



# Sparse and risk diversification portfolio selection

Qian Li<sup>1</sup> · Wei Zhang<sup>2</sup>

Received: 14 September 2021 / Accepted: 15 July 2022 / Published online: 31 July 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

Portfolio risk management has become more important since some unpredictable factors, such as the 2008 financial crisis and the recent COVID-19 crisis. Although the risk can be actively managed by risk diversification, the high transaction cost and managerial concerns ensue by over diversifying portfolio risk. In this paper, we jointly integrate risk diversification and sparse asset selection into mean-variance portfolio framework, and propose an optimal portfolio selection model labeled as JMV. The weighted piecewise quadratic approximation is considered as a penalty promoting sparsity for the asset selection. The variance associated with the marginal risk regard as another penalty term to diversify the risk. By exposing the feature of JMV, we prove that the KKT point of JMV is the local minimizer if the regularization parameter satisfies a mild condition. To solve this model, we introduce the accelerated proximal gradient (APG) algorithm [Wen in *SIAM J. Optim* 27:124–145, 2017], which is one of the most efficient first-order large-scale algorithm. Meanwhile, the APG algorithm is linearly convergent to a local minimizer of the JMV model. Furthermore, empirical analysis consistently demonstrate the theoretical results and the superiority of the JMV model.

**Keywords** Sparse portfolio selection · Non-convex regularization · Accelerated proximal algorithm · Linear convergence

---

✉ Qian Li  
liqian15123329166@163.com

Wei Zhang  
lindelfeel@gmail.com

<sup>1</sup> School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai 201620, China

<sup>2</sup> School of Mathematics, South China University of Technology, Guangzhou 510640, China

## 1 Introduction

Whether the 2008 financial crisis or the recent COVID-19 crisis have a massive impact on companies and industries, making the dramatic fluctuation and recession of the market. In order to safeguard the portfolio and its value, the portfolio risk management becomes particularly important. The global minimum variance portfolio model (GMV) [14] is widely used in a volatile market, which belongs to the Markowitz mean-variance (MV) portfolio framework [21]. Although the GMV model has a good performance in the out-of-sample tests, it still suffers from estimation error of the covariance matrix [8, 10, 16] and tends to highly concentrated portfolios on a few assets [22]. Indeed, diversifying risk according to the risk contributors plays an crucial role in modern portfolio risk management.

The marginal risk was first introduced by CreditMetrics [23] to measure the risk contribution of a given asset, which is defined as the difference between the risk of the portfolio and the risk of the portfolio without the given asset. Specifically, Zhu et al. [35] defined the marginal risk by decomposing the covariance matrix of the asset return, and proposed a portfolio selection model with marginal risk control. It has been found by empirical study the model with marginal risk control is a suitable analytical tool for active portfolio risk management. Li et al. [19] used a factor model to capture the systematic risk and proposed the concepts of marginal systematic risk and relative marginal systematic risk. Then these two concepts were integrated respectively into the (MV) formulation to construct portfolio optimization model for actively allocating the systematic risk. The above models were solved by the branch-and-bound method, which converges slowly when solving the large scale problems [28]. To overcome the challenge of computational cost, an optimal trade-off model was proposed for portfolio selection with the effect of systematic risk diversification, which can be solved by the efficient accelerated gradient algorithm [17]. Several other approaches on risk diversification are proposed in [20, 24–27].

Although integrating the marginal (systematic) risk into portfolio selection conducts risk diversification, there are still some drawbacks in practical applications. A important issues is the considerable transaction cost and managerial concerns by over diversifying portfolio risk. A directly approach is that introducing the cardinality constraint to limit the total number of positions. There are many efficient methods to solve the cardinality constraint portfolio selection problem such as branch-and-bound method [6] and nonmonotone projected gradient (NPG) method [31]. Another natural approach is to augment the objective function with a penalty on the portfolio weight vector. The most famous convex penalty approach is adding an  $\ell_1$  norm penalty to the Markowitz framework [1, 7], which encourages sparse and stable portfolios [1, 10]. While the  $\ell_1$  norm penalty is ineffective in inducing sparsity with the budget and no short selling constraints, an alternative is the use of weighted  $\ell_1$  norm penalty [11]. To promote sparsity and countervail the shortcomings of convex penalty related to large biased coefficient values [12], Fastrich et al. [11] apply non-convex penalties, including  $\ell_q$ -penalty [4, 5, 32], smoothly clipped absolute deviation [9],

log-sum penalty [3] and minimax concave penalty [33], to identify sparse and stable portfolios with desirable out-of-sample properties. Recently, Li et al. [18] introduced the weighted piecewise quadratic approximation (PQA) function as a new penalty in the MV portfolio framework. By utilizing PQA penalty, one can not only promotes sparsity and has a good out-of-sample performance, but also design a more efficient large-scale optimization algorithm to achieve a local minimizer.

Motivated by the challenges of portfolio risk management and realistic investment, Zhao et al. [34] developed a robust conditional value at risk optimal portfolio rebalancing model with both embedded sparsity and diversification, and then proposed an effective ADMM algorithm to solve this model. In this paper, under the mean-variance framework, we jointly integrate asset selection, risk diversification and some other investment constrains into consideration. An efficient first-order large-scale algorithm based on the structure of optimization problem is presented. Empirical analysis is constructed on the historical market data. The main contributions of this paper are described as follows.

First, we propose an optimal portfolio selection model which jointly considering risk diversification and sparse asset selection, abbreviated as the JMV model. The weighted piecewise quadratic approximation is considered as a penalty promoting sparsity for the asset selection. On the other hand, the risk diversification is also formulated as a penalty based on the definitions of the marginal risk and variance. Besides, by exposing the special structure of the JMV model, we prove that the KKT point of JMV is the local minimizer if the regularization parameter satisfies a mild condition. The numerical tests demonstrate that this condition can be easily satisfied. Then, we introduce the accelerated proximal gradient (APG) algorithm [30] to solve the JMV model. Meanwhile, a algorithm depended on projection onto the probability simplex is presented to solve the subproblem in APG algorithm. Under some mild conditions, the APG algorithm is linearly convergent to a local minimizer of JMV model. Empirical analysis not only demonstrate the theoretical results, but also show that JMV model has a better out-of-sample performance and achieves a better balance among risk diversification, sparsity and some other practical investment factors when compared with the existing models. Furthermore, the efficiency of APG algorithm on JMV model is illustrated in numerical experiments.

The reminder of the paper is organized in this way. In Sect. 2, we review the background of risk diversification and sparsity penalty, which serve as the preliminaries for our optimization model. In Sect. 3, we present an optimal portfolio selection model which jointly considering risk diversification and sparse asset selection. Then, we prove that the KKT point of the JMV model is the local minimizer. In Sect. 4, we present an accelerated proximal gradient (APG) algorithm for solving the JMV model. The convergence property of APG algorithm for the JMV model is derived. To assess the investment performance of the JMV model, some empirical analysis is carried out in Sect. 5. We also conduct some numerical experiments in Sect. 6 to show the efficiency of the algorithm. Finally, a conclusion of this paper is in Sect. 7.

