Heuristic Approaches for Portfolio Optimization †

Manfred Gilli (manfred.gilli@metri.unige.ch)
Department of Econometrics, University of Geneva, 1211 Geneva 4, Switzerland.

Evis Këllezi (evis.kellezi@metri.unige.ch)
Department of Econometrics and FAME, University of Geneva, 1211 Geneva 4, Switzerland.

Abstract. Constraints on downside risk, measured by shortfall probability, expected shortfall, semi-variance etc., lead to optimal asset allocations which differ from the mean-variance optimum. The resulting optimization problem can become quite complex as it exhibits multiple local extrema and discontinuities, in particular if we also introduce constraints restricting the trading variables to be integers, constraints on the holding size of assets or on the maximum number of different assets in the portfolio. In such situations classical optimization methods fail to work efficiently and heuristic optimization techniques can be the only way out. The paper shows how a particular optimization heuristic, called threshold accepting, can be successfully employed to solve complex portfolio choice problems.

Keywords: Portfolio Optimization, Downside Risk Measures, Heuristic Optimization, Threshold Accepting.

JEL codes: G11, C61, C63

1. Introduction

The fundamental goal of an investor is to optimally allocate his investments between different assets. The pioneering work of Markowitz (1952) introduced the mean-variance optimization as a quantitative tool which allows to make this allocation by considering the trade-off between risk, measured by the variance of the future asset returns, and return. The assumptions of the normality of the returns and of the quadratic investor's preferences allow to simplify the problem in a relatively easy to solve quadratic program.

Notwithstanding its popularity, this approach has also been subject to a lot of criticism. Alternative approaches attempt to conform the fundamental assumptions to reality by dismissing the normality hypothesis in order to account for the fat-tailedness and the asymmetry of the asset returns. Consequently, other measures of risk, like e.g.

† Preliminary and incomplete version. Presented at the sixth International Conference on Computing in Economics and Finance of the Society for Computational Economics, Barcelona, July 6–8, 2000.

Value at Risk (VaR), expected shortfall, mean absolute deviation, semi-variance etc. are employed, leading to problems that can not always be reduced to standard linear or quadratic programs. The resulting optimization problem often becomes quite complex as it exhibits multiple local extrema and discontinuities, in particular if we introduce constraints restricting the trading variables to be integers, constraints on the holding size of assets, on the maximum number of different assets in the portfolio, etc.

In such situations classical optimization methods fail to work efficiently and heuristic optimization techniques can be the only way out. They are relatively easy to implement and computationally attractive. This paper builds on work by Dueck and Winker (1992) who first applied a heuristic optimization technique, called Threshold Accepting, to portfolio choice problems. We show how this technique can be successfully employed to solve complex portfolio choice problems where risk is characterized by Value at Risk and Expected Shortfall.

In Section 2 we outline the different frameworks for portfolio choice as well as the most frequently used risk measures. Section 3 gives a general representation of the threshold accepting heuristic we use. The performance and efficiency of the algorithm is discussed in Section 4 by, first, comparing it with the quadratic programming solutions in the mean-variance framework and, second, applying the algorithm to problems minimizing the portfolio expected shortfall or VaR conditional to some return constraints. Section 6 concludes.

2. Approaches to the portfolio choice problem

2.1. The mean-variance approach

The mean-variance optimization is certainly the most popular approach to portfolio choice. In this framework, the investor is faced with a trade-off between the profitability of his portfolio, characterized by the expected return, and the risk, measured by the variance of the portfolio returns. The first two moments of the portfolio future return are sufficient to define a complete ordering of the investors preferences. This strong result is due to the simplistic hypothesis that the investors' preferences are quadratic and the returns are normally distributed.

Denoting by x_i, $i = 1, \ldots, n_A$, the amount invested in asset i out of an initial capital v^0 and by r_i, $i = 1, \ldots, n_A$, the log-returns for each asset over the planning period, then the expected return on the portfolio defined by the vector $x = (x_1, x_2, \ldots, x_{n_A})'$ is given as

$$\mu(x) = \frac{1}{v^0} \sum_{i=1}^{n_A} E(r_i)\, x_i = \frac{1}{v^0}\, x' E(r)\,.$$

The variance of the portfolio return is

$$\sigma^2(x) = x'\, Q\, x\,,$$

where Q is the matrix of variances and covariances of the vector of returns r.

