International Journal of Electronics Signals and Systems

Volume 2 | Issue 1 Article 11

July 2012

CONSTRAINT ROBUST PORTFOLIO SELECTION BY MULTIOBJECTIVE EVOLUTIONARY GENETIC ALGORITHM

S. K. MISHRA
Dept of Electronics and Communication Engineering National Institute of Technology, Rourkela, 769008, INDIA, sudhansu.nit@gmail.com

G. PANDA
School of Electrical Sciences , Indian Institute of Technology, Bhubaneswar, INDIA, ganapati.panda@gmail.com

S. MEHER
Dept of Electronics and Communication Engineering National Institute of Technology, Rourkela, 769008, INDIA, smeher@nitrkl.ac.in

R. MAJHI
School of Management , National Institute of Technology, Warangal, INDIA, ritanjalimajhi@gmail.com

Recommended Citation

MISHRA, S. K.; PANDA, G.; MEHER, S.; and MAJHI, R. (2012) "CONSTRAINT ROBUST PORTFOLIO SELECTION BY MULTIOBJECTIVE EVOLUTIONARY GENETIC ALGORITHM," *International Journal of Electronics Signals and Systems*: Vol. 2 : Iss. 1 , Article 11.
DOI: 10.47893/IJESS.2012.1066
Available at: https://www.interscience.in/ijess/vol2/iss1/11

CONSTRAINT ROBUST PORTFOLIO SELECTION BY MULTIOBJECTIVE EVOLUTIONARY GENETIC ALGORITHM

[1]S.K. MISHRA, [2]G.PANDA, [3]S. MEHER & [4]R.MAJHI

[1, 3] Dept of Electronics and Communication Engineering
National Institute of Technology, Rourkela, 769008, INDIA
[2]School of Electrical Sciences , Indian Institute of Technology, Bhubaneswar, INDIA
[4] School of Management , National Institute of Technology, Warangal, INDIA
Email: Sudhansu.nit@gmail.com, ganapati.panda@gmail.com, smeher@nitrkl.ac.in, ritanjalimajhi@gmail.com

Abstract -The problem of portfolio selection is a very challenging problem in computational finance and has received a lot of attention in last few decades. Selecting an asset and optimal weighting of it from a set of available assets is a critical issue for which the decision maker takes several aspects into consideration. Different constraints like cardinality constraints, minimum buy in thresholds and maximum limit constraint are associated with assets selection. Financial returns associated are often strongly non-Gaussian in character, and exhibit multivariate outliers. Taking these constraints into consideration and with the presence of these outliers we consider a multi-objective problem where the percentage of each available asset is so selected that the total profit of the portfolio is maximized while total risk is minimized. Nondominated Sorting Genetic Algorithm-II is used for solving this multiobjective portfolio selection problem. Performance of the proposed algorithm is carried out by performing different numerical experiments using real-world data.

Keywords: *Multiobjective optimization, Pareto optimal solutions, portfolio asset selection problem, non-dominated sorting, elitism, decision making, constraint handling.*

I.INTRODUCTION

A portfolio is a collection of assets held by a private individual or an institution. The portfolio selection seeks an optimal way to distribute a given budget on a set of available assets. Massive investment to different products like pension funds, banking insurance policies, stock exchange and other series of financial assets is one of the complex problems in financial management. The choice of an appropriate investment portfolio is an important task for a portfolio manager. Optimal selection of stock exchange assets as well as the optimal investment for each asset is a well known portfolio selection problem. Portfolio selection is a complex task as it depends on various factors such as assets interrelationships, preference of the decision makers and resource allocation. When investing money in a set of stock exchange assets, the investors are interested in obtaining the maximum profit of an investment and minimum risk simultaneously. This optimization problem has many constraints like (i) the number of assets a portfolio can contain is fixed and finite (ii) the minimum and maximum amount of possible investments for each chosen assets. (iii) the maximum number of assets that the portfolio manager can select out of all the assets.(iv) the outliers present in the data.

