Empirical Studies of Resampling Strategies in Noisy Evolutionary Multi-Objective Optimization

Shasha Zhou
College of Computer Science and Engineering
University of Electronic Science and Technology of China
Chengdu, China
shashazhou.chn@gmail.com

Ke Li
Department of Computer Science
University of Exeter
Exeter, UK
k.li@exeter.ac.uk

***Abstract*—Optimization problems are ubiquitous in real-world engineering scenarios where the goals are to enhance interested aspects such as efficiency, productivity, and profitability. However, solving practical optimization problems could be non-trivial, partly due to the presence of a wide range of noises, including environmental noises, model biases, time-domain variations, measurement uncertainties and many other uncontrolled variables. In this paper, we empirically study the effect of noise range, sample size and resampling type on the solution quality of MOEAs when noise is added to decision variables. Our empirical results, conducted on three commonly used MOEAs, i.e. NSGA-II, MOEA/D and IBEA, demonstrate that noise range has more significant impact on the robustness of optimization algorithms compared to sample size and resampling type. In addition, we introduce the concept of *bad point*, which is able to illustrate how noise affects the performance of different MOEAs.**

***Index Terms*—robustness, multi-objective, evolutionary optimization, noisy optimization**

I. Introduction

Optimization problems are ubiquitous in real-world engineering scenarios where the goals are to enhance interested aspects such as efficiency, productivity, and profitability. In both scientific research and practical applications, Multi-Objective Optimization Problems (MOPs) [1] are a common occurrence [2]–[4]. Over the years, the Multi-Objective Evolutionary Algorithm (MOEA) has proven to be an effective and promising approach to tackle MOPs. Thanks to the efforts of the community, numerous mature methods have been developed, which enable us to obtain a representative subset of the overall Pareto front (PF) in complex situations. MOEAs can be classified into three categories based on evolutionary mechanisms, i.e. domination-based EA [5], [6], indicator-based EA [7]–[9] and decomposition-based EA [10]–[12].

However, solving practical optimization problems could be non-trivial, partly due to the presence of a wide range of uncertainties. The uncertainty in optimization problems can occur in objective functions, decision variables, function parameters. Thus, the noisy multi-objective optimization can be divided into three types. The first type is when noise is added on decision variables, which affects the objective (fitness) evaluations [13], [14]. Secondly, noise in fitness function could result in errors in performance estimation [15], [16]. The last type of uncertainty stems from the fluctuation of environmental parameters subject to varying environmental and operational condition [17], which causes the specifications of optimization problem change over time [18]. In this work, we concern the problems with noise in decision variables.

The presence of these noises could have various impacts on the performance of optimization algorithms. Firstly, noise can cause inaccurate fitness function evaluation, e.g., underestimation of the fitness value of a high-quality solution, or vice versa. This would then lead to inappropriate selection of parents for reproduction, resulting in the loss of promising individuals [19]. In addition, when the population is dominated by a large number of poor solutions, it may deviate from the true Pareto front due to the presence of noise. This would impede the convergence of the population, and thus lead to the generation of various sub-optimal solutions [20]. Finally, optimization algorithms may overestimate the fitness of a solution when trying to reach the optimal point, the degree of which is directly proportional to the intensity of noise [21].

Resampling also known as explicit averaging, is a popular and general technique that is commonly used in the literature for handling evaluation noise by reducing noise strength [13], [22], which independently evaluates the candidate solution κ times. Assuming that the noise follows a normal distribution with standard deviation of σ_ϵ, mean of κ independent evaluations of the solution $\mathbf{x}$ has a standard deviation of $\sigma_\epsilon/\sqrt{\kappa}$. Moreover, resampling can provide an estimate of the noise strength of each solution. It can also reveal the impact of noise, e.g., by analyzing the change in the rank of solutions after resampling.

Although there is general agreement that resampling can improve performance in noisy problems, there remains many crucial questions to be solved, like the impact of noise, sampling methods, sample size and etc. In this work, we empirically study the effect of noise range, sample size and resampling type on the solution quality of MOEAs when noise is added to decision variables. Our empirical results, conducted on three commonly used MOEAs, i.e. NSGA-II, MOEA/D and IBEA, demonstrate that noise range has more significant impact on the robustness of optimization algorithms compared to sample size and resampling type. In addition, we introduce the concept of *bad point*, which is able to illustrate how noise affects the performance of different MOEAs.

