STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **10**, June 2022, pp 775–788.
Published online in International Academic Press (www.IAPress.org)

Penalty ADM Algorithm for Cardinality Constrained Mean-Absolute Deviation Portfolio Optimization

Temadher Alassiry Almaadeed[1], Tahereh Khodamoradi[2], Maziar Salahi [2] Abdelouahed Hamdi [1]

[1] *Department of Mathematics, Statistics and Physics, College of Arts and Sciences, Qatar University, Doha, Qatar*
[2] *Department of Applied Mathematics, Faculty of Mathematical Sciences, University of Guilan, Rasht, Iran*

Abstract In this paper, we study the cardinality constrained mean-absolute deviation portfolio optimization problem with risk-neutral interest rate and short-selling. We enhance the model by adding extra constraints to avoid investing in those stocks without short-selling positions. Also, we further enhance the model by determining the short rebate based on the return. The penalty alternating direction method is used to solve the mixed integer linear model. Finally, numerical experiments are provided to compare all models in terms of Sharpe ratios and CPU times using the data set of the NASDAQ and S&P indexes.

Keywords MAD model; Cardinality constrained; PADM method.

AMS 2010 subject classifications 90C11, 91G10
DOI: 10.19139/soic-2310-5070-1312

1. Introduction

The Mean-Variance (MV) model of Harry Markowitz [36] gives the best possible trade-off between mean of returns and variance of returns. Since it is a quadratic optimization problem and the estimation of the covariance could be difficult for large data sets, the mean-absolute deviation (MAD) model is proposed by Konno and Yamazaki [30] defining the absolute deviation of the rate of return as the measure of risk instead of variance. Simaan [40] has given a comparison of the MAD model and the MV model. Feinstein and Thapa [14] have given a reformulation of the MAD model with fewer constraints, which is computationally better. Later, Chang [10] presented a new formulation of the one by Feinstein and Thapa. For further details of the MAD model, we refer to [34].

With all these advantages, in the MAD model some constraints such as short-selling and cardinality constraints that are more realistic, were ignored. Short-selling is borrowing stocks and selling them. The investor must lend stocks with the same value to the lender of stocks. The first short-selling model in portfolio theory is proposed by Lintner [33]. Konno et al. used the long-short strategy in the MAD ?model under non-convex transaction costs [29]. They observed that this strategy finds a portfolio that is much better compared to the portfolio with only the long strategy. Cardinality constraints restrict the number of stocks in the portfolio and the threshold constraints also restrict the proportion of each stock in portfolio within certain interval. Gao and Li [15] studied Cardinality Constrained Mean-Variance (CCMV) portfolio optimization problem with short-selling. The authors in [31] extended the MAD model that takes the cardinality and the threshold constraints with short-selling. Their new model is presented as a mixed integer problem and DC (Difference of Convex functions) algorithm is used to

*Correspondence to: Abdelouahed Hamdi (Email: abhamdi@qu.edu.qa).Department of Mathematics, Statistics and Physics, College of Arts and Sciences, Qatar University, Doha, Qatar.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

solve it [3, 37, 42, 43]. Other important studies of the MAD model are those of Angelelli et al. [4], Byrne and Lee [9], Tavakoli [5], Li et al. [32], Silva et al. [39], Carnia et al. [23]. Another feature that one can use in the model is the so-called risk-neutral interest rate [6, 25, 35]. Jacobs et al. [21, 22] applied short-selling strategy with risk-neutral interest rate in portfolio selection model. Most recently, in [26, 27] also the authors have used short-selling with risk-neutral interest rate in the CCMV model and discussed some drawbacks of the model. Alternating direction method (ADM) is an iterative procedure that solves optimization models by alternatingly solving two simpler subproblems [7]. There are some strategies that have been studied to make the ADM more practical and efficient in [11, 20, 24, 41]. Also, an extension of the ADM called penalty ADM (PADM) proposed in [44] that is appropriate for biconvex sets and optimization with biconvex functions [18], k-means clustering problems [8], mixed-integer nonlinear problems [16, 17, 38], and bilevel problems [28]. Motivated by the fact that most MAD models do not take into account more realistic constraints, the contributions of the paper are as follows. First, we present an enhanced version of the MAD model including short-selling, risk-neutral interest rate and cardinality constraints. The model is enhanced by adding extra constraints to positive investment in those stocks having short positions. Also, these constraints avoid to invest short and long in the same stock. We further propose an improved version of it, in which the short rebate is determined based on the stock's return. Thus, it leads to a model with fewer constraints and binary variables. Finally, the PADM is applied to solve the proposed model.

The rest of this paper is as follows. In Section 1, we present the MAD model with short-selling and risk-neutral interest rate in details.In Section 2, we also extend the model with the cardinality constraints, short-selling and risk-neutral interest rate. In Section 3, we present the details of the PADM for solving the proposed mixed integer models with its convergence.Finally, in Section 4 some numerical results are provided to show the efficiency of the proposed models.

2. Extended MAD model

The MAD model with short-selling and risk-neutral interest rate is as follows [26]:

$$
\begin{aligned}
\min_{x,u}\ & \lambda\,(\frac{1}{T}\sum_{t=1}^{T}|\sum_{j=1}^{N}(r_{t,j}-r_j)x_j|)-(1-\lambda)(\sum_{j=1}^{N}x_jr_j-r_ch_jx_j)\\
\text{s.t.}\ & \sum_{j=1}^{N}x_j=1,\\
& \varepsilon_j\le x_j\le\delta_j,\quad j=1,...,N,\\
& if\ \ x_j\ge 0,\ then\ \ h_j=0,\\
& if\ \ x_j<0,\ then\ \ 0<h_j<1,
\end{aligned}
\tag{1}
$$

which is equivalent to

$$
\begin{aligned}
\min_{x,u}\quad & \lambda\,(\frac{1}{T}\sum_{t=1}^{T}u_t)-(1-\lambda)(\sum_{j=1}^{N}x_jr_j-r_ch_jx_j)\\
\text{s.t.}\quad & u_t+\sum_{j=1}^{N}(r_{t,j}-r_j)x_j\ge 0,\quad t=1,...,T,\\
& u_t-\sum_{j=1}^{N}(r_{t,j}-r_j)x_j\ge 0,\quad t=1,...,T,
\end{aligned}
\tag{2}
$$

