A Clonal Selection Algorithm for the design of an optimal Investment Portfolio

Juan Fernando García-Mejía
CU UAEM Atlacomulco
Universidad Autónoma del Estado de México
Atlacomulco,México
fgarciam@uaemex.mx

Diana Barrera Guzma
CU UAEM Atlacomulco
Universidad Autónoma del Estado de México
Atlacomulco,México
dbarrerag579@alumno.uaemex.mx

Elsa Teresita Rodriguez-López
CU UAEM Atlacomulco
Universidad Autónoma del Estado de México
Atlacomulco,México
etrodriguezl@uaemex.mx

Mineli de Jesus Navarrete
CU UAEM Atlacomulco
Universidad Autónoma del Estado de México
Atlacomulco,México
mdejesusn@uaemex.mx

Yenit Martinez-Garduño
CU UAEM Atlacomulco
Universidad Autónoma del Estado de México
Atlacomulco,México
ymartinezg@uaemex.mx

Carlos Eduardo Torres-Reyes
CU UAEM Atlacomulco
Universidad Autónoma del Estado de México
Atlacomulco,México
cetorresr@uaemex.mx

***Abstract*— One of the applications of artificial intelligence is combinatorial or numerical optimization by means of computational algorithms. This proposal documents the obtainment process of an optimal investment portfolio of 10 assets of the Mexican Stock Exchange from the diversification of the amounts to be invested, which is calculated by a population based metaheuristic called immunological algorithm of clonal selection, a technique supported in the evolutionary principle of the immune system of upper mammals, whose objective function was formulated as maximizing the gain-risk ratio described by the Markowitz Theory of Investment Portfolios, the possible solution vectors, it is to say, the amounts to be invested, are encoded with real numbers. In this proposal, the results obtained were analyzed with statistical tests which allowed to determine the optimal parameters and stability of the proposed algorithm that made possible to obtain the maximum profit-risk ratio.**

Keywords— investment portfolio, numerical optimization, Clonal Selection Algorithm.

I. INTRODUCTION

One way to define the concept of optimization is through mathematical language as shown in general terms by expressions (1), (2), (3) [1]

$$\underset{\vec{x} \in \Re^n}{minimize} \quad f_i(\vec{x}) \quad \forall \quad (i = 1,2,\dots,M) \tag{1}$$

$$subject\ to \quad \phi_j(\vec{x}) = 0, \forall \quad (j = 1,2,\dots,J) \tag{2}$$

$$\psi_k(\vec{x}) \leq 0, \quad \forall (k = 1,2,\dots,M) \tag{3}$$

where $f_i(\vec{x})$,$\phi_j(\vec{x})$ and $\psi_k(\vec{x})$ are the functions of the design vector (4)

$$\vec{x} = (x_1, x_2, \dots x_n)^T \tag{4}$$

where the components x_n of $\vec{x}$ are called decision variables, they can be real or discrete and determine respectively if it is a numerical or combinatorial optimization. The possible functions formed with equation 1 are called objetive functions, in the case of $M = 1$, there is only a single function. The space $\Re^n$ formed by decision variables is called search space while the space formed by the equation 1 is called solution space.

The equalities and inequalities described respectively in [1] equations 2 and 3 can be either linear or nonlinear and are called constraints. It should be noted that it is possible to write the objective functions as maximization problems and the inequalities can also be written as ≥ 0 since the maximization of $f(x)$ is equivalent to minimization of and is equivalent to $-g(x) \geq 0$ [1]

A numerical optimization problem that can be solved by stochastic algorithms is the construction of an investment portfolio, which is defined as a set of assets, that means accounting documents issued by public or private entities, which entitle to receive a certain future income. The purpose of investment portfolios administration is to find ways of diversifying the amounts to be invested in the assets in such a way as to enable, as far as possible, to guarantee maximum profit and to reasonably reduce the risk of a possible loss with regard to the amount in which the assets were acquired. One way to achieve this is by diversifying the amounts to be invested, it is to say the money earmarked for the purchase of each of the assets that constitute the portfolio. One way to achieve this diversification is through the Optimal Portfolio Theory developed by Harry Markowitz [2].