The rest of this paper is organized in the following. Section II provides some preliminary knowledge related to this work. The experimental setup is introduced in Section III, and the results are discussed in Section IV. At the end, Section V concludes this paper.

II. Preliminaries

In this section, we start from a formal problem definition of the noisy multi-objective optimization problems (NMOPs) considered in this paper. Then, we overview the existing noisy optimization algorithms.

A. Problem Formulation

In this paper, we consider adding noise into decision space, and a minimization NMOP can be stated as follows:

$$\begin{array}{ll} \text{minimize} & \mathbf{F}(\mathbf{x}) = (f_1(\mathbf{x}+\boldsymbol{\delta}), \dots, f_m(\mathbf{x}+\boldsymbol{\delta}))^T \\ \text{subject to} & \mathbf{x} \in \Omega \end{array}, \quad (1)$$

where $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^T$ is additive noise in decision space, and $-\delta_i^{max} \leq \delta_i \leq \delta_i^{max}, i = 1, \dots, n$. $\Omega = \prod_{i=1}^{n}[a_i, b_i] \subseteq \mathbb{R}^n$ is the decision (variable) space, $\mathbf{x} = (x_1, \dots, x_n)^T \in \Omega$ is a candidate solution. $\mathbf{F} : \Omega \rightarrow \mathbb{R}^m$constitutes m conflicting objective functions, and $\mathbb{R}^m$ is called the objective space. The optimal solutions to NMOPs are denoted as Pareto set (PS), and the corresponding objective values are denoted as Pareto front (PF). The goal of noisy optimization is to find the global minimum of the true objective functions.

B. Literature Review

Although uncertainties are quite common in real life, there is a little research on how to deal with NMOPs. The existing noisy multi-objective optimization algorithm can be roughly categorized under five categories [23]: (1) resampling, (2) implicit averaging, (3) robust selection, (4) specialized search strategies and (5) alternative fitness estimation methods. Among them, averaging methods are robust, simple and can be easily integrated with any population-based search method.

In resampling, also known as explicit averaging, a candidate solution is evaluated κ times independently, in which κ is known as the sample size. The true fitness is estimated using some statistical indicators, including mean, median, max. However, the cost of resampling is expensive. Given a population size of N, the number of function evaluations per iteration is $N\kappa$. Thus, the cost of resampling is expensive. Researchers have designed some methods to reduce the total number of evaluations to achieve the same performance. Fieldsend and Everson [24] proposed rolling tide evolutionary algorithm (RTEA), which only samples the non-dominated solutions. There is also an accumulative sampling strategy developed in [22]. A new method of fitness evaluation based on fitness inheritance is proposed in [25], where a weighted fitness value of parents is regarded as one of the child.

The implicit averaging increases the population size to compensate for the impact of noise. For a sufficient large population, [26] proves that it can maintain more diversity so that the larger differences in ability can be discerned with a modest number of evaluations.

Noises usually affect the judgment of the dominance relationship between two solutions when we face a NMOP, leading to many promising solutions to be deleted. Therefore, the third category aims to design robust selection operators. For example, stochastic dominance criteria [27], [28] are adopted to identify the degree of dominance of two solutions with a probabilistic estimate. A rolling tide characteristic [24] is introduced into an elitist multi-objective optimization to avoid deception by noise while selecting optimal solutions and identifying the solution for resampling. Buche et. al [29] proposed the concept of dominance-dependent lifetime for each population member, which varies inversely with the number of solutions it dominated. Thus, the strategy defends the overall population against the impact of the misleading fitness of unreliable solutions.

Increasing the diversity of a population is the fourth category, which is conducive to providing individuals with more traits in the selection process. Goh and Tan [30] adopted an experiential learning-directed perturbation (ELDP) strategy to govern the candidate movement toward the direction of fitness improvement on the basis of the information acquired from the last few generations for faster convergence. Meneghini et. al [31] introduce a coevolutionary approach for robust MOO (C-RMOEA/D), which works with two populations. One representing solutions and minimizing the objectives, another representing uncertainties and maximizing the objectives in a worst case scenario.

The last way to solve noise is alternative fitness estimation methods. The Welch confidence interval [32] is used to determine if there is a significant difference between two solutions. An indicator-based multi-objective optimization algorithm [33] proposes a novel fitness evaluation method based on ϵ-index.

Although above NMOEAs alleviate the influence of noises to some extent, most algorithms only focus on the noise in the objective function. Actually, noises exist in all aspects of life, there remains many crucial questions to be solved in noisy optimization.