$$
\begin{aligned}
&\sum_{j=1}^{N} x_j = 1,\\
&u_t \geq 0, \quad t = 1, ..., T,
\end{aligned}
$$

$$
\begin{aligned}
&\varepsilon_j \leq x_j \leq \delta_j, \quad j = 1, ..., N,\\
&if \ \ x_j \geq 0, \ then \ \ h_j = 0,\\
&if \ \ x_j < 0, \ then \ \ 0 < h_j < 1,
\end{aligned}
$$

where N and T denote the number of stocks and the end of investment time, respectively, x_j is the proportion of investment in jth stock, $r_{t,j}$ is the return of the jth stock at time t, $(t = 1, ..., T; j = 1, ..., N)$, r_j is the expected return of the jth stock $(j = 1, ..., N)$ and r_c is risk-neutral interest rate. Also, ε_j and δ_j are the lower and upper bounds of the jth stock, respectively, and $\lambda \in [0, 1]$ is the risk aversion parameter. The ε_j is negative for short-selling. The term $r_c \sum_{j=1}^{N} h_j x_j$ shows the short rebate, where

$$0 < h_j < 1, \quad \forall j,$$

denotes the portion of the investor of the interest on the proceeds from the short-sale of stock j. On the other hand, for stocks that are not in the short-selling position, we add constraints $r_j x_j \geq 0$, $(j = 1, 2, ..., N)$ to model (2) in order to have the proportion of investment to be positive. To linearize (2), let $p_j = h_j x_j$, and add the following constraints:

$$
\begin{aligned}
&p_j - M\, w_j \leq 0,\\
&c\, x_j - p_j + M\, w_j \leq M,\\
&-c\, x_j + p_j \leq 0,\\
&p_j + M\, w_j \geq 0,\\
&x_j \geq -M\, w_j,\\
&x_j \leq M\,(1 - w_j),\\
&w_j \in \{0, 1\},
\end{aligned}
\tag{3}
$$

where M is a large positive constant. If $x_j \leq 0$, then from (3) $w_j = 1$ and thus $p_j = c\, x_j$. Also, if $x_j \geq 0$, we have $w_j = 0$ and thus $p_j = 0$. Therefore, we get the following model:

$$
\begin{aligned}
\min_{x,u,p,w} \ & \lambda\, (\frac{1}{T} \sum_{t=1}^{T} u_t) - (1 - \lambda)(\sum_{j=1}^{N} x_j r_j - r_c p_j)\\
\text{s.t. } & u_t + \sum_{j=1}^{N} (r_{t,j} - r_j) x_j \geq 0, \quad t = 1, ..., T,\\
& u_t - \sum_{j=1}^{N} (r_{t,j} - r_j) x_j \geq 0, \quad t = 1, ..., T,\\
& \sum_{j=1}^{N} x_j = 1,\\
& r_j x_j \geq 0, \quad j = 1, ..., N,\\
& \varepsilon_j \leq x_j \leq \delta_j, \quad j = 1, ..., N,\\
& p_j - M\, w_j \leq 0, \qquad j = 1, ..., N,
\end{aligned}
\tag{4}
$$

$$
\begin{aligned}
& c\,x_j - p_j + M\,w_j \le M, \qquad j = 1,...,N, \\
& -c\,x_j + p_j \le 0, \qquad j = 1,...,N, \\
& p_j + M\,w_j \ge 0, \qquad j = 1,...,N, \\
& x_j \ge -M\,w_j, \qquad j = 1,...,N, \\
& x_j \le M\,(1 - w_j), \qquad j = 1,...,N, \\
& w_j \in \{0,1\}, \qquad j = 1,...,N, \\
& u_t \ge 0, \qquad t = 1,...,T,
\end{aligned}
$$

where h_j's are equal to $c \in (0,1)$, which is a constant. As we see, model (4) has N binary variables and $8N + 2T + 1$ constraints that might be computationally inefficient for large N and T values. Thus, let us determine the short rebate (h_j) in model (2) based on the return. If $r_j \ge 0$, we let $h_j = 0$ and if $r_j < 0$, then we let $0 < h_j < 1$. The enhanced variant of model (4) with less constraints and no binary variables is as follows:

$$
\begin{aligned}
\min_{x,u}\ & \lambda\,(\frac{1}{T}\sum_{t=1}^{T}\ u_t) - (1-\lambda)(\sum_{j=1}^{N}\ x_j r_j - r_c h_j x_j) \\
\text{s.t. } & u_t + \sum_{j=1}^{N}\ (r_{t,j} - r_j)x_j \ge 0, \quad t = 1,...,T, \\
& u_t - \sum_{j=1}^{N}\ (r_{t,j} - r_j)x_j \ge 0, \quad t = 1,...,T, \\
& \sum_{j=1}^{N} x_j = 1, \\
& r_j x_j \ge 0, \quad j = 1,...,N, \\
& \varepsilon_j \le x_j \le \delta_j, \quad j = 1,...,N, \\
& u_t \ge 0, \quad t = 1,...,T.
\end{aligned} \tag{5}
$$

Lemma 1
Let V^* and U^* be the optimal objective functions values of (4) and (5), respectively. Then $V^* = U^*$.

Proof
In both models, based on $r_j x_j \ge 0$, for $r_j \ge 0$ we have $x_j \ge 0$ then $h_j = 0$, and for $r_j < 0$ we have $x_j < 0$ then $h_j = c$. Let S_1 and S_2 be the set of feasible points of models (4) and (5), respectively. Since any feasible point in (4) is also feasible for (5), we conclude that $S_1 \subseteq S_2$. Now, let (x^*, u^*) be the optimal solution of (5). One can see, in (4), when $x_j \ge 0$, from $x_j \le M(1 - w_j)$ and $x_j \ge -M w_j$, we conclude $w_j = 0$. Similarly, when $x_j < 0$ we conclude $w_j = 1$. Further, if $w_j = 0$, from $p_j - M w_j \le 0$ and $p_j + M w_j \ge 0$ we have $p_j = 0$. Also if $w_j = 1$, from $-c x_j + p_j \le 0$ and $c x_j - p_j + M w_j \le M$ we have $p_j = c x_j$.

