Sparse Portfolio Selection via Non-convex Fraction Function

Angang Cui\(^1\) · Jigen Peng\(^1,\ast\) · Chengyi Zhang\(^2\) · Haiyang Li\(^2\) · Meng Wen\(^2\)

Received: date / Accepted: date

Abstract In this paper, a continuous and non-convex promoting sparsity fraction function is studied in two sparse portfolio selection models with and without short-selling constraints. Firstly, we study the properties of the optimal solution to the problem \((\text{FP}_{a,\lambda,\eta})\) including the first-order and the second optimality condition and the lower and upper bound of the absolute value for its nonzero entries. Secondly, we develop the thresholding representation theory of the problem \((\text{FP}_{a,\lambda,\eta})\). Based on it, we prove the existence of the resolvent operator of gradient of \(P_a(x)\), calculate its analytic expression, and propose an iterative fraction penalty thresholding (IFPT) algorithm to solve the problem \((\text{FP}_{a,\lambda,\eta})\). Moreover, we also prove that the value of the regularization parameter \(\lambda > 0\) can not be chosen too large. Indeed, there exists \(\bar{\lambda} > 0\) such that the optimal solution to the problem \((\text{FP}_{a,\lambda,\eta})\) is equal to zero for any \(\lambda > \bar{\lambda}\). At last, inspired by the thresholding representation theory of the problem \((\text{FP}_{a,\lambda,\eta})\), we propose an iterative nonnegative fraction penalty thresholding (INFPT) algorithm to solve the problem \((\text{FP}_{a,\lambda,\eta}^\geq)\). Empirical results show that our methods, for some proper \(a > 0\), perform effective in finding the sparse portfolio weights with and without short-selling constraints.

Keywords Markowitz mean-variance model · Sparse portfolio selection · Short-selling · Non-convex fraction function · IFPT algorithm · INFPT algorithm

Mathematics Subject Classification (2010) 65K05 · 90C26 · 90C90

\* Corresponding author
Jigen Peng
E-mail: jgpengxjtu@126.com;
1 School of Mathematics and Statistics, Xi’an Jiaotong University, Xi’an, 710049, China
2 School of Science, Xi’an Polytechnic University, Xi’an, 710048, China
1 Introduction

The classical mean-variance (M-V) portfolio selection model [1], also known as Markowitz mean-variance model constructed in a frictionless world, has been widely used in economic modeling of finance markets and asset pricing. In M-V model, the return and the risk of a portfolio are measured by the mean and the variance of the portfolio random returns, respectively, and it aims to find the optimal asset weight vector that minimizes the portfolio variance, subject to the constraint that the portfolio exhibits a desired portfolio return. It means investors need to invest in a large number of assets. The M-V portfolio theory believes that using diversified portfolio investment can effectively control the portfolio risk. When the number of assets is typically large, it means investors need to invest in a large number of assets and the solution of the M-V model is usually non-zero on almost all of the components. However, the large number of assets always lead to the high transaction costs and complexity of portfolio management and the M-V model becomes numerically unstable [8]. Therefore, almost all the investors can only invest in a limited number of assets.

The number restriction on assets motivates many researchers to study the sparse M-V portfolio selection problem, that is, get a sparse asset allocation (solution) with better out-of-sample performances and to reduce the transaction costs and the complexity of portfolio management. This sparse problem is often called cardinality constrained portfolio optimization, and some variations thereof, have been fairly intensively studied in [2–12]. For the sake of uniformity, in this paper, we call it the sparse portfolio selection problem. Unfortunately, this sparse problem, motivated by the need of inducing sparsity on the selected portfolio to reduce transaction costs, complexity of portfolio management, and instability of the solution is a difficult, in fact NP-hard, combinatorial problem (see [4]). In [8], the $\ell_1$-norm is proposed to promote the sparsity of assets in the portfolio, as argued by authors, helps inducing sparsity of the selected portfolio and can be a remedy to the high instability of classic methods for portfolio selection when short-selling is permitted. However, $\ell_1$-regularization approach is not effective in promoting sparsity in presence of budget and no-short-selling constraints [11]. Moreover, it tends to lead to biased estimation by shrinking all the entries toward to zero simultaneously, and sometimes results in over-penalization as the $\ell_1$-norm in compressed sensing (see [13]).

Inspired by the good performance of the non-convex fraction function in image restoration [14], and based on authors’ recent researches on fraction regularization in compressed sensing [15], we propose two sparse fraction portfolio selection models with and without short-selling constraints in this paper. The proposed sparse fraction portfolio selection model with short-selling constraint can generate optimal portfolios with better sparsity than the portfolio selection models using $\ell_1$-regularization do, and the sparse fraction portfolio selection model without short-selling constraint can also perform effective in finding the sparse portfolio weights.
The rest of the paper is organized as follows. In Section 2, we review some sparse portfolio selection models, and then present two sparse portfolio selection models by introducing fraction regularization on portfolio weights. The Section 3 is devoted to discussing the properties of the optimal solution to the regularization problem \((FP_{a,\lambda,\eta})\) including the first-order and the second optimality condition and the lower and upper bound of the absolute value for its nonzero entries. Moreover, we also proved that the value of the regularization parameter \(\lambda\) can not be chosen too large. Indeed, there exists \(\lambda_0 > 0\) such that the optimal solution to the problem \((FP_{a,\lambda,\eta})\) is equal to zero for any \(\lambda > \lambda_0\). In Section 3, we propose the IFPT algorithm to solve the problem \((FP_{a,\lambda,\eta})\) and, inspired by the thresholding representation of the IFPT algorithm, the INFPT algorithm is given to solve the problem \((FP_{a,\lambda,\eta}^\geq)\). In Section 4, we present the experiments with a series of portfolio selection applications to demonstrate the effectiveness of the new algorithms. We conclude this paper in Section 5.

2 Sparse portfolio selection models

2.1 The M-V portfolio selection model

Let \(r_t = (r_{1,t}, r_{2,t}, \cdots, r_{n,t})^\top \in \mathbb{R}^n\) be the vector of asset returns at time \(t\), \(t = 1, 2, \cdots, T\), \(E(r_t) = \mu\) and \(Q = E[(r_t - \mu)(r_t - \mu)^\top]\) be the mean return vector and the covariance matrix of asset returns, where \(r_{it}\) is the return of asset \(i\) at time \(t\). The traditional Markowitz portfolio selection model (see [1]) can be expressed as follows

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad \sigma := x^\top Q x \\
\text{s.t.} & \quad \mu^\top x = \beta \\
& \quad e_n^\top x = 1,
\end{align*}
\]

where \(x = (x_1, x_2, \cdots, x_n)^\top \in \mathbb{R}^n\) is the vector of asset weights, \(e_n \in \mathbb{R}^n\) is the vector of all ones, \(\beta\) is the minimum expected return from the portfolio that is expected by an investor. Note also that, if the non-short-selling (without short-selling) constraint \(x \geq 0\) is added to problem (1), we can recast this problem as portfolio selection model without short-selling constraint

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad \sigma := x^\top Q x \\
\text{s.t.} & \quad \mu^\top x = \beta \\
& \quad e_n^\top x = 1 \\
& \quad x \geq 0.
\end{align*}
\]  

Since \(Q = E[(r_t - \mu)(r_t - \mu)^\top]\), we have

\[
x^\top Q x = E[|\beta - x^\top r_t|^2] = \frac{1}{T} \|Rx - \beta e_T\|_2^2
\]
where $R = (r_1, r_2, \cdots, r_T)^T \in \mathbb{R}^{T \times n}$. Then the Markowitz portfolio selection model (1) can be expressed as follows

$$
\min_{x \in \mathbb{R}^n} \frac{1}{T} \|Rx - \beta e_T\|_2^2 \tag{3}
$$

s.t. $\mu^T x = \beta$

$e_n^T x = 1$

Usually, both the vector $\mu$ and the matrix $Q$ are not known analytically but can be estimated using historical data.

2.2 The sparse portfolio selection model

The sparsity requirement comes from the real world practice, where the administration of a portfolio made up of a large number of assets, possibly with very small holdings for some of them, is clearly not desirable because of the transactions costs and the complexity of management. The sparsity restricted model is often called cardinality constrained portfolio selection problem by limiting the number of assets in the portfolio, and defined as follows

$$
\min_{x \in \mathbb{R}^n} \frac{1}{T} \|Rx - \beta e_T\|_2^2 \tag{4}
$$

s.t. $\mu^T x = \beta$

$e_n^T x = 1$

$$\|x\|_0 \leq k,$$

where the $\|x\|_0$ is the $\ell_0$-norm of $x$ indicates the number of nonzero components of $x$, and the parameter $k$ is the chosen limit of assets to be held in the portfolio.

For the sake of uniformity, we call it the sparse portfolio selection model in this paper. Unfortunately, sparse problem (4) motivated by the need of inducing sparsity on the selected portfolio is a difficult, in fact NP-hard, combinatorial problem (see [4]).

In [8], an important sparse portfolio selection model, based on $\ell_1$-norm regularization, is proposed to promote the sparsity of assets in the portfolio, as argued by authors, helps induce sparsity of the selected portfolio and can be a remedy to the high instability of classic methods for portfolio selection when short-selling is permitted. This $\ell_1$-norm regularization portfolio selection model can be viewed as the following mathematical form

$$
\min_{x \in \mathbb{R}^n} \frac{1}{T} \|Rx - \beta e_T\|_2^2 + \lambda \|x\|_1 \tag{5}
$$

where $\lambda > 0$ is a regularization parameter, and $\|x\|_1 = \sum_{i=1}^n |x_i|$. Problem (5) is called sparse and stable M-V portfolio selection model in [8], however, sparsity of the resulting portfolio is not guaranteed from problem (5), since the $\ell_1$-norm of the asset weights will result in a constant value of one when
asset weights is nonnegative. Moreover, it tends to lead to biased estimation by shrinking all the entries toward to zero simultaneously, and sometimes results in over-penalization as the $\ell_1$-norm in compressed sensing (see [13]).

