Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH Open Access

# Proximal linearized method for sparse equity portfolio optimization with minimum transaction cost

Hong Seng Sim[1], Wendy Shin Yie Ling[2], Wah June Leong[2] and Chuei Yee Chen[2*]

*Correspondence:
cychen@upm.edu.my
[2]Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, 43400, UPM Serdang, Selangor, Malaysia
Full list of author information is available at the end of the article

**Abstract**

In this paper, we propose a sparse equity portfolio optimization model that aims at minimizing transaction cost by avoiding small investments while promoting diversification to help mitigate the volatility in the portfolio. The former is achieved by including the $\ell_0$-norm regularization of the asset weights to promote sparsity. Subjected to a minimum expected return, the proposed model turns out to be an objective function consisting of discontinuous and nonconvex terms. The complexity of the model calls for proximal method, which allows us to handle the objective terms separately via the corresponding proximal operators. We develop an efficient algorithm to find the optimal portfolio and prove its global convergence. The efficiency of the algorithm is demonstrated using real stock data and the model is promising in portfolio selection in terms of generating higher expected return while maintaining good level of sparsity, and thus minimizing transaction cost.

**Mathematics Subject Classification:** 90C26; 91G10; 65K10

**Keywords:** Portfolio optimization; Sparse portfolio; Minimum transaction cost; Mean–variance model; Proximal method

## 1 Introduction

Introduced by Markowitz [22] in 1952, mean–variance optimization (MVO) has been widely used in the selection of optimal investment portfolios. The success of MVO is attributed to the simplicity of its quadratic objective function, which in turn can be optimized by quadratic programming (QP) methods that are widely available. However, MVO has flaws on its own and its implementation in portfolio optimization has been heavily criticized by academics and professionals [25]. One of its flaws, as pointed out by Michaud [23], is its sensitivity towards input parameters, thus maximizing the errors associated with these inputs. This was proven theoretically and computationally by Best and Grauer [2], where a slight change in the assets' expected return or correlations resulted in large changes in portfolio weights. This has led to a number of studies investigating different strategies on risk measures and return rates [17, 18, 20, 27]. Through the literature, it is evident that MVO remains to be one of the most successful frameworks due to the absence of models that are simple enough to be cast as a QP problem.

Over the past one decade or so, the success of robust optimization techniques has allowed researchers to consider nonquadratic objective function and regularization for portfolio optimization. Consequently, the work by Daubechies et al. [11] showed that the usual quadratic regularizing penalties can be replaced by weighted $\ell_p$-norm penalties with $p \in [1, 2]$. Two specific cases in portfolio optimization, namely least absolute shrinkage and selection operator (lasso) when $p = 1$ and ridge regression when $p = 2$, were considered by Brodie et al. [8] and DeMiguel et al. [13], respectively. While the ridge regression regularization minimizes the sample variance subject to the constraint which leads to diversification, lasso regularization encourages sparse portfolios which in turn leads to the minimization of transaction cost. Such regularizations have been studied notably by Chen et al. [9], De Mol [12], and Fastrich et al. [14].

In reality, financial institutions charge their customers transaction fees for trading on the stock market. The two most common ways to charge their customers are based on a fixed transaction fee and/or a proportion of the investment amount, whichever is higher. In general, a large number of transactions will result in higher transaction cost, likely to be caused by small investments that incur fixed transaction fees. Transaction cost, in this sense, will have an effect on the portfolio optimization and the frequency of time rebalancing the portfolio. On the other hand, diversification is the practice of spreading the investments around so that the exposure to any one type of asset is limited. This practice can help to mitigate the risk and volatility in the portfolio, potentially upsizing the number of investment components and thus increasing the number of transactions. Therefore, a more realistic model is needed to strike a balance between diversification and minimization of transaction costs for optimal portfolio selection.

Due to the complexity of the objective function and the regularization that are involved, many existing studies in the literature employ the alternating direction method of multipliers (ADMM), which was first introduced by Gabay and Mercier [16] in 1976. It was not until the recent decade that ADMM has received much attention in machine learning problems. The essence of ADMM is that it allows one to handle the objective terms separately when they can only be approximated using proximal operators. Its appealing features in large-scale convex optimization problems include ease of implementation and relatively good performance (see, for instance, Boyd et al. [7], Fazel et al. [15], and Perrin and Roncalli [25]). Some of the examples of ADMM-like algorithms in portfolio optimization can be found in Chen et al. [9], Dai and Wen [10], and Lai et al. [19], where they are used to solve $\ell_p$-regularizing problems when $p \in [1, 2]$. Though the $\ell_0$-norm is ideal for sparsity problems, the regularization results in a discontinuous and nonconvex problem, thus solving it computationally will turn out to be complicated.

