SPARSE MARKOWITZ PORTFOLIO SELECTION BY USING
STOCHASTIC LINEAR COMPLEMENTARITY APPROACH

QIYU WANG
Department of Applied Mathematics, The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong

HAILIN SUN∗
School of Economics and Management, Nanjing University of Science and Technology
Nanjing, 210094, China

(Communicated by Duan Li)

ABSTRACT. We consider the framework of the classical Markowitz mean-variance (MV) model when multiple solutions exist, among which the sparse solutions are stable and cost-efficient. We study a two-phase stochastic linear complementarity approach. This approach stabilizes the optimization problem, finds the sparse asset allocation that saves the transaction cost, and results in the solution set of the Markowitz problem. We apply the sample average approximation (SAA) method to the two-phase optimization approach and give detailed convergence analysis. We implement this methodology on the data sets of Standard and Poor 500 index (S&P 500), real data of Hong Kong and China market stocks (HKCHN) and Fama & French 48 industry sectors (FF48). With mock investment in training data, we construct portfolios, test them in the out-of-sample data and find their Sharpe ratios outperform the $\ell_1$ penalty regularized portfolios, $\ell_p$ penalty regularized portfolios, cardinality constrained portfolios, and $1/N$ investment strategy. Moreover, we show the advantage of our approach in the risk management by using the criteria of standard deviation (STD), Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR).

1. Introduction. Markowitz[19] ushered the era of modern portfolio management by his mean-variance model. In this model, he initiated a way for portfolio selection that balances the variance and return by mean-variance framework. Specifically, for $m$ securities, the traditional Markowitz portfolio is to find a portfolio that has minimal variance for a given expected return $\rho$ by solving the following problem:

$$\begin{align*}
\min & \quad w^T Cw \\
\text{s.t.} & \quad w^T \mu = \rho \\
& \quad w^T 1_m = 1,
\end{align*}$$

(1)

where $w = (w_1, w_2, ..., w_m)^T$, $1_m$ is an $m$-dimensional vector with all entries being one, random variable $r_i : \Xi \rightarrow \mathbb{R}$ is the return of the $i$th security, $\mu_i = \mathbb{E}[r_i(\xi)]$ is its expected return, the return of the securities $r(\xi) = (r_1(\xi), r_2(\xi), ..., r_m(\xi))^T$, the

2010 Mathematics Subject Classification. Primary: 90C15, 90C90; Secondary: 91G10.

Key words and phrases. Stochastic programming, sparse linear complementarity problem, sparse Markowitz portfolio selection, sample average approximation, convergence analysis.

* Corresponding author.
expected return of the securities $\mu = (\mu_1, \mu_2, \ldots, \mu_m)^T$, the covariance matrix of the returns $C = E[(r(\xi) - \mu)(r(\xi) - \mu)^T]$. Note that $C$ is an $m \times m$ positive semi-definite matrix.

In practice, it is challenging to select portfolios computed from the Markowitz MV model. Firstly, when the assets in the investment universe are large-scale, the transaction cost and the trading complexity are considerable. Secondly, when the distribution of the stock returns is non-normal, a problem of overfitting arises in the estimation of unknown parameters, see $[21]$. Last but not least, the Markowitz portfolio problem is unstable, i.e., small changes in the data would result large changes in solution, as reported in $[6]$. 

Due to these facts, there is an ensemble of recently developed sparse Markowitz portfolio theories in the literature as follows.

1. Regularized method: In $[6]$, $[9]$, $[11]$, $[14]$, $[29]$ and $[31]$, they proposed approaches which use penalty to regularize the Markowitz portfolio optimization such as $\ell_1$ penalty.

   $\min w^T Cw + \lambda \| w \|_1 \\
   \text{s.t.} \quad w^T 1_m = 1, \quad w^T \mu = \rho$, \quad (2)

   or in $[9]$, for $p \in (0, 1)$,

   $\min w^T Cw + \lambda \| w \|_p^p \\
   \text{s.t.} \quad w^T 1_m = 1, \quad w^T \mu = \rho$. \quad (3)

2. Cardinality constrained portfolio selection (CCPS): Many researchers have investigated the CCPS by improving the branch-and-bound method, in $[1]$, $[3]$, developing a branch-and-bound method based on a novel geometric approach in $[18]$, constructing a tight semidefinite program (SDP) problem approach in $[33]$. Several other scholars studied the CCPS by heuristic approaches in $[8]$, $[16]$, a positive programming approach in $[27]$, and a nonmonotone projected gradient (NPG) method in $[28]$. The core of the model is to add a cardinality constraint in the Markowitz portfolio selection model to select a cardinality number of assets in the portfolio,

   $\min w^T Cw \\
   \text{s.t.} \quad w^T 1_m = 1, \quad w^T \mu = \rho \\
   \| w \|_0 \leq k$.

Although these several works are used to enhance the Markowitz portfolio selection, they have problems, such as the change of the objective function, imposition of additional constraints. Then there is an idea arising from those frameworks. How about finding a sparse solution from the optimal solution set of the classic Markowitz MV model? When the covariance matrix is semi-definite, i.e., when there exist multiple solutions of the Markowitz portfolio optimization problem, we study a novel method to select a sparse portfolio solution of the classic Markowitz MV model.

Our approach is traced back to inspirations from $[7]$. Chen & Xiang $[12]$ gave a way to find sparse solutions of the linear complementarity problem (LCP). By this method, we study a two-phase stochastic linear complementarity approach (two-phase approach) to find an approximate sparse solution of MV model by solving two simple convex problems in two phases. Note that finding a sparse portfolio over the solution set of MV model is an NP-hard problem, see $[23]$. One of the advantages
of the method is that we can approximate the NP-hard problem by solving a convex quadratic optimization problem and a linear optimization problem. Moreover, due to the discontinuity of $\ell_0$ norm, it may not be easy to know how to guarantee the convergence when the SAA method is applied to the problem. Therefore, another contribution of this paper is that we establish the convergence analysis when we apply the SAA method with the two-phase approach to find a sparse portfolio.

We use the two-phase approach to compute the sparse Markowitz portfolio of a randomly generated example, and three empirical examples, which are S&P 500 portfolios, HKCHN cross market portfolios and FF48 portfolios. We test their out-of-sample performance. From preliminary results, our approach has better performance compared with the reported well-performed $\ell_1$ penalty regularized portfolio, $\ell_p$ penalty regularized portfolio, cardinality constrained portfolio and $1/N$ investment strategy [15].

Our approach outperforms traditional Markowitz portfolios in the sense that: the sparse portfolio saves transaction fees and reduces investment complexity. Normally, the transaction cost is positive linearly related to the number of assets allocated. The sparse solutions from our approach naturally produce a cost efficient investment.

Moreover, our approach outperforms other sparse portfolios in the sense that the sparse solution of the two-phase approach is located in the solution set of the classical Markowitz portfolio optimization which facilitates risk management. The merit of variance minimization of the Markowitz portfolio is inherited by the sparse solution from the two-phase approach in the sense that the sparse solution is exactly a solution of the classic Markowitz portfolio optimization problem.

From the results of our empirical examples, our sparse portfolio could outperform the comparative portfolios from the perspective of Sharpe ratio. Therefore, by our approach, we find a sparse and stable Markowitz portfolio such that the portfolio is well-posed to achieve the goal of classical MV model and preserves the stability and sparsity properties.