2.3 The new sparse portfolio selection model

Inspired by the good performance of the non-convex fraction function in image restoration and compressed sensing (see, e.g., [14, 15]), we take the non-convex function $P_a(x)$ to substitute the $\ell_1$-norm in problem (5).

The function $P_a(x)$ is defined as

$$P_a(x) = \sum_{i=1}^{n} \rho_a(x_i), \quad a > 0 \quad (6)$$

where

$$\rho_a(t) = \frac{a|t|}{a|t| + 1} \quad (7)$$

is the fraction function, and the parameter $a \in (0, +\infty)$. It is easy to verify that $\rho_a(t)$ is increasing and concave in $t \in [0, +\infty)$. With the change of parameter $a$, we have

$$\lim_{a \to +\infty} \rho_a(t) = \begin{cases} 0, & \text{if } t = 0; \\ 1, & \text{if } t \neq 0, \end{cases} \quad (8)$$

and the non-convex function $P_a(x)$ interpolates the $\ell_0$-norm of vector $x$:

$$\lim_{a \to +\infty} P_a(x) = \lim_{a \to +\infty} \sum_{x_i \neq 0} \rho_a(x_i) = \|x\|_0. \quad (9)$$

![Figure 1](image_url) Fig. 1 The behavior of the fraction function $p_a(t)$ for various values of $a$. 
By this substitution, we translate the problem (5) with short-selling constraint into the following problem

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad \frac{1}{T} \|Rx - \beta e_T\|_2^2 + \lambda P_a(x) \\
\text{s.t.} & \quad \mu^T x = \beta \\
& \quad e_n^T x = 1.
\end{align*}
\] (10)

If short-selling is not permitted (without short-selling constraint), the model for the problem (10) has the form

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad \frac{1}{T} \|Rx - \beta e_T\|_2^2 + \lambda P_a(x) \\
\text{s.t.} & \quad \mu^T x = \beta \\
& \quad e_n^T x = 1 \\
& \quad x \geq 0.
\end{align*}
\] (11)

Meanwhile, problems (10) and (11) could be expressed in the matrix-vector form

\[
\begin{align*}
(F_{Pa,\lambda}) & \quad \min_{x \in \mathbb{R}^n} \quad \frac{1}{T} \|Rx - \beta e_T\|_2^2 + \lambda P_a(x) \\
\text{s.t.} & \quad Ax = b
\end{align*}
\] (12)

and

\[
\begin{align*}
(F_{Pa,\lambda}^>) & \quad \min_{x \in \mathbb{R}^n} \quad \frac{1}{T} \|Rx - \beta e_T\|_2^2 + \lambda P_a(x) \\
\text{s.t.} & \quad Ax = b \\
& \quad x \geq 0,
\end{align*}
\] (13)

where \( b = (\beta, 1)^T \in \mathbb{R}^2, A = (\mu, e_n)^T \in \mathbb{R}^{2 \times n} \).

The penalty function problems for \((F_{Pa,\lambda})\) and \((F_{Pa,\lambda}^>)\) are given by

\[
\begin{align*}
(F_{Pa,\lambda,\eta}) & \quad \min_{x \in \mathbb{R}^n} \quad \frac{1}{T} \|Rx - \beta e_T\|_2^2 + \lambda P_a(x) + \eta \|Ax - b\|_2^2 \\
(F_{Pa,\lambda,\eta}^>) & \quad \min_{x \in \mathbb{R}^n, \exists x \geq 0} \quad \frac{1}{T} \|Rx - \beta e_T\|_2^2 + \lambda P_a(x) + \eta \|Ax - b\|_2^2
\end{align*}
\] (14) (15)

where \( \eta > 0 \) is the penalty parameter.

3 Properties of the problem \((F_{Pa,\lambda,\eta})\)

In this section, we discuss some properties of the problem \((F_{Pa,\lambda,\eta})\) including the first-order and the second optimality condition and the lower and upper bounds of the absolute value for its nonzero entries. Moreover, we also prove that the value of the regularization parameter \( \lambda \) can not be chosen too large. Indeed, there exists \( \bar{\lambda} > 0 \) such that the optimal solution to the problem \((F_{Pa,\lambda,\eta})\) is equal to zero for any \( \lambda > \bar{\lambda} \).
3.1 Lower and upper bounds of the optimal solution

**Theorem 1** (The first-order optimality condition) Let \( x^* \) be any solution to the problem \((FP_{a,\lambda,\eta})\) and for any \( h \in \mathbb{R}^n \) with \( \text{supp}(h) \subseteq \text{supp}(x^*) \) \((\text{supp}(x^*) \) represents the support of vector \( x^*\), \( \# = \{x_i | x_i \neq 0\} \)),

\[
\frac{2}{T} \langle \beta e_T - Rx^*, Rh \rangle + 2\eta \langle b - Ax^*, Ah \rangle = \lambda \sum_{i \in \text{supp}(x^*)} \frac{a_i \text{sign}(x^*_i)}{(1 + a|x^*_i|)^2}. \tag{16}
\]

**Proof** Let \( x^* \) be any solution to the problem \((FP_{a,\lambda,\eta})\). Then, for all \( \tau \in \mathbb{R} \) and \( h \in \mathbb{R}^n \), the following inequality holds

\[
\frac{1}{T} \|Rx^* - \beta e_T\|_2^2 + \lambda P_a(x^*) + \eta \|Ax^* - b\|_2^2 \\
\leq \frac{1}{T} \|R(x^* + \tau h) - \beta e_T\|_2^2 + \lambda P_a(x^* + \tau h) + \eta \|A(x^* + \tau h) - b\|_2^2,
\]

equivalently,

\[
\frac{\tau^2}{T} \|Rh\|_2^2 + \eta \tau^2 \|Ah\|_2^2 + \frac{2\tau}{T} \langle Rx^* - \beta e_T, Rh \rangle + 2\eta \tau \langle Ax^* - b, Ah \rangle + \lambda (P_a(x^* + \tau h) - P_a(x^*)) \geq 0. \tag{17}
\]

If \( \text{supp}(h) \subseteq \text{supp}(x^*) \), then for a small enough \( \tau \) the vector \( x^*, x^* + \tau h \) and \( x^* - \tau h \) have the same sign, and

\[
P_a(x^* + \tau h) - P_a(x^*) = \sum_{i \in \text{supp}(x^*)} \left( \frac{a_i |x^*_i + \tau h_i|}{1 + a|x^*_i| + \tau h_i} - \frac{a_i |x^*_i|}{1 + a|x^*_i|} \right).
\]

Dividing by \( \tau > 0 \) both sides of the inequality (18) and letting \( \tau \to 0 \) yield

\[
\frac{2}{T} \langle Rx^* - \beta e_T, Rh \rangle + 2\eta \langle Ax^* - b, Ah \rangle + \lambda \sum_{i \in \text{supp}(x^*)} \frac{a_i \text{sign}(x^*_i)}{(1 + a|x^*_i|)^2} \geq 0. \tag{18}
\]

Obviously, the above inequality also holds for \( -h \) which leads to the equality (16). This completes the proof.

Choosing \( h \) as the \( i \)-th base vector \( e_i \) for each \( i = 1, 2, \cdots, n \) in (16), we can derive the following corollary.

**Corollary 1** Suppose that \( x^* \) is the solution of the problem \((FP_{a,\lambda,\eta})\). Then, for \( i \in \text{supp}(x^*) \),

\[
\frac{2}{T} \langle R^T (\beta e_T - Rx^*) \rangle_i + 2\eta \langle A^T (b - Ax^*) \rangle_i = \frac{a \lambda}{(1 + a|x^*_i|)^2}. \tag{19}
\]

Following the analysis adopted above, we can further establish the following optimality condition.
Theorem 2 (The second-order optimality condition) Every solution $x^*$ to the problem $(FP_{a,\lambda,\eta})$ satisfies the following condition:

(1) For all $h \in \mathbb{R}^n$ with $\text{supp}(h) \subseteq \text{supp}(x^*)$,

$$\frac{1}{T} \|Rh\|_2^2 + \eta \|Ah\|_2^2 \geq \lambda \sum_{i \in \text{supp}(x^*)} \frac{2a^2h_i^2}{(1 + a|x^*_i|)^3}. \tag{20}$$

(2) Moreover, for all $i \in \text{supp}(x^*)$ and $a > \frac{\sqrt{\lambda}}{\sqrt{T} \|R_i\|_2^2 + \eta \|A_i\|_2^2}$, it holds that

$$|x^*_i| \geq \frac{\sqrt{\lambda}}{\sqrt{T} \|R_i\|_2^2 + \eta \|A_i\|_2^2} - \frac{1}{a}. \tag{21}$$

Proof (1) Let $\text{supp}(h) \subseteq \text{supp}(x^*)$. Then, incorporating the equality (16) into the inequality (18) yields that, for all $\tau \in \mathbb{R}$,

$$\tau^2 \|Rh\|_2^2 + \eta \tau^2 \|Ah\|_2^2 - \lambda \sum_{i \in \text{supp}(x^*)} \frac{\tau a h_i \text{sign}(x^*_i)}{(1 + a|x^*_i|)^2} + \lambda (P_a(x^* + \tau h) - P_a(x^*)) \geq 0,$n

or equivalently

$$\frac{1}{T} \|Rh\|_2^2 + \eta \|Ah\|_2^2 \geq \lambda \sum_{i \in \text{supp}(x^*)} \frac{\tau a h_i \text{sign}(x^*_i)}{(1 + a|x^*_i|)^2} - (P_a(x^* + \tau h) - P_a(x^*)). \tag{22}$$

Hence, letting $\tau \to 0$ on the right-hand of inequality above, we have the inequality (20).

