Multi-Objective Portfolio Optimization

Parastoo Azizeddin & Amirreza Hatamipour

Sharif University of Technology
Department of Electrical Engineering
Convex Optimization II Final Project

***Abstract* – Financial portfolio selection has been one of the most important decision-making field studied in modern finance since 1950s. One of the major concerns in Financial portfolio selection is the problem of finding the most appropriate stocks to hold, based on fluctuations of each of the stocks. Investors are always interested in managing their portfolios, reaching high returns with low risks. Risk is a criterion indicating the uncertainty about the oncoming return. In the financial portfolio selection problem, the investor usually considers several conflicting objectives such as return, risk, and liquidity. Return in the stock market is the aggregate of company dividend and its stock price changes in the market. By diversification of risk and yield maximization, a desirable investment occurs. In this project, we intend to reach an optimal portfolio, and we use Iran Stock Market data as a case study. We use 2 different Genetic Algorithms to solve this problem: A) SPEA-2. B) NSGA-2**

***Keywords* – Portfolio Optimization, Multi-objective Optimization, SPEA-2 algorithm, NSGA-2 algorithm**

I. INTRODUCTION

Different exact techniques have been used to solve the portfolio optimisation problems. These techniques usually involve the exploration of the large number of combinations of states which increases exponentially with the size of problem becoming computationally intractable. Furthermore, many of these techniques are inept in handling the nonlinear objective and constraint functions and several assumptions are generally required to make the problem solvable using reasonable computational resources. Alternatively, some heuristic-based and evolutionary techniques can approximate solutions for problem instances of NP-hard problems in a reasonable time. Those techniques can tackle the optimisation problems in polynomial time with a traded-off of their optimality. In some circumstances of the real world problems, the speed to reach acceptable approximate solutions is very critical. Feasible near-optimum solutions are acceptable but untimely are not. Simple heuristics, based on greedy search algorithms, tend to stop in inferior local optima.

Genetic Algorithms (GA) are population based heuristic algorithms. In GA, solutions are represented as chromosomes that to be breed by crossover or modified by mutation. Selection processes are used to find optimal or near-optimal solutions imitating the natural selection of survival of the fittest [1]. Busetti [2] compared GA with tabu search and found that GA performs better for portfolio optimization problems for the problem setting considered. Streichart et al. [3] applied the Multi-Objective Evolutionary Algorithm (MOEA) to solve portfolio optimization problem. Earlier Tettamanzi et al. (cited in [4]) transformed the multi objective optimization problem into a single-objective problem by using a trade-off function (therefore not a true multi-objective). Mukherjee et al. [5] introduced an alternative hybrid encoding for evolutionary algorithms, which combines both ‘continuous’ real value and ‘discrete’ binary value together. The algorithm then compared with the different EA representations. When the algorithm and the other EAs without Larmarckism (the genetic encoding can be adapted and changed not only by mating and mutations but also during evaluations). was applied on the problem with only cardinality constraints, the algorithms performed better than those of standard EAs. Subbu et al.

II. Problem Definition

Investors are always interested in managing their portfolios, reaching high returns with low risks. Risk is a criterion indicating the uncertainty about the oncoming return. Based on their preferences and wealth, they choose how to maximize their utility while lowering the level of risk. Also, due to the difference in risk aversion among investors, we can formulate the problem in the following three different ways:

- **Form 1:** Minimizing the risk with the constraint of having lower limit on return value **?**, **?**:

$$\begin{aligned} \text{Minimize} \quad & F_{\Omega}(w) \\ \text{s.t} \quad & \mathbb{E}(R(w)) \geq R^{*} \\ & 1^{T} w = 1 \\ & w_j \geq 0, \forall_j = 1, 2, ..., n \end{aligned}$$