In this paper, we propose a new algorithmic framework to maximize the sparsity within the entire portfolio while promoting diversification, i.e., to minimize the $\ell_0$- and $\ell_2$-norm of the asset weights, respectively, subject to a minimum expected return via MVO. We first transform the constrained problem into an unconstrained one, to find a nonsmooth and nonconvex objective term. The technique of ADMM allows us to handle these terms separately, but nevertheless yields convergence to its optimal solution. Numerical results using real data are also provided to illustrate the reliability of the proposed model and its efficiency in generating higher expected return while minimizing transaction cost when compared to the standard MVO.

This paper is organized as follows: In Sect. 2, we present a model for sparse equity portfolio optimization with minimum transaction cost and establish the proximal linearized method for $\ell_0$-norm minimization. Subsequently, in Sect. 3, we present an ADMM-like algorithm to find the optimal portfolio of the proposed model, together with its convergence analysis. To illustrate the reliability and efficiency of our method, we present the numerical results using real stock data in Sect. 4. Finally, the conclusion of the paper is presented in Sect. 5.

## 2 Proximal linearized method for $\ell_0$-norm minimization

We begin with a universe of $n$ assets under consideration, with mean return vector $\mu \in \mathbb{R}^n$ and the covariance matrix $V \in \mathbb{R}^{n\times n}$. Let $x \in \mathbb{R}^n$ be the vector of asset weights in the portfolio. Our objective is to maximize the portfolio return $\mu^T x$ and minimize the variance of portfolio return $x^T Vx$, while maintaining a certain level of diversification $\|x\|_2^2$ and minimizing transaction cost $\|x\|_0$. The variance of the portfolio return is the measure of risk inherent in investing in a portfolio, and we shall denote this as variance risk throughout this paper. The portfolio is said to be purely concentrated if there exists an $i$ such that $x_i = 1$ and equally weighted if $x_i = \frac{1}{n}$ for all $i$. We assume that the capital is fully invested, thus $e^T x = 1$ where $e \in \mathbb{R}^n$ is an all-one vector. The standard MVO [22] is as follows:

$$\min_{x\in\mathbb{R}^n} \quad \frac{1}{2}x^T Vx - \gamma \mu^T x \tag{1}$$

$$\text{subject to} \quad x \succeq 0, \tag{2}$$

$$e^T x = 1, \tag{3}$$

where $\gamma > 0$ is a parameter for leveraging the expected return and the inequality (2) is the no short selling restriction, with the notation $\succeq$ representing component-wise inequality between vectors. Here, diversification is of general importance to reduce portfolio risk without necessarily reducing portfolio return. While diversification does not mean that we add more money into our investment, it certainly does reduce our investment value as investment in each equity incurs transaction cost. Our proposed method takes into consideration having diversified investments, but at the same time avoiding small investments that might result in unnecessary transaction costs due to diversification. To do so, we consider the sparsity measure of the vector $x \in \mathbb{R}^n$ given by its $\ell_0$-norm:

$$\|x\|_0 := \text{number of nonzero components of } x_i.$$

A sparse equity portfolio optimization with minimum transaction cost (SEPO-$\ell_0$) is stated as follows:

$$\min_{x\in\mathbb{R}^n} \quad \frac{\beta_1}{2}x^T Vx - \mu^T x + \frac{\beta_2}{2}\|x\|_2^2 + \|x\|_0 \tag{4}$$

$$\text{subject to} \quad \mu^T x \geq r, \tag{5}$$

$$x \succeq 0, \tag{6}$$

$$e^T x = 1, \tag{7}$$

where $\beta_1 > 0$ is a parameter for leveraging the portfolio variance risk, $\beta_2 > 0$ is a parameter for leveraging portfolio diversification, and $r \geq 0$ is the minimum guaranteed return ratio with $r \leq \max\{\mu_i\}$. Though it is standard to introduce a parameter for leveraging the expected return $\mu^T x$ in the objective function, here we consider a more direct inequality constraint (5) where one can easily decide on the minimum expected return while maximizing it via the objective function. Note that minimizing $\ell_0$-norm in (4) promotes sparsity within the portfolio, since the values of $x_i$ are forced to be zero except for the large ones, thus reducing the transaction cost.

Our model (4)–(7) poses computational difficulties due to the nonconvexity and discontinuity of the $\ell_0$-norm, the minimum guaranteed return constraint (5), and no short selling constraint (6). Instead of dealing with the problem in its entirety, we employ the ADMM such that the smooth and nonsmooth terms can be handled separately. This calls for a brief introduction to proximal operators and Moreau envelope [26]:

**Definition 2.1** Let $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function and $\sigma > 0$ be a parameter. The proximal operator of $\psi$ is defined as

$$\operatorname{prox}_{\sigma\psi}(x) = \underset{y\in\mathbb{R}^n}{\arg\min}\left\{\psi(y) + \frac{1}{2\sigma}\|y - x\|_2^2\right\}. \tag{8}$$

Its Moreau envelope (or Moreau–Yosida regularization) is defined by

$$\operatorname{env}_{\sigma\psi}(x) = \inf_{y\in\mathbb{R}^n}\left\{\psi(y) + \frac{1}{2\sigma}\|y - x\|_2^2\right\}. \tag{9}$$

The parameter $\sigma$ can be interpreted as a trade-off between minimizing $\psi$ and being close to $x$. Moreau envelope, specifically, is a way to smooth a nonsmooth function, and it can be shown that the optimal value of $\operatorname{env}_{\sigma\psi}(x)$ is also the optimal value of $\operatorname{prox}_{\sigma\psi}(x)$.

