



Ca' Foscari
University
of Venice

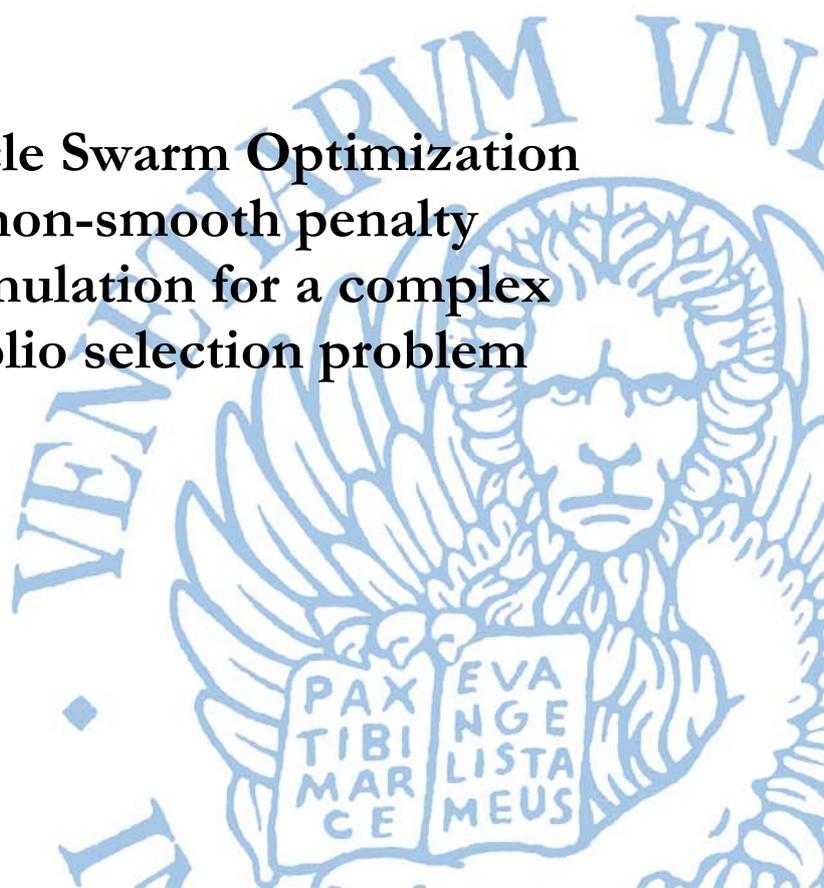
Department
of Economics

Working Paper

**Marco Corazza, Giovanni
Fasano and Riccardo Gusso**

**Particle Swarm Optimization
with non-smooth penalty
reformulation for a complex
portfolio selection problem**

ISSN: 1827/3580
No. 10/WP/2011





Particle Swarm Optimization with non-smooth penalty reformulation for a complex portfolio selection problem

Marco Corazza

*Department of Economics
Ca' Foscari University of Venice*

and

Advanced School of Economics in Venice

Giovanni Fasano

*Department of Management
Ca' Foscari University of Venice*

and

CNR - INSEAN (Italian Ship Model Basin)

Riccardo Gusso

*Department of Economics
Ca' Foscari University of Venice*

First Draft: July 2011

Abstract

In the classical model for portfolio selection the risk is measured by the variance of returns. It is well known that, if returns are not elliptically distributed, this may cause inaccurate investment decisions. To address this issue, several alternative measures of risk have been proposed. In this contribution we focus on a class of measures that uses information contained both in lower and in upper tail of the distribution of the returns. We consider a nonlinear mixed-integer portfolio selection model which takes into account several constraints used in fund management practice. The latter problem is NP-hard in general, and exact algorithms for its minimization, which are both effective and efficient, are still sought at present. Thus, to approximately solve this model we experience the heuristics Particle Swarm Optimization (PSO). Since PSO was originally conceived for unconstrained global optimization problems, we apply it to a novel reformulation of our mixed-integer model, where a standard exact penalty function is introduced.

Keywords

Portfolio selection, coherent risk measure, fund management constraints, NP-hard mathematical programming problem, PSO, exact penalty method, SP100 index's assets.

JEL Codes

C61, C63, G11.

Address for correspondence:

Marco Corazza

Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy

Phone: (+39) 041 2346921

Fax: (+39) 041 2349176

E-mail: corazza@unive.it

This Working Paper is published under the auspices of the Department of Economics of the Ca' Foscari University of Venice. Opinions expressed herein are those of the authors and not those of the Department. The Working Paper series is designed to divulge preliminary or incomplete work, circulated to favour discussion and comments. Citation of this paper should consider its provisional character.

1 Introduction

Making effective portfolio selection in real stock markets is a not-so-easy task for at least the following three reasons.

First, we need to gauge the risk by measures that, on one hand, satisfy appropriate formal properties (namely, the coherence ones) and that, on the other hand, better couple with the non normal returns distributions which characterize the stock markets. Moreover, it should be desirable that these risk measures were suitably parameterized with respect to the investors' risk attitude. In other terms, we need personalizable coherent risk measures that well fit a non Gaussian (financial) world.

Second, we have to take into account those practises and rules of the portfolio management industry that can affect the portfolio selection process (for instance, the use of bounds for the minimum and the maximum number of stocks to trade). Generally, such aspects are formalized in terms of constraints that very often yield the corresponding mathematical programming problem to be NP-hard.

Third, the portfolio selection problems arising from the joint use of the considered risk measures and practises and rules are possibly highly nonlinear, nondifferentiable, nonconvex and mixed-integer, which contributes to the NP-hardness. Therefore, the development of ad hoc solution approaches is usually needed to find the optimal solutions of such problems. But generally the portfolio management industry does not possess the mathematical knowledge and/or the research capabilities to develop such approaches; further, it could be not convenient for it to build up a team of external experts. As a consequence, it could happen that part of the investors' demand remains unsatisfied or (worse) is satisfied by the use of inappropriate solution technologies. So, we need also a "universal" global optimizer, that is a solution methodology able to cope with a large variety of real portfolio selection problems.

In this paper we deal with such issues, and we show solution proposals for managing each of them. We perform various checks of the overall resulting solution procedure by applying it to the selection of large portfolios. In particular, we tackle the above mentioned issues as follows.

First, as measure of risk of the portfolio returns we consider a recently proposed coherent risk measure based on the combination of upper and lower moments of different orders of the returns distribution (see [6]). This measure shows to be able to effectively manage non Gaussian distributions of asset returns and to appropriately reflect different investors' risk attitudes (see section 2 for details). Particularly, it permits to take into account, following a personal investors' weighting, both the risk contained in the "bad tail", the left one, of the portfolio returns, and the chances contained in the "good tail", the right one, of the same portfolio. Further, to the best of our knowledge, this is the first application of such a risk measure to the selection of large portfolios.

