PROGRESS REPORT

METAHEURISTIC APPROACH TO SOLVE PORTFOLIO SELECTION PROBLEM : AN APPLICATION TO ISTANBUL STOCK EXCHANGE

Taylan KABBANI
taylan.kabbani@ozu.edu.trl

Graduate School of Science and Engineering
Data Science
November 28, 2020

Contents

1 Problem Definition

Financial markets are at the heart of the modern economy and they provide an avenue for the sale and purchase of assets such as bonds, stocks, foreign exchange, and derivatives. The prime objective of any investor when investing capital in the stock market is to minimize the risk involved in the trading process and maximize the profits generated, and this objective can be met by optimally choosing a portfolio (grouping of stocks) in which the capital among stocks is invested in such a proportion that the profit is maximum and the risk is minimum, this known as *Portfolio Optimization*

The traditional financial risk management approach is based on *mean–variance model* of portfolio theory (**Markowitz**)[1], which uses historical mean return and co-variance of stocks to optimize a portfolio, therefore we can divide the Portfolio Optimization problem (POP) into two stages. The first stage is to forecast the future return (beliefs about the future performances) of available securities based on historical data, and the second stage is to distribute the capital among the chosen assets in a way to minimizes risk and maximizes profits.

In this project, specific constraints will be introduced to the basic *Markowitz* model in order to make it more adherent to the real world trading mechanisms. The addition of these constraint will turn the model from Quadratic Programming (QP) problem to a Mixed Integer Quadratic Programming (MIQP) problem, which is a NP-Hard problem that can be optimally solved using *Metaheuristic* approaches.

The Constrained Multi-Objective Portfolio Selection will be applied on new problem instances (Istanbul Stock exchange), and will be optimally solved using a Metaheuristics approach, (**Genetic Algorithm or Tabu Search**).

2 Literature Review

Chang et al. [3] used three different metaheuristic approaches (Genetic Algorithm, Simulated Annealing and Tabu Search) to solve a single-objective portfolio optimization problem (with cardinality & quantity constraints. The reported that no individual heuristic was found to be dominating across five data sets. **Schaerf et al.** [4] explored the use of Tabu Search on single-objective constrained POP,also they proposed new algorithms that combine different neighborhood relations. According to their results, the developed solver finds the optimal solution in several instances and is at least comparable to other state-of-the-art methods for the others. As further constraints, **Soleimani et al.** [5] introduced sector capitalisation as additional constraint to reduce investment risk, they used GA and compared the results of a small instance problem to results in LINGO (which can generate global optima for small problems) and reported that the difference between LINGO's global optimum and GA's best objective is just 2.9%. By combining local search techniques (SD & FD) and exact mathematical programming (Quadratic programming solver) **Gaspero et al.** [7] proposed a hybrid local search algorithm, according to their results, the developed solver finds the optimal solution in several instances and is at least comparable to other state-of-the-art methods for the others

Surveying the techniques used in literature ([3],[5],[8]-[10]) to solve the constrained portfolio optimization problem reveals that Genetic Algorithm is the most used technique to solve the problem, where couple of studies ([3],[4],[11]) considered using Tabu search.

3 Mathematical Model

. The unconstrained *Markowitz model* is represented first. Subsequently, the proposed constraints are imposed to deliver the final model.

3.1 Unconstrained Markowitz model

Markowitz model demonstrated that the investor's interest in minimizing risk can be approximated by minimizing the co-variance between chosen stocks, i.e., minimizing the chance of loss by choosing stocks that do not move together, so if some stocks in portfolio are not performing well, then other stocks (having low co-variance with poorly performing stocks) in portfolio can cover the loss [2].

Given a set of n available stocks , $A = \{a_1, a_2, ..., a_n\}$, let μ_i be the expected return (increase in stock price) for stock a_i and x_i the portion of the capital invested in stock a_i, and each pair of stocks (a_i, a_j) has a real value of covariance σ_{ij}. Let μ_p represent the desired return from the portfolio, thus the formulation of the unconstrained problem is given as:

$$min \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \sigma_{ij} \tag{1}$$

$$S.t.$$

$$\sum_{i=1}^{n} \mu_i x_i \geq \mu_p \tag{2}$$

$$\sum_{i=1}^{n} x_i = 1 \tag{3}$$

$$0 \leq x_i \leq 1 \;\; i = 1, 2, ..., n \tag{4}$$

Eq. (1) minimises the total variance (risk) associated with the portfolio, Eq. (2) ensures that the portfolio has an expected return of or bigger than μ_p. Eq. (3) ensures that 100% of the capital is being invested and short-selling is not feasible.