Throughout this paper, we use the following notation. $x^Ty$ denotes the scalar products of two vectors $x$ and $y$, $\|\cdot\|_1$ and $\|\cdot\|_0$ denote the $\ell_1$ norm and $\ell_0$ norm of a vector respectively. $d(x, S) := \inf_{x' \in S} \|x - x'\|$ denotes the distance from point $x$ to set $S$. For two sets $S_1$ and $S_2$, $\mathbb{D}(S_1, S_2) := \sup_{x \in S_1} d(x, S_2)$ denotes the deviation of set $S_1$ from set $S_2$ and
\[
\mathbb{H}(S_1, S_2) := \max(\mathbb{D}(S_1, S_2), \mathbb{D}(S_2, S_1))
\]
denotes the Hausdorff distance between two sets $S_1$ and $S_2$.

This paper is organized as follows. In Section 2, we introduce the two-phase approach to find a sparse Markowitz portfolio. In Section 3, we analyze the convergence of the SAA method with the two-phase approach. In Section 4, we show the numerical and empirical applications of the model. Section 5 is a conclusion.

2. Model construction and two-phase stochastic linear complementarity approach. We consider the sparse Markowitz portfolio optimization problem as follows:
\[
\begin{align*}
\min & \quad \| w \|_0 \\
\text{s.t.} & \quad w \in \mathcal{H},
\end{align*}
\]
where $\mathcal{H}$ is the optimal solution set of the Markowitz portfolio optimization problem

$$
\begin{align*}
\min_{w} & \quad w^T C w \\
\text{s.t.} & \quad w^T \mu = \rho \\
& \quad w^T 1_m = 1 \\
& \quad \|w\|_1 \leq \eta,
\end{align*}
$$

where $\eta \geq 1$.

Compared with problem (1), problem (5) has an additional constraint to limit the arbitrage $\|w\|_1 \leq \eta$. Short selling, which is defined as “the sale of a security that the seller does not own or that the seller owns but does not deliver” by the SEC, is considered crucial for effective arbitrage. In fact, short sale is often constrained. In other words, arbitrage is limited, see implications in [14], [17], [26]. Moreover, when $\eta = 1$, problem (5) is equivalent to the classical Markowitz portfolio model without short-selling:

$$
\begin{align*}
\min_{w} & \quad w^T C w \\
\text{s.t.} & \quad w^T \mu = \rho \\
& \quad w^T 1_m = 1, \\
& \quad w \geq 0,
\end{align*}
$$

and when $\eta = \infty$, problem (5) is equivalent to (1).

Problem (4) is a nonconvex noncontinuous optimization problem and may not be easy to solve. Inspired by the method in [12], we provide the two-phase stochastic linear complementarity approach to approximate problem (4) in the following subsection.

2.1. Two-phase stochastic linear complementarity approach. The basic idea of this approach is that we reformulate the feasible set of problem (4) (solution set of problem (5)) as a linear system and find its least $\ell_1$ norm to approximate its sparse solution. To this end, we consider the reformulation of problem (5)

$$
\begin{align*}
\min & \quad (w^+ - w^-)^T C (w^+ - w^-) \\
\text{s.t.} & \quad (w^+ - w^-)^T \mu = \rho \\
& \quad (w^+ - w^-)^T 1_m = 1, \\
& \quad w^+ \geq 0, w^- \geq 0 \\
& \quad 1^T 2_m (w^+, w^-) \leq \eta.
\end{align*}
$$

Firstly, we will show the equivalence between problem (5) and problem (7). Secondly, we will study how to find a sparse solution from the optimal solution set of problem (7).

**Proposition 1.** Problem (5) and problem (7) are equivalent in the sense that (i) If $w^*$ is an optimal solution of (5), then $((w^+)^*,(w^-)^*)$ is a solution of problem (7), where $(w^+)^* = \max(w^*, 0)$, $(w^-)^* = \max(-w^*, 0)$; (ii) if $((w^+)^*,(w^-)^*)$ is an optimal solution of (7), then $w^* = (w^+)^* - (w^-)^*$ is an optimal solution of problem (5); (iii) the least $\ell_1$ norm solution and sparse solution of problem (7) must be the least $\ell_1$ norm solution and sparse solution of problem (5) and vice versa.

**Proof.** (i) Let $w^*$ be an optimal solution of problem (5). Then $(\hat{w}^+, \hat{w}^-)$ is a feasible solution of (7), where $\hat{w}^+ = \max(w^*, 0)$ and $\hat{w}^- = \max(-w^*, 0)$. Let us prove that $(\hat{w}^+, \hat{w}^-)$ is an optimal solution of (7). Assume for the sake of a contradiction that $(\tilde{w}^+, \tilde{w}^-)$ is not an optimal solution of problem (7), then there exists $(\tilde{w}^+, \tilde{w}^-)$ such that

$$
(\tilde{w}^+ - \tilde{w}^-)^T C (\tilde{w}^+ - \tilde{w}^-) < (\hat{w}^+ - \hat{w}^-)^T C (\hat{w}^+ - \hat{w}^-).
$$

(8)
Let $\bar{w} = (\bar{w}^+ - \bar{w}^-)$. Since $\|\bar{w}\|_1 \leq \|\bar{w}^+\|_1 + \|\bar{w}^-\|_1 = 1\|_{2m}(\bar{w}^+, \bar{w}^-) \leq \eta$, $\bar{w}$ is a feasible solution of problem (5), then (8) implies

$$\bar{w}^T C\bar{w} < (w^*)^T Cw^*,$$

a contradiction that $w^*$ is an optimal solution of problem (5).

(ii) Let $v_1$ and $v$ be the optimal values of problem (5) and problem (7) respectively. It is obvious that for any feasible solution $\bar{w}$ of problem (5), $(\bar{w}^+, \bar{w}^-)$ is a feasible solution of problem (7) with the same objective function value, that is

$$\bar{w}^T C(\bar{w}^+ - \bar{w}^-) = \bar{w}^T C\bar{w},$$

where $\bar{w}^+ = \max(\bar{w}, 0)$ and $\bar{w}^- = \max(-\bar{w}, 0)$. Then $v_1 \geq v$. Let $((w^+)^*, (w^-)^*)$ be an optimal solution of problem (7), $w^* = (w^+)^* - (w^-)^*$. Note that $w^*$ is a feasible solution of problem (5). Let $w'$ be an optimal solution of problem (5). Then

$$v \leq v_1 = w^T Cw' \leq (w^*)^T Cw^* = ((w^+)^* - (w^-)^*)^T C((w^+)^* - (w^-)^*) = v,$$

which means $(w^*)^T Cw^* = w^T Cw'$ and $w^*$ is an optimal solution of problem (5).

(iii) By (i) and (ii), we have that for any optimal solution $w^*$ of (5), any $(w^+, w^-)$ such that $1\|_{2m}(w^+, w^-) \leq \eta$, $w^+ - w^- = w^*$ and $w^+ \geq 0$, $w^- \geq 0$ is the optimal solution of (7). Then let $\ell^*$ be the least $\ell_1$ norm solution of (5). Then we claim $(w^*_{\ell^*}, w^-_{\ell^*})$ is a least $\ell_1$ norm solution of (7) where $w^*_{\ell^*} = \max\{w^*_{\ell^*}, 0\}$ and $w^-_{\ell^*} = \max\{-w^*_{\ell^*}, 0\}$. Otherwise, there exists an optimal solution $(w^+, w^-)$ of (7) such that

$$1:\|w^+ - w^-\| \leq 1\|_{2m}(w^+, w^-) < 1\|_{2m}(w^*_{\ell^*}, w^-_{\ell^*}) = 1:\|_{m}w^*_{\ell^*},$$

which contradicts the fact that $w^*_{\ell^*}$ is the least $\ell_1$ norm solution of problem (5).