- **Form 2:** Maximizing the return with the constraint of having upper limit on risk value **?**:

$$\begin{aligned} \text{Maximize} \quad & \mathbb{E}(R(w)) \\ \text{s.t} \quad & F_{\Omega}(w) \leq L^{*} \\ & 1^{T} w = 1 \\ & w_j \geq 0, \forall_j = 1, 2, ..., n \end{aligned}$$

- **Form 3:** Design a multi-objective problem for maximizing return while minimizing risk level **?**:

$$\begin{aligned} \text{Maximize} \quad & \mathbb{E}(R(w)) \\ \text{Minimize} \quad & F_{\Omega}(w) \\ \text{s.t d} \quad & 1^{T} w = 1 \\ & w_j \geq 0, \forall_j = 1, 2, ..., n \end{aligned}$$

where $F_{\Omega}(w)$ is a function of risk level measurements, L^{*} is the highest acceptable risk level, and $\mathbb{E}(R(w))$ is expected return

of investment.
In fact, all investors should solve problem 3, but it's difficult to optimize two objectives simultaneously. In this project, we intend to find a solution for problem 3 by using multi-objective approach.

A. Multi-objective Optimization Definition

Multi-objective formulations are realistic models for many complex engineering optimization problems. In many real-life problems, objectives under consideration conflict with each other, and optimizing a particular solution with respect to a single objective can result in unacceptable results with respect to the other objectives. A reasonable solution to a multi-objective problem is to investigate a set of solutions, each of which satisfies the objectives at an acceptable level without being dominated by any other solution [6].
For multiple-objective problems, the objectives are generally conflicting, preventing simultaneous optimization of each objective. Many, or even most, real engineering problems actually do have multiple objectives, i.e., minimize cost, maximize performance, maximize reliability, etc. These are difficult but realistic problems [6].
A general multi-objective optimization problem can be described as a vector function that maps a tuple of parameters (decision variables) to a tuple of objectives. Formally [7]:

$$\begin{gathered}\text{min/max } x = f(x) = (f_1(x), f_2(x), ..., f_n(x)) \\ \text{subject to } x = (x_1, x_2, ..., x_m) \in X \\ y = (y_1, y_2, ..., y_n) \in Y\end{gathered}$$

where x is called the decision vector, X is the parameter space, y is the objective vector, and Y is the objective space.

B. Pareto-Optimal Definition

For multi-objective optimization problems, objective functions may be optimized separately from each other and the best solution can be found for each objective dimension. However, suitable solutions for all the functions can seldom be found. This is because in most cases the objective functions are in conflict with each other.Hence, optimizing x with respect to a single objective often results in unacceptable results with respect to the other objectives. Therefore, a perfect multi-objective solution that simultaneously optimizes each objective function is almost impossible [7].
The family of solutions of a multi-objective optimization problem is composed of all those potential solutions such that the components of the corresponding objective vectors cannot be all simultaneously improved. This is known as the concept of Pareto optimality. [7]
In the following lines we indicate the mathematical concept of Pareto Optimality [8]: MOOP (multi-objective optimization problem) is generally defined to obtain decision variables

$$x = (x_1, x_2, x_3, \ldots\ldots\ldots x_p) \in S$$

in p dimensional of model space S, while simultaneously optimizing of a vector

$$f(x) = [f_1(x), f_2(x), f_3(x) \ldots\ldots f_N(x)]$$

including all objective functions.
In MOOP, all objective functions in $f(x)$ vector may not be optimized simultaneously. Therefore Pareto optimality sometimes called as Pareto dominance approach is used to find set of possible solutions. According to Pareto approach, if x is better solution than y in minimization problem, $f(x)$ dominates $f(y)$, if and only if

$$f_k(x) \leq f_k(y), k = 1, \ldots, N$$

including all objective functions [9]. If there is no other $\bar{x} \in X$satisfies a condition such that $f(\bar{x}) < f(x)$ without deteriorating any other objective function, non-dominated solutions called Pareto optimal set $P^* \in S$ or Pareto front $PF^* = f(x) \in \mathbb{R}x \in P^*$ exists in an objective function space as schematically illustrated in the figure bellow:

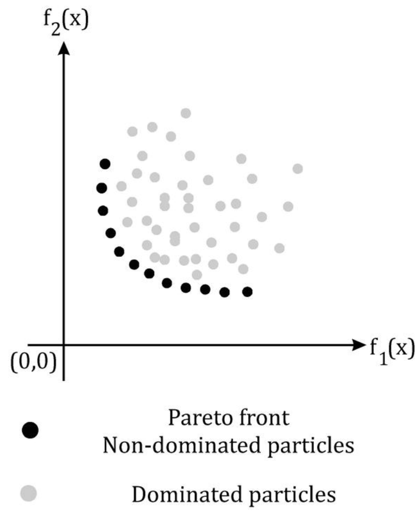

Fig. 1. Conceptual representation of the non-dominated and dominated particles in objective function space.

C. Multi-objective optimization in Portfolio optimization

Now we write the problem in multi-objective format:

$$\begin{gathered}\text{Minimize} \quad (-\mathbb{E}(R(w)), F_\Omega(w)) \\ \text{s.t} \quad 1^T w = 1 \\ \text{s.t} \quad x \geq 0\end{gathered}$$

Markowitz defined the expected return as a linear function and risk as a quadratic function like mentioned bellow:

$$\mathbb{E}(R(w))$$

D. Our Novel Contribution

We changed the objective function in order to have better portfolio optimization. We added the term called "Score" to the objective function which is a linear term of type $score(i) * x_i$ to take account of the preference of the shares. This data is gathered from Iran Stocks website. So the problem will be changed to the following form:

$$\begin{gathered}\text{Minimize} \quad (-\mathbb{E}(R(w)), F_\Omega(w), -Score(x)) \\ \text{s.t} \quad 1^T w = 1 \\ \text{s.t} \quad x \geq 0\end{gathered}$$

III. Solving Methods

A. NSGA-II

NSGA-II is a solid multi-objective algorithm, widely used in many real-world applications. While today it can be considered as an outdated approach, nsga2 has still a great value, if not as a solid benchmark to test against. NSGA-II genererates offsprings using a specific type of crossover and mutation and then selects the next generation according to nondominated-sorting and crowding distance comparison [10]. The algorithm follows the general outline of a genetic algorithm with a modified mating and survival selection. In NSGA-II , first, individuals are selected frontwise. By doing so, there will be the situation where a front needs to be split because not all individuals are allowed to survive. In this splitting front, solutions are selected based on crowding distance.

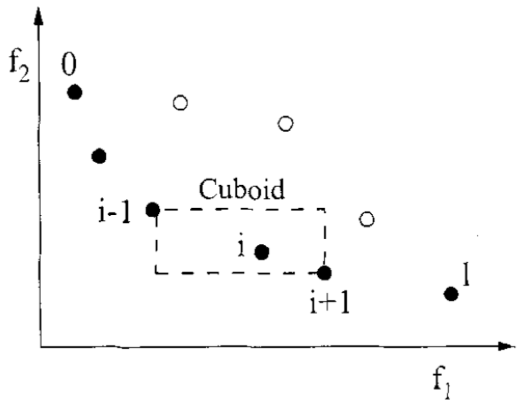

Fig. 2. Crowding-distance calculation. Points marked in filled circle are solutions of the same nondominated front.

The crowding distance is the Manhatten Distance in the objective space. However, the extreme points are desired to be kept every generation and, therefore, get assigned a crowding distance of infinity.
Furthermore, to increase some selection pressure, NSGA-II uses a binary tournament mating selection. Each individual is first compared by rank and then crowding distance.

