# Robust and Sparse Portfolio: Optimization Models and Algorithms 

Hongxin Zhao ${ }^{1, *}$, Yilun Jiang ${ }^{2}$ and Yizhou Yang ${ }^{3}$<br>1 School of Mathematics and Statistics, Beijing Jiaotong University, Beijing 100044, China<br>2 Department of Economic Management, Shijiazhuang Institute of Railway Technology, Shijiazhuang 050000, China; yljiangstat@163.com<br>3 Personnel Department, Shijiazhuang University, Shijiazhuang 050035, China; joeyyangnz@gmail.com<br>* Correspondence: hxzhao@bjtu.edu.cn

Citation: Zhao, H.; Jiang, Y.; Yang, Y. Robust and Sparse Portfolio: Optimization Models and Algorithms. Mathematics 2023, 11, 4925. https://doi.org/10.3390/ math11244925

Academic Editors: Jing Yao, Xiang
Hu and Jingchao Li
Received: 8 November 2023
Revised: 7 December 2023
Accepted: 8 December 2023
Published: 11 December 2023


Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).


#### Abstract

The robust and sparse portfolio selection problem is one of the most-popular and -frequently studied problems in the optimization and financial literature. By considering the uncertainty of the parameters, the goal is to construct a sparse portfolio with low volatility and decent returns, subject to other investment constraints. In this paper, we propose a new portfolio selection model, which considers the perturbation in the asset return matrix and the parameter uncertainty in the expected asset return. We define three types of stationary points of the penalty problem: the Karush-KuhnTucker point, the strong Karush-Kuhn-Tucker point, and the partial minimizer. We analyze the relationship between these stationary points and the local/global minimizer of the penalty model under mild conditions. We design a penalty alternating-direction method to obtain the solutions. Compared with several existing portfolio models on seven real-world datasets, extensive numerical experiments demonstrate the robustness and effectiveness of our model in generating lower volatility.


Keywords: portfolio optimization; robustness; sparsity; uncertainty set; penalty-alternating-direction method

MSC: 91G10; 90C90; 90C30

## 1. Introduction

In 1952, Harry M. Markowitz [1] published the classic "Portfolio Selection" in The Journal of Finance, which ushered in a new era of financial mathematical analysis. Markowitz pointed out that investors who care about return and risk should hold portfolios located at the efficient boundary of mean-variance, which is the famous mean-variance portfolio (MVP) selection model. Since then, many portfolio selection strategies have been proposed by referring to the MVP and its variants. However, MVPs exhibit instability due to estimation errors in the input parameters [2], especially in large-scale conditions. The instability means that the solution obtained under sample fluctuation may be optimal for a given sample, but it is not optimal from the perspective of risk. For more comments on this model, we refer to [3-6] and the references therein.

This paper focuses attention on sample fluctuations and parameter uncertainty in the portfolio selection problem. We now review some relevant methods for the parameter uncertainty. Among various approaches, the attractive one is the robust portfolio (RP), which corresponds to a robust optimization, since it does not use any information about the probability distribution of the uncertain parameters. RP we considered is a conservative approach that minimizes the loss function within an uncertainty set and then solves the problem under the worst-case scenario. In the last two decades, robust portfolio selection problems have gained the increasing interest of researches. These researches constructed well-known optimal portfolios from the perspective of robust optimization [7-10]. In this way, Goldfarb and Iyengar [11] formulated and solved RP problems. They introduced the
uncertainty structures for the input parameters, then they showed that the RP problems corresponding to the second-order cone programs and these uncertainty structures correspond to confidence regions employed to estimate the market parameters. Given the uncertainty in the mean and covariance matrix of the asset return, Lobo and Boyd [12] computed the maximum risk of a portfolio in a numerically efficient way. They proved that this is a semi-definite programming problem and is readily solved by interior-point methods for convex optimization. Min et al. [13] proposed the hybrid RP models under ellipsoidal uncertainty sets, and they considered both the best-case and the worst-case counterparts. Won and Kim [14] considered RP problems involving a trade-off between the worst-case utility and the worst-case regret, or the largest difference between the best utility achievable under the model and that achieved by a given portfolio. They showed that the entire optimal trade-off curve can be found via solving a series of semi-definite programs under the ellipsoidal uncertainty model. Some research works [15,16] concentrated on the application of robust optimization on basic mean-variance, mean value-at-risk (mean-VaR), and mean conditional-value-at-risk (mean-CVaR) problems, but did not consider variants of the problem like robust index tracking, robust and sparse portfolio selection problems, and so on. More relevant works can be found in [17-20] and the references therein.

RPs have a wide range of applications, among these, one essential step is the construction of uncertainty sets. Two types of uncertainty sets are widely used, namely the box uncertainty set and the ellipsoidal uncertainty set. Tütüncü and Koenig [21] used symmetric box uncertainty sets defined as $U_{\mu}=\left\{\mu \in \mathbb{R}^{n} \mid \mu^{L} \leq \mu \leq \mu^{U}\right\}$ and $U_{\Sigma}=\left\{\Sigma \in \mathbb{R}^{n \times n} \mid \Sigma^{L} \leq \Sigma \leq \Sigma^{U}, \Sigma \succeq 0\right\}$, where $\mu^{L} \in \mathbb{R}^{n}$ and $\mu^{U} \in \mathbb{R}^{n}$ are the lower and upper bounds of mean vector $\mu, \Sigma^{L} \in \mathbb{R}^{n \times n}$ and $\Sigma^{U} \in \mathbb{R}^{n \times n}$ are the lower and the upper bounds of the covariance matrix $\Sigma$, respectively, and $\Sigma$ is positive semi-definite. Khodamoradi et al. [22] used box uncertainty sets for a cardinal-constrained mean-variance portfolio problem which allows short selling. Swain and Ojha [10] analyzed the robust version of the mean-variance portfolio problem and mean-semi-variance portfolio problem with box uncertainty sets. Alternatively, Fabozzi et al. [23] defined an ellipsoidal uncertainty set for the expected asset return as $U_{\mu}=\left\{\mu \mid(\mu-\bar{\mu}) \Sigma^{-1}(\mu-\bar{\mu})^{\top} \leq \epsilon^{2}\right\}$, where $\bar{\mu}$ is the nominal asset return and $\epsilon^{2}$ is a small scalar, which controls the size of the uncertainty set. However, they did not consider the uncertainty of the covariance matrix, thus the solution was robust only against perturbations in the asset return vector. Pinar [24] developed a multi-period robust mean-variance portfolio problem with an ellipsoidal uncertainty set while allowing short selling. As we all know, the estimation error is more sensitive to the mean vector than the covariance matrix. On the other hand, dealing with the uncertainty in the covariance matrix is more complicated than dealing with the uncertainty set of the mean vector. Thus, in this paper, we consider two types of uncertainty sets for the mean vector.

Financial data have some remarkable features, such as multicollinearity and a heavy tail. Therefore, the perturbations of these data should not be underestimated. By referring to Brodie et al. [2], who transferred the MVP into a Lasso-type portfolio, we consider the perturbations in the asset return matrix and design its uncertainty set. In addition, from the perspective of transaction costs and administrative expenses, more assets are not always better. Therefore, it is also necessary to consider sparsity when constructing a portfolio [25-27]. After these discussions, a natural question follows: How do we find better RPs that not only reduce the undesired impact of parameter uncertainty, but also improve sparsity and reduce cost?

Following the above considerations, this paper proposes a sparsity constrained robust portfolio optimization model with parameter uncertainty and data perturbation. Specifically, we consider the perturbation in the asset return matrix and the parameter uncertainty in the expected asset return. By using the equivalence of robustness and regularization, the Lasso-type objective function can be converted into the sum of a square root and the $\ell_{1}$ norm. We consider two kinds of uncertainty sets: the box uncertainty set and the ellipsoidal uncertainty set. For its penalty model, we define three types of stationary points:
the Karush-Kuhn-Tucker (KKT) point, the strong KKT point, and the partial minimizer. Under mild constraint qualification (CQ), we prove that any local minimizer of the penalty model is a KKT point. Moreover, the global minimizer of the penalty model is proven to be a partial minimizer and, then, a stronger KKT point under Slater's CQ. Finally, a penalty alternating direction method is proposed to obtain a portfolio, and its convergence is established. We confirm the effectiveness of our approach by comparing with nine widely studied portfolio models on seven real-world data sets. The numerical results show that the portfolios we proposed have less volatility, that is less risk. Moreover, our portfolio strategies can yield higher Sharpe ratios when the appropriate parameters are selected.

This paper is organized as follows. Some notations and preliminaries used in this paper are given in the next section. The model of robust and sparse portfolios and the analysis of their optimization theory are stated in Section 3. Two types of uncertainty sets of mean vectors are presented in Section 4 . The optimization algorithm named the penalty alternating direction method is established in Section 5. Extensive numerical experiments are conducted in Section 6. Conclusions are drawn in Section 7.

## 2. Notations and Preliminary

We use $\mathbb{R}$ and $\mathbb{R}^{n}$ and $\mathbb{R}^{m \times n}$ to denote the set of real numbers and the $n$-dimensional and $m \times n$-dimensional Euclidean space. We use boldfaced small letters to denote vectors, e.g., $\mathbf{w} \in \mathbb{R}^{n}$ is a column vector with $n$ elements $w_{i}, i=1, \ldots, n$. The transpose of $\mathbf{w}$ is denoted as $\mathbf{w}^{T}$, which is a row vector. In particular, $\mathbf{1}_{n}$ is the vector of all ones of size $n$. For a vector $\mathbf{a} \in \mathbb{R}^{n}$, we define its absolute value vector by $|\mathbf{a}|:=\left(\left|a_{1}\right|, \cdots,\left|a_{n}\right|\right)$. We use capital letters to denote matrices, e.g., $A \in \mathbb{R}^{m \times n}$ and $a_{i j}$ denote the $(i, j)$-th entry of $A$. Given an index $\Gamma \subset\{1, \ldots, n\}, \mathbf{a}_{\Gamma}$ denotes the sub-vector of $\mathbf{a}$. We write the Euclidean norm of $\mathbf{w}$ by $\|\mathbf{w}\|_{2}$, the $\ell_{1}$ norm by $\|\mathbf{w}\|_{1}$, and the infinity norm by $\|\mathbf{w}\|_{\infty}$. For two vectors $\mathbf{a} \in \mathbb{R}^{n}$ and $\mathbf{b} \in \mathbb{R}^{n},\langle\mathbf{a}, \mathbf{b}\rangle$ denotes the standard inner product.