On the other hand, we consider $w_j^* = 0$ for $x_j^* \ge 0$ and $w_j^* = 1$ for $x_j^* < 0$ and by considering $p_j^* = h_j x_j^*$ (if $x_j^* \ge 0$, then $h_j = 0$; and if $x_j^* < 0$, then $h_j = c$), (x^*, u^*, w^*, p^*) is feasible for (4). Now, we show it is also optimal for (4). Suppose by contradiction that $(x^{**}, u^{**}, w^{**}, p^{**})$ is an optimal solution for (4), then $\lambda\,(\frac{1}{T}\sum_{t=1}^{T}\ u_t^{**}) - (1-\lambda)(\sum_{j=1}^{N}\ x_j^{**} r_j - r_c p_j^{**}) \le \lambda\,(\frac{1}{T}\sum_{t=1}^{T}\ u_t^{*}) - (1-\lambda)(\sum_{j=1}^{N}\ x_j^{*} r_j - r_c p_j^{*})$. Since $p_j^{**} = h_j x_j^{**}$ and $p_j^* = h_j x_j^*$, (if $x_j^* \le 0$ and $x_j^{**} \le 0$, then $h_j = 0$, and if $x_j^* \ge 0$ and $x_j^{**} \ge 0$, then $h_j = c$) we have $\lambda\,(\frac{1}{T}\sum_{t=1}^{T}\ u_t^{**}) - (1-\lambda)(\sum_{j=1}^{N}\ x_j^{**} r_j - r_c h_j x_j^{**}) \le \lambda\,(\frac{1}{T}\sum_{t=1}^{T}\ u_t^{*}) - (1-\lambda)(\sum_{j=1}^{N}\ x_j^{*} r_j - r_c h_j x_j^{*})$. This is in contradiction with the optimality of (x^*, u^*) for (5), since (x^{**}, u^{**}) is also feasible for (5). This complates the proof. □

As we see, models (4) and (5) have equal optimal objective values while model (5) has much less constraints and no binary variables.

3. Models with cardinality constraintes

In this section, we also include cardinality constraintes in models (4) and (5) to restrict stocks' number in the portfolio as follows:

$$
\begin{aligned}
\min_{x,z,u,p,w}\ & \lambda\,(\frac{1}{T}\sum_{t=1}^{T} u_t) - (1-\lambda)(\sum_{j=1}^{N} x_j r_j - r_c p_j) \\
\text{s.t. } & u_t + \sum_{j=1}^{N} (r_{t,j} - r_j)x_j \geq 0,\ \ t = 1, ..., T, \\
& u_t - \sum_{j=1}^{N} (r_{t,j} - r_j)x_j \geq 0,\ \ t = 1, ..., T, \\
& \sum_{j=1}^{N} x_j = 1, \\
& \sum_{j=1}^{N} z_j = K, \\
& r_j x_j \geq 0,\ \ j = 1, ..., N, \\
& \varepsilon_j z_j \leq x_j \leq \delta_j z_j,\ \ j = 1, ..., N, \\
& p_j - M\, w_j \leq 0,\ \ j = 1, ..., N, \\
& c\, x_j - p_j + M\, w_j \leq M,\ \ j = 1, ..., N, \\
& -c\, x_j + p_j \leq 0,\ \ j = 1, ..., N, \\
& p_j + M\, w_j \geq 0,\ \ j = 1, ..., N, \\
& x_j \geq -M\, w_j,\ \ j = 1, ..., N, \\
& x_j \leq M\,(1 - w_j),\ \ j = 1, ..., N, \\
& w_j \in \{0, 1\},\ \ j = 1, ..., N, \\
& z_j \in \{0, 1\},\ \ j = 1, ..., N, \\
& u_t \geq 0,\ \ t = 1, ..., T,
\end{aligned}
\tag{6}
$$

and

$$
\begin{aligned}
\min_{x,u}\ & \lambda\,(\frac{1}{T}\sum_{t=1}^{T} u_t) - (1-\lambda)(\sum_{j=1}^{N} x_j r_j - r_c h_j x_j) \\
\text{s.t. } & u_t + \sum_{j=1}^{N} (r_{t,j} - r_j)x_j \geq 0,\ \ t = 1, ..., T, \\
& u_t - \sum_{j=1}^{N} (r_{t,j} - r_j)x_j \geq 0,\ \ t = 1, ..., T,
\end{aligned}
$$

$$\sum_{j=1}^{N} x_j = 1, \tag{7}$$
$$\sum_{j=1}^{N} z_j = K,$$
$$r_j x_j \geq 0, \;\; j = 1, ..., N,$$
$$\varepsilon_j z_j \leq x_j \leq \delta_j z_j, \;\; j = 1, ..., N,$$
$$z_j \in \{0, 1\}, \;\; j = 1, ..., N,$$
$$u_t \geq 0, \;\; t = 1, ..., T,$$

where K is the desired number of stocks in the portfolio and z_j's are zero-one variables. If $z_j = 1$, stock j is in portfolio and if $z_j = 0$ it does not.

Lemma 2
Let W^* and G^* denote the optimal objective values of models (6) and (7), respectively. Then $W^* = G^*$.

Proof
In both models, based on $r_j x_j \geq 0$, for $r_j \geq 0$ we have $x_j \geq 0$ then $h_j = 0$, and for $r_j < 0$ we have $x_j < 0$ then $h_j = c$. Let S_1 and S_2 be the set of feasible points of models (6) and (7), respectively. Since any feasible point in (6) is also feasible for (7), $S_1 \subseteq S_2$. Now, let (x^*, u^*, z^*) be the optimal solution of (7). As we see, in (6), when $x_j \geq 0$, from $x_j \leq M(1 - w_j)$ and $x_j \geq -M w_j$, we conclude $w_j = 0$. Similarly, when $x_j < 0$ we conclude $w_j = 1$. Further, if $w_j = 0$, from $p_j - M w_j \leq 0$ and $p_j + M w_j \geq 0$ we have $p_j = 0$. Also if $w_j = 1$, from $-c x_j + p_j \leq 0$ and $c x_j - p_j + M w_j \leq M$ we have $p_j = c x_j$.