3.2 Constrained Markowitz model

- **Multi-objective**: While the previous single-objective model considers only minimal risk for a given expected return it's desirable to also seek a maximum return for a given expected level of risk, this can be achieved by subsuming the expected return constraint into the objective function via a weighting approach which yields a profitable algorithmic approach. The objective function Eq.(1) will be come Eq.(5) Where λ represent the risk aversion of the investor, when set to 0 the investor is taking high risk in order to maximize the profits, if 1, the investor is not willing to take any risk but profits will not be necessarily maximized, Eq.(10)

- **The cardinality constraint**: This constraint will determine the number of assets k the portfolio must include.A binary variable z_i is introduced to denote whether an asset is being selected or not. Eq.(7) & Eq.(11)

- **Quantity constraint**: Sets boundaries for the weights of included assets by specifying a lower (ε) and upper (δ) bounds allowed for the allocated proportions to each asset in the portfolio, Eq.(8)

- **Proposed Model**:

$$min\ \lambda[\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \sigma_{ij}] - (1-\lambda)\sum_{i=1}^{n}\mu_i x_i \tag{5}$$

$$\sum_{i=1}^{n} x_i = 1 \tag{6}$$

$$\sum_{i=1}^{n} z_i = K \tag{7}$$

$$\varepsilon z_i \leq x_i \leq \delta z_i \quad i = 1,2,...,n \tag{8}$$

$$0 \leq x_i \leq 1 \quad i = 1,2,...,n \tag{9}$$

$$0 \leq \lambda \leq 1 \tag{10}$$

$$z_i = \begin{cases} 1, & \text{if asset i is held} \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

$$z_i \in \{1,0\}, \quad i = 1,2,...,n \tag{12}$$

- Where:
  * n : number of stocks in dataset.
  * λ : The risk aversion of the investor
  * x_i : Proportion of capital invested in asset i
  * σ_{ij} : The co-variance between asset i and j
  * μ_i : The expected return of the asset i
  * K : the number of assets in the portfolio
  * ε : The minimum proportion invested in asset
  * δ : The maximum proportion invested in asset

This formulation is a mixed quadratic and integer programming problem for which efficient algorithms do no exist

4 Data Generation & Benchmark data

Financial Data: The monthly stock price data for number of companies (will be determined later) in Istanbul Stock exchange will be imported for a period of 4 years (from 2016-01-01 to 2020-04-01) using Yahoo Finance[6] which provides a python model to get historical stock data. The data consists of 1069 instances and 7 features, namely date, the highest price of the day, the lowest price of the day, open price, close price, volume and adjacent close price.

The computational experiments to test the algorithm will be done with one of the five sets of benchmark data that have been already used in [3],[4] and [7] where constraint parameters exactly will be set as [3]. These data correspond to weekly prices from March 1992 to September 1997 and they come from the indices: Hang Seng in Hong Kong, DAX 100 in Germany, FTSE 100 in UK, S&P 100 in USA and Nikkei 225 in Japan. The number N of different assets considered for each one of the test problems is 31, 85, 89, 98 and 225, respectively.`http://people.brunel.ac.uk/~mastjjb/jeb/orlib/portinfo.html)`

4.1 Efficient Frontier

The efficient frontier represents the set of optimal portfolios that offer the expected return for minimum associated risk [13]. Therefore, solving the unconstrained quadratic problem (Eqs. (1)-(4)) for varying values of the expected return (μ_p) we obtain the so-called *unconstrained efficiency frontier*(UEF), i.e. the curve represents the set of Pareto-optimal (non-dominated) portfolios. Fig. 1 shows such a UFE for one of the benchmark problems of **Chang et al.**[3](assets drawn from the UK FTSE market index)
Given that the constraint problem has never been solved exactly (Eqs.(5)-(12)), the quality of the heuristic method will be measured in average percentage loss w.r.t. the UEF (available from the web site), i.e., taking the sets of Pareto optimal portfolios obtained by solving the constrained POP using the proposed heuristic approach we can trace out the heuristic efficient frontier and compare it to the UEF [3].

