# Computer-aided Multi-Optimization Through Genetic Algorithms for Funding Allocation

Antonio Carlos Pinto Dias Alves
Banco do Brasil, Diretoria de Gestão de Riscos
Rua Senador Dantas 105, sala 1704
Brazil

*ABSTRACT:* - In this paper we consider the problem of finding the efficient frontier associated with the standard mean-variance portfolio optimization model. We extend the standard model to include cardinality constraints that limit a portfolio to have a specified number of assets, and to impose limits on the proportion of the portfolio held in a given asset (if any of the asset is held). We illustrate the differences that arise in the shape of this efficient frontier when such constraints are present. We present some heuristic algorithms based upon genetic algorithms. We used a new operator and posed a multi-objective optimization function to achieve the results.

Keywords: portfolio optimization, efficient frontier

## 1. INTRODUCTION

The selection of an appropriate portfolio of assets in which to invest is an essential component of fund management. Although a large proportion of portfolio selection decisions are taken on a qualitative basis, quantitative approaches to selection are becoming more widely adopted. Markowitz [34,35] set up a quantitative framework for the selection of a portfolio.

In this framework it is assumed that asset returns follow a multivariate normal distribution. This means that the return on a portfolio of assets can be completely described by the expected return and the variance (risk). For a particular universe of assets, the set of portfolios of assets that offer the minimum risk for a given level of return form the efficient frontier. The portfolios on the efficient frontier can be found by quadratic programming (QP). The strengths of this approach are that QP solvers are available and efficient in terms of computing time. The solutions are optimal and the selection process can be constrained by practical considerations which can be written as linear constraints. The weaknesses are of two kinds: (1) the underlying assumption of multivariate normality is not sustainable. The distribution of individual asset returns tends to exhibit a higher probability of extreme values than is consistent with normality (statistically this is known as leptokurtosis). This departure from multivariate normality means that distribution moments higher than the first two moments (expected return and variance) need to be considered to fully describe portfolio behaviour. (2) integer constraints that limit a portfolio to have a specified number of assets, or to impose limits on the proportion of the portfolio held in a given asset (if any of the asset is held) cannot easily be applied. Constraints of this type are of practical significance. This paper examines the use of three standard heuristic methods in portfolio selection.

The method considered here is a genetic algorithm The attraction of this approache is that it is effectively independent of the objective function adopted. This means that the Markowitz quadratic objective function can potentially be replaced in the light of the first set of weaknesses identified above. In addition, the imposition of integer constraints is straightforward. In this paper the heuristics that we have developed are described and their performance compared with that of QP for the construction of the unconstrained efficient frontier (UEF). This approach allows the closeness of the heuristic solutions to optimality to be measured. The performance of the heuristic methods in constructing the efficient frontier in the presence of a constraint fixing the number of assets in the selected portfolio is demonstrated. This frontier is called the cardinality constrained efficient frontier (CCEF).

## 2. FORMULATION

In this section we formulate the cardinality constrained mean-variance portfolio optimisation problem. We first formulate the unconstrained portfolio optimisation problem and illustrate how to calculate the efficient frontier. We then comment on the approaches presented in the literature that have used a different objective function. Finally we formulate the cardinality constrained problem.

### 2.1 Unconstrained problem

Let: N be the number of assets available
R* be the desired expected return Then the decision variables are: The proportion held of asset i (i=1,...,N) and using the standard Markowitz mean-variance approach [14,15,34,35,42] we have that the unconstrained portfolio optimisation problem is:

minimise (A) subject to conditions (B)

Equation (A) minimises the total variance (risk) associated with the portfolio whilst equations (B) ensures that the portfolio has an expected return of R*. Equations (B) also ensures that the proportions add to one. This formulation is a simple nonlinear (quadratic) programming problem for which computationally effective algorithms exist so there is (in practice) little difficulty in calculating the optimal solution for any particular data set. Note here that the above formulation can be expressed in terms of a correlation between assets i and j and the standard deviations si, sj in returns for these assets.