(2) If we replace $h$ in inequality (22) with the base vector $e_i$ for every $i \in \text{supp}(x^*)$, then we have the component-wise inequality

$$\frac{1}{T} \|R_{i,i}\|_2^2 + \eta \|A_{i,i}\|_2^2 \geq \frac{\lambda}{\tau^2} \left( \frac{a \text{sign}(x^*_i)}{(1 + a|x^*_i|)^2} - \frac{a|x^*_i|}{1 + a|x^*_i|} \right),$$

where $R_{i,i}$ and $A_{i,i}$ represent the $i$-th column of the matrix $R$ and $A$, respectively. Particularly, above inequality is available for $\tau = -x^*_i$. So, we have

$$\frac{1}{T} \|R_{i,i}\|_2^2 + \eta \|A_{i,i}\|_2^2 \geq \frac{\lambda}{(x^*_i)^2} \left( \frac{a|x^*_i|}{(1 + a|x^*_i|)^2} + \frac{a|x^*_i|}{1 + a|x^*_i|} \right) \geq \frac{\lambda a^2}{(1 + a|x^*_i|)^2}.$$n

It follows that

$$\frac{1}{T} \|R_i\|_2^2 + \eta \|A_i\|_2^2 \geq \frac{\lambda a^2}{(1 + a|x^*_i|)^2}.$$

From the inequality above, the inequality (21) immediately follows. This completes the proof.
Theorem 3 Suppose that $x^*$ is the optimal solution to the problem $(FP_{a,\lambda,\eta})$. If $\lambda > \frac{1}{T}\|\beta e\|_2^2 + \eta\|b\|_2^2$, then

$$\|x^*\|_\infty \leq \frac{\frac{1}{T}\|\beta e\|_2^2 + \eta\|b\|_2^2}{a(\lambda - (\frac{1}{T}\|\beta e\|_2^2 + \eta\|b\|_2^2))}.$$  (23)

Proof Let $x^*$ be the optimal solution to the problem $(FP_{a,\lambda,\eta})$. Then we have

$$f(x^*) = \frac{1}{T}\|Rx^* - \beta e\|_2^2 + \lambda P_a(x^*) + \eta\|Ax^* - b\|_2^2 \leq f(0) = \frac{1}{T}\|\beta e\|_2^2 + \eta\|b\|_2^2.$$  

Hence $\lambda P_a(x^*) \leq \frac{1}{T}\|\beta e\|_2^2 + \eta\|b\|_2^2$, which implies that

$$\frac{a\|x^*\|_\infty}{1 + a\|x^*\|_\infty} \leq \frac{\frac{1}{T}\|\beta e\|_2^2 + \eta\|b\|_2^2}{\lambda}.$$  

If $\lambda > \frac{1}{T}\|\beta e\|_2^2 + \eta\|b\|_2^2$, then

$$\|x^*\|_\infty \leq \frac{\frac{1}{T}\|\beta e\|_2^2 + \eta\|b\|_2^2}{a(\lambda - (\frac{1}{T}\|\beta e\|_2^2 + \eta\|b\|_2^2))}.$$  

This completes the proof.

3.2 Large regularization parameter $\lambda$ leads to zero solution

Before we embark to this discussion, we should declare that the results derived in this following discussion are worst-case ones, implying that the kind of guarantees we obtained are over-pessimistic for all possibilities.

Lemma 1 Let $x^*$ of sparsity $r$ be the optimal solution of the problem $(FP_{a,\lambda,\eta})$, the matrix $\tilde{R}$ be the submatrix of $R$ corresponding to $\text{supp}(x^*)$ and the matrix $\tilde{A}$ be the submatrix of $A$ corresponding to $\text{supp}(x^*)$. Then the matrix $\frac{2}{T}\tilde{R}^\top \tilde{R} + 2\eta \tilde{A}^\top \tilde{A}$ is positive definite.

Proof Without loss of generality, we assume

$$x^* = (x_1^*, x_2^*, \ldots, x_r^*, 0, \ldots, 0)^\top.$$  

Let $z^* = (x_1^*, x_2^*, \ldots, x_r^*)^\top \in \mathbb{R}^r$, $\tilde{R} \in \mathbb{R}^{r \times r}$ be the sub-matrix of $R$, whose columns in matrix $R$ corresponding to $z^*$, and $\tilde{A} \in \mathbb{R}^{2 \times r}$ be the sub-matrix of $A$, whose columns in matrix $A$ corresponding to $z^*$. Define a function $g : \mathbb{R}^r \rightarrow \mathbb{R}$ by

$$g(z^*) = \frac{1}{T}\|\tilde{R}z^* - \beta e\|_2^2 + \lambda P_a(z^*) + \eta\|\tilde{A}z^* - b\|_2^2.$$  (24)

We have

$$f(X^*) = \frac{1}{T}\|Rx^* - \beta e\|_2^2 + \lambda P_a(x^*) + \eta\|Ax^* - b\|_2^2$$

$$= \frac{1}{T}\|\tilde{R}z^* - \beta e\|_2^2 + \lambda P_a(z^*) + \eta\|\tilde{A}z^* - b\|_2^2$$  

$$= g(z^*).$$  (25)
Since function $g$ is continuously differentiable at $z^*$, moreover, in a neighborhood of $z^*$,
\[
    g(z^*) = f(x^*) \leq \min_{z \in \mathbb{R}^n} \left\{ f(x) | x_i = 0, i = r + 1, r + 2, \ldots, n \right\} = \min_{z \in \mathbb{R}^n} g(z),
\]
which implies that $z^*$ is a local minimizer of the function $g$. Hence, the second order necessary condition for

\[
    \min_{z \in \mathbb{R}^n} g(z)
\]
holds at $z^*$. The second order necessary condition at $z^*$ gives that the matrix
\[
    \frac{2}{T} \tilde{R}^\top \tilde{R} + 2\eta \tilde{A}^\top \tilde{A} - \text{Diag} \left( \frac{2\lambda a^2}{(a|z^*| + 1)^2} \right), \quad i = 1, 2, \ldots, r
\]
is positive semi-definite, and the matrix
\[
    M = \text{Diag} \left( \frac{2\lambda a^2}{(a|z^*| + 1)^2} \right)
\]
is positive. Therefore, the matrix $\frac{2}{T} \tilde{R}^\top \tilde{R} + 2\eta \tilde{A}^\top \tilde{A}$ must be positive definite. This completes the proof.

Nextly, we shall show that the value of the regularization parameter $\lambda$ of the problem $(FP_{a,\lambda,\eta})$ can not be chosen too large.

**Theorem 4** Let
\[
    \bar{\lambda} = \frac{1}{T} \| \beta e_T \|_2^2 + \eta \| b \|_2^2 + \frac{1}{a T} \| \tilde{R}^\top e_T + \eta \tilde{A}^\top b \|_\infty \]
\[
    + \frac{1}{a} \sqrt{\left\| \frac{\beta}{T} \tilde{R}^\top e_T + \eta \tilde{A}^\top b \right\|_\infty^2 + 2a \left( \frac{1}{T} \| \beta e_T \|_2^2 + \eta \| b \|_2^2 \right) \left\| \frac{\beta}{T} \tilde{R}^\top e_T + \eta \tilde{A}^\top b \right\|_\infty^2}.
\]

Then for all $\lambda \geq \bar{\lambda}$, the problem $(FP_{a,\lambda,\eta})$ admits the zero solution.

**Proof** By the proof of Lemma 1, the first order necessary condition for

\[
    g(z) = f(x) \leq \min_{x \in \mathbb{R}^n} \left\{ f(x) | x_i = 0, i = r + 1, r + 2, \ldots, n \right\}
\]

holds at $x^*$. The first order necessary condition gives that the matrix
\[
    \frac{2}{T} \tilde{R}^\top \tilde{R} + 2\eta \tilde{A}^\top \tilde{A} - \text{Diag} \left( \frac{2\lambda a^2}{(a|z^*| + 1)^2} \right), \quad i = 1, 2, \ldots, r
\]
is positive semi-definite, and the matrix
\[
    M = \text{Diag} \left( \frac{2\lambda a^2}{(a|z^*| + 1)^2} \right)
\]
is positive. Therefore, the matrix $\frac{2}{T} \tilde{R}^\top \tilde{R} + 2\eta \tilde{A}^\top \tilde{A}$ must be positive definite. This completes the proof.

Nextly, we shall show that the value of the regularization parameter $\lambda$ of the problem $(FP_{a,\lambda,\eta})$ can not be chosen too large.

**Theorem 4** Let
\[
    \bar{\lambda} = \frac{1}{T} \| \beta e_T \|_2^2 + \eta \| b \|_2^2 + \frac{1}{a T} \| \tilde{R}^\top e_T + \eta \tilde{A}^\top b \|_\infty \]
\[
    + \frac{1}{a} \sqrt{\left\| \frac{\beta}{T} \tilde{R}^\top e_T + \eta \tilde{A}^\top b \right\|_\infty^2 + 2a \left( \frac{1}{T} \| \beta e_T \|_2^2 + \eta \| b \|_2^2 \right) \left\| \frac{\beta}{T} \tilde{R}^\top e_T + \eta \tilde{A}^\top b \right\|_\infty^2}.
\]

Then for all $\lambda \geq \bar{\lambda}$, the problem $(FP_{a,\lambda,\eta})$ admits the zero solution.