We now provide some existing results of optimization that are crucial for the theory of this paper. For the convenience of expression, we define the following convex programming:

$$
\begin{array}{rl}
\min _{\mathbf{x} \in \mathbb{R}^{n}} & f(\mathbf{x})  \tag{1}\\
\text { s.t. } & g_{i}(\mathbf{x}) \geq 0, i=1, \ldots, m, \mathbf{x} \in \Omega
\end{array}
$$

where $\Omega$ is a nonempty convex set, $f$ is a convex function, and the $g_{i}(\mathbf{x})$ s are concave functions. For problem (1), Slater's CQ builds a bridge between its solution and the KKT point (the point satisfying the conditions in Theorem 1).

Definition 1 ([28], Definition 4.17). Slater's $C Q$ holds in problem (1) if there exists $\mathbf{u} \in \Omega$ such that $g_{i}(u)>0$ for all $i=1, \ldots, m$.

Theorem 1 ([28], Theorem 4.18). Suppose that Slater's CQ holds in problem (1). Then, $\mathbf{x}^{*}$ is an optimal solution to problem (1) if and only if there exist non-negative Lagrange multipliers $\left(\lambda_{1}, \ldots, \lambda_{m}\right) \in \mathbb{R}^{m}$ such that

$$
0 \in \partial f\left(\mathbf{x}^{*}\right)-\sum_{i=1}^{m} \lambda_{i} \partial g_{i}\left(\mathbf{x}^{*}\right)+N\left(\mathbf{x}^{*} ; \Omega\right)
$$

and $\lambda_{i} \partial g_{i}\left(\mathbf{x}^{*}\right)=0$ for all $i=1, \ldots, m$, where $\partial f\left(\mathbf{x}^{*}\right)$ denotes the classical sub-differential set ([28], Definition 2.30) of $f$ at $\mathbf{x}^{*}$ and $N\left(\mathbf{x}^{*} ; \#\right)$ denotes the classical normal cone ([28], Definition 2.9) of \# at $\mathbf{x}^{*}$.

We also introduce some crucial terminologies and results for sparsity nonlinear programming:

$$
\begin{array}{rl}
\min _{\mathbf{x} \in \mathbb{R}^{n}} & f(\mathbf{x}) \\
\text { s.t. } & g_{i}(\mathbf{x}) \geq 0, i=1, \ldots, m  \tag{2}\\
& h_{j}(\mathbf{x})=0, j=1, \ldots, l \\
& \|\mathbf{x}\|_{0} \leq s
\end{array}
$$

where $f$ is a convex function and $g$ and $h$ are continuously differentiable. A restricted linear independence constraint qualification (R-LICQ) used for sparsity nonlinear programming (2) was defined by [29] as follows.

Definition 2 ([29], Definition 2.4). We say that the R-LICQ holds at $\mathbf{x}^{*}$, where $\mathbf{x}^{*}$ is feasible for the problem (2):

- When $\left\|\mathbf{x}^{*}\right\|_{0}=s, \nabla g_{i}\left(\mathbf{x}^{*}\right), i \in I\left(\mathbf{x}^{*}\right), \nabla h_{j}\left(\mathbf{x}^{*}\right), j=1, \ldots, l$, are linearly independent.
- When $\left\|\mathbf{x}^{*}\right\|_{0}<s, \nabla_{\Gamma^{*}} g_{i}\left(\mathbf{x}^{*}\right), i \in I\left(\mathbf{x}^{*}\right), \nabla_{\Gamma^{*}} h_{j}\left(\mathbf{x}^{*}\right), j=1, \ldots, l$, are linearly independent.

Based on the R-LICQ, the following decomposition result holds.
Theorem 2 ([29], Proposition 2.5). Let $\mathbf{x}^{*}$ be a feasible point of problem (2) and the R-LICQ hold at $\mathbf{x}^{*}$. Then,

$$
\hat{N}\left(\mathbf{x}^{*} ; S \cap Q\right)=\hat{N}\left(\mathbf{x}^{*} ; S\right)+\hat{N}\left(\mathbf{x}^{*} ; Q\right)
$$

where $S:=\left\{x:\|x\|_{0} \leq s\right\}, Q:=\left\{x: g_{i}(\mathbf{x}) \geq 0, i=1, \ldots, m, h_{j}(x)=0, j=1, \ldots, l\right\}$, and $\hat{N}\left(\mathbf{x}^{*} ; \#\right)$ denotes the Frechét normal cone ([30], Definition 6.3) of \# at $\mathbf{x}^{*}$, which degenerates into the classical norm cone described in Theorem 1 if \# is a convex set.

For the partial problem (10) of the portfolio model (6) in Section 3.1, the R-LICQ holds automatically at $\mathbf{x}^{*}$, where $m=0, l=1$, and $h(\boldsymbol{x}):=\mathbf{1}^{T} \mathbf{x}-1$. Next, we establish the relationship between the local minimizer of problem (2) and its KKT point (the point satisfying the KKT system in Theorem 3).

Theorem 3. Suppose that $\mathbf{x}^{*}$ is a local minimizer of problem (2) and the R-LICQ holds at $\mathbf{x}^{*}$. Then, there exist non-negative Lagrange multipliers $\left(\lambda_{1}^{*}, \ldots, \lambda_{m}^{*}\right) \in \mathbb{R}_{+}^{m}$ and $\left(\mu_{1}^{*}, \ldots, \mu_{m}^{*}\right) \in \mathbb{R}^{l}$ such that

$$
\left\{\begin{array}{l}
0 \in \partial f\left(\mathbf{x}^{*}\right)-\sum_{i=1}^{m} \lambda_{i} \partial g_{i}\left(\mathbf{x}^{*}\right)+\sum_{j=1}^{l} \lambda_{i} \partial h_{j}\left(\mathbf{x}^{*}\right)+\hat{N}\left(\mathbf{x}^{*} ; S\right)  \tag{3}\\
g_{i}(\mathbf{x}) \geq 0, \lambda_{i} g_{i}(\boldsymbol{x})=0, i=1, \ldots, m \\
h_{j}(\mathbf{x})=0, j=1, \ldots, l \\
\|\mathbf{x}\|_{0} \leq s
\end{array}\right.
$$

Proof. It follows from Theorem 6.12 of [30] that

$$
0 \in \partial f\left(\mathbf{x}^{*}\right)+\hat{N}\left(\mathbf{x}^{*} ; S \cap Q\right) .
$$

Combing Theorem 2 with the proof of Theorem 3.2 of [29], this result holds.
This result is different from Theorem [29]. We allow the objective function of problem (2) to be non-differentiable. The analysis process of this result is completely consistent with that of Theorem [29].

## 3. Model and Optimization Theory

In this section, we first propose a robust and sparse portfolio model (4) with an uncertainty set constraint and a sparsity constraint. For the convenience of the numerical
calculation, we consider its $\ell_{1}$ norm penalization variant (6). We define three types of stationary points of the penalization variant: the KKT point, the strong KKT point, and the partial minimizer. The relationships of these stationary points and the local/global minimizer of the penalization problem (6) are established in Section 3.2.

### 3.1. Robust and Sparse Portfolio Model

Consider $n$ risky assets, denoting the asset return at period $t$ by $\mathbf{r}_{t}=\left(r_{1}, \ldots, r_{n}\right)^{\top} \in \mathbb{R}^{n}$. The expected return vector of different assets is denoted by $E\left(\mathbf{r}_{t}\right)=\mu$, and the covariance matrix is denoted by $E\left[\left(\mathbf{r}_{t}-\mu\right)\left(\mathbf{r}_{t}-\mu\right)^{\top}\right]=V$. In the traditional Markowitz portfolio selection problem, the portfolio construction is based on the trade-off between risk and return. For a given level of acceptable portfolio return $\rho=\mathbf{w}^{\top} \mu$, the mean-variance optimization can be formulated as

$$
\min _{\mathbf{w} \in \mathbb{R}^{n}} \frac{1}{2} \mathbf{w}^{\top} V \mathbf{w}, \quad \text { s.t. } \quad \mathbf{w}^{\top} \mu=\rho, \quad \mathbf{w}^{\top} \mathbf{1}_{n}=1
$$

and its aim is to find a portfolio that has minimal risk for a given expected return. A significant model that has been developed from the Markowitz model is the Lasso-type portfolio proposed by Brodie et al. [2], which is given as:

$$
\min _{\mathbf{w} \in \mathbb{R}^{n}} \frac{1}{T}\left\|\rho \mathbf{1}_{T}-R \mathbf{w}\right\|_{2}^{2}+\alpha\|\mathbf{w}\|_{1}, \quad \text { s.t. } \quad \mathbf{w}^{\top} \bar{\mu}=\rho, \quad \mathbf{w}^{\top} \mathbf{1}_{n}=1,
$$

where $\bar{\mu}=\frac{1}{T} \sum_{t=1}^{T} \mathbf{r}_{t}, \alpha$ is the penalty parameter, and $R \in \mathbb{R}^{T \times n}$ is the asset return matrix. Brodie et al. [2] confirmed that the $\ell_{1}$ norm can produce a sparse portfolio, and this method can stabilize the problem. In this paper, we start with the square root Lasso-type portfolio, while adding more consideration about the perturbation in asset return matrix $R$ and the parameter uncertainty in $\mu$. We propose the following robust and sparse portfolio selection model:

$$
\begin{align*}
\min _{\mathbf{w} \in \mathbb{R}^{n}} \max _{\Delta \in U_{0}} & \left\|\rho \mathbf{1}_{T}-(R+\Delta) \mathbf{w}\right\|_{2} \\
\text { s.t. } & \min _{\mu \in U} \mathbf{w}^{T} \mu \geq \rho,  \tag{4}\\
& \mathbf{w}^{T} \mathbf{1}_{n}=1, \mathbf{w} \geq 0, \\
& \|\mathbf{w}\|_{0} \leq s,
\end{align*}
$$

where $\Delta$ is the data perturbation matrix and $U_{0}=\left\{\Delta \in \mathbb{R}^{T \times n}:\left\|\Delta_{i}\right\|_{2} \leq \alpha, \forall i \in\{1, \ldots, n\}\right\}$. The uncertainty set of the asset return is denoted by $U$, and we will discuss two selections of $U$ in the last section.

In [31] (Chapter 2), they showed the equivalence of robustness and regularization. Specifically, they precisely characterized the conditions on the model of uncertainty and loss function under which robustness is equivalent to regularization for linear regression.