On the other hand, we consider $w_j^* = 0$ for $x_j^* \geq 0$ and $w_j^* = 1$ for $x_j^* < 0$ and by considering $p_j^* = h_j x_j^*$ (if $x_j^* \geq 0$, then $h_j = 0$; and if $x_j^* < 0$, then $h_j = c$), $(x^*, u^*, z^*, w^*, p^*)$ is feasible for (6). Now, we indicate it is also optimal for (6). Suppose by contradiction that $(x^{**}, u^{**}, z^{**}, w^{**}, p^{**})$ is an optimal solution for model (6), then $\lambda\, (\frac{1}{T} \sum_{t=1}^{T} \; u_t^{**}) - (1 - \lambda)(\sum_{j=1}^{N} \; x_j^{**} r_j - r_c p_j^{**}) \leq \lambda\, (\frac{1}{T} \sum_{t=1}^{T} \; u_t^{*}) - (1 - \lambda)(\sum_{j=1}^{N} \; x_j^{*} r_j - r_c p_j^{*})$. Since $p_j^{**} = h_j x_j^{**}$ and $p_j^* = h_j x_j^*$, (if $x_j^* \leq 0$ and $x_j^{**} \leq 0$, then $h_j = 0$, and if $x_j^* \geq 0$ and $x_j^{**} \geq 0$, then $h_j = c$) we have $\lambda\, (\frac{1}{T} \sum_{t=1}^{T} \; u_t^{**}) - (1 - \lambda)(\sum_{j=1}^{N} \; x_j^{**} r_j - r_c h_j x_j^{**}) \leq \lambda\, (\frac{1}{T} \sum_{t=1}^{T} \; u_t^{*}) - (1 - \lambda)(\sum_{j=1}^{N} \; x_j^{*} r_j - r_c h_j x_j^{*})$. This is in contradiction with the optimality of (x^*, u^*, z^*) for (7), since (x^{**}, u^{**}, z^{**}) is also feasible for (7). This complates the proof. $\square$

As we see, the optimal objective function values of (6) and (7) are equal, while model (7) has fewer constraints and binary variables. Though model (7) is solved faster than model (6) as can be seen in Section 4, but for higher dimensions and some K values it is still slow. Thus, in the next section, we discuss the details of PADM to solve it.

4. The PADM

Here, we give the details of the PADM for solving model (7). To do so, we reformulate model (7) as follows:

$$
\begin{aligned}
\min_{x,u,w,z}\ & f(x,u,w,z) := \lambda\,(\frac{1}{T}\sum_{t=1}^{T}\ u_t) - (1-\lambda)(\sum_{j=1}^{N}\ x_j r_j - r_c h_j x_j) \\
\text{s.t. } & u_t + \sum_{j=1}^{N}\ (r_{t,j} - r_j)x_j \geq 0,\ \ t = 1, ..., T, \\
& u_t - \sum_{j=1}^{N}\ (r_{t,j} - r_j)x_j \geq 0,\ \ t = 1, ..., T, \\
& \sum_{j=1}^{N} x_j = 1, \\
& \sum_{j=1}^{N} z_j = K, \\
& r_j x_j \geq 0,\ \ j = 1, ..., N, \\
& \varepsilon_j z_j \leq w_j \leq \delta_j z_j,\ \ j = 1, ..., N, \\
& u_t \geq 0,\ \ t = 1, ..., T, \\
& z_j \in \{0,1\},\ \ j = 1, ..., N, \\
& x = w.
\end{aligned}
\tag{8}
$$

Now let

$$
C = \{(x_j, u_t):\ u_t + \sum_{j=1}^{N}\ (r_{t,j} - r_j)x_j \geq 0, u_t - \sum_{j=1}^{N}\ (r_{t,j} - r_j)x_j \geq 0, \\
\sum_{j=1}^{N} x_j = 1, r_j x_j \geq 0, u_t \geq 0\}
$$

and

$$
D = \{(z_j, w_j):\ \sum_{j=1}^{N} z_j = K, \sum_{j=1}^{N} w_j = 1, \varepsilon_j z_j \leq w_j \leq \delta_j z_j,\ z_j \in \{0,1\}\}.
$$

Since constraint $\sum_{j=1}^{N} x_j = 1$ is considered in the set C, it does not necessarily need to be included in the set D. However, computational experiments showed that the presence of this constraint in both subproblems remarkably improves the PADM solution. Now, we define the penalty sumproblem of model (8) as follows:

$$
\begin{aligned}
\min_{x,u,w,z}\ & \Phi(x,u,w,z;\gamma) := \lambda\,(\frac{1}{T}\sum_{t=1}^{T}\ u_t) - (1-\lambda)(\sum_{j=1}^{N}\ x_j r_j - r_c h_j x_j) + \gamma||x - w||_1 \\
\text{s.t. } & (x,u) \in C, \\
& (w,z) \in D,
\end{aligned}
\tag{9}
$$

where $\gamma > 0$ is the penalty parameter. The PADM steps for a given starting point $w^{s,l}$ and initial value γ^s for a penalty parameter are as follows:
Step 1: $(x^{s,l+1}, u^{s,l+1}) \in \underset{(x,u)\in C}{\operatorname{argmin}}\ \Phi(x, u, w^{s,l}; \gamma^s)$.

Step 2: $(w^{s,l+1}, z^{s,l+1}) \in \underset{(w,z)\in D}{\operatorname{argmin}} \ \Phi(x^{s,l+1}, w, z; \gamma^s)$.
Step 3: Update γ^s.
In Step 1, we solve the following problem:

$$\begin{aligned} \min_{x,u} \ & \Phi(x, u, w^{s,l}; \gamma^s) \\ \text{s.t. } & u_t + \sum_{j=1}^{N} (r_{t,j} - r_j)x_j \geq 0, \quad t = 1, ..., T, \\ & u_t - \sum_{j=1}^{N} (r_{t,j} - r_j)x_j \geq 0, \quad t = 1, ..., T, \\ & \sum_{j=1}^{N} x_j = 1, \\ & r_j x_j \geq 0, \quad j = 1, ..., N, \\ & \varepsilon_j \leq x_j \leq \delta_j, \quad j = 1, ..., N, \\ & u_t \geq 0, \quad t = 1, ..., T, \geq 0. \end{aligned} \tag{10}$$