**Proof** By the proof of Lemma 1, the first order necessary condition for

\[
    g(z) = f(x) \leq \min_{x \in \mathbb{R}^n} \left\{ f(x) | x_i = 0, i = r + 1, r + 2, \ldots, n \right\}
\]

holds at $x^*$. The first order necessary condition gives that the matrix
\[
    \frac{2}{T} \tilde{R}^\top \tilde{R} + 2\eta \tilde{A}^\top \tilde{A} - \text{Diag} \left( \frac{2\lambda a^2}{(a|z^*| + 1)^2} \right), \quad i = 1, 2, \ldots, r
\]
is positive semi-definite, and the matrix
\[
    M = \text{Diag} \left( \frac{2\lambda a^2}{(a|z^*| + 1)^2} \right)
\]
is positive. Therefore, the matrix $\frac{2}{T} \tilde{R}^\top \tilde{R} + 2\eta \tilde{A}^\top \tilde{A}$ must be positive definite. This completes the proof.
Sparse Portfolio Selection via Non-convex Fraction Function

Because the matrix \( \frac{2}{T} \tilde{R}^T e_T + 2\eta \tilde{A}^T b \) is positive definite (see Lemma 1), and hence

\[-(z^*)^T \left( \frac{2}{T} \tilde{R}^T e_T + 2\eta \tilde{A}^T b \right) + (z^*)^T \text{Diag}(\text{sign}(z^*)) \frac{\lambda a}{(a|z^*_i| + 1)^2} < 0,\]
equivalently,

\[
\sum_{i=1}^r \left( \frac{\lambda a|z^*_i|}{(a|z^*_i| + 1)^2} - 2 \left( \frac{\beta}{T} \tilde{R}^T e_T + \eta \tilde{A}^T b \right)_i z^*_i \right) < 0. \tag{28}
\]

Since

\[
\lambda > \frac{1}{T} \| \beta e_T \|_2^2 + \eta \| b \|_2^2 + \frac{1}{a} \left\| \frac{\beta}{T} R^T e_T + \eta A^T b \right\|_\infty
\]

\[
+ \frac{1}{\lambda} \left( \frac{\beta}{T} \tilde{R}^T e_T + \eta \tilde{A}^T b \right) \| b \|_2^2 + 2a \left( \frac{1}{T} \| \beta e_T \|_2^2 + \eta \| b \|_2^2 \right) \left\| \frac{\beta}{T} R^T e_T + \eta A^T b \right\|_\infty,
\]
we obtain

\[
a\lambda^2 - 2 \left[ a \left( \frac{1}{T} \| \beta e_T \|_2^2 + \eta \| b \|_2^2 \right) + \left\| \frac{\beta}{T} R^T e_T + \eta A^T b \right\|_\infty \right] \lambda
\]

\[
+ a \left( \frac{1}{T} \| \beta e_T \|_2^2 + \eta \| b \|_2^2 \right)^2 \geq 0, \tag{29}
\]
which implies that

\[
a\lambda^2 - \frac{a\left( \frac{1}{T} \| \beta e_T \|_2^2 + \eta \| b \|_2^2 \right)^2}{\lambda} \geq 2 \left\| \frac{\beta}{T} R^T e_T + \eta A^T b \right\|_\infty. \tag{30}
\]

Together with

\[
\frac{\lambda a}{(a|z^*_i| + 1)^2} \geq \frac{a\left( \frac{1}{T} \| \beta e_T \|_2^2 + \eta \| b \|_2^2 \right)^2}{\lambda}
\]
and

\[
\left| \left( \frac{\beta}{T} \tilde{R}^T e_T + \eta \tilde{A}^T b \right)_i \right| \leq \left\| \frac{\beta}{T} R^T e_T + \eta A^T b \right\|_\infty, \tag{31}
\]
we obtain

\[
\frac{\lambda a}{(a|z^*_i| + 1)^2} - 2 \left| \left( \frac{\beta}{T} \tilde{R}^T e_T + \eta \tilde{A}^T b \right)_i \right| \geq 0. \tag{32}
\]

Hence, for any \( i \in \{1, 2, \cdots, r\}, \)

\[
\frac{\lambda a|z^*_i|}{(a|z^*_i| + 1)^2} - 2 \left( \frac{\beta}{T} \tilde{R}^T e_T + \eta \tilde{A}^T b \right)_i z^*_i \geq 0,
\]
which is a contradiction with (28), as claimed. This completes the proof.
3.3 Problem \((FP_{a, \lambda, \eta})\) solves problem \((FP_{a, \lambda})\) for any \(\eta \rightarrow \infty\)

Let

\[
C_\lambda(x)_{(Ax=b)} = \frac{1}{T} \|Rx - \beta e_T\|_2^2 + \lambda P_a(x)
\]

and

\[
C_{\lambda, \eta}(x) = \frac{1}{T} \|Rx - \beta e_T\|_2^2 + \lambda P_a(x) + \eta \|Ax - b\|_2^2.
\]

Theorem 5
Suppose \(x^{[\lambda]}\) is the unique minimizer of the problem \((FP_{a, \lambda})\). Then, for each fixed \(\eta > 0\), the minimizer \(x^{[\lambda, \eta]}\) of the problem \((FP_{a, \lambda, \eta})\) converges to \(x^{[\lambda]}\) as \(\eta \rightarrow +\infty\).

Proof
Since \(x^{[\lambda, \eta]}\) minimizes \(C_{\lambda, \eta}\), it follows that

\[
C_{\lambda, \eta}(x^{[\lambda, \eta]}) \leq C_{\lambda, \eta}(x^{[\lambda]}) = C_\lambda(x^{[\lambda]})_{(Ax=b)}.
\]

Consequently

\[
\lambda P_a(x^{[\lambda, \eta]}) \leq C_{\lambda, \eta}(x^{[\lambda, \eta]}) \leq C_\lambda(x^{[\lambda]})_{(Ax=b)},
\]

so that \(P_a(x^{[\lambda, \eta]})\) is bounded, uniformly in \(\eta\). This implies that the set \(\{x^{[\lambda, \eta]} : \eta > 0\}\) must have accumulation points that can be written as

\[
\tilde{x} = \lim_{n \rightarrow +\infty} x^{[\lambda, \eta_n]},
\]

where \(\eta_n \rightarrow +\infty\) as \(n \rightarrow +\infty\).

On the other hand, we also have

\[
\|Ax^{[\lambda, \eta]} - b\|_2^2 \leq \frac{1}{\eta} C_{\lambda, \eta}(x^{[\lambda, \eta]}) \leq \frac{1}{\eta} C_\lambda(x^{[\lambda]})_{(Ax=b)} \rightarrow 0 \text{ as } \eta \rightarrow +\infty.
\]

This implies that any accumulation point \(\tilde{x}\), of the type described above, satisfies \(A \tilde{x} = b\). It follows that

\[
C_\lambda(\tilde{x})_{(Ax=b)} = \lim_{n \rightarrow +\infty} C_\lambda(x^{[\lambda, \eta_n]})_{(Ax^{[\lambda, \eta_n]} - b\|_2^2 \leq \frac{1}{\eta_n} C_\lambda(x^{[\lambda]})_{(Ax^{[\lambda]} = b)}}
\]

\[
\leq \lim_{n \rightarrow +\infty} C_{\lambda, \eta_n}(x^{[\lambda, \eta_n]})
\]

\[
\leq \lim_{n \rightarrow +\infty} C_\lambda(x^{[\lambda]})_{(Ax^{[\lambda]} = b)}
\]

\[
= C_\lambda(x^{[\lambda]})_{(Ax^{[\lambda]} = b)}.
\]

Since \(\tilde{x}\) is any accumulation point of \(\{x^{[\lambda, \eta]} : \eta > 0\}\) for \(\eta \rightarrow +\infty\), and \(x^{[\lambda]}\) is the unique minimizer of \(C_\lambda(x)_{(Ax=b)}\), it follows that \(\tilde{x} = x^{[\lambda]}\). Since this is true for an arbitrary accumulation point of \(\{x^{[\lambda, \eta]} : \eta > 0\}\) the type described above, it follows that

\[
\lim_{\eta \rightarrow +\infty} x^{[\lambda, \eta]} = x^{[\lambda]}.
\]

This completes the proof.
4 Algorithms for solving problems $(FP_{a,\lambda,\eta})$ and $(FP_{a,\lambda,\eta}^\geq)$

In this section, we develop the thresholding representation theories of the problems $(FP_{a,\lambda,\eta})$ and $(FP_{a,\lambda,\eta}^\geq)$. Based on them, we propose the IFPT algorithm and the INFPT algorithm to solve the problems $(FP_{a,\lambda,\eta})$ and $(FP_{a,\lambda,\eta}^\geq)$ for all $a > 0$, respectively.

4.1 IFPT algorithm for solving the problem $(FP_{a,\lambda,\eta})$

In this subsection, the IFPT algorithm is proposed to solve the problem $(FP_{a,\lambda,\eta})$ for all $a > 0$. Before we embark to this discussion, some crucial results need to be introduced for our later use.

Define a function of $\beta \in \mathbb{R}$ as

$$f_\lambda(\beta) = (\beta - \gamma)^2 + \lambda p_a(\beta),$$

and

$$\text{prox}_{\lambda a}^\gamma(\gamma) \triangleq \arg\min_{\beta \in \mathbb{R}} f_\lambda(\beta).$$

Lemma 2 (see [15,16]) The operator $\text{prox}_{\lambda a}^\gamma$ defined in (36) can be expressed as

$$\text{prox}_{\lambda a}^\gamma(\gamma) = \begin{cases} g_{a,\lambda}(\gamma), & \text{if } |\gamma| > t_{a,\lambda}^1; \\ 0, & \text{if } |\gamma| \leq t_{a,\lambda}^1, \end{cases}$$

where $g_{a,\lambda}$ is defined as

$$g_{a,\lambda}(\gamma) = \text{sign}(\gamma) \left( \frac{1 + |\gamma|}{2} (1 + 2 \cos(\frac{\phi(\gamma)}{3}) - \frac{\phi(\gamma)}{3}) - 1 \right),$$

$$\phi(\gamma) = \arccos \left( \frac{27\lambda a^2}{4(1 + a|\gamma|)^2} - 1 \right),$$

and the threshold value satisfies

$$t_{a,\lambda}^1 = \begin{cases} \frac{\lambda}{2} a, & \text{if } \lambda \leq \frac{1}{a^2}; \\ \frac{\sqrt{\lambda}}{2} a, & \text{if } \lambda > \frac{1}{a^2}. \end{cases}$$

Definition 1 Let $x \in \mathbb{R}^n$, the iterative thresholding operator $G_{\lambda,P_a}$ can be defined as

$$G_{\lambda,P_a}(x) = \left( \text{prox}_{\lambda a}^\gamma(x_1), \text{prox}_{\lambda a}^\gamma(x_2), \ldots, \text{prox}_{\lambda a}^\gamma(x_n) \right)^T,$$

where $\text{prox}_{\lambda a}^\gamma$ is defined in Lemma 2.
The iterative thresholding operator $G_{\lambda,P_n}$ simply applies the operator $\text{prox}_{\lambda}^T$ defined in Lemma 2 to a vector, and effectively shrink them towards zero. It is clear that if many of the entries of the vector $x$ are below the threshold value $t^*_{\alpha,\lambda}$, the sparsity of $G_{\lambda,P_n}(x)$ may be considerably lower than the sparsity of vector $x$.