Definition 3. Let $g: \mathbb{R}^{T} \rightarrow \mathbb{R}$ and $h: \mathbb{R}^{n} \rightarrow \mathbb{R}$ be the norm, then the induced norm $\|\cdot\|_{(h, g)}$ is defined as

$$
\|\Delta\|_{(h, g)}=\max _{\mathbf{w} \in \mathbb{R}^{n}} \frac{g(\Delta \mathbf{w})}{h(\mathbf{w})}
$$

Theorem 4 ([31], Chapter 2). If $r, q \in[1, \infty]$, then

$$
\min _{\mathbf{w}} \max _{\Delta \in U_{\left(\ell_{q}, \ell_{r}\right)}}\|\mathbf{y}-(R+\Delta) \mathbf{w}\|_{r}=\min _{\mathbf{w}}\|\mathbf{y}-R \mathbf{w}\|_{r}+\alpha\|\mathbf{w}\|_{q},
$$

where $U_{\left(\ell_{q}, \ell_{r}\right)}=\left\{\Delta:\|\Delta\|_{\left(\ell_{q}, \ell_{r}\right)} \leq \alpha\right\}$. Moreover, if $U_{0}=\left\{\Delta:\left\|\Delta_{i}\right\|_{2} \leq \alpha, \forall i \in\{1, \ldots, n\}\right\}$, then $U_{\left(\ell_{1}, \ell_{2}\right)}=U_{0}$, and this implies

$$
\min _{\mathbf{w}} \max _{\left\|\Delta_{i}\right\|_{2} \leq \alpha}\|\mathbf{y}-(R+\Delta) \mathbf{w}\|_{2}=\min _{\mathbf{w}}\|\mathbf{y}-R \mathbf{w}\|_{2}+\alpha\|\mathbf{w}\|_{1} .
$$

From the relationship of the robustness and the regularization, problem (4) can be rewritten as

$$
\begin{align*}
\min _{\mathbf{w} \in \mathbb{R}^{n}} & \left\|\rho \mathbf{1}_{T}-R \mathbf{w}\right\|_{2}+\alpha\|\mathbf{w}\|_{1} \\
\text { s.t. } & \min _{\mu \in U} \mathbf{w}^{T} \mu \geq \rho  \tag{5}\\
& \mathbf{w}^{T} \mathbf{1}_{n}=1, \mathbf{w} \geq 0 \\
& \|\mathbf{w}\|_{0} \leq s
\end{align*}
$$

Under this transformation, the problem (5) actually enjoys robustness. We plan to use an alternating penalty method to solve problem (5). To ensure the implementation of the alternating penalty method, we add a copy constraint $\mathbf{w}=\mathbf{v}$ to the problem (5) and, then, move it to the objective function by means of the $\ell_{1}$ norm penalty, then the penalization formulation is

$$
\begin{array}{rl}
\min _{\mathbf{w}, \mathbf{v} \in \mathbb{R}^{n}} & f(\mathbf{w}, \mathbf{v}):=\left\|\rho \mathbf{1}_{T}-R \mathbf{w}\right\|_{2}+\alpha\|\mathbf{w}\|_{1}+\beta\|\mathbf{w}-\mathbf{v}\|_{1} \\
\text { s.t. } & \mathbf{w} \in \Omega_{1}:=\left\{\mathbf{w} \mid \min _{\mu \in U} \mathbf{w}^{T} \mu \geq \rho, \mathbf{w} \geq 0\right\}  \tag{6}\\
& \mathbf{v} \in \Omega_{2}:=\left\{\mathbf{v} \mid \mathbf{v}^{T} \mathbf{1}_{n}=1,\|\mathbf{v}\|_{0} \leq s\right\} .
\end{array}
$$

We conduct its optimality analysis in the next subsection.

### 3.2. Optimization Theory

We now analyze the optimality of the penalization problem (6). Obviously, the objective function of the problem (6) is a lower semi-continuous and coercive function. Theorem 5 in the next subsection provides the existence of optimal solutions. The selection of the uncertainty set $U$ is discussed in Section 4.

This subsection provides a few theoretical results of the problem (6) including the existence of the solution and three classes of the first-order necessary optimal condition.

Theorem 5. For any given $\alpha \in \mathbb{R}_{+}$and $\beta \in \mathbb{R}_{+}$, the optimal solutions of the problem (6) can be attained.

Proof. It is clear that $f$ is a proper, closed, and coercive function and $\Omega_{1} \times \Omega_{2}$ is a nonempty closed set satisfying $\Omega_{1} \times \Omega_{2} \cap \operatorname{dom}(f) \neq \varnothing$. It follows from Theorem 2.14 of [32] that this theorem holds.

We now define a class of KKT points of the problem (6). For the convenience of expression and the generality of optimality, we write $\min _{\mu \in U} \mathbf{w}^{T} \mu \geq \rho$ as $g(\mathbf{w}) \geq 0$ and suppose that $g$ is a concave function and is not necessarily differentiable. Indeed, the quadratic uncertainty set and the absolute uncertainty set introduced in Section 4 satisfy these terminologies.

Definition 4. The point $\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right) \in \Omega_{1} \times \Omega_{2}$ is called a KKT point of the problem (6), if there exist Lagrange multipliers $\lambda_{1}^{*} \in \mathbb{R}_{+}$and $\lambda_{2}^{*} \in \mathbb{R}$ such that the following system holds:

$$
\left\{\begin{array}{l}
0 \in \partial_{\mathbf{w}} f\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right)-\lambda_{1}^{*} \partial_{\mathbf{w}} g\left(\mathbf{w}^{*}\right)+N\left(\mathbf{w}^{*} ; \mathbb{R}_{+}\right)  \tag{7}\\
0 \in \partial_{\mathbf{v}} f\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right)-\lambda_{2}^{*} \mathbf{1}+\hat{N}\left(\mathbf{v}^{*} ; S\right) \\
g\left(\mathbf{w}^{*}\right) \geq 0, \lambda^{*} g\left(\mathbf{w}^{*}\right)=0 \\
\mathbf{1}^{T} \mathbf{v}^{*}=1,\left\|\mathbf{v}^{*}\right\|_{0} \leq s
\end{array}\right.
$$

Although the functions corresponding to the quadratic uncertainty set and the absolute uncertainty set introduced in Section 4 are all concave and may both be non-differentiable and Slater's CQ automatically holds for both functions, we still considered Slater's CQ as a condition of Theorem 6 for the sake of generality. Moreover, it is stated in Section 2 that the R-LICQ of $\Omega_{2}$ holds at every point. Then, only under the condition that Slater's CQ holds, the relationship between the local minimizer of the problem (6) and the KKT point of the problem (6) can be established.

Theorem 6. Let $\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right) \in \Omega_{1} \times \Omega_{2}$ be a local minimizer of the problem (6). If Slater's $C Q$ holds on $\Omega_{1}$, then it is a KKT point of the problem (6).

Proof. On the one hand, since $\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right) \in \Omega_{1} \times \Omega_{2}$ is a local minimizer of the problem (6), $\mathbf{w}^{*}$ is a local minimizer of the following optimization:

$$
\begin{array}{rl}
\min _{\mathbf{w} \in \mathbb{R}^{n}} & f\left(\mathbf{w}, \mathbf{v}^{*}\right)=\left\|\rho \mathbf{1}_{T}-R \mathbf{w}\right\|_{2}+\alpha\|\mathbf{w}\|_{1}+\beta\left\|\mathbf{w}-\mathbf{v}^{*}\right\|_{1}  \tag{8}\\
\text { s.t. } & \mathbf{w} \in \Omega_{1} .
\end{array}
$$

Notice that $f\left(\mathbf{w}, \mathbf{v}^{*}\right)$ is a convex function about $\mathbf{w}$ and $\Omega_{1}$ is a convex set. Then, problem (8) is a convex optimization. Since Slater's $C Q$ holds on $\Omega_{1}$, it follows from Theorem 1 that there exists a Lagrange multiplier $\lambda_{1}^{*} \in \mathbb{R}_{+}$such that

$$
\left\{\begin{array}{l}
0 \in \partial_{\mathbf{w}} f\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right)+\lambda_{1}^{*} \partial_{\mathbf{w}} g\left(\mathbf{w}^{*}\right)+N\left(\mathbf{w}^{*} ; \mathbb{R}_{+}\right)  \tag{9}\\
g\left(\mathbf{w}^{*}\right) \geq 0, \lambda_{1}^{*} g\left(\mathbf{w}^{*}\right)=0
\end{array}\right.
$$

On the other hand, $\mathbf{v}^{*}$ is a local minimizer of the following optimization:

$$
\begin{array}{rl}
\min _{\mathbf{v} \in \mathbb{R}^{n}} & f\left(\mathbf{w}^{*}, \mathbf{v}\right)=\left\|\rho \mathbf{1}_{T}-R \mathbf{w}^{*}\right\|_{2}+\alpha\left\|\mathbf{w}^{*}\right\|_{1}+\beta\left\|\mathbf{w}^{*}-\mathbf{v}\right\|_{1}  \tag{10}\\
\text { s.t. } & \mathbf{v} \in \Omega_{2} .
\end{array}
$$

Since the R-LICQ of $\Omega_{2}$ holds at every point, it follows from Theorem 3 that there exists a Lagrange multiplier $\lambda_{2}^{*} \in \mathbb{R}$ such that

$$
\left\{\begin{array}{l}
0 \in \partial_{\mathbf{v}} f\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right)+\lambda_{2}^{*} \mathbf{1}+\hat{N}\left(\mathbf{v}^{*} ; S\right)  \tag{11}\\
\mathbf{1}^{T} \mathbf{v}^{*}=1,\left\|\mathbf{v}^{*}\right\|_{0} \leq s
\end{array}\right.
$$

Combing the system (9) and (11), this theorem holds.
Again, problem (10) can be simply written as

$$
\begin{gathered}
\min _{\mathbf{v} \in \mathbb{R}^{n}}\left\|\mathbf{w}^{*}-\mathbf{v}\right\|_{1} \\
\text { s.t. } \mathbf{v} \in \Omega_{2}
\end{gathered}
$$

and it has a closed-form solution; see [33], i.e.,

$$
v_{i}^{*}= \begin{cases}\frac{w_{*}^{*}}{\left(\mathbf{w}_{s}^{*}\right)^{T} \mathbf{1}_{s}}, & \text { if } i \in I_{s}^{*}  \tag{12}\\ 0, & \text { otherwise },\end{cases}
$$

where $I_{s}^{*}:=\left\{i \mid w_{\langle 1\rangle}^{*} \geq \ldots \geq w_{\langle s\rangle}^{*}\right\}$ and $w_{\langle i\rangle}^{*}$ denotes the $i$-th largest absolute value among the $n$ elements of $\mathbf{w}^{*}$. Thus, we can define a class of strong KKT points of the problem (6) as follows.