Second, as the professional practices and rules are concerned, following the indications of a north-eastern Italian company skilled in automatic financial trading systems, we address our analysis to the use of bounds for the minimum and the maximum number of

stock to trade¹, and the minimum and the maximum capital percentage to invest in each asset. These kinds of bounds are often considered by the fund manager industry as they constitute a tool for control, although in an indirect way, the transaction costs. All these practices/rules are formalized in terms of constraints that surely make NP-hard the corresponding mathematical programming problem (see again section 2 for details). Notice that the portfolio selection problem so arising is new in the specialized literature.

Third, the resulting selection problem, which takes into account also standard constraints (namely, the minimum return constraint and the budget one) is nonlinear, non-differentiable and mixed-integer. At present, for such a general scheme of mathematical programming problem (which is NP-hard, see [19]) both efficient and effective solution algorithms do not exist. Thus, in order to both investigate the numerical complexity of our portfolio selection problem and to provide a computationally cheap reliable solution to it, we adopt an exact penalty method (see [29, 20, 8, 14]) combined in an original manner with a recently proposed bio-inspired population-based metaheuristic, the Particle Swarm Optimization (PSO) (see [16] as, likely, first contribution on it). In short, our solution approach runs as follows (see section 3 for details):

- a standard exact penalty scheme transforms the considered mixed-integer portfolio selection problem into an equivalent nondifferentiable unconstrained minimization problem;
- then, as also the latter model is nonlinear, nondifferentiable and non convex, for its minimization a derivative-free algorithm is a possible solution method. Among the various approaches proposed in the specialized literature, we consider the PSO in order to approximately computing a global minimizer of the overall exact penalty-based model.

More generally, the choice of bio-inspired metaheuristics as global optimizers is also motivated by the fact that *they are more universal and less exacting with respect to an optimization problem*» (see [10], page 9).

Of course, PSO is not the only bio-inspired metaheuristics able to deal with minimization problems like ours. As possible alternatives we recall the Differential Evolution (DE) and the Genetic Algorithms (GAs); see the final remarks for some considerations about them as global minimizers for portfolio selection problems. From a methodological point of view it is to notice that our combined use of an exact penalty scheme with the PSO is not frequent in the literature. Indeed, for the solution of constrained problems the PSO is often modified with hybrid variants, which are suitably adapted to cope even with nonlinear constraints. In this respect we first provide a theoretical result ensuring the correspondence between the solutions of the original mathematical programming problem and the solutions of the exact penalty-based model (see again section 3 for details). Further, we also develop a simple original approach for the initialization of the particles' parameters. Its utilization offers numerical evidence of improvements in the convergence to a global minimum (see section 4 for details). Finally, notice that the solution approach we propose (exact penalty

¹These boundings are known as cardinality constraints.

scheme + PSO) is independent of the characteristics of the objective function and of the constraints. In other terms, from a methodological point of view our proposal can play the role of universal global (approximate) optimizer, for a very large variety of portfolio selection problems.

The remainder of this paper is organized as follows. In the next section first we illustrate the coherent risk measures we use, then we present our portfolio selection problem. In section 3 first we recall the basics of PSO, second we apply the exact penalty method for the reformulation of our portfolio selection problem, then we provide the theoretical result above mentioned. In section 4 we apply our overall solution procedure to the selection of large portfolios based on the set of assets constituting the Standard&Poor's SP100 index. We test various settings of the solution procedure, where we use a simple approach for the initialization of the particles' parameters. Then, we apply the so-set solution procedure to different time periods from August 2004 to October 2009, in order to detect possible differences in the optimal portfolio composition, and we critically present the obtained results. Some final remarks are given in the last section.

2 Portfolio selection and risk measures

The basic idea in the portfolio selection problem is to select stocks in order to maximize the portfolio performance and at the same time to minimize its risk. This implies that for a formal approach to the latter problem, a correct definition of performance and risk of the portfolio is required. While there is a general agreement about the measurement of performance by the expected value of the future return of the portfolio, the discussion regarding an adequate measure of risk is still open.

In its pioneering work [17] Markowitz proposed to use the variance of portfolio return to measure its risk, and this idea has been used for a long time in financial practice. However, it is well known that the mean-variance model leads to optimal investment decisions only if investment returns are elliptically distributed or alternatively if the utility function of investors is quadratic. The main shortcomings of quadratic utility functions have been pointed out since their introduction (see [27]), and it is a stylized fact that the distributions of returns of financial instruments present asymmetry and "fat tails". These considerations have opened the way for the research on alternative measures of risk, along with their properties: a recent characterization of them is presented in [23].

One crucial fact should be taken into account for a correct specification of the risk measure: while variance gives the same weight to positive and negative deviations from the mean, several empirical studies have shown (see e.g. [3]) that investors treat them in different ways. This has led to the definition of risk measures that are focused on the "bad tail" of the distribution of the returns, as for example the semivariance (see [2]), the lower partial moments (see [12]) and the minimax ones (see [28]). On the other hand, some other risk measures were based on a quantile of the "bad tail", as the well known Value-at-Risk (VaR) (see [18]).

Since the introduction of the notion of coherent risk measure in [1], along with the specification of the properties for a measure in this class (monotonicity, positive homogeneity,

translation invariance and sub-additivity), there has been a growing interest for the previously introduced measures. In particular, properties similar to those for coherent risk measures have been studied also for the other risk measures. The Conditional Value-at-Risk (CVaR) (see [24]) is possibly the most famous measure obtained by the research in this direction; other examples based on lower partial moments are reported in [11].

More recently Chen and Wang in [6] have investigated the possibility of building a new class of coherent risk measures, by combining upper and lower moments of different orders. This approach seems to have several advantages with respect to others considered so far. Indeed, on one hand these measures better couple with non normal distributions than ones based only on first order moments. On the other hand, they better reflect investors' risk attitude, for at least a couple of reasons. First they are less affected by estimation risk than measures that use only information from the lower part of the return distribution. Moreover, according with the conclusions presented in [6], their use in the portfolio selection problem allows for more realistic and robust results, compared with the ones obtained using CVaR. In this contribution we use the class of risk measures in [6] for our portfolio selection problem. Our problem also takes into account several constraints, often used in fund management practice. In particular, we focus on handling the cardinality constraints, which yield a final model in the class of nonlinear mixed-integer programming problems.

In the next sections we analyze the results obtained both in terms of our algorithm efficiency and in terms of financial meanings.