### 2.2 Efficient frontier

By resolving the above QP for varying values of R* we can trace out the efficient frontier, a smooth non-decreasing curve that gives the best possible tradeoff of risk against return, i.e. the curve represents the set of Pareto-optimal (non-dominated) portfolios. Throughout this paper we refer this curve as the unconstrained efficient frontier (UEF). For the unconstrained case it is standard practice to trace out the efficient frontier by introducing a weighting parameter and considering:

minimise (X) subject to constraints (Y)

There is a case that represents maximise expected return (irrespective of the risk involved) and the optimal solution will involve just the single asset with the highest return. In another case one represents minimise risk (irrespective of the return involved) and the optimal solution will typically involve a number of assets. Values of □ satisfying 0<□<1 represent an explicit tradeoff between risk and return, generating solutions between the two extremes □=0 and □=1. As before, by resolving this QP for varying values of □, we can trace out the efficient frontier. To see that this is so consider a particular value of □, e.g. □=0.25. By varying □ (varying the slope of the iso-profit lines) and solving equations (5)-(7) we can trace out *exactly* the same efficient frontier curve as we would obtain by solving equations for varying values of R*.

### 2.3 Other objectives

Departures from the standard Markowitz mean-variance approach presented above include the following considerations: (a) whether variance is considered to be an adequate measure of the risk associated with the portfolio or not; and (b) including transaction costs associated with changing from a current portfolio to a new portfolio.

### 2.4 Constrained problem

In order to extend our formulation to the cardinality constrained case let: K be the desired number of assets in the portfolio be the minimum proportion that must be held of asset i (i=1,...,N) if any of asset I is held and i be the maximum proportion that can be held of asset i (i=1,...,N) if any of asset i is held where we must have 0≤□i≤□i≤1 (i=1,...,N). In practice □i represents a "min-buy" or "minimum transaction level" for asset i and □i limits the exposure of the portfolio to asset i. Introducing zero-one decision variables:
zi = 1 if any of asset i (i=1,...,N) is held = 0 otherwise the cardinality constrained portfolio optimisation problem is minimise (D) subject to some conditions

Equation (D) minimises the total variance (risk) associated with the portfolio whilst conditions ensures that the portfolio has an expected return of R*. Conditions also ensures that the proportions and that exactly K assets are held. The objective function involving as it does the covariance matrix, is positive semi-definite [8,9,18,43,46] and hence we are minimising a convex function. Note here that we have explicitly chosen to formulate this problem with an equality (rather than an inequality ≤) with respect to the number of assets in the portfolio. This is because if we can solve the equality constrained case then any situation involving inequalities (lower or upper limits on the number of assets in the portfolio) can be easily dealt with.

Under our approach the decision-maker will be faced with a different CCEF for each value of K and must explicitly consider the tradeoffs involved in deciding which portfolio to adopt. An illustration of this is given in Section 5.5 below.

### 2.5 Practical constraints

There are a number of constraints that can be added to our constrained formulation to better reflect practical portfolio optimisation. *(a) Class constraints* Let □m, m=1,...,M be M sets of assets that are mutually exclusive. Class constraints limit the proportion of the portfolio that can be invested in assets in each class. Let Lm be the lower proportion limit and Um be the upper proportion limit for class m then the class constraints are:

Lm ≤wi ≤Um m=1,...,M

Such constraints typically limit the "exposure" of the portfolio to assets with a common characteristic. For example typical classes might be oil stocks, utility stocks, telecommunication stocks, etc. The heuristics presented in this paper do not deal with constraints of this type. *(b) Assets in the portfolio* Assets which must be in the portfolio can be accommodated in our formulation simply by setting zi to one for any such asset i. Although we do not present it below the changes required to our heuristics to deal with this are trivial.

## 3. CONSTRAINED EFFICIENT FRONTIER

One aspect of constrained portfolio optimisation that appears to have received no attention in the literature is the fact that in the presence of constraints of the type we have considered above the efficient frontier is markedly different from the UEF. In this section we illustrate this.

### 3.1 Minimum proportion constraints

To illustrate the effect of imposing a nonzero minimum proportion the efficient frontier for the case where □i=0.24 and □i=1, (i=1,2,3,4). It is clear that again the efficient frontier is discontinuous and has portions that are invisible to an exact approach based upon weighting.

## 4. HEURISTIC ALGORITHMS

In this section we outline the three heuristic algorithms based upon genetic algorithms, tabu search and simulated annealing that we have developed for finding the CCEF. We also discuss here any application of these techniques to portfolio optimization previously reported in the literature. Note here that all of our heuristics use the weighted formulation presented in Section 3.2 above.

With respect to the first of these two reasons the CCEF, as found by the heuristics presented below, for the four asset example (K=2) considered previously. Hence, for this example, our heuristics have found all portions (visible or invisible) of the CCEF. With respect to the second of these reasons it might appear that we would be in a better position to design a heuristic algorithm if the equality constraint relating to return were changed to an inequality, i.e. to wi□i ≥R*.