III. Experiment Setup

The experimental settings used in this paper are outlined as follows.

A. Test Problems

To evaluate the impact of sampling method, totally 9 test benchmark problems are selected from the ZDT and DTLZ test suit, i.e. ZDT1$\sim$ZDT4, ZDT6 and DTLZ1$\sim$DTLZ4. For each ZDT instance, the number of objectives m is 2. For each DTLZ instance, the number of objectives m is set as 3. As suggested in [34], the number of decision variables is set as $n = 30$ for ZDT1$\sim$ZDT3, and $n = 10$ for ZDT4 and ZDT6. According to the recommendations in [35], the number of decision variables is set as $n = m + r - 1$ for DTLZ test instances, where $r = 5$ for DTLZ1 and $r = 10$ for DTLZ2, DTLZ3 and DTLZ4.

Manuscript 1037 submitted to 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Received April 23, 2023.

B. Test Algorithms

In our empirical study, we adopt three widely used multi-objective evolutionary algorithms, including NSGA-II, MOEA/D and IBEA as test algorithms. All algorithms are implemented in C++.

C. Parameter Settings

The parameter settings of our experiments are summarized as follows.

1) *Settings for Reproduction Operations*: The crossover probability is $p_c = 1.0$ and its distribution index is $\eta_c = 30$. The mutation probability is $p_m = 1/n$ and its distribution index is $\eta_m = 20$.
2) *Population Size*: The population size is predefined as 100 for each algorithm.
3) *Number of Runs and Stopping Condition*: Each algorithm is independently run 30 times on each test instance. The stopping condition of an algorithm is a predefined number of generations, summarized in Appendix A.

D. Performance Metrics

The inverted generational distance (IGD) and HV metrics are chosen to assess the performance of the algorithms. Both the IGD and HV metrics evaluate the convergence and diversity of a solution set simultaneously. A smaller IGD or a larger HV typically indicates better convergence and diversity. The HV metric is shown to be sensitive to the specification of the worst point, especially for irregular PF shapes. And the IGD metric tends to favor a set of solutions with a similar distribution to the reference set.

E. Sampling Methods

seeking to get conclusions in a more general perspective based on comprehensive experimental results, four sampling methods are chosen in our empirical study, i.e. random sampling, Monte Carlo sampling (MC) [36], Latin hypercube sampling (LHS) [37] and Sobol' sampling [38].

- *Random Sampling*: Random sampling randomly selects a point in noisy space without any predefined criteria.
- *Monte Carlo Sampling*: In Monte Carlo sampling, a mathematical model is developed to represent the system or process of interest, and random values are generated based on certain probability distributions.
- *Latin Hypercube Sampling*: Latin hypercube sampling (LHS) is a statistical sampling technique used to generate a set of representative samples from a multivariate distribution. The goal of LHS is to produce a more efficient sampling scheme than simple random sampling, while still maintaining randomness and minimizing bias.
- *Sobol' Sampling*: Sobol' sampling is designed to ensure the generated points are evenly distributed in the sample space, thereby minimizing the gaps or clusters of points that can occur with traditional random sampling. Sobol' sampling is particularly effective in high-dimensional spaces where random sampling can become inefficient or inaccurate due to the curse of dimensionality.

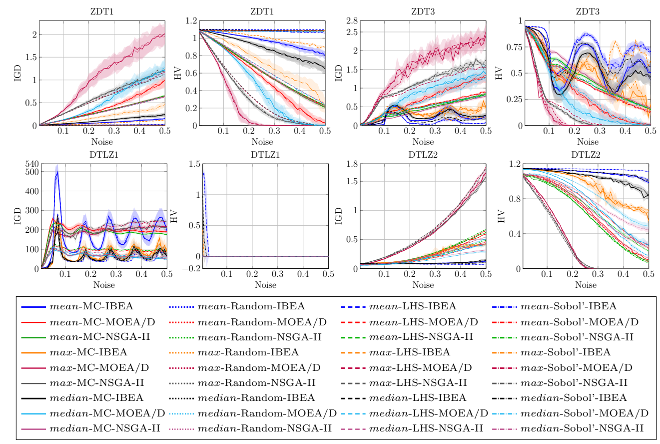

Fig. 1. Line charts of the trend of IGD and HV values with size of noise δ_i^{max} changing on ZDT1, ZDT3, DTLZ1 and DTLZ2 instances when sample size κ fixed as 30. The data are averaged over 30 random instances.

