or the nondominated set of points in a combined pool of all generations-wise populations obtained from a run. Since the metric measures the distance between the optimal front and the obtained Pareto front, the smaller values are desirable. Mathematically it can be defined as:

$$\gamma = \sum_{i=1}^{N} d_i \,, \tag{1}$$

Where N is the number of nondominated solutions found by the algorithm, and d_i for *i-th* solution is:

$$d_i = min_{j=1}^{|P^*|} \sqrt{\sum_{k=1}^{M} \left(\frac{f_k(i) - f_k(j)}{f_k{}^{max} - f_k{}^{min}} \right)^2} . \tag{2}$$

Here *M* denotes the number of objectives and $f_k{}^{max}$, $f_k{}^{min}$ are the maximum and minimum function values of the *k-th* objective function in *P** respectively.

3-2-2 Diversity Metric

Deb et al. (2002) also introduced a diversity metric to gauge the extent of spread achieved among the obtained solutions. Assuming that there are N solutions on the best-nondominated front, this metric is given by:

$$\Delta = \frac{d_f + d_l + \sum_{i=1}^{N-1} |d_i - \bar{d}|}{d_f + d_l + (N-1)\bar{d}} . \tag{3}$$

Where $d_i \ (i = 1, .., N-1)$ is the Euclidean distance between consecutive solutions in the obtained nondominated set and $\bar{d}$ is the average of all d_i. Here d_f and d_l refer to the Euclidean distances between the extreme solutions and the boundary solutions of the obtained nondominated set.

The proposed algorithm is also compared statistically with the classic NSGA II using the non-parametric Wilcoxon Signed-Rank, Matched Pairs Test conducted by IBM SPSS software package. This pairwise test aims to find any significant difference between the two algorithms (Ali, Pant, & Siarry, 2012). However, the results obtained from this test with the significance level of 0.05, show that there is no significant difference between the modified algorithm and the classic NSGA-II in terms of diversity and convergence.

4- Results and Discussion

In this section, results of five test problems obtained from the proposed algorithm (Modified NSGA-II) are compared with the results of classic NSGA-II using the performance metrics. Table 2 represent the mean and variance of the values of convergence metric, diversity metric and the average number of Pareto Front (PF) points of 10 runs of 250 generations for each problem.

TABLE II. STATISTICS OF THE RESULTS ON TEST PROBLEMS

Algorithm		Convergence Metric	Divergence Metric	Mean PF Points
MOP2	NSGA II	**0.010322915**±2.07703E-07	**0.573570278**±0.001800206	97.892
	Modified NSGA II	0.010471533±**1.03984E-07**	0.5776085± **0.001245595**	**98.672**
FON	NSGA II	**0.001182639**±8.34482E-09	0.539811932±**0.000174501**	98.664
	Modified NSGA II	0.001242241±**7.32416E-09**	**0.535733962**±0.000460322	**99.14**

Algorithm		Convergence Metric	Divergence Metric	Mean PF Points
LAU	NSGA II	0.006814716±8.78907E-08	**0.556221519**±0.00126668	99.212
	Modified NSGA II	**0.006334897±7.45159E-08**	0.556871087±**0.001077994**	**99.316**
LIS	NSGA II	**0.006349297±2.34976E-06**	0.556200579±0.001759527	99.504
	Modified NSGA II	0.037480899±1.29259E-05	**0.555742102±0.000794065**	**99.636**
MOP4	NSGA II	**0.001761318±9.84682E-08**	**0.575570093±0.00210237**	94.2
	Modified NSGA II	0.002156692±1.03617E-07	0.585793203±0.002499827	**97.572**

In terms of convergence, even though the differences are minimal, the modified version needs some small improvements, since it is mostly dominated by the classic NSGA-II in four out of five examples. As for the divergence metric, the proposed algorithm has smaller variances than the classic NSGA-II, hence making our algorithm more robust. However only in two out of five problems, the divergence mean of Modified version outperforms NSGA II. Figs 2-6 illustrate the obtained nondominated Pareto front and the optimal Pareto front for each test problems. It can be seen from the figures that our algorithm finds the true Pareto Front for each problem precisely.