Moreover, let $(w^+_{\ell^*}, w^-_{\ell^*})$ be the least $\ell_1$ norm solution of (7), it is obvious that $(w^+_{\ell^*})^T w^-_{\ell^*} = 0$. Otherwise, $1\|_{2m}(w^+_{\ell^*}, w^-_{\ell^*}) > 1\|_{2m}(\bar{w}^+_{\ell^*}, \bar{w}^-_{\ell^*})$, where $\bar{w}^+_{\ell^*} = \max(w^+ - w^-_{\ell^*}, 0)$, $\bar{w}^-_{\ell^*} = \max(w^+_{\ell^*} - w^-_{\ell^*}, 0)$ and $(\bar{w}^+_{\ell^*}, \bar{w}^-_{\ell^*})$ is a feasible solution of (7), which contradicts the fact that $(w^+_{\ell^*}, w^-_{\ell^*})$ is the least $\ell_1$ norm solution of problem (7).

Moreover, $w^*_{\ell^*} = (w^+_{\ell^*} - w^-_{\ell^*})$ is the least $\ell_1$ norm solution of (5). Otherwise, there exists an optimal solution of $\bar{w}_{\ell^*}$ of (5) such that

$$1\|_{2m}(\bar{w}^+_{\ell^*}, 0), \max(-\bar{w}^-_{\ell^*}, 0)) = 1:\|_{m}\bar{w}^*_{\ell^*}, \|w^*_{\ell^*}\|_{m} = 1\|_{m}(w^*_{\ell^*}, w^-_{\ell^*}),$$

which contradicts the fact that $(w^*_{\ell^*}, w^-_{\ell^*})$ is the $\ell_1$ norm solution of problem (7).

The proof for sparse solution is similar to the least $\ell_1$ norm solution, we omit the details. \hfill \Box

It is obvious that problem (7) can be rewritten as:

$$\min_{w^+} \begin{pmatrix} (w^+)^T, (w^-)^T \end{pmatrix} H((w^+)^T, (w^-)^T)^T$$

s.t. $G \begin{pmatrix} (w^+) \\ (w^-) \end{pmatrix} \geq b$

$$w^+ \geq 0, w^- \geq 0,$$

where $H = \begin{pmatrix} C & -C \\ -C & C \end{pmatrix}$, $G = \begin{pmatrix} \mu^T \\ \eta \\ -\mu^T \\ \mu^T \\ \eta \end{pmatrix}$, $b = \begin{pmatrix} \rho \\ 1 \\ -\rho \\ 1 \\ -1 \end{pmatrix}$. Moreover, problem (9) can be rewritten as an LCP:

$$x^T(Mx + q) = 0, Mx + q \geq 0, x \geq 0,$$
where \( M = \begin{pmatrix} H & -G^T \\ G & O_{5 \times 5} \end{pmatrix} \), \( q = \begin{pmatrix} O_{2m \times 1} \\ -b \end{pmatrix} \), \( x = \begin{pmatrix} w^+ \\ w^- \end{pmatrix} \), \( y = (y_1, y_2, y_3, y_4, y_5)^T \).

We use LCP(\( q, M \)) to denote LCP(\( 10 \)) with the related \( q \) and \( M \).

Hereafter, for simplicity, we use \( x = (w^+, w^-, y) \) to denote \( x = ((w^+)^T, (w^-)^T, y^T)^T \). Problem (9) and problem (10) are equivalent in the sense that (i) if \( ((w^+)^*, (w^-)^*) \) is a solution of (9) then there is \( y^* \in R^5 \) such that \( ((w^+)^*, (w^-)^*, y^*) \) is a solution of (10); (ii) if \( ((w^+)^*, (w^-)^*, y^*) \) is a solution of (10) then \( w^* \) is a solution of (9).

By Theorem 3.1.7 in [13], since \( M \) is a positive semi-definite matrix, the solution set of the LCP(\( q, M \)), denoted by the SOL(\( q, M \)), can be written as:

\[
\text{SOL}(q, M) = \{ x \in R^{2m+4} | Mx + q \geq 0, H \begin{pmatrix} w^+ \\ w^- \end{pmatrix} = \gamma, q^Tx = \gamma \} \tag{11}
\]

where \( \gamma = H \begin{pmatrix} (w^+)^* \\ (w^-)^* \end{pmatrix}, \) \( x^* = ((w^+)^*, (w^-)^*, y^*) \) is an arbitrary solution of the LCP(\( q, M \)).

By a given solution \( x^* \) of the LCP(\( q, M \)), all the solutions \( (w^+, w^-) \) of the Markowitz portfolio optimization with corresponding Lagrange multipliers \( y \) are given in the SOL(\( q, M \)). Then we try to look for a sparse solution in the solution set SOL(\( q, M \)). We can find the sparse Markowitz portfolio of problem (4) by solving

\[
\begin{array}{ll}
\min & \| (w^+, w^-) \|_0 \\
\text{s.t.} & Mx + q \geq 0 \\
& x \geq 0 \\
& H \begin{pmatrix} w^+ \\ w^- \end{pmatrix} = \gamma \\
& q^Tx = \gamma,
\end{array} \tag{12}
\]

where \( \gamma = H \begin{pmatrix} (w^+)^* \\ (w^-)^* \end{pmatrix}, \) \( x^* = ((w^+)^*, (w^-)^*, y^*) \) is an arbitrary solution of the LCP(\( q, M \)).

Note that \( \| \cdot \|_1 \) is a good approximation of \( \| \cdot \|_0 \) solution, which has been widely used such as [6]. Then we can approximate problem (12) by the following linear programming:

\[
\begin{array}{ll}
\min & (w^+, w^-)^T 1_{2m} \\
\text{s.t.} & Mx + q \geq 0 \\
& x \geq 0 \\
& H \begin{pmatrix} w^+ \\ w^- \end{pmatrix} = \gamma \\
& q^Tx = \gamma,
\end{array} \tag{13}
\]

where \( \gamma = H \begin{pmatrix} (w^+)^* \\ (w^-)^* \end{pmatrix}, \) \( x^* = ((w^+)^*, (w^-)^*, y^*) \) is a solution of the LCP(\( q, M \)).

It will be very interesting to consider the upper bound of the \( \ell_0 \) norm of the classical Markowitz portfolio optimization sparse solutions. We will discuss it in the following proposition.

**Proposition 2.** Consider problem (4), the number of nonzero entries in a solution of problem (4) is at most \( \text{rank}(C) + 3 \), where \( C \) is the covariance matrix.
Proof. \( \mathcal{H} \) denotes the solution set of problem (5). Let \( \vec{w} \) and \((w^+, w^-)\) be the sparse solution of problem (5) and (7). By Proposition 1, \( \|\vec{w}\|_0 = \|(w^+, w^-)\|_0 \). Note that problem (7) can be written as the LCP \((q, M)\). From \( \text{rank}(G)=3 \), it is obvious that \( \text{rank}(M) \leq \text{rank}(H) + 3 = \text{rank}(C) + 3 \), the last equation is from the definition of \( H \). Then by [12, Theorem 2.1] and Remark 1, we have
\[
\|\vec{w}\|_0 = \|(w^+, w^-)\|_0 \leq \text{rank}(C) + 3.
\]

\[ \Box \]

2.2. Equivalence between the least \( \ell_1 \) norm solution and the sparse solution. Under some conditions, we can show that a solution of problem (12) can be found exactly by solving problem (13) if we replace \((w^+, w^-)\) by \( x \) in the objective functions for both two problems (In this subsection, we consider problem (12) and (13) with objective function \( \|x\|_0 \) and \( \|x\|_1 \) respectively). To this end, we need the following definition from [7]. Here \( |t| \) is the number of elements of the set \( t \).