Markowitz set up a quantitative framework for the selection of a portfolio [1,2]. This framework uses the mean variance of historical returns of many assets to measure its expected return and risk. Konno and Yamazaki [3] proposed the mean absolute deviation (MAD) of portfolio which is taken as the risk measure. The possible asymmetry of return is taken into account by Konno Shirakawa and Yamazaki [4] who extended the MAD approach to include skewness in the objective function. Negative semi-variance proposed by Markowitz [2] is one of the several objective functions that considered downside risk. But the entire algorithms remain silent about different constraint associated with portfolio. One of the constraint i.e. the cardinality constraint is approached by some of the researches, Mansini and Speranza [5, 6], and Young[7]. In these papers the MOEAs lacked generality. This multiobjective decision making (MODM) problem with constraint and outliers is solved by NSGA-II algorithm. In addition to that, the present work was did with in depth analysis in examining how the cardinality constraints affect the evolution search process on the measurement of different metric and the efficient frontier attained.

The reminder of the paper is organized as follows. Section 2 outlines the multi-objective optimization formulation of portfolio selection. In Section 3 some of the multi-objective evolutionary techniques used in this paper are dealt. For comparing different multiobjective algorithm, different metric proposed by various authors are presented in section 4. Section 5 deals with the simulation study using real-life data. The results in terms Pareto fronts between risk and return are shown in Section 6. The paper concludes in section 7 with a summary and some ideas for further research work direction.

II. MULTIOBJECTIVE OPTIMIZATION: BASIC CONCEPTS AND A BRIEF OVERVIEW.

Most of the practical optimization problems require decision by simultaneously fulfilling more

than one goals. These goals are the minimization or maximization of functions generally contradicts in nature. It is not possible to find a single solution for such multiobjective problems. A multiobjective optimization problem (MOOP) is defined as the problem of computing finding a vector of decision variables that satisfies some restrictions and optimize a vector function whose elements represent the value of the functions. The generalised multiobjective optimization problem may be formulated as:
Maximize or minimize

$$f_m(x) \quad m = 1,2,3........,M \tag{1}$$

Subjected $$g_j(x) \geq 0 \quad j = 1,2,3.....,J \tag{2}$$

$$h_k(x) = 0 \quad k = 1,2,3........,K \tag{3}$$

$$x_i^L \leq x_i \leq x_i^U \quad i = 1,2,3.......,n \tag{4}$$

Where x is represents a vector of decision variables $x = (x_1, x_2, ..., x_n)^T$ and will optimize the vector function, $\vec{f}(\vec{x}) = \{ f_1(\vec{x}), f_2(\vec{x}), ..., f_m(\vec{x})\}^T$

Where $f_m(x)$ are the x objective functions. The values x and x_i^U represent the minimum and maximum acceptable values for the variable x_i respectively and define the boundary of the search space. The J inequalities g_j and the K equalities h_k are known as constraint functions.

Pareto Optimality: A point $\vec{x}^* \in \Omega$ is Pareto optimal if for every $\vec{x} \in \Omega$ and $I = \{1,2,3,....,k\}$ either $\forall_{i \in I} (f_i(\vec{x}) = f_i(\vec{x}^*))$ or, there is at least one $i \in I$ such that $f_i(\vec{x}) > f_i(\vec{x}^*)$. The symbols f and Ω represents the objective function and the feasible region $(\Omega \in S)$ of the whole search space S respectively. In other words, $\vec{x}^*$ is Pareto optimal if there exists no feasible vector $\vec{x}$ which would decrease some criteria without causing a simultaneous increase in at least one other criterion.

Pareto dominance: A vector $\vec{u} = \{u_1, u_2, ...u_k\}^T$ is said to be dominate $\vec{v} = \{v_1, v_2, ..., v_k\}^T$ that is $\vec{v} > \vec{u}$ if and only if u is partially less than v i.e.

$$\forall i \in \{1,2,.....,k\}, u_i \leq v_i \wedge \exists i \in \{1,2,.......,k\} : u_i < v_i \tag{5}$$

Pareto optimal set: For a given MOP $\vec{f}(x)$, the Pareto optimal set p^* is defined as,

$$p^* := \{x \in \Omega \mid \neg \exists x' \in \Omega, \vec{f}(x') \leq \vec{f}(x)\} \tag{6}$$

The solution of a MOOP is a set of vectors which are not dominated by any other vector, and which are Pareto-equivalent to each other. This set is known as the Pareto-optimal set.