Thus the mean-variance efficient portfolios, defined as having the highest expected return for a given variance and the minimum variance for a given expected return, are obtained by solving the following quadratic program

$$\begin{array}{rcl} \min\limits_x & \tfrac{1}{2}\, x'\, Q\, x & \\ \sum_j x_j r_j & \geq & \rho\, v^0 \\ \sum_j x_j & = & v^0 \\ x_j^\ell \leq x_j \leq x_j^u & & j = 1, \ldots, n_A\,. \end{array} \tag{1}$$

for different values of ρ, where ρ is the required return on the portfolio. The vectors $x_j^\ell, x_j^u, j = 1, \ldots, n_A$ represent constraints on the minimum and maximum holding size of the individual assets.

The implementation of the Markowitz model with n_A assets requires n_A estimates of expected returns, n_A estimates of variances and $n_A(n_A - 1)/2$ correlation coefficients.

Several efficient algorithms exist for computing the mean-variance portfolios. Early successful parametric quadratic programming methods include the critical-line algorithm and the simplex method.

2.2. Scenario generation

An alternative approach to the above optimization setting is the scenario analysis where uncertainty about the future returns is modeled through a set of possible realizations, called scenarios. Scenarios of future outcomes can be generated relying on a model, past returns or experts' opinions.

A simple approach is to use empirical distributions computed from past returns as equiprobable scenarios. Observations of returns over n_S overlapping periods of length Δt are considered as the n_S possible outcomes (or scenarios) of the future returns and a probability of $1/n_S$ is assigned to each of them.

Assume that we have T historical prices p^h, $h = 1,\ldots,T$ of the assets under consideration. For each point in time, we can compute the realized return vector over the previous period of length Δt, which will further be considered as one of the n_S scenarios for the future returns on the assets. Thus, for example, a scenario r_j^s for the return on asset j is obtained as

$$r_j^s = \log(p_j^{t+\Delta t}/p_j^t)\,. \tag{2}$$

For each asset, we obtain as many scenarios as there are overlapping periods of length Δt, i.e. n_S. In this setting problem (1) becomes

$$\begin{aligned} \min_x \ \frac{1}{n_S}\sum_{s=1}^{n_S}\left(\sum_{j=1}^{n_A} r_j^s x_j - \frac{1}{n_S}\sum_{j=1}^{n_A}\sum_{s=1}^{n_S} r_j^s x_j\right)^2 & \\ \frac{1}{n_S}\sum_{j=1}^{n_A}\sum_{s=1}^{n_S} r_j^s x_j \ \geq \ \rho\, v^0 & \\ \textstyle\sum_j x_j \ = \ v^0 & \\ x_j^\ell \leq x_j \leq x_j^u \qquad j = 1,\ldots,n_A\,. & \end{aligned} \tag{3}$$

2.3. Mean downside-risk framework

If we denote by v the future portfolio value, i.e. the value of the portfolio by the end of the planning period, then the probability

$$P(v < \mathtt{VaR}) \tag{4}$$

that the portfolio value falls below the VaR level, is called the *shortfall probability*. The conditional mean value of the portfolio given that the portfolio value has fallen below VaR, called the *expected shortfall*, is defined as

$$E(v \mid v < \mathtt{VaR})\,. \tag{5}$$

Other risk measures used in practice are the *mean absolute deviation*

$$E(|v - Ev|)$$

and the *semi-variance*

$$E((v - Ev)^2 \mid v < Ev)$$

where we consider only the negative deviations from the mean.

Maximizing the expected value of the portfolio for a certain level of risk characterized by one of the measures defined above leads to alternative ways of describing the investor's problem (e.g. Leibowitz and

Kogelman (1991), Lucas and Klaassen (1998) and Palmquist, Uryasev and Krokhmal (1999)). Earlier related work had suggested a safety-first approach (see e.g. Arzac and Bawa (1977) and Roy (1952)).