Definition 2.1. An \( m \times n \) matrix \( P \) is said to satisfy the \( s \)-restricted isometry property (RIP) with a restricted isometry constant \( \delta_s \), if for every \( m \times |t| \) sub-matrix \( P_t \) of \( P \) and for every vector \( z \in R^{|t|} \) with \( |t| \leq s \),
\[
(1 - \delta_s)\|z\|_2^2 \leq \|P_t z\|_2^2 \leq (1 + \delta_s)\|z\|_2^2.
\]

\( P \) is said to satisfy the \( s, s' \)-restricted orthogonality (RO) with a restricted orthogonality constant \( \theta_{s,s'} \) for \( s + s' \leq n \) if for all sub-matrices, \( P_t \in R^{m \times |t|} \) and \( P_{t'} \in R^{m \times |t'|} \) of \( P \), with \( |t| \leq s \) and \( |t'| \leq s' \), for all the vectors \( z \in R^{|t|} \) and \( z' \in R^{|t'|} \),
\[
|(P_t z, P_{t'} z')| \leq \theta_{s,s'}\|z\|_2\|z'\|_2
\]
holds for all the disjoint sets \( t \) and \( t' \).

The following theorem is from [12].

Theorem 2.2. Let \( \hat{x} \) be the optimal solution of linear programming problem (13) with \( \|\hat{x}\|_0 \leq s \). Then
1. if \( H \) satisfies the RIP with a restricted isometry constant \( \delta_{2s} < 1 \), then \( \hat{x} \) is the unique sparse solution of the LCP \((q, M)\);
2. if \( H \) satisfies the RIP and RO with
\[
\delta_s + \theta_{s,s'} + \theta_{s,2s'} < 1,
\]
then \( \hat{x} \) is the unique solution of (13) and the unique sparse solution of the LCP \((q, M)\).

Theorem 2.2 shows the relationship between problems (12) and (13).

Therefore, we approach an NP-hard problem by solving two simple problems (10) and (13) in two phases.

Remark 1. In this subsection, we considers the \( \|x_0\| \) where \( x = (w^+, w^-, y) \), but we actually want to consider \( \|w\|_0 \). The reason we consider \( \|x\|_0 \) is that, under the conditions of Theorem 2.2, its solution can be obtained by solving problem (13). Moreover, assuming that \( w^* \) is the optimal solution of problem (4) and \( \bar{x} = ((\vec{w})^+, (\vec{w})^-, \bar{y}) \) is the optimal solution of problem (12) with objective function \( \|x\|_0 \), it is obvious that
\[
\|(w^+)^*, (w^-)^*, \bar{y})\|_0 \geq \|(w^+, w^-, \bar{y})\|_0 = \|\bar{x}\|_0 \geq \|(\vec{w}^+, \vec{w}^-)\|_0 \geq \|((w^+)^*, (w^-)^*)\|_0,
\]
for any $\hat{y} \in R^5$. Then $0 < \|\hat{x}\|_0 - \|(w^+)^*,(w^-)^*\|_0 \leq 5$. In case that the dimension of $w$ is very large, problem (12) is a good approximation of problem (4).

But the RIP and RO conditions in Theorem 2.2 are not easy to satisfy. In the rest of paper, we consider problem (12) and (13) with objective function $\|(w^+, w^-)\|_0$ and $\|(w^+, w^-)\|_1$ respectively.

3. The SAA method and convergence analysis. In this section, we use the SAA method to model the randomness of the Markowitz portfolio and consider the convergence analysis between the original problem and the approximation problem.

Let $\{r^j = (r_{1i}^j, \ldots, r_{mi}^j)\}_{j=1}^N$ be the i.i.d. samples of random variable $r$. Then the SAA mechanism of the LCP sparse Markowitz portfolio optimization is:

$$\min \quad (w^+, w^-)^T 1_{2m}$$
$$\text{s.t.} \quad M_N x + q \geq 0$$
$$H_N \left( \begin{array}{c} w^+ \\ w^- \end{array} \right) = \varsigma_N$$
$$q^T x = \gamma_N,$$

where

$$C_N = \frac{1}{N} \sum_{j=1}^N (r^j - \mu_N)(r^j - \mu_N)^T,$$

$$\mu_{i,N} = \frac{1}{N} \sum_{j=1}^N r_{ji}^j, \quad \mu_N = (\mu_{1,N}, \ldots, \mu_{m,N})^T, \quad M_N = \left( \begin{array}{c} H_N \\ G_N \end{array} \right), \quad H_N = \left( \begin{array}{cc} C_N & -C_N \\ -C_N & C_N \end{array} \right), \quad G_N = \left( \begin{array}{cc} \mu_N^T & -\mu_N^T \\ -\mu_N^T & \mu_N^T \\ -1^T_{m} & 1^T_{m} \\ -1^T_{m} & -1^T_{m} \end{array} \right),$$

$$\varsigma_N = H_N ((w^+)^T_N, (w^-)^T_N), \quad y_N = (y_{1,N}, \ldots, y_{5,N})^T,$$

$$\gamma_N = q^T x_N$$

and $x_N = ((w^+)^T_N, (w^-)^T_N, y_N^T)^T$ is a solution of the following LCP

$$x^T M_N x + q^T x = 0, \quad M_N x + q \geq 0, x \geq 0. \quad (15)$$

Note that the LCP$(q, M_N)$ (15) is driven from the first order necessary condition of the following SAA form of problem (9):

$$\min \quad (w^+, w^-)^T H_N ((w^+)^T, (w^-)^T)$$
$$\text{s.t.} \quad G_N \left( \begin{array}{c} w^+ \\ w^- \end{array} \right) \geq b$$
$$w^+ \geq 0, w^- \geq 0. \quad (16)$$

It follows that we consider the convergence analysis between the original problem (13) and its SAA form (14). Although problem (13) is a stochastic linear problem, some of its parameters are solutions of the LCP$(q, M)$ (10). So it is hard to give the convergence analysis between problem (13) and its SAA form directly. We also prove it in two steps. In the first step, we prove the convergence analysis of KKT pairs between problem (9) and its SAA form (16). In the second step, we prove the convergence between the optimal solutions and the optimal values of (13) and (14). We need the following Slater condition.

**Assumption 1.** There exists a feasible point $(w^+_0, w^-_0)$ such that $(w^+_0, w^-_0) > 0$, $1_{2m}(w^+_0, w^-_0) < \eta$, $(w^+_0 - w^-_0)^T \mu = \rho$ and $(w^+_0 - w^-_0)^T 1_m = 1.$
Let \( f(w^+, w^-) = ((w^+)^T, (w^-)^T)H((w^+)^T, (w^-)^T)^T, \)
\[
f_N(w^+, w^-) = ((w^+)^T, (w^-)^T)H_N((w^+)^T, (w^-)^T)^T,
\]
\[D = (G^T, I_{2m \times 2m})^T, \] and \( D_N = (G_N^T, I_{2m \times 2m})^T, \)
and \( I \) is an identity matrix. Then
\[
g(w^+, w^-) = D \left( \begin{array}{c} w^+ \\ w^- \end{array} \right) - (b^T, 0_{1 \times 2m})^T,
\]
and \( g_N(w^+, w^-) = D_N \left( \begin{array}{c} w^+ \\ w^- \end{array} \right) - (b^T, 0_{1 \times 2m})^T,
\]
reformulate the constraints \( G \left( \begin{array}{c} w^+ \\ w^- \end{array} \right) - b, \ G_N \left( \begin{array}{c} w^+ \\ w^- \end{array} \right) - b \) and \((w^+, w^-) \geq 0\).