Suppose now we are given a problem

$$\min \psi(x) + \phi(x),$$

where $\psi, \phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ are closed proper functions, of which both $\psi$ and $\phi$ can be nonsmooth. Under the ADMM algorithm, each iteration $k$ takes on an alternating nature with the proximal operators of $\psi$ and $\phi$ being evaluated separately:

$$\begin{aligned}
&x^{k+1} \in \operatorname{prox}_{\sigma\psi}\left(z^k - u^k\right),\\
&y^{k+1} \in \operatorname{prox}_{\sigma\phi}\left(x^{k+1} + u^k\right),\\
&u^{k+1} := u^k + x^{k+1} - z^{k+1}.
\end{aligned}$$

Viewing the above as a fixed point iteration, the ADMM scheme results in $x = z$ such that

$$\begin{aligned}
&x = \operatorname{prox}_{\sigma\psi}(x - u),\\
&y = \operatorname{prox}_{\sigma\phi}(x + u).
\end{aligned}$$

Turning our attention back to our problem (4)–(7), we first denote the set $R$ associated with the inequality constraint in (5) by

$$R = \{x \in \mathbb{R}^n : \mu^T x \geq r\}, \tag{10}$$

and the indicator function of $R$ by

$$I_R(x) = \begin{cases} 0, & x \in R, \\ \infty, & x \notin R. \end{cases} \tag{11}$$

The feasible set for constraint (6) is given by

$$S = \{x \in \mathbb{R}^n : x \geq 0\}, \tag{12}$$

with its indicator function

$$I_S(x) = \begin{cases} 0, & x \in S, \\ \infty, & x \notin S. \end{cases} \tag{13}$$

We now define the augmented Lagrangian corresponding to problem (4)–(7) as

$$\begin{aligned} \mathcal{L}(x, \lambda, \rho) = {} & \frac{\beta_1}{2} x^T V x - \mu^T x + \frac{\beta_2}{2} \|x\|_2^2 + \|x\|_0 + I_R(x) + I_S(x) \\ & + \lambda\left(e^T x - 1\right) + \frac{\rho}{2}\left(e^T x - 1\right)^2, \end{aligned} \tag{14}$$

where $\lambda$ is the usual Lagrange multiplier and $\rho > 0$ is the penalty parameter for the equality constraint $e^T x = 1$. To obtain convergence, we may set $\rho$ to be a constant that is larger than the threshold of the problem [1]. Our problem (14), with a threshold of $\rho = 4$, can be rewritten as $\mathcal{L}(x, \lambda)$ where $x$ and $\lambda$ are updated via

$$\begin{aligned} x^{k+1} &= \arg\min_x \mathcal{L}\left(x, \lambda^k\right), \\ \lambda^{k+1} &= \lambda^k + \rho\left(e^T x^{k+1} - 1\right). \end{aligned}$$

Problem (14) can now be viewed as the following minimization problem:

$$\min_{x,\lambda} \quad P(x, \lambda) + Q(x), \tag{15}$$

where $P(x, \lambda)$ consists of the smooth terms given by

$$P(x, \lambda) = \frac{\beta_1}{2} x^T V x - \mu^T x + \frac{\beta_2}{2} \|x\|_2^2 + \lambda\left(e^T x - 1\right) + \frac{\rho}{2}\left(e^T x - 1\right)^2, \tag{16}$$

and $Q(x)$ comprises the nonsmooth terms, namely

$$Q(x) = \|x\|_0 + I_R(x) + I_S(x). \tag{17}$$

For the purpose of our discussion on the proximal method, we let $\lambda$ be a fixed value, say $\hat{\lambda}$, which we use in the following minimization problem:

$$\min_x \quad P(x,\hat{\lambda}) + Q(x). \tag{18}$$

Our proximal method for minimizing the objective function in (18) can be viewed as the proximal regularization of $P$ linearized at a given point $x^k$:

$$x^{k+1} \in \arg\min_x \left\{ Q(x) + \left(x - x^k\right)^T \nabla P\left(x^k\right) + \frac{1}{2t}\left\|x - x^k\right\|^2 \right\}, \tag{19}$$

where $t > 0$ and $\nabla$ denotes the derivative operator. Invoking simple algebra and ignoring the constant terms, (19) can be written as

$$x^{k+1} \in \arg\min_x \left\{ Q(x) + \frac{1}{2t^k}\left\|x - \left(x^k - t^k \nabla P\left(x^k\right)\right)\right\|^2 \right\}. \tag{20}$$