2.1 Our portfolio selection model

Let X be a real valued random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let us denote $\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}$, with $p \in [1, +\infty[$, where $\mathbb{E}[\cdot]$ indicates the expected value of a random variable. Then, the measures of risk introduced in [6] are defined as:

$$\rho_{a,p}(X) = a\|(X - \mathbb{E}[X])^+\|_1 + (1 - a)\|(X - \mathbb{E}[X])^-\|_p - \mathbb{E}[X], \quad (1)$$

where $a \in [0, 1]$, $X^- = \max\{-X, 0\}$ and $X^+ = (-X)^-$. For a and p fixed, any risk measure of this class is then a convex combination of the two coherent risk measures based on lower partial moments $\|(X - \mathbb{E}[X])^-\|_1 - \mathbb{E}[X]$ and $\|(X - \mathbb{E}[X])^-\|_p - \mathbb{E}[X]$. Thus, it is a coherent risk measure (see [11]). For a detailed description of its properties we refer the reader to [6]. We only remark here that $\rho_{a,p}$ is non-decreasing with respect to p and non-increasing with respect to a . Thus, the value of these parameters can be adjusted to reflect different attitudes of the investors towards risk.

Now we describe the portfolio selection model we consider. Suppose we have N assets to choose from, and for $i = 1, \dots, N$ let $x_i \in \mathbb{R}$ be the weight of i -th asset in the portfolio, with $X^T = (x_1, \dots, x_N)$. Let $Z^T = (z_1, \dots, z_N) \in \{0, 1\}^N$ be a binary vector, such that $z_i = 1$ if the asset i is included in the portfolio, $z_i = 0$ otherwise. Moreover, for $i = 1, \dots, N$, let r_i be a real valued random variable that represents the return of asset i , with \hat{r}_i its expected value, i.e. $\hat{r}_i = \mathbb{E}[r_i]$. Then, the random variable $R \in \mathbb{R}$ that represents the return of the

whole portfolio can be expressed as

$$R = \sum_{i=1}^N x_i r_i,$$

with expected value

$$\hat{R} = \sum_{i=1}^N x_i \hat{r}_i.$$

Then, considering (1), our goal is to minimize $\rho_{\alpha,p}(R)$, subject to several constraints. Of course the first ones to consider are the constraints regarding the minimum desirable expected return of the portfolio, i.e.

$$\hat{R} \geq l, \text{ with } l > 0,$$

and the usual budget constraint

$$\sum_{i=1}^N x_i = 1.$$

Moreover, as stated in the previous section, we also introduce the following cardinality constraint: we select a (not too) small subset of the available assets. The latter choice summarizes a quite common problem for a fund manager, who has to build a portfolio by choosing from several hundreds of assets. When the number of selected assets is too large, several practical accounting problems may arise, which can increase transaction costs. By using the latter cardinality constraint we implicitly consider transaction costs in our model. The resulting constraint is explicitly given by

$$K_d \leq \sum_{i=1}^N z_i \leq K_u, \text{ where } 1 \leq K_d \leq K_u \leq N.$$

Further, we require that any of the selected assets must not constitute a too large or too small fraction of the portfolio, i.e.

$$z_i d \leq x_i \leq z_i u, \text{ where } 0 \leq d \leq u \leq 1,$$

and d, u represent respectively the minimum and maximum fraction allowed. Of course, to ensure compatibility with the cardinality constraint, the constants d and u must satisfy

$$d \leq \frac{1}{K_d} \text{ and } u \geq \frac{1}{K_u}. \quad (2)$$

Then, our overall portfolio selection problem can be written as follows:

$$\begin{aligned}
& \min_{X,Z} \rho_{a,p}(R) \\
& \text{s.t.} \quad \hat{R} \geq l \\
& \quad \sum_{i=1}^N x_i = 1 \\
& \quad K_d \leq \sum_{i=1}^N z_i \leq K_u \\
& \quad z_i d \leq x_i \leq z_i u, \quad \text{with } i = 1, \dots, N, \\
& \quad z_i(z_i - 1) = 0, \quad \text{with } i = 1, \dots, N,
\end{aligned} \tag{3}$$

where the last N constraints are introduced to model the relations $z_i \in \{0, 1\}$, with $i = 1, \dots, N$. It is clear that if in the last N constraints we have $z_i = 0$, then the variable x_i does not play any role in the solution of problem (3), i.e. $x_i = 0$. Conversely, $z_i = 1$ implies that potentially the i -th asset will contribute to the final portfolio, with $x_i \in [d, u]$. Finally, the constraints $z_i(z_i - 1) = 0$, with $i = 1, \dots, N$, represent just one (and possibly not the best) reformulation of the integrality constraints $z_i \in \{0, 1\}$, with $i = 1, \dots, N$. We do not investigate further the latter issue, since it is not a focus of this paper.

Of course, (3) is a nonlinear and nonconvex mixed-integer problem, which in general admits several local solutions, but we want to possibly seek global solutions and not simply local minimizers. However, detecting precise solutions of (3) may be heavily time consuming in case exact methods are adopted. Thus, at present we experience the heuristic technique PSO on a non-smooth reformulation of problem (3). The next section is devoted to detail the PSO heuristics.

3 PSO for non-smooth reformulation of the portfolio selection problem

Particle Swarm Optimization is an iterative heuristics for the solution of nonlinear global optimization problems (see [16]). It is based on a biological paradigm, which is inspired by the flight of birds in a flock. In particular, the basic idea of PSO (see also [22] for a tutorial) is to replicate the behaviour of shoals of fishes or flocks of birds, when they cooperate in the search for food. On this purpose every member of the swarm explores the search area keeping memory of its best position reached so far, and it exchanges this information with the neighbors in the swarm. Thus, the whole swarm is supposed to converge eventually to the best global position reached by the swarm members.

In its mathematical counterpart the paradigm of a flying flock may be formulated as follows: given a minimization problem, find a global minimum (best global position) in a nonlinear minimization problem. Every member of the swarm (namely a particle) represents a possible solution of the minimization problem, and it is initially positioned randomly in

the feasible set of the problem. Every particle is also initially assigned a random velocity, which is used to determine its initial direction of movement.

For a more formal description of PSO let us consider the global optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}),$$

where $f : \mathbb{R}^d \mapsto \mathbb{R}$ is the objective function in the minimization problem. Suppose we apply PSO for its solution, where M particles are considered. At the k -th step of the PSO algorithm three vectors are associated to each particle $j \in \{1, \dots, M\}$:

- $\mathbf{x}_j^k \in \mathbb{R}^d$, the position at step k of particle j ;
- $\mathbf{v}_j^k \in \mathbb{R}^d$, the velocity at step k of particle j ;
- $\mathbf{p}_j \in \mathbb{R}^d$, the best position visited so far by the j -th particle.