Now by using $(x^{s,l+1}, u^{s,l+1})$, the optimal solution of (10), in Step 2 we solve the following problem:

$$\begin{aligned} \min_{w,z} \quad & \Phi(x^{s,l+1}, w, z; \gamma^s) \\ \text{s.t.} \quad & \sum_{j=1}^{N} w_j = 1, \\ & \sum_{j=1}^{N} z_j = K, \\ & \varepsilon_j z_j \leq w_j \leq \delta_j z_j, \\ & z_j \in \{0, 1\}. \end{aligned} \tag{11}$$

The PADM algorithm can be outlined as follows:

The PADM Algorithm

1: Choose appropriate starting points $(x^{0,0}, w^{0,0})$ and penalty parameter $\gamma^0 > 0$.
2: For s=0,1,..., maxiter **do**
3: Set l=0.
4: while $(x^{s,l}, w^{s,l})$ is not a partial minimum of (9) with $\gamma = \gamma^s$ **do**
5: Compute $(x^{s,l+1}, u^{s,l+1})$ by solving (10).
6: Compute $(w^{s,l+1}, z^{s,l+1})$ by solving (11).
8: end while
9: Choose new penalty parameter $\gamma^{s+1} \geq \gamma^s$
10: end for.

Definition 1
([12]): Let (x^*, u^*, w^*, z^*) be in the feasible set of (8). Then it is called a partial minimum for (8), if it satisfies

$$f(x^*, u^*, w^*, z^*) \leq f(x, u, w^*, z^*),$$

for all (x, u, w^*, z^*) in the feasible set of (8) and

$$f(x^*, u^*, w^*, z^*) \leq f(x^*, u^*, w, z),$$

for all (x^*, u^*, w, z) in the feasible set of (8).

With partial minimum, if the coupling constraints are satisfied, we stop with a feasible solution of model (8). Otherwise, we update the penalty parameter and solve the next penalty problem to find a partial minimum.

Lemma 3
Let (x^*, u^*, w^*, z^*) be a partial minimum of $\Phi(x, u, w, z; \gamma)$ for a fixed $\gamma \geq 0$ and let it is feasible for model (8). Then (x^*, u^*, w^*, z^*) is also a partial minimum of model (8).

Proof
Let $(x, u) \in C$ be such that (x, u, w^*, z^*) is feasible for model (8). Then we have

$$\begin{aligned} f(x, u, w^*, z^*) &= f(x, u, w^*, z^*) + \gamma ||x - w^*||_1 \\ &= \Phi(x, u, w^*, z^*; \gamma) \geq \Phi(x^*, u^*, w^*, z^*; \gamma) \\ &= f(x^*, u^*, w^*, z^*) + \gamma ||x^* - w^*||_1 \\ &= f(x^*, u^*, w^*, z^*). \end{aligned}$$

Similarly, for all $(w, z) \in D$ such that (x^*, u^*, w, z) is feasible for model (8) we have

$$\begin{aligned} f(x^*, u^*, w, z) &= f(x^*, u^*, w, z) + \gamma ||x^* - w||_1 \\ &= \Phi(x^*, u^*, w, z; \gamma) \geq \Phi(x^*, u^*, w^*, z^*; \gamma) \\ &= f(x^*, u^*, w^*, z^*) + \gamma ||x^* - w^*||_1 \\ &= f(x^*, u^*, w^*, z^*). \end{aligned}$$

Therefore, (x^*, u^*, w^*, z^*) is a partial minimum of model (8). □

In the sequel, we prove the convergence of PADM Algorithm.

Theorem 1
Suppose that $\gamma^s \nearrow \infty$ and (x^s, u^s, w^s, z^s) be a sequence of partial minima of (9) generated by the PADM Algorithm with $(x^s, u^s, w^s, z^s) \longrightarrow (x^*, u^*, w^*, z^*)$. Then, (x^*, w^*) is a partial minimizer of $||x - w||_1$.

Proof
Let (x^s, u^s, w^s, z^s) be a partial minimum of $\Phi(x, u, w, z; \gamma)$. Thus

$$\Phi(x, u, w^s, z^s; \gamma^s) \geq \Phi(x^s, u^s, w^s, z^s; \gamma^s),$$

for all $(x, u) \in C$ and

$$\Phi(x^s, u^s, w, z; \gamma^s) \geq \Phi(x^s, u^s, w^s, z^s; \gamma^s),$$

for all $(w, z) \in D$, which are equivalent to

$$f(x, u, w^s, z^s) + \gamma^s ||x - w^s||_1 \geq f(x^s, u^s, w^s, z^s) + \gamma^s ||x^s - w^s||_1, \tag{12}$$

for all $(x, u) \in C$ and

$$f(x^s, u^s, w, z) + \gamma^s ||x^s - w||_1 \geq f(x^s, u^s, w^s, z^s) + \gamma^s ||x^s - w^s||_1, \tag{13}$$

for all $(w, z) \in D$. Since the sequence γ^s is unbounded, by dividing (12) and (13) by γ^s, we have

$$\frac{1}{\gamma^s} f(x, u, w^s, z^s) + \frac{\gamma^s}{\gamma^s} ||x - w^s||_1 \geq \frac{1}{\gamma^s} f(x^s, u^s, w^s, z^s) + \frac{\gamma^s}{\gamma^s} ||x^s - w^s||_1,$$

for all $(x, u) \in C$ and

$$\frac{1}{\gamma^s} f(x^s, u^s, w, z) + \frac{\gamma^s}{\gamma^s} ||x^s - w||_1 \geq \frac{1}{\gamma^s} f(x^s, u^s, w^s, z^s) + \frac{\gamma^s}{\gamma^s} ||x^s - w^s||_1,$$

for all $(w, z) \in D$. Then for $s \to \infty$ we obtain

$$||x - w^*||_1 \geq ||x^* - w^*||_1,$$

for all $(x, u) \in C$ and

$$||x^* - w||_1 \geq ||x^* - w^*||_1,$$

for all $(w, z) \in D$. This proves the result. □

5. Numerical experiments

In this section, first we compare the performance of models (4)-(7) in terms of in-sample and out-of-sample Sharpe ratios by the data of S&P index on Information Technology[†] for 2015-2017 with 48 stocks. For the out-of-sample performance, we apply the rolling-horizon procedure [13]. The in-sample Sharpe ratio (SR) formula is