Nextly, we shall show that the optimal solution to the problem $(FP_{a,\lambda,\eta})$ can also be expressed a thresholding operation.

For any fixed positive parameters $\lambda > 0$, $\varphi > 0$, $\eta > 0$, $a > 0$ and $x, z \in \mathbb{R}^n$, let

$$C_{\lambda,\eta,\varphi}(x, z) = \varphi[C_{\lambda,\eta}(x) - \frac{1}{T}\|Rx - Rz\|_2^2 - \eta\|Ax - Az\|_2^2] + \|x - z\|_2^2 \quad (41)$$

be the surrogate function of the function $C_{\lambda,\eta}(x)$ defined in (34). Clearly, $C_{\lambda,\eta,\varphi}(x, x) = \varphi C_{\lambda,\eta}(x)$.

**Theorem 6** For any fixed $\lambda > 0$, $\varphi > 0$, $\eta > 0$, $a > 0$ and $z \in \mathbb{R}^n$, $\min_{x \in \mathbb{R}^n} C_{\lambda,\eta,\varphi}(x, z)$ equivalents to

$$\min_{x \in \mathbb{R}^n} \left\{ \|x - B_\varphi(z)\|_2^2 + \lambda \varphi P_a(x) \right\}$$

where $B_\varphi(z) = z + \frac{\varphi}{T}R^T(\beta e_T - Rz) + \varphi \eta A^T(b - Az)$.

**Proof** By the definition, $C_{\lambda,\eta,\varphi}(x, z)$ can be rewritten as

$$C_{\lambda,\eta,\varphi}(x, z) = \|x - (z + \frac{\varphi}{T}R^T(\beta e_T - Rz) + \varphi \eta A^T(b - Az))\|_2^2 + \lambda \varphi P_a(x)$$

$$+ \|z\|_2^2 - \|z + \frac{\varphi}{T}R^T(\beta e_T - Rz) + \varphi \eta A^T(b - Az)\|_2^2 + \|z\|_2^2 + \frac{\varphi}{T}\|\beta e_T\|_2^2$$

$$- \frac{\varphi}{T}\|Rz\|_2^2 + \varphi \eta \|b\|_2 - \varphi \eta \|Az\|_2^2$$

which implies that $\min_{x \in \mathbb{R}^n} C_{\lambda,\eta,\varphi}(x, z)$ for any fixed $\lambda > 0$, $\varphi > 0$, $\eta > 0$, $a > 0$ and $z \in \mathbb{R}^n$ equivalents to

$$\min_{x \in \mathbb{R}^n} \left\{ \|x - B_\varphi(z)\|_2^2 + \lambda \varphi P_a(x) \right\}.$$ 

This completes the proof.

**Theorem 7** For any fixed $\lambda > 0$ and $0 < \varphi < \frac{1}{\sqrt{n}\|R\|_2 + \|\lambda x\|_2^2}$. If $x^*$ is the optimal solution of $\min_{x \in \mathbb{R}^n} C_{\lambda,\eta}(x)$, then $x^*$ is also the optimal solution of $\min_{x \in \mathbb{R}^n} C_{\lambda,\eta,\varphi}(x, x^*)$, that is

$$C_{\lambda,\eta,\varphi}(x^*, x^*) \leq C_{\lambda,\eta,\varphi}(x, x^*).$$
Proof By the definition of $C_{\lambda,\eta,\phi}(x, z)$, we have

$$
C_{\lambda,\mu,\phi}(x, x^*) = \phi \left[ C_{\lambda,\eta}(x) - \frac{1}{T} \| Rx - Rx^* \|_2^2 - \eta \| Ax - Ax^* \|_2^2 + \| x - x^* \|_2^2 \right]
= \phi \left[ \frac{1}{T} \| Rx - \beta e \|_2^2 + \lambda P_a(x) + \eta \| Ax - b \|_2^2 \right] - \frac{\varphi}{T} \| Rx - Rx^* \|_2^2
\geq \phi \frac{1}{T} \| Rx - \beta e \|_2^2 + \lambda P_a(x) + \eta \| Ax - b \|_2^2
= \varphi C_{\lambda,\eta}(x)
\geq \varphi C_{\lambda,\eta}(x^*)
= C_{\lambda,\eta,\phi}(x^*, x^*).
$$

This completes the proof.

Theorem 7 told us that $x^*$ is the optimal solution to $\min_{x \in \mathbb{R}^n} C_{\lambda,\eta,\phi}(x, z)$ with $z = x^*$, as long as, $x^*$ is the optimal solution of the problem $(FP_{a,\lambda,\eta})$. Combined with Theorem 6, we derive the most important conclusion in this paper, which underlies the algorithm to be proposed.

**Theorem 8** Let $x^*$ be the optimal solution of the problem $(FP_{a,\lambda,\eta})$. Then $x^*$ is also the optimal solution of the following minimization problem

$$
\min_{x \in \mathbb{R}^n} \left\{ \| x - B_\phi(x^*) \|_2^2 + \lambda \varphi P_a(x) \right\}.
$$

Combining Lemma 2, Definition 1 and Theorem 8, the thresholding representation of the problem $(FP_{a,\lambda,\eta})$ can be immediately concluded as the following description.

**Corollary 2** Let $x^* \in \mathbb{R}^n$ be the optimal solution of the problem $(FP_{a,\lambda,\eta})$. Then it can be given by

$$
x^* = G_{\lambda,\phi,P_a}(B_\phi(x^*)) \tag{42}
$$

where $G_{\lambda,\phi,P_a}$ and $\text{prox}_{\lambda \phi}$ are obtained by replacing $\lambda$ with $\lambda \varphi$ in $G_{\lambda,P_a}$ and $\text{prox}_{\lambda \phi}$.

With the representation (42), the IFPT algorithm for solving the problem $(FP_{a,\lambda,\eta})$ can be naturally proposed as following:

$$
x^{k+1} = G_{\lambda,\phi,P_a}(B_\phi(x^k)) \tag{43}
$$

where $B_\phi(x^k) = x^k + \frac{\varphi}{T} R^T (\beta e - Rx^k) + \varphi \eta A^T (b - Ax^k)$, which means that, in per iteration, every entries of vector $x^{k+1}$ satisfies

$$
x^{k+1}_i = \begin{cases} 
g_{a,\lambda \varphi}(x_i^k), & \text{if } |x_i^k| > t^*_{a,\lambda \varphi}, \\
0, & \text{if } |x_i^k| \leq t^*_{a,\lambda \varphi}, 
\end{cases} \tag{44}
$$

for $i = 1, \cdots, n$, where $g_{a,\lambda \varphi}$ and $t^*_{a,\lambda \varphi}$ are all defined in Lemma 2 which obtained by replacing $\lambda$ with $\lambda \varphi$ in $g_{a,\lambda}$ and $t^*_{a,\lambda}$, respectively.
**Theorem 9** (Convergence results of IFPT algorithm) Let \( \{x^k\} \) be the sequence generated by the FP algorithm with the step size \( \varphi \) satisfying \( 0 < \varphi < \frac{1}{\|R\|_2 + \eta \|A\|_2^2} \). Then

1) The sequence \( C_{\lambda, \eta}(x^k) \) is decreasing.

2) \( \{x^k\} \) is asymptotically regular, i.e., \( \lim_{k \to \infty} \|x^{k+1} - x^k\|_2^2 = 0 \).

3) Any accumulation point of \( \{x^k\} \) is a stationary point of the problem \( (FP_{\alpha, \lambda, \eta}) \).

**Proof** 1) By the proof of Theorem 8, we have

\[
C_{\lambda, \eta, \varphi}(x^{k+1}, x^k) = \min_{x \in \mathbb{R}^n} C_{\lambda, \eta, \varphi}(x, x^k).
\]

Combined with the definition of \( C_{\lambda, \eta}(x) \) and \( C_{\lambda, \eta, \varphi}(x, z) \), we have

\[
C_{\lambda, \eta}(x^{k+1}) = \frac{1}{\varphi} \left[ C_{\lambda, \eta, \varphi}(x^{k+1}, x^k) - \|x^{k+1} - x^k\|_2^2 \right] + \frac{1}{T} \|Rx^{k+1} - Rx^k\|_2^2 \\
+ \eta \|Ax^{k+1} - Ax^k\|_2^2.
\]

Since \( 0 < \varphi < \frac{1}{\|R\|_2 + \eta \|A\|_2^2} \), we get

\[
C_{\lambda, \eta}(x^{k+1}) = \frac{1}{\varphi} \left[ C_{\lambda, \eta, \varphi}(x^{k+1}, x^k) - \|x^{k+1} - x^k\|_2^2 \right] \\
+ \frac{1}{T} \|Rx^{k+1} - Rx^k\|_2^2 + \eta \|Ax^{k+1} - Ax^k\|_2^2 \\
\leq \frac{1}{\mu} \left[ C_{\lambda, \eta, \varphi}(x^k, x^k) - \|x^{k+1} - x^k\|_2^2 \right] \\
+ \frac{1}{T} \|Rx^{k+1} - Rx^k\|_2^2 + \eta \|Ax^{k+1} - Ax^k\|_2^2 \]

\[
\leq C_{\lambda, \eta}(x^k).
\]

That is, the sequence \( \{x^k\} \) is a minimization sequence of function \( C_{\lambda, \eta}(x) \), and

\[
C_{\lambda, \eta}(x^{k+1}) \leq C_{\lambda, \eta}(x^k)
\]

for all \( k \geq 0 \).