## 2 Preliminaries

In this section, we first recall some relevant portfolio selection models and review the background of risk diversification and sparsity penalty, which serve as the preliminaries for our optimization model.

Suppose there are  $n$  risky assets with the random returns  $r = (r_1, \dots, r_n)^T$  in the financial market. The mean vector and the covariance matrix of  $r$  are denoted as  $\mu$  and  $\Sigma$ , respectively. The Markowitz mean-variance portfolio selection model [21] is formulated as follows:

$$\begin{aligned} \text{MV} \quad & \min \quad x^T \Sigma x - \tau \mu^T x \\ & \text{s.t.} \quad e^T x = 1, \quad x \geq 0. \end{aligned}$$

where  $\tau \geq 0$  is the parameter to balance the risk and return of the portfolio and  $e$  is a vector in which all elements are ones. If  $\tau = 0$ , MV model reduces to the global minimum variance portfolio model, which always has a better performance than MV model in the out-of-sample test. To diversify the risk, the marginal risk is proposed so that the risk contribution of a specific asset can be quantified.

**Definition 1** [35] The marginal risk of asset  $i$  in a portfolio  $x = (x_1, x_2, \dots, x_n)^T$ , denoted by  $MR_i(x)$ , is defined as:

$$MR_i(x) = x^T \Sigma_i x = \sigma_{ii} x_i^2 + 2 \sum_{j \neq i}^n \omega_{ij} x_i x_j \sigma_{ij}, \tag{1}$$

where  $\omega_{ij} = \sigma_{ii} / (\sigma_{ii} + \sigma_{jj})$  for  $i \neq j$ ,  $\omega_{ii} = 1$  and

$$\Sigma_i = \begin{pmatrix} 0 & \cdots & \omega_{i1} \sigma_{i1} & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ \omega_{i1} \sigma_{i1} & \cdots & \sigma_{ii} & \cdots & \omega_{in} \sigma_{in} \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & \omega_{in} \sigma_{in} & \cdots & 0 \end{pmatrix}.$$

Based on this definition,  $\Sigma_i$  is generally an indefinite matrix which have one positive and one negative eigenvalue since it can be decomposed as:

$$\Sigma_i = \alpha_i u_i u_i^T - \gamma_i v_i v_i^T$$

where  $\alpha_i > 0$  and  $-\gamma_i < 0$  are the two non-zero eigenvalues, and  $u_i$  and  $v_i$  are the corresponding orthogonal unit eigenvectors. As we all know, there are various portfolio selection model with marginal risk control. We introduce

$$R(x) = \sum_{i=1}^n (MR_i(x) - \theta)^2$$

to measure the risk concentration, where  $\theta$  is a parameter to balance the average risk contributions of the selected assets. The smaller the quantity  $R(x)$  is, the more

uniformly the risk is distributed among the selected assets. However, this portfolio type usually increases the number of nonzero weights, which implies high transaction cost and managerial concerns.

To control the transaction and monitor cost, sparsity plays an important role in the formulation of investment portfolios. Naturally, the cardinality constraint  $\|x\|_0 \leq K$  is introduced to limit the total number of positions, where  $0 \leq K \leq n$ . The cardinality constraint is always approximated as convex or non-convex regularization. The piecewise quadratic approximation function is proposed as a non-convex regularization to encourage sparse solutions which theoretical and numerical superiority are demonstrated in [18]. Based on the weighted  $\ell_1$  norm and the background of portfolio management, we introduce the non-convex weighted piecewise quadratic approximation function:

$$S(x) = -x^T V^T Vx + 2\|Vx\|_1 = -x^T Wx + 2\|Vx\|_1,$$

where  $V = \text{diag}(w_1, \dots, w_n)$ ,  $w_i > 0$  is the individual regularization weight parameter of asset  $i$ . In practice, the investors can restrict the investment proportions of assets that are predicted to be more volatile or unfavorable. Fortunately, the weight parameter  $w_i$  can be used to control the investment proportion of asset  $i$ .

For simplicity, the support of  $x \in \mathbb{R}^n$  is  $\text{supp}(x) := \{i \mid x_i \neq 0, i = 1, 2, \dots, n\}$ .  $\nabla f(x)$  is the gradient of  $f(x)$ ,  $\nabla^2 f(x)$  is the Hessian matrix of  $f(x)$ . For any matrix  $A, B \in \mathbb{R}^{n \times n}$ ,  $A \geq B$  if and only if  $A - B$  is a positive semidefinite matrix,  $A > B$  if and only if  $A - B$  is a positive definite matrix. For any index set  $S \subset \{1, 2, \dots, n\}$ , we denote  $A_{SS}$  as the sub-matrix of  $A$  with the rows and columns restricted to  $S$ .  $x_S$  represents the sub-vector consisting of only the components  $x_i, i \in S$ .

### 3 JMV model and theory

In this section, we present an optimal portfolio selection model which jointly considering risk diversification and sparse asset selection. And then we develop the theoretical results on local optimality of the proposed model.

In this paper, we propose the following portfolio selection model which jointly considering risk diversification and sparsity:

$$\begin{aligned} \text{JMV} \quad & \min \quad x^T \Sigma x - \tau \mu^T x + \lambda_1 R(x) + \lambda_2 S(x) \\ & \text{s.t.} \quad e^T x = 1, \quad x \geq 0. \end{aligned}$$

where  $\lambda_1, \lambda_2 \geq 0$  are the regularization parameters that control the degrees of risk diversification and sparsity, respectively.

If  $\lambda_2 = 0$ , the JMV model is the risk diversification portfolio selection:

$$\begin{aligned} \text{RDMV} \quad & \min \quad x^T \Sigma x - \tau \mu^T x + \lambda_1 R(x) \\ & \text{s.t.} \quad e^T x = 1, \quad x \geq 0. \end{aligned}$$

If  $\lambda_1 = 0$ , the JMV model is the sparsity portfolio selection:

$$\begin{aligned} \text{SMV} \quad & \min \quad x^\top \Sigma x - \tau \mu^\top x + \lambda_2 S(x) \\ & \text{s.t.} \quad e^\top x = 1, \quad x \geq 0. \end{aligned}$$

Denoting  $q = 2\lambda_2 \nu e - \tau \mu$ , the JMV model can be rewritten as follows:

$$\begin{aligned} \min \quad & x^\top \Sigma x + q^\top x + \lambda_1 R(x) - \lambda_2 x^\top W x \\ & \text{s.t.} \quad e^\top x = 1, \quad x \geq 0. \end{aligned}$$

Denote the objection function as  $F(x)$ . The KKT conditions of this model are:

$$\begin{aligned} e^\top x = 1, \quad x \geq 0, \quad \phi_i x_i = 0, \quad \phi_i \geq 0, \quad i = 1, \dots, n, \\ 2\Sigma x + 4\lambda_1 \sum_{i=1}^n MR_i(x)\Sigma_i x - 4\lambda_1 \theta \Sigma x - 2\lambda_2 W x + q - \nu e - \phi = 0, \quad \nu \in \mathbb{R}, \end{aligned}$$

where  $\phi$  and  $\nu$  are Lagrange multipliers. Let  $x^*$  be the stationary point of the JMV model,  $L = \{i \mid x_i = 0, \phi_i = 0, i = 1, 2, \dots, n\}$ ,  $E$  is the complement of  $L$ ,  $\sigma$  denote the smallest eigenvalue of  $\Sigma_{EE}$ , and  $\omega$  is the largest eigenvalue of  $W_{EE}$ . In the following, we will show that the stationary point  $x^*$  is a local minimizer of JMV under a suitable regularization parameter.