Note that in our experiment, if a solution $\mathbf{x}$ is outside the decision space after adding noise $\boldsymbol{\delta}$, it would be modified to the feasible domain boundary value, which can be written as:

$$x_i' = \begin{cases} a_i & \text{if } x_i + \delta_i < a_i \\ b_i & \text{if } x_i + \delta_i > b_i \\ x_i + \delta_i & \text{else} \end{cases}. \quad (2)$$

IV. Empirical Studies

We seek to answer the following three research questions (*RQ*s) through our experimental evaluation:

- *RQ1*: How does the size of noise effect the performance of noisy optimization?
- *RQ2*: How does the sample size effect the performance of noisy optimization?
- *RQ3*: How do the test instances and evolutionary algorithms effect the performance of noisy optimization?

A. Impact of Noise

1) ***Methods***: To address *RQ1*, this subsection aims at examining the influence of diverse levels of noise on the efficacy of three optimization algorithms, namely NSGA-II, MOEA/D and IBEA. In this subsection, sample size κ is set as 30, which will be studied in Section IV-B, and the maximum noise size $\delta_i^{max}(i = 1, \ldots, n)$ ranges from 0.005 to 0.5. To approximate the authentic fitness values, the mean, maximum and median values are utilized for each individual in this study. To ensure statistical significance, each experiment is repeated 30 times, and the corresponding results on ZDT1, ZDT3, DTLZ1 and DTLZ2 are shown in Fig. 1. The outcomes on other five test instances can be found in Appendix B.
2) ***Results and Analysis***: The line charts presented in Fig. 1 clearly demonstrate that the introduction of noise has a detrimental effect on the performance of MOEAs in terms of the IGD and HV metrics. The observed trends indicate that as the noise size increases, the values of the IGD metric exhibit an upward trend, whereas those

Manuscript 1037 submitted to 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Received April 23, 2023.

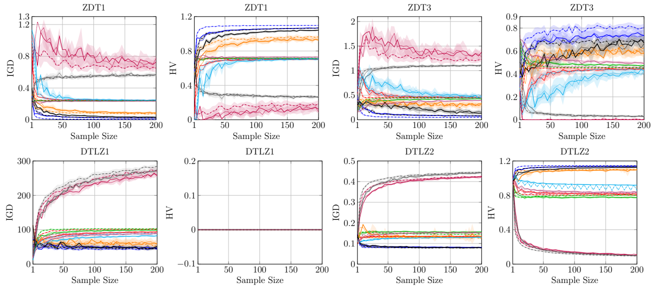

Fig. 2. Line charts of the trend of IGD and HV values with sample size κ changing on ZDT1, ZDT3, DTLZ1 and DTLZ2 instances when δ_i^{max} fixed as 0.2. The data are averaged over 30 random instances. The legend keep the same with Fig. 1.

of the HV metric tend to decrease overall. Specifically, for the ZDT1 test instance, the changes in the IGD and HV metrics are directly proportional to the noise size. However, for the ZDT2 and DTLZ1 test instances, the changes in the metrics are more nuanced, with a significant changed in metric values initially followed by a gradual plateauing as noise size increases. Conversely, the trend observed for DTLZ2 is opposite to that of ZDT2 and DTLZ1.

It is noteworthy that the DTLZ1 test instance appears to be highly sensitive to noise, with the metric values reaching a maximum of 500 when δ_i^{max} is 0.08. This suggests that the heterogeneity of the noise ranges in the decision space, and their corresponding effects on the objective space, may have a significant impact on the performance of MOEAs. Additionally, the HV metric values for the DTLZ1 test instance are zero, as the reference point in this experiment is set to 1 for each objective function, and the pareto front identified by the MOEAs under noisy conditions exceeds this range.

Response to *RQ1*: This subsection presents the findings of an empirical study that investigates the performance of MOEAs under varying size of noise. The results indicate that the efficiency of MOEAs deteriorates with increasing noise size when sample size is held constant. Moreover, it was observed that the impact of noise varies across different test instances, which may be attributed to the heterogeneity of noise ranges in the decision space and their corresponding effects on the objective space.

B. Impact of Sample Size

1) ***Methods***: In order to figure out the answer to *RQ2*, a set of experiments is conducted in which the sample size (κ) is varied from 1 to 200, while the maximum noise level (δ_i^{max}) is fixed at 0.2 for each test instance. The results of this experiment on ZDT1, ZDT3, DTLZ1 and DTLZ2 instances are depicted in Fig. 2, and additional results can be found in Appendix C. To further investigate the relationship between sample size and noise size on the performance of the IGD and HV values, we present heat maps on ZDT1 using mean values in Fig. 3 and Fig. 4, with further results available in Appendix D.