We consider the convergence analysis between KKT pairs between problem (9) and its SAA form (16). The KKT condition of problem (9) is:
\[
0 = 2H \left( \begin{array}{c} w^+ \\ w^- \end{array} \right) - DT \left( \begin{array}{c} y \\ s \end{array} \right),
\]
\[
0 = \min\{g(w^+, w^-), \left( \begin{array}{c} y \\ s \end{array} \right) \},
\]
and the KKT condition of problem (16) is:
\[
0 = 2H_N \left( \begin{array}{c} w^+ \\ w^- \end{array} \right) - D_N^T \left( \begin{array}{c} y_N \\ s_N \end{array} \right),
\]
\[
0 = \min\{g_N(w^+, w^-), \left( \begin{array}{c} y_N \\ s_N \end{array} \right) \},
\]
where \( y, s, y_N \) and \( s_N \) are Lagrangian multipliers corresponding to the constraints
\( G \left( \begin{array}{c} w^+ \\ w^- \end{array} \right) \geq b, (w^+, w^-) \geq 0, G_N \left( \begin{array}{c} w^+ \\ w^- \end{array} \right) \geq b_N \) and \((w^+, w^-) \geq 0\) respectively.

**Proposition 3.** Suppose Assumption 1 holds. Then there exists a sufficiently large compact set \( \mathcal{C} \subset R^{1m+5} \) such that (i) the intersection of \( \mathcal{C} \) and the set of KKT pairs of the true problem (9), denoted by \( Y^* \), is nonempty; (ii) for \( N \) sufficiently large, the intersection of \( \mathcal{C} \) and the set of KKT pairs of the SAA problem (16), denoted by \( Y_N \), is nonempty; (iii) for every \( \epsilon > 0 \), there exists \( N(\epsilon) > 0 \) such that
\[
\mathcal{H}(Y_N, Y^*) \leq \epsilon
\]
for \( N \geq N(\epsilon) \).

The result is directly implied by Proposition 2.3 in [30]. Note that in Proposition 2.3 of [30], they use the condition named "no nonzero abnormal multipliers constraint qualification (NNAMCQ)" to bound the Lagrangian multipliers. This constraint qualification is well known, see [5] and [32].

Here the NNAMCQ may not hold in problem (9) since we rewrite every equality constraint in (7) as two inequality constraints. Note that problems (7) and (9) are equivalent and \((y_1 - y_3)\) and \((y_2 - y_4)\) are corresponding Lagrangian multipliers of equality constraints, \(y_5\) and \(s\) are corresponding Lagrangian multipliers of inequality constraints in problem (7). Moreover, by Assumption 1, problem (7) satisfies MFCQ, which implies NNAMCQ. Then \((y_1 - y_3, y_2 - y_4, y_5, s)\) and \((y_1 - y_3, y_2 - y_4, y_5, s)\) are uniformly bounded for \( N \) sufficiently large and there exists sufficiently large compact set \( \mathcal{C} \) such that \( Y^* \) and \( Y_N \) are nonempty.
Note that the KKT pairs of problem (9) and problem (16) are the solutions of the \( \text{LCP}(q, M) \) in (10) and the \( \text{LCP}(q, M_N) \) in (15). Proposition 3.1 shows the convergence analysis between the first phase problem of our two-phase method and implies \( \varsigma_N \to \varsigma \) and \( \gamma_N \to \gamma \) as \( N \to \infty \) almost surely.

Then we move to the convergence analysis of the second phase problem (13) and its SAA problem (14). We use the Chapter 6, Proposition 6 and Remark 8 in [4] to show the result. Let \( v^* \) and \( v_N \) denote the optimal value of the true problem (13) and the SAA problem (14) respectively.

**Proposition 4.** Suppose Assumption 1 holds. Then there exists a sufficiently large compact set \( K \in \mathbb{R}^{2m+5} \) such that (i) the intersection of \( K \) and the optimal solution set of problem (13), denoted by \( S \), is nonempty; (ii) for \( N \) sufficiently large the intersection of \( K \) and optimal solution set of problem (14), denoted by \( S_N \), is nonempty; (iii) we have \( v_N \to v^* \) and \( D(S_N, S) \to 0 \) w.p.1 as \( N \to \infty \) w.p.1.

**Proof.** Obviously, the optimal solution set of problem (13) is nonempty and \( v^* \) is finite. Note that the feasible sets of problem (9) and problem (16) are closed and belong to the compact set \( \{w^+, w^- \in \mathbb{R}^m : 1^T \eta (w^+ + w^-) \leq \eta, w^+, w^- \geq 0\} \). Moreover, we have uniform boundness of \( (y_1-y_3, y_2-y_4, y_5) \) and \( (y_{1,N}-y_{3,N}, y_{2,N}-y_{4,N}, y_{5,N}) \) for \( N \) sufficiently large by Proposition 2.2 in [30] and the discussion below Proposition 3, we have \( S \) and \( S_N \) are nonempty for \( N \) sufficiently large.

Moreover, by Proposition 3, we have that \( \varsigma_N \to \varsigma \) and \( \gamma_N \to \gamma \) as \( N \to \infty \) w.p.1. Then by the uniform law of large numbers [4, Chapter 6, Proposition 7], the constraint functions of problem (14) uniformly converge to the constraint functions of problem (13). Note that problem (13) and problem (14) are linear problems, by Proposition 6 and Remark 8 in Chapter 6 of [4], we have \( v_N \to v^* \) and \( D(S_N, S) \to 0 \) w.p.1 as \( N \to \infty \) w.p.1. \( \square \)

4. **Applications.** In this section, we demonstrate the two-phase approach by a randomly generated example and three empirical examples. We use Matlab R2014a, in a computer with Intel Core 2 Due CPU E8500 3.16GHz for the randomly generated example portfolio construction, Hong Kong and China cross market portfolio construction and FF48 portfolio construction; Matlab R2015a in the service machine with the Intel Xeon E7-4890v2 processor, 2.8GHz, 37.5M Cache, 15 Cores per CPU, 4 CPU, 60 Core in total for S&P500 portfolio construction.

In the examples, we compute the first phase optimization problem by “quadprog” and the second phase optimization problem by “linprog”. For the comparative portfolios, we compute the \( \ell_1 \) norm penalty regularized portfolio by “CVX”\(^1\). For portfolios with cardinality constraint, we compute with a Matlab tool box Yalmip “optimize”\(^2\). For the \( \ell_p \) penalty regularized portfolio, we approximate the non-smooth objective function with a smoothing function which has been extensively investigated in [2],[10] and compute it by “fmincon”. The distance of the comparative portfolios to the Markowitz portfolio feasible solution set is also calculated. Together, we also compute and compare the VaR and CVaR of these portfolios.

4.1. **Randomly generated example.** We randomly generate sample returns of nine asset and apply the two-phase method to construct sparse portfolios.

\(^1\)Downloaded from http://cvxr.com/cvx/download/.

\(^2\)Downloaded from http://users.isy.liu.se/johanl/yalmp/. Copyright owned by Johan Lofberg.
Example 1. For nine assets in the investment universe, we generate the multivariate normal random variables with mean \( \mu \) and covariance \( \Sigma \) as the sample of the asset returns. We set every three assets in a group with the same expected return, the same STD, and fully correlated.

Explicitly, our assets have been arranged in three groups. Asset 1, asset 2 and asset 3 in Group a; asset 4, asset 5, and asset 6 in Group b; asset 7, asset 8 and asset 9 in Group c. The mean returns of the nine assets are \( \mu = [10 \ 10 \ 10 \ 1 \ 1 \ 1 \ 3 \ 3 \ 3] \), and the variance of the nine assets are \( \sigma^2 = [5 \ 5 \ 5 \ 1 \ 1 \ 1 \ 2 \ 2 \ 2] \). The correlations for assets in the same group is 1 and correlations of assets inter-groups are \( \rho_{ab} = 0.5 \), \( \rho_{ac} = 0.3 \) and \( \rho_{bc} = 0.8 \).

We calculate and analyze the true optimal value and the approximate optimal value. The true optimal value and the SAA optimal value (STD) for various sample sizes, are displayed respectively in Table 1.