Definition 5. The point $\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right) \in \Omega_{1} \times \Omega_{2}$ is called a strong KKT point of the problem (6), if there exists a Lagrange multiplier $\lambda^{*} \in \mathbb{R}_{+}$such that the following system holds:

$$
\left\{\begin{array}{l}
0 \in \partial_{\mathbf{w}} f\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right)+\partial_{\mathbf{w}} g\left(\mathbf{w}^{*}\right)+N\left(\mathbf{w}^{*} ; \mathbb{R}_{+}\right),  \tag{13}\\
g\left(\mathbf{w}^{*}\right) \geq 0, \lambda^{*} g\left(\mathbf{w}^{*}\right)=0 \\
v_{i}^{*}= \begin{cases}\frac{w_{*}^{*}}{\left(\mathbf{w}_{s}^{*}\right)^{T} \mathbf{1}_{s}}, & \text { if } i \in I_{s}^{*} \\
0, & \text { otherwise } .\end{cases}
\end{array}\right.
$$

It is easy to prove that, if $\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right)$ is a strong KKT point of the problem (6), then it is a KKT point of the problem (6). The following result provides the relationship between the global minimizer of the problem (6) and the strong KKT point of the problem (6).

Theorem 7. Let $\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right) \in \Omega_{1} \times \Omega_{2}$ be a global minimizer of the problem (6). If Slater's $C Q$ holds on $\Omega_{1}$ at $w^{*}$, then it is a strong KKT point of the problem (6).

Proof. The part of $\mathbf{w}^{*}$ in (13) follows from (7). We only need to discuss the part of $\mathbf{v}^{*}$ in (13). Since $\mathbf{v}^{*}$ is the global minimizer of (10), it follows from (12) that the part of $\mathbf{v}^{*}$ in (13) holds.

Note that the local minimizer of the problem (6) cannot be guaranteed to be a strong KKT point.

Finally, we introduce the third stationary point of the problem (6), which is called the partial minimizer.

Definition 6. The point $\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right) \in \Omega_{1} \times \Omega_{2}$ is called a partial minimizer of the problem (6), if it satisfies

$$
f\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right) \leq f\left(\mathbf{w}, \mathbf{v}^{*}\right), \forall \mathbf{w} \in \Omega_{1}, f\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right) \leq f\left(\mathbf{w}^{*}, \mathbf{v}\right), \forall \mathbf{v} \in \Omega_{2} .
$$

Clearly, any global minimizer of the problem (6) is a partial minimizer. Moreover, on the one hand, the partial problem (8) is a convex optimization, and Slater's CQ ensures that its KKT point and global minimizer are consistent. On the other hand, the partial problem (10) has a closed-form solution. Thus, the equivalence relationship between the KKT point of the problem (6) and the partial minimizer of the problem (6) can be established under Slater's CQ.

Theorem 8. Let $\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right) \in \Omega_{1} \times \Omega_{2}$ be a feasible point of the problem (6). Suppose that Slater's CQ holds on $\Omega_{1}$. Then, $\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right)$ is a partial minimizer of the problem (6) if and only if $\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right)$ is a strong KKT point of the problem (6).

Proof. Suppose that $\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right)$ is a strong KKT point of the problem (6), then

$$
0 \in \partial_{\mathbf{w}} f\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right)+\partial_{\mathbf{w}} g\left(\mathbf{w}^{*}\right)+N\left(\mathbf{w}^{*} ; \mathbb{R}_{+}\right), g\left(\mathbf{w}^{*}\right) \geq 0, \text { and } \lambda_{1}^{*} g\left(\mathbf{w}^{*}\right)=0
$$

Since Slater's CQ holds at $\mathbf{w}^{*}, \mathbf{w}^{*}$ is a global minimizer of the problem (8). Then, we have that $f\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right) \leq f\left(\mathbf{w}, \mathbf{v}^{*}\right) \forall \mathbf{w} \in \Omega_{1}$. Moreover, it follows from the definition of the strong KKT point of the problem (6) that $\mathbf{v}^{*}$ is a global minimizer of the problem (10). Then, we have that $f\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right) \leq f\left(\mathbf{w}^{*}, \mathbf{v}\right), \forall \mathbf{v} \in \Omega_{2}$. Thus, $\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right)$ is a partial minimum of the problem (6). The opposite conclusion clearly holds.

## 4. The Uncertainty Set U

In Subsection 3.2, we rewrite the uncertainty set constraint as $g(\boldsymbol{w}) \geq 0$, where $g$ is a generalized concave function and is not necessarily differentiable. This section introduces two mainstream formulations for the uncertainty set in asset mean return vector $\mu$ (see [34]), which corresponds to the quadratic uncertainty set and the absolute uncertainty set, respectively.

### 4.1. The Quadratic Uncertainty Set

The first one is the quadratic formulation, $U=\left\{\mu \mid(\mu-\bar{\mu})^{T} \Omega(\mu-\bar{\mu}) \leq \kappa^{2}\right\}$, where $\bar{\mu}$ is the nominal expected return and $\kappa$ is the error. Assume that asset returns are independent and identically distributed and $\mu-\bar{\mu}$ follows a normal distribution with mean value 0 and covariance matrix $\Omega$, where $\Omega$ is the covariance matrix of errors in the expected asset return. In Yin et al. [35], they discussed the choice of uncertainty matrix $\Omega$ in the quadratic uncertainty set and proposed the selection criteria. In the quadratic uncertainty case, $\min _{\mu \in U} \mathbf{w}^{T} \mu$ in problem (5) is equivalent to the following problem:

$$
\max _{\mu \in U} \mathbf{w}^{T} \bar{\mu}-\mathbf{w}^{T} \mu .
$$

Solving the above problem, we obtain

$$
\mu=\bar{\mu}-\sqrt{\frac{\kappa^{2}}{\mathbf{w}^{T} \Omega \mathbf{w}}} \Omega \mathbf{w} .
$$

Then, the problem (5) is rewritten as

$$
\begin{align*}
\min _{\mathbf{w}, \mathbf{v} \in \mathbb{R}^{n}} & \left\|\rho \mathbf{1}_{T}-R \mathbf{w}\right\|_{2}+\alpha\|\mathbf{w}\|_{1} \\
\text { s.t. } & \bar{\mu}^{T} \mathbf{w}-\kappa \sqrt{\mathbf{w}^{T} \Omega \mathbf{w}} \geq \rho,  \tag{14}\\
& \mathbf{w}^{T} \mathbf{1}_{n}=1, \mathbf{w} \geq 0 \\
& \|\mathbf{w}\|_{0} \leq s .
\end{align*}
$$

Here, $g(\mathbf{w})=\bar{\mu}^{T} \mathbf{w}-\kappa \sqrt{\mathbf{w}^{T} \Omega \mathbf{w}}-\rho$. The penalization form of problem (14) can be rewritten as

$$
\begin{aligned}
\min _{\mathbf{w}, \mathbf{v} \in \mathbb{R}^{n}} & \left\|\rho \mathbf{1}_{T}-R \mathbf{w}\right\|_{2}+\alpha\|\mathbf{w}\|_{1}+\beta\|\mathbf{w}-\mathbf{v}\|_{1} \\
\text { s.t. } & \mathbf{w} \in \Omega_{1}=\left\{\mathbf{w} \mid \kappa\|\sqrt{\Omega} \mathbf{w}\|_{2} \leq \bar{\mu}^{T} \mathbf{w}-\rho, \mathbf{w} \geq 0\right\} \\
& \mathbf{v} \in \Omega_{2}=\left\{\mathbf{v} \mid \mathbf{v}^{T} \mathbf{1}_{n}=1,\|\mathbf{v}\|_{0} \leq s\right\} .
\end{aligned}
$$

According to the proposition of Yin et al. [35], we choose $\Omega \propto \operatorname{diag}(V)$. By using this uncertainty matrix, it is expected to reduce the sensitivity to the inputs, as well as keep the original volatility unchanged.

### 4.2. The Absolute Uncertainty Set

Fabozzi et al. [23] used the absolute uncertainty set in mean returns that ask that the sum of absolute spreads between estimated and possible mean returns should not be too large. The absolute formulation is $U=\left\{\mu\left|\sum_{i}\right| \mu_{i}-\bar{\mu}_{i} \left\lvert\, \leq \frac{\kappa \bar{\sigma}}{\sqrt{T}}\right.\right\}$. In this case,

$$
\mu^{T} \mathbf{w}-\bar{\mu}^{T} \mathbf{w} \geq-\sum_{i}\left|\mu_{i}-\bar{\mu}_{i}\right| \max \left(\left|\mathbf{w}_{i}\right|\right) \geq-\frac{\kappa \bar{\sigma}}{\sqrt{T}} \max \left(\left|\mathbf{w}_{i}\right|\right) ;
$$

thus,

$$
\mu^{T} \mathbf{w} \geq \bar{\mu}^{T} \mathbf{w}-\frac{\kappa \bar{\sigma}}{\sqrt{T}} \max \left(\left|\mathbf{w}_{i}\right|\right)
$$

Then, the problem (5) is equivalent to

$$
\begin{align*}
\min _{\mathbf{w}, \mathbf{v} \in \mathbb{R}^{n}} & \left\|\rho \mathbf{1}_{T}-R \mathbf{w}\right\|_{2}+\alpha\|\mathbf{w}\|_{1} \\
\text { s.t. } & \bar{\mu}^{T} \mathbf{w}-\frac{\kappa \bar{\sigma}}{\sqrt{T}} \max \left(\left|\mathbf{w}_{i}\right|\right) \geq \rho,  \tag{15}\\
& \mathbf{w}^{T} \mathbf{1}_{n}=1, \mathbf{w} \geq 0 \\
& \|\mathbf{w}\|_{0} \leq s .
\end{align*}
$$

Here, $g(\mathbf{w})=\bar{\mu}^{T} \mathbf{w}-\kappa \bar{\sigma} \max \left(\left|\mathbf{w}_{i}\right|\right) / \sqrt{T}-\rho$. Similarly, the penalization from of problem (15) can be written as

$$
\begin{aligned}
\min _{\mathbf{w}, \mathbf{v} \in \mathbb{R}^{n}} & \left\|\rho \mathbf{1}_{T}-R \mathbf{w}\right\|_{2}+\alpha\|\mathbf{w}\|_{1}+\beta\|\mathbf{w}-\mathbf{v}\|_{1} \\
\text { s.t. } & \mathbf{w} \in \Omega_{1}=\left\{\mathbf{w}| | w_{i} \left\lvert\, \leq \frac{\sqrt{T}}{\kappa \bar{\sigma}}\left(\mathbf{w}^{T} \bar{\mu}-\rho\right)\right., i=1, \ldots, n, \mathbf{w} \geq 0\right\} \\
& \mathbf{v} \in \Omega_{2}=\left\{\mathbf{v} \mid \mathbf{v}^{T} \mathbf{1}_{n}=1,\|\mathbf{v}\|_{0} \leq s\right\} .
\end{aligned}
$$

## 5. Optimization

This section introduces a penalty alternating direction method (PADM) to solve problem (5).