Using Definition 2.1, the iterative scheme consists of a proximal step at a resulting gradient point which gives us the proximal gradient method:

$$x^{k+1} \in \operatorname{prox}_{\alpha^k Q}\left(x^k - \alpha^k \nabla P\left(x^k\right)\right), \tag{21}$$

where $\alpha^k > 0$ is a suitable step size. Note that if $\nabla P$ is Lipschitz continuous with constant $L_c$, then the proximal gradient method is known to converge at a rate of $\mathcal{O}(1/k)$ with fixed step size $\alpha \in (0, 1/L_c]$ (see Parikh and Boyd [24]). In the case when $L_c$ is not known, the step sizes can be chosen via line search methods. In the context of line search methods, the largest possible step size $\alpha = 1$ is more desirable. Therefore, proximal gradient methods usually have a fixed step size $\alpha = \min\{1, 1/L_c\}$. In our case, the Lipschitz continuity of $\nabla P$ gives

$$\begin{aligned} \left\|\nabla P(x) - \nabla P(y)\right\|_2 &= \left\|\beta_1 V(x-y) + \beta_2(x-y) + \rho(x-y)\right\|_2 \\ &\le \left\|\beta_1 V + \beta_2 I + \rho e e^T\right\|_F \|x-y\|_2 \end{aligned} \tag{22}$$

for all $x, y \in \mathbb{R}^n$ where $I$ denotes the identity matrix and $\|\cdot\|_F$ denotes the Frobenius norm. Since the Lipschitz constant of (22) is not easily accessible, we can estimate it in the following way:

$$\begin{aligned} L_c &\le \beta_1 \|V\|_F + \beta_2 \|I\|_F + \rho \left\|ee^T\right\|_F \\ &= \beta_1 \left(\operatorname{tr}\left(VV^T\right)\right)^{1/2} + \beta_2\sqrt{n} + \rho n =: \tilde{L}_c, \end{aligned} \tag{23}$$

where tr denotes the matrix trace. Since $\tilde{L}_c > 1$, it is clear that $\min\{1, 1/\tilde{L}_c\}$ will always return the value $1/\tilde{L}_c$. We shall henceforth fix our step size $\alpha = 1/\tilde{L}_c$. Our choice of step size follows from the well-known descent property below:

**Lemma 2.1** (Descent property [6]) *Let $\psi : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function with gradient $\nabla\psi$ assumed to be $L_c$-Lipschitz continuous. Then, for any $\tilde{L}_c \ge L_c$,*

$$\psi(x) \le \psi(y) + (x-y)^T \nabla\psi(y) + \frac{\tilde{L}_c}{2}\|x-y\|^2, \quad \forall x, y \in \mathbb{R}^n. \tag{24}$$

Using the proximal operator defined in Definition 2.1, the minimization of (19) is equivalent to the following step:

$$x^{k+1} \in \operatorname{prox}_{\alpha Q}\left(x^k - \alpha \nabla P\left(x^k\right)\right), \tag{25}$$

where $\alpha = \frac{1}{\tilde{L}_c}$. The choice of $\tilde{L}_c$ also guarantees a sufficient decrease of our objective function under the proximal methods:

**Lemma 2.2** (Sufficient decrease property [6]) *Let $\psi : \mathbb{R}^n \to \mathbb{R}$ be a $C^1$ function with its gradient $\nabla\psi$ being Lipschitz continuous with modulus $L_c$. Let $\phi : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper and lower semicontinuous function with $\inf_{\mathbb{R}^n} \phi > -\infty$. Suppose $\tilde{L}_c$ is chosen such as $\tilde{L}_c > L_c$. Then, for any $x \in \operatorname{dom} \phi$ and any $\hat{x} \in \mathbb{R}^n$ defined by*

$$\hat{x} \in \operatorname{prox}_{\alpha\phi}\left(x - \alpha\nabla\psi(x)\right), \quad \alpha = \frac{1}{\tilde{L}_c}, \tag{26}$$

*we have*

$$\psi(\hat{x}) + \phi(\hat{x}) \le \psi(x) + \phi(x) - \frac{1}{2}(\tilde{L}_c - L_c)\|\hat{x} - x\|^2. \tag{27}$$

Note that dom $\phi$ in Lemma 2.2 defines the set of points for which proper and lower semicontinuous function $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ takes on a finite value:

$$\operatorname{dom} \phi = \left\{x \in \mathbb{R}^n : \phi(x) < +\infty\right\}.$$

In view of Lemma 2.2, we turn to our nonsmooth term $Q(x)$, for which we continue to invoke the ADMM algorithm on (25):

$$\begin{cases} z^{k+1} \in \operatorname{prox}_{\sigma\|\cdot\|_0}(z^k - \alpha\nabla P(z^k)), \\ y^{k+1} = \operatorname{prox}_{I_R}(z^{k+1}), \\ x^{k+1} = \operatorname{prox}_{I_S}(y^{k+1}). \end{cases} \tag{28}$$

The convergence of our iteration is guaranteed since the ADMM method ensures the convergence of the objective function to its optimal value [7].