For example, if the risk profile of the investor is determined in terms of VaR, a mean-VaR efficient portfolio would be solution of the following optimization problem:

$$\begin{aligned} \max_x \ & Ev \\ P(v < \texttt{VaR}) \ & \leq \ \beta \\ \textstyle\sum_j x_j \ & = \ v^0 \\ x_j^\ell \leq x_j \leq x_j^u & \qquad j = 1, \ldots, n_A \,. \end{aligned} \tag{6}$$

In other words, such an investor is trying to maximize the future value of his portfolio, requiring that the probability that the future value of his portfolio falls below VaR is not greater than β.

If the uncertainty in the future asset returns is handled via scenario generation, the above optimization can be further explicited as follows:

$$\begin{aligned} \min_x \ & -\frac{1}{n_S}\sum_{s=1}^{n_S} v^s \\ \#\{s \,|\, v^s < \texttt{VaR}\} \ & \leq \ \beta\, n_S \\ \textstyle\sum_j x_j \ & = \ v^0 \\ x_j^\ell \leq x_j \leq x_j^u & \qquad j = 1, \ldots, n_A \,. \end{aligned} \tag{7}$$

Furthermore, it would be realistic to consider an investor who cares not only for the shortfall probability, but also for the extent to which his portfolio value can fall below the VaR level. In this case, the investor's risk profile is defined via a constraint on the expected shortfall tolerated ν if the portfolio value falls below VaR. Then the mean-expected shortfall efficient portfolios are solutions of the following program for different values of ν:

$$\begin{aligned} \max_x \ & Ev \\ E[v \,|\, v < \texttt{VaR}] \ & \geq \ \nu \\ \textstyle\sum_j x_j \ & = \ v^0 \\ x_j^\ell \leq x_j \leq x_j^u & \qquad j = 1, \ldots, n_A \,. \end{aligned} \tag{8}$$

Again if the future returns are generated by scenarios, the optimization problem becomes:

$$\begin{aligned} \min_x \; -\frac{1}{n_S}\sum_{s=1}^{n_S} v^s & \\ \tfrac{1}{\#\{s\,|\,v^s<\mathtt{VaR}\}} \sum_{s|v^s<\mathtt{VaR}} v^s \;&\geq\; \nu \\ \#\{s\,|\,v^s<\mathtt{VaR}\} \;&\leq\; \beta\, n_S \\ \textstyle\sum_j x_j \;&=\; v^0 \\ x_j^\ell \leq x_j \leq x_j^u \qquad & j=1,\ldots,n_A\,. \end{aligned} \tag{9}$$

3. The threshold accepting optimization heuristic

Heuristic approaches prove useful in situations where the classical optimization methods fail to work efficiently. Heuristic optimization techniques like simulated annealing (Kirkpatrick *et al.*, (1983)) and genetic algorithms (Holland (1975)) are used with increasing success in a variety of disciplines. The reason for their success is that they are relatively easy to implement and that the cost of computing power is no longer a matter of concern.

Threshold accepting (TA) has been introduced by Dueck and Scheuer (1990) as a deterministic analog to simulated annealing. It is a refined local search procedure which escapes local minima by accepting solutions which are not worse by more than a given threshold. The algorithm is deterministic in the sense that we fix a number of iterations and explore the neighbourhood with a fixed number of steps during each iteration. The threshold is decreased successively and reaches the value of zero in the last round.

The threshold accepting algorithm has the advantage of an easy parameterization, it is robust to changes in problem characteristics and works well for many problem instances.

Let us formalize our optimization problem as $f : \mathcal{X} \rightarrow \mathbb{R}$ where $\mathcal{X}$ is a discrete set and where we may have more then one optimal solution defined by the set

$$\mathcal{X}_{\min} = \{x \in \mathcal{X} \mid f(x) = f_{\mathrm{opt}}\} \tag{10}$$

with

$$f_{\mathrm{opt}} = \min_{x\in\mathcal{X}} f(x)\,. \tag{11}$$

The threshold accepting heuristic described in algorithm 1 will after completion provide us with a solution $x \in \mathcal{X}_{\min}$ or a solution close to an element in $\mathcal{X}_{\min}$. The complexity of the algorithm is $\mathcal{O}(niter \times steps)$.

Algorithm 1 Pseudo-code for the threshold accepting algorithm.