Pareto front: For a given MOOP $\vec{f}(x)$ and Pareto optimal set p^*,the Pareto front pf^* is defined as:

$$pf^* := \{\vec{u} = \vec{f} = (f_1(x), f_2(x), ..., f_k(x)) \mid x \in p^*\} \tag{7}$$

The Pareto optimal set when grouped generates a discontinuous plot known as the Pareto front or Pareto border. The generalized concept is given in 1986 by Pareto [8]. It is difficult to find an analytical expression of the line or surface that contains these points. The procedure to generate the Pareto fronts is to compute the feasible points Ω and the corresponding $f(\Omega)$. When there are sufficient numbers of points, it is possible to determine the nondominated points and to produce the Pareto front. Hence the computation of complete Pareto front involves large computational complexity due to the presence of large number of suboptimal Pareto fronts. It requires the solution to be diverse to cover maximum possible regions.

III. Multi-Objective Formulation of Portfolio

The basic mean-variance portfolio selection problem can be formalized as:

$$\text{Min } V(w) = W^T Q W \tag{8}$$

$$\text{Max} W^T \mu = E \tag{9}$$

$$W^T e = 1 \tag{10}$$

$$0 \leq w_i \leq 1 \text{ and } i = 1,2...,N \tag{11}$$

Where N is the number of assets available, Q denotes the covariance matrix of all investment alternatives, μ_i is the expected return of asset i and e is the unit vector. The decision variables w_i determines what share of the budget should be distributed in asset i. Here $W = \{w_1 w_2 w_3 ... w_N\}$.

Equation 1 and 2 give the two competing objectives which are to be optimized. Equations 3 and 4 show the constraints for a feasible portfolio which means that first all the available money is to be invested and secondly all investments must be positive i.e. no short sales are allowed. The constraints given in equation 3 can be met by normalizing the weights

$$s = \sum_{i=1}^{n} w_i \quad (11)$$

Then the new values for each element of weight vector are normalized.

$$w'_i = \frac{w_i}{s} \quad (12)$$

There are some real world constraint that portfolio manager must consider while solving the portfolio optimization problem. One example of this constraint is cardinality constraint. Let K be the maximum number of assets the portfolio manager can invest money out of N available asset. Then K is called as cardinality constraint.

$$\sum_{i=1}^{N} Z_i = K \quad (13)$$

The decision variable $Z_i \in \{0,1\}$

The variable $Z_i = 1$ if any asset $i(i = 1,2,....,N)$ is held and $Z_i = 0$ if it not held. This equation ensures that exactly K asset of N available asset.

The multi-objective portfolio selection problem involves two competing objectives (i) minimize the total variance, denoting the risk associated with the portfolio expressed in (1) (ii) maximize the return of the portfolio shown in (2). Along with this the maximum number of assets that the portfolio manager can select out of all the assets and the outliers present are to be considered.

The problem is thus to find portfolios amongst K asset of the N available assets that satisfy these two objectives simultaneously with the presence of outlier.

IV. MULTIOBJECTIVE EVOLUTIONARY ALGORITHMS

The classical optimization techniques are ineffective for solving constrained optimization problem such as portfolio management. This shortcoming has motivated researchers to develop multi-objective optimization using evolutionary techniques. Based on basic concepts from the biological model of evolution, the search dynamic of multi-objective evolution algorithm (MOEA) is guided by biologically inspired evolutionary operators like selection, crossover and mutation. The crossover and mutation operator change and create potential solutions while the selection operator provides the convergence property. When MOEA is applied for portfolio optimization, issues like representation, variation operator and constraint handling techniques are considered. MOEA maintains a population of chromosome, where each of them represents a potential solution to the portfolio optimization problem. One chromosome represented by a weight vector, provides the composition of the portfolio.

The pioneering work [9] in the practical application of genetic algorithm to MOOP is the vector evaluated genetic algorithm VEGA. For similar applications a number of algorithms based on genetic algorithm such as NSGA[10] ,NPGA[11], PESA-II[12], NSGA-II [13], RDGA [14] and DMOEA [15] have been proposed in literature. The NSGA-II proposed in [11] is an useful alternative and popular algorithm which alleviates various shortcomings of NSGA.

Dev and Pratab have proposed NSGA II where selection criteria are based on the crowding comparison operator. Here the pool of individuals is split into different fronts and each front has assigned a specific rank. All individuals from a front F_i are ordered according to a crowding measure which is equal to the sum of distance to the two closest individuals along each objective. The environmental selection is processed based on these ranks. The archive is formed by the non dominated individuals from each front and it begins with the best ranking front. Here the new population obtained after environmental selection is used for selection crossover and mutation to create a new population. It uses a binary tournament selection operator. These algorithms are dealt in sequel.