In (28), the proximal operator of the $\ell_0$-norm can be expressed in its component-wise form as

$$\operatorname{prox}_{\sigma\|\cdot\|_0}(x) = \begin{cases} \{0\}, & \text{if } x_i < \sqrt{2\sigma}, \\ \{0, x_i\}, & \text{if } x_i = \sqrt{2\sigma}, \\ \{x_i\}, & \text{if } x_i > \sqrt{2\sigma}. \end{cases} \tag{29}$$

Note that $\operatorname{prox}_{\sigma\|\cdot\|_0}(x)$ is known as a hard thresholding operator since it forces the vector's components $x_i$ except the large ones to be zero [26]. In other words, a larger $\sigma$ results in higher sparsity and less penalization for moving away from $x$. Doing so ensures that our portfolio selections avoid small investments.

Meanwhile, the proximal operator of the indicator function $I_R$ is reduced to Euclidean projection onto $R$:

$$\operatorname{prox}_{I_R}(x) = \begin{cases} x, & \text{if } \mu^T x \geq r, \\ \frac{r}{\mu^T x} x, & \text{if } \mu^T x < r. \end{cases} \tag{30}$$

The proximal operator of the indicator function $I_S$ is the projection of the vector $x$ onto $\mathbb{R}_+$:

$$\operatorname{prox}_{I_S}(x) = x_+, \tag{31}$$

where $x_+$ is taken component-wise with each negative $x_i$ being replaced by a zero. In view of (28), we have

$$x_i^{k+1} = \max\left\{0, y_i^{k+1}\right\}, \quad i = 1, 2, \ldots, n.$$

## 3 Alternating proximal algorithm and its convergence

In this section, we present an ADMM algorithm to find the optimal portfolio of the proposed SEPO-$\ell_0$ model (4)–(7) and establish its global convergence.

**SEPO-$\ell_0$ Algorithm**

**Step 0.** Given $\beta_1, \beta_2, \sigma, r, V, \mu, \rho, \alpha$, initial point $(x^0, \lambda^0)$, and convergence tolerance $\varepsilon$. Set $k := 0$.

**Step 1.** Compute $z^{k+1} \in \operatorname{prox}_{\sigma\|\cdot\|_0}(z^k - \alpha \nabla P(z^k, \lambda^k))$.

**Step 2.** Compute $y^{k+1} = \operatorname{prox}_{I_R}(z^{k+1})$.

**Step 3.** Compute $x^{k+1} = \operatorname{prox}_{I_S}(y^{k+1})$.

**Step 4.** Compute $\lambda^{k+1} = \lambda^k + \rho(e^T x^{k+1} - 1)$.

**Step 5.** If $\|\nabla P(x^{k+1}, \lambda^{k+1})\| < \varepsilon$ or $k > 10000$, stop. Else, set $k := k + 1$ and go to Step 1.

We have seen in Sect. 2 how the proposed proximal method guarantees the descent of the solution. To proceed with the convergence of SEPO-$\ell_0$ algorithm, we begin with Assumption A for any objective function $\mathcal{L}\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ where $\mathcal{L} = \psi + \phi$:

**Assumption A**

(i) $\psi : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function where its gradient $\nabla\psi$ is Lipschitz continuous with modulus $L_c$.

(ii) $\phi\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper and lower semicontinuous function.

(iii) $\inf_{\mathbb{R}^n} \psi > -\infty$ and $\inf_{\mathbb{R}^n} \phi > -\infty$.

SEPO-$\ell_0$ algorithm also results in nice convergence properties of (14):

**Lemma 3.1** (Convergence properties [6]) *Suppose that $\mathcal{L}\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is an objective function such that Assumption A holds. Let $\{x^k\}_{k\in\mathbb{N}}$ be a sequence generated by SEPO-$\ell_0$ algorithm. Then, the sequence $\{\mathcal{L}(x^k, \lambda^k) : k \in \mathbb{N}\}$ is nonincreasing and, in particular,*

$$\mathcal{L}\left(x^k\right) - \mathcal{L}\left(x^{k+1}\right) \geq \frac{1}{2}(\tilde{L}_c - L_c)\left\|x^{k+1} - x^k\right\|^2. \tag{32}$$