Initialize *niter* and *steps*
Initialize sequence of thresholds th_r, $r = 1, 2, \ldots, niter$
Generate starting point $x^0 \in \mathcal{X}$
for $r = 1$ to *niter* **do**
    for $i = 1$ to *steps* **do**
        Generate $x^1 \in \mathcal{N}_{x^0}$ (neighbour of x^0)
        if $f(x^1) < f(x^0) + th_r$ **then**
            $x^0 = x^1$
        end if
    end for
end for

The parameters of the algorithm are the number of iterations *niter*, the number of steps per iteration *steps* and the sequence of thresholds *th*. In practice we start with the definition of the objective function, which can be a non-trivial task if f comprises several dimensions. Second we construct a mapping $\mathcal{N} : \mathcal{X} \to 2^{\mathcal{X}}$ which defines for each $x \in \mathcal{X}$ a neighbourhood $\mathcal{N}(x) \subset \mathcal{X}$. Third we define the sequence of thresholds by exploring the neighbourhood of randomly selected elements $x \in \mathcal{X}$.

These different steps of the implementation and parameterization of the algorithm will be illustrated with the application presented in the next section.

4. Application

The working of the TA algorithm is first illustrated to solve a standard mean-variance optimization problem for which the solution is also computed with the quadratic programming algorithm which will be used as a benchmark. Second we apply the TA algorithm to a non-convex optimization problem with integer variables and a variety of constraints such as holding and trading size.

4.1. Mean-variance optimization

In the following application we consider an investment opportunity set of ten assets from the Swiss market index (SMI) and cash. The annual mean return r and the matrix of variances and covariances Q are based on the closing prices of the last 90 trading days before 30–6–1999.

The mean-variance optimization problem has already been defined in (1). The following is a reformulation of the problem where the initial capital v^0 has been normalized to one:

$$\begin{aligned} &\min_{\omega} \ \tfrac{1}{2}\,\omega'\,Q\,\omega \\ &\omega' r \ \geq \ \rho \\ &\iota'\omega \ = \ 1 \\ &\omega_j^l \leq \omega_j \leq \omega_j^u \qquad j = 1, \ldots, n_A + 1\,. \end{aligned}$$

The composition of the portfolio is defined by the shares $\omega_i = x_i / v^0$ and ω_{n_A+1} is the proportion of cash in the portfolio. The risk-free return of cash is r_{n_A+1}.

Definition of objective function

The variance can now be minimized by exploring with the threshold accepting algorithm 1 the elements in the set $\mathcal{X}$ which satisfy the constraints. However a better way is to accept in the search process solutions which violate the return constraint. This can be done by minimizing the following objective function

$$F(\omega) = V(\omega) + p\ (\rho - R(\omega))$$

where p is a penalty function defined as

$$p = \begin{cases} \frac{V_{\max} - V_{\min}}{\rho - \overline{R}} & \text{if} \quad \rho > R(\omega) \\ 0 & \text{otherwise.} \end{cases}$$

$V(\omega)$ and $R(\omega)$ denote respectively the variance and the return of a portfolio defined by ω. The values for $V_{\max}$, $V_{\min}$ and $\overline{R}$ which define the scaling constant $(V_{\max} - V_{\min})/(\rho - \overline{R})$ are estimated from 1000 randomly drawn portfolios.

Definition of neighbourhood

To generate a point x^1 in the neighbourhood $\mathcal{N}_{x^0}$ of a given point x^0 we draw with a probability $1/(n_A + 1)$ two assets i and j out of all n_A assets and cash. The amount of i and j in the portfolio is ω_i, respectively ω_j. We then sell a fraction q of asset i, i.e. $q\,\omega_i$ and buy for the corresponding amount asset j. After this move the amount of i and j in the portfolio is $(1-q)\omega_i$, respectively $\omega_j + q\,\omega_i$. The fraction q is a fixed parameter.

In order to avoid short selling and respect the constraints on the holding size of the assets the procedure for the selection of a neighbour solution must be refined. Algorithm 2 describes the procedure of the selection of a neighbour-solution in detail.

Algorithm 2 Definition of neighbourhood.