Data missing case: Consider a case where the price data of some trading days is missing for some assets, which is very often encountered in the real practice. Then the covariance matrix of the underlying assets could not be easily constructed. In [22], a quadratically convergent Newton method is introduced to compute the nearest covariance matrix. Using the above sample, we randomly delete some data, then apply the method in [22] to construct the nearest covariance matrix of the portfolio optimization problem. We then use the two-phase approach to solve the reconstructed problem.

The optimal value (variance) at phase 2 in the case of data missing is 6.23 (the equal STD 2.4960) in Table 1 Column 11. The two-phase stochastic approach could be accustomed to the case of asset prices data missing.

Convergence analysis: We also conduct tests for different sample size \( N = 100, 500, 1000, \cdots, 10000 \). For each sample size, we conduct 100 independent exercises. The true optimal value and approximate optimal values with \( N \) increasing are shown in Table 1. The convergence of the optimal values with \( N \) increasing is shown in Figure 1.

| N   | true 500 | 1500 | 3000 | 4500 | 6000 | 7500 | 9000 | 10000 | dmissing |
|-----|----------|------|------|------|------|------|------|-------|----------|
| Val | 2.489    | 2.490| 2.487| 2.487| 2.487| 2.489| 2.488| 2.490  | 2.486    | 2.496    |

Table 1. Convergence analysis of SAA sparse portfolio optimal value (STD) for Example 4.1

4.2. Empirical applications. In this subsection, we conduct empirical tests. The data sets that we use are returns of S&P 500, 50 Hong Kong and mainland China stocks, and FF48.

We use a rolling window procedure for out-of-sample comparison. \( T \) is the total period of returns in the data set; \( \tau \) is the length of the rolling window; \( os \) is the length of the out-of-sample testing window, which is the holding period. The data frequency is daily or monthly, depending on the data set.

Our testing scheme is that we use \( \tau \) returns as the training data, then use the subsequent \( os \) returns as the forecasting data to compute the out-of-sample return, STD, Sharpe ratio and sparsity. We do the exercise from the beginning of our sampling data and roll ahead. For example, our first portfolio selection takes place
at the end of the first $\tau$ trading days or months. We use the $\tau$ historical returns to estimate covariance matrix $C$ and mean $\mu$ by SAA method. We then solve the portfolio optimization problem by using the estimated parameters, targeting the required return and compute the weights of optimal solutions. Once a portfolio is thus determined, it is held for the subsequent $os$ trading days or months from $(\tau + 1)$ to $(\tau + os)$, and its returns are recorded. We repeat the same process, using returns from 2 to $(\tau + 1)$ to construct the portfolio in test 2. The portfolio is observed for the subsequent $os$ trading days or months from $(\tau + 2)$ to $(\tau + os + 1)$ and their returns are recorded.

To reduce the effects of chance factors, we repeat the same exercise always with a rolling window as described above for 20 rolling times (out-of-sample performance tests). For a given period (whether it is the full period, or the subperiods), all the daily or monthly returns corresponding to this period are used to compute the average return and its STD.

The criteria that we pay attention to are portfolio STD, Sharpe ratio. Sharpe ratio, which measures the risk-adjusted return, is a ratio of return and STD in [25],

$$SR = \frac{r_p}{\sigma_p},$$  \hspace{1cm} (17)

where $r_p$ is portfolio return; $\sigma_p$ is portfolio STD. We compute the

$$(\sigma_p)^2 = \frac{1}{20 os} \sum_{t=\tau}^{19+\tau os-1} \sum_{s=0}^{os-1} (w_t^T r_{t+s+1} - r_p)^2,$$  \hspace{1cm} (18)

with

$$r_p = \frac{1}{20 os} \sum_{t=\tau}^{19+\tau os-1} \sum_{s=0}^{os-1} w_t^T r_{t+s+1}. $$  \hspace{1cm} (19)

It measures the trade-off between returns and volatilities of the portfolios.

We also compute each test Sharpe ratio,

$$SR_t = \frac{r_p}{\sigma_p^t}, \hspace{0.5cm} t = \tau, ..., \tau + 19$$  \hspace{1cm} (20)
where $r_t^p$ is portfolio return; $\sigma_t^p$ is portfolio STD at test $t$. We compute the

$$\left(\sigma_t^p\right)^2 = \frac{1}{os} \sum_{s=0}^{os-1} (w_t^T r_{t+s+1} - r_p)^2,$$

(21)

with

$$r_t^p = \frac{1}{os} \sum_{s=0}^{os-1} w_t^T r_{t+s+1}. $$

(22)

We compare the value of risk measure between different portfolios by using VaR and CVaR. VaR is a widely used risk measure to evaluate the market risk.

The definition of VaR [20] is:

$$\text{VaR}_\alpha = \inf \{ l \in \mathbb{R} : P(L \leq l) \geq \alpha \},$$

which says that VaR at confidence level $\alpha$ is given by the smallest number $l$ so that the probability that the loss $L$ exceeds $l$ is at most $(1 - \alpha)$ and the definition of CVaR [24] is:

$$\text{CVaR}_\alpha = \frac{1}{1 - \alpha} \int_{\alpha}^{1} \text{VaR}_\gamma(L) d\gamma.$$

We use the historical simulation method to calculate VaR and CVaR. For the computation of VaR and CVaR, the portfolios are thus held for the subsequent $os'$ days or months after the training period and their daily or monthly out-of-sample returns are observed. The confidence level of the VaR and CVaR is $\alpha = 99\%$.

We also investigate the distance of the $\ell_1$ penalty regularized portfolio, cardinality constrained portfolio solution and $\ell_p$ penalty regularized portfolio to the optimal Markowitz portfolio solution set, which is defined as the shortest distance between the comparative portfolio solution to the optimal solution set:

$$\min \| z - w \|_2 \quad \text{s.t.} \quad z \in \mathcal{H},$$

(23)

where $\mathcal{H}$ is the solution set of the Markowitz portfolio optimization problem and $w$ is a solution of the comparative portfolio. The optimal value of the optimization problem (23) is the distance of the comparative portfolio solution to the optimal solution set.

**Example 2.** S&P 500. The data set contains the returns of S&P 500 index component stocks of big companies by market capitalization listed on the NYSE or NASDAQ.

Our sampling data is from Compustat. Sampling data are daily returns of S&P 500 from January 2001 to August 2001. The required return is set as the average of the training sample returns. Our empirical analysis relies on a “rolling-sample” procedure. In this example, $\tau = 100$, $os = 5$ for STD, Sharpe ratio computation, $os' = 20$, $\alpha = 99\%$ for VaR and CVaR computation.

We likewise carry out computation for our comparative portfolios. The empirical results of the forecasting return, STD, Sharpe ratio, sparsity, Var and CVaR are in Table 2. The sparsity of the MV model is contained in the bracket of the LCP sparse portfolio column. Abbreviations are LCP sparse portfolio (LCPSP), $\ell_1$ penalty regularized portfolio with tuning parameter $\lambda = 0.1$ ($\ell_1 0.1$), cardinality constrained portfolio with cardinality number 100 (CCPS100) and $1/N$ investment strategy.
The one-by-one test results of Sharpe ratio and sparsity are displayed in Figure 2.

| S&P 500 | LCPSP | $\ell_1$ 0.1 | CCPS100 | 1/N |
|---------|-------|-------------|---------|-----|
| return  | 0.001 | 0.000823    | 0.0014  | -0.00003 |
| STD     | 0.0024| 0.002287    | 0.0084  | 0.0074 |
| Sharpe  | 0.3989| 0.359736    | 0.1694  | -0.0045 |
| VaR     | 0.0041| 0.004289    | 0.0115  | 0.0131 |
| CVaR    | 0.0046| 0.004292    | 0.0147  | 0.0131 |
| sparsity| 89(406)| 66.25       | 58.6    | 500  |
| distance| 1.00E-05| 3.50E-07   |         |      |

Table 2. S&P 500 Portfolio return, STD, Sharpe Ratio, sparsity, VaR, CVaR and distance.

