



Bachelor Thesis Project (BTP) Report

Shivam Pandey

Roll Number: 160010003

Month of Submission: May 2020

Supervisor: Prof. Somenath Biswas, Computer Science and Engineering

*This report is submitted towards partial fulfillment of the requirements for the award of the
Bachelor of Technology (B.Tech) degree in Computer Science and Engineering at the
Indian Institute of Technology Goa.*

Acknowledgements

To begin with, I would like to thank God and my family for being my constant source of inner strength and support. I can achieve nothing without their blessings.

I would like to specially thank *Prof. Somenath Biswas* for all his guidance on my Bachelor Thesis Project (BTP). I have learnt a lot from his wisdom and scientific expertise in the field of Computer Science. I also express my sincere gratitude to our Director, *Prof. B. K. Mishra* and our Dean, *Prof. Sachin D. Kore*, for always being there as pillars of support and source for motivation. I would also like to thank all professors at IIT Goa for encouraging and promoting me in all aspects. Lastly, I really thank my friends *Abhishek Varghese*, *Shreyash Tripathi*, *Apurv Varshney*, *Shivam Kumar* and *Jaruko Toppo* for all the informative talks and exchange of ideas, which also helped me a lot in building the concepts and working at this project. This project would not have been completed without the support of all the aforementioned people and I am truly indebted to all of them.

Abstract

Optimization is the practice of finding workable solutions to a given problem, until no new better solution is found. Multi-objective evolutionary algorithms have been extensively studied by many researchers for about a century now. In this research project, we try to delve deep into one of the most popular evolutionary algorithm, namely Non-dominated Sorting Genetic Algorithm-II (NSGA-II). We commence by defining the problem statement. Thereafter, we will talk about the principal features of NSGA-II, mainly about the diversity of solutions obtained from the algorithm. Next, we try to further explore NSGA-II and bring some modifications to the existing algorithm by considering diversity in a more detailed manner. Then, we finally compare the results obtained from our modified algorithm and the results obtained from traditional NSGA-II. Some future directions are discussed in the last section, which still remain unexplored.

Study of Multi-Objective Optimization using Evolutionary Algorithms

Shivam Pandey

Introduction & Problem Statement

Optimization is a method of finding and comparing feasible solutions, till no better solution can be found. A solution is termed as good or bad on the basis of an objective, which in real life can be the cost of manufacturing, fuel efficiency of a car, profit earned in business, etc. In real life, we usually come across optimization problems having several objectives. A multi-objective optimization problem (MOOP) entails more than one objective to be optimized, often conflicting objectives, and has a pareto-optimal set of solutions. Pareto-optimal set is defined as the *non-dominated*¹ set of the entire feasible search space. Multi-Objective Evolutionary Algorithms (MOEAs) seek to approximate the *pareto-front*² of the MOOP, by evolving a population of solutions, inspired from the biological evolution using selection, reproduction, recombination and mutation. Classical optimization methods can at best find one solution in one simulation run, thereby making those methods inconvenient to solve MOOPs. Evolutionary approaches to MOOPs, on the other hand, are capable of searching for multiple optimal solutions concurrently in a single run, due to their population-approach. This ability of an evolutionary algorithm (EA) to find multiple optimal solutions in one simulation run makes EAs unique in solving MOOPs.

My Bachelor Thesis Project (BTP) work focuses on studying and exploring existing MOEAs like Non-dominated Sorting Genetic Algorithm-II (NSGA-II) [2], Pareto Archived Evolution Strategy (PAES) [4], and formulating possible improvement ideas along with combining the best ideas from existing MOEAs to propose a better algorithm, in terms of *quality of solutions*³, running time of the algorithm, or preferably both.

¹Among a set of solutions P , the non-dominated set of solutions P' , are those that are not dominated by any member of the set P .

²The curve formed in the objective space by joining these pareto-optimal solutions is known as a **pareto front**.

³The MOEA converges as **close** as possible to the pareto front and maintains a **good diversity** amongst the solutions.

Literature Review

^[1] In the initial phase of my work, the formulation of a MOOP was studied. Most MOEAs use the concept of *dominance* and *pareto-optimality* in their search. A solution x is said to dominate the other solution y , if both the following conditions are true:

- The solution x is no worse than y in all objectives.
- The solution x is strictly better than y in at least one objective.

After studying the terminologies under the literature of EAs, the general structure of a genetic algorithm (GA) was studied and how they are implemented to solve an optimization problem, may it be a single-objective problem or a multi-objective problem.