B. SPEA-2

The Strength Pareto Evolutionary-2 Algorithm (SPEA-2) is a technique for finding or approximating the Pareto-optimal set for multi-objective optimization problems. As the name indicates it is different from the basic EA. SPEA uses the Pareto concept when assigning fitness to the population in order to cover multi-objective problems.The steps are shown in the figure bellow:

Algorithm 1 (SPEA2 Main Loop)

Input:	N	*(population size)*
	$\overline{N}$	*(archive size)*
	T	*(maximum number of generations)*
Output:	$\boldsymbol{A}$	*(nondominated set)*

Step 1: ***Initialization**: Generate an initial population $\boldsymbol{P}_0$ and create the empty archive (external set) $\overline{\boldsymbol{P}}_0 = \emptyset$. Set $t = 0$.*

Step 2: ***Fitness assignment**: Calculate fitness values of individuals in $\boldsymbol{P}_t$ and $\overline{\boldsymbol{P}}_t$ (cf. Section 3.1).*

Step 3: ***Environmental selection**: Copy all nondominated individuals in $\boldsymbol{P}_t$ and $\overline{\boldsymbol{P}}_t$ to $\overline{\boldsymbol{P}}_{t+1}$. If size of $\overline{\boldsymbol{P}}_{t+1}$ exceeds $\overline{N}$ then reduce $\overline{\boldsymbol{P}}_{t+1}$ by means of the truncation operator, otherwise if size of $\overline{\boldsymbol{P}}_{t+1}$ is less than $\overline{N}$ then fill $\overline{\boldsymbol{P}}_{t+1}$ with dominated individuals in $\boldsymbol{P}_t$ and $\overline{\boldsymbol{P}}_t$ (cf. Section 3.2).*

Step 4: ***Termination**: If $t \geq T$ or another stopping criterion is satisfied then set $\boldsymbol{A}$ to the set of decision vectors represented by the nondominated individuals in $\overline{\boldsymbol{P}}_{t+1}$. Stop.*

Step 5: ***Mating selection**: Perform binary tournament selection with replacement on $\overline{\boldsymbol{P}}_{t+1}$ in order to fill the mating pool.*

Step 6: ***Variation**: Apply recombination and mutation operators to the mating pool and set $\boldsymbol{P}_{t+1}$ to the resulting population. Increment generation counter ($t = t + 1$) and go to Step 2.*

Fig. 3. Algorithm SPEA-2 steps

Fitness assignment step evaluates how much each point is dominated [11]. The fitness inversely is related to the number of points that each point is dominating. Figure bellow indicates the fitness assignment in a maximization problem.

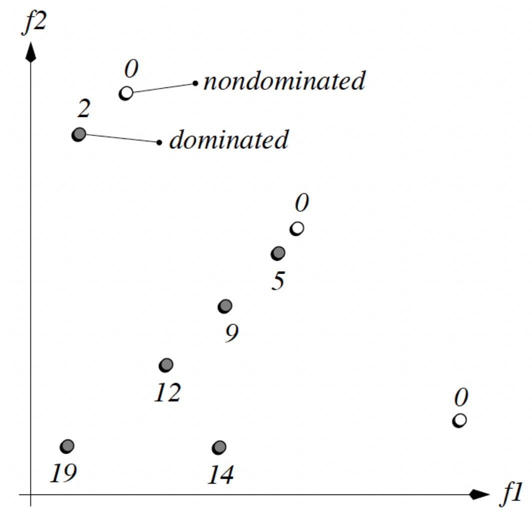

Fig. 4. Fitness assignment schemes in SPEA2 for a maximization problem with two objectives f_1 and f_2

The next step is environmental selection which lets the better point in step 2 get higher probability for attendance in the next generation. The archive set in this step is filled with points that are more probable to be in the Pareto optimal set [11].

Finally, SPEA-2 output is a Pareto optimal set that has a random nature due to the initialization step. The solution of the problem also depends on the population and archive size, while the maximum generation number is not as important as the two other input parameters. It is clear that how much the population (and archive) size is greater, the output is more accurate.