**Theorem 1** *Let  $x^*$  be the stationary point of the JMV model. If the regularization parameters satisfy  $0 \leq 4\lambda_1 \leq 1/\theta$  and  $0 \leq 2\lambda_2 \leq \sigma/\omega$ , then  $x^*$  is a local minimizer of the JMV model.*

**Proof of Theorem 1** By the second-order optimality condition, to justify the theorem, we only need to show that there is a feasible direction  $h$  where  $e^\top h = 0, x^* + h \geq 0, \{h_i = 0 \mid i \in L\}$  and  $\|h\| < \epsilon, \epsilon > 0$ , such that  $h^\top \nabla^2 F(x^*) h > 0$ . For this purpose, we first get

$$\begin{aligned} \nabla^2 F(x^*) &= 2\Sigma - 2\lambda_2 W + 4\lambda_1 \sum_{i=1}^n (2\Sigma_i x^* x^{*\top} \Sigma_i + x^{*\top} \Sigma_i x^* \Sigma_i - \theta \Sigma_i) \\ &= (\Sigma - 2\lambda_2 W) + \left( 8\lambda_1 \sum_{i=1}^n \Sigma_i x^* x^{*\top} \Sigma_i + 4\lambda_1 \sum_{i=1}^n x^{*\top} \Sigma_i x^* \Sigma_i + (1 - 4\lambda_1 \theta) \Sigma \right) \end{aligned}$$

By  $h$  is a feasible direction, we have

$$h^\top (\Sigma - 2\lambda_2 W) h = h_E^\top (\Sigma_{EE} - 2\lambda_2 W_{EE}) h_E.$$

If  $2\lambda_2 \leq \sigma/\omega$ , then  $h^\top (\Sigma - 2\lambda_2 W) h > 0$ .

On the other hand,

$$\begin{aligned}
 & h^\top \left( 8\lambda_1 \sum_{i=1}^n \Sigma_i x^* x^{*\top} \Sigma_i + 4\lambda_1 \sum_{i=1}^n x^{*\top} \Sigma_i x^* \Sigma_i + (1 - 4\lambda_1 \theta) \Sigma \right) h \\
 & \geq \frac{8\lambda_1}{n} (x^{*\top} \Sigma h)^2 + (1 - 4\lambda_1 \theta) h^\top \Sigma h - 4\lambda_1 \sum_{i=1}^n (x^{*\top} \Sigma_i x^*)^2 \sum_{i=1}^n (h^{*\top} \Sigma_i h^*)^2 \\
 & \geq \frac{8\lambda_1 \sigma}{n} \|x^*\|^2 \|h\|^2 + (1 - 4\lambda_1 \theta) \sigma \|h\|^2 - 4\lambda_1 \sum_{i=1}^n (x^{*\top} \Sigma_i x^*)^2 \sum_{i=1}^n \alpha_i^2 \|h\|^4 \\
 & > 0
 \end{aligned}$$

If  $0 \leq 4\lambda_1 \leq 1/\theta$  and

$$\|h\|^2 < \frac{1}{4\lambda_1 \sum_{i=1}^n (x^{*\top} \Sigma_i x^*)^2 \sum_{i=1}^n \alpha_i^2} \left( \frac{8\lambda_1 \sigma}{n} \|x^*\|^2 + (1 - 4\lambda_1 \theta) \sigma \right),$$

then the above last inequality hold.

In conclusion, if  $0 \leq 4\lambda_1 \leq 1/\theta$  and  $0 \leq 2\lambda_2 \leq \sigma/\omega$ ,  $x^*$  is a local minimizer of the JMV model. □

In numerical experiments, we will illustrate that setting  $0 \leq 4\lambda_1 \leq 1/\theta$  and  $0 \leq 2\lambda_2 \leq \sigma/\omega$  are reasonable and enough to get the desired optimal portfolios.

### 4 Accelerated proximal gradient algorithm

In this section, we present an accelerated proximal gradient (APG) algorithm for solving the JMV model.

#### 4.1 APG algorithm for JMV

Denote  $f(x) = x^\top \Sigma x + \lambda_1 R(x) - \lambda_2 x^\top W x$ ,

$$g(x) = \begin{cases} q^\top x, & x \in X, \\ +\infty, & x \notin X, \end{cases}$$

where  $X = \{x \in \mathbb{R}^n \mid e^\top x = 1, x \geq 0\}$ . Then, the JMV model can be expressed as following:

$$\min_{x \in \mathbb{R}^n} F(x) = f(x) + g(x). \tag{2}$$

Actually,  $X$  is equivalent to the feasible set of the above optimization problem. We use  $\mathcal{X}$  to denote the set of stationary points of  $F(x)$ . We choose a sufficiently large constant  $l$  such that

$$f_1(x) = f(x) + \frac{l}{2} \|x\|^2, \quad f_2(x) = \frac{l}{2} \|x\|^2$$

are continuously differentiable and convex. On the feasible set  $X$ ,  $\nabla f_1$  and  $\nabla f_2$  are Lipschitz continuous with modulus  $L > 0$  and  $l > 0$ , respectively. Moreover, by taking a larger  $L$  if necessary, we assume throughout that  $L > l$ . Thus,  $f(x) = f_1(x) - f_2(x)$  has Lipschitz continuous gradient, the Lipschitz continuity constants of  $\nabla f$  is  $L$ .

The accelerated proximal algorithm for solving the JMV model is as following:

---

**Algorithm 1** APG Algorithm

---

**Input:**  $\beta_0 = 0, \beta_k \in [0, \sqrt{\frac{L}{L+l}}], k \geq 1$  and  $\epsilon > 0$ .

**Step 0:** Take  $x^{-1} = x^0 \in \mathbb{R}^n, k = 0$ .

**Step 1:** Compute

$$y^k = (1 + \beta_k)x^k - \beta_k x^{k-1},$$

$$x^{k+1} = \text{Prox}_{\frac{1}{L}g} \left( y^k - \frac{1}{L} \nabla f(y^k) \right) \tag{3}$$

**Step 3:** If  $\|x^{k+1} - x^k\| \leq \epsilon$ , **Stop**. Otherwise,  $k := k+1$  and go to **Step 1**.

---

By the definition of the proximal operator, we note that the Eq. (3) is equivalently given by

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \left\{ \langle \nabla f(y^k), x \rangle + \frac{L}{2} \|x - y^k\|^2 + g(x) \right\}. \tag{3}$$

This subproblem is the key iteration of APG algorithm. Next, we first introduce how to solve the subproblem and then analysis the convergence of the APG algorithm for JMV model.