Select two assets i and j with probability $1/(n_A+1)$
$t = q\,\omega_i$
if $(\omega_i - t) \geq \omega_i^l$ **then**
  $\omega_i = \omega_i - t$
else
  $t = \omega_i$
  $\omega_i = 0$
end if
if $(\omega_j + t) \leq \omega_j^u$ **then**
  $\omega_j = \omega_j + t$
else
  $\omega_{nA+1} = \omega_{nA+1} + t$
end if

Definition of thresholds

In order to define the sequence of thresholds we compute the empirical distribution of the distance of the objective function evaluated at random points and its neighbours. Figure 1 shows this empirical distribution computed from 5000 random points. The quantiles which determine the sequence of thresholds are then $10^{-3}\left[\,22.5\ \ 3.4\ \ 1.1\ \ 0.7\ \ 0.4\ \ 0\,\right]$.

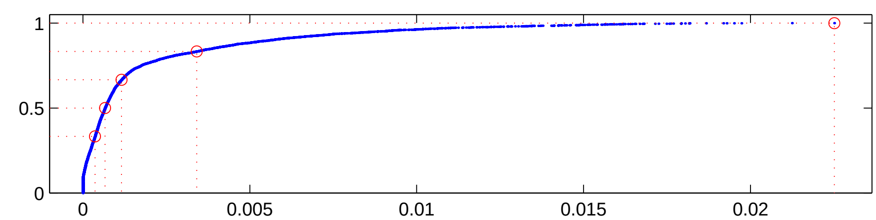

Figure 1. Empirical distribution of distance between x^0 and neighbours x^1.

Choosing *niter* $= 6$ and *steps* $= 1000$ we have determined all the parameters of our TA algorithm. The following figure illustrates how the algorithm searches its way to the solution. At the optimal solution the expected return and the variance are practically the same for the QP and TA algorithm. The optimal portfolio contains asset 3, 5 and 8 and cash (column 11). The weights of the assets in the optimal portfolio for both algorithms are given in Figure 3.

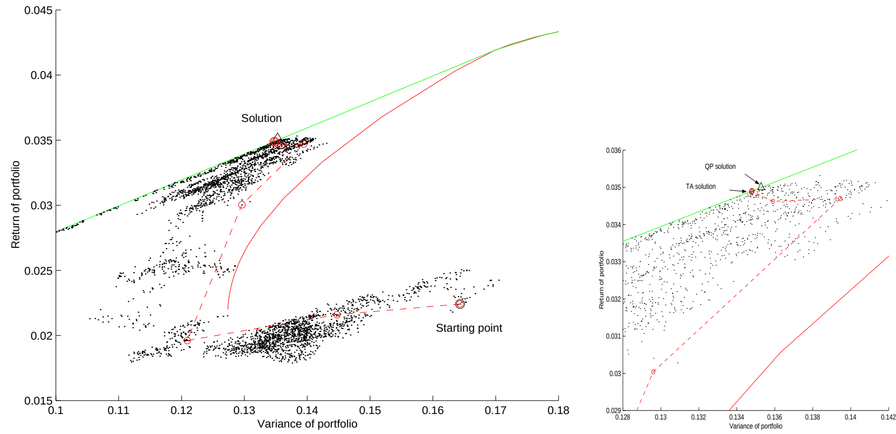

Figure 2. Working of the TA algorithm. Efficient frontier with cash (upper line) and without cash (lower curve).

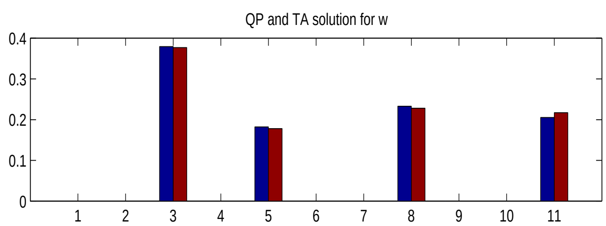

Figure 3. Composition of the optimal portfolio for QP and TA.

4.2. Mean downside risk optimization

Our second illustration of the working of the TA algorithm is a non-convex optimization problem with integer variables and a variety of constraints such as holding and trading size.

In the following the quantity of each asset in the portfolio is defined by an integer number. The generation of neighbours $x^1 \in \mathcal{N}_x^0$ to a given solution x^0 is again performed by drawing randomly two assets i and j. We then sell k_i assets i, transfer the amount to the cash and buy k_j assets j from cash. In order to make sure that each transfer is of approximatively of the same amount, the number of assets k_i and k_j to be transferred are defined as $k_i = \lceil \frac{p_{\max}}{p_i} \rceil$ and $k_j = \lceil \frac{p_{\max}}{p_j} \rceil$. This procedure is summarized in algorithm 3 where we omitted the details necessary to check for short selling and holding constraints.