Moreover, $pbest_j = f(\mathbf{p}_j)$ denotes the value of the objective function in the position \mathbf{p}_j of the j -th particle. The overall PSO algorithm, as in the version with inertia weight proposed in [25], works as follows:

1. Set $k = 1$ and evaluate $f(\mathbf{x}_j^k)$ for $j = 1, \dots, M$. Set $pbest_j = +\infty$ for $j = 1, \dots, M$.
2. If $f(\mathbf{x}_j^k) < pbest_j$ then set $\mathbf{p}_j = \mathbf{x}_j^k$ and $pbest_j = f(\mathbf{x}_j^k)$.
3. Update position and velocity of the j -th particle, with $j = 1, \dots, M$, as

$$\mathbf{v}_j^{k+1} = w^{k+1} \mathbf{v}_j^k + \mathbf{U}_{\phi_1} \otimes (\mathbf{p}_j - \mathbf{x}_j^k) + \mathbf{U}_{\phi_2} \otimes (\mathbf{p}_{g(j)} - \mathbf{x}_j^k) \quad (4)$$

$$\mathbf{x}_j^{k+1} = \mathbf{x}_j^k + \mathbf{v}_j^{k+1} \quad (5)$$

where $\mathbf{U}_{\phi_1}, \mathbf{U}_{\phi_2} \in \mathbb{R}^d$ and their components are uniformly randomly distributed in $[0, \phi_1]$ and $[0, \phi_2]$ respectively, the symbol \otimes denotes component-wise product and $\mathbf{p}_{g(j)}$ is the best position in a neighborhood of the j -th particle.

4. If a convergence test is not satisfied then set $k = k + 1$ and go to 2.

The values of ϕ_1 and ϕ_2 strongly affect the strength of the attractive forces towards the personal and the neighborhood best positions explored so far. Thus, in order to (possibly) yield the convergence of the swarm, they have to be set carefully in accordance with the value of the inertia weight w^k . The parameter w^k is generally linearly decreasing with the number of steps, i.e.

$$w^k = w_{max} + \frac{w_{min} - w_{max}}{K} k,$$

where common values for w_{max} and w_{min} are respectively 0.9 and 0.4, while K is usually the maximum number of steps allowed.

Another widely adopted version of the PSO algorithm is the one with constriction coefficients (see [7]), where the updating velocity rule (4) is replaced by

$$\mathbf{v}_j^{k+1} = \chi \left[\mathbf{v}_j^k + \mathbf{U}_{\phi_1} \otimes (\mathbf{p}_j - \mathbf{x}_j^k) + \mathbf{U}_{\phi_2} \otimes (\mathbf{p}_{g(j)} - \mathbf{x}_j^k) \right], \quad (6)$$

with $\chi = \frac{2}{\phi - 2 + \sqrt{\phi^2 - 4\phi}}$, $\phi = \phi_1 + \phi_2$, and $\phi > 4$.

As stated before, we can think that at step k the j -th particle moves as subject to two attractive vectors: the direction towards its previous best position (namely $(\mathbf{p}_j - \mathbf{x}_j^k)$) and the direction towards the best position in a suitable subset of the swarm (namely $(\mathbf{p}_{g(j)} - \mathbf{x}_j^k)$). We recall that $g(j)$ denotes the index of the particle with the best position reached so far, in a neighborhood of the j -th particle. The specification of the neighborhood topology is then a choice to set. In our implementation we have considered the so called *gbest* topology, that is $g(j) = g$ for every $j = 1, \dots, M$, and g is the index of the best particle in the whole swarm, that is $g = \arg \min_{j=1, \dots, M} \mathbf{p}_j$. This choice implies that the whole swarm is used as the neighborhood of each particle.

We remark that the original formulation of PSO was conceived for unconstrained problems. Thus, in general using PSO formulae (4)-(5), when constraints are included in the formulation, is improper. Indeed, in the latter case the algorithm above cannot prevent from generating infeasible particles' positions, unless specific adjustments are adopted. When constraints are included, different strategies were proposed in the literature (see also [22]) to ensure that at any step of PSO, feasible positions are generated. Most of them involve repositioning of the particles, as for example the bumping and the random positioning strategies proposed in [30], or introducing some external criteria to rearrange the components of the particles, as the ones specific for cardinality constraints proposed in [9] and [26]. However in this paper we decided to use PSO as in its original formulation, that is as a tool for the solution of unconstrained optimization problems. The latter choice is mainly motivated by the necessity of avoiding both a possible misleading application of specific metaheuristics to handle nonlinear constraints, and the careless setting of ad hoc coefficients.

To this purpose, first we have reformulated our problem into an unconstrained one, using the nondifferentiable ℓ_1 penalty function method described in [29, 14]. The latter approach is known in the literature of constrained optimization as exact penalty method, where the term exact refers to the correspondence between the minimizers of the original constrained problem and the minimizers of the unconstrained (penalized) one. An example in the literature where PSO is applied to minimize a penalty function is given in [5]. However, in the latter paper no exact penalty functions are used and no integer unknowns are introduced. In addition, the updating rule for the penalty parameter seems to be far too "problem dependent".

On the contrary, in our approach we reformulate as follows the problem (3) (which has $N + 1$ equality constraints and $2N + 3$ inequality constraints), using the nondifferentiable penalty function:

$$\min_{X, Z} P(X, Z; \varepsilon) \quad (7)$$

where

$$\begin{aligned}
P(X, Z; \varepsilon) = & \rho_{a,p}(R) + \frac{1}{\varepsilon} \left[\max\{0, l - \hat{R}\} + \left| \sum_{i=1}^N x_i - 1 \right| \right. \\
& + \max \left\{ 0, K_d - \sum_{i=1}^N z_i \right\} + \max \left\{ 0, \sum_{i=1}^N z_i - K_u \right\} \\
& + \sum_{i=1}^N \max\{0, z_i d - x_i\} + \sum_{i=1}^N \max\{0, x_i - z_i u\} \\
& \left. + \sum_{i=1}^N |z_i(1 - z_i)| \right]
\end{aligned} \tag{8}$$

and ε is the penalty parameter.

The correct choice of ε ensures the correspondence between the solutions of problems (7) and (3) (see also [21]), which is summarized by the result which follows.