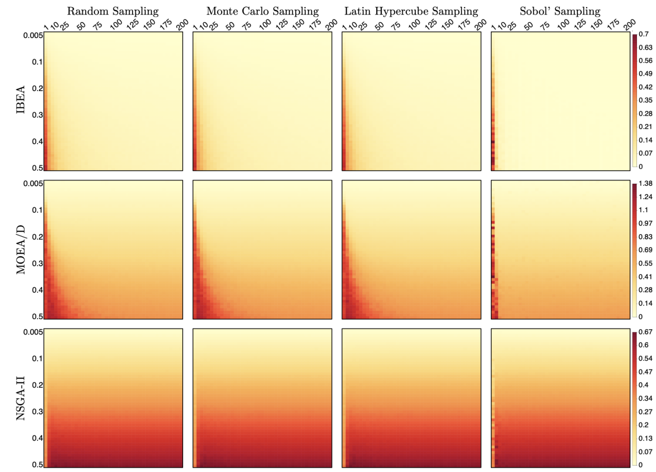

Fig. 3. IGD values found by three algorithms with more than $2,000$ combinations of δ_{max} and κ on ZDT1 instance using mean values.

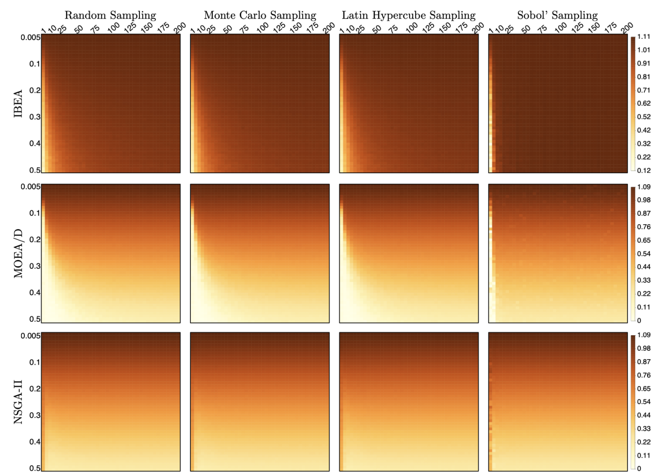

Fig. 4. HV values found by three algorithms with more than $2,000$ combinations of δ_{max} and κ on ZDT1 instance using mean values.

2) ***Results and Analysis***: From Fig. 2, it is clear to see that the performance of noisy optimization is relatively insensitive to the sample size, as long as the sample size is not too small. specifically, for fixed noise size, the metric values exhibit significant fluctuations κ is very small ($\kappa \leq 20$), but then become stable for larger sample size. Furthermore, the line charts suggest that the choise of sample type does not have a substantial impact on the algorithm's performance.

The presented heat maps in figures 3 and 4 corroborate our earlier observations, revealing that the robustness of the optimization algorithms is primarily impacted by the max noise range, while exhibiting a relatively minor dependence on the sample size. Furthermore, the comparative analysis of the performance across different sample types confirms a consistent trend, though with variations in the sensitivity to noise size and sample size across different algorithms. More specifically, IBEA shows a consistent performance in terms of IGD and HV

Manuscript 1037 submitted to 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Received April 23, 2023.

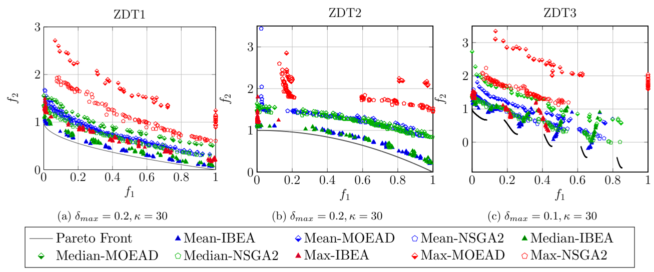

Fig. 5. Plots of population distributions optimized by IBEA, MOEA/D and NSGA-II using random sampling estimated by mean, max and median values on ZDT1, ZDT2 and ZDT3 instances.