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

A. *Optimal Portfolio Theory*

Markowitz's modern investment portfolio theory considers that a denominated investor entity owns a certain amount of money with which will buy a certain amount of assets and will keep them for a certain amount of time. Mathematically the construction of an investment portfolio can be expressed by means of equations (5), (6) and (7). The expression (5) models the portfolio efficiency $\overline{r_p}$, in (6) the calculation of loss risk is shown σ_p^2 and (7), represents the portfolio constraint, it means that it is not possible to invest more than 100% of the available money [3]

.

$$\overline{r_p} = \sum_{i=1}^{N} X_i * \overline{r_i} \tag{5}$$

$$\overline{r_p} = X_1 * \overline{r_1} + X_2 * \overline{r_2} + \cdots + X_2 * \overline{r_2}$$

$$\sigma_p^2 = \sum_{i=1}^{N} \sum_{i=1}^{N} X_i * X_j * \sigma_{ij} \tag{6}$$

$$\sum_{i=1}^{N} X_i = 1 \tag{7}$$

Where:

$\overline{r_p}$ is expected gain

X_i is proportion of the initial value invested in the *i-th* portfolio asset

$\overline{r_i}$ is the performance of the *i-th* asset

N is the number of assets

σ_p^2 is the estimated portfolio risk

σ_{ij} is the covariance of returns between *i-th* and *j-th*

It is important to note that σ_{ij} represents the asset values variations in each timescale $t$which is determined in months or years.

B. *State of the art*

According to the specialized literature the problem of designing an investment portfolio by means of the diversification of the assets that form it can be solved through the techniques described in Fig 1. In this proposal the calculation of amounts is done by a population based metaheuristic, thematic that is addressed in the review of the state of the art

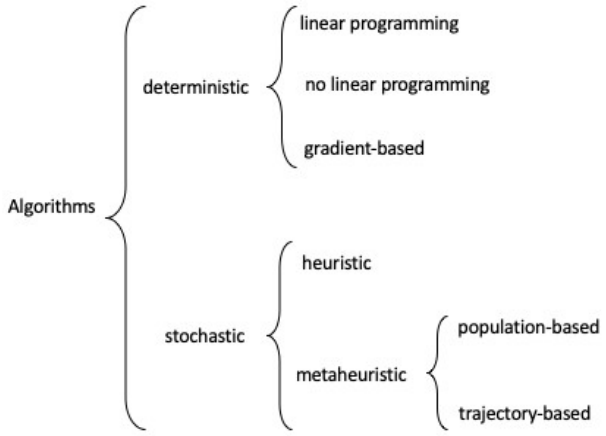

Fig 1 Optimization Techniques

An example of population based algorithms is presented in [4], where the amounts to be invested in a portfolio of 11 assets are obtained by means of a Genetic Algorithm (GA) in its single objective version with a binary encoding. Moreover, in [5] an Imperialist Competitive Algorithm (ICA) is used respectively with 10 assets, this work use a coding based on real numbers. In cases developed in [4] and [5] it must be noted that the assets selected from suggestions from financial specialists are listed on the Mexican Stock Exchange.

There are other approaches of diversification of investment portfolios from evolutionary algorithms, an example of this is [6] where construction and diversification is performed from a multi-target Genetic Algorithm specifically of NSGAII (Non-dominated Sorting Genetic Algorithm) with which they select assets and diversify the amounts to be invested in them. A classic approach used in the literature is the one selected in [7] where the investment portfolio problem is treated from a single objective approach following the premise of [6] using two stages, the first finds the assets, the second, calculates the amount to invest in them, codified both proposals with binary numbers.

On the other hand, in [8] an immunological algorithm of clonal selection is used which allows to select, from a set of portfolios previously built and diversified, which of them offers the best return on investment, it should be noted that the problem uses a real coding. In addition, the proposal described in [8] is raised as a multi-objective optimization problem.