### 5.1. Alternating Direction Methods

We first discuss the optimization of the problem (6). Due to the complexity of this problem, alternating direction methods (ADMs) can be used to solve this problem. The framework of ADMs is described as follows:

Next, we state the general convergence result of Algorithm 1, and one can refer to Geissler et al. [36] for a proof (Theorem 8) and for further details about this method.

```
Algorithm 1 ADM: Alternating Direction Method.
    1: Set the problem parameters: \(\alpha, \kappa, \rho, T>0\), asset return matrix \(R \in \mathbb{R}^{T \times n}\), and nominal
    expected return vector \(\bar{\mu} \in \mathbb{R}^{n}\). Initialize \(\varepsilon>0,\left(\mathbf{w}^{0}, \mathbf{v}^{0}\right)\) and penalty parameter \(\beta>0\).
    Set the iteration index \(k:=0,1, \ldots\).
    2: Compute
\[
\begin{equation*}
\mathbf{w}^{k+1} \in \arg \min _{\mathbf{w}}\left\{f\left(\mathbf{w}, \mathbf{v}^{k}\right): \mathbf{w} \in \Omega_{1}\right\}, \tag{16}
\end{equation*}
\]
and
\[
\begin{equation*}
\mathbf{v}^{k+1} \in \arg \min _{\mathbf{v}}\left\{f\left(\mathbf{w}^{k+1}, \mathbf{v}\right): \mathbf{v} \in \Omega_{2}\right\} \tag{17}
\end{equation*}
\]
3: If \(\frac{\left\|\mathbf{w}^{k+1}-\mathbf{w}^{k}\right\|_{2}+\left\|\mathbf{v}^{k+1}-\mathbf{v}^{k}\right\|_{2}}{\left\|\mathbf{w}^{k}\right\|_{2}+\left\|\mathbf{v}^{k}\right\|_{2}} \leq \varepsilon\), then stop with \(\left(\mathbf{w}^{k}, \mathbf{v}^{k}\right)\) being an output point of (6).
```

Theorem 9. Let $\left\{\left(\mathbf{w}^{k}, \mathbf{v}^{k}\right)\right\}$ be a sequence generated by Algorithm 1. Then, the following holds:
(a) $\left\{\left(\mathbf{w}^{k}, \mathbf{v}^{k}\right)\right\}$ is bounded.
(b) Any limiting point $\left\{\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right)\right\}$ of $\left\{\left(\mathbf{w}^{k}, \mathbf{v}^{k}\right)\right\}$ is a partial minimizer of the problem (6).
(c) If Slater's CQ holds on $\Omega_{1}$, the limiting point of $\left\{\left(\mathbf{w}^{k}, \mathbf{v}^{k}\right)\right\}$ is also a strong KKT of the problem (6).

Proof. (a) It follows from Algorithm 1 that

$$
f\left(\mathbf{w}^{k+1}, \mathbf{v}^{k+1}\right) \leq f\left(\mathbf{w}^{k+1}, \mathbf{v}^{k}\right) \leq f\left(\mathbf{w}^{k}, \mathbf{v}^{k}\right) .
$$

Since $f$ is a coercive function, then the level set of $f$ is bounded. Thus, $\left\{\left(\mathbf{w}^{k}, \mathbf{v}^{k}\right)\right\}$ is bounded.
(b) Clearly, $\left\{f\left(\mathbf{w}^{k}, \mathbf{v}^{k}\right)\right\}$ is a decreasing sequence and $f\left(\mathbf{w}^{k}, \mathbf{v}^{k}\right) \geq 0$, then there exists a value $f^{*}$ such that $\lim _{k \rightarrow \infty} f\left(\mathbf{w}^{k}, \mathbf{v}^{k}\right)=f^{*}$. Suppose that $\left\{\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right)\right\}$ is a limiting point of $\left\{\left(\mathbf{w}^{k}, \mathbf{v}^{k}\right)\right\}$. Then, there exists a sequence $\left\{k_{j}\right\}$ such that $\lim _{j \rightarrow \infty} k_{j}=\infty, \lim _{j \rightarrow \infty}\left(\mathbf{w}^{k_{j}}, \mathbf{v}^{k_{j}}\right)=$ $\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right)$ and $\lim _{j \rightarrow \infty} f\left(\mathbf{w}^{k_{j}}, \mathbf{v}^{k_{j}}\right)=f\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right)=f^{*}$, where the last equality holds since $\lim _{k \rightarrow \infty} f\left(\mathbf{w}^{k}, \mathbf{v}^{k}\right)=f^{*}$. Without loss of generality, let $\lim _{k \rightarrow \infty}\left(\mathbf{w}^{k}, \mathbf{v}^{k}\right)=\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right)$. It follows from Algorithm 1 that

$$
\mathbf{w}^{k+1} \in \arg \min _{\mathbf{w} \in \Omega_{1}} f\left(\mathbf{w}, \mathbf{v}^{k}\right) .
$$

Since $f$ is continuous with respect to $\mathbf{w}$, then taking $k \rightarrow \infty$, we have that

$$
\mathbf{w}^{*} \in \arg \min _{\mathbf{w} \in \Omega_{1}} f\left(\mathbf{w}, \mathbf{v}^{*}\right) .
$$

Similarly,

$$
\mathbf{v}^{*} \in \arg \min _{\mathbf{v} \in \Omega_{2}} f\left(\mathbf{w}^{*}, \mathbf{v}\right)
$$

Thus, the limiting point $\left\{\left(\mathbf{w}^{*}, \mathbf{v}^{*}\right)\right\}$ of $\left\{\left(\mathbf{w}^{k}, \mathbf{v}^{k}\right)\right\}$ is a partial minimizer of the problem (6).
(c) Under Slater's CQ, the partial minimizer of the problem (6) is a strong KKT point of the problem (6) and the opposite also holds. Thus, this result holds.

### 5.2. The Optimization for the Partial Problem (8)

We now discuss the optimization of the partial problem (8) at the $k$-th iteration of the ADM. Some non-exact penalty methods and smoothing methods can be used to solve this problem. Here, we obtain $\mathbf{w}^{k+1}$ by solving the following optimization:

$$
\begin{array}{rl}
\min _{\mathbf{w} \in \mathbb{R}^{n}} & f\left(\mathbf{w}, \mathbf{v}^{k}\right)=\left\|\rho \mathbf{1}_{T}-R \mathbf{w}\right\|_{2}+\alpha\|\mathbf{w}\|_{1}+\beta\left\|\mathbf{w}-\mathbf{v}^{k}\right\|_{1}+\gamma\left|g(\mathbf{w})_{-}\right|  \tag{18}\\
\text {s.t. } & \mathbf{w} \geq 0,
\end{array}
$$

where $\gamma>0$ is a penalty parameter. Let

$$
\psi_{\mu}(t)=\left\{\begin{array}{ll}
|t|, & |t| \geq \mu, \\
\frac{t^{2}}{2 \mu}+\frac{\mu}{2}, & |t|<\mu,
\end{array} \quad \phi_{\mu}(t)=\frac{1}{2}\left(t+\sqrt{t^{2}+\mu}\right)\right.
$$

where $\mu>0$ is a smoothing parameter. Then, a class of the smoothing optimization of problem (18) can be given as follows:

$$
\begin{align*}
\min _{\mathbf{w} \in \mathbb{R}^{n}} & f_{\mu}\left(\mathbf{w}, \mathbf{v}^{k}\right)=\sqrt{\left\|\rho \mathbf{1}_{T}-R \mathbf{w}\right\|_{2}^{2}+\mu}+ \\
& \alpha \sum_{i=1}^{n} \psi_{\mu}\left(w_{i}\right)+\beta \sum_{i=1}^{n} \psi_{\mu}\left(w_{i}-v_{i}^{k}\right)+\gamma \phi_{\mu}(-g(\mathbf{w}))  \tag{19}\\
\text { s.t. } & \mathbf{w} \geq 0
\end{align*}
$$

The projection gradient method (PGM) can be used to solve the problem (19), and its iteration formula is

$$
\mathbf{w}^{j+1, k}=P_{\mathbb{R}_{+}}\left(\mathbf{w}^{j, k}+\eta \nabla f_{\mu}\left(\mathbf{w}^{j, k}, \mathbf{v}^{k}\right)\right),
$$

where $\eta>0$ denotes the step length at the $j$-th iteration of the PGM at the $k$-th iteration of the ADM and $P_{\#}(t)$ denotes the projection point of $t$ onto \#. The framework of the above method is called the penalty projection gradient method (PPGM) and can be described in Algorithm 2.

```
Algorithm 2 PPGM: Penalty Projection Gradient Method.
    Set the problem parameters: \(\alpha, \kappa, \rho, T, \beta, \gamma_{\max }>0\), asset return matrix \(R \in \mathbb{R}^{T \times n}\), and
    nominal expected return vector \(\bar{\mu} \in \mathbb{R}^{n}\). Initialize penalty parameters \(\gamma^{0}>0, \tau>1\).
    Set the iteration index \(j=0,1, \cdots\).
    Computing \(\mathbf{w}^{j+1, k}=P_{\mathbb{R}_{+}}\left(\mathbf{w}^{j-1, k}+\eta \nabla f_{\mu}\left(\mathbf{w}, \mathbf{v}^{k}\right)\right)\).
    If \(\mathbf{w}^{j+1, k}\) satisfies \(\frac{\left\|\mathbf{w}^{j+1, k}-\mathbf{w}^{j, k}\right\|_{2}}{\left\|\mathbf{w}^{j, k}\right\|_{2}} \leq \varepsilon\) and \(g\left(\mathbf{w}^{j, k}\right) \geq-\varepsilon\), then stop and \(\mathbf{w}^{k}=\mathbf{w}^{j, k}\).
    If \(\frac{\left\|\mathbf{w}^{j+1, k}-\mathbf{w}^{j, k}\right\|_{2}}{\left\|\mathbf{w}^{j, k}\right\|_{2}} \leq \varepsilon\) and \(g\left(w^{j, k}\right)<-\varepsilon\), then choose new penalty parameter \(\gamma=\)
    \(\min \left\{\tau \gamma, \gamma_{\max }\right\}\). Otherwise, return to Step 2.
```


### 5.3. Penalty Alternating Direction Method

At the end of this section, we describe the PADM for the general problem (5). At iteration $l$, set the value of penalty parameter $\beta^{l}$ and obtain $\left(\mathbf{w}^{l}, \mathbf{v}^{l}\right)$ by the ADM with $\beta^{l}$. If the inequality $\left\|\mathbf{w}^{l}-\mathbf{v}^{l}\right\|_{1} \leq$ tol holds, where tol is a small positive constant, we stop with a feasible solution of problem (6). Otherwise, the penalty parameter $\beta^{l}$ is updated to $\beta^{l+1}$. In this way, the PADM generates a sequence of the partial minimizer of problem (6) with $\beta^{l}$. The framework of the PADM is formally stated in Algorithm 3.