IV. Methodology & Approach

A. Data

We collected the aggregate monthly return of 10 shares in the Iran stock market including: Mellat bank (MLT), hiweb (HW), Tondgooyan Petrochemical (TOP), Iran national copper industry (NCI),Persian Golf Petrochemical (PGP), Iran khodro Investigation (IKHI), Ghadir Investigation (GHI), Shooyandeh Industry (SHI),Bandar Abbas oil refining(BAOR) , and Barekat Pharmacy (BPH) [12]. The companies were chosen randomly. For each of 10, the data collected for 12 months i.e. from Tir 1400 to Tir 1401. After accumulating data, the vector λ was made indicating the average aggregate monthly return of the 10 shares (a vector with length 10). Also, the matrix Q was calculated as the co-variance matrix performing the risk. The co-variance estimation was naturally the summation of square deviations of data from the mean. We formed return vector λ and co-variance matrix Q to make the two first objective functions for return and risk respectively. The third objective Score(x) which represents score related to each share was formed. Finally, all the three discussed objectives were made for passing to the SPEA-2 and NSGA-II algorithm.

B. Implementation & Codes

We used SPEA-2 and NGSA-II algorithm open source code implementation by Mostapha Kalami in MATLAB [13], [14]. The codes are simple to work, but it does not cover the constraints. As we have a budget constraint $1^T w = 1$, we manipulated the code and improved the functions in order to support the budget constraint.

We tested the problem for two cases. The first case is a normal version where the objectives are just the return and the risk. The second case uses all the objectives including return, risk, and score. In the following part we discuss about the two case results.

C. Results

At first we test NGSA-II algorithm on our data: The normal case with a 2 term objective function Pareto optimal set is plotted in the figure below. The horizontal axis indicates the negative of expected return, and the vertical one is related to the risk.

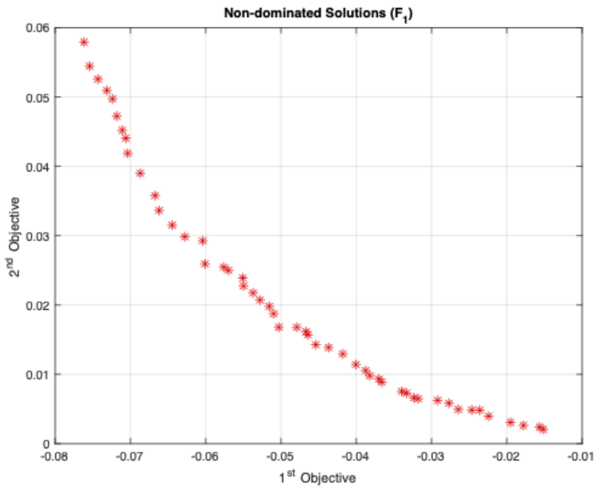

Fig. 5. Non-dominated solution for the two objectives of return and risk

	Solution
Retrun	0.04809
Risk	0.021445

TABLE I

TABLE 1: Example of normal case using NGSA-II.

The second case of having 3 term objective function Non-dominated solution is plotted in the figure below. The first axis indicates negative of expected return, and the second one is related to the risk. Also, the vertical axis is related to the score.

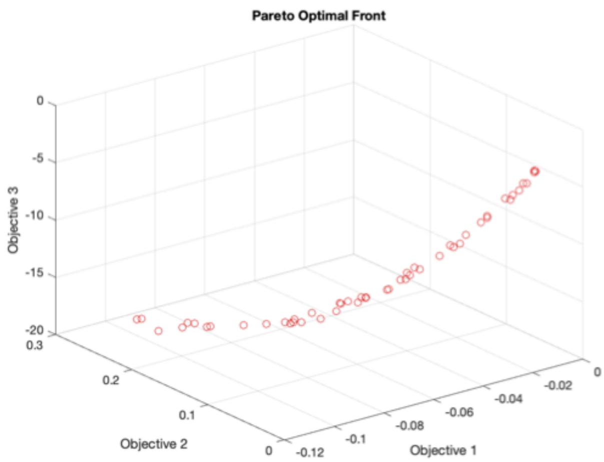

Fig. 6. Non-dominated solution for the two objectives of return and risk

	Solution
Retrun	0.061405
Risk	0.078709
Score	11.345

TABLE II

TABLE 1: Example of normal case using NGSA-II.