Proposition 3.1 *Consider the problem (3) with $\rho_{a,p}(R)$ continuous on \mathbb{R}^N . Consider the Exact Penalty function $P(X, Z; \varepsilon)$ in (8). Let (X^*, Z^*) be a strict local minimizer of problem (3) where the KKT conditions (see the appendix for details) are satisfied, with the generalized Lagrange multipliers λ_i^* , with $i = 1, \dots, N + 1$ (for the equality constraints) and σ_j^* , with $j = 1, \dots, 2N + 3$ (for the inequality constraints). Then, for any $\varepsilon > \varepsilon^*$ the solution (X^*, Z^*) is also a local minimizer of $P(X, Z; \varepsilon)$, where*

$$\varepsilon^* = \left\| \begin{array}{c} \lambda^* \\ \sigma^* \end{array} \right\|_{\infty}.$$

□

Observe that the penalty function $P(X, Z; \varepsilon)$ is clearly nondifferentiable because of the ℓ_1 -norm in (8). This also motivates the choice of using PSO for its minimization, since PSO evidently does not require the derivatives of $P(X, Z; \varepsilon)$. We avoid to go into details (see [21, 29]), however the latter choice turns to be of great interest on those problems where illconditioning may arise. We also remark that the threshold value ε^* is unknown. Nevertheless, acceptable values of this threshold can often be found by appropriate numerical investigation, provided that a constraints qualification condition is satisfied in (X^*, Z^*) .

Of course, since PSO is a heuristics, the minimization of the penalty function $P(X, Z; \varepsilon)$ theoretically does not ensure that a global minimum of the problem (3) is detected. Nevertheless, PSO often provides a suitable compromise between the performance (i.e. a satisfactory estimate of a global minimizer for the problem (3)) and the computational cost.

4 Numerical results

In order to test our approach, in this section we consider data of daily close prices $\{p_{i,t}\}$ of asset i at time t , for $i = 1, \dots, 100$ and $t = 1, \dots, T$, where T is the time horizon

considered. The assets considered are included in the Standard&Poor's SP100 index² from August 2004 to October 2009. Several subsets of these data have then been selected, to analyze differences in the optimal portfolio composition with respect to the time period. The price series have been used to compute each stock return

$$r_{i,t} = \frac{p_{i,t+1} - p_{i,t}}{p_{i,t}}.$$

Using the same idea of [6] we estimate the risk measure for any portfolio $X = (x_1, \dots, x_N)^T$ as

$$\begin{aligned} \rho_{a,p}(R) = & \frac{a}{T} \left[\sum_{t=1}^T \left(\sum_{i=1}^N (r_{i,t} - \hat{r}_i) x_i \right)^+ \right] \\ & + (1-a) \left\{ \frac{1}{T} \sum_{t=1}^T \left[\left(\sum_{i=1}^N (r_{i,t} - \hat{r}_i) x_i \right)^- \right]^p \right\}^{\frac{1}{p}}, \end{aligned} \quad (9)$$

where \hat{r}_i is estimated using the historical data, that is

$$\hat{r}_i = \frac{1}{T} \sum_{t=1}^T r_{i,t}.$$

According with problem (3), the minimum level of desired return l in (8) has been set to the global average of stock returns, i.e. $l = \sum_{i=1}^N \hat{r}_i$. Moreover, to reflect a realistic problem of portfolio selection, we set the values $d = 0.02$ and $u = 0.20$ in (8). In order to analyze the impact of the cardinality constraints on the selection problem, from both the computational and the economic point of view, we have chosen $K_d = 5$, while we have considered two different values for K_u : $K_u = 50$ (the maximum value allowed according to (2)), and $K_u = 30$. The PSO algorithm to solve (8) has been implemented in MATLAB 7, and the experiments have been performed on a workstation Acer Aspire M1610 with a Intel Core 2 Duo E4500 processor.

We first performed a set of preliminary tests (reported below) with the aim of assessing a proper version of PSO to use. We experienced different values for the penalty parameter ε and for the coefficients of PSO, including the number M of particles in the swarm. Then, we set the values of the parameters³ a , p of the risk measure in (9), as $a = 0.5$ and $p = 2$, and we ran the two versions of the algorithm (i.e. with decreasing inertia weight and constriction coefficients respectively), using one-year data of returns. Since the evaluation of the objective function $P(X, Z; \varepsilon)$ in (8) is relatively inexpensive, we stopped PSO iterations when either of the following stopping criteria was satisfied:

²Due to the lack of available data for the whole period considered, Mastercard and Philip Morris have been replaced by Verisign and Molson Coors Brewing Company.

³Experiments with different values for a and p showed no difference with regard to the results of this preliminary phase.

- a) the maximum number of 10000 steps was outreached;
- b) $|fbest^{k+1} - fbest^k| < 10^{-8}$ for 2000 consecutive steps, where $fbest^k$ is the best value of the fitness function $f = P(X, Z; \varepsilon)$ at k -th iteration.

We noticed that, in the majority of the tests, the version of PSO with constriction coefficients (6) showed earlier convergence of the swarm to a global best position with worse values of the fitness function, compared to versions of PSO using a decreasing inertia weight. Thus, we decided to adopt the latter variant of PSO for the subsequent experiments.

In Table 1 we report the results in terms of the averaged best value of the fitness function and its standard deviation, normalized to take into account the effects of using different values of ε . In Table 2 we show the results in terms of the averaged ratio between the final (F_f) and initial (F_i) value of the fitness function, and of the average computational time, for different numbers of particles used. The results in Tables 1 and 2 are averaged over 10 runs, and the results in Table 2 refer to the case $\varepsilon = 10^{-6}$, which is the best value from Table 1.

ε	Normalized fitness	Standard deviation
1	0.388728255	0.134295545
0.1	0.332573168	0.136412572
0.01	0.337188413	0.061398093
0.001	0.372167277	0.253545145
0.0001	0.413884094	0.155254628
0.00001	0.381672803	0.200184247
0.000001	0.260743870	0.099544544
0.0000001	0.341627127	0.185689216

Table 1: Results for different choices of the parameter ε in (8).

M	Fitness	Ratio of decrease (F_f/F_i)	Time (seconds)
50	444393.5727	0.015960233	54.2939
100	60160.8053	0.002162255	109.2537
200	6423.0849	0.000233782	158.0136

Table 2: Results for different choices of the number M of particles.

This suggested us to select $\varepsilon = 10^{-6}$ and $M = 200$. The number of particles is then quite high, but this concurs with the general evidence that larger populations perform better for higher dimensional problems (see also [4] and [22]).

Then, we repeated the computation in order to select the best values for the acceleration coefficients ϕ_1, ϕ_2 , while for the initial and final values of the inertia weight w_{max}, w_{min} in 4 we used 0.9 and 0.4 as suggested by the current literature. The results are shown in Table 3, and the best performance was obtained with $\phi_1 = \phi_2 = 1.85$.