2) Let \( \theta = 1 - \frac{1}{T} \|R\|_2^2 + \varphi \eta \|A\|_2^2 \). Then \( \theta \in (0, 1) \) and

\[
\frac{\varphi}{T} \|Rx^{k+1} - Rx^k\|_2^2 + \varphi \eta \|Ax^{k+1} - Ax^k\|_2^2 \leq (1 - \theta) \|x^{k+1} - x^k\|_2^2.
\]

By (45), we have

\[
\frac{1}{\varphi} \left[ \|x^{k+1} - x^k\|_2^2 \right] - \frac{1}{T} \|Rx^{k+1} - Rx^k\|_2^2 - \eta \|Ax^{k+1} - Ax^k\|_2^2 \\
\leq C_{\lambda, \eta}(x^k) - C_{\lambda, \eta}(x^{k+1}).
\]
Combining (46) and (47), we get

\[
\sum_{k=1}^{n} \|x^{k+1} - x^k\|_2^2 \leq \frac{1}{\theta} \sum_{k=1}^{n} \|x^{k+1} - x^k\|_2^2 \cdot \frac{1}{\theta} \sum_{k=1}^{n} \psi \|Rx^{k+1} - Rx^k\|_2^2
\]

\[
\leq \frac{\psi}{\theta} \sum_{k=1}^{n} \{C_{\lambda,\eta}(x^k) - C_{\lambda,\eta}(x^{k+1})\}
\]

\[
= \frac{\psi}{\theta} (C_{\lambda,\eta}(x^1) - C_{\lambda,\eta}(x^{n+1}))
\]

\[
\leq \frac{\psi}{\theta} C_{\lambda,\eta}(x^1).
\]

Thus, the series \(\sum_{k=1}^{\infty} \|x^{k+1} - x^k\|_2^2\) is convergent, which implies that

\[
\|x^{k+1} - x^k\|_2^2 \to 0 \quad \text{as} \quad k \to \infty.
\]

3) Let \(\{x^{k_j}\}\) be a convergent subsequence of \(\{x^k\}\), and the limit point denoted as \(x^*\), i.e.,

\[
x^{k_j} \to x^* \quad \text{as} \quad k_j \to \infty.
\] (48)

From

\[
\|x^{k_{j+1}} - x^*\|_2 \leq \|x^{k_{j+1}} - x^{k_j}\|_2 + \|x^{k_j} - x^*\|_2
\]

and

\[
\|x^{k_{j+1}} - x^{k_j}\|_2 + \|x^{k_j} - x^*\|_2 \to 0 \quad \text{as} \quad k_j \to \infty,
\]

we get

\[
x^{k_{j+1}} \to x^* \quad \text{as} \quad k_j \to \infty.
\] (49)

By iteration (43), it follows that

\[
x^{k_{j+1}} = G_{\lambda,\eta}(P_a(B_\phi(x^{k_j}))
\]

and combined with Theorem 5, we have

\[
\|x^{k_{j+1}} - B_\phi(x^{k_j})\|_2^2 + \lambda \phi P_a(x^{k_{j+1}}) \leq \|x - B_\phi(x^{k_j})\|_2^2 + \lambda \phi P_a(x).
\]

Taking limit and using the continuity of the function \(P_a\) as well as (48) and (49), we can immediately get that

\[
\|x^* - B_\phi(x^*)\|_2^2 + \lambda \phi P_a(x^*) \leq \|x - B_\phi(x^*)\|_2^2 + \lambda \phi P_a(x)
\]

for any \(x \in \mathbb{R}^n\), which implies that \(x^*\) minimizes the function

\[
\|x - B_\phi(x^*)\|_2^2 + \lambda \phi P_a(x),
\] (50)

and we can conclude that

\[
x^* = G_{\lambda,\eta}(P_a(B_\phi(x^*))
\]

This completes the proof.
An important question we should face is that the solutions of a regularization problem depends seriously on the setting of the regularization parameter $\lambda$, and the selection of proper regularization parameters is always a very hard problem. In most and general cases, a "trial and error" method, say, the cross-validation method, is still an accepted, or even unique, choice. Nevertheless, when some prior information is known for a problem, it is realistic to set the regularization parameter more reasonably and intelligently.

To make this clear, let us suppose that the portfolio is required to be $r$-sparsity, that is, the portfolio should consist of $r$ assets. Let $x^\ast$ be the optimal solution to the regularization problem $(FP_{a,\lambda,\eta})$ and $|B_\varphi(x^\ast)|_i$ be the $i$-th largest value among the absolute elements of $B_\varphi(x^\ast)$. Without loss of generality, we set

$$|B_\varphi(x^\ast)|_1 \geq |B_\varphi(x^\ast)|_2 \geq \cdots \geq |B_\varphi(x^\ast)|_n.$$  

Then, the following inequalities hold:

$$|B_\varphi(x^\ast)|_i > t^*_{a,\lambda,\varphi} \Leftrightarrow i \in \{1, 2, \cdots, r\};$$

$$|B_\varphi(x^\ast)|_j \leq t^*_{a,\lambda,\varphi} \Leftrightarrow j \in \{r + 1, r + 2, \cdots, n\},$$

where $t^*_{a,\lambda,\varphi}$ is the threshold value which is defined in Lemma 2 which obtained by replacing $\lambda$ with $\lambda\varphi$ in $t^*_{a,\lambda}$.

According to $t^2_{a,\lambda,\varphi} \leq t^1_{a,\lambda,\varphi}$, we have

$$\begin{cases} |B_\varphi(x^\ast)|_r \geq t^*_{a,\lambda,\varphi} \geq t^2_{a,\lambda,\varphi} = \sqrt{\lambda\varphi} = \frac{1}{2a}; \\ |B_\varphi(x^\ast)|_{r+1} < t^*_{a,\lambda,\varphi} \leq t^1_{a,\lambda,\varphi} = \frac{\lambda\varphi}{4a}. \end{cases}$$  \hspace{1cm} (51)

which implies

$$\frac{2|B_\varphi(x^\ast)|_{r+1}}{a\varphi} \leq \lambda \leq \frac{(2a|B_\varphi(x^\ast)|_r + 1)^2}{4a^2\varphi}. \hspace{1cm} (52)$$

For convenience, we denote by $\lambda_1$ and $\lambda_2$ the left and the right of above inequality respectively. And a choice of $\lambda$ is

$$\lambda = \begin{cases} \lambda_1, & \text{if } \lambda_1 \leq \frac{1}{a\varphi}; \\ \lambda_2, & \text{if } \lambda_1 > \frac{1}{a\varphi}. \end{cases}$$

In practice, we approximate $x^\ast$ by $x^k$ in (52), say, we can take

$$\lambda = \begin{cases} \lambda_{1,k} = \frac{2|B_\varphi(x^\ast)|_{r+1}}{a\varphi}, & \text{if } \lambda_{1,k} \leq \frac{1}{a\varphi}; \\ \lambda_{2,k} = (\frac{(2a|B_\varphi(x^\ast)|_r + 1))^2}{4a^2\varphi}, & \text{if } \lambda_{1,k} > \frac{1}{a\varphi}. \end{cases}$$  \hspace{1cm} (53)

in applications.

When doing so, the IFPT algorithm will be adaptive and free from the choice of regularization parameter.

Notice that (53) is valid for any $\varphi$ satisfying

$$0 < \varphi < \frac{1}{\frac{1}{2}\|R\|^2 + \eta\|A\|^2}.$$
In general, we can take
\[ \varphi = \frac{1 - \varepsilon}{\|R\|^2 + \eta\|A\|^2} \]
with any small \( \varepsilon \in (0, 1) \) below.

There is one more thing needed to be mentioned that the threshold value
\[ t_{a, \lambda \varphi}^* = \begin{cases} \frac{\lambda a}{\sqrt{\lambda}} - \frac{1}{2a}, & \text{if } \lambda = \lambda_{2,k}; \\ \frac{\lambda a}{\sqrt{\lambda}} - \frac{1}{2a}, & \text{if } \lambda = \lambda_{1,k}. \end{cases} \]

**Algorithm 1**: IFPT algorithm

Initialize: Choose \( x_0 \in \mathbb{R}^n, \varphi = \frac{1 - \varepsilon}{\|R\|^2 + \eta\|A\|^2} \) and \( a = a_0 \) (\( a_0 \) is a given positive number);

while not converged do

Evaluate \( B_{\varphi}(x^k) = x^k + R^T(\beta e_T - Rx^k) + \varphi \eta A^T(b - Ax^k) \);

\[ \lambda_{1,k} = \frac{2\|B_{\varphi}(x^k)\|_1 + 1}{2\|B_{\varphi}(x^k)\|_2 + 1}; \quad \lambda_{2,k} = \frac{(2\|B_{\varphi}(x^k)\|_1 + 1)^2}{4\|B_{\varphi}(x^k)\|_2^2}; \]

if \( \lambda_{1,k} \leq \frac{1}{\sqrt{\lambda}} \) then

\[ \lambda = \lambda_{1,k}; \quad t_{a, \lambda \varphi}^* = \frac{\lambda a}{\sqrt{\lambda}} \]

for \( i = 1 : n \)

1. \( |B_{\varphi}(x^k)_i| > t_{a, \lambda \varphi}^* \), then \( x_{k+1}^i = g_{\lambda \varphi}(B_{\varphi}(x^k)_i); \)
2. \( |B_{\varphi}(x^k)_i| \leq t_{a, \lambda \varphi}^* \), then \( x_{k+1}^i = 0; \)

else

\[ \lambda = \lambda_{2,k}; \quad t_{a, \lambda \varphi}^* = \sqrt{\lambda} - \frac{1}{2a} \]

for \( i = 1 : n \)

1. \( |B_{\varphi}(x^k)_i| > t_{a, \lambda \varphi}^* \), then \( x_{k+1}^i = g_{\lambda \varphi}(B_{\varphi}(x^k)_i); \)
2. \( |B_{\varphi}(x^k)_i| \leq t_{a, \lambda \varphi}^* \), then \( x_{k+1}^i = 0; \)

end

\( k \rightarrow k + 1 \)

end while

return: \( x_{k+1} \)

---

### 4.2 INFPT algorithm for solving \((FP_{a, \lambda, \eta} \geq \_)}\)

Inspired by the IFPT algorithm given in subsection 4.1, we propose the INFPT algorithm to solve the problem \((FP_{a, \lambda, \eta} \geq \_)}\) for all \( a > 0 \).