**Example 3.** We randomly select 50 stocks which includes 25 Hong Kong stocks and 25 China stocks.

**Data construction:** These stocks include 20 stocks in the Shanghai-Hong Kong Stock Connect Scheme and 30 A&H share stocks. We collect the daily adjusted close price from October 2013 to January 2014 of the 50 stocks and calculate the daily returns of stocks constructed from the collected price data.

These stocks include Hong Kong Stock Exchange and Clearing Ltd., China railway group Ltd., GOME Electrical Appliances Holding Ltd., Digital China Holdings Ltd., Brightoil Petroleum Holdings Ltd., Shandong Weigao Group Medical Polymer Co, Aluminum Corporation of China Ltd., Shanghai Electric Group Co. Ltd., China Everbright Ltd., Shenzhen International Holdings Ltd. listed in Hong Kong market, and China Merchants Bank, SAIC Motor, Gansu Yasheng Industrial, China Space-sat, Kweichow Moutai, Sichuan Roal and Bridge, Daqin Railway, Huaxia Bank, Industrial and Commercial Bank of China, Bank of China listed in the China market. A and H shares include China Vanke Co Ltd, Ping An insurance, China Pacific Insurance, Huaneng Power International Inc, Anhui Conch Cement, Luoyang Glass, China Minsheng Bank, First Tractor, China CITIC Bank, Jingwei Textile, Jiangsu Express, Guangzhou Automobile, Shanghai Pharmaceuticals, Jincheng Machinery Electric Co. Ltd., and CSSC Offshore & Marine Engineering.

We use the same data rolling scheme. The required return is set as the average of the total returns. The training data window length is $\tau = 25$. The holding period data length is $os = 10$. For VaR and CVaR computation, we use $os' = 20$ for observation and set confidence level $\alpha = 99\%$.

The empirical results of the forecasting return, STD, Sharpe ratio, sparsity, Var and CVaR are in Table 3. The sparsity of the MV model is contained in the bracket. The one-by-one rolling results of the tests are displayed in Figures 3(a) and 3(b). Our portfolio outperforms $\ell_1$ portfolio (with tuning parameter 0.1), cardinality constrained portfolio (with the cardinality number 20 and 25), $\ell_p$ portfolio (with $p=0.5$, tuning parameter 0.015) and $1/N$ investment strategy, denoted by $\ell_1$ 0.1, CCPS 20, CCPS 25, $\ell_p$ 0.015, and $1/N$ respectively.

---

*A Chinese company could raise capital by issuing A shares and H shares. In other words, the equity structure of a company could be comprised of A share, H share and other shares. H share: shares of company incorporated in mainland China that are traded on the HKEx.*
After the Shanghai-Hong Kong Stock Connect, there will be soon Shenzhen-Hong Kong Stock Connect. The sparse portfolio allocation research is thus of great importance for cross market investment.

| HKCHN | LCPSP | \(\ell_1\) 0.1 | CCPS 20 | CCPS 25 | \(\ell_p\) 0.015 | 1/N |
|-------|-------|----------|--------|--------|--------------|-----|
| return | 0.001296 | 0.00044 | 0.001348 | 0.001583 | 0.000537 | -0.00171 |
| STD   | 0.007345 | 0.007115 | 0.013617 | 0.012368 | 0.010248 | 0.006009 |
| Sharpe| 0.1764   | 0.061903 | 0.099   | 0.128   | 0.0524     | -0.2841 |
| VaR   | 0.013248 | 0.011514 | 0.023592 | 0.024182 | 0.014756 | 0.010944 |
| CVaR  | 0.01408  | 0.011807 | 0.026692 | 0.024382 | 0.015791 | 0.010944 |
| sparsity | 27(49) | 14 | 19.45 | 23.1 | 24.1 |
| distance | 0.0024 | 0.003406 | 0.00012 | 0.144709 |

**TABLE 3.** Hong Kong and Mainland China Cross Market Portfolio return, STD, Sharpe Ratio, sparsity, VaR and CVaR and distance.

**Example 4.** FF48. It includes 48 industry sector portfolios (abbreviated to FF48).

Our sampling data is from Fama and French. Sampling data are monthly returns of 48 industry sector value-weighted portfolios in percentage from September 2005 to January 2011.

We use the same data rolling scheme. The required return is set as the average of the total returns on each test. The size of the rolling window is \(\tau = 25\). The out-of-sample performance tests used \(os = 10\). For VaR and CVaR computation, we use \(os' = 20\) for observation and set confidence level \(\alpha = 99\%\). Our sample data is rolling ahead monthly. We roll the window and repeat the procedure for 20 tests to generate the average aggregate results.

The empirical results of the forecasting return, STD, Sharpe ratio, sparsity, Var and CVaR are in Table 4. The sparsity of the MV model is contained in the bracket. The one-by-one test results of Sharpe ratio and sparsity are displayed in Figures 4(a) and 4(b). Our portfolio outperforms \(\ell_1\) portfolio (with tuning parameter 0.1), cardinality constrained portfolio (with the cardinality number 18 and 24), \(\ell_p\) portfolio (with \(p=0.5\), tuning parameter 0.015) and 1/N investment strategy, denoted by \(\ell_1\) 0.1, CCPS 18, CCPS 24, \(\ell_p\) 0.015, and 1/N respectively.

| FF48 | LCPSP | \(\ell_1\) 0.1 | CCPS18 | CCPS 24 | \(\ell_p\) 0.015 | 1/N |
|------|-------|----------|--------|--------|--------------|-----|
| return | -0.1201 | -0.13259 | -0.6413 | -0.3736 | -0.3334 | -0.7838 |
| STD   | 5.9265  | 5.917113 | 8.5639 | 7.5703 | 5.3545 | 8.002 |
| Sharpe| -0.0203 | -0.0224 | -0.0749 | -0.0493 | -0.0623 | -0.098 |
| VaR   | 8.5242  | 14.59896 | 12.8905 | 9.8635 | 10.0079 | 7.827 |
| CVaR  | 10.3835 | 14.86594 | 15.514 | 12.8947 | 13.5395 | 10.0896 |
| sparsity | 29(48) | 24.35 | 21.45 | 7.2 | 31.2 | 48 |
| distance | 0.0044 | 0.2232 | 0.5608 | 0.0938 |

**TABLE 4.** FF48 Portfolio return, STD, Sharpe Ratio, VaR, CVaR, sparsity and distance.

\(^4\)Source: Hong Kong Wenhui
4.3. Discussion of the empirical results. In this subsection, we discuss the out-of-sample performance of our approach and our comparing approaches. Tables 2-4 report the out-of-sample performance of the portfolios constructed from the S&P 500, Hong Kong and mainland China cross-market stocks and FF48 data sets. We illustrate each test’s Sharpe ratio performance and sparsity by bar graph in Figures 2-4. In the case when the MV model has multiple solutions, our method combines the advantage of the MV model and its sparsity modification methods (\(\ell_1\) regularization, \(\ell_p\) regularization, cardinality constrained portfolio selection methods and so on).

On the one hand, we can observe from the tables and figures in Section 4.2 that the number of assets selected in the our approach are sparser than that in the classical Markowitz portfolios. That means we could construct sparse Markowitz mean-variance portfolios with reduced transaction cost in this way. Note that the simple structure of sparse portfolios enables the portfolio construction to depend less on the human management. Moreover, the sparsity of the sparse Markowitz portfolio solution could provide a reference for the setting of \(k\) in the cardinality constrained portfolio selection model.