### 4.2 Solving the subproblem

The subproblem (3) can be expressed as the following optimization problem:

$$\begin{aligned} \min \quad & \frac{1}{2} \|x - B(y)\|^2 \\ \text{s.t.} \quad & e^\top x = 1, \quad x \geq 0, \end{aligned} \tag{4}$$

where

$$B(y) = \left[ I - \frac{1}{L} (2\Sigma + 4\lambda_1 \sum_{i=1}^n MR_i(y)\Sigma_i - 4\lambda_1 \theta \Sigma - 2\lambda_2 W) \right] y - \frac{1}{L} q.$$

This is to compute the Euclidean projection of a point  $B(y)$  onto the probability simplex. We adopt the algorithm in [29] to solve this problem.



### 4.3 Convergence of APG algorithm

By the convergence properties of the APG algorithm in [30], we can get that any accumulation point of the sequence  $\{x^k\}$  generated by APG algorithm is a stationary point of JMV model. Furthermore, under the following Assumption 1, the local linear convergence rate of the sequence  $\{x^k\}$  and  $\{F(x^k)\}$  are hold.

**Assumption 1** [30]

- (i) For any  $\xi \geq \inf_{x \in \mathbb{R}^n} F(x)$ , there exist  $\epsilon > 0$  and  $\tau > 0$  such that

$$\text{dist}(x, \mathcal{X}) \leq \nu \left\| \text{Prox}_{\frac{1}{L}g} \left( x - \frac{1}{L} \nabla f(x) \right) - x \right\|$$

whenever  $\left\| \text{Prox}_{\frac{1}{L}g} \left( x - \frac{1}{L} \nabla f(x) \right) - x \right\| < \epsilon$  and  $F(x) \leq \xi$ , where

$$\text{dist}(x, \mathcal{X}) = \inf_{y \in \mathcal{X}} \|x - y\|.$$

- (ii) There exists  $\delta > 0$ , such that  $\|x - y\| \geq \delta$  whenever  $x, y \in \mathcal{X}, F(x) \neq F(y)$

Next, we first justify the above assumption is satisfied in the JMV model.

**Lemma 1** *If the regularization parameters satisfy  $0 \leq 4\lambda_1 \leq 1/\theta$  and  $0 \leq 2\lambda_2 \leq \sigma/\omega$ , then the objective function  $F(x)$  of JMV model satisfies the Assumption 1.*

**Proof of Theorem 1** i) First, we prove that if there exists  $\epsilon > 0$  such that, for any  $x \in X$  with  $\left\| \text{Prox}_{\frac{1}{L}g} \left( x - \frac{1}{L} \nabla f(x) \right) - x \right\| < \epsilon$ , then it has a  $x^* \in \mathcal{X}$  such that  $\text{supp}(x) = \text{supp}(x^*)$ . By contraction, if the claim does not hold, there would exist an  $S \subseteq \{1, 2, \dots, n\}$  and a sequence of vectors  $\{x^1, x^2, \dots\}$  satisfying  $\text{supp}(x^r) = S$  for all  $r$  and  $x^r - z^r \rightarrow 0$  where  $z^r = \text{Prox}_{\frac{1}{L}g} \left( x^r - \frac{1}{L} \nabla f(x^r) \right)$ , and yet there is no  $x^* \in \mathcal{X}$  for  $\text{supp}(x^*) = S$ . By [29], the projection  $z^r$  can be easily determined through  $B(x^r)$ . Denote

$$\rho = \max \{ 1 \leq i \leq n : \eta_i + \frac{1}{i} (1 - \sum_{j=1}^i \eta_j) \},$$

where  $\eta_1 \geq \eta_2 \geq \dots \eta_n$  is the sorted sequence of the elements of  $B(x^r)$ . The elements of the projection can be written as

$$z_i^r = \max \{ B(x^r)_i + \delta, 0 \}, \quad i = 1, \dots, n, \quad \text{where } \delta = \frac{1}{\rho} (1 - \sum_{j=1}^{\rho} u_j).$$

That is to say, only the dimensions corresponding to the largest  $\rho$  elements of  $B(x^r)$  are nonzero. Since  $\{(x^r, z^r)\}$  are bounded, then every one of its cluster points  $(\hat{x}, \hat{z})$  satisfies

$$\hat{z} = \hat{x} = \text{Prox}_{\frac{1}{L}g} \left( \hat{x} - \frac{1}{L} \nabla f(\hat{x}) \right).$$

As a result, the largest  $\rho$  elements of  $B(x^r)$  will be in the same dimensions as in  $B(\hat{x})$  when  $r$  is large enough,

$$\text{supp}(x^r) = \text{supp}(\hat{x}) = S.$$

Moreover,  $\hat{x}$  is a stationary point of JMV model, i.e.  $\hat{x} \in \mathcal{X}$ , which contradicts our earlier hypothesis. Therefore, it has a  $x^* \in \mathcal{X}$  such that  $\text{supp}(x) = \text{supp}(x^*)$ .

Next, we prove that the Assumption 1 i) holds. When  $\|z - x\| < \epsilon$ ,  $z = \text{Prox}_{\frac{1}{L}g} \left( x - \frac{1}{L} \nabla f(x) \right)$ , there exists a  $x^* \in \mathcal{X}$  such that  $\text{supp}(x) = \text{supp}(x^*)$  and  $\text{dist}(x, \mathcal{X}) = \|x - x^*\|$ . Since

$$\|x - x^*\| \leq \|x - z\| + \|z - x^*\| \tag{5}$$

and

$$\begin{aligned} \|z - x^*\| \leq & \left\| \left[ I - \frac{1}{L} (2\Sigma + 4\lambda_1 \sum_{i=1}^n (2\Sigma_i \xi \xi^\top \Sigma_i \right. \right. \\ & \left. \left. + \xi^\top \Sigma_i \xi \Sigma_i) - 4\lambda_1 \theta \Sigma - 2\lambda_2 W) \right] (x - x^*) \right\| \end{aligned}$$

where  $\xi$  is between  $x$  and  $x^*$ . By Theorem 1, we know that  $h^\top \nabla F(x^*) h > 0$ , where  $\text{supp}(h) = S$ . So if  $\epsilon$  is small enough, we have

$$0 < \left( 2\Sigma + 4\lambda_1 \sum_{i=1}^n (2\Sigma_i \xi \xi^\top \Sigma_i + \xi^\top \Sigma_i \xi \Sigma_i) \right)_{SS} < LI.$$

Thus

$$\|z - x^*\| \leq c \|x - x^*\|, \quad 0 < c < 1. \tag{6}$$

Combining (5) and (6), we obtain that there exists  $\nu > 0$  such that

$$\|x - x^*\| \leq \nu \|x - z\| = \nu \left\| x - \text{Prox}_{\frac{1}{L}g} \left( x - \frac{1}{L} \nabla f(x) \right) \right\|.$$

This proves i).

ii) Based on  $\nabla F(x^*)$  is positive definite, it is easy to get that ii) holds. □

In what follows, we can get the linear convergence rate of the APG algorithm for JMV model.