*Moreover,*

$$\sum_{k=1}^{\infty} \left\| x^{k+1} - x^k \right\|^2 < \infty, \tag{33}$$

*and hence*

$$\lim_{k\to\infty} \left\| x^{k+1} - x^k \right\| = 0. \tag{34}$$

*Proof* Without loss of generality, we let $\lambda$ be a fixed constant and work with $\mathcal{L}(x) = P(x) + Q(x)$ in place of $\mathcal{L}(x, \lambda)$, where $P(x)$ and $Q(x)$ are given by (16) and (17), respectively. Note that $P(x)$ is differentiable and its gradient is Lipschitz continuous with modulus $L_c$. Invoking SEPO-$\ell_0$ algorithm and by Lemma 2.2, we have

$$P\left(x^{k+1}\right) + Q\left(x^{k+1}\right) \le P\left(x^k\right) + Q\left(x^k\right) - \frac{1}{2}(\tilde{L}_c - L_c)\left\| x^{k+1} - x^k \right\|^2, \tag{35}$$

where $\tilde{L}_c$ is given by (23). Writing $\mathcal{L}(x^k) = P(x^k) + Q(x^k)$ in (35) and rearranging it leads to (32), which asserts that the sequence $\{\mathcal{L}(x^k, \lambda^k) : k \in \mathbb{N}\}$ is nonincreasing.

Note that $P$ and $Q$ are bounded below (see Assumption A), and hence $\mathcal{L}$ converges to some $\underline{\mathcal{L}}$. Let $N \in \mathbb{N}_+$. We sum up (32) from $k = 0$ to $k = N - 1$ to get

$$\begin{aligned}
\sum_{k=0}^{N-1} \left\| x^{k+1} - x^k \right\|^2 &\le \frac{2}{\tilde{L}_c - L_c} \sum_{k=0}^{N-1} \left( \mathcal{L}\left(x^k\right) - \mathcal{L}\left(x^{k+1}\right) \right) \\
&= \frac{2}{\tilde{L}_c - L_c} \left( \mathcal{L}\left(x^0\right) - \mathcal{L}\left(x^N\right) \right) \\
&\le \frac{2}{\tilde{L}_c - L_c} \left( \mathcal{L}\left(x^0\right) - \underline{\mathcal{L}} \right).
\end{aligned}$$

It follows that (33) and (34) hold when we take the limit as $N \to \infty$. □

Before we present the result that sums up the properties of the sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by SEPO-$\ell_0$ algorithm starting from the initial point $x^0$, we first give some basic notations. We denote by crit $\mathcal{L}$ the set of critical points of $\mathcal{L}$ and $\omega(x^0)$ the set of all limit points, where

$$\omega\left(x^0\right) = \left\{ \begin{array}{c} \overline{x} \in \mathbb{R}^n : \exists \text{ an increasing sequence of integers } \{k_j\}_{j\in\mathbb{N}} \\ \text{such that } x^{k_j} \to \overline{x} \text{ as } j \to \infty \end{array} \right\}.$$

Given any set $\Omega \subset \mathbb{R}^n$ and any point $x \in \mathbb{R}^n$, the distance from $x$ to $\Omega$ is denoted and defined by

$$\operatorname{dist}(x, \Omega) := \inf\left\{ \|y - x\| : y \in \Omega \right\}.$$

When $\Omega = \emptyset$, we invoke the usual convention that $\inf \emptyset = \infty$ and hence $\operatorname{dist}(x, \Omega) = \infty$ for all $x$.

**Lemma 3.2** (Properties of limit points [6]) *Suppose that* $\mathcal{L}\colon \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ *is an objective function such that Assumption* A *holds. Let* $\{x^k\}_{k\in\mathbb{N}}$ *be a bounded sequence generated by SEPO-*$\ell_0$ *algorithm. Then, the following hold*:

(a) $\omega(x^0)$ *is a nonempty, compact, and connected set.*
(b) $\omega(x^0) \subset \operatorname{crit} \mathcal{L}$.
(c) $\lim_{k\to\infty} \operatorname{dist}(x^k, \omega(x^0)) = 0$.
(d) *The objective function* $\mathcal{L}$ *is finite and constant on* $\omega(x^0)$.

*Proof* See Bolte et al. [6]. □

What remains is its global convergence, which we shall establish by means of the Kurdyka–Łojasiewicz (KL) property [6] as an extension of Łojasiewicz gradient inequality [21] for nonsmooth functions. We first show that the objective function (14) is semialgebraic and therefore is a KL function. This, in turn, is crucial in giving us the convergence of the sequences generated via SEPO-$\ell_0$ algorithm. We begin by recalling notations and definitions concerning subdifferential (see, for instance, [6, 26]) and KL property.