NSGA II Algorithm:

1. Initialize population
2. Generate random parent population p_0 of size N
3. Evaluate objective Values
4. Assign fitness (or rank) equal to its non dominated level
5. Generate offspring Population Q_0 of size N with binary tournament selection, recombination and mutation.
6. For $t = 1$ to Number of Generations

6.a. Combine Parent and Offspring Populations

6.b. Assign Rank (level) based on Pareto Dominance.

6.c. Generate sets of non-dominated fronts

6.d. until the parent population is filled do

6.e.1 Determine Crowding distance between points on each front F_i

6.e.2 Include the ith non dominated front in the next parent population (P_{t+1})

6.e.3 check the next front for inclusion

6.f Sort the front in descending order using Crowded comparison operator

6.g Choose the first N - card (P_{t+1}) elements from front and include them in the next parent

population $\left(P_{t+1}\right)$
6.h Using binary tournament selection, recombination and mutation create next generation
7. Return to 6

V. PERFORMANCE MEASURE FOR COMPARISON

1. S metric. It measures the spread of candidate solution throughout nondominated vectors found. Schott [16] introduced this metric, measuring the distance neighboring vectors in the nondomionated vectors found. This metric is defined as:

$$S \stackrel{\Delta}{=} \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\bar{d}-d_i\right)^2} \tag{14}$$

Where

$$d_i = \min_j\left(\left|f_1^i\left(\vec{x}\right)-f_1^j\left(\vec{x}\right)\right|+\left|f_2^i\left(\vec{x}\right)-f_2^j\left(\vec{x}\right)\right|\right)$$

and $i, j = 1,2,\ldots,n$ (15)

$\bar{d}$ = mean of all d_i and n is the number of nondominated vectors found so far. A value of zero for this metric indicates all members of the Pareto front currently available are equidistantly spaced. The S metric indicates the extent of objective space dominated by a given nondominated set A. If the S metric of a non dominated front f_1 is less than another front f_2 then f_1 is better than f_2. It has been proposed by Zitzler .

2. Δ metric. This metric called as spacing metric (Δ) measures how evenly the points in the approximation set are distributed in the objective space. This formulation introduced by K. Deb[13] is given by

$$\Delta = \frac{d_f + d_l + \sum_{i=1}^{N-1}\left|d_i - \bar{d}\right|}{d_f + d_l + (N-1)\bar{d}} \tag{16}$$

Where d_i be the Euclidean distance between consecutive solutions in the obtained nondominated set of solutions. $\bar{d}$ is the average of these distances. d_f and d_l are the Euclidean distance between the extreme solutions and the boundary solutions of the obtained non dominated set and N is the number of solutions from nondominated set. The low value for Δ indicate a better diversity and hence better is the algorithm.

3. Generation distance (GD): The concept of generation distance was introduced by Van Veldhuizen and Lamont [17]. It estimates the distance of elements of nondominated vectors found, from those efficient Pareto optimal set and is defined as:

$$GD = \frac{\sqrt{\sum_{i=1}^{n} d_i^{\,2}}}{n} \tag{17}$$

Where n is the number of vectors in the set of nondominated solution which are called as candidate solutions. d_i is the Euclidean distance between each of these and the nearest member of the global efficient Pareto front. If $GD = 0$, all the candidate solutions are in global efficient Pareto front and any other value of GD indicates how far are the solutions from the global efficient Pareto front. The more value of GD means the elements are more away from the global efficient Pareto front.

4. Inverted generation distance (IGD): This quality indicator is used to measure how far the elements are in the global efficient Pareto front from those non-dominated vectors found from proposed algorithm and is introduced by Van Veldhuizen [17]. If $IGD = 0$, all the candidate solutions are in the global efficient Pareto front covering all its extension.

VI. SIMULATION STUDIES

In this section we present the simulation results obtained when searching the general efficient frontier that resolves the problem formulated in equation 1 and 2 and with the presence of associated cardinality constraint.