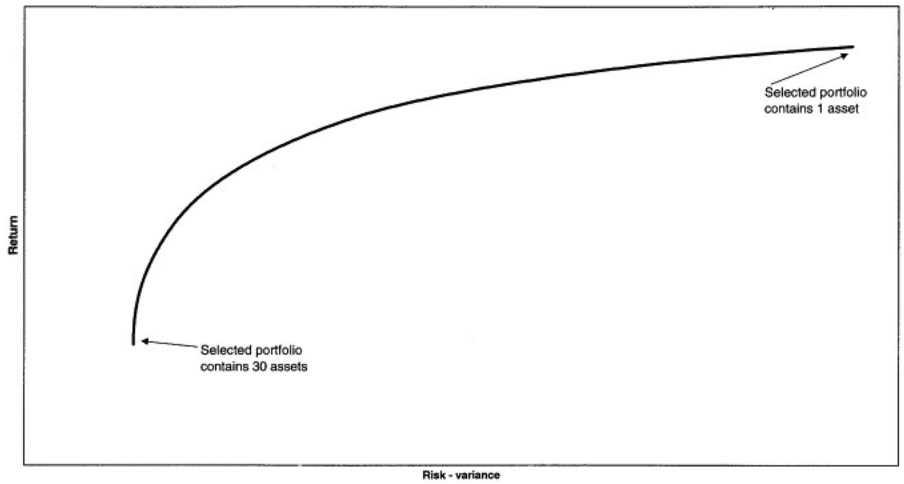

Figure 1: An example efficient frontier

5 Local Search Approach for Portfolio Selection

For representing a solution, I make use of two sequences $L = a_1, ..., a_l$ and $S = x_1, ..., xl$ such that a_l is the stock in the solution (portfolio) and x_i is the fraction of a_i, where length L satisfies Eq.(7) and S satisfies Eq.(6), Eq.(8) and Eq.(9).

5.1 Setting Initial Solution

In the literature [3,4], the initial feasible solution is selected as the best among randomly generated portfolios with K assets. However, this will increase the computation time especially for large number of generated portfolios and large number of K assets, furthermore, we might start at a portfolio very far from an optimal one which also increases the computation time of solving the problem. In this project I propose a *Greedy approach* (similar to Fractional Knapsack Problem) based on *Sharpe Ration*[12], which is basically the ratio of the expected return of an asset to its standard deviation.

Algorithm 1: Initial Solution

Input: Expected returns, Standard deviations, Number of iteration to perform(T)
Output: {stock_1: x_1, stock_2: x_2,..., stock_n: x_n}

```
L = []                                   // Starting with empty set of assets
S=[]                                     // Starting with empty set of fractions
T = Integer                              // Number of specified Iterations
Best_solution = {}
V = Infinite                             // Best value found for the objective function
for Each available asset do
    Calculate Sharpe_ratio = Asset Return / Standard Deviation
Rank available assets based on their Sharpe_ratio in a descending order
Add the first K assets to L              // Constraint in Eq.(7)
for t=1 to T do
    S = {}                               // Current solution in the iteration
    f = 0                                // Current value for the objective function
    for Each asset in L do
        // Constraint in Eq.(8) & Eq.(9)
        x_i = randomly assign a value between ε and δ
    Renormalize weights in S to add up to one   // Constraint in Eq.(6)
    if f < V & length(S) = k then
        Update the best solution
        V = f
        Best_solution = S
return L, S
```

To illustrate the effectiveness of the proposed greedy algorithm to find an initial solution, I solve a small problem (toy problem) which had been solved using exact method, the General Algebraic Modeling System (GAMS) in [14]. As no exact methods are available to solve the constrained POP, the constraints are being relaxed and set as in [14], by setting parameters as following: Lambda = 1, Capital = 1000, k=3, epsilon = 0.2, delta = 0.9, T= 10000, where T is the number of iterations to generate random weights and evaluate them, so by increasing T we might get better solution but with computation trade-off it. The obtained results, i.e., proportions of the capital to be invested in each stock, are very close the ones found by the exact method, which a good indication of the efficiency of the proposed approach. Solution found by GAMS : {0.497 IBM, 0.0 WMT, 0.502 SEHI}, Solution found by the Greed algorithm : {0.442 IBM, 0.1036 WMT, 0.454 SEHI}

The algorithm code (python 3) can be found here, and an illustrative notebook is available here.

5.2 Neighborhood Generation

Given that the variables we are trying to optimize are continuous, *neighbor solutions* can be defined as the *move* applied to the chosen asset's weight. The move can be *Increase* or *Decrease* the weight by a real-value q where $0 < q < 1$. For example, if the move is to increase the weight of an asset a_i , its weight will be $x_i := x_i(1+q)$, if the move is decreasing $x_i := x_i(1-q)$. So for all assets present in the portfolio of K assets we increase and decrease, therefor, the number of neighbors which we need to evaluate is 2K. The following neighborhood relation is a modified version of (Chang et al.) [3]:

- **Move** ($Increase \uparrow, Decrease \downarrow, Replace \updownarrow$):

Step.1 The quantity of a chosen asset's weight is increased ($x_i := x_i(1+q)$) or decreased ($x_i := x_i(1-q)$).