Using the same data set as for the previous problem but considering an investment opportunity set of 20 assets (including cash) we now

Algorithm 3 Definition of neighbourhood in case of integer variables.

Randomly select asset i to sell
$x_i = x_i - k_i$
cash = cash + $k_i\, p_i^0$
Randomly select asset j to buy
$x_j = x_j + k_j$
cash = cash − $k_j\, p_j^0$

solve the mean-VaR problem defined in (7). To compute the capital v^s at the end of the planning period we use simulated prices p^s, computed as

$$p^s = p^0\, r^s \qquad s = 1, \dots, n_S$$

where the rate of return r^s has been defined in (2). We assume an initial capital $v^0 = 800\,000$, a shortfall probability $\beta = 0.05$ and $VaR = 750\,000$. Figure 4 shows the results of the TA algorithm with the setting $niter = 6$, $steps = 1500$ and $k = 4$.

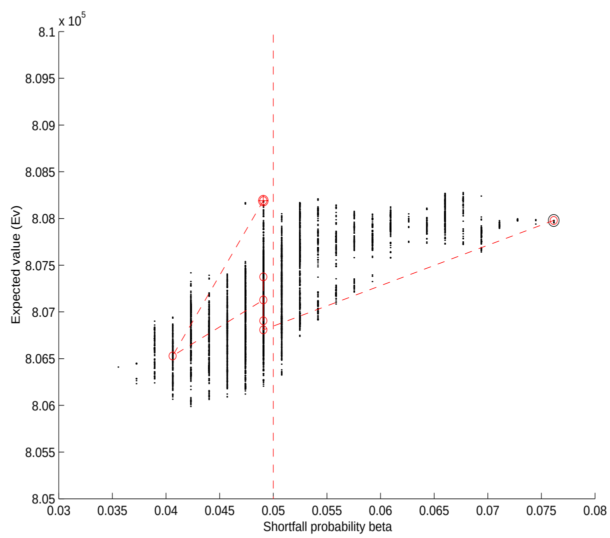

Figure 4. Search path of the TA algorithm in the β, $E(v)$ plane.

The solution verifies $VaR = 749\,760$, $E(v) = 808\,190$ and an empirical shortfall probability of .0491. The composition of the optimal portfolio is given in figure 5.

In figure 4 we observe that the solutions lie in planes. The reason for this is the integer formulation of the problem.

Figure 6 illustrates the working of the TA algorithm in the β, VaR plane and in figure 7 we see its working in the $E(v)$, VaR plane.

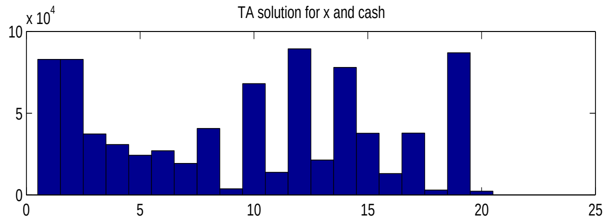

Figure 5. Optimal portfolio computed by TA for the mean VaR problem.

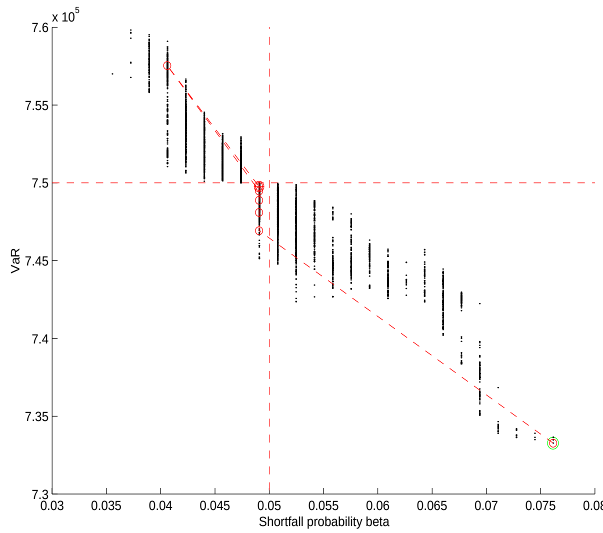

Figure 6. Search path of the TA algorithm in the β, VaR plane.