**Definition 3.1** Let $\phi\colon \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ be a proper and lower semicontinuous function. The (limiting) subdifferential of $\phi$ at $x \in \operatorname{dom}\phi$ is denoted and defined by

$$\partial\phi(x) = \begin{Bmatrix} u \in \mathbb{R}^n\colon \exists x^k \to x, \phi\big(x^k\big) \to \phi(x), u^k \to u, \\ \liminf_{y\to x^k} \frac{\phi(y)-\phi(x^k)-\langle u^k, y-x^k\rangle}{\|y-x^k\|} \geq 0 \end{Bmatrix}. \tag{36}$$

The point $x$ is called a (limiting) critical point of $\phi$ if $0 \in \partial\phi(x)$.

It follows that $0 \in \partial\phi(x)$ if $x \in \mathbb{R}^n$ is a local minimizer of $\phi$. For a continuously differentiable $\phi$, $\partial\phi(x) = \{\nabla\phi\}$, and hence we have the usual gradient mapping $\nabla\phi$ from $x \in \operatorname{dom}\phi$ to $\nabla\phi(x)$. If $\psi$ is convex, the subdifferential (36) turns out to be the classical Fréchet subdifferential [26].

Let $\eta \in (0,\infty]$ and denote by $\Phi_\eta$ the class of all concave and continuous functions $\varphi\colon [0,\eta) \to \mathbb{R}_+$ that are continuously differentiable on $(0,\eta)$ and continuous at 0 with $\varphi(0) = 0$ and $\varphi'(s) > 0$ for all $s \in (0,\eta)$.

**Definition 3.2** (Kurdyka–Łojasiewicz (KL) property) Let $\phi\colon \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ be a proper and lower semicontinuous function. The function $\phi$ is said to have the Kurdyka–Łojasiewicz (KL) property at $\bar{u} \in \operatorname{dom}\partial\phi := \{u \in \mathbb{R}^n\colon \partial\phi(u) \neq \emptyset\}$ if there exist $\eta \in (0,+\infty]$, a neighborhood $U$ of $\bar{u}$ and a function $\varphi \in \Phi_\eta$ such that for all $u \in U \cap [\phi(\bar{u}) < \phi(u) < \phi(\bar{u}) + \eta]$, the following inequality holds:

$$\varphi'\big(\phi(u) - \phi(\bar{u})\big)\operatorname{dist}\big(0, \partial\phi(u)\big) \geq 1. \tag{37}$$

Moreover, $\phi$ is called a KL function if it satisfies the KL property at each point of $\operatorname{dom}\phi$.

The definition above uses the sublevel sets: Given $a, b \in \mathbb{R}$, the sublevel sets of a function $\phi$ are denoted and defined by

$$[a \leq \phi \leq b] := \big\{x \in \mathbb{R}^n\colon a \leq \phi(x) \leq b\big\}.$$

A similar definition holds for $[a < \phi < b]$. The level sets of $\phi$ are denoted and defined by

$$[\phi = a] := \{x \in \mathbb{R}^n : \phi(x) = a\}.$$

Closely related to the KL function is the semi-algebraic function, which is crucial in the proof of the convergence property of our proposed method.

**Definition 3.3** (Semi-algebraic sets and functions)

(i) A subset $\Omega \subset \mathbb{R}^n$ is called semi-algebraic if there exists a finite number of real polynomial functions $p_{ij}$ and $q_{ij}$ such that

$$\Omega = \bigcup_{j=1}^{p} \bigcap_{i=1}^{q} \{u \in \mathbb{R}^n : p_{ij}(u) = 0 \text{ and } q_{ij}(u) < 0\}. \tag{38}$$

(ii) A function $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called semi-algebraic if its graph

$$\{(u, t) \in \mathbb{R}^{n+1} : \phi(u) = t\} \tag{39}$$

is a semi-algebraic subset of $\mathbb{R}^{n+1}$.

It follows that semi-algebraic functions are indeed KL functions, and the result below is a nonsmooth version of the Łojasiewicz gradient inequality.

**Theorem 3.1** ([4, 5]) *Let $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. If $\phi$ is semi-algebraic, then it is a KL function.*

Theorem 3.1 allows us to avoid the technicality in proving the KL property. Instead, one can make use of the broad range of semi-algebraic functions and sets [3, 6]. Some of the examples of semi-algebraic functions include real polynomial functions, and indicator functions of semi-algebraic sets. Apart from that, finite sums and products of semi-algebraic functions, as well as scalar products, are all semi-algebraic.

We are now ready to give the global convergence result of the proposed model (4)–(7).

**Theorem 3.2** (Global convergence) *Suppose the objective function $\mathcal{L} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a KL function such that Assumption A holds. Then the sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by SEPO-$\ell_0$ algorithm converges to a critical point $x^*$.*

*Proof* See Bolte et al. [6]. □

By virtue of Theorem 3.2, we now show that each term in (14) is semi-algebraic since the finite sum of semi-algebraic functions is also semi-algebraic. It is obvious that the function in (14) is a sum of a smooth function $P(x)$, $\ell_0$-norm, and indicator functions. The function $P(x)$ given by (16) is a linear combination of linear and quadratic functions, and hence $P(x)$ is a real polynomial function, which in turn is semi-algebraic.

