

Uncertainty Modelling for Project Portfolio Problem using Interval Analysis

Fausto Balderas¹, Eduardo Fernandez², Claudia Gómez³, Gilberto Rivera⁴, Laura Cruz-Reyes³, Nelson Rangel V⁵.

¹ *Computer Science in the Graduate Division, National Mexican Institute of Technology/Tijuana Institute of Technology, 22500, Baja California, Mexico.
fausto.balderas@itcm.edu.mx*

² *Faculty of Civil Engineering, Autonomous University of Sinaloa, 80040, Sinaloa, Mexico.
eddyf@uas.edu.mx*

³ *Postgraduate & Research Division, National Mexican Institute of Technology/Madero Institute of Technology, 89440, Tamaulipas, Mexico.
claudia.gomez@itcm.edu.mx, lauracruzreyes@itcm.edu.mx,
automatas2002@yahoo.com.mx*

⁴ *Institute of Engineering and Technology, Autonomous University of Juarez City, 32310, Chihuahua, Mexico.
gilberto.rivera@uacj.mx*

⁵ *Catedrático CONACyT, División de Estudios de Posgrado e Investigación, Tecnológico Nacional de México, Instituto Tecnológico de Ciudad Madero
nrangelva@conacyt.mx*

Abstract. The problem of selection of project portfolio is often approached by organizations. The decision maker, along with the decision analyst must often provide information about the portfolio values, that most of the time come with uncertainty. An interval number is a set of numbers, which the exact value is not known. In this paper, we apply the interval approach to the problem of the project portfolio, modeling the imprecision in resources available, projects and the total resource, as well as restrictions relevant to the problem. The experiment is performed on two fronts: uncertainty modeling approach with interval analysis and without the interval approach in order to compare results. The results show that modeling the uncertainty and imprecision with interval numbers, were able to overcome results to those found without the application of interval approach.

Key words: project portfolio selection, uncertainty, interval approach.

1. Introduction

The decision analysis for portfolios (DAP) can be defined as a set of theories, methods and practices that help the actors of the decision-making process select a group options from a very large set of them, using mathematical modeling, and considering relevant restrictions, preferences, uncertainty and imprecision [12]

Let us suppose we have a set A composed of N projects, each one described by estimations of its impacts and resource consumption. A portfolio is a subset of A that can be modeled as a vector of binary components $X = \langle x_1, x_2, \dots, x_n \rangle$, where the value 1 of the component x_i indicates that the i -th project will be financed.

The contribution of projects to portfolio measures is often not accurately known, that is, there exists imprecision due to lack of knowledge about future states of nature (probabilistic uncertainty), or due to a simple lack of information (vagueness)

The use of multi-objective optimization techniques is necessary, which have been addressed by Multi-Objective Evolutionary Algorithms (MOEAs). The MOEAs dealt simultaneously with a set of possible solutions, allowing them to obtain an approximation to the Pareto frontier [2]. One of the algorithms mainly used to solve multi-objective problems is the Non-Dominated Sorting Genetic Algorithm II (NSGA-II, [4]), which has gained much popularity in solving problems efficiently and to a low computational cost.

Interval analysis is a method originated independently by Sunaga in [14] and Moore in [8] and developed ever since the 1950s by a score of mathematicians as an approach to putting bounds on rounding errors in mathematical computations. It represents a numerical quantity as an interval within which the true value of the quantity lies.

This paper intends to contribute in the study of the problem of project selection using mathematical modeling integrated with interval analysis, and considering relevant constraints, uncertainty, and imprecision, to deal with the project cost, and the total of the available resources, six experiments are proposed that model different reaction attitudes that the DM when facing uncertainty.

2. An overview of different approaches for modeling uncertainty in the project portfolio problem.

Uncertainty is known as the lack of knowledge about what will happen in the future. The term may be used in different situations, but his usage is common in the statistical and economic field, where particular circumstances make it impossible to produce an accurate diagnosis of what will happen. Uncertainty is also applied in the decision-making process. Two kinds of uncertainty can be distinguished:

Firstly, the uncertainty associated with future uncertain states of nature, which cause variability in the benefits of the projects. Secondly, the imprecision associated with the vagueness; the uncertain knowledge of the non-stochastic character.

The probabilistic modeling has been used, mainly to model the variability of project impacts; the majority of the works have presented it introducing additional criteria that try to minimize the risk at a certain extent, in the multi-objective optimization problem. Various papers differ from the way in which this measure is defined [e.g., 1, 5, 7, 9, 10, 13].