5. Concluding remarks

In this paper we tried to illustrate how heuristic optimization algorithms like the threshold accepting method can be successfully applied to solve realistic non-convex portfolio optimization problems. We showed that, in the cases where these problems contain non-linear and non-convex constraints, the heuristic methods are the only reasonable way out. Examples of these situations can be problems where constraints on downside risk preferences are introduced, where the solutions are required to be integers, etc.

We mainly focus on the cases where the distribution of the asset future returns are modelled by equally weighted scenarios of past returns. The sensitivity of optimized portfolios with respect to alternative scenario generations procedures should be further investigated.

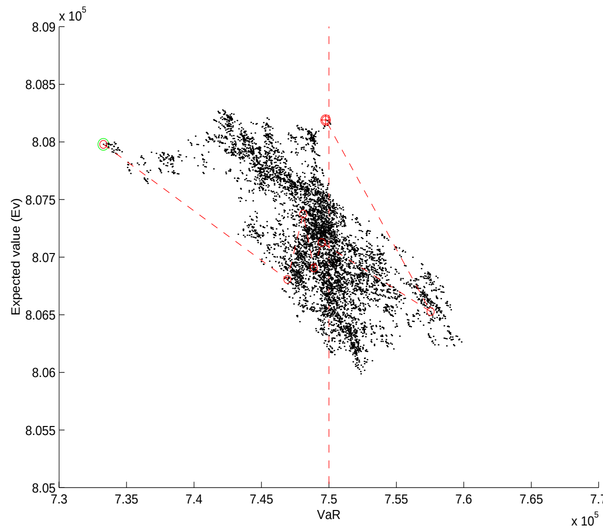

Figure 7. Search path of the TA algorithm in the $E(v)$, VaR plane.

Acknowledgements

This paper benefited significantly from conversations with Peter Winker. We would like to thank him for valuable suggestions and comments. Financial support from the Swiss National Science Foundation (project 12–5248.97) is gratefully acknowledged.

References

Arzac, V.S. and V.S. Bawa, 1977. Portfolio Choice and Equilibrium in Capital Markets with Safety-First Investors. *Journal of Financial Economics* 4, 277–288.

Dembo, R.S. and D. Rosen, 1999. The Practice of Portfolio Replication: A Practical Overview of Forward and Inverse Problems. *Annals of Operations Research* 85, 267–284.

Dueck, G. and T. Scheuer, 1990. Threshold accepting: A general purpose algorithm appearing superior to simulated annealing. *Journal of Computational Physics* 90, 161–175.

Dueck, G. and P. Winker, 1992. New Concepts and Algorithms for Portfolio Choice. *Applied Stochastic Models and Data Analysis* 8, 159–178.

Holland, J.H., 1975. *Adaptation in Natural and Artificial Systems.* University of Michigan Press.

Kirkpatrick, S., C. Gelatt and M. Vecchi, 1983. Optimization by simulated annealing. *Science* 220, 671–680.

Koskosides, Y. and A. Duarte, 1997. A scenario-based approach for active asset allocation. *Journal of Portfolio Management*, 74–85.

Leibowitz, M.L. and S. Kogelman, 1991. Asset Allocation under Shortfall Constraints. *Journal of Portfolio Management* Winter, 18-23.

Lucas, A. and P. Klaassen, 1998. Extreme Returns, Downside Risk, and Optimal Asset Allocation. *Journal of Portfolio Management* 25, 71–79.

Markowitz, H., 1952. Portfolio Selection. *Journal of Finance* 7, 77–91.

Palmquist, J., S. Uryasev and P. Krokhmal, 1999. Portfolio Optimization with Conditional Value-at-Risk Objective and Constraints. Research Report #99–14, Center for Applied Optimization, University of Florida.

Roy, A.D., 1952. Safety-First and the Holding of Assets. *Econometrica* 20, 431–449.

Winker, P., 2001. *Optimization Heuristics in Econometrics and Statistics: A simple approach for complex problems with Threshold Accepting*. Wiley, New York. (to appear).