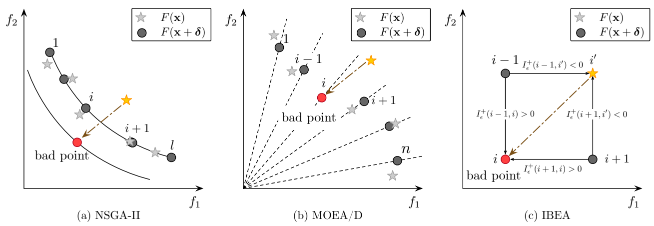

Fig. 6. Examples to reveal the impact of bad point in NSGA-II, MOEA/D and IBEA.

values, except in instances with particularly small sample size. In contrast, MOEA/D exhibits a relatively higher demand for sample size with increasing noise size. Additionally, our results indicate that IBEA outperforms the other two algorithms, while NSGA-II performs relatively poorly.

Response to *RQ2*: Here we empirically investigate the correlation between noise size and sample size. In a nutshell, the performance of noisy optimization is relatively insensitive to the sample size and sample type.

C. Impact of Test Instances and Evolutionary Algorithms

1) ***Methods***: Based on our analyses presented in Section IV-A and Section IV-B, we conclude that various optimization algorithms exhibit differing levels of robustness to noise. Notably, IBEA exhibits a higher degree of robustness than other algorithms. Specifically, we present plots of the optimized results obtained on ZDT1 $\sim$ ZDT3 using different algorithms, to explore the relationships between them. Additionally, we introduce the concept of *bad point* to provide insight into the impact of noise on various evolutionary algorithms.
2) ***Results and Analysis***: Our investigation of the population distributions presented in Fig. 5 reveals that the populations optimized by IBEA are closer to the pareto front, while MOEA/D performs the worst across all three test instances. Additionally, we observe that when estimating fitness values using the maximum value method, the results are more conservative than when using mean or median methods.
To explore the influence of noise on different evolutionary algorithms, we introduce the concept of *bad point*, defined as a poor solution assigned with a good fitness value. We investigate the impact of *bad points* on NSGA-II, MOEA/D, and IBEA, and illustrate our findings in Fig. 6. In this figure, we use star markers to represent individuals in the population, which are replaced by circle points in the objective space when evaluated in a noisy environment. And the colored markers in the figure are *bad points*. Our finding show that the varying degrees of robustness among different algorithms can be attributed to the their distinct optimization mechanisms. Specifically, NSGA-II (Fig. 6(a)) may be dominated by a *bad point* in a noiseless environment, resulting in the persistence of inferior solutions into the next generation and the rejection of superior solutions. In contrast, MOEA/D (Fig. 6(b)) uses a random selection of two individuals for crossover operator, making it less susceptible to the influence of *bad points*. IBEA, as an indicator-based EA, takes all individuals in the population into consideration when calculating the fitness function. As illustrated in Fig. 6(c), when a *bad point* moves to the red circle point, it does not affect fitness values of the other two individuals in the population.

Response to *RQ3*: Our empirical study in this subsection demonstrates that different evolutionary algorithms exhibit varying degrees of robustness in the presence of noise in decision variables. And to illustrate the reason behind this variation, we introduce the concept of *bad point*.

V. Conclusion

In this paper, we study the effects of some resampling strategies in noisy multi-objective evolutionary optimization. The experiments designed in this paper are devoted to exploiting the factors that influence the result of MOEAs when noise are presented in decision variables. To sum up, the size of noise, test instances and algorithms have significant impacts on the performance of RMOEAs, while the sample size and type have little influences. There are three core takeaways from our experiments.

- The efficiency of MOEAs degenerates with the increase in noise size.
- The impact of noise size varies across different test instances, this could be attributed to the heterogeneity of noise ranges in the decision space and their corresponding effects on the objective space.
- We propose the concept of *bad point* to illustrate the effect of noise on different optimization algorithms.

In future, we seek to make further research about MOPs in noisy environment as follows. 1) reduce the influence of high level noise on optimization performance; 2) effects of more complex noisy situation, like considering constraints; 3) researches on realistic applications, e.g. noise in hyper parameter optimization.

Acknowledgment

This work was supported by UKRI Future Leaders Fellowship (MR/S017062/1), EPSRC (2404317), NSFC (62076056), Royal Society (IES/R2/212077) and Amazon Research Award.

APPENDIX

The supplementary material is in https://tinyurl.com/62v24cm2.

REFERENCES

[1] C. A. C. Coello, "Evolutionary multi-objective optimization: a historical view of the field," *IEEE Comput. Intell. Mag.*, vol. 1, no. 1, pp. 28–36, 2006.