## Algorithm 3 PADM: Penalty Alternating Direction Method.

Set the problem parameters: $\alpha, \kappa, \rho, T, \beta_{\max }>0$, asset return matrix $R \in \mathbb{R}^{T \times n}$, and nominal expected return vector $\bar{\mu} \in \mathbb{R}^{n}$. Initialize penalty parameters $\beta^{0}>0, \tau>1$. Set the iteration index $l=0,1, \cdots$.
Obtain $\left(\mathbf{w}^{l}, \mathbf{v}^{l}\right)$ by the ADM with $\beta^{l}$.
If $\left(\mathbf{w}^{l}, \mathbf{v}^{l}\right)$ satisfies $\left\|\mathbf{w}^{l}-\mathbf{v}^{l}\right\|_{1} \leq$ tol, then stop with $\left(\mathbf{w}^{l}, \mathbf{v}^{l}\right)$. Otherwise, choose new penalty parameter $\beta^{l+1}=\min \left\{\tau \beta^{l}, \beta_{\max }\right\}$, and return to Step 2.

## 6. Numerical Results

This section shows extensive numerical experiments. In Section 6.1, we first present six real data sets, explain some existing models to be compared with and describe the performance measures to be used. In Section 6.2, we demonstrate that our methods lead to robust and sparse portfolios. In Section 6.3, we compare nine popular portfolios in terms of out-of-sample (OOS) performance measures. Finally, in Section 6.4, we show the cumulative return of different portfolio strategies. All of our computations are conducted in the Matlab R2019a environment, on a PC with an Intel(R) Core(TM) i5-7200U CPU ( $2.50 \mathrm{GHz}, 4 \mathrm{CPUs}$ ) and 4G RAM processors.

### 6.1. Models of Comparison, Data, and Performance Measures

(a) Eleven portfolio models compared. We compare the OOS performance of 11 portfolio models across six real data sets of weekly and monthly returns. Those models are well studied, and we divide them into four groups, which are summarized in Table 1. The first group is the robust and sparse portfolio strategies developed in this paper. The second group includes some well-studied portfolio strategies. The third group includes three benchmark portfolio strategies. The last group consists of two portfolios that use the shrinkage technique to estimate the covariance matrix.

Table 1. List of portfolio strategies considered.

| Group | Model | Abbr. | Refer. |
| :--- | :--- | :--- | :--- |
| $(1)$ | Robust and sparse portfolios with |  |  |
|  | quadratic uncertainty set |  |  |
|  | absolute uncertainty set | RSQ | this paper |
|  | Some well-studied portfolio strategies with |  |  |
| $(2)$ | $\ell_{1}$ regularization | RSA | this paper |
|  | $\ell_{1,2}$ regularization | Brodie et al. [2] |  |
|  | Elastic Net regularization | L12 | Zhao et al. [37] |
|  | upper and lower bound | EN | Yen and Yen [38] |
|  | Benchmarks' portfolio strategies with |  |  |
| $(3)$ | short-sales constrained | Behr et al. [39] |  |
|  | short-sales unconstrained | SC | Jagannathan and Ma [5] |
|  | equally weighted (1/N) portfolio | SU | Jagannathan and Ma [5] |
|  | Shrinkage of covariance | EW | DeMiguel et al. [40] |
| (4) | sample covariance and identity matrix | SCID | Olivier and Wolf [41] |
|  | sample covariance and 1-factor matrix | SC1F | Olivier and Wolf [3] |

(b) Seven data sets tested. Table 2 lists some real-world data sets: DJIA [42], NASDAQ [43], S\&P [44,45], Russell2000 [46], Russell3000 [47], and FF100 [38]. All the data are obtained from Yahoo finance (https:/ / finance.yahoo.com/ , accessed on 10 January 2023) and Ken French's website (https:/ /mba.tuck.dartmouth.edu/pages/faculty/ken.french/ data_library.html, accessed on 10 January 2023). In all cases, we remove those assets that have missing values.

Table 2. Information of the seven real data sets.

| $\#$ | Data Sets | Stocks | Time Period | Source | Frequency |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | DJIA | 29 | $01 / 10 / 2017-30 / 10 / 2022$ | Yahoo finance | Weekly |
| 2 | NASDAQ | 95 | $01 / 10 / 2017-30 / 10 / 2022$ | Yahoo finance | Weekly |
| 3 | SP500 | 336 | $01 / 10 / 2017-30 / 10 / 2022$ | Yahoo finance | Weekly |
| 4 | Russell2000 | 1340 | $01 / 10 / 2017-30 / 10 / 2022$ | Yahoo finance | Weekly |
| 5 | Russell3000 | 2166 | $01 / 10 / 2017-30 / 10 / 2022$ | Yahoo finance | Weekly |
| 6 | SP100 | 71 | $01 / 10 / 2017-30 / 10 / 2022$ | Yahoo finance | Weekly |
| 7 | FF100 | 100 | $11 / 1999-06 / 2022$ | K.French | Monthly |

(c) Measuring the OOS performance and its setup. We largely follow the "rollingwindow" procedures in $[2,37]$ to conduct our comparison. Let $T$ be the length of a data set and $\tau$ be the window length (e.g., $\tau=120$ ) used to construct the optimal portfolio by a model. In each period $(t+1), t=\tau, \ldots, T-1$, we compute different portfolios over the previous $\tau$ periods. We then compute the OOS return in the $(t+1)$-th period based on the obtained portfolio. We repeat this procedure until we reach the end of the data set. In this way, we will obtain a series of $(T-\tau)$ portfolio vectors for each model listed in Table 1. To make it precise, let $\mathbf{w}_{t}^{s}$ be the optimal portfolio obtained by the portfolio strategy $s$ over the date from $t-\tau+1, \ldots, t$. The OOS return in the $t+1$ period is computed as $\mathbf{r}_{t+1}^{s}=\mathbf{w}_{t}^{S T} \mathbf{r}_{t+1}$, where $\mathbf{r}_{t+1}$ is the return in the $(t+1)$-th period. Thus, we obtain a time series of $(T-\tau-1)$ periods OOS returns for all strategies. Note that we use the traditional "rolling-window" procedures for the numerical analysis, and some new methods could provide new ideas for the analysis of portfolio selection problems, see [48].

The OOS performance of each portfolio strategy is assessed by using four quantities: (i) the OOS portfolio variance $\left(\widehat{\sigma}^{2}\right)$, (ii) the OOS portfolio Sharpe ratio $(\widehat{S R})$, (iii) portfolio turnover (TURN), and (iv) the average short positions (ASP). The specific definitions can be found in DeMiguel et al. [6], Yen and Yen [38], and Zhao et al. [37]. We evaluate the cumulative return (CR). The CR of a portfolio scores the total payoffs that are yielded by
the investment strategy across the investment periods without considering any risk or cost, see Shen et al. [49]. We also consider some quantities studied in [38] on the profiles of the portfolio weights: PAP represents the proportion of active positions and PZP is the proportion of zero positions, respectively, defined as $P A P_{t}=\frac{\left|S_{t}^{1}\right|}{N}, P Z P_{t}=\frac{\left|S_{t}^{0}\right|}{N}$, where $S_{t}^{1}=\left\{i: w_{i, t} \neq 0\right\}$ and $S_{t}^{0}=\left\{i: w_{i, t}=0\right\}$.

### 6.2. Robust and Sparse Portfolio

This section shows the weight of robust and sparse portfolios. We use the DJIA data set and the sparse levels $s_{1}=\lceil 30 \% n\rceil$ and $s_{2}=\lceil 50 \% n\rceil$. The parameter $\alpha=\beta=10 \lambda$, and the value of $\lambda$ varies from $10^{-2}$ to $10^{1}$.

Figure 1 shows the portfolio weights, PAP, and PZP. The two plots in the top panel correspond to a robust portfolio under the quadratic uncertainty set, and the sparsity is $s_{1}$. The two plots in the bottom panel correspond to a robust portfolio under the absolute uncertainty set, and the sparsity is $s_{2}$. With the increase of penalty parameter $\lambda$, the portfolio weights tended to be sparse. The PAP and PZP indicate that we can obtain sparse portfolios that satisfy the specified sparsity.


Figure 1. Portfolio weights.
Figure 2 shows the sparse portfolio. We use four different data sets. The sparsity level on DJIA is $s_{1}=\lceil 30 \% n\rceil$, on NASDAQ and FF100 is $s_{2}=\lceil 10 \% n\rceil$, and on Russell2000 is $s_{3}=\lceil 1 \% n\rceil$. We solve the robust portfolio under the quadratic uncertainty set to show the results. We obtain the portfolio with the specified sparsity and the distribution of different asset weight values.

### 6.3. Out-of-Sample Performance

The Sharpe ratio considers return and risk at the same time; it is a comprehensive measurement for us to observe the performance of a portfolio. Thus, we first test the Sharpe ratio of different portfolio strategies. We use the SP100 data set. The parameter $\alpha=2 \beta$ and the value of $\beta$ varies from 10 to $10^{1.5}$. The sparsity level $s_{1}=\lceil 15 \% n\rceil$ and $s_{2}=\lceil 5 \% n\rceil$.

By comparing with two benchmark portfolios, Figure 3 shows that the RSQ and RSA can produce a higher Sharpe ratio when choosing a suitable penalty parameter.


Figure 2. Sparse solutions.