Step.2 All other weights in the portfolio are changed accordingly so as to maintain constraint in E.q(6)

Step.3 If the weight of an asset falls below the minimum (less that ε) it is replaced by a new asset which has a high Sharpe Ration (calculated in Algorithm 1.) with a weight equals to ε.

Step.4 If the move is to increase the weight but the weight $x_i := x_i(1+q) > \delta$ then x_i is set to δ.

- This Move never change the number of assets in the portfolio therefore the length of the portfolio will always satisfy constraint in Eq.(7)

- The q value here, represents the step of the move and as mention earlier it will be between 0 and 1. In (Chang et al.) [3] they used fixed value of q to decrease or increase weights by 0.1. Here, i propose to assign q randomly for each move which i believe will reduce the computing time, where we start with large step sizes (q is big) and slowly reduce the step size.

6 Future plans

1. Apply the Initial solution algorithm to data instances and compare with a benchmark.
2. Decide what Local search technique to use (Tabu search or Genetic Algorithm) and develop the algorithm to be suited for the problem.
3. Test the algorithm on benchmark data mentioned in Section (4)
4. Generate and clean the data of Istanbul Stock exchange, to apply the algorithm to it.
5. Develop a ML algorithm to predict the future expected returned of an asset (this step might not be performed)
6. Writing the code of the algorithm using Python3
7. Apply the algorithm to Istanbul Stock exchange data.

7 References

1. Markowitz, H., 1952. Portfolio Selection. The Journal of Finance, 7(1), p.77.
2. Pareek, M. and Thakkar, P., 2015. Surveying stock market portfolio optimization techniques. 2015 5th Nirma University International Conference on Engineering (NUiCONE),.
3. Chang T-J, Meade N, Beasley JE, Sharaiha YM. Heuristics for cardinality constrained portfolio optimisation. Comput Oper Res 2000;27(13):1271–302.
4. Schaerf A. Local search techniques for constrained portfolio selection problems.Comput Econ 2002;20(3):177–90. 0104017
5. Soleimani H, Golmakani HR, Salimi MH. Markowitz-based portfolio selection with minimum transaction lots, cardinality constraints and regarding sector capitalization using genetic algorithm. Expert Syst Appl 2009;36(3):5058–63.
6. Yahoo-finance. (n.d.). Retrieved from https://pypi.org/project/yahoofinance/
7. Gaspero LD, Tollo GD, Roli A, Schaerf A. Hybrid metaheuristics for constrained portfolio selection problems. Quant Finance 2011;11(10):1473–87.

8. J.-S. Chen, J.-L. Hou, S.-M. Wu, and Y.-W. Chang-Chien, “Constructing investment strategy portfolios by combination genetic algorithms,”Expert Systems with Applications, vol. 36, no. 2, Part 2, pp. 3824 –3828, 2009.

9. T.-J. Chang, S.-C. Yang, and K.-J. Chang, “Portfolio optimization problems in different risk measures using genetic algorithm,” Expert Systems with Applications, vol. 36, no. 7, pp. 10529 – 10537, 2009.

10. H. Dastkhan, N. S. Gharneh, and H. Golmakani, “A linguistic-based portfolio selection model using weighted max-min operator and hybrid genetic algorithm,” Expert Systems with Applications, vol. 38, no. 9, pp. 11735 – 11743, 2011.

11. Aldaihani, M. M., & Deehani, T. M. (2010). Mathematical models and a tabu search for the portfolio management problem in the Kuwait stock exchange. International Journal of Operational Research, 7(4), 445. doi:10.1504/ijor.2010.032421

12. Investopedia. 2020. How To Use The Sharpe Ratio To Analyze Portfolio Risk And Return. https://www.investopedia.com/terms/s/sharperatio.asp

13. Ganti, A. (2020, September 16). Efficient Frontier Definition. Retrieved November 26, 2020, from https://www.investopedia.com/terms/e/efficientfrontier.asp

14. Ahmed, S., n.d. The Optimization Process: An Example Of Portfolio Optimization. [online] Www2.isye.gatech.edu. Available at: ¡https://www2.isye.gatech.edu/ sahmed/isye6669/notes/p [Accessed 28 November 2020].