When evaluating projects, the amount of resources sufficient to realize the project is not known exactly and the budget is frequently overestimated by the DM [6]. The DM often feels more comfortable expressing the imprecision or qualitative judgments than establishing a probability distribution function or a measure of risk. The probability distributions have been used in project portfolios to model the variability of the impacts of the projects, but these have not been used (at least in the work that we know) for modeling variability in the resources that the project consumes, or in the available resources.

3. Uncertainty in the portfolio problem

A portfolio is represented as a one-dimensional vector $\vec{X} = \langle x_1, x_2, \dots, x_N \rangle$; of size N . To give values to x_i we will applied the well-known model for assigning 0-1, where x is the indicator variable of the portfolio, if $x_i = 1$, represents that the project i is supported inside the portfolio, otherwise $x_i = 0$.

Let's consider N projects that fulfill minimal requirements of acceptability that compete for financing. An important element of the portfolio problems is the interaction between the projects that can be in the sense of the benefits or in the sense of resources that are consumed, but in this paper, we will consider only independent projects, i.e. we will assume that there is no synergy of benefits or resources.

P_G represents the total amount of available financial resources. J is the total number of project objectives. The information about a project will be in a vector form, Equation (1):

$$\vec{p} = \langle c_i, o_1, \dots, o_j \rangle, \quad (1)$$

Where the first element, represent the project cost (c_i) and the elements of the remaining columns represent contributions to the project objectives (o_j), that will be given as interval numbers.

Once the projects that are in the portfolio are known, the cost associated with the finance of X is denoted as C_l and, under the premise of non-interaction of resources, it can be calculated using Equation (2) that represents the sum of the costs of all the proposals favored in the X portfolio:

$$C_l = \sum_{i=1}^N c_i * x_i; i = 1, 2, \dots, N. \quad (2)$$

Under the condition of assumed independence, the integrated contribution of all the projects in the portfolio X to the j -th objective, can be calculated by the Equation (3):

$$z_j = \sum_{i=1}^N o_{i,j} * x_i; j = 1, 2, \dots, J. \quad (3)$$

The best portfolio is obtained by solving:

$$\begin{aligned} & \text{Maximizing } z_1(\vec{x}), z_2(\vec{x}), \dots, z_p(\vec{x}), \\ & \vec{x} \in R_F, \end{aligned} \quad (4)$$

where R_F , is the space of feasible portfolios limited by $C_l \leq P_G$ and the area and region constraints.

4. Method of solution for the multi-objective portfolio problem

To work out a solution to the problem enunciated in Section 3, the NSGA-II was implemented, obtaining the Pareto frontier with solutions not dominated among themselves. The final decision to choose which is the best result depends on the DM and the DM-analyst.

4.1 The NSGA-II working with interval analysis.

The NSGA-II is considered one of the benchmarks of the multi-objective optimization to solve problems preferably of two or three objectives. It is based on the creation of non-dominated fronts establishing an elitism on the first front in addition includes an indicator of diversity called crowding distance.

Dominance between two interval vectors: Let \vec{p} and \vec{q} be two interval vectors we say that \vec{p} dominates \vec{q} (denoted by $\vec{p} \succ \vec{q}$) if $p_i \geq q_i$ for all i values and there is at least one i such that $p_i > q_i$.

A fundamental step of our proposal is to generalize the NSGA-II for the treatment of interval analysis multi-objective optimization problems, and then we propose to adapt the most important strategies to interval analysis (Algorithm 1):

Algorithm 1. *fast – non – dominated – sort*(P)

```

1: for each  $\vec{p} \in P$ 
2:    $S_p = \emptyset$ 
3:    $n_p = 0$ 
4:   for each  $\vec{q} \in P$ 
5:     if ( $\vec{p} < \vec{q}$ )
6:        $S_p = S_p \cup \{\vec{q}\}$ 
7:     else if ( $\vec{q} < \vec{p}$ ) then
8:        $n_p = n_p + 1$ 
9:   if  $n_p = 0$  then
10:     $rank_p = 1$ 
11:     $F_1 = F_1 \cup \{\vec{p}\}$ 
12:   $i = 1$ 
13:  while  $F_i \neq \emptyset$ 
14:     $Q = \emptyset$ 
15:    for each  $\vec{p} \in F_i$ 

```

\vec{p} and \vec{q} are interval vectors

then if \vec{p} dominates \vec{q}

Add \vec{q} to the set of solutions dominated by \vec{p}

Increment the domination counter of \vec{p}

\vec{p} belongs to the first front

Initialize the front counter

Used to store the members of the next front

```

16:   for each  $\vec{q} \in S_p$ 
17:      $n_q = n_q - 1$ 
18:     if  $n_q = 0$  then  $\vec{q}$  belongs to the next front
19:        $rank_q = i + 1$ 
20:        $Q = Q \cup \{\vec{q}\}$ 
21:  $i = i + 1$ 
22:  $F_i = Q$ 