$\phi_1 = \phi_2 = \phi/2$	Average fitness	Standard deviation
1.25	564611.56	261486.21
1.50	13439.35	11620.83
1.75	6047.39	5361.74
1.85	1822.09	1963.27
2.00	22763505.40	859690.31

Table 3: Results with different choices of the parameter ϕ .

After this preliminary phase, we solved the portfolio selection problems for different values of the parameters a and p of the risk measure $\rho_{a,p}$, and K_u , considering one year data of daily returns of different time periods. We wanted in fact to study both the capability of PSO to find a global minimum for the optimization problem and the economic meaning of the portfolios obtained, while considering different scenarios and different attitudes towards risk.

We set then the maximum number of the algorithm steps to 20000 and, for every combinations of the parameters and the dataset, we did first 25 runs of the algorithm, each with different random initial positions and velocities. Since the standard deviation of the values of the fitness function was still high (an example of the values of the fitness function at the end of the first 25 runs is shown in Table 4) and we found a different optimal portfolio for every run, each corresponding to a possible local minimum, we decided to iterate the procedure in the following way: we did other 25 runs of the algorithm, with again random initial velocities for all particles, but we used the 25 global best positions found in the previous phase as initial positions for 25 particles, while the remaining 175 ones were set again randomly. At the end of this second phase we obtained convergence to the same global best position for each of the 25 runs (in general not corresponding to the best position of the previous 25 ones, see Table 5) and we assumed the latter to be the global minimum X^* of the optimization problem. We remark that the difference between the global best fitness $P(X^*, Z^*; \varepsilon)$ and the risk measure for $\rho_{a,p}(X^*)$ is negligible⁴, except for the case $\rho_{0.5,5}$ with $K_u = 30$ and data of period 2006-07 (see Table 5). This means that the use of the nondifferentiable penalty function is effective, in order to impose the satisfaction of the constraints, including the cardinality constraints.

As it could be guessed from Table 5, it is interesting to remark that the monotonicity properties expected by theoretical results ([6, Theorem 2.3]) are fulfilled. Tables 6 and 7 also confirm the latter statement, and highlight that PSO is effective to detect a good approximation of a global minimum for $P(X, Z; \varepsilon)$. In this regard PSO provides an efficient compromise between the correctness of the solution found and the resources used in the computation.

Interestingly enough, we also notice that $\rho_{a,p}(X^*)$ is slightly decreasing with K_u (except for the case $\rho_{0.75,2}(X^*)$ with data of period 2004-05). This is consistent with the fact that

⁴At least in the cases where l is set equal to the global average of all stock returns, while there may be some problems when l is increased. We guess that in the latter case the maximum number of stocks to trade, K_u , is not enough large to permit the satisfaction of the minimum return constraint.

PSO has performed a good exploration of the feasible set when $K_u = 30$, that is a subset of the one with $K_u = 50$.

Run	1	2	3	4	5
Best fitness	90.35	249.29	16.25	96.25	311.97
Run	6	7	8	9	10
Best fitness	247.11	4.96	57.84	50.78	384.16
Run	11	12	13	14	15
Best fitness	46.22	4.06	0.62	1.25	0.47
Run	16	17	18	19	20
Best fitness	2.17	0.22	112.70	118.46	0.29
Run	21	22	23	24	25
Best fitness	3.67	3.75	2.09	186.67	0.54

Table 4: Best fitness after the first 25 runs, with $a = 0.5, p = 1, K_u = 50$, data of period 2006-07.

In order to analyze the financial meaning of the portfolios obtained, we solved using PSO another portfolio selection problem, replacing $\rho_{a,p}(X)$ with variance as measure of risk, and keeping the same set of constraints of problem 3. We first checked the diversification of the portfolios, by comparing the number of assets among them, as shown in Table 8. It appears that when the cardinality constraint is in more relaxed form, that is $K_u = 50$, the diversification obtained using $\rho_{a,p}$ is higher than using variance, and it is also slightly increasing with p . Again, this is also consistent with the results obtained in [6], where the cardinality constraint was not explicitly introduced, and the comparison was made with respect to CVaR.

$K_u = 50$	$p = 1$	$p = 2$	$p = 5$
Best fitness among the first 25 runs	0.2208	0.6356	0.0662
Global best fitness	0.0020	0.0029	0.0044
$K_u = 30$	$p = 1$	$p = 2$	$p = 5$
Best fitness among the first 25 runs	1.4742	0.4846	0.2062
Global best fitness	0.0021	0.0031	0.0134

Table 5: Comparison between the best fitness in the first 25 runs, and the global best fitness after other 25 runs, with $a = 0.5$, data of period 2006-07.

Finally, we compared the performance of the portfolios selected using $\rho_{a,p}$ with those portfolios selected using the variance. Following indications from financial practice, we proceeded in the following way: we used one-year data of daily returns for portfolio selection, by minimizing $P(X, Z; \varepsilon)$ in (8); then, we invested the selected portfolios for the three next months. After that we repeated the selection, and we re-invested the resulting portfolios for other three months, and so on for other two quarters. In this way we analyzed the performance of the portfolios along one entire year. We considered two initial starting

periods of one year length (in order to compute the portfolios for the test in the subsequent first quarter), that is August 2004-July 2005 and February 2007-January 2008, with the aim of analyzing the impact of different macroeconomic conditions on the performance of the portfolios. The results are shown in Tables 9-12.

2004-05	$p = 1$	$p = 2$	$p = 5$
$\rho_{0.5,p}; K_u = 50$	0.001951	0.002816	0.004513
Number of assets	43	45	40
$\rho_{0.5,p}; K_u = 30$	0.002333	0.003219	0.004808
Number of assets	23	28	28
2006-07	$p = 1$	$p = 2$	$p = 5$
$\rho_{0.5,p}; K_u = 50$	0.002012	0.002924	0.004379
Number of assets	44	45	45
$\rho_{0.5,p}; K_u = 30$	0.002094	0.003099	0.004339
Number of assets	30	30	29

Table 6: Monotonicity of $\rho_{a,p}(X^*)$ for $a = 0.5$ and different values of p and K_u , with one year data from two time periods.