**Theorem 2** *Let  $\{x^k\}$  be the sequence generated by the APG algorithm, if  $0 \leq 4\lambda_1 \leq 1/\theta$  and  $0 \leq 2\lambda_2 \leq \sigma/\omega$ , then  $\{x^k\}$  is linearly convergent to a local minimizer of the JMV model.*

## 5 Empirical analysis

In this section, some empirical analysis are conducted on the proposed JMV, not only to evaluate its out-of-sample performance by employing real market data, but also to demonstrate the validity of our theoretical analysis on JMV.

### 5.1 Data and models

In the following empirical analysis, we compare the portfolio performance of the proposed JMV with these models:

- MV: The Markowitz minimum-variance portfolio model [21].
- EW: Equally-weighted risk contributions [20].
- LMV: The weighted  $\ell_1$  regularized portfolio selection model [11].
- SMV: The sparse portfolio selection model.
- RDMV: The risk diversification portfolio selection model.

The optimal solutions of the MV and LMV models are computed by the optimization package CVX [13]. The equally-weighted risk contributions solution is solved as in [20]. The accelerated proximal gradient algorithm are used to solve the SMV, RDMV and JMV models, where we terminate the algorithms when an  $\epsilon$ -optimal solution is achieved or the number of iterations exceed 3000. The error precision is set to  $\epsilon = 10^{-5}$ .

To compare the above models, three different data sets are considered, including the weekly returns of the S &P 500, as well as the monthly returns of the 48 and 100 Fama French portfolios. The detail information of data set are list in Table 1. Note that the investment period includes the 2008 financial crisis, which is selected on purpose to test the portfolios performance in an unstable investment environment. Consequently, their ability to resist dramatic fluctuations can be compared.

We set  $\tau = 0$  in the tested models. The weight parameters  $w$  in the LMV, SMV and JMV models are determined by considering specific financial time series properties, as suggested in [11]. The acceleration parameters in the APG algorithm are selected as

$$\beta_k \equiv 0.98 \sqrt{\frac{L}{L + l}}$$

The parameter  $\theta$  is set as

**Table 1** Information of data set

No	Data set	Time period	Source	Frequency
1	S &P 500	5/1/2006 – 13/12/2011	Datastream	Weekly
2	FF-48	01/2000 – 02/2020	K. French	Monthly
3	FF-100	01/2000 – 02/2020	K. French	Monthly

$$\theta = \frac{\bar{x}^\top \Sigma \bar{x}}{\bar{K}},$$

where  $\bar{x}$  is the optimal portfolio solution of the MV model,  $\bar{K} = \|\bar{x}\|_0$ .

### 5.2 Performance measures

We demonstrate the theoretical results and evaluate the out-of-sample performance of the proposed JMV by the two tests included in-sample test on JMV and out-of-sample evaluation via realistic investment.

We use backtesting to compare the out-of-sample performances of the optimal portfolios generated by the tested models. The tests are performed in a rolling horizon fashion as follows: We choose a window with the size of  $\Gamma = 100$  to construct the estimation. The tested models are then respectively solved to generate portfolio strategies for the following 10 weeks. At  $\Gamma + 11$ , the estimation is updated using the data from 11 up to  $\Gamma + 10$ . The models are re-solved to produce new optimal portfolios for the following 10 weeks. The above procedure repeats until the end of the out-of-sample period.

We utilize the optimal portfolios generated by the tested models to compute the following performance measures. The out-of-sample mean ( $\bar{r}$ ), the out-of-sample risk ( $s^2$ ) and the out-of-sample Sharpe ratio (Sh) are defined as:

$$\bar{r} = \frac{1}{T - \Gamma} \sum_{t=\Gamma}^{T-1} r_{t+1}^\top x_t, \quad s = \sqrt{\frac{1}{T - \Gamma - 1} \sum_{t=\Gamma}^{T-1} (r_{t+1}^\top x_t - \bar{r})^2}, \quad \text{Sh} = \frac{\bar{r}}{s}.$$

where  $r_t$  is the random return vector at time  $t$ ,  $x_t$  is the optimal portfolio vector at time  $t$ . The number of selected assets for investment (No) is computed as:

$$\text{No} = \frac{1}{\Gamma} \sum_{i=1}^{\Gamma} |\text{supp}(x_i)|.$$

The larger the value of the number, the more assets need to manage. Suppose that there are  $M$  selected assets and the maximum marginal risk (MMR) in selected assets is:

$$\text{MMR} = \max_{i=1,2,\dots,M} \text{MR}_i(x).$$

The smaller the value of MMR, the better effect of risk diversification. The turnover (TO) represents the average weekly trading volume, which is defined as:

$$\text{TO} = \frac{1}{T - \Gamma - 1} \sum_{t=\Gamma}^{T-1} \|x_{t+1} - x_t\|_1.$$

Taking the transaction fee into account, a large turnover will wipe out the gains of portfolios. Let  $W_0$  is the initial wealth, the cumulative profit ( $W$ ) is

$$W = W_0 \prod_{t=\Gamma}^{T-1} (1 + r_{t+1}^\top x_t).$$

Transaction cost is closely related to the turnover, in our empirical analysis, we consider that the transaction cost is 1% of the trading volume for each trade (selling and buying) of asset. Then the net profit can be computed as

$$\text{Net Profit} = \text{Profit} - \text{Transaction Cost}.$$

### 5.3 Empirical results

#### 5.3.1 In-sample test on JMV

We consider a portfolio selection problem with historical data from the S &P 500, where the weekly returns of 50 randomly selected stocks from 5th, January 2006 to 27th, September 2010 are used to construct the test model. We set a grid of 90 ascending values from 8000 to 39600 and from  $8 \times 10^{-5}$  to  $3.96 \times 10^{-5}$  for the regularization parameter  $\lambda_1$  and  $\lambda_2$  in the JMV model, respectively.

Figure 1 shows how the sparsity, Sharpe ratio, risk and max marginal risk of the optimal portfolios generated by JMV vary with the regularization parameters  $\lambda_1$  and  $\lambda_2$ . Unsurprisingly, we can see that when the value of  $\lambda_1$  increases, the max marginal