4.1 Genetic algorithms

A genetic algorithm (GA) can be described as an "intelligent" probabilistic search algorithm. The theoretical foundations of GAs were originally developed by Holland [24]. GAs are based on the evolutionary process of biological organisms in nature. During the course of evolution, natural populations evolve according to the principles of natural selection and "survival of the fittest". Individuals which are more successful in adapting to their environment will have a better chance of surviving and reproducing, whilst individuals which are less fit will be eliminated. This means that the *genes* from the highly fit individuals will spread to an increasing number of individuals in each successive generation. The combination of good characteristics from highly adapted parents may produce even more fit offspring. In this way, species evolve to become increasingly better adapted to their environment. A GA simulates these processes by taking an initial population of individuals and applying genetic operators in each reproduction. In optimisation terms, each individual in the population is encoded into a string or *chromosome* which represents a possible *solution* to a given problem. The fitness of an individual is evaluated with respect to a given objective function. Highly fit individuals or *solutions* are given opportunities to reproduce by exchanging pieces of their genetic information, in a *crossover* procedure, with other highly fit individuals. This produces new "offspring" solutions (i.e. *children*), which share some characteristics taken from both parents. Mutation is often applied after crossover by altering some genes in the strings. The offspring can either replace the whole population (*generational* approach) or replace less fit individuals (*steady-state* approach). This evaluation-selection-reproduction cycle is repeated until a satisfactory solution is found.

A more comprehensive overview of GAs can be found in [4,38,40,41]. Arnone, Loraschi and Tettamanzi [3] presented a GA for the unconstrained portfolio optimisation problem, but with the risk associated with the portfolio being measured by downside risk rather than by variance. Computational results were presented for one problem involving 15 assets. Loraschi, Tettamanzi, Tomassini and Verda [32] presented a distributed GA for the unconstrained portfolio optimisation problem based on an island model where a GA is used with multiple independent subpopulations (each run on a different processor) and highly-fit individuals occasionally migrate between the subpopulations. Computational results were presented for one problem involving 53 assets comparing their distributed GA with the GA presented in [3].

4.2 Genetic algorithm heuristic

In our GA heuristic the chromosome representation of a solution has two distinct parts, a set Q of K distinct assets and K real numbers si . We need an iterative procedure to ensure that the constraints relating to the upper limits □i are satisfied. Note here that Algorithm 1 can be viewed as a heuristic for solving the with a given set of K assets. Whilst, obviously, this QP could be solved optimally this would not lead to a computationally efficient heuristic (examining as we do a large number of possible solutions). Note here that the strategy adopted in Algorithm 1, namely to change (if possible) the GA solution into a feasible solution for the original problem, is a strategy that we have used, with success, in our previous GA work [7,13]. We used a population size of 100. Parents were chosen by binary tournament selection which works by forming two pools of individuals, each consisting of two individuals drawn from the population randomly. The individuals with the best fitness, one taken from each of the two tournament pools, are chosen to be parents. Children in our GA heuristic are generated by uniform crossover. In uniform crossover two parents have a single child. If an asset i is present in both parents it is present in the child (with an associated value si randomly chosen from one or other parent). If an asset i is present in just one parent it has probability 0.5 of being present in the child. Children are also subject to mutation, multiplying by 0.9 or 1.1. We used a steady-state population replacement strategy. With this strategy each new child is placed in the population as soon as it is generated (replacing a suitably chosen member of the population). In our GA we choose to replace the member of the population with the worst objective function value.

## 5. COMPUTATIONAL RESULTS

In this section we present computational results for the heuristic algorithm we have presented above for finding the CCEF.

### 5.1 Test data sets

To test our heuristics we constructed five test data sets by considering the stocks involved in Bovespa. We had 252 values for each stock from which to calculate (weekly) returns and covariances and the size of our five test problems ranged from N=31 to N=225.

### 5.2 Unconstrained efficient frontier

In order to initially test the effectiveness of our GA heuristics we first used them to find the UEF. Adopting this approach has the advantage that (as mentioned in Section 2 above) the UEF can be exactly calculated via QP so our heuristic results can be compared with benchmark optimal solutions. The reason for doing this comparison is simply that for the CCEF we have no way of calculating the exact efficient frontier for problems of the size we are considering, and hence no way of benchmarking our heuristics against the exact solution. We would anticipate that, unless our heuristics are able to find the UEF to a reasonable degree of accuracy, they are unlikely to be able to find the CCEF.