All the computational experiments have been computed with a set of benchmark data available online and obtained from OR-Library being maintained by Prof. Beasley. Five data sets port 1 to port 5 represent the portfolio problem. Each data set corresponds to a different stock market of the world. The test data comprises of weekly prices from March 1992 to September 1997 from the following indices: Hang Seng in Hong Kong, DAX 100 in Germany, FTSE 100 in UK, S&P 100 in USA and Nikkei in Japan. For each set of test data, the numbers of different assets are 31,85,89,98 and 225. In the paper we have used the first data set which corresponds to Hang Seng stock having 31 assets. The data can be found from http://people.brunel.ac.uk /~mastjjb/jeb/orlib /portinfo.html. In the paper only cardinality constraints as provided in equations 13 have been used. Along with this there are some outliers in the input data i.e. the weekly data of return. In the work we have selected different number of assets form the Hang Seng stock where there are 31 assets. The NSGA II has population size of 100, number of generations 100, crossover rate 0.8 and mutation rate 0.05. The number of real-coded variables is equal to number of assets and

VII. The Pareto Fronts Obtained by NSGA-II Algorithm

The standard efficient frontier corresponding to Hang Seng benchmark problem and the unconstraint efficient front generated by four algorithms are depicted in Figs.

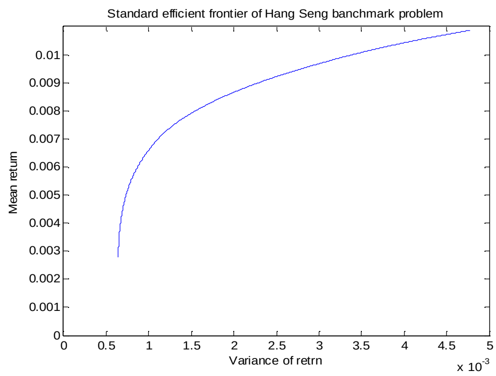

Fig 1. Plots of UEF for Hang Sang

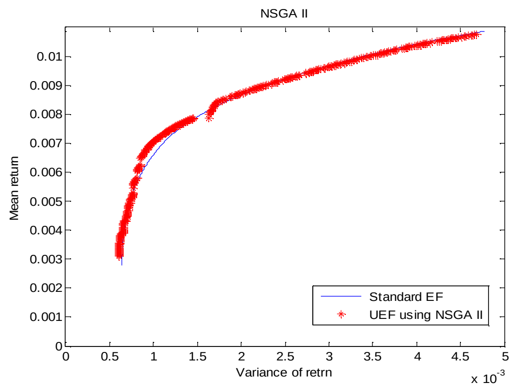

Fig 2. Plots of Pareto fronts achieved by NSGA II

If decision maker is restricted to select only five, ten, fifteen or twenty number of assets in his portfolio out of all the 31 assets then the Pareto curb obtained is shown in the figure 3.

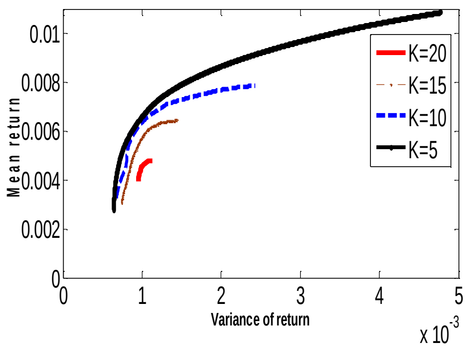

Fig 3. Plots of Pareto fronts achieved at different cardinality constraint.

Table 2

	NC	K=5	K=10	K=15	K=20
S matric	0.0000045	0.0000098	0.0000354	0.0000657	0.0009643
Delta matric	0.5315743	0.6742157	0.7564327	0.8998673	0.9676854
GD matric	0.0007156	0.0009956	0.0018989	0.0099753	0.0267876
IGD matric	0.0073746	0.0097854	0.0196453	0.0843251	0.3215461

Table 2 demonstrates the values of performance metrics. when cardinality constraint increases these metrics values increases. From the graph shows the value of S metric.