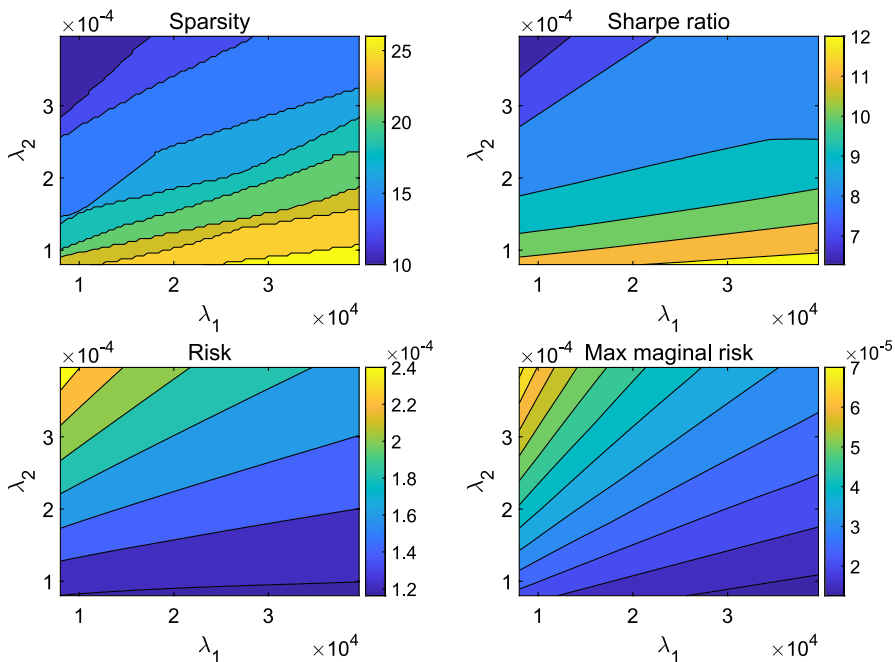


Fig. 1 No, Sh,  $s^2$  and MMR for Optimal Portfolios Generated by JMV

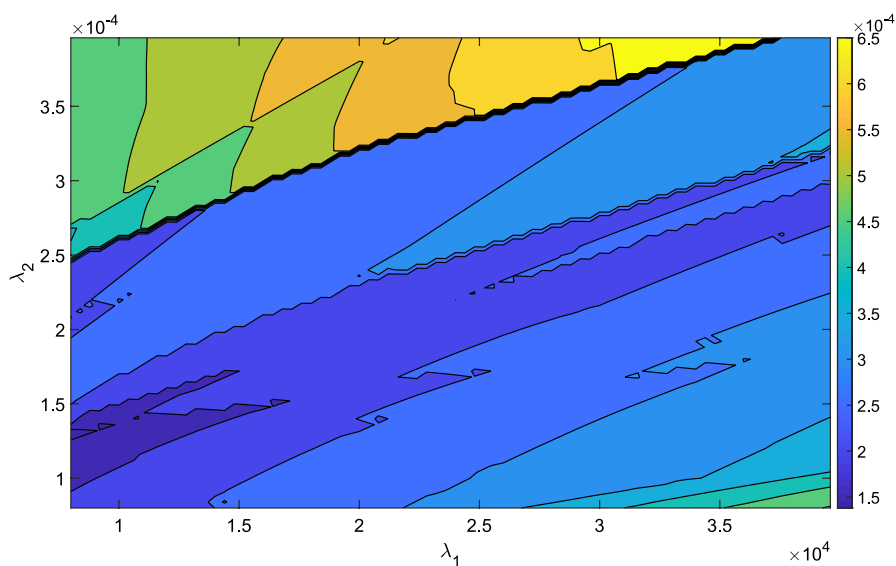
risk is decreasing and the number of nonzero assets in the optimal portfolio is both decreasing. When the value of  $\lambda_2$  increases, the max marginal risk and the number of nonzero assets in the optimal portfolio are all increasing. Moreover, the values of the sparsity, Sharpe ratio, risk and max marginal risk are more sensitive to the parameter  $\lambda_1$ . With the increase of  $\lambda_1$ , the Sharpe ratio of the optimal portfolio decreases and the risk of the optimal portfolio increases. On the other hand, with the increase of  $\lambda_2$ , the Sharpe ratio of the optimal portfolio increases and the risk of the optimal portfolio decreases.

Note that, in more than 50% area, the value of the Sharpe ratio is basically similar from 8.36 to 9.15, while the values of the sparsity and the max marginal risk are quite different. This illustrates that JMV model can generate a portfolio with a sufficient sparsity level, a good effect of risk diversification and a large Sharpe ratio.

Moreover, to demonstrate the theoretical guarantee of the local minimizer of JMV, Fig. 2 shows how the smallest eigenvalue of Hessian matrix for nonzero assets vary with  $\lambda_1$  and  $\lambda_2$  in this test. From Fig. 2, we can see that the smallest eigenvalue of Hessian matrix are lager than zero, which means the JMV model can achieve the local minimizer in numerical experiments.

### 5.3.2 Out-of-sample evaluation via realistic investment

To further demonstrate the strength of JMV, we report the real market performances of different models via backtesting on the three data sets shown in Table 1. The initial wealth is set as 100 at the beginning of backtesting. For each rolling window, a sufficiently wide range of regularization parameter in tested models is implemented to produce an ensemble of portfolios containing different



**Fig. 2** The smallest eigenvalue of Hessian matrix for different  $\lambda_1$  and  $\lambda_2$

numbers of nonzero portfolios and effect of risk diversification. Among these, we select the best regularization parameters in tested models which produces a portfolio with the largest Sharpe ratio.

Fig. 3 illustrates the portfolio values from 6 December 2007 to 13 December 2011 for S &P 500. During the time period from December, 2007 to January, 2009, the portfolio values of all the tested models are decreasing, where the EW decreasing fastest. Notice that the period from December, 2007 to January, 2009 is mainly a period featured by financial crisis. Moreover, we set  $\tau = 0$  in out-of-sample analysis which means that we focus on risk minimization in the MV model, so the performance of the JMV model is a little better than that of the MV model. The financial market begin growing in the beginning of 2009. From the beginning of 2009 to the end of 2010, the portfolio values of the JMV and RDMV models preform better than those of the EW, LMV, SMV and MV models, among which the performance of MV is the worst. At the end of 2010 the economy began to grow steadily, the performance of the EW, SMV and LMV models is getting better. Figures 4 and 5 illustrate the portfolio values from June 2008 to February 2020 for FF 48 and FF 100, respectively. They exhibit that the portfolio values of JMV perform the best during the whole period, compared with other tested models.

In addition, the results are shown in Table 2, including the number of nonzero portfolios, turnover and net profit of the optimal portfolios generated by all the tested models for S &P 500, FF 48, and FF 100. Looking at the number of nonzero portfolios in Table 2, we note that the risk of the portfolio generated by JMV model is neither concentrated on a few assets nor distributed to many assets. SMV model can produce optimal portfolios with the fewest number of nonzero portfolios for FF48 and FF100, and fewer number than MV, EW, RDMV and JMV for S &P 500. With regard to the turnover, equally-weighted risk contributions portfolio is the lowest. Because equally-weighted risk contributions portfolio assigns the capital equally to all stocks, which making it difficult to manage.