^[2] An existing elitist MOEA, particularly known as NSGA-II, with computational complexity $O(MN^2)$ (here, M = number of objectives, N = population size), was studied and implemented on a two-objective optimization benchmark problem: Schaffer's study (SCH)^[3] (here, we took a minimization problem). The two objective functions considered for the minimization problem were x^2 and $(x-2)^2$, where the decision variable space : $x \in (-\infty, +\infty)$. It is evident that the optimal set of solutions for this problem would be $x \in [0, 2]$, hence we would expect the algorithm to converge to the pareto-front (non-dominated solutions) in the objective space after sufficient number of generations are evolved. In fig.1 and fig.2 shown below, the NSGA-II was run for about 3000 generations on a population of 40 individuals.

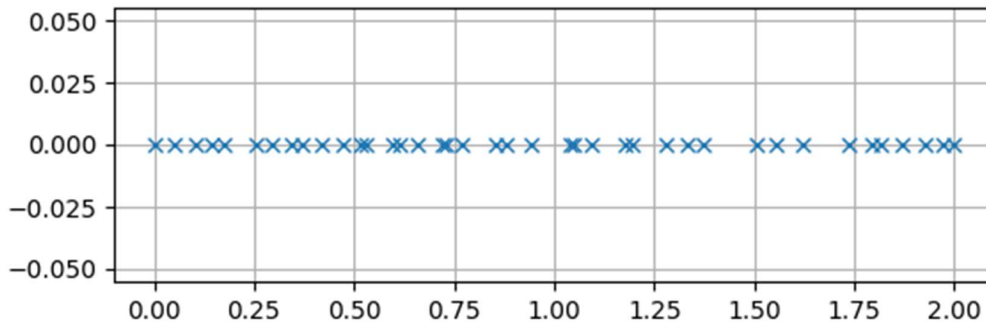


Fig. 1 Visualization of Decision Variable Space (or Data Space)

In Fig.1, each of the crosses denote one individual in the final population the algorithm outputs, which is evidently close to the optimal solution set, i.e. $x \in [0, 2]$.

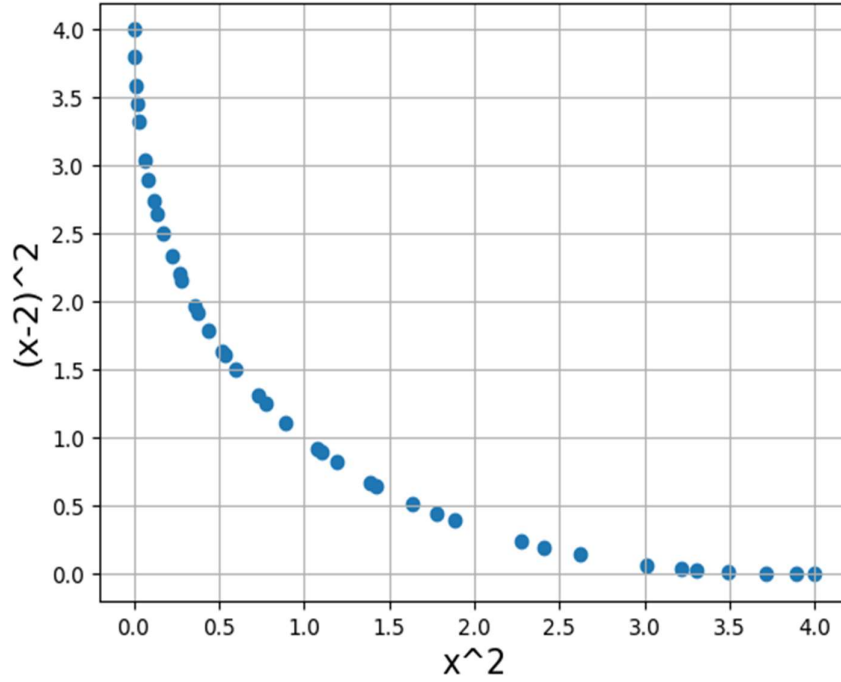


Fig.2 Visualization of Objective Space

In Fig. 2, a reasonably good approximation to the pareto-front was obtained in the objective space, along with reasonable diversity which is expected from NSGA-II. Hence, the run of NSGA-II was successful on this particular MOOP. With this successful execution of NSGA-II, a concrete understanding was developed about the working of MOEAs.