Figure 3. The Sharpe ratio.
Table 3 reports the OOS performance by using four quantities defined in Section 6.1. We set $\alpha=\beta=10$ and the sparsity level $s_{1}=\lceil\lceil 30 \% n\rceil$ (on the DJIA, NASDAQ, SP500, and FF100 data sets) and $s_{2}=\lceil\lceil 30 \% n\rceil$ (on the Russell2000 and Russell3000 data sets). We can observe that the RSA and RSQ portfolios achieve the smallest variances across all portfolio strategies, i.e., on average with $10.84(\%)^{2}$ and $11.09(\%)^{2}$, respectively. This means they are less volatile, i.e., less risky. SU, SC1F, and SCID have the highest variance on average, $995.73(\%)^{2}, 442.81(\%)^{2}$ and $404.20(\%)^{2}$ in this setting. The variance of the remaining portfolio strategies is $11.94(\%)^{2}(\mathrm{~L} 1), 11.98(\%)^{2}(\mathrm{EN}), 16(\%)^{2}(\mathrm{~L} 12), 27.78(\%)^{2}$ (SC), $28.71(\%)^{2}$ (EW), and $71.11(\%)^{2}$ (Box), respectively. In addition, we observe that the Sharpe ratios of the various portfolios on average are $12.34 \%$ (SC), $11.82 \%$ (EW), $11.77 \%$ (RSA), $11.70 \%$ (RSQ), $11.61 \%$ (L12), $11.21 \%$ (EN), $11.18 \%$ (L1), $10.58 \%$ (SCID), $10.09 \%$ (SC1F),
9.23\% (Box), and 7.92\% (SU). We see that the RSA and RSQ portfolios do not result in a significantly different OOS Sharpe ratio when compared with SC and EW; however, they are higher than the rest of the portfolio strategies.

Table 3. Portfolio out-of-sample variance $\left(\widehat{\sigma}^{2}\right)\left((\%)^{2}\right)$, Sharpe ratio $(\widehat{S R})$, turnover (TURN), and the average short positions (ASP).

|  |  | DJIA | NASDAQ | SP500 | Russell2000 | Russell3000 | FF100 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | $\mathrm{n}=29$ | $\mathrm{n}=95$ | $\mathrm{n}=336$ | $\mathrm{n}=1340$ | $\mathrm{n}=2166$ | $\mathrm{n}=100$ |
| RSA | var | 5.3741 | 5.9065 | 11.5604 | 12.1813 | 12.7837 | 17.2215 |
|  | SR | 0.0703 | 0.1807 | 0.1008 | 0.0063 | 0.0267 | 0.3215 |
|  | TURN | 0.1194 | 0.1721 | 0.1430 | 0.1033 | 0.1021 | 0.1698 |
|  | ASP | -1.11e-18 | $2.22 \mathrm{e}-18$ | 0 | -3.08e-18 | -4.01e-18 | -6.28e-18 |
| RSQ | var | 5.3949 | 5.8633 | 12.5936 | 12.7041 | 12.7702 | 17.2097 |
|  | SR | 0.0737 | 0.1889 | 0.0863 | 0.0059 | 0.0258 | 0.3216 |
|  | TURN | 0.1063 | 0.1711 | 0.1334 | 0.1112 | 0.1003 | 0.1698 |
|  | ASP | -6.66e-18 | -3.33e-18 | $-1.43 \mathrm{e}-17$ | -9.25e-18 | -9.25e-18 | 0 |
| L12 | var | 9.5931 | 7.1442 | 32.1907 | 14.8927 | 13.7879 | 18.3918 |
|  | SR | 0.0860 | 0.1768 | 0.1000 | 0.0072 | 0.0295 | 0.2971 |
|  | TURN | 0.0220 | 0.0286 | 0.0369 | 0.0675 | 0.0605 | 0.0296 |
|  | ASP | -0.0216 | -0.0209 | -0.0240 | -0.0057 | -0.0050 | -0.0276 |
| L1 | var | 8.0373 | 6.7078 | 13.1228 | 14.2679 | 13.0166 | 16.4937 |
|  | SR | 0.0864 | 0.1831 | 0.0414 | 0.0042 | 0.0279 | 0.3279 |
|  | TURN | 0.1063 | 0.0677 | 0.0369 | 0.1250 | 0.1136 | 0.0629 |
|  | ASP | -0.0036 | -0.0012 | -0.0124 | -0.0039 | 0.0020 | -0.0195 |
| EN | var | 8.5871 | 6.5362 | 12.9449 | 14.2074 | 13.0817 | 16.5823 |
|  | SR | 0.0911 | 0.1790 | 0.0418 | 0.0029 | 0.0402 | 0.3181 |
|  | TURN | 0.0256 | 0.0472 | 0.0408 | 0.1208 | 0.1181 | 0.0510 |
|  | ASP | -0.0257 | -0.0198 | -0.0272 | 0.0015 | 0.0011 | -0.0257 |
| BOX | var | 9.9181 | 8.7749 | $3.59 \mathrm{e}+02$ | 7.3636 | 7.0004 | 34.5878 |
|  | SR | 0.0250 | 0.0768 | -0.1204 | -0.0495 | -0.0023 | 0.6244 |
|  | TURN | 0.8322 | 1.7146 | 3.2709 | 0.2820 | 0.2973 | 5.1895 |
|  | ASP | 1.1172 | 2.7850 | 6.5137 | 0.5744 | 0.5886 | 8.4327 |
| SC | var | 10.0132 | 8.0089 | 86.1291 | 24.3110 | 16.9741 | 21.2942 |
|  | SR | 0.0871 | 0.1891 | 0.1202 | 0.0233 | $0.0399$ | 0.2812 |
|  | TURN | 0.0388 | 0.0313 | 0.0411 | 0.0512 | 0.0481 | 0.0247 |
|  | ASP | $1.38 \mathrm{e}-16$ | $1.23 \mathrm{e}-16$ | $-1.52 \mathrm{e}-16$ | 3.12e-16 | -4.19e-16 | 0 |
| SU | var | 9.9181 | 15.0721 | $5.91 \mathrm{e}+03$ | 12.1721 | 9.9999 | 17.2369 |
|  | SR | 0.0250 | 0.1800 | -0.1270 | -0.0482 | -0.0167 | 0.4624 |
|  | TURN | 0.8322 | 2.7947 | 3.9150 | 0.3919 | 0.3722 | 5.2429 |
|  | ASP | 1.1172 | 2.5580 | 5.8870 | 0.5542 | 0.5494 | 6.3456 |
| EW | var | 11.1339 | 8.1780 | 89.8252 | 21.3370 | 19.5742 | 22.2346 |
|  | SR | 0.0701 | 0.1790 | 0.1151 | 0.0194 | 0.0337 | 0.2922 |
|  | TURN | 0.0208 | 0.0253 | 0.0400 | 0.0429 | 0.0381 | 0.0252 |
|  | ASP | $1.13 \mathrm{e}-16$ | $1.13 \mathrm{e}-16$ | $-1.12 \mathrm{e}-16$ | $4.52 \mathrm{e}-16$ | -3.39e-16 | 0 |
| SCID | var | 6.9600 | 6.9782 | $2.38 \mathrm{e}+03$ | 7.3676 | 7.0304 | 16.8727 |
|  | SR | 0.0294 | 0.1097 | -0.1259 | -0.0497 | -0.0025 | 0.6740 |
|  | TURN | 0.4115 | 0.8589 | 0.9940 | 0.2803 | 0.2884 | 1.5393 |
|  | ASP | 0.6362 | 1.6735 | 2.8275 | 0.5727 | 0.5871 | 3.5605 |
| SC1F | var | 6.2670 | 6.5347 | $2.6140 \mathrm{e}+03$ | 7.3790 | 7.0367 | 15.6203 |
|  | SR | 0.0380 | 0.0902 | -0.1253 | -0.0510 | -0.0034 | 0.6570 |
|  | TURN | 0.3032 | 1.1648 | 0.8972 | 0.2964 | 0.2962 | 1.5858 |
|  | ASP | 0.4690 | 1.3556 | 2.1771 | 0.5722 | 0.5863 | 2.8086 |

As for the portfolio turnover, unsurprisingly, the EW portfolio strategy exhibit the lowest turnover of all portfolio strategies, amounting to $3.20 \%$. The RSA and RSQ portfolio strategies have moderate levels of turnover on average, $13.49 \%$ and $13.20 \%$. The highest average turnover is generated by the SU portfolio and, then, by the Box portfolio, amounting on average to $225.81 \%$ and $193.10 \%$, meaning that they are very costly. The turnover of the remaining portfolio strategies range between $11.85 \%$ (L12), $25.76 \%$ (L2), $16.66 \%$ (L1), and $13.98 \%(\mathrm{EN})$, respectively. The high turnover of SU and Box was reflect in the enormous average short positions of over $283.52 \%$ and $333.52 \%$ on average across the six data sets. The second two highest average short positions are by SCID and SC1F, respectively, amounting to $164.29 \%$ and $132.81 \%$. The average short positions of the SC and EW portfolios are on average approximately $0 \%$ across the six data sets. The average short positions of the RSQ and RSA portfolio strategies also tend to zero. Therefore, considering the moderate turnover and the average short positions, the proposed RSQ and RSA strategies represent a practically implementable method that outperform the portfolio strategies listed in Table 1.

### 6.4. Cumulative Return

In this subsection, we show the CR of several portfolio strategies. We use the FF100 data set. The sparsity level $s=\lceil 10 \% n\rceil$. The parameter $\alpha=\beta=10$. According to the OOS performance, we choose RSQ, RSA, L12, L1, EN, EW, and SC to compare the CR.

Figure 4 shows the curves of the CR over the corresponding investment periods for the different portfolio strategies. Apparently, RSQ and RSA outperform the others with visible margins. However, RSA and RSQ do not produce significant differences. This result suggest that, compared with the other portfolios, the sparse portfolios RSA and RSQ grow more steadily together with a reduced volatility across most of the investment periods.


Figure 4. The cumulative return.

## 7. Conclusions

Portfolio selection has been a fertile area for robust optimization techniques. We proposed a robust and sparse portfolio selection optimization model by considering the perturbation in the asset return matrix and the parameter uncertainty in the expected asset return. We used the equivalence of robustness and regularization to deal with the perturbation in the asset return matrix. To deal with the uncertainty in the expected asset return, we considered two kinds of uncertainty sets and solved the worst-case scenario. We defined three types of stationary points of the penalty problem and then analyzed the relationship between these stationary points and local/global minimizers. Then, we
designed the penalty alternating direction method to solve each problem. Although there is no theoretical guarantee for the equivalence between problems (5) and (6), as well as problems (8) and (18), we confirmed the effectiveness of our approach by comparing with nine widely studied portfolio models on seven real-world data sets. Extensive numerical experiments confirm that the portfolios we proposed have lower volatility, that is less risk. Moreover, our portfolio strategies can yield higher Sharpe ratios when the appropriate parameters are selected.