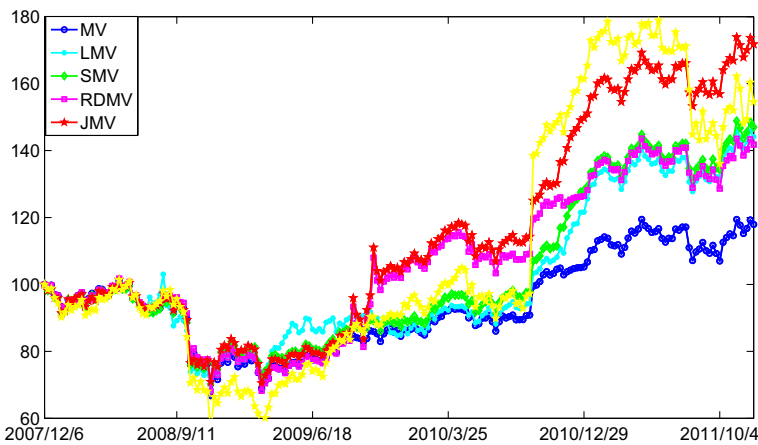


Fig. 3 Evolution of portfolio values for S &P 500

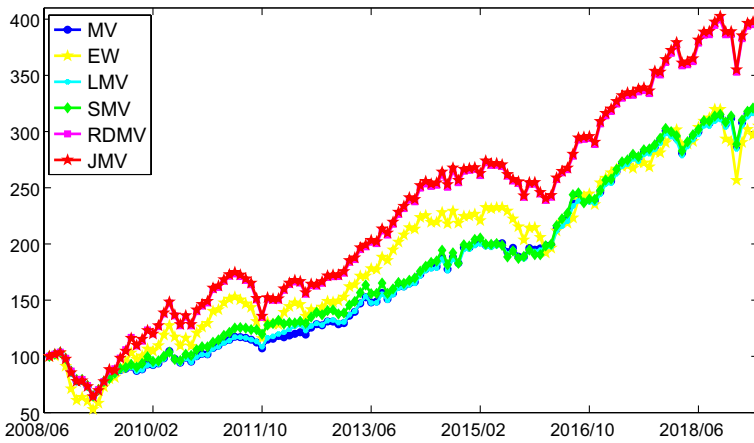


Fig. 4 Evolution of portfolio values for FF 48

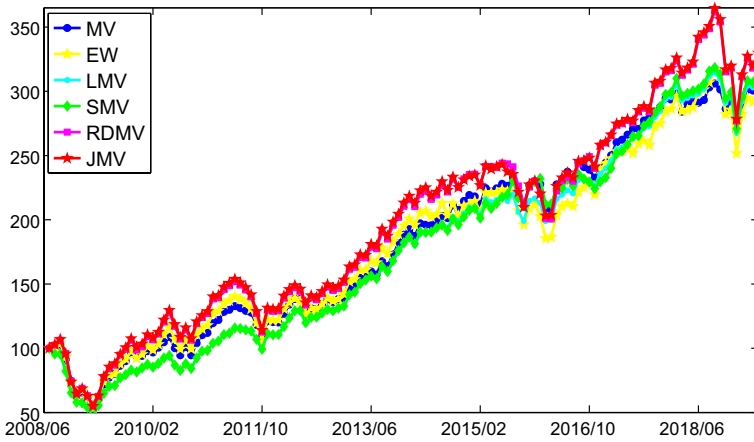


Fig. 5 Evolution of portfolio values for FF 100

Table 2 Recovery results of a signal with different number of samples

		MV	EW	LMV	SMV	RDMV	JMV
S & P 500	No	11.905	50	7.810	9.238	14.476	10.761
	To	0.303	0.0864	0.992	0.635	0.598	0.895
	Net profit	19.863	57.840	46.390	49.356	44.034	73.846
FF 48	No	5.923	48	5.231	2.615	39.846	38.523
	To	0.223	0.041	0.389	0.226	0.584	0.605
	Net profit	227.713	204.997	225.164	229.997	305.725	308.349
FF 100	No	4.9231	98	6.231	4.231	68.077	60.846
	To	0.527	0.032	0.574	0.819	0.6845	0.701
	Net profit	210.809	202.093	216.338	220.392	228.053	229.800



JMV is comparatively similar with RDMV in the aspect of turnover. However, in the terms of profit, the JMV model generates the largest net profit.

### 6 Numerical experiments

In this section, we compare the computational results of APG algorithm on JMV with the other two important first-order algorithms, including the proximal gradient algorithm [15] and the fast iterative shrinkage-thresholding algorithm (FISTA) [2]. The APG algorithm and FISTA are the accelerated first-order algorithms, the difference is that the acceleration parameters are different. The numerical experiments are carried out using MATLAB 2018(a) on a PC with 2.50GHZ CPU processor and 8GB of RAM.

The randomly generated problems are used to test these algorithms. The test problems are randomly generated by the same method mentioned in Li et al. [19]: The mean vector  $\mu = (\mu_1, \mu_2, \dots, \mu_n)^T$ . The variance matrix

$$\Sigma = \left( \sum_{l=1}^m \sum_{k=1}^m \eta_{ik} \sigma_{kl} \eta_{jl} \right)_{n \times n} + \text{diag}(\sigma_{\epsilon_1}, \sigma_{\epsilon_2}, \dots, \sigma_{\epsilon_n}).$$

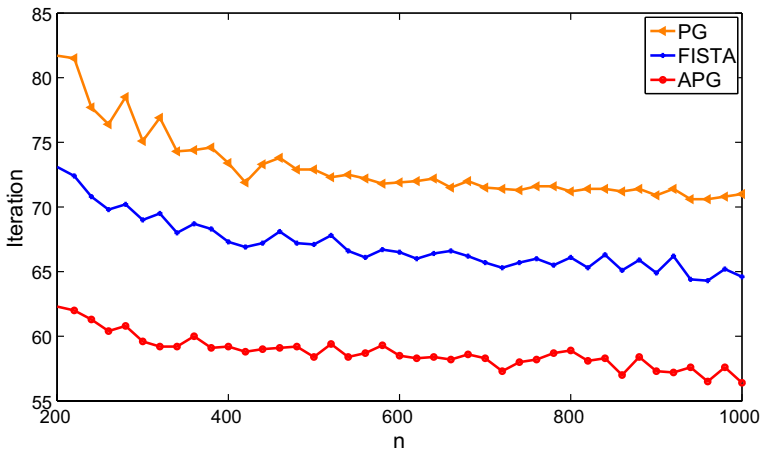
- $\mu_i \in [0, 0.03]$  and  $\sigma_{\epsilon_i} \in [0, 0.002]$  are generated by the uniform distribution,  $i = 1, 2, \dots, n$ .
- $\sigma_{ij}$  is calculated with randomly sampled series from  $[0, 0.03]$ ,  $i = 1, 2, \dots, m$ ,  $j = 1, 2, \dots, m$ ,  $m = n/10$ .
- $\eta_{ij} \in [0.3, 2]/m$  is randomly generated by the uniform distribution,  $i = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, m$ .

In this part, the regularization parameters are set as  $\lambda_1 = 100$ ,  $\lambda_2 = 0.00001$ . The weighted parameters are set as  $w_i = 1/2$ ,  $i = 1, 2, \dots, n$ , while the error precision is set as  $\epsilon = 10^{-5}$ .

The comparison results of three first-order algorithms are summarized in Fig. 6, where  $n$  is the dimension of the problem and ‘‘Iteration’’ is the average number of iterations consumed to solve the ten randomly generated test problems. We can see that the accelerated first-order algorithms outperform the PG algorithm. Moreover, the APG algorithm is always the fastest one.