Experimental Analysis & Improvement Idea

The major experimental analysis throughout this survey of NSGA-II lies at considering the diversity of solutions in both, objective space and data space. In MOEAs, one of the most important factor which contributes to the obtained set of non-dominated solutions to be called “good” is *diversity*⁴. It should be noted that NSGA-II^[2] does not take into account the diversity in *data space*, but only in *objective space*. This motivated further study to consider the diversity (individually and simultaneously) in both, data space and objective space, and compare the results with traditional NSGA-II. The expectation is that if we perform the evolution of each generation by taking into account the crowding distances in both, data space and objective space, then we might obtain a good amount of uniformity in both the spaces.

⁴*Diversity* signifies how uniformly the solutions are spread throughout the search space.

Before we move towards the implementation of the algorithm, we define the metric which will be used to measure the diversity in a given search space. With the use of this metric and graphical visualizations acquired from simulation run, we will give a detailed analysis of the results obtained.

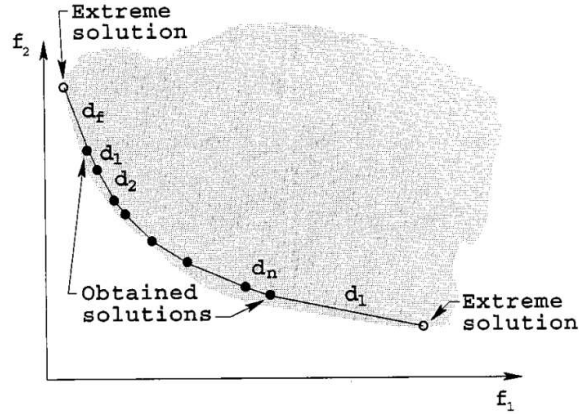


Fig.3 Diversity Metric Δ (Courtesy [2])

The diversity metric Δ calculates the magnitude of spread achieved in the obtained solutions. The following metric ^[2] is used to measure the non-uniformity in the distribution:

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

It should be noted that for the most uniformly spread out set of non-dominated solutions, the numerator of Δ would become zero, and thus, $\Delta = 0$. For any other distribution, we will obtain $\Delta > 0$. Please note that Δ will always be non-negative, which is self-evident from its mathematical expression. So, it is to be realized that the more diverse (uniformly spread) the set of solutions will be, the closer the value of Δ will be to zero. Hence, smaller the value of Δ , more diverse is the solution set.

Now, to run our algorithm, we shall consider a slightly more complex MOOP, which is called as the *Min-Ex* problem. It is a two-objective, two-variable minimization problem, defined below:

Min-Ex Problem

$$\begin{aligned} \text{Minimize} \quad & f_1(x_1, x_2) = x_1 \\ \text{Minimize} \quad & f_2(x_1, x_2) = (1 + x_2) / x_1 \\ \text{subject to} \quad & 0.1 \leq x_1 \leq 1 \quad \& \quad 0 \leq x_2 \leq 5 \end{aligned}$$

Simulation Run

We will run three different algorithms on the *Min-Ex* problem, which includes the traditional NSGA-II and the other two algorithms, which are slightly modified versions of the original NSGA-II. We define the following three names for three different algorithms to be implemented for experimentation:

- *Version-I* → Traditional NSGA-II algorithm.
- *Version-II* → Modified algorithm, which will evaluate the crowding distances only in *data space*.
- *Version-III* → Modified algorithm, which will evaluate the crowding distances in both, *data space* and *objective space* (here, we add up the crowding distances in data space and objective space for each individual.).

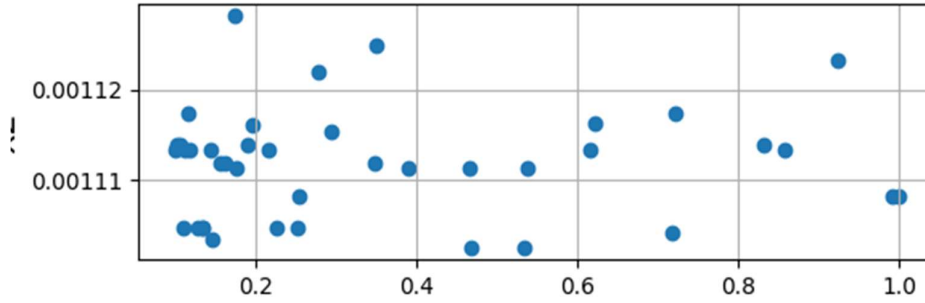


Fig.4A Data Space (Run of *Version-I*)