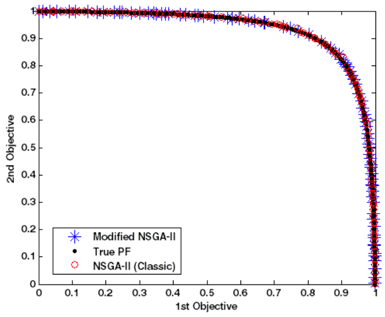

Fig 2. True PF and nondominated Solutions on FON

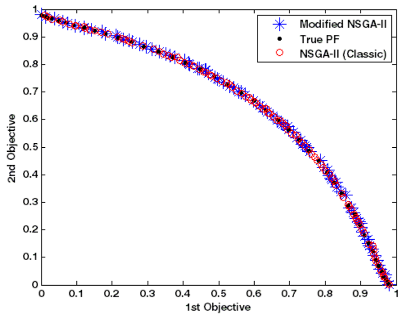

Fig 3. True PF and nondominated Solutions on MOP2

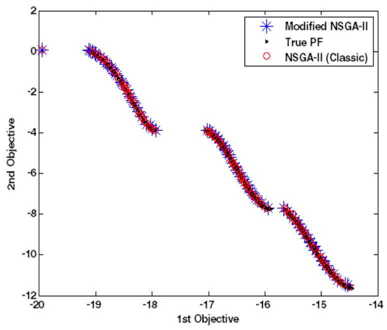

Fig 4. True PF and nondominated Solutions on MOP4

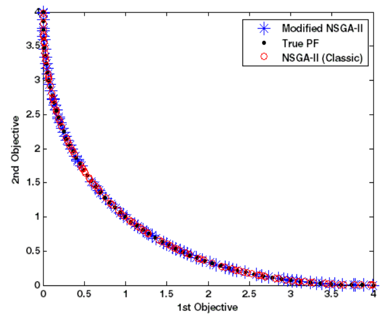

Fig 5. True PF and nondominated Solutions on LAU

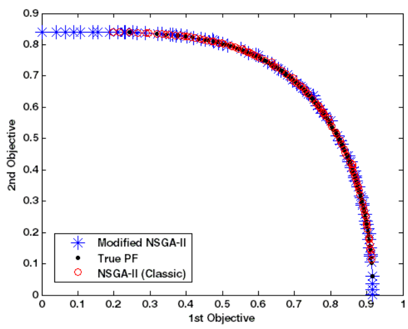

Fig 6. True PF and nondominated Solutions on LIS

On the other hand, the mean of PF points found by the proposed algorithm is greater than the classic NSGA-II in all cases; indicating a better exploration of search space. (See Table 2-6)

4-1- Analysis of PF Solutions

As an step by step analysis of the results, the number of PF points found by the two algorithms for the first 10 generations are shown in Fig. 7-10. From these figures, it is clear that nearly in all the generations DOE-based NSGA-II finds more PF points than the classic version. This shows a more thorough search in the algorithm which leads to better population and thus faster convergence.

4-2- Test on the Population Size

To test the effect of population size on DOE points, each problem is solved 10 times with population sizes 10, 50 and 100. The results for the number of DOE points found in each generation with different population sizes are shown in Figs.11-14.

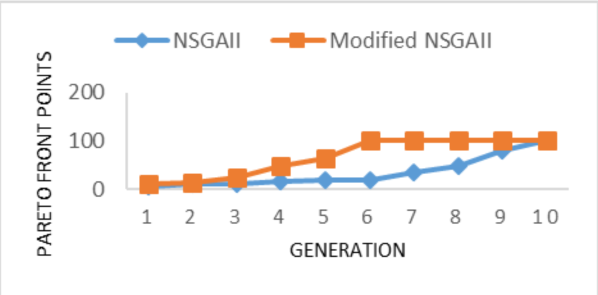

Fig 7. Superiority of Modified NSGA-II in finding PF Points over the classic NSGA-II on problem FON

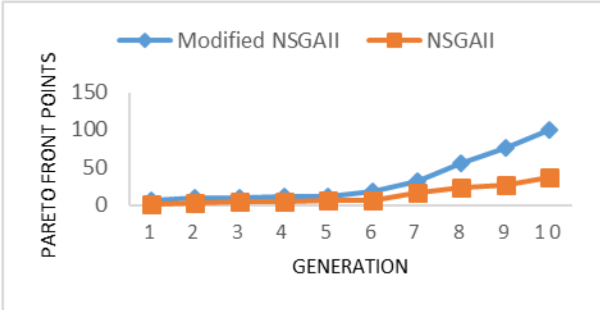

Fig 8. Superiority of Modified NSGA-II in finding PF Points over the classic NSGA-II on problem MOP2

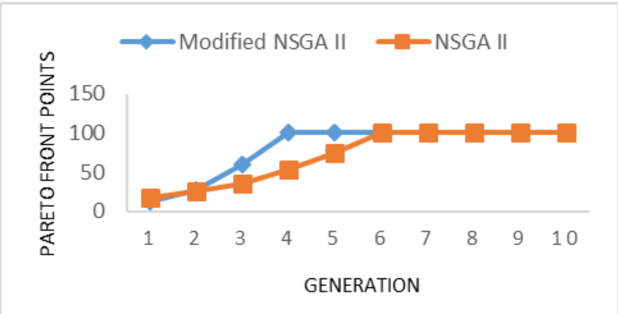

Fig 9. Superiority of Modified NSGA-II in finding PF Points over the classic NSGA-II on problem LAU