```

To sort the population of size N according to the level of non-dominance, each solution must be compared with all the solutions in the population to find out if it is dominated (Lines 4 to 10). For the set of solutions of the population P , on line 5, comparison is performed by vector of objectives corresponding to the solution in turn to determine if \vec{p} dominates \vec{q} (\vec{p} and \vec{q} are interval vectors), if so, this solution is included in any structure, to identify which solutions were dominated by \vec{p} . Otherwise, i.e. if \vec{q} dominate \vec{p} , increments the value of n_p , this variable indicates the number of solutions that \vec{p} did not dominates (Lines 7 and 8). In the lines 9 to 11 since the evaluation of the solution is known \vec{p} in the previous process, and if there are no solutions that dominate it, the solution \vec{p} will be part of the first front F_0 . To find the individuals of the following front, the solutions of the first front are temporarily discounted and the process is carried out again (Lines 13 to 22). The procedure is repeated to find the other fronts; the worst case would be that there is only one solution in each front.

5. Computer Experiments

This section presents the necessary experiments to validate the effectiveness of our proposal, for this, we worked with the characterization of uncertainty without using interval analysis and later the interval analysis was included in order to make a quality analysis comparing the results between treatments.

In addition, the necessary experiments are described to validate the quality of the algorithm NSGA-II applying the interval analysis and without applying the interval analysis, we compare the results of the two algorithms. To carry out these experiments are described below the conditions, under which the experiments are performed:

- A. Test Environment.** The algorithm NSGA-II was implemented in Java Programming Language. In order to compute the results, the program was run in a computer with the following characteristics: Intel Core i7 3.5 GHz CPU, 16 GB of RAM, and Mac OS X Yosemite 10.10.4 as the operating system.
- B.** The solutions were obtained from 30 independent runs of NSGA-II and NSGA-II applying interval analysis.
- C. Algorithm Configuration.** In order to set the algorithm NSGA-II, we experiment with the configuration given by Cruz-Reyes et al. in [3]: population size = 100, number of generations = 500, mutation probability = .05 and cross-over probability = 1.

5.1 Description of the case of study

In the project portfolio problem, we have a set A of N projects, the DM should describe the contributions of the objectives, which refer to the number of persons benefited by each project; the costs of those projects must also be described. As for the general information of the problem, there is a budget to distribute (P_G) that had to be analyzed, evaluated and calculated by the DM previously. The proposals are classified into three areas (K) by their nature, and by their impact location into two regions (M). The DM requires a certain level of balance in the supported portfolio. Now, in this section the information where the treatment of uncertainty is included shall be described in detail.

A portfolio is a subset of A that indicates which projects are contained within. There are 28 projects in area one, 41 to the area two, 31 the area three; the projects are divided equitably into two regions. The available budget is 250 thousand dollars.

In Table 1 we can see that for project 1, the DM assumes that it is going to cost \$6,125, and the objective 1 will be benefited with a value of 12 555 and the objective 2 in 9 145, this information is the one estimated for the projects.

Table 1. Estimated Information of 1 of the 100 available projects.

Id	c_i	Area	Region	o_{i1}	o_{i2}
1	\$6 125	2	1	12 555	9 145

The DM has to deal with the uncertainties encountered in the instance, therefore it is considered to treat these imprecisions reflected in intervals, as it can be observed in Table 2, which shows the same project from Table 1, but now reflected with intervals in the estimated values. The problem is not only to provide the estimated information; but the question is how to reflect the uncertainty found. In this work, the uncertainty is modeled in the cost of the projects (column 2), and the benefits of the projects (columns 5 and 6), as intervals. Also, the available budget is presented in the form of interval, calculated using a reduction of 4% ($P_{G_{min}} = \$240\ 000$ dollars) and in the same way it increases 4% ($P_{G_{max}} = \$260\ 000$ dollars).

Table 2. Information about 1 of the 100 available projects reflecting the imprecision in the form of intervals.

Id	$[c_{i_{min}}, c_{i_{max}}]$	AreaRegion	$[o_{i1_{min}}, o_{i1_{max}}]$	$[o_{i2_{min}}, o_{i2_{max}}]$	
1	[\$6 025 , \$6 225]	2	1	[12 405 , 12 705]	[9 095 , 9 195]

We took the first step to solve this problem without applying the intervals; the purpose of this first stage is to apply certain heuristics to model the effects of uncertainty in the instance data. Afterwards, the second stage was carried out where we presented the results of the experiments designed to validate the advantages of dealing with the uncertainty applying interval analysis.