II. METHOD

This proposal used a Clone Selection Algorithm coded in RStudio, whose canonic version was developed in 2002 by Nunes de Castro and Fernando J. Von Zuben of the Universidad Campinas de Brazil, from the biological theory of the immune system developed by Frank Macfarlane Burnet, an Australian biologue that in 1978 demonstrated the function of lymphocytes (immune cells that are produced in the bone marrow) and their response to antigens (foreign substances such as chemicals, bacteria or viruses from outside the body) that attack the body, mainly in the upper mammals. Burnet's

theory, in turn, is described from the parallel with Darwin's theory of evolution [9][10].

Burnet's theory of clonal selection immune system stipulates that lymphocytes react to a certain invasive antigen, so that this stimulation conditions a proliferation (cloning) of the lymphocyte series, those who have a higher affinity are subjected to a series of changes (mutations), the mutated clones that have the highest ability to neutralize the antigen in question, once it has been neutralized, are destroyed, except for a small number of them that are integrated into a set of lymphocytes called immune memory, which is used when a new antigen attacks the body, repeating the process iteratively [11]

A. Clonal Selection Algorithm

The computational equivalent of the clone selection theory developed by Nunes de Castro and Fernando J. Von Zuben is commonly called CLONALG the figure 2 shows the flowchart of this technique, while table I shows an equivalence biological to computational concepts in CLONAG terms

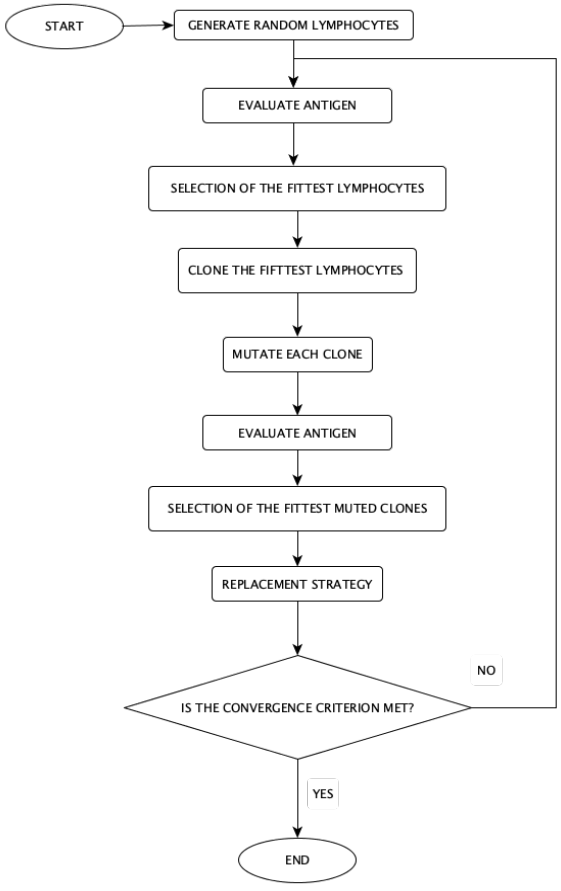

Fig 2 CLONALG flow chart [10]

TABLE I. BIOLOGICAL EQUIVALENCES TO CLONALG

Theory of clone selection immune system	CLONALG algorithm
antigen	An antigen is the equivalent of the objective function
lymphocyte	It is a possible solution to the objective function
affinity	Value obtained from the objective function given the action of an antigen

B. Methodological scheme

The Fig. 3 shows the methodological scheme that is proposed to determine the investment amounts that guarantee the best risk-gain ratio. Statistical analysis is necessary to validate the performance of the algorithm and determine which parameters allow the design of the optimal portfolio.