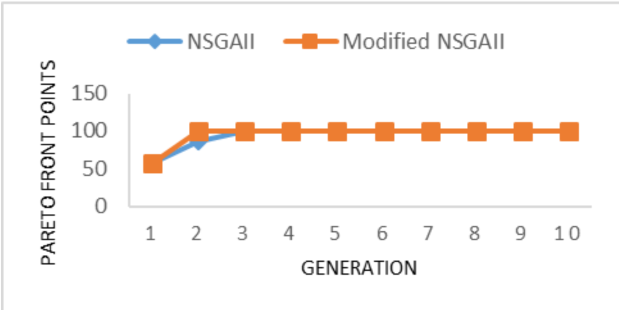

Fig 10. Superiority of Modified NSGA-II in finding PF Points over the classic NSGA-II on problem LIS

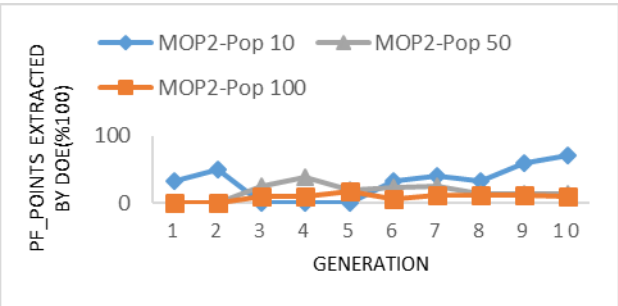

Fig 11. Increase in Population Size leads to a decrease in PF_DOE points on problem MOP2.

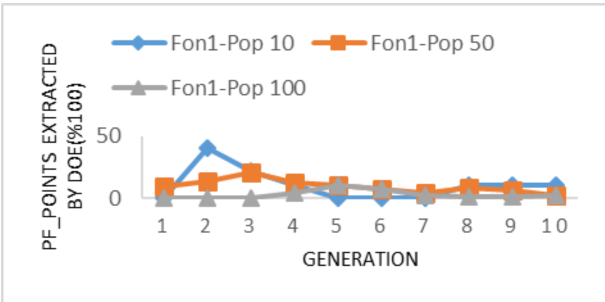

Fig 12. Increase in Population Size leads to a decrease in PF_DOE points on problem FON.

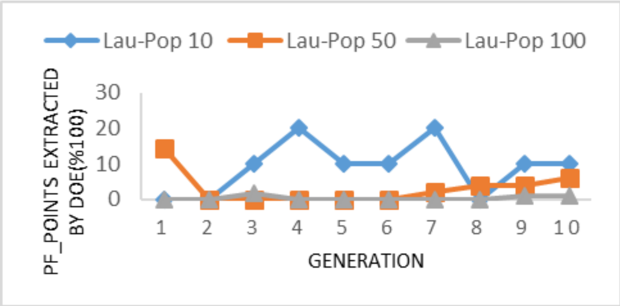

Fig 13. Increase in Population Size leads to a decrease in PF_DOE points on problem LAU.

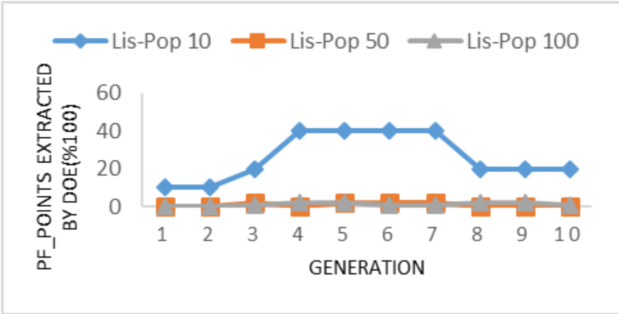

Fig 14. Increase in Population Size leads to a decrease in PF_DOE points on problem LIS.

It can be concluded from the Fig. 11-14 that DOE points play a major part in the Pareto front of smaller populations. However as the population size increases, few DOE points can be seen in the nondominated Pareto front, leading to a small percentage of them over the generations.

5- Conclusion

In this study, we proposed a modified NSGA-II for solving multi-objective optimization problems. Since the performance of evolutionary algorithms such as NSGA-II largely depends on the population of individuals, the algorithm aims to improve the estimated population at each generation hence speeding up the convergence process. For this purpose, in each generation, the Modified NSGA-II adds new optimized points to the population using DOE regression model. The basic idea behind this is that feeding better solutions to the algorithm results in better and faster convergence of the algorithm. The numerical results show that the proposed version finds more Pareto front points than the classic NSGA-II. Additionally, in some problems it performs quite well in convergence and diversity. However it is clear that some improvements can be made to counter the premature convergence nature of the algorithm resulted by adding DOE points. In terms of diversity, the proposed version is more robust than the classic NSGA-II since the diversity metric variances in our algorithm are often lower than the real-coded NSGA-II. As a future study, other statistical approaches can be added to modify the NSGA-II operators for higher performance; also adding random local search in the modified algorithm is suggested as another future work.