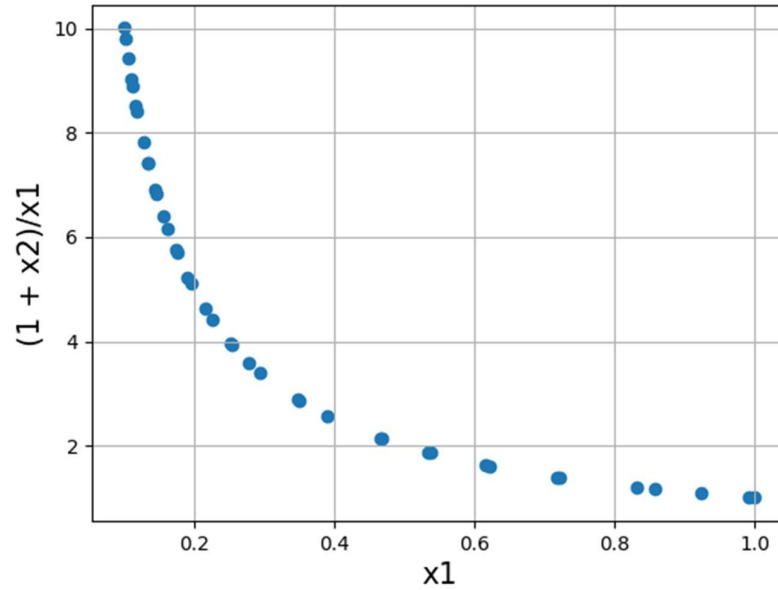


Fig.4B Objective Space (Run of *Version-I*)

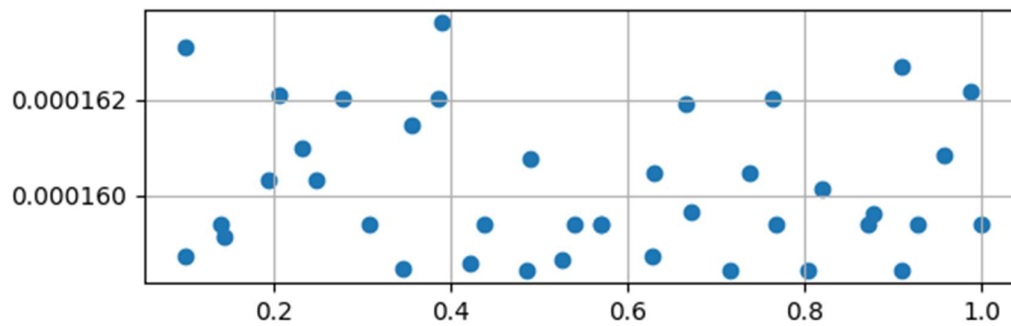


Fig.5A Data Space (Run of *Version-II*)