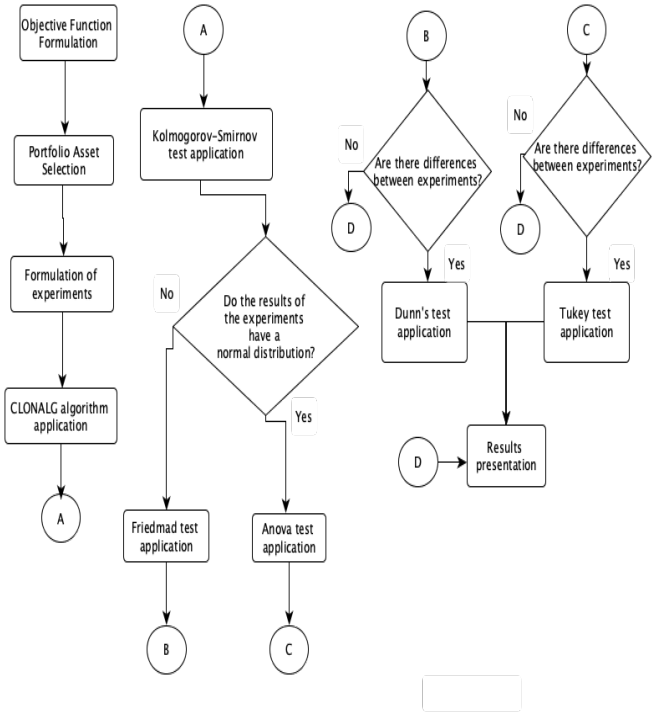

Fig 3 Methodological scheme

As mentioned in previous sections of this document, an immunological algorithm of clonal selection, being an evolutionary algorithm has an objective function named antigen, which for this case was constructed from equations (5) and (6) obtaining (8)

$$maximize \ \frac{\sum_{i=1}^{N} X_i * \overline{r_i}}{\sum_{i=1}^{N} \sum_{j=1}^{N} X_i * X_j * \sigma_{ij}} \qquad (8)$$

Each of the lymphocytes for this proposal is composed of a vector whose components represent the amounts to be invested in the portfolio and have the form described in expression (9)

$$\vec{X} = (X_1, X_2 \dots \dots X_N)^T \quad (9)$$

Where $X_1$is the amount to invest in the stock 1, $X_2$is the amount to invest in the stock 2 and X_N is the amount to invest in the *N-th* stock that for this proposal has a value of 10

On the other hand, the assets that constitute the portfolio which is the subject of study in this proposal, were selected by a financial specialist who calculated the risk matrix of 10 stocks of the Mexican Stock Exchange in a time window t of 1 year. Table II shows the estimated value when calculating diversification using CLONALG.

TABLE II. INVESTMENT PORTFOLIO TO OPTIMIZE

Stock	Stock value
Aeromex	40.02
Bimbo	51.27
Cemex	9.57
Elektra	351.7
Herdez	42.99
Ienova	73.92
Kimber	40.75
Oma	85.73
Soriana	40.51
Walmex	42.26

The number of clones each lymphocyte will have is proportional to its affinity and is calculated by expression (10) while the mutation operation of the clones requires first of all to get a mutation percentage which is obtained through the expression (11) and then apply the expression (12) [12]

$$Cl_n = round\left(\frac{Nl^*\beta}{n} + 0.5\right) \quad (10)$$

Where

Cl_n number of clones for the *n-th* lymphocyte
Nl is total number of lymphocytes
β cloning factor (constant) = 0.1
n number of lymphocyte to clone

$$\alpha = exp\left(-\rho f^*(C_i)\right) \quad (11)$$

where
$f^*(X_i) \in [0,1]$ is the normalized fitness of (Cl_i)
ρ is the decay constant which determines the shape of the mutation rate,

$$\widetilde{C_i} = \begin{cases} C_i + \lambda(C_{r1}\text{-}C_{r2}), & \forall\, \alpha \le p_m \\ C_i, & otherwise \end{cases} \quad (12)$$

$\widetilde{C_i}$ *i-th* mutated clone
C_i clone to mutate
λ mutation factor $rand[0,1]$
p_m mutation probability

Table III describes the experiments to be carried out in this proposal, which are executed 40 times each, according to the central limit theorem, in order to perform a statistical study to determine the statistical behavior