$$SR = \frac{\mu - r_c}{\delta},$$

where μ, r_c and δ are the expected portfolio return, risk-neutral interest rate and the mean-absolute deviation, respectively, and the out-of-sample Sharpe ratio ($\widehat{SR}$) is

$$\widehat{SR} = \frac{\widehat{\mu} - r_c}{\widehat{\delta}},$$

where

$$\widehat{\delta} = \frac{1}{T - \tau - 1} \sum_{t=\tau}^{T-1} \mid (x_t' \, r_{t+1} - \widehat{\mu}) \mid,$$

$$\widehat{\mu} = \frac{1}{T - \tau} \sum_{t=\tau}^{T-1} (x_t' \, r_{t+1}),$$

and x_t is the proportion of investment at time t, $t = \tau, \tau + 1, ..., T - 1$, τ is the length of the estimation time window, r_{t+1} denotes the stock return and T is the total number of returns in the data set. By considering the stocks monthly return from $2/1/2015$ to $29/12/2017$ and an estimation window of $\tau = 36$ data, for 3 years, we use the 2018 data for the out-of-sample.

We compare the Sharpe ratios of models (4) and (5). The results are reported in Table 1, for $r_c = 0.03$, $c = 0.1$, and taking $-\varepsilon_j = \delta_j = 0.1$ that are the lower and upper bounds of the proportion of investment in any stock. We performed all computations in MATLAB R2017a on a 2.50 GHz laptop with 4 GB of RAM, and we used CVX 2.2

Table 1. In-sample and out-of-sample portfolio Sharpe ratios of models (4) and (5) for Information Technology of S&P index data with $\lambda = 0.5$.

	Model (4) ($x \geq 0$, $r_c = 0$, $c = 0$)	Model (4) ($r_c = 0$, $c = 0$)	Model (4)	Model (5)
In-sample	0.3020	0.3797	0.4190	0.4190
Out-of-sample	-0.0237	0.0742	0.0776	0.0776

software using MATLAB [19] to solve the models. In this table, $c = 0$ and $r_c = 0$, means no risk-neutral interest rate and $x \geq 0$ shows also that no short-selling.

As we see, in Column 4, the in-sample and out-of-sample portfolio Sharpe ratios of (4) with risk-neutral interest rate and short-selling are higher than the one in Column 2 (Column 3) without risk-neutral interest rate and short-selling (without risk-neutral interest rate and with short-selling). Therefore, the short-selling and risk-neutral interest rate increase the in-sample and out-of-sample portfolio Sharpe ratios. On the other hand, models (4) and (5) with these two factor in columns 4 and 5 have similar in-sample and out-of-sample portfolio Sharpe ratios while model (5) is solved much faster as can be seen in Table 3.

We also compare the Sharpe ratios of models (6) and (7). To solve the mixed-integer models, we used Mosek in CVX. The results are reported in Table 2 for different K values.

Table 2. In-sample and out-of-sample portfolio Sharpe ratios of models (6) and (7) for Information Technology of S&P index data with $\lambda = 0.5$.

Desired number of stocks		Model (6) ($x \geq 0$, $r_c = 0$, $c = 0$)	Model (6) ($r_c = 0$, $c = 0$)	Model (6)	Model (7)
$K = 10$	In-sample	0.2893	0.2893	0.2893	0.2893
	Out-of-sample	-0.1127	-0.1127	-0.1127	-0.1127
$K = 20$	In-sample	0.3020	0.3620	0.4020	0.4020
	Out-of-sample	-0.0232	0.0736	0.0761	0.0761
$K = 30$	In-sample	0.3020	0.3797	0.4190	0.4190
	Out-of-sample	-0.0231	0.0718	0.0786	0.0786
$K = 40$	In-sample	0.3020	0.3797	0.4190	0.4190
	Out-of-sample	-0.0237	0.0742	0.0776	0.0776

As one can see, model (6) in Column 4, without risk-neutral interest rate and with short-selling has higher in-sample and out-of-sample Sharpe ratios than the one without these two factors in Column 3. The in-sample and out-of-sample Sharpe ratios of (6) in Column 5 with both risk-neutral interest rate and short-selling are higher than the one in Column 4 with short-selling and without risk-neutral interest rate. On the other hand, models (6) and (7) with risk-neutral interest rate and short-selling (Columns 5 and 6) have similar in-sample and out-of-sample Sharpe ratios.

[†]https://finance.yahoo.com

To compare the CPU times of models (4)-(7), we used the data set of the NASDAQ? and S&P indexes from December 2006 until March 2008 when $T = 61$, $r_c = 0.03$, $c = 0.1$, and taking $-\varepsilon_j = \delta_j = 0.3$. The results for models (4) and (5) are summarized in Table 3. These results show that model (5) is solved much faster than model (4).

Table 3. Comparison of the CPU times of models (4) and (5) for NASDAQ and S&P indexes data with $\lambda = 0.5$.

Indexes	Number of stocks	Model (4)	Model (5)
$S\&P$	$N = 280$	$1.0011e + 03$	1.454
	$N = 322$	$1.0036e + 03$	1.862
NASDAQ	$N = 456$	1.0065e+03	2.177
	$N = 570$	1.0085e+03	2.182

The corresponding results for models (6) and (7) are reported in Table 4 for different K values. These results indicate that (7) is solved much faster than (6), but it still needs significant amount of time for higher dimensions and some K values. Thus we compare the performance of PADM Algorithm in terms of objective values and CPU times for model (9), with model (7) that is solved by Mosek in CVX for the data set of the NASDAQ? and S&P indexes from December 2006 until March 2008 when $T = 61$, $r_c = 0.03$, $c = 0.1$, and taking $-\varepsilon_j = \delta_j = 0.3$. In model (9), the initial penalty parameter is set to 0.1 and it is updated with the factor 10. The inner loop is stopped when $||(x^{s,l}, w^{s,l}) - (x^{s,l-1}, w^{s,l-1})||_\infty \leq 10^{-5}$. The PADM also terminates with a partial minimum if $||x - w||_1 \leq 10^{-5}$. The results for both models are summarized in Table 5. In this table, the gap in the last column is $\dfrac{|f - \widehat{f}|}{|\widehat{f}| + 1}$ where f and $\widehat{f}$ are the objective values of models (7) and (9), respectively. These results show that except for two K values, model (9) with the PADM Algorithm is solved much faster than model (7) with gaps that are small, specially for larger values of K. Since PADM always finds a partial minimum that is close to the optimal solution and as gaps show, we conclude that the results of PADM is competitive to the results of solving model (7) by Mosek in CVX.