On the other hand, the empirical results of the examples show that the two-phase optimization scheme could find sparse solutions in the optimal solution set of the Markowitz MV model. The out-of-sample results indicate that the LCP sparse portfolio from our approach can keep the property of the classical Markowitz portfolio very well. That is in all the tests, the STDs of LCP sparse portfolio are reasonably small and in most of the tests, its STDs are the smallest. Moreover, not only the performance of the Sharpe ratio, VaR and CVaR are reasonably good, but also the Sharpe ratios in all the periods are stabler than the other portfolios in Figures 2(a), 3(a) and 4(a).

5. Conclusions. We study a two-phase stochastic linear complementarity approach to seek for a sparse asset allocation of the Markowitz portfolio. In contrast to \(\ell_1\) penalty regularized portfolio, \(\ell_p\) penalty regularized portfolio, cardinality constrained models and \(1/N\) investment strategy, the two-phase approach finds the sparse Markowitz portfolio selection with efficient investment in accordance with the Markowitz minimum variance portfolio structure and preserves the stability of the model. The convergence analysis showed that SAA method is effective with this two-phase portfolio optimization approach. The application demonstrated the sparsity, and the superior performance of our approach from the perspective of Sharpe ratio, STD, VaR, and CVaR.

Acknowledgments. We would like to thank Prof. Chen Xiaojun for advising us to study the two-phase stochastic linear complementarity approach and providing us with precious suggestions very much. We thank the valuable suggestions from Associate Editor and two referees. The first author was supported by the National Science Foundation of China (11301222). The second author was supported by the National Natural Science Foundation of China (No. 11401308) and Natural Science Foundation of Jiangsu Province, China (No. BK20140768).
Figure 2. S&P 500 Portfolio (a) Sharpe Ratio, (b) Sparsity. The bar from the left to the right in each test stands for LCP sparse portfolio, $\ell_1 0.1$, CCPS100 and $1/N$.

Figure 3. Hong Kong and Mainland China Cross Market Portfolio (a) Sharpe Ratio, (b) Sparsity. The bar from the left to the right in each test stands for LCP sparse portfolio, $\ell_1 0.1$, CCPS20, CCPS25, $\ell_p 0.015$ and $1/N$. 
Figure 4. FF48 Portfolio (a) Sharpe Ratio, (b) Sparsity. The bar from the left to the right in each test stands for LCP sparse portfolio, $\ell_1 0.1$, CCPS18, CCPS24, $\ell_p 0.015$ and $1/N$.

REFERENCES

[1] D. Bertsimas and R. Shioda, Algorithms for cardinality-constrained quadratic optimization, *Computational Optimization and Applications*, 43 (2009), 1–22.

[2] W. Bian and X. Chen, Neural network for nonsmooth, nonconvex constrained minimization via smooth approximation, *IEEE Transactions on Neural Networks and Learning Systems*, 25 (2014), 545–556.

[3] P. Bonami and M. A. Lejeune, An exact solution approach for portfolio optimization problems under stochastic and integer constraints, *Operations Research*, 57 (2009), 650–670.

[4] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer-Verlag, New York, 2000.

[5] J. M. Borwein and Q. J. Zhu, A survey of subdifferential calculus with applications, *Nonlinear Analysis: Theory, Methods & Applications*, 38 (1999), 687–773.

[6] J. Brodie, I. Daubechies, C. De Mol, D. Giammone and I. Loris, Sparse and stable markowitz portfolios, *Proceedings of the National Academy of Sciences*, 106 (2009), 12267–12272.

[7] E. J. Candes and T. Tao, Decoding by linear programming, *IEEE Transactions on Information Theory*, 51 (2005), 4203–4215.

[8] F. Cesarone, A. Scozzari and F. Tardella, A new method for mean-variance portfolio optimization with cardinality constraints, *Annals of Operations Research*, 205 (2013), 213–234.

[9] C. Chen, X. Li, C. Tolman, S. Wang and Y. Ye, Sparse portfolio selection via quasi-norm regularization, preprint, arXiv:1312.6350.

[10] X. Chen, Smoothing methods for nonsmooth, nonconvex minimization, *Mathematical Programming*, 134 (2012), 71–99.

[11] X. Chen, L. Guo, Z. Lu and J. Ye, An augmented Lagrangian method for non-Lipschitz nonconvex programming, *SIAM Journal on Numerical Analysis*, 55 (2017), 168–193.
[12] X. Chen and S. Xiang, Sparse solutions of linear complementarity problems, *Mathematical Programming*, 159 (2016), 539–556.
[13] R. W. Cottle, J. -S. Pang and R. E. Stone, *The Linear Complementarity Problem*, Academic Press, Boston, MA, 1992.
[14] V. DeMiguel, L. Garlappi, F. J. Nogales and R. Uppal, A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms, *Management Science*, 55 (2009), 798–812.
[15] V. DeMiguel, L. Garlappi and R. Uppal, Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?, *Review of Financial Studies*, 22 (2009), 1915–1953.
[16] G. F. Deng, W. T. Lin and C. C. Lo, Markowitz-based portfolio selection with cardinality constraints using improved particle swarm optimization, *Expert Systems with Applications*, 39 (2012), 4558–4566.
[17] D. W. Diamond and R. E. Verrecchia, Constraints on short-selling and asset price adjustment to private information, *Journal of Financial Economics*, 18 (1987), 277–311.
[18] J. Gao and D. Li, Optimal cardinality constrained portfolio selection, *Operations Research*, 61 (2013), 745–761.
[19] H. Markowitz, Portfolio selection, *The Journal of Finance*, 7 (1952), 77–91.
[20] A. J McNeil, R. Frey and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, 2015.
[21] R. C. Merton, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics*, 8 (1980), 323–361.
[22] H. Qi and D. Sun, A quadratically convergent Newton method for computing the nearest correlation matrix, *SIAM Journal on Matrix Analysis and Applications*, 28 (2006), 360–385.
[23] B. K. Natarajan, Sparse approximate solutions to linear systems, *SIAM Journal on Computing*, 24 (1995), 227–234.
[24] R. T. Rockafellar and S. Uryasev, Optimization of conditional value-at-risk, *Journal of Risk*, 2 (2000), 21–42.
[25] W. F. Sharpe, The Sharpe ratio, *The Journal of Portfolio Management*, 21 (1994), 49–58.
[26] A. Shleifer and R. W. Vishny, The limits of arbitrage, *The Journal of Finance*, 52 (1997), 35–55.
[27] Y. Tian, S. Fang, Z. Deng and Q. Jin, Cardinality constrained portfolio selection problem: A completely positive programming approach, *Journal of Industrial and Management Optimization*, 12 (2016), 1041–1056.
[28] F. Xu, Z. Lv and Z. Xu, An efficient optimization approach for a cardinality-constrained index tracking problem, *Optimization Methods and Software*, 31 (2016), 258–271.
[29] F. Xu, G. Wang and Y. Gao, Nonconvex L1/2 regularization for sparse portfolio selection, *Pacific Journal of Optimization*, 10 (2014), 163–176.
[30] H. Xu and D. Zhang, Monte Carlo methods for mean-risk optimization and portfolio selection, *Computational Management Science*, 9 (2012), 3–29.
[31] L. Xue, S. Ma and H. Zou, Positive-definite ℓ1-penalized estimation of large covariance matrices, *Journal of the American Statistical Association*, 107 (2012), 1480–1491.
[32] J. Ye, Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints, *SIAM Journal on Optimization*, 10 (2000), 943–962.
[33] X. Zheng, X. Sun and D. Li, Improving the performance of MIQP solvers for quadratic programs with cardinality and minimum threshold constraints: A semidefinite program approach, *INFORMS Journal on Computing*, 26 (2014), 690–703.

Received January 2016; 1st revision August 2016; 2nd revision February 2017.

E-mail address: qiuy.wang@connect.polyu.hk
E-mail address: hlsun@njust.edu.cn