**Definition 2** Given any vector \( v \in \mathbb{R}^n \), define the projection map on \( \mathbb{R}^n_+ \) by
\[ \mathcal{P}_+(v) \triangleq \arg \min_{\vartheta \in \mathbb{R}^n_+} \{ \|\vartheta - v\|^2 : \vartheta \geq 0 \} = \max\{0, v\}. \]

(54)

**Theorem 10** Let \( v \in \mathbb{R}^n \). Then
\[ \mathcal{G}_{\lambda, P_{\lambda}}(\mathcal{P}_+(v)) \triangleq \arg \min_{x \in \mathbb{R}^n} \left\{ \|x - v\|^2 + \lambda P_{\lambda}(x) : x \geq 0 \right\} \]

where \( \mathcal{G}_{\lambda, P_{\lambda}} \) is defined in Definition 1 and \( \mathcal{P}_+ \) is defined in Definition 2.
Proof Given any vector $v \in \mathbb{R}^n$, let us introduce the following notations

$$x_+ = x_{I^+} \quad \text{and} \quad x_- = x_{I^-},$$

where

$$I^+ = \{ i \mid i \in \{1, 2, \ldots, n\}, \ x_i \geq 0 \}$$

and

$$I^- = \{ i \mid i \in \{1, 2, \ldots, n\}, \ x_i < 0 \}.$$

Observe that the following relations hold

(i) $\|x\|_2^2 = \|x_+\|_2^2 + \|x_-\|_2^2$;

(ii) $\|(x - v)_+\|_2^2 + \|(x - v)_-\|_2^2 = \|x - P_+(v)\|_2^2$;

(iii) $\|x_-\|_2^2 = 0 \iff x_i = 0 \ \forall i \in I^-$,

where the second relation follows from relation (i) and the fact that $(P_+(v))_i = v_i$ for any $i \in I^+$ and $(P_+(v))_i = 0$ for any $i \in I^-$. By these facts, we can get that

$$\hat{x} = G_{\lambda, p_+}(P_+(v))$$

if and only if

$$\hat{x} = \arg \min_{x \in \mathbb{R}^n} \left\{ \|x - v\|_2^2 + \lambda P_+(x) : \ x \geq 0 \right\}$$

$$= \arg \min_{x \in \mathbb{R}^n} \left\{ \|(x - v)_+\|_2^2 + \|(x - v)_-\|_2^2 + \lambda P_+(x) : \ x \geq 0 \right\}$$

$$= \arg \min_{x \in \mathbb{R}^n} \left\{ \|(x - v)_+\|_2^2 + \|(x - v)_-\|_2^2 - 2 \sum_{i \in I^-} x_i v_i + \lambda P_+(x) : \ x \geq 0 \right\}$$

$$= \arg \min_{x \in \mathbb{R}^n} \left\{ \|(x - v)_+\|_2^2 + \lambda P_+(x) : \ x_i = 0 \ \forall i \in I^-, \ x \geq 0 \right\}$$

$$= \arg \min_{x \in \mathbb{R}^n} \left\{ \|(x - v)_+\|_2^2 + \lambda P_+(x) : \ \|x_-\|_2^2 = 0 \right\}$$

$$= \arg \min_{x \in \mathbb{R}^n} \left\{ \left(\|(x - v)_+\|_2^2 + \|x_-\|_2^2\right) + \lambda P_+(x) \right\}$$

$$= \arg \min_{x \in \mathbb{R}^n} \left\{ \|x - P_+(v)\|_2^2 + \lambda P_+(x) \right\}$$

$$= G_{\lambda, p_+}(P_+(v)).$$

This completes the proof.

By Theorem 10 and inspired by iteration (43), the procedure of the INFPT algorithm for solving the regularization problem $(FP^Z_{a,\lambda,\eta})$ can be inductively defined as

$$x^{k+1} = G_{\lambda, p_+}(P_+(B_\varphi(x^k)))$$

where $B_\varphi(x^k) = x^k + \lambda R^T(\beta e_T - Rx^k) + \varphi A^T(b - Ax^k)$.

The difference between INFPT algorithm and IFPT algorithm is that the operator $G_{\lambda, p_+}(P_+(\cdot))$ acts only on the nonnegative part of the real line since all the $P_+(\cdot)$ are nonnegative.
In addition, the regularization parameter \( \lambda \) in INFPT algorithm can be selected as similarly as the IFPT algorithm. When doing so, the INFPT algorithm will also be adaptive and free from the choice of the regularization parameter.

Similarly, we suppose that the nonnegative vector \( x^*_\tilde{r} \) of sparsity \( \tilde{r} \) is the optimal solution to the regularization problem \((FP^\geq_{\alpha,\lambda,\eta})\), and \(|P_+(B_\varphi(x^k))|_i\) be the \( i \)-th largest value among the elements of the nonnegative vector \( P_+(B_\varphi(x^k)) \).

Without loss of generality, we set
\[
|P_+(B_\varphi(x^k))|_1 \geq |P_+(B_\varphi(x^k))|_2 \geq \cdots \geq |P_+(B_\varphi(x^k))|_n \geq 0. \tag{57}
\]

Then the following inequalities hold:
\[
|P_+(B_\varphi(x^k))|_i > \tilde{t}^*_a,\lambda \varphi \Leftrightarrow i \in \{1, 2, \cdots, \tilde{r}\},
\]
\[
|P_+(B_\varphi(x^k))|_j \leq \tilde{t}^*_a,\lambda \varphi \Leftrightarrow j \in \{
\tilde{r} + 1, \tilde{r} + 2, \cdots, n\}.
\]

In accordance with the selection of the regularization parameter in IFPT algorithm, the optimal regularization parameter for the INFPT algorithm can be selected as
\[
\lambda = \begin{cases} 
\lambda_{1,k} = \frac{2(\lambda_1 P_+(B_\varphi(x^k)))_{\tilde{r}+1}}{2\alpha_\varphi + (\lambda_1 P_+(B_\varphi(x^k)))_{\tilde{r}+1}}, & \text{if } \lambda_{1,k} \leq \frac{1}{\alpha_\varphi}; \\
\lambda_{2,k} = \frac{2(\lambda_1 P_+(B_\varphi(x^k)))_{\tilde{r}+1}^2}{4\alpha_\varphi}, & \text{if } \lambda_{1,k} > \frac{1}{\alpha_\varphi}.
\end{cases} \tag{58}
\]

We also take
\[
\varphi = \frac{1 - \varepsilon}{\frac{1}{\|R\|^2 + \eta\|A\|^2}}
\]
with any small \( \varepsilon \in (0, 1) \) below, and
\[
\tilde{t}^*_a,\lambda \varphi = \begin{cases} 
\frac{\lambda_1 \alpha_\varphi}{\sqrt{\lambda \varphi}} - \frac{\lambda_1 \alpha_\varphi}{2\alpha_\varphi}, & \text{if } \lambda = \lambda_{1,k}; \\
\frac{\lambda_1 \alpha_\varphi}{\sqrt{\lambda \varphi}} - \frac{\lambda_1 \alpha_\varphi}{2\alpha_\varphi}, & \text{if } \lambda = \lambda_{2,k}.
\end{cases}
\]

5 Numerical experiments

In this section, we apply the IFPT algorithm and INFPT algorithm described above to construct the optimal (sparse) portfolios with and without short-selling constraints, and carry out a series of simulations to evaluate their out-of-sample performance. The tests and comparisons are performed on two sets of portfolios from Fama and French web site\(^1\): the 48 industry portfolios (FF48) and 100 portfolios formed on size and book-to-market (FF100), ranging from July 1976 to June 2006. Tests use these two real market data during a period of 30 years from July 1976 to June 2006, and the time period is divided into 6 equal sub-periods. To determine \( R \) and \( \mu \), we use the historical returns from July 1971 until June 1976. We then solve the optimization problems using
\[^1\]http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
Algorithm 2: INFPT algorithm

Initialize: Choose $x^0 \in \mathbb{R}^n, \varphi = \frac{1}{\sqrt{\|R\|_2^2 + \eta\|A\|_2^2}}$ and $a = a_0$ ($a_0$ is a given positive number);

while not converged do

$P_+(B_{\varphi}(x^k)) = P_+(x^k + \frac{\varphi}{2}R^T(b_{k} - Rx_k) + \varphi A^T(Ax_k))$;

$\lambda_{1,k} = \frac{2|P_+(B_{\varphi}(x^k))|_{r+1}}{a\varphi}$;

if $\lambda_{1,k} \leq \frac{1}{\sqrt{a^2 \varphi}}$ then

$\lambda = \lambda_{1,k}; t_{a,\lambda \varphi} = \frac{\lambda \varphi}{2a}$

for $i = 1 : n$

1. $P_+(B_{\varphi}(x^k))_i > t_{a,\lambda \varphi},$ then $x_{k+1}^i = g_{\lambda \varphi}(P_+(B_{\varphi}(x^k)))_i$;

2. $P_+(B_{\varphi}(x^k))_i \leq t_{a,\lambda \varphi},$ then $x_{k+1}^i = 0$;

else

$\lambda = \lambda_{2,k}; t_{a,\lambda \varphi} = \max\{\sqrt{\lambda \varphi} - \frac{1}{2a}, 0\}$

for $i = 1 : n$

1. $P_+(B_{\varphi}(x^k))_i > t_{a,\lambda \varphi},$ then $x_{k+1}^i = g_{\lambda \varphi}(P_+(B_{\varphi}(x^k)))_i$;

2. $P_+(B_{\varphi}(x^k))_i \leq t_{a,\lambda \varphi},$ then $x_{k+1}^i = 0$;

end

$k \rightarrow k + 1$

end while

return: $x^{k+1}$

this matrix and vector, targeting an annualized return $\beta$, equal to the average historical return, from July 1971 until June 1976, obtained by a portfolio in which all industry sectors are given the equal weight $1/n$. The performance of each portfolio is evaluated by looking at its out-of-sample total return, $m$, its out-of-sample variance, $\sigma$, and its out-of-sample Sharpe ratio, $S = \frac{m}{\sigma}$.