Table 4. Comparison of the CPU times of models (6) and (7) for NASDAQ and S&P indexes data with $\lambda = 0.5$.

Indexes	Desired number of stocks	Model (6)	Model (7)
S&P (N=168)	K=10	$> 5e + 03$	53.169
	K=20	$> 5e + 03$	87.047
	K=50	$> 5e + 03$	982.866
	K=70	$> 5e + 03$	382.515
	K=100	$> 5e + 03$	25.359
NASDAQ (N=200)	K=10	$> 5e + 03$	683.402
	K=20	$> 5e + 03$	809.982
	K=50	$> 5e + 03$	2.0016e+03
	K=70	$> 5e + 03$	3.0019e+03
	K=100	$> 5e + 03$	985.905

6. Conclusions

This paper investigated MAD model with short-selling, risk-neutral interest rate, and cardinality constraints. Several improved variants of it are proposed. Numerical results on real data of the S&P 500 index, Information Technology, indicated that using short-selling and risk-neutral interest rate led to better Sharpe ratios compared to the classical one. Also, the results on the data of NASDAQ and S&P indexes showed that the improved models are solved significantly faster than the classical ones. To solve models with cardinality constraints, we applied the PADM that is much faster than Mosek in CVX while having lower gaps in almost all instances. One may consider

Table 5. Comparison of objective functions and CPU times of models (7) and (9) with the data of NASDAQ index data, with different number of stocks and desired number of stocks when $x^u = -x^l = 0.3$ and penalty parameter equals to 10^{-1}.

N	K	Objective values		CPU times		
		Model (7)	Model (9)	Model (7)	Model (9)	Gap %
$N = 168$	K=10	-0.0049	-0.0018	63.869	16.489	0.31
	K=20	-0.0232	-0.0138	201.1172	7.094	0.92
	K=50	-0.0648	-0.0548	1.0064e+03	7.239	0.94
	K=70	-0.0816	-0.0701	1.0012e+03	9.349	1.07
	K=100	-0.0930	-0.0914	28.645	6.759	0.15
	K=150	-0.0931	-0.0931	**2.113**	**4.033**	1.3417e-12
$N = 200$	K=10	-0.0061	-0.0023	683.402	16.451	0.83
	K=20	-0.0236	-0.0131	809.982	10.036	1.02
	K=50	-0.0672	-0.0581	2.0016e+03	8.041	0.86
	K=70	-0.0877	-0.0771	3.0019e+03	8.666	0.97
	K=100	-0.1080	-0.0973	985.905	6.072	0.97
	K=150	-0.1132	-0.1132	**1.895**	**2.246**	1.6986e-11
$N = 500$	K=10	-0.0139	-0.0001	1.5105e+03	19.612	1.38
	K=20	-0.0413	-0.0284	$> 5e+03$	8.745	1.24
	K=50	-0.1088	-0.0913	$> 5e+03$	9.095	1.58
	K=70	-0.1463	-0.1306	$> 5e+03$	9.079	1.37
	K=100	-0.1940	-0.1812	$> 5e+03$	12.676	1.07
	K=150	-0.2601	-0.2425	$> 5e+03$	13.615	1.40
	K=200	-0.3057	-0.2894	21.148	6.497	1.25
	K=300	-0.3495	-0.3495	5.426	2.031	4.9281e-14
$N = 1000$	K=10	-0.0152	-0.0030	$> 5e+03$	14.894	1.20
	K=20	-0.0470	-0.0259	$> 5e+03$	14.724	2.02
	K=50	-0.1290	-0.1017	$> 5e+03$	11.812	2.42
	K=70	-0.1774	-0.1480	$> 5e+03$	93.978	2.50
	K=100	-0.2428	-0.2118	$> 5e+03$	10.899	2.50
	K=150	-0.3360	-0.3148	$> 5e+03$	16.051	1.59
	K=200	-0.4131	-0.4017	21.373	8.078	0.81
	K=300	-0.5443	-0.5345	21.288	7.412	0.63

applying other algorithms such as harmony search and particle swarm optimization to the enhanced model and evaluate their performance [1].

REFERENCES

1. N. M. Bala, and S. bin Safei, *A Hybrid Harmony Search and Particle Swarm Optimization Algorithm (HSPSO) for Testing Non-functional Properties in Software System*, Statistics, Optimization & Information Computing, 2021.
2. A. Beck, and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences, vol. 2, no. 1, pp. 183–202, 2009.
3. L.T.H. An and P.D. Tao, *Large-scale molecular optimization from distance matrices by a DC optimization approach*, SIAM Journal on Optimization, vol. 14, pp. 77–114, 2003.
4. E. Angelelli, R. Mansini, and M.G. Speranza, *A comparison of MAD and CVaR models with real features*, Journal of Banking & Finance, 32, pp. 1188–1197, 2008.
5. M.R.T. Baghdadabad, *An empirical analysis of funds' alternative measures in the mean absolute deviation (MAD) framework*, International Journal of Emerging Markets, 10, p. 726, 2015.
6. J.R. Birge and R.Q. Zhang, *Risk-neutral option pricing methods for adjusting constrained cash flows*, The Engineering Economist, 44, pp. 36–49, 1999.
7. A. Bnouhachem, A. Hamdi, and M. Xu, *A new LQP alternating direction method for solving variational inequality problems with separable structure*, Optimization, 65, pp. 2251–2267, 2016.
8. J.P. Burgard, C.M. Costa,and M. Schmidt, *Decomposition methods for robustified k-means clustering problems: If less conservative does not mean less bad*, Tech. Rep., 2020.
9. P. Byrne, S. Lee, *et al.* *Sector, region or function a MAD reassessment of real estate diversification in great britain*, Journal of Property Investment & Finance, 29, pp. 167–189, 2011.