In this case with NSGA-2 algorithm when we add new term score to the objective function,the return will be higher with higher amount of risk.

Here we test SPEA-2 algorithm and the results are shown in the following The normal case with a 2 term objective function Pareto optimal set is plotted in the figure below. The horizontal axis indicates the negative of expected return, and the vertical one is related to the risk. In this case, we gathered an example of optimal portfolio in TABLE III.

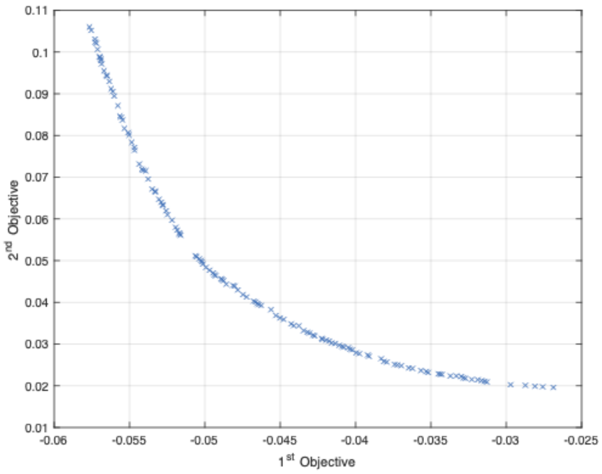

Fig. 7. Pareto optimal set for the two objectives of return and risk

	Solution
MLT	0.0167
HW	0.0018
TOP	0.0004
NCI	0.0000
PGP	0.0581
IKHI	0.0014
GHI	0.2871
SHI	0.0009
BAOR	0.6333
BPH	0.0003
Retrun	0.046179
Risk	0.050178

TABLE III

TABLE 3: Example of normal case.

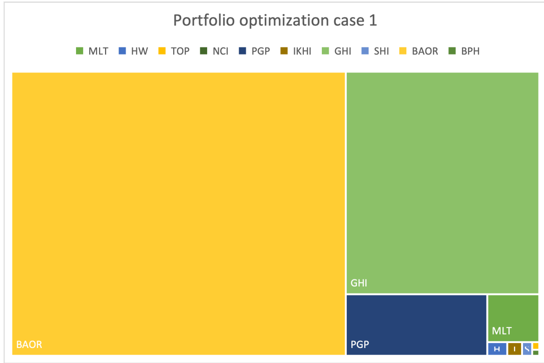

Fig. 8. Portfolio in case 1

The second case of having 3 term objective function Pareto optimal set is plotted in the figure below. The first axis indicates negative of expected return, and the second one is related to the risk. Also, the vertical axis is related to the score. In this case, we gathered an example of optimal portfolio in TABLE IV.

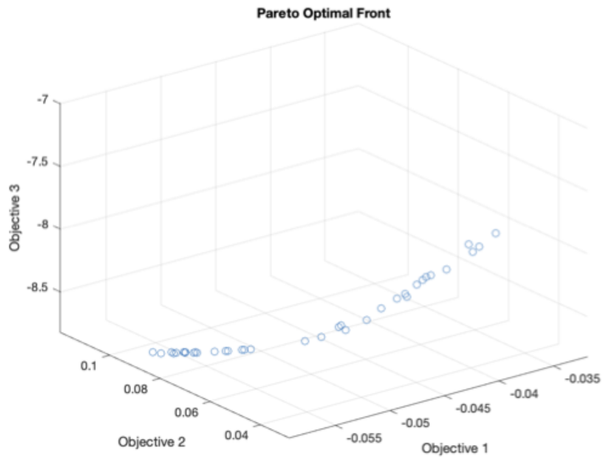

Fig. 9. Pareto optimal set for the three objectives of return, risk, and score

	Solution
MLT	0.0362
HW	0.0000
TOP	0.0026
NCI	0.0000
PGP	0.0000
IKHI	0.0229
GHI	0.3034
SHI	0.0000
BAOR	0.6348
BPH	0.0000
Retrun	0.045311
Risk	0.046755
Score	8.007

TABLE IV

TABLE 4: Example of our developed case.