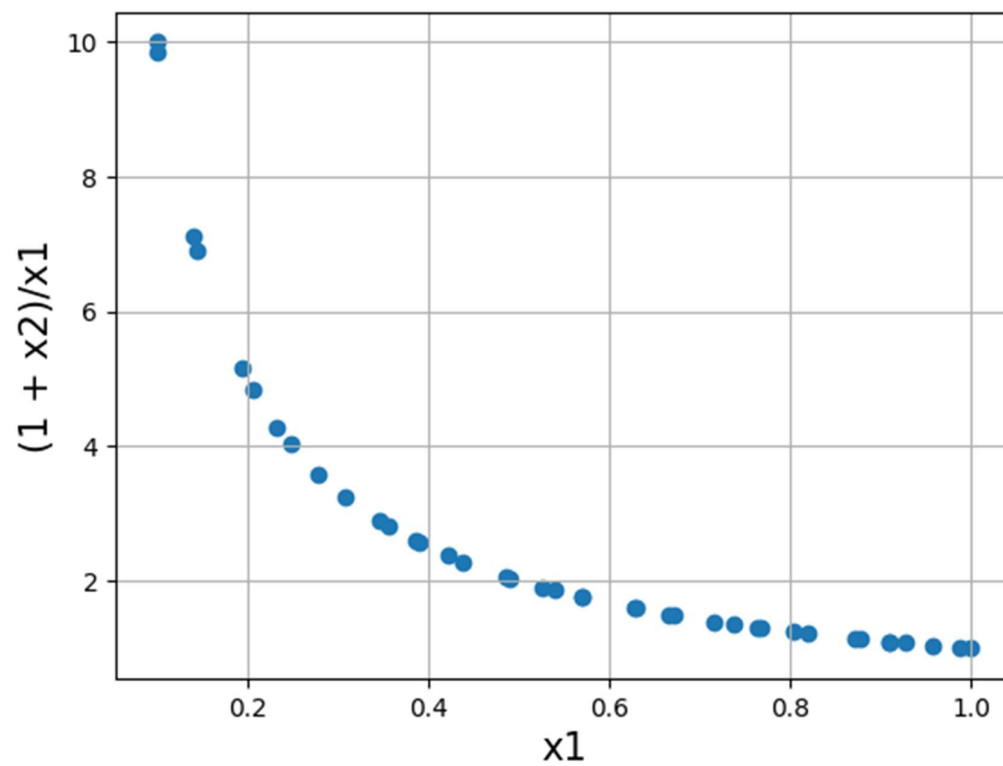


Fig.5B Objective Space (Run of *Version-II*)

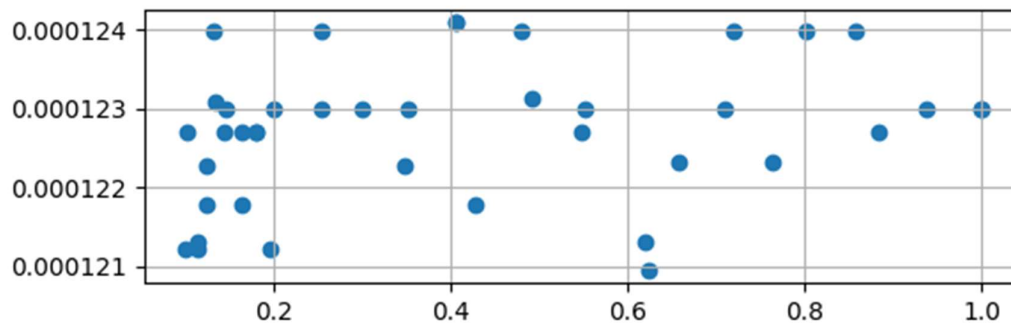


Fig.6A Data Space (Run of *Version-III*)

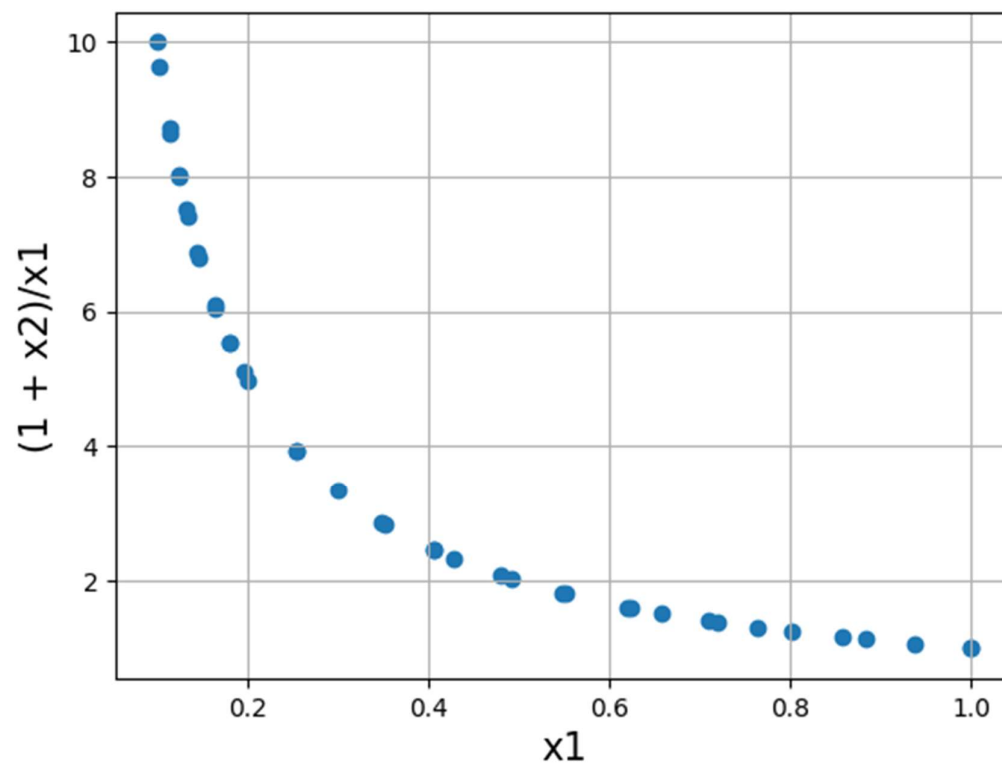


Fig.6B Objective Space (Run of *Version-III*)

Discussion and Results

We ran each version of the algorithm five times over the *Min-Ex* problem and then calculated the mean diversity metric in both the spaces for each version. Please note that we have used the population size of 40, mutation rate of 0.2 and evolved for 3000 generations in a single simulation run.