### 7 Conclusion

In this paper, we have utilized the weighted piecewise quadratic approximation function as a penalty term to promote sparsity, and the variance associated with the marginal risk as another penalty term to diversify the risk. Then, the optimal portfolio selection model which jointly considering risk diversification and sparse asset selection have been proposed. By exposing the feature of the JMV model, it is proved that



**Fig. 6** Comparisons of iterations of three first-order algorithms

the KKT point of JMV is the local minimizer under a mild condition. To solve this model, we have introduced the accelerated proximal gradient (APG) algorithm and presented the linear convergence rate of APG algorithm for solving JMV.

To demonstrate the validity and usefulness of this model, empirical analysis have been constructed on the historical datas of the S & P 500, FF48 and FF100. In-sample test is shown that JMV can attain portfolios with a sufficient sparsity level, a good effect of risk diversification and a large Sharpe ratio. Compared with the MV, EW, LMV, SMV, and RDMV models, JMV has a better out-of-sample performance and achieves a better balance among risk diversification, sparsity and some other practical investment factors. Numerical experiments also illustrate the superiority of APG algorithm on JMV model.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China (Grants No. 11901382) and the Shanghai Chenguang Project (Grants No. 19CG67).

## References

1. Brodie, J., Daubechies, I., Mol, C.D., Giannone, D., Loris, I.: Sparse and stable markowitz portfolios. *Proc. Natl. Acad. Sci. U.S.A.* **106**(30), 12267–12272 (2007)
2. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**, 183–202 (2009)
3. Candès, E.J., Wakin, M.B., Boyd, S.P.: Enhancing sparsity by reweighted  $\ell_1$  minimization. *J. Fourier Anal. Appl.* **14**(5), 877–905 (2008)
4. Chartrand, R.: Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Process Lett.* **14**(10), 707–710 (2007)
5. Chartrand, R., Staneva, V.: Restricted isometry properties and nonconvex compressive sensing. *Inverse Probl.* **24**(3), 20–35 (2008)
6. Cui, X.T., Zheng, X.J., Zhu, S.S., Sun, X.L.: Convex relaxations and MIQCQP reformulations for a class of cardinality-constrained portfolio selection problems. *J. Glob. Optim.* **56**(4), 1409–1423 (2013)
8. DeMiguel, V., Nogales, F.J.: Portfolio selection with robust estimation. *Oper. Res.* **57**(3), 560–577 (2009)

7. DeMiguel, V., Garlappi, L., Nogales, F., Uppal, R.: A generalized approach to portfolio optimization: improving performance by constraining portfolio norms. *Manage. Sci.* **55**(5), 798–812 (2009)
9. Fan, J., Li, R.Z.: Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **96**(456), 1348–1360 (2001)
10. Fan, J.Q., Zhang, J.J., Ke, Y.: Vast portfolio selection with gross-exposure constraints. *J. Am. Stat. Assoc.* **107**(498), 592–606 (2012)
11. Fastrich, B., Paterlini, S., Winker, P.: Constructing optimal sparse portfolios using regularization methods. *Comput. Manag. Sci.* **12**(3), 417–434 (2015)
12. Gasso, G., Rakotomamonjy, A., Canu, S.: Recovering sparse signals with a certain family of non-convex penalties and DC programming. *IEEE Trans. Signal Process.* **57**(12), 4686–4698 (2009)
13. Grant, M., Boyd, S.: CVX: Matlab software for disciplined convex programming, Version 2.1. <http://cvxr.com/cvx/> (2014)
14. Jagannathan, R., Ma, T.S.: Risk reduction in large portfolios: why imposing the wrong constraints helps. *J. Finance.* **58**(4), 1651–1684 (2003)
15. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* **16**, 964–979 (1979)
16. Ledoit, O., Wolf, M.: Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empir. Finance.* **10**(5), 603–621 (2003)
19. Li, Y.J., Zhu, S.S., Li, D.H., Li, D.: Active allocation of systematic risk and control of risk sensitivity in portfolio optimization. *Eur. J. Oper. Res.* **228**(3), 556–570 (2013)
17. Li, Q., Bai, Y.Q., Yan, X., Zhang, W.: Portfolio selection with the effect of systematic risk diversification: formulation and accelerated gradient algorithm. *Optim. Methods Softw.* **34**(3), 612–633 (2019)
18. Li, Q., Bai, Y.Q., Yu, C.J., Yuan, Y.X.: A new piecewise quadratic approximation approach for  $\ell_0$  norm minimization problem. *Sci. China Math.* **62**(1), 185–204 (2019)
20. Maillard, S., Roncalli, T., Teiletche, J.: The properties of equally weighted risk contribution portfolios. *J. Portf. Manag.* **36**, 60–70 (2010)
21. Markowitz, H.: Portfolio selection. *J. Finance.* **7**(1), 77–91 (1952)
22. Michaud, R.O.: The Markowitz optimization enigma: is ‘optimized optimal’? *Financ. Anal. J.* **45**(1), 43–54 (1989)
23. Morgan, J.P.: CreditMetrics technical document. J.P. Morgan, New York (1997)
24. Qian, E.: Risk parity portfolios: efficient portfolios through true diversification. Panagora Asset Manag. Technical report: <https://www.panagora.com/assets/PanAgora-Risk-Parity-Portfolios-Efficient-Portfolios-Through-True-Diversification.pdf> (2005)
25. Qian, E.: Risk parity and diversification. *J. Invest.* **20**(1), 119–127 (2011)
26. Roncalli, T.: Introduction to risk parity and budgeting. Chapman & Hall/CRC financial mathematics series. CRC Press, Boca Raton (2014)
27. Spinu, F.: An algorithm for computing risk parity weights. SSRN. <http://ssrn.com/abstract=2297383> (2013)
28. Sun, X.L., Zheng, X.J., Li, D.: Recent advances in mathematical programming with semi-continuous variables and cardinality constraint. *China J. Oper. Res.* **1**(1), 55–77 (2013)
29. Wang, W. R., Carreira-Perpiñán, M. A.: Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application. [arXiv: org/abs/1309.1541](https://arxiv.org/abs/1309.1541) (2013)
30. Wen, B., Chen, X.J., Pong, T.K.: Linear convergence of proximal gradient algorithm with extrapolation for a class of nonconvex nonsmooth minimization problems. *SIAM J. Optim.* **27**(1), 124–145 (2017)
32. Xu, Z.B., Hai, Z., Yao, W., Chang, X.Y., Yong, L.:  $\ell_{1/2}$  regularization. *Sci. China Inf. Sci.* **53**(6), 1159–1169 (2010)
31. Xu, F.M., Lu, Z.S., Xu, Z.B.: An efficient optimization approach for a cardinality-constrained index tracking problem. *Optim. Methods Softw.* **31**(2), 258–271 (2016)
33. Zhang, C.H.: Nearly unbiased variable selection under minimax concave penalty. *Ann. Stat.* **38**(2), 894–942 (2010)
34. Zhao, Z.H., Xu, F.M., Du, D.L., Wang, M.H.: Robust portfolio rebalancing with cardinality and diversification constraints. *Quant. Financ.* **21**(20), 1707–1721 (2021)
35. Zhu, S.S., Li, D., Sun, X.L.: Portfolio selection with marginal risk control. *J. Comput. Financ.* **14**(1), 3–28 (2010)

---

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.