We note that the robust optimization (RO) mainly consider the uncertainty sets of parameters and thus it do not consider any distribution information of the data. This characteristic makes RO attractive, but at the same time, this method loses the comprehensive characterization of the data. Recently, distributed robust optimization (DRO) has attracted widespread attention and research. Although DRO takes into account the distribution information of the data, the cost paid is that it is difficult to solve. We will consider how to apply DRO to sparse portfolio problems, while considering the distribution information of financial data and improving the sparsity. The most direct extension is the distributed robust portfolio optimization with the $\ell_{0}$ norm constraint, which is a worthwhile and challenging issue.

Author Contributions: Conceptualization, H.Z.; methodology, H.Z. and Y.J.; software, H.Z., Y.J., and Y.Y.; formal analysis, H.Z. and Y.J.; writing-original draft preparation, H.Z.; writing-review and editing, H.Z., Y.J., and Y.Y.; visualization, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.
Data Availability Statement: The data source has been presented in Section 6 of the paper.
Conflicts of Interest: The authors declare no conflict of interest.

## References

1. Markowitz, H. Portfolio selection. J. Financ. 1952, 7, 142-149
2. Brodie, J.; Daubechies, I.; De Mol, C.; Giannone, D.; Loris, I. Sparse and stable markowitz portfolios. Proc. Natl. Acad. Sci. USA 2009, 106, 12267-12272. [CrossRef] [PubMed]
3. Ledoit, O.; Wolf, M. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J. Empir. Financ. 2003, 10, 603-621. [CrossRef]
4. Ledoit, O.; Wolf, M. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. Rev. Financ. Stud. 2017, 30, 4349-4388. [CrossRef]
5. Jagannathan, R.; Ma, T. Risk reduction in large portfolios: Why imposing the wrong constraints helps. J. Financ. 2003, 58, 1651-1683. [CrossRef]
6. DeMiguel, V.; Garlappi, L.; Nogales, F.J.; Uppal, R. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. Manag. Sci. 2009, 55, 798C812. [CrossRef]
7. Ben-Tal, A.; Nemirovski, A.; Roos, C. Robust solutions of uncertain quadratic and conic-quadratic problems. SIAM J. Optim. 2002, 13, 535-560. [CrossRef]
8. El Ghaoui, L.; Oustry, F.; Lebret, H. Robust solutions to uncertain semidefinite programs. SIAM J. Optim. 1998, 9, 33-52. [CrossRef]
9. Lee, Y.; Kim, M.J.; Kim, J.H.; Jang, J.R.; Chang, K.W. Sparse and robust portfolio selection via semi-definite relaxation. J. Oper. Res. 2020, 71, 687-699. [CrossRef]
10. Swain, P.; Ojha, A.K. Robust approach for uncertain portfolio allocation problems under box uncertainty. In Recent Trends in Applied Mathematics: Select Proceedings of AMSE; Springer: Singapore, 2019; pp. 347-356.
11. Goldfarb, D.; Iyengar, G. Robust portfolio selection problems. Math. Oper. Res. 2003, 28, 1-38. [CrossRef]
12. Lobo, M.S.; Boyd, S. The Worst-Case Risk of a Portfolio. Unpublished Manuscript. Available online: Available online: http: / /faculty.fuqua.duke.edu/(2000)\%7Emlobo/bio/researchfiles/rsk-bnd.pdf (accessed on 13 July 2022).
13. Min, L.; Dong, J.; Liu, J.; Gong, X. Robust mean-risk portfolio optimization using machine learning-based trade-off parameter. Appl. Soft Comput. 2021, 113, 107948. [CrossRef]
14. Won, J.H.; Kim, S.J. Robust trade-off portfolio selection. Optim. Eng. 2020, 21, 867-904. [CrossRef]
15. Fabozzi, F.J.; Huang, D.; Zhou, G. Robust portfolios: Contributions from operations research and finance. Ann. Oper. Res. 2010, 176, 191-220. [CrossRef]
16. Scutella, M.G.; Recchia, R. Robust portfolio asset allocation and risk measures. Ann. Oper. Res. 2013, 204, 145-169. [CrossRef]
17. Xidonas, P.; Steuer, R.; Hassapis, C. Robust portfolio optimization: A categorized bibliographic review. Ann. Oper. Res. 2020, 292, 533-552. [CrossRef]
18. Ghahtarani, A.; Saif, A.; Ghasemi, A. Robust portfolio selection problems: A comprehensive review. Oper. Res. 2022, 22, 3203-3264. [CrossRef]
19. Leyffer, S.; Menickelly, M.; Munson, T.; Vanaret, C.; Wild, S.M. A survey of nonlinear robust optimization. INFOR: Inf. Syst. Oper. Res. 2020, 58, 342-373. [CrossRef]
20. Zhao, Z.; Xu, F.; Du, D.; Meihua, W. Robust portfolio rebalancing with cardinality and diversification constraints. Quant. Financ. 2021, 21, 1707-1721. [CrossRef]
21. Tütüncü, R.H.; Koenig, M. Robust asset allocation. Ann. Oper. Res. 2004, 132, 157-187. [CrossRef]
22. Khodamoradi, T.; Salahi, M.; Najafi, A.R. Robust CCMV model with short selling and risk-neutral interest rate. Phys. A Stat. Mech. Its Appl. 2020, 547, 124429. [CrossRef]
23. Fabozzi, F.J.; Kolm, P.N.; Pachamanova, D.A.; Focardi, S.M. Robust portfolio optimization. J. Portf. Manag. 2007, 33, 40. [CrossRef]
24. Pınar, M.Ç. On robust mean-variance portfolios. Optimization 2016, 65, 1039-1048. [CrossRef]
25. Busse, J.A.; Chordia, T.; Jiang, L.; Tang, Y. Transaction costs, portfolio characteristics, and mutual fund performance. Manag. Sci. 2021, 67, 1227-1248. [CrossRef]
26. Hautsch, N.; Voigt, S. Large-scale portfolio allocation under transaction costs and model uncertainty. J. Econom. 2019, 212, 221-240. [CrossRef]
27. Yu, J.R.; Chiou, W.J.P.; Lee, W.Y.; ; Lin, S.J. Portfolio models with return forecasting and transaction costs. Int. Rev. Econ. Financ. 2020, 66, 118-130. [CrossRef]
28. Mordukhovich, B.S.; Nguyen, M.M. An Easy Path to Convex Analysis and Applications; Morgan and Claypool Publishers Series: San Rafael, CA, USA, 2014.
29. Pan, L.; Xiu, N.; Fan, J. Optimality conditions for sparsity nonlinear programming. Sci. China Math. 2017, 60, 759-776. [CrossRef]
30. Rockafellar, R.T.; Wets, R.J. Variational Analysis; Springer: Berlin/Heidelberg, Germany, 1998.
31. Bertsimas, D.; Dunn, J. Machine Learning under a Modern Optimization Lens; Dynamic Ideas LLC: Charlestown, MA, USA, 2019.
32. Beck, A. First-Order Methods in Optimization; Society for Industrial and Applied Mathematics and Mathematical Optimization Society: Philadelphia, PA, USA, 2017.
33. Costa, C.M.; Kreber, D.; Schmidta, M. An alternating method for cardinality-constrained optimization: A computational study for the best subset selection and sparse portfolio problems. Informs J. Comput. 2022, 34, 2968-2988. [CrossRef]
34. Heckel, T.; de Carvahlo, R.; Lu, X.; Perchet, R. Insights into robust optimization: Decomposing into mean-variance and risk-based portfolios. J. Invest. Strateg. 2016, 6, 1-24. [CrossRef]
35. Yin, C.; Perchet, R.; Soupé, F. A practical guide to robust portfolio optimization. Quant. Financ. 2021, 21, 911-928. [CrossRef]
36. Geissler, B.; Morsi, A.; Schewe, L.; Schmidt, M. Penalty alternating direction methods for mixed-integer optimization: A new view on feasibility pumps. SIAM J. Optim. 2017, 27, 1611-1636. [CrossRef]
37. Zhao, H.; Kong, L.; Qi, H.D. Optimal portfolio selections via $\ell_{12}$-norm regularization. Comput. Optim. Appl. 2021, 80, 853-881. [CrossRef]
38. Yen, Y.M.; Yen, T.J. Solving norm constrained portfolio optimization via coordinate-wise descent algorithms. Comput. Stat. Data Anal. 2014, 76, 737-759. [CrossRef]
39. Behr, P.; Guettler, A.; Miebs, F. On portfolio optimization: Imposing the right constraints. J. Bank. Financ. 2013, 37, 1232-1242. [CrossRef]
40. DeMiguel, V.; Garlappi, L.; Uppal, R. Optimal versus naive diversification: How in-efficient is the $1 / n$ portfolio strategy? Rev. Financ. Stud. 2009, 22, 1915-1953. [CrossRef]
41. Ledoit, O.; Wolf, M. A well-conditioned estimator for large-dimensional covariance matrices. J. Multivar. Anal. 2004, 88, 365-411. [CrossRef]
42. Lai, Z.R.; Yang, P.Y.; Fang, L.; Wu, X. Short-term sparse portfolio optimization based on alternating direction method of multipliers. J. Mach. Learn. Res. 2018, 19, 2547-2574.
43. Chou, R.K.; Chung, H. Decimalization, trading costs, and information transmission between etfs and index futures. J. Futur. Mark. 2006, 26, 131-151. [CrossRef]
44. Kan, R.; Wang, X.; Zhou, G. Optimal portfolio choice with estimation risk: No risk-free asset case. Manag. Sci. 2022, 68, 2047-2068. [CrossRef]
45. Mutunge, P.; Haugl, ; D. Minimizing the tracking error of cardinality constrained portfolios. Comput. Oper. Res. 2018, 90, 33-41. [CrossRef]
46. Fan, J.; Zhang, J.; Yu, K. Vast portfolio selection with gross-exposure constraints. J. Am. Stat. Assoc. 2012, 107, 592-606. [CrossRef]
47. Teng, Y.; Yang, L.; Yu, B.; Song, X. A penalty PALM method for sparse portfolio selection problems. Optim. Methods Softw. 2017, 32, 126-147. [CrossRef]
48. Wang, Y.; Gao, S.; Yu, Y.; Cai, Z.; Wang, Z. A gravitational search algorithm with hierarchy and distributed framework. Knowl.Based Syst. 2021, 218, 106877. [CrossRef]
49. Shen, W.; Wang, J.; Ma, S. Doubly regularized portfolio with risk minimization. In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, Québec City, QC, Canada, 27-31 July 2014.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