TABLE III. EXPERIMENTAL SET

Test	lymphocytes	lymphocytes to clone	Mutation rate	Probability of mutation
1	100	10	Random [0,1]	0.1
2	100	10	Random [0,1]	0.2
3	100	10	Random [0,1]	0.3
4	100	10	Random [0,1]	0.4
5	100	10	Random [0,1]	0.5
6	100	10	Random [0,1]	0.6
7	100	10	Random [0,1]	0.7
8	100	10	Random [0,1]	0.8
9	100	10	Random [0,1]	0.9
10	100	10	Random [0,1	Random [0,1]

III. RESULTS

After applying the 40 executions to the set of experiments with the CLONALG algorithm, the Kolmogorov-Sminov (K-S) test is applied. As a result of this statistical test it is observed that the 10 experiments have a different behavior from a normal distribution, therefore according to the proposed methodology the Friedman test is applied, which aims to determine whether there are significant differences in each of the tests in the experimental set. The result of this statistical test has a significance of p = 0.001 < 0.005. Therefore the Dunn's test must be applied, to determine which of them is statistically significant. The result reported by this test indicates that all the experiments have statistically significant differences, then in order to be able to decide which are the best parameters that allow THE CLONALG to diversify the portfolio in question, descriptive statistics are used, whose results with respect to the risk-gain ratio described in expression 8 is shown in Table 8. The highest value of the experiment averages is placed in yellow, the

minimum standard deviation in green, and finally the minimum value of the standard deviation and average ratio is placed in red.

TABLE IV. DESCRIPTIVE STATISTICS OF THE EXECUTION OF CLONALG

Test	mean	standard deviation	standard deviation/ mead
1	8855.08522	67.200 52553	0.007588 919
2	8857.012154	79.090 12187	0.008929 662
3	8857.00869	82.166 50705	0.009277 004
4	8850.378821	103.96 34951	0.011746 785
5	8850.131529	88.049 22913	0.009948 918
6	8842.648922	99.290 22207	0.011228 561
7	8819.109768	111.96 1199	0.012695 295
8	8823.127166	102.24 32571	0.011588 097
9	8854.558638	63.684 43964	0.007192 277
10	8866.116586	63.944 61611	0.007212 246

Based on Table IV, the one obtained in test 9 is selected as an optimal portfolio, since it has the most stable behavior, it means that it is deviated less than the average regarding the executions, therefore it is considered more stable than the rest of the experiments. Fig 4 shows the evolutionary behavior of CLONALG and Table V shows the averages to be invested in the diversified portfolio. Fig. 9 shows the histogram of test 9 where a skewed distribution to the right and a high repeatability of the results obtained are confirmed.

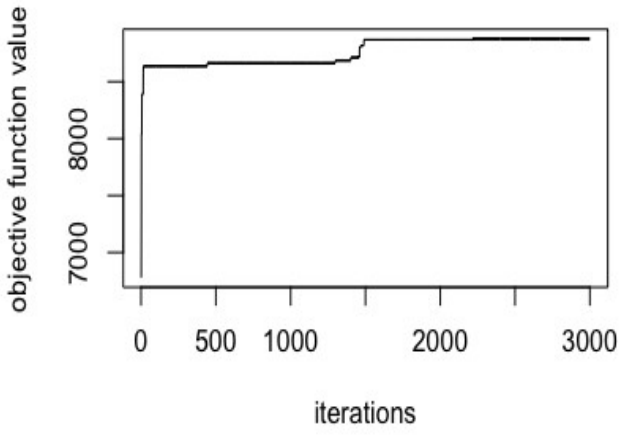

Fig 4 Evolutionary response of test 9

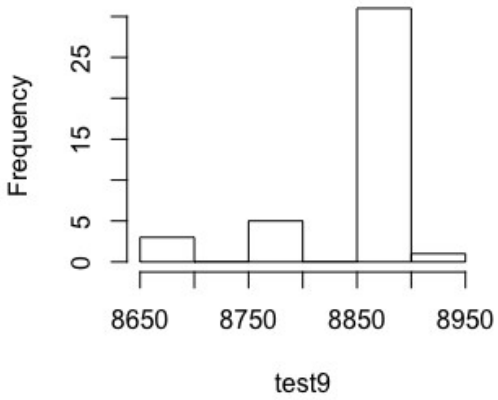

Fig 5 Histogram of test 9

TABLE V. DIVERSIFIED INVESTMENT PORTFOLIO WITH THE PERCENTAGES TO INVEST

Stock	Amount to invesment
Aeromex	0.076631954
Bimbo	0.00027424159
Cemex	0.0063488394
Elektra	0.38579492
Herdez	0.1508086
Ienova	0.11485962
Kimber	9.1590199e-05
Oma	0.10589933
Soriana	4.8529663e-05
Walmex	0.15924238