2004-05	$a = 0$	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$
$\rho_{a,2}; K_u = 50$	0.003853	0.003448	0.002816	0.002509	0.002135
N. of assets	43	45	45	46	43
$\rho_{a,2}; K_u = 30$	0.004029	0.003473	0.003219	0.002434	0.002306
N. of assets	29	29	28	27	29
2006-07	$a = 0$	$a = 0.25$	$a = 0.5$	$a = 0.75$	$a = 1$
$\rho_{a,2}; K_u = 50$	0.004194	0.003587	0.002924	0.002319	0.001820
N. of assets	45	41	45	48	48
$\rho_{a,2}; K_u = 30$	0.004227	0.003652	0.003099	0.002345	0.001855
N. of assets	30	30	30	29	29

Table 7: Monotonicity of $\rho_{a,p}(X^*)$ for $p = 2$ and different values of a and K_u , with one year data from two time periods.

2004-05	$p = 1$	$p = 2$	$p = 5$	σ^2
$K_u = 50$	43	45	45	42
$K_u = 30$	29	28	28	28
2006-07	$p = 1$	$p = 2$	$p = 5$	σ^2
$K_u = 50$	44	45	45	42
$K_u = 30$	30	30	29	28

Table 8: Comparison of number of assets in portfolios using $\rho_{0.5,p}$ and σ^2 , with one year data from two time periods.

Period length	$\rho_{0.5,1}$	$\rho_{0.5,2}$	$\rho_{0.5,5}$	σ^2
3 months	1,67%	1,33%	0,67%	-0,18%
6 months	3,22%	4,00%	4,55%	3,59%
9 months	0,47%	-1,41%	1,14%	3,35%
12 months	-3,35%	-2,79%	-1,24%	-2,59%

Table 9: Portfolio returns with different measures of risk; case $K_u = 50$ for the time period 2005-06.

Period length	$\rho_{0.5,1}$	$\rho_{0.5,2}$	$\rho_{0.5,5}$	σ^2
3 months	-1,54%	1,68%	0,54%	3,17%
6 months	4,68%	5,36%	3,05%	5,36%
9 months	2,13%	0,96%	1,14%	0,96%
12 months	-3,23%	-1,89%	-0,93%	-1,89%

Table 10: Portfolio returns with different measures of risk; case $K_u = 30$ for the time period 2005-06.

Period length	$\rho_{0.5,1}$	$\rho_{0.5,2}$	$\rho_{0.5,5}$	σ^2
3 months	7,23%	3,20%	3,19%	5,13%
6 months	-8,32%	-6,64%	-6,79%	-8,24%
9 months	-23,82%	-25,13%	-22,81%	-22,30%
12 months	-10,44%	-9,77%	-9,38%	-7,11%

Table 11: Portfolio returns with different measures of risk; case $K_u = 50$ for the time period 2008-09.

We notice that the behaviour of the portfolio selected using $\rho_{0.5,2}$ is quite similar to the one of the portfolio selected using the variance, and this is more evident when $K_u = 30$ in the period 2005-06, where in two cases we obtain the same return and the differences in the composition of the two portfolios are negligible. Instead, the portfolios selected using $\rho_{0.5,1}$ and $\rho_{0.5,5}$ appear to correspond respectively to a more aggressive and conservative investor: in particular during the financial markets crisis in the period 2008-09, the portfolio selected using $\rho_{0.5,5}$ shows less losses than the ones obtained using variance. Also these results are in agreement with the theoretical meaning of the parameter p described in [6].

Period length	$\rho_{0.5,1}$	$\rho_{0.5,2}$	$\rho_{0.5,5}$	σ^2
3 months	3,42%	3,63%	3,94%	3,53%
6 months	-8,10%	-9,18%	-6,89%	-10,24%
9 months	-30,50%	-24,42%	-19,84%	-19,52%
12 months	-13,15%	-11,85%	-10,51%	-12,10%

Table 12: Portfolio returns with different measures of risk; case $K_u = 30$ for the time period 2008-09.

5 Final remarks

In this paper we have first proposed a partially novel formulation, given by a combination between a nondifferentiable exact penalty method and the PSO, for the selection of large portfolios characterized by an upper-and-lower-moments-based coherent risk measure and a mixed-integer formulation. Although the obtained results are satisfactory, this solution approach seems to offer opportunities for possible improvements and extensions. In particular:

- As the reformulation of problem (3) is concerned, we point out that other possible reformulations of that mathematical programming problem may be considered, both smooth and non-smooth. To this purpose, resorting to continuously differentiable penalty functions appears particularly promising. This method is substantially characterized by theoretical properties equivalent to the ones characterizing the penalty scheme used in this paper (see [13] for details);
- As the initialization of the particles' positions and velocities is concerned, we guess that the performances of our simple approach can be significantly improved by resorting to a theoretical-based procedure recently proposed in [4]. By so doing we expect for improvements both in solution quality and in computational time;
- As stated in section 1, PSO is not the only bio-inspired metaheuristic able to deal with minimization problems like (8). Currently, in order to compare different bio-inspired metaheuristics as global minimizers of complex portfolio selection problems, we have started to use GAs. The very first preliminary results (not included here) suggest that the optimal portfolio compositions obtained by PSO and GAs are quite similar, but that PSO needs a computational time which is significantly lower than the one needed by GAs;
- Finally we recall that, from a methodological point of view, the solution approach we propose can play the role of universal global (approximate) optimizer for a large variety of complex portfolio selection problems. So, it can constitute a flexible tool for the fund management industry in order not to leave unsatisfied demand.

It remains obvious that, in order to carefully detect values and drawback of such a solution approach, further investigations are necessary with respect to different risk measures, constraints and data. It will be done in future researches.

6 Appendix

Consider the general constrained optimization problem

$$\begin{aligned} \min \quad & f(x) \\ & h_i(x) = 0 \quad i = 1, \dots, p, \\ & g_j(x) \leq 0 \quad j = 1, \dots, q. \end{aligned} \tag{10}$$

Suppose that at the feasible point x^* some inequality constraints (thereof subscripts are in the subset $\mathcal{A}(x^*)$) are satisfied as equalities, i.e.

$$g_j(x^*) = 0, \quad j \in \mathcal{A}(x^*).$$

We say that for the problem (10) the condition *LICQ* (Linear Independent Constraint Qualification) holds at x^* if the vectors

$$\left\{ \nabla h_1(x^*), \dots, \nabla h_p(x^*), \nabla g_j(x^*)|_{j \in \mathcal{A}(x^*)} \right\}$$

are linearly independent. Then we can now define the following first order optimality conditions for the minimizer x^* of (10).