In order to understand the effect of sparsity on the performance of resulting portfolios, the value of $k$ in the following tables specifies the number of assets in a portfolio. In all the experiments, we set $a = 1$.

Firstly, we present the numerical results of IFPT algorithm in FF48 problem with short-selling constraint, and compare them with those obtained with the $\ell_1$-norm regularization portfolio selection model (solved by LARS algorithm [8,17]). For the sake of simplicity, we renamed the $\ell_1$-norm regularization portfolio selection model (5) solved by LARS algorithm as $\text{La}(\ell_1)$. And then we show the performance of INFPT algorithm in finding the sparse portfolio weights in FF100 problem without short-selling constraint.

Tables 1 and 2 report the numerical results of IFPT algorithm and $\text{La}(\ell_1)$ in FF48 problem with short-selling constraint and vary $k$ from 6 to 20 with step size 2. The numerical results show that the performance of the IFPT algorithm is better than the performance of $\text{La}(\ell_1)$ in all periods except in the period 07/76-06/06. Table 3 reports the performance of INFPT algorithm in problem FF100 without short-selling constraint. It can be observed from Table 3 that the INFPT algorithm performs effectively in finding the sparse portfolio weights in FF100 problem without short-selling constraint.
Sparse Portfolio Selection via Non-convex Fraction Function

| Item | k = 6 | k = 8 | k = 10 | k = 12 |
|------|-------|-------|--------|--------|
|       | IFPT (ℓ₁) | IFPT (ℓ₁) | IFPT (ℓ₁) | IFPT (ℓ₁) |
| 07/76-06/81 | 9.09 | 7.09 | 9.17 | 9.81 |
| 07/81-06/86 | 7.35 | 6.14 | 8.20 | 8.37 |
| 07/86-06/91 | 3.19 | 3.49 | 3.72 | 3.90 |
| 07/91-06/96 | 8.36 | 8.86 | 9.49 | 9.48 |
| 07/96-06/01 | 3.72 | 4.17 | 4.54 | 4.86 |
| 07/01-06/06 | 1.61 | 1.63 | 2.48 | 2.66 |

Table 1 Comparison results of Sharpe Ratio S between IFPT algorithm and La(ℓ₁) with k = 6, 8, 10, 12 in problem FF48 with short-selling constraint.

| Item | k = 14 | k = 16 | k = 18 | k = 20 |
|------|-------|-------|--------|--------|
|       | IFPT (ℓ₁) | IFPT (ℓ₁) | IFPT (ℓ₁) | IFPT (ℓ₁) |
| 07/76-06/81 | 10.33 | 10.03 | 9.49 | 9.24 |
| 07/81-06/86 | 6.59 | 7.13 | 6.79 | 6.60 |
| 07/86-06/91 | 3.83 | 3.92 | 3.58 | 3.52 |
| 07/91-06/96 | 9.22 | 9.36 | 9.06 | 9.29 |
| 07/96-06/01 | 3.75 | 3.15 | 3.47 | 3.31 |
| 07/01-06/06 | 2.61 | 3.00 | 2.56 | 2.34 |
| 07/76-06/06 | 23.86 | 28.28 | 28.58 | 30.06 |

Table 2 Comparison results of Sharpe Ratio S between IFPT algorithm and La(ℓ₁) with k = 14, 16, 18, 20 in problem FF48 with short-selling constraint.

| Period | k = 6 | k = 8 | k = 10 | k = 12 | k = 14 | k = 16 | k = 18 | k = 20 |
|--------|-------|-------|--------|--------|-------|-------|--------|--------|
| 07/76-06/81 | 3.41 | 3.52 | 3.59 | 3.60 | 3.60 | 3.65 | 3.69 | 3.79 |
| 07/81-06/86 | 4.50 | 4.68 | 4.70 | 4.71 | 4.96 | 4.99 | 5.02 |
| 07/86-06/91 | 2.14 | 2.11 | 2.09 | 2.07 | 2.07 | 2.07 | 2.07 |
| 07/91-06/96 | 12.35 | 12.61 | 12.59 | 12.60 | 12.70 | 12.74 | 12.76 |
| 07/96-06/01 | 3.69 | 4.17 | 4.17 | 4.14 | 4.13 | 4.05 | 4.01 |
| 07/01-06/06 | 1.47 | 1.38 | 1.44 | 1.52 | 1.74 | 1.76 | 2.02 |
| 07/76-06/06 | 22.12 | 22.75 | 22.87 | 23.05 | 23.41 | 23.50 | 24.81 |

Table 3 The performance of INFPT algorithm in problem FF100 without short-selling constraint with k = 6, 8, 10, 12, 14, 16, 18, 20.

6 Conclusions

The sparsity requirement in portfolio selection problems comes from the real world practice, where the administration of a portfolio made up of a large number of assets, possibly with very small holdings for some of them, is clearly not desirable because of the transactions costs and the complexity of management. In this paper, a continuous and non-convex sparsity promoting fraction function is studied in two sparse portfolio selection models with and without short-selling constraints in terms of theory, algorithms and computation. Firstly, we study the properties of the optimal solution to the problem \((FP_{n,\lambda,\eta})\) including the first-order and the second optimality condition and the lower and upper bound of the absolute value for its nonzero entries. Secondly, the IFPT algorithm is proposed to solve the problem \((FP_{n,\lambda,\eta})\) for all \(a > 0\). Moreover, we also prove that the value of the regularization parameter \(\lambda > 0\) can not be
chosen too large. Indeed, there exists $\lambda > 0$ such that the optimal solution to the problem $(FP_{a,\lambda,\eta})$ is equal to zero for any $\lambda > \lambda$. At last, inspired by the thresholding representation of the IFPT algorithm, the INFPT algorithm is proposed to solve the problem $(FP_{a,\lambda,\eta})$ for all $a > 0$. Empirical results show that our methods perform effective in finding the sparse portfolio weights in FF48 and FF100 problems with and without short-selling constraints.

Acknowledgements We would like to thank editorial and referees for their comments which help us to enrich the content and improve the presentation of the results in this paper. The work was supported by the National Natural Science Foundations of China (11771347, 11131006, 41390450).

References

1. H. M. Markowitz, Portfolio selection. Journal of Finance, 7(1): 77-91, 1952.
2. R. Armañanzas, J. A. Lozano, A multiobjective approach to the portfolio optimization problem. Proceedings of the 2005 IEEE Congress on Evolutionary Computation, 2: 1388-1395, 2005.
3. D. Bertsimas, R. Shioda, Algorithms for cardinality-constrained quadratic optimization. Computational Optimization and Applications, 43: 1-22, 2009.
4. D. Bienstock, Computational study of a family of mixed-integer quadratic programming problems. Mathematical Programming, 74: 121-140, 1996.
5. T. Chang, N. Meade, J. E. Beasley, and Y. M. Sharaiha, Heuristics for cardinality constrained portfolio optimization. Computers and Operations Research, 27(15): 1271-1302, 2000.
6. Y. Crama, M. Schyns, Simulated annealing for complex portfolio selection problems. European Journal of Operational Research, 150(3): 546-571, 2003.
7. J.E. Fieldsend, J. Matafko, and M. Peng, Cardinality constrained portfolio optimisation, Lecture Notes in Computer Science (LNCS 3177), Z.R. Yang, R. Everson and H. Yin (Eds.), Springer, 2004, pp. 788-793.
8. J. Brodie, I. Daubechies, C. D. Molc, D. Giannoned, and I. Lorisce. Sparse and stable markowitz portfolios. European Central Bank, Working Paper Series, 2008.
9. J. Y. Gotoh, A. Takeda, On the role of norm constraints in portfolio selection. Computational Management Science, 8: 323-353, 2011.
10. Y. M. Yena, T. J. Yen, Solving norm constrained portfolio optimization via coordinate-wise descent algorithms. Computational Statistics and Data Analysis, 76: 737-759, 2014.
11. D. Lorenzo, G. Liuzzi, F. Rinaldi, F. Schoen and M. Sciandrone, A concave optimization based approach for sparse portfolio selection. Optimization Method and Software, 27: 983-1000, 2012.
12. Y. Teng, L. Yang, B. Yu and X. Song, Penalty PALM Method for Sparse Portfolio Selection Problems, 32(1): 126C147, 2017.
13. I. Daubechies, M. Defrise and D. M. Christopher, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Communications on Pure and Applied Mathematics, 57(11): 1413-1457, 2004.
14. D. Geman, G. Reynolds, Constrained restoration and recovery of discontinuities. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(3): 367-383, 1992.
15. H. Li, Q. Zhang, A. Cui and J. Peng. Minimization of fraction function penalty in compressed sensing. https://arxiv.org/pdf/1705.06048.pdf.
16. F. Xing, Investigation on solutions of cubic equations with one unknown. Journal of the Central University for Nationalities (Natural Sciences Edition), 12(3): 207-218, 2003.
17. B. Efron, T. Hastie, T. Johnstone and R. Tibshirani, Least angle regression. The Annals of Statistics, 32(2): 407-499, 2004.