IV. CONCLUSIONS

The proposal presented in this document can be considered innovative in the field of evolutionary algorithms by the following stances:

• Specialized literature addressing the use of statistical techniques to study algorithm behavior in terms of probability distribution is scarce. A differentiating aspect in this work.

• CLONALG has not been used in diversifying investment portfolios from an objective function. Becoming this metaheuristic an interesting option.

• It is possible to theorize that test 9 of the experimental set is the best option given its high value of mutation probability, but it is still possible to study the effects on the number of original lymphocytes, increase or decrease in the number of lymphocytes to proliferate to mention a few areas of opportunity.

In the future it is considered the hybridization of the algorithm, which could bring the following benefits to consider:

• Improves convergence, which in this case is slow, but faster than the articles reviewed in the state of the art

• Statistical results could have a normal probability distribution

V. References

[1] X.-S. Yang, *Engineering optimization: an introduction with metaheuristic applications*. Hoboken, NJ: John Wiley, 2010.

[2] J. C. Francis and D. Kim, *"Modern portfolio theory: foundations, analysis, and new developments"*. 2013.

[3] J. C. Francis and D. Kim, *"Modern portfolio theory: foundations, analysis, and new developments"*. 2013.

[4] V. Fernandez Cortes, D. Valle Cruz, and P. E. Lizola Margolis , "Optimización de Portafolios de Inversión con Algoritmos Genéticos," *Journal of Research in Accounting and Management Science*, vol. 4, no. 2, pp. 111–125.

[5] J. F. García Mejía, G. A. Perez, C. E. Torres Reyes, and Y. Martínez Garduño, "Diseño de un portafolio de inversión por medio de un Algoritmo Competitivo Imperialista," *Revista Aristas: Investigación Básica y Aplicada*, vol. 6, no. 12, pp. 29–33, Apr. 2018.

[6] De Greiff, Samuel, Rivera, Juan Carlos."Optimización de portafolios de inversión con costos de transacción utilizando un algoritmo genético multiobjetivo: caso aplicado a la Bolsa de Valores de Colombia". *Estudios Gerenciales*, vol 6, no 12, pp 74-87. March 2018 https://dx.doi.org/10.18046/j.estger.2018.146.2812

[7] L. Yu, S. Wang, and K. K. Lai, "Portfolio Optimization Using Evolutionary Algorithms," *Reflexing Interfaces*, pp. 235–245.

[8] Bin Shalan, S.A. & Ykhlef, M "Solving portfolio optimization problem by clonal selection algorithm". Arab J Sci Eng (2015) vol 40 no.2407. https://doi.org/10.1007/s13369-015-1744-4, june 2015

[9] L. N. de Castro and F. J. Von Zuben, "Learning and optimization using the clonal selection principle," in *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 3, pp. 239-251, June 2002. doi: 10.1109/TEVC.2002.1011539

[10] D. Ribatti, *Milestones in immunology: based on collected papers*. London, United Kingdom: Academic Press, an imprint of Elsevier, 2017.

[11] A. I. Tauber and S. H. Podolsky, *The generation of diversity: clonal selection theory and the rise of molecular immunologgy*. Cambridge (Mass.): Harvard University Press, 2000.

[12] V Cutello, G Nicosia, and MPavone. "Real coded clonal selection algorithm for unconstrained global optimization using a hybrid inversely proportional hypermutation operator." *Proceedings of the 2006 ACM symposium on Applied computing (SAC '06). Association for Computing Machinery*, New York, NY, USA, pp 950–954. DOI:https://doi.org/10.1145/1141277.1141501