References

Ali, M., Pant, M., & Siarry, P. (2012). “An efficient differential evolution based algorithm for solving multi-objective optimization problems”. *European Journal of Operational Research.* Vol. 217, No. 2, pp. 404-416.

Bandyopadhyay, S., & Bhattacharya, R. (2011). “Applying modified NSGA-II for bi-objective supply chain problem”. *Journal of Intelligent Manufacturing.* Vol. 24, No. 4, pp. 707–716.

Bandyopadhyay, S., & Bhattacharya, R. (2013). “Solving multi-objective parallel machine scheduling problem by modified NSGA-II”. *Applied Mathematical Modelling.* Vol. 37, No. 10-11, pp. 6718–6729.

Blickle , T., & Thiele, L. (1995). “A Comparison of Selection Schemes used in Genetic Algorithms”. Zurich: Computer Engineering and Communication Networks Lab, Swiss Federal Institute of Technology, unpublished.

Cali, M., Santarelli, M., & Leone, P. (2007). “Design of experiments for fitting regression models on the tubular SOFC :screening test, response surface analysis and optimization”. *International Journal of Hydrogen Energey.* Vol. 32, No. 3, pp. 343-358.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). "A fast and elitist multiobjective genetic algorithm:NSGA-II". *IEEE Transactions on Evolutionary Computation.* Vol. 6, No. 2, pp. 182-197.

Dette, H., Lopez, I., Rodriguez, I., & Pepelyshev, A. (2006). "Maximin efficient design of experiment for exponential regression models". *Journal of Statistical Planning and Inference.* Vol. 136, No. 12, pp. 4397-4418.

Dhanalakshmi, S., Kannan, S., Mahadevan, K., & Baskar, S. (2011). "Application of modified NSGA-II algorithm to combined economic and emission dispatch problem." *International Journal of Electrical Power & Energy Systems.* Vol. 33, No. 4, pp. 992–1002.

Etghania, M., Shojaeefardb, M., Khalkhalib, A., & Akbarib, M. (2013). "A hybrid method of modified NSGA-II and TOPSIS to optimize performance and emissions of a diesel engine using biodiesel". *Applied Thermal Engineering.* Vol. 59, No. 1-2, pp. 309–315.

Ishibuchi, H., Tsukamoto, N., & Nojima, Y. (2009). "Empirical analysis of using weighted sum fitness functions in NSGAII for many-objective 0/1 knapsack problems." 11th International conference on Computer Modelling and Simulation, Cambridge: UKSIM, pp. 71–76.

Jensen, M. (2003). "Reducing the run-time complexity of multiobjective EAs: the NSGA-II and other algorithms". *IEEE Transactions on Evolutionary Computation.* Vol. 7, No. 5, pp. 503–515.

Li, Y., & Hickernell, F. (2014). "Design of experiments for linear regression models when the gradiant information is available". *Journal of Statistical Planning and Interference.* Vol. 144, No. 13, pp. 141–151. doi:10.1016/j.jspi.2013.01.010

Liu, Y., Zhou, C., & Ye, W. (2005). "A fast optimization method of using nondominated sorting genetic algorithm (NSGA-ll) and 1-nearest neighbor (1 NN) classifier for numerical model calibration". IEEE International Conference on Granular Computing. Vol. 2, China: Beijing, pp. 544–549.

Maocai, W., Guangming, D., & Hanping, H. (2010). "Improved NSGA-II algorithm for optimization of constrained functions". International Conference on Machine Vision and Human-Machine Interface, Kaifeng, China, pp. 673–675.

Montgomery, D. (2005). "Design and Analysis of Experiments" (6th ed.). United States of America: John Wiley & Sons, Inc.

Pindoriya, N., & Srinivasan, D. (2010). "Modified NSGA-II for day-ahead multi-objective thermal generation scheduling". IPEC, 2010 Conference Proceedings, Singapore, pp. 752–757.

Pires, D., Antunes, C., & Martins, A. (2012). "NSGA-II with local search for a multi-objective reactive power compensation problem". *International Journal of Electrical Power & Energy Systems.* Vol. 1, No. 43, pp. 313–324.

Ramesh, S., Kannan, S., & Baskar, S. (2012). "Application of modified NSGA-II algorithm to multi-objective reactive power planning". *Applied Soft Computing.,* Vol. 12, No. 2, pp. 741–753.