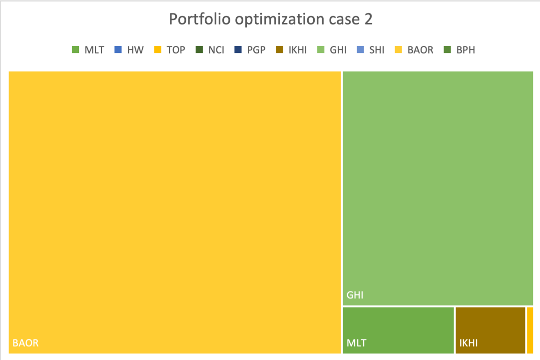

Fig. 10. Portfolio in case 2

These results fully coincide with our intuitive knowledge. As we expect, when we add new term score to the objective function, the risk will be lower than the normal case, but the return will not change a lot.

V. Conclusion

We could successfully form a three-objective portfolio optimization problem to control return, risk, and score. Additionally, we used SPEA-2 and NSGA-II algorithm to solve this problem, and they appeared very successful for this kind of problem. The results interpretation is fully meaningful, and indicates the correctness of the solutions. Finally, using SPEA-2 algorithm for solving portfolio optimization problem with several objectives appears effective and more extension to the problems is needed. Improving the objectives and adding new ones beside testing more data on this framework are suitable topics for following research. One of the terms that I think would be effective in portfolio optimization is the term liquidity that is based on average daily number of transactions for each asset.

REFERENCES

[1] D. Maringer, "Portfolio management with heuristic optimization. advanced in computational management,"
[2] F. R. Busetti, "Metaheuristic approaches to realistic portfolio optimisation," *http://arxiv.org/ftp/cond-mat/papers/0501/0501057.pdf*.
[3] H. Z. Streichert, Ulmer, "Comparing discrete and continuous genotypes on the constraintcd portfolio selection problem," *http:///www-ra.informatik.uni-tuebingen.de/*.
[4] A. L. S. Arnone and Tettamanzi, "A genetic approach to portfolio selection," journal on neural and mass-parallel computing and information systems," *http:///www-ra.informatik.uni-tuebingen.de/*.
[5] H. U. F. Streichert and A. Zell, "Hybrid representation for compositional optimization and parallelizing moeas," *http://drops.dagstuhl.de/opus/volltexte/2005/251/pdf/04461*.
[6] A. Konak, D. W. Coit, and A. E. Smith, "Multi-objective optimization using genetic algorithms: A tutorial," *Reliability engineering & system safety*, vol. 91, no. 9, pp. 992–1007, 2006.
[7] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach," *IEEE transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 257–271, 1999.
[8] E. Büyük, "Pareto-based multiobjective particle swarm optimization: examples in geophysical modeling," 2021.
[9] K. Deb, "Multi-objective optimisation using evolutionary algorithms: an introduction," in *Multi-objective evolutionary optimisation for product design and manufacturing*, pp. 3–34, Springer, 2011.
[10] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002.
[11] E. Zitzler, M. Laumanns, and L. Thiele, "Spea2: Improving the strength pareto evolutionary algorithm," *TIK-report*, vol. 103, 2001.
[12] "Tehran securities exchange technology management co. – tsetmc.,"
[13] M. K. Heris, "Algorithm 2 (spea2) in matlab," *https://yarpiz.com/74/ypea122-spea2*.
[14] M. K. Heris, "Nsga-ii in matlab," *https://yarpiz.com/56/ypea120-nsga2*.