5.2 Treatment of the uncertainty without the application of the interval analysis.

To work with the treatment of uncertainty it is necessary to model the attitude of the DM before uncertainty. Six experiments are proposed that model different reaction attitudes that the DM when facing risk, where the total availability of resources, and the total cost of the projects varies.

In experiments 1, 2 and 3, the value of P_G is calculated as $\frac{P_{G_{min}} + \frac{(P_{G_{min}} + P_{G_{max}})}{2}}{2}$. Therefore $P_G = \$245,000$ dollars.

Experiment 1: Taken to the extreme the cost of the projects, producing:

1. $P_G = \$245\ 000$ dollars.
2. $c_i = c_{i_{max}}$.

Experiment 2: In half of the projects the cost is taken to the upper end, in the remaining projects the midpoint is considered, thus yielding:

1. $P_G = \$245\ 000$ dollars.
2. 50% of all the projects: $c_i = c_{i_{max}}$.
3. 50% of all the projects: $c_i = \frac{(c_{i_{max}} + c_{i_{min}})}{2}$.

Experiment 3: The midpoints of cost are considered for all projects, yielding:

1. $P_G = \$245\ 000$ dollars.
2. $c_i = \frac{(c_{i_{max}} + c_{i_{min}})}{2}$.

In experiments 4, 5 and 6, the value of P_G is taken as $P_{G_{min}}$. Therefore $P_G = \$240\ 000$ dollars.

Experiment 4: For all the projects the cost is taken to the upper end.

1. $P_G = \$240\ 000$ dollars.
2. $c_i = c_{i_{max}}$.

Experiment 5: In half of the projects the cost is taken to the upper end and in the remaining projects the midpoint is considered, yielding:

1. $P_G = \$240\ 000$ dollars.

2. 50% of all the projects: $c_i = c_{i_{max}}$.
3. 50% of all the projects: $c_i = \frac{(c_{i_{max}}+c_{i_{min}})}{2}$.

Experiment 6: The midpoints of cost for all projects are considered, yielding:

1. $P_G = \$240\ 000$ dollars.
2. $c_i = \frac{(c_{i_{max}}+c_{i_{min}})}{2}$.

Notice that in this work the uncertainty is not contemplated in the objectives, but uncertainty is taken into account in the cost of each project. The number of projects, areas and regions, is the same one for the six experiments.

5.3 Results

This section describes the results of the comparison made among the different experiments (treatment of the uncertainty without the application of the interval analysis) with the purpose of determining whether the solution proposed provides competitive or better results when incorporating uncertainty to the estimated information (treatment of uncertainty applying interval analysis).

In the MOEAs, the algorithm performance is measured in terms of non-dominated solutions inside the last generation; this set is known as the approximation to the Pareto frontier. Validation of the results of both experiments was performed through two types of analysis of non-dominated solutions:

Quality Analysis: It determines if the solutions obtained by applying the interval analysis to the treatment of uncertainty provides the best quality solution. This is accomplished by comparing the non-dominated solutions obtained from the NSGA-II by applying interval analysis against the non-dominated solutions obtained from modeling the DM's attitude without applying the interval analysis.

Cardinality: We verify the number of projects contained in the portfolio, inside the column *Card.

In Tables 3 and 4, the results of the entire Pareto front were taken, we shown the comparison of the results obtained with the NSGA-II with Interval Analysis (Table 3) again each of the six heuristics (Table 4) described in the previous section.

Table 3. Results of the analysis realized with NSGA-II with Interval Analysis.

Algorithm	Number of solutions of the non set without repeated	Number of solutions of the non-dominated set	Average *Card
NSGA-II	1233	45	39.06
Interval Analysis			

Table 4. Results of the analysis realized with NSGA-II without Interval Analysis.

Experiment	Solution set without repeated	Number of solutions of the non-dominated set	Average *Card
1	1329	88	37.56
2	1304	76	37.90
3	1324	88	38.31
4	1318	85	37.00
5	1332	92	37.28
6	1266	86	37.69

Finally, the results of non-dominated solutions of the two stages are joined, finding that regardless of the approach, the results of the algorithm Interval Analysis NSGA-II remain at a strong 100% for all cases; and the results for the other algorithm NSGA-II, none of the solutions is maintained in the non-dominated solutions. Finally, the cardinality of the solutions obtained is analyzed, finding that the solutions obtained by the algorithm Interval Analysis NSGA-II in average are better.