[2] A. A. Montaño, C. A. C. Coello, and E. Mezura-Montes, "Multiobjective evolutionary algorithms in aeronautical and aerospace engineering," *IEEE Trans. Evol. Comput.*, vol. 16, no. 5, pp. 662–694, 2012.

[3] A. Ponsich, A. L. Jaimes, and C. A. C. Coello, "A survey on multiobjective evolutionary algorithms for the solution of the portfolio optimization problem and other finance and economics applications," *IEEE Trans. Evol. Comput.*, vol. 17, no. 3, pp. 321–344, 2013.

[4] H. R. Cheshmehgaz, M. I. Desa, and A. Wibowo, "A flexible three-level logistic network design considering cost and time criteria with a multi-objective evolutionary algorithm," *J. Intell. Manuf.*, vol. 24, no. 2, pp. 277–293, 2013.

[5] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, 2002.

[6] M. Kim, T. Hiroyasu, M. Miki, and S. Watanabe, "SPEA2+: improving the performance of the strength pareto evolutionary algorithm 2," in *PPSN'04: Proc. of the 2004 Parallel Problem Solving from Nature*, ser. Lecture Notes in Computer Science, vol. 3242. Springer, 2004, pp. 742–751.

[7] E. Zitzler and S. Künzli, "Indicator-based selection in multiobjective search," in *PPSN'04: Proc. of the 2004 Parallel Problem Solving from Nature*, ser. Lecture Notes in Computer Science, vol. 3242. Springer, 2004, pp. 832–842.

[8] J. Bader and E. Zitzler, "Hype: An algorithm for fast hypervolume-based many-objective optimization," *Evol. Comput.*, vol. 19, no. 1, pp. 45–76, 2011.

[9] N. Beume, B. Naujoks, and M. T. M. Emmerich, "SMS-EMOA: multiobjective selection based on dominated hypervolume," *Eur. J. Oper. Res.*, vol. 181, no. 3, pp. 1653–1669, 2007.

[10] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Trans. Evol. Comput.*, vol. 11, no. 6, pp. 712–731, 2007.

[11] K. Li, K. Deb, Q. Zhang, and S. Kwong, "An evolutionary many-objective optimization algorithm based on dominance and decomposition," *IEEE Trans. Evol. Comput.*, vol. 19, no. 5, pp. 694–716, 2015.

[12] M. Wu, K. Li, S. Kwong, Q. Zhang, and J. Zhang, "Learning to decompose: A paradigm for decomposition-based multiobjective optimization," *IEEE Trans. Evol. Comput.*, vol. 23, no. 3, pp. 376–390, 2019.

[13] K. Deb and H. Gupta, "Introducing robustness in multi-objective optimization," *Evol. Comput.*, vol. 14, no. 4, pp. 463–494, 2006.

[14] Z. He, G. G. Yen, and J. Lv, "Evolutionary multiobjective optimization with robustness enhancement," *IEEE Trans. Evol. Comput.*, vol. 24, no. 3, pp. 494–507, 2020.

[15] R. Liu, Y. Li, H. Wang, and J. Liu, "A noisy multi-objective optimization algorithm based on mean and wiener filters," *Knowl. Based Syst.*, vol. 228, p. 107215, 2021.

[16] R. Liu, N. Li, and F. Wang, "Noisy multi-objective optimization algorithm based on gaussian model and regularity model," *Swarm Evol. Comput.*, vol. 69, p. 101027, 2022.

[17] M. Asafuddoula, H. K. Singh, and T. Ray, "Six-sigma robust design optimization using a many-objective decomposition-based evolutionary algorithm," *IEEE Trans. Evol. Comput.*, vol. 19, no. 4, pp. 490–507, 2015.

[18] Y. Jin and J. Branke, "Evolutionary optimization in uncertain environments-a survey," *IEEE Trans. Evol. Comput.*, vol. 9, no. 3, pp. 303–317, 2005.

[19] P. D. Stroud, "Kalman-extended genetic algorithm for search in non-stationary environments with noisy fitness evaluations," *IEEE Trans. Evol. Comput.*, vol. 5, no. 1, pp. 66–77, 2001.

[20] H.-G. Beyer, "Evolutionary algorithms in noisy environments: Theoretical issues and guidelines for practice," *Comput. Methods Appl. Mech. Eng.*, vol. 186, no. 2-4, pp. 239–267, 2000.