The following table shows the values of *mean diversity metric in objective space* (Δ_o) and *mean diversity metric in data space* (Δ_d) :

Table 1: Mean diversity metric in data space and objective space when run on *Min-Ex* problem with a population of 40 individuals and evolving for 3000 generations.

Min-Ex	Δ_d	Δ_o
<i>Version-I</i>	0.9640 (Fig.4A)	0.6971 (Fig.4B)
<i>Version-II</i>	0.5556 (Fig.5A)	1.084 (Fig.5B)
<i>Version-III</i>	0.8492 (Fig.6A)	0.8798 (Fig.6B)

From table 1, it can be diagnosed that the diversity metrics Δ_d and Δ_o took quite promising values. Firstly, we will only discuss about the set of solutions obtained in data space from all three versions. We would have expected that the most diverse data space is obtained from the run of *Version-II*, since it takes into account the crowding distances in data space only. This is very much evident from Fig.5A and the value of $\Delta_d(\text{Version-II})$. Observing Fig.4A, Fig.5A, Fig.6A, it is clearly seen that the most uniformly spread out data space is of Fig.5A, followed by Fig.6A and then Fig.4A. Mathematically, this can be seen from the following inequalities:

$$\Delta_d(\text{Version-II}) = 0.5556 < \Delta_d(\text{Version-III}) = 0.8492 < \Delta_d(\text{Version-I}) = 0.9640$$

Now, if we do similar analysis on the diversity of objective space obtained from the three versions, we again acquire our expected result. It is clearly evident from Fig.4B, Fig.5B, Fig.6B, that the most diverse pareto-front is obtained from the run of *Version-I* (Fig.4B), and the most non-uniformity is seen from the run of *Version-II* (Fig.5B). Mathematically, this can be seen from following inequalities:

$$\Delta_o(\text{Version-I}) = 0.6971 < \Delta_o(\text{Version-III}) = 0.8798 < \Delta_o(\text{Version-II}) = 1.084$$

Conclusion and Future Directions

After performing the above experimental analysis and survey, we recognize that quite promising results were obtained. The principal improvement idea which we discussed throughout this report was to achieve diversity in both, *data space* and *objective space*. Since, the traditional NSGA-II algorithm does not take into account the diversity of data space (or, decision variable space), so we tried slight modifications to original NSGA-II and implemented it on a MOOP, namely *Min-Ex* problem.

Our main expectation was to achieve the diversity in data space, along with objective space. The implementation of the *Version-III* proved to be quite rewarding, since it showed diversity in both the spaces. It is to be noted that the time complexity of *Version-III* remains the same as original NSGA-II. We conclude that this modification can be used in place of the traditional NSGA-II to achieve a more diverse set of solutions in both, data space and objective space.

There are more aspects which can be looked upon to improve the quality of solutions. We can take into account the idea of “archives” [4]. Sometimes, it is useful to keep track of good solutions that have been visited in the past, even if they are no longer active in offspring generation. Then, the reproducing population can perform much more active exploration, instead of being filled up with old solutions. This idea can be further incorporated in the modification of *Version-III*, and is yet to be explored.

References

- [1] K. Deb, Multiobjective Optimization Using Evolutionary Algorithms. Chichester, U.K.: Wiley, 2001.
- [2] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, “A fast and elitist multi-objective genetic algorithm: NSGA-II”, Proc. Parallel Problem Solving from Nature VI, 2000
- [3] J. D. Schaffer, “Multiple objective optimization with vector evaluated genetic algorithms,” in Proceedings of the First International Conference on Genetic Algorithms, J. J. Grefenstette, Ed. Hillsdale, NJ: Lawrence Erlbaum, 1987, pp. 93–100.
- [4] J. Knowles and D. Corne, “The Pareto archived evolution strategy: A new baseline algorithm for multiobjective optimization,” in Proceedings of the 1999 Congress on Evolutionary Computation. Piscataway, NJ: IEEE Press, 1999, pp. 98–105.