6. Discussion

In this work, we presented a new tool useful to describe and represent the uncertainty associated with the problems of decision-making in real life. The study of the project portfolio problem was treated using mathematical modeling with interval analysis and considering relevant constraints, preferences, uncertainty and imprecision; to deal with the cost of projects and the total resources available.

The uncertainty and imprecision, were handled in the total availability of the resources, with two quantities of available budget and in the total cost of the projects, through 6 experiments without the application of interval analysis. Therefore, it is proposed to handle the uncertainty with interval analysis, where the total availability value of resources is provided the form of an interval number, as well as the total cost of the projects, generating a single execution in the treatment of uncertainty.

The solution method proposed to solve the portfolio problem consists of generalize the NSGA-II for the interval analysis.

After presenting the results, our proposal gives evidence of obtaining good compromise solutions in terms of our quality metrics and robustness. Therefore, we can say that:

- i. In a single execution, interval non-dominated solutions were found able to prevail with the analyses without the interval analysis.
- ii. The best solutions were obtained in the cardinality of the portfolios.
- iii. The treatment with the interval analysis addresses the uncertainty attempting to model the variability in the resources that the project consumes and in the available resources.

As future work, we are going to: (a) increase the number of objectives, (b) include preferences of the DM, (c) implement another algorithm.

Acknowledgements

This work has been partially supported by: a) Consolidation National Lab Project 280712, Project supported by CONACyT; b) Fronteras de la Ciencias Project 1340, Project supported by CONACyT; and, c) Project 3058 from the program Cátedras CONACyT.

References

1. Badri, M.A., Davis, D., Davis, D. A comprehensive 0-1 goal programming model for project selection. *International Journal of Project Management*, 19 (2001) 243-252.
2. Cruz-Reyes, L., Fernandez, E., Olmedo, R., Sanchez, P., Navarro, J. Preference Incorporation into evolutionary multiobjective optimization using preference information implicit in a set of assignment examples. In *Fourth International Workshop on Knowledge Discovery, Knowledge Management and Decision Support*. Atlantis Press, (2013) 179-187.
3. Cruz-Reyes, L., Fernandez, E., Gomez, C., Sanchez, P. Preference Incorporation into Evolutionary Multiobjective Optimization Using a Multi-Criteria Evaluation Method. In *Recent Advances on Hybrid Approaches for Designing Intelligent Systems*. Springer International Publishing, (2014) 533-542.
4. Deb, K., Agrawal, S., Pratap, A., Meyarivan, T. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. *Lecture notes in computer science*, 1917 (2000) 849-858.
5. Dickinson, M.W., Thornton, A.C., Graves, S. Technology portfolio management: Optimizing interdependent projects on multiple time periods. *IEEE Transactions on Engineering Management*, 48 (4) (2001) 518-527.
6. López Irarragorri, Fernando, Igor Litvinchev, Miguel Mata Pérez and Elisa Schaeffer. Usando Herramientas de MILP para Estudiar el Modelo de Selección de Portafolios R&D para Casos de Grandes Carteras de Proyectos en el Sector Social. *Computación y Sistemas*, 2008, vol. 12, no 002.
7. Klapka, J., Pinos, P. Decision support system for multicriterial R&D and information systems projects selection. *European Journal of Operational Research*, 140 (2) (2002) 434-446.
8. Moore, R.E.: Interval arithmetic and automatic error analysis in digital computing. Ph. D. Dissertation, Department of Mathematics, Stanford University, Stanford, CA (1962)
9. Radulescu, C.Z., Radulescu, M. Project portfolio selection models and decision support. *Studies in Informatics and*

- Control, 10 (4) (2001) 275-286.
10. Rabbani, M., Aramoon, M., Baharian, G. A multi-objective particle swarm optimization for project selection problem. *Expert Systems with Applications*, 37(1) (2010) 315-321.
 11. Rivera, G., Cruz, L., Fernandez, E., Gomez, C., Perez, F. Many-objective portfolio optimization of interdependent projects with 'a priori' incorporation of decision-maker preferences. *Appl. Math*, 8(4) (2014) 1517-1531.
 12. Salo, A., Keisler, J., Morton, A. *Portfolio Decision Analysis. Improved methods for resource allocation*, International Series in Operations Research & Management Science. Springer New York, 162 (2011) 3-27.
 13. Santhanam, R., Kyparisis, J. A multiple criteria decision model for information system project selection. *Computers and Operations Research*, 22-8 (1995) 807-818.
 14. Sunaga, T.: *Theory of interval algebra and its application to numerical analysis*. RAAG Memoires 2, (1958) 29-46.