Proposition 6.1 (KKT Conditions). *Consider the problem (10), where the functions f , h and g are continuously differentiable. Suppose that x^* is a local minimizer of (10), where the LICQ holds. Then, there exists a unique Lagrange multiplier vector $(\lambda^{*T}, \sigma^{*T}) \in \mathbb{R}^{p+q}$ such that*

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^p \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^q \sigma_j^* \nabla g_j(x^*) &= 0 \\ h_i(x^*) &= 0, & i &= 1, \dots, p \\ \sigma_j^* g_j(x^*) &= 0, & j &= 1, \dots, q \\ g_j(x^*) &\leq 0, & j &= 1, \dots, q \\ \sigma_j^* &\geq 0, & j &= 1, \dots, q. \end{aligned}$$

□

Observe that the constraints qualification condition *LICQ* in Proposition 6.1 substantially ensures that there exist the functions $\lambda = \lambda(x)$ and $\sigma = \sigma(x)$, with $\lambda^* = \lambda(x^*)$ and $\sigma^* = \sigma(x^*)$, which can be explicitated by the Implicit Function Theorem, at least in a neighborhood of x^* . Equivalently, the condition *LICQ* can be replaced by several other qualification conditions (see also [14, 21, 29]).

Acknowledgements. M. Corazza and R. Gusso wish to thank the *European Social Fund* and the *Regione del Veneto* for the support received (research project *Metodologie informatiche ed algoritmi bio-ispirati per l'ottimizzazione e la gestione in azienda* [Computer science methodologies and bio-inspired algorithms for optimization and management in company], code 2120/1/11/1268/2008). G. Fasano wishes to thank the *CNR – INSEAN (Italian Ship Model Basin)* for the support received.

References

- [1] P. Artzner, F. Delbaen, J.M. Eber, and D. Heath, *Coherent measures of risk*, *Mathematical Finance* **9** (1999), no. 3, 203–228.

- [2] E. Ballester, *Mean-semivariance efficient frontier: A downside risk model for portfolio selection*, Applied Mathematical Finance **12** (2005), no. 1, 1–15.
- [3] B.F.J. Borges and L. Knetsch, *Tests of market outcomes with asymmetric valuations of gains and losses: Smaller gains, fewer trades, and less value*, Journal of Economic Behavior & Organization **33** (1998), no. 2, 185–193.
- [4] E.F. Campana, G. Fasano, and A. Pinto, *Dynamical analysis for the selection of parameters and initial population, in particle swarm population*, Journal of Global Optimization **48** (2009), no. 3, 347–397.
- [5] W. Chen, R. Zhang, Y. Cai, and F. Xu, *Particle swarm optimization for constrained portfolio selection problems*, in: Proceedings of the Fifth International Conference on Machine learning and Cybernetics (2006), pp. 2425–2429.
- [6] Z. Chen and Y. Wang, *Two-sided coherent risk measures and their application in realistic portfolio optimization*, Journal of Banking and Finance **32** (2008), no. 12, 2667–2673.
- [7] M. Clerc and J. Kennedy, *The particle swarm - explosion, stability, and convergence in a multidimensional complex space*, IEEE Transactions on Evolutionary Computation **6** (2002), no. 1, 58–73.
- [8] A.R. Conn, *Constrained optimization using a nondifferentiable penalty function*, SIAM Journal on Numerical Analysis **10** (1973), no. 4, 760–784.
- [9] T. Cura, *Particle swarm optimization approach to portfolio optimization*, Nonlinear Analysis: Real World Applications **10** (2009), no. 4, 2396–2406.
- [10] V. Feoktistov, *Differential Evolution. In Search of Solutions*, Springer, New York (2006).
- [11] T. Fischer, *Risk capital allocation by coherent risk measures based on one-sided moments*, Insurance: Mathematics and Economics **32** (2003), no. 1, 135–146.
- [12] P.C. Fishburn, *Mean-risk analysis with risk associated with below-target returns*, The American Economic Review **67** (1977), no. 2, 116–126.
- [13] R. Fletcher, *A Class of Methods for Nonlinear Programming with Termination and Convergence Properties*, in: Integer and Nonlinear Programming (J. Abadie, ed.), North-Holland, Amsterdam, 1970, pp. 157–173.
- [14] ———, *Practical Methods of Optimization*, John Wiley & Sons, Glichester, 1991.
- [15] R. Jansen and R. van Dijk, *Optimal benchmark tracking with small portfolios*, The Journal of Portfolio Management **28** (2002), no. 2, 33–39.
- [16] J. Kennedy and R.C. Eberhart, *Particle Swarm Optimization*, in: Proceedings of the IEEE International Conference on Neural Networks IV (1995), pp. 1942–1948.

- [17] H.M. Markowitz, *Portfolio selection*, The Journal of Finance **7** (1952), no. 1, 77–91.
- [18] A.J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, Princeton, 2005.
- [19] C.H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Prentice-Hall, New Jersey, 1982.
- [20] T. Pietrzykowski, *An exact potential method for constrained maxima*, SIAM Journal on Numerical Analysis **6** (1969), no. 2, 299–304.
- [21] G. Di Pillo and L. Grippo, *Exact penalty functions in constrained optimization*, SIAM Journal on Control and Optimization **27** (1989), no. 6, 1333–1360.
- [22] R. Poli, J. Kennedy, and T. Blackwell, *Particle Swarm Optimization – An overview*, Swarm Intelligence **1** (2007), no. 1, 33–57.
- [23] S. Rachev, S. Ortobelli, S. Stoyanov, F.J. Fabozzi, and A. Biglova, *Desirable properties of an ideal risk measure in portfolio theory*, International Journal of Theoretical and Applied Finance **11** (2008), no. 1, 19–54.
- [24] R.T. Rockafellar and S. Uryasev, *Conditional value-at-risk for general loss distributions*, Journal of Banking and Finance **26** (2002), no. 7, 1443–1471.
- [25] Y. Shi and R. Eberhart, *A modified particle swarm optimizer*, in: Proceedings of the IEEE International Conference on Evolutionary Computation, 1998, pp. 69–73.
- [26] N. Thomaidis, T. Angelidis, V. Vassiliadis, and G. Dounias, *Active portfolio management with cardinality constraints: An application of Particle Swarm Optimization*, New Mathematics and Natural Computation **5** (2009), no. 3, 535–555.
- [27] R.F. Wipperfurth, *Utility implications of portfolio selection and performance appraisal model*, Journal of Financial and Quantitative Analysis **6** (1971), no. 3, 913–924.
- [28] M.R. Young, *A minimax portfolio selection rule with linear programming solution*, Management Science **44** (1998), no. 5, 673–683.
- [29] W.I. Zangwill, *Non-linear programming via penalty functions*, Management Science **13** (1967), no. 5, 344–358.
- [30] W.J. Zhang, X.F. Xie, and D.C. Bi, *Handling boundary constraints for numerical optimization by particle swarm flying in periodic search space*, in: Proceedings of the 2004 Congress on Evolutionary Computation, 2004, pp. 2307–2311.