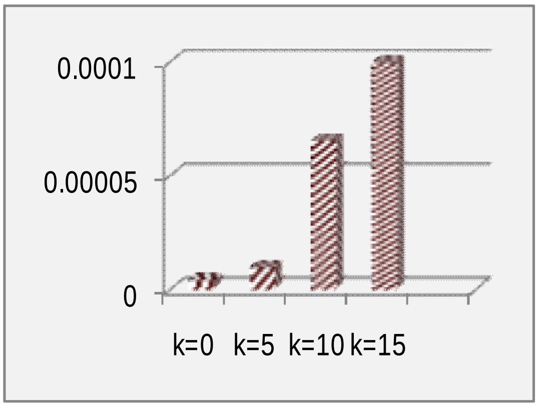

Fig.4. S **matric for different cardinality constraint**

VIII. Conclusion

The paper makes a comparative performance study on portfolio management task employing Nondominated Sorting Genetic Algorithm-II. The data set which corresponding to Hang-Seng stock is used for carrying out simulation based experiments. Experimental results reveal that the NSGA-II algorithm perform satisfactorily to solve the constraint portfolio selection problem with the presence of outliers. Future work includes introduction of different operators for local search in the existing models which allow better exploration and exploitation of the search space when applied to portfolio optimization problem. Another possible future research direction is to handle different real world constraints like minimum buy in thresholds or maximum limit constraints, which would make the problem more complex and then devising improved optimization tools to effectively solve it.

REFERENCES

[1]. H .M. Markowitz. Portfolio Selection, Journal of Finance7(1952)77-91

[2]. H.M. Markowitz, Portfolio Selection: efficient diversification of investments. New York: Yale University Press. John Wiley & Sons, (1991).

[3]. H .Konno and H.Yamazki, Mean absolute-deviation portfolio optimization model and its application toTokyo Stock Market. Management Science 37 (1991) 519-531.

[4]. H.Konno, H.Shrirakawa and H.Yamazki , A mean-absolute deviation-skewness portfolio optimization model. Annals of Operation research 45(1993)205-220.

[5]. R . Mansini and M.G. Speranza, Heuristic algorithms for the portfolio selection problem with minimum transaction lots,Working paper (1997) available from the second author at Dip.di Metodi Quantitativi, Universita di Brescia, C.da.S Chiara 48 /b,25122 Brescia , Italy.

[6]. H. Kellerer ,R. Mansini and M.G. Speranza, On selecting a portfolio with fixed costs and minimum transaction lots. Working paper(1997) available from the third author at Dip.di Metodi Quantitativi, Universita di Brescia, C.da .S Chiara 48/b,25122 Brescia, Italy.

[7]. M.R. Young, a minimax portfolio selection rule with linear programming solution. Management Science 44(1998) 673-683.

[8]. Vilfredo Pareto. Cours. D, Economie Politique , Volume I and II. F. Rouge, Lausanne.1896.

[9]. Schaffer J.D. Multiple objective optimization with vector evaluated genetic algorithms. In Genetic Algorithms and their Applications: Proceedings of the international conference ongenetic algorithm, Lawrence Erlbaum, (1985) 93-100.

[10]. Srinivas N, Deb K, Multiobjective optimization using nondominated sorting in genetic algorithms. J Evol Comput (1994) 221-248.

[11]. Horn. J, Nafpliotis. N, Goldberg D.E, A niched pareto genetic algorithm for multiobjective optimization, In: Proceedings of the first IEEE conference on evolutionary computation, IEEE world congress on computation intelligence, 27-29 June, Orlando, FL, USA,(1994) 82-87.

[12]. Corne D, Jerram NR, Knowles J, Oates J. PESA-II: region-based selection in evolutionary multiobjective optimization. In:Proceeding conference (GECCO-2001), San Francisco, CA, 2001.

[13]. Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evolutionary Computing 6(2), (2002) 182–197.

[14]. Lu H, Yen G.G, Rank-density-based multiobjective genetic algorithm and benchmark test function study, IEEE Trans Evolutionary Computing 7(4), (2003), 325–343.

[15]. Yen G.G, Lu H, Dynamic multiobjective evolutionary algorithm: adaptive cell-based rank and density estimation. IEEE Trans Evolutionary Computing 7(3), (2003) 253–274.

[16]. J. R. Schott, Fault tolerant design using single and multicriteria genetic algorithm optimization, M.S. thesis, Dept. Aeronautics and Astronautics, Massachusetts Inst. Technol, Cambridge, MA, May 1995.

[17]. D. A. Van Veldhuizen and G. B. Lamont, Multiobjective evolutionary algorithm research: A history and analysis, Dept. Elec. Comput. Eng., Graduate School of Eng., Air Force Inst. Technol., Wright PattersonAFB, OH, Tech. Rep. TR-98-03, (1998).

◈◈◈