10. C.T.Chang, *A modified goal programming approach for the mean-absolute deviation portfolio optimization model*, Applied Mathematics and Computation,171, pp. 567–572, 2005.
11. X. Chang, S. Liu, and X. Li, *Modified alternating direction method of multipliers for convex quadratic semidefinite programming*, Neurocomputing,214, pp. 575–586, 2016.
12. C.M. Costa, D. Kreber, and M. Schmidt, *An alternating method for cardinality-constrained optimization: A computational study for the best subset selection and sparse portfolio problem* .
13. Z. Dai and F. Wen, *A generalized approach to sparse and stable portfolio optimization problem*, The Journal of Industrial & Management Optimization, 14, pp. 1651–1666, 2018.
14. C.D. Feinstein and M.N. Thapa, *A reformulation of a mean-absolute deviation portfolio optimization model*, Management Science, 39, pp.1552–1553, 1993.
15. J. Gao and D. Li, *Optimal cardinality constrained portfolio selection*, Operations Research, 61, pp. 745–761, 2013.
16. B. Geißler, A. Morsi, L. Schewe, and M. Schmidt, *Solving power-constrained gas transportation problems using an mip-based alternating direction method*, Computers & Chemical Engineering, 82, pp. 303–317, 2015.
17. B. Geißler, A. Morsi, L. Schewe, and M. Schmidt, *Penalty alternating direction methods for mixed-integer optimization: A new view on feasibility pumps*, SIAM Journal on Optimization, 27, pp. 1611–1636, 2017.
18. J. Gorski, F. Pfeuffer, and K. Klamroth, *Biconvex sets and optimization with biconvex functions: a survey and extensions*, Mathematical Methods of Operations Research, 66, pp. 373–407, 2007.
19. M. Grant, S. Boyd, and Y. Ye, *Cvx: Matlab software for disciplined convex programming, version 2.0 beta*, 2013.
20. A. Hamdi, *Decomposition for structured convex programs with smooth multiplier methods*, Applied Mathematics and Computation, 169, pp. 218–241, 2005.
21. B.I. Jacobs, K.N. Levy, and H.M. Markowitz, *Portfolio optimization with factors, scenarios, and realistic short positions*, Operations Research, 53, pp. 586–599, 2005.
22. B.I. Jacobs, K.N. Levy, and H.M. Markowitz, *Trimability and fast optimization of long-short portfolios*, Financial Analysts Journal, 62, pp. 36–46, 2006.
23. C.E. Kalfin Sukono, *Optimization of the mean-absolute deviation portfolio investment in some mining stocks using the singular covariance matrix method*, in *Journal of Physics: Conference Series*, Vol. 1315. IOP Publishing, p. 012002, 2019.
24. S.A. Karbasy and M. Salahi, *A hybrid algorithm for the two-trust-region subproblem*, Computational and Applied Mathematics, 38, pp. 1–19, 2019.
25. I.G. Kawaller, *A note: Debunking the myth of the risk-free return*, The Journal of Futures Markets, 7, pp. 327–331, 1987.
26. T. Khodamoradi, M. Salahi, and A.R. Najafi, *Cardinality-constrained portfolio optimization with short selling and risk-neutral interest rate*, Decisions in Economics and Finance, pp. 1–18, 2020.
27. T. Khodamoradi, M. Salahi, and A.R. Najafi, *A note on CCMV portfolio optimization model with short selling and risk-neutral interest rate*, Statistics, Optimization & Information Computing, 8, pp. 0–9, 2020.
28. T. Kleinert and M. Schmidt, *Computing feasible points of bilevel problems with a penalty alternating direction method*, INFORMS Journal on Computing, 2020.
29. H. Konno, K. Akishino, and R. Yamamoto, *Optimization of a long-short portfolio under nonconvex transaction cost*, Computational Optimization and Applications, 32, pp. 115–132, 2005.
30. H. Konno and H. Yamazaki, *Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market*, Management Science, 37, pp. 519–531, 1991.
31. H.A. Le Thi and M. Moeini, *Long-short portfolio optimization under cardinality constraints by difference of convex functions algorithm*, Journal of Optimization Theory and Applications, 161, pp. 199–224, 2014.
32. P. Li, Y. Han, and Y. Xia, *Portfolio optimization using asymmetry robust mean absolute deviation model*, Finance Research Letters, 18, pp. 353–362, 2016.
33. J. Lintner, *Security prices, risk, and maximal gains from diversification*, The Journal of Finance, 20, pp. 587–615, 1965.
34. R. Mansini and M.G. Speranza, *Linear and Mixed Integer Programming for Portfolio Optimization*, Springer, 2015.
35. D. Maringer, *Portfolio Management With Heuristic Optimization*, Springer Science & Business Media, 2006.
36. H. Markowitz, *Portfolio selection*, The Journal of Finance, 7, pp. 77–91, 1952.
37. M.A. Noor, K.I. Noor, A. Hamdi, and E.H. El-Shemas, *On difference of two monotone operators*, Optimization Letters, 3, pp. 329–335, 2009.
38. L. Schewe, M. Schmidt, and D. Weninger, *A decomposition heuristic for mixed-integer supply chain problems*, Operations Research Letters, 48, pp. 225–232, 2020.
39. L.P.d. Silva, D. Alem, and F.L.d. Carvalho, *Portfolio optimization using mean absolute deviation (MAD) and conditional value-at-risk (CVaR)*, Production, 27, pp. 0–0, 2017.
40. Y. Simaan, *Estimation risk in portfolio selection: the mean variance model versus the mean absolute deviation model*, Management Science, 43, pp. 1437–1446, 1997.
41. R. Taleghani and M. Salahi, *An ADMM-factorization algorithm for low rank matrix completion.*, Applications & Applied Mathematics, 14, pp. 1145–1156, 2019.
42. P.D. Tao, *Convex analysis approach to DC programming: theory, algorithms and applications*, Acta Mathematica Vietnamica, 22, pp. 289–355, 1997.
43. P.D. Tao, *et al.*, *The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems*, Annals of Operations Research, 133, pp. 23–46, 2005.
44. Y. Teng, L. Yang, B. Yu, and X. Song, *A penalty PALM method for sparse portfolio selection problems*, Optimization Methods and Software, 32, pp. 126–147, 2017.