#### 5.2.1 UEF calculation

In the computational results presented below we took 2000 return (R*) values, hence calculating 2000 distinct points on the continuous exact UEF. For adjacent points we used linear interpolation to approximate the exact UEF. This calculation of the percentage deviation of each portfolio from the (linearly interpolated) exact UEF is not as trivial as it might at first sight appear and we consider this below.

#### 5.2.2 Percentage deviation calculation

There are two basic issues: (a) how we measure the percentage deviation ("distance") of a portfolio from a continuous efficient frontier; and (b) how we quantify this percentage deviation in the case of a linearly interpolated efficient frontier.

With respect to second issue mentioned above, quantifying the percentage deviation in the case of a linearly interpolated efficient frontier we, in order to work in commensurate units, used the portfolio standard deviation (rather than variance) in computing percentage deviation.

#### 5.2.3 Results

With regard to all the computational results reported in this paper we examined 50 different  values (E=50, Algorithms 2-4). With regard to the number of iterations T* (see agorithms 2-4) we used T*=1000N for the GA heuristic. These values mean that (excluding initialisation) each heuristic vluates exactly 1000N solutions each value of X. . We show, for each of our five data sets and each of our three heuristics:

(a) the median percentage error (b) the mean percentage error (c) the total computer time in seconds.

Note here however that, as all of our algorithms are heuristics, we can provide no guarantees as to the quality (deviation from the exact CCEF) of any particular portfolio in the set H. In particular a portfolio in H could be on the exact CCEF or could be dominated by another portfolio (not in H) which is on the exact CCEF. We show, for each of our five data sets and each of our three heuristics: (a) the median percentage error (b) the mean percentage error (c) the number of (undominated) efficient points.

For some data sets there are considerable differences in the percentage error measures, indicating that the algorithms give significantly different results. Hence we would envisage that a sensible approach to the cardinality constrained portfolio optimisation problem in practice would be to run all three heuristics and to pool their results in an obvious fashion (i.e. combine the three sets of undominated points given by the three algorithms together into one set and eliminate from this new set those points which are dominated). Note here that we stated before (Section 4) that using our heuristics it is possible to gain information about those portions of the CCEF that would be invisible to an exact approach based upon weighting.

## 6. CONCLUSIONS

In this paper we have considered the problem of calculating the efficient frontier for the cardinality constrained portfolio optimisation problem. We highlighted the differences that arise in the shape of this efficient frontier as compared with the unconstrained efficient frontier. Computational results were presented for genetic algorithms for finding the cardinality constrained efficient frontier. These indicated that a sensible approach was to pool their results.

## REFERENCES


[1] E.H.L. Aarts and J.K. Lenstra (editors), *Local search in combinatorial optimization*. Wiley, (1997).

[2] C.J. Adcock and N. Meade, A simple algorithm to incorporate transaction costs in quadratic optimisation. *European Journal of Operational Research* 79 (1994) 85-94.

[3] S. Arnone, A. Loraschi and A. Tettamanzi, A genetic approach to portfolio selection. *Neural Network World* 6 (1993) 597-604.

[4] T. Bäck, D.B. Fogel and Z. Michalewicz (editors), *Handbook of evolutionary computation*. Oxford University Press, (1997).

[5] J.E. Beasley, OR-Library: distributing test problems by electronic mail. *Journal of the Operational Research Society* 41 (1990) 1069-1072.

[6] J.E. Beasley, Obtaining test problems via Internet. *Journal of Global Optimization* 8 (1996) 429-433.

[7] J.E. Beasley and P.C. Chu, A genetic algorithm for the set covering problem. *European Journal of Operational Research* 94 (1996) 392-404.

[8] D. Bienstock, Computational study of a family of mixed-integer quadratic programming problems. In *Integer Programming and Combinatorial Optimization: 4th International IPCO Conference, Copenhagen, Denmark, May 1995*

[9] E.J. Elton and M.J. Gruber, *Modern portfolio theory and investment analysis*. Wiley, (1995).

[10] H. Markowitz, Portfolio selection, *Journal of Finance* 7 (1952) 77-91.

[11] H.M. Markowitz, *Portfolio selection: efficient diversification of investments*. Wiley, (1959).