[21] M. Rattray and J. Shapiro, "Noisy fitness evaluation in genetic algorithms and the dynamics of learning," in *Proc. of the 4th Workshop on Foundations of Genetic Algorithms. San Diego, CA, USA, August 5 1996*, R. K. Belew and M. D. Vose, Eds. Morgan Kaufmann, 1996, pp. 117–139.

[22] T. Park and K. R. Ryu, "Accumulative sampling for noisy evolutionary multi-objective optimization," in *GECCO'11: Proc. of the 2004 Genetic and Evolutionary Computation Conference*, N. Krasnogor and P. L. Lanzi, Eds. ACM, 2011, pp. 793–800.

[23] P. Rakshit, A. Konar, and S. Das, "Noisy evolutionary optimization algorithms - A comprehensive survey," *Swarm Evol. Comput.*, vol. 33, pp. 18–45, 2017.

[24] J. E. Fieldsend and R. M. Everson, "The rolling tide evolutionary algorithm: A multiobjective optimizer for noisy optimization problems," *IEEE Trans. Evol. Comput.*, vol. 19, no. 1, pp. 103–117, 2015.

[25] L. T. Bui, H. A. Abbass, and D. Essam, "Fitness inheritance for noisy evolutionary multi-objective optimization," in *GECCO'05: Proc. of the 2005 Genetic and Evolutionary Computation Conference*, H. Beyer and U. O'Reilly, Eds. ACM, 2005, pp. 779–785.

[26] P. J. Darwen and J. B. Pollack, "Co-evolutionary learning on noisy tasks," in *CEC'99: Proc. of the 1999 Congress on Evolutionary Computation*. IEEE, 1999, pp. 1724–1731.

[27] J. Teich, "Pareto-front exploration with uncertain objectives," in *EMO'01: Proc. of the 2001 Evolutionary Multi-Criterion Optimization*, ser. Lecture Notes in Computer Science, E. Zitzler, K. Deb, L. Thiele, C. A. C. Coello, and D. Corne, Eds., vol. 1993. Springer, 2001, pp. 314–328.

[28] P. Rakshit and A. Konar, "Extending multi-objective differential evolution for optimization in presence of noise," *Inf. Sci.*, vol. 305, pp. 56–76, 2015.

[29] D. Buche, P. Stoll, R. Dornberger, and P. Koumoutsakos, "Multiobjective evolutionary algorithm for the optimization of noisy combustion processes," *IEEE Trans. Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 32, no. 4, pp. 460–473, 2002.

[30] C. K. Goh and K. C. Tan, "An investigation on noisy environments in evolutionary multiobjective optimization," *IEEE Trans. Evol. Comput.*, vol. 11, no. 3, pp. 354–381, 2007.

[31] I. R. Meneghini, F. G. Guimarães, and A. Gaspar-Cunha, "Competitive coevolutionary algorithm for robust multi-objective optimization: The worst case minimization," in *CEC'16: Proc. of the 2016 IEEE Congress on Evolutionary Computation*. IEEE, 2016, pp. 586–593.

[32] A. Syberfeldt, A. H. C. Ng, R. I. John, and P. R. Moore, "Evolutionary optimisation of noisy multi-objective problems using confidence-based dynamic resampling," *Eur. J. Oper. Res.*, vol. 204, no. 3, pp. 533–544, 2010.

[33] M. Basseur and E. Zitzler, "A preliminary study on handling uncertainty in indicator-based multiobjective optimization," in *Applications of Evolutionary Computing, EvoWorkshops 2006: EvoBIO, EvoCOMNET, EvoHOT, EvoIASP, EvoINTERACTION, EvoMUSART, and EvoSTOC*, ser. Lecture Notes in Computer Science, vol. 3907. Springer, 2006, pp. 727–739.

[34] E. Zitzler, K. Deb, and L. Thiele, "Comparison of multiobjective evolutionary algorithms: Empirical results," *Evol. Comput.*, vol. 8, no. 2, pp. 173–195, 2000.

[35] 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. C. Jain, and R. R. Goldberg, Eds. Springer, 2005, pp. 105–145.

[36] W. K. Hastings, "Monte carlo sampling methods using markov chains and their applications," 1970.

[37] M. Stein, "Large sample properties of simulations using latin hypercube sampling," *Technometrics*, vol. 29, no. 2, pp. 143–151, 1987.

[38] I. Sobol', "On the distribution of points in a cube and the approximate evaluation of integrals," *USSR Computational Mathematics and Mathematical Physics*, vol. 7, no. 4, pp. 86–112, 1967.