As a specific example given by Bolte et al. [6], the $\ell_0$-norm is the sparsity measure of the vector $x \in \mathbb{R}^n$, which is indeed semi-algebraic. In particular, the graph of $\|\cdot\|_0$ is given by

a finite union of product sets:

$$\operatorname{graph}\|\cdot\|_0 = \bigcup_{I\subset\{1,\dots,n\}} \left(\prod_{i=1}^{n} J_i^I\right) \times \{n-|I|\}, \tag{40}$$

where for any given $I \subset \{1,\dots,n\}$, $|I|$ denotes the cardinality of $I$ and

$$J_i^I = \begin{cases} \{0\}, & \text{if } i \in I, \\ \mathbb{R}\setminus\{0\}, & \text{otherwise.} \end{cases}$$

It is obvious that (40) is a piecewise linear set, hence the claim. Lastly, the indicator functions $I_R(x)$ and $I_S(x)$ defined by (11) and (13), respectively, are also semi-algebraic, since the feasible sets (10) and (12) are convex.

## 4 Numerical experiments and results

In this section, we study the efficiency of the proposed portfolio optimization model, SEPO-$\ell_0$, in maximizing portfolio return and minimizing transaction costs. We test our algorithm on real data of stock prices and returns of 100 companies across 10 different sectors in China, collected from January 2019 to June 2019. These data are in turn used to generate the covariance matrix, which gives us the portfolio variance as in our objective function (4). We start with equally-weighted portfolio, i.e., $x_i^0 = \frac{1}{n}$ for all $i$. We set $\varepsilon = 10^{-7}$ and stop the algorithm when $\|\nabla P(x^{k+1}, \lambda^{k+1})\| < \varepsilon$ or $k > 10000$. All computational results are obtained by running Matlab R2021a on Windows 10 (Intel Core i7 1065G7 16 GB CPU @ 1.30 $\sim$ 1.50 GHz).

For testing purposes, we set our penalty parameter $\rho = 5$ and tuning parameter $\beta_2 = 1$. The latter means that we set our weight on the portfolio diversification as constant. Meanwhile, the value of $\beta_1$ is chosen to be relatively smaller than $\beta_2$. For illustration, we present our results for minimum guaranteed return ratio $r = 0.1$ and $r = 0.2$.

In Table 1, we present the computational results of the expected return, variance risk, and sparsity ratio under the proposed SEPO-$\ell_0$ model and standard MVO model for different values of $\beta_1$, when we set the minimum guaranteed ratio to be 0.1 and 0.2, respectively. Note that though we leveraged on the variance risk when $\beta_1 = 1$, the portfolio selection under SEPO-$\ell_0$ manages to generate expected return of 0.3455 and 0.4014 when $r = 0.1$ and $r = 0.2$, respectively. Meanwhile, the standard MVO is only able to generate expected return of 0.1560 when we set $r = 0.1$. The variance risks, however, are higher under SEPO-$\ell_0$ due to the sparsity, as compared to maximum diversification of the standard MVO. From the table, we can see that our model offers good sparsity ratio between 0.30 and 0.61 when $r = 0.1$ and between 0.52 and 0.72 when $r = 0.2$. This simply means that out of 100 stocks considered under minimum expected return ratio $r = 0.1$, one will only need to invest in the selected 39–70 stocks where the algorithm returns nonzero $x_i$'s. Despite the sparse portfolio selection and increased risk, we can see that the proposed model is more promising in terms of a higher expected return.

We also compare the expected return and variance risk for the SEPO-$\ell_0$ and standard MVO for $r = 0.1$ by using a scatterplot seen in Fig. 1. The downward trend of the portfolio expected return and variance risk mimic the standard MVO as $\beta_1$ approaches 1. Note that

**Table 1** The values of portfolio expected return, variance risk, and sparsity ratio for different $\beta_1$ with minimum guaranteed return ratio $r = 0.1$ and $r = 0.2$ under SEPO-$\ell_0$ and standard MVO

| $\beta_1$ | $r = 0.1$ | | | | | | $r = 0.2$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | SEPO-$\ell_0$ | | | Standard MVO | | | SEPO-$\ell_0$ | | |
| | E.R. | V.R. | Spar | E.R. | V.R. | Spar | E.R. | V.R. | Spar |
| 0.1 | 0.6355 | 3.2835 | 0.58 | 0.6889 | 2.3603 | 0 | 0.7441 | 4.3108 | 0.72 |
| 0.2 | 0.6279 | 2.9577 | 0.61 | 0.5735 | 1.5138 | 0 | 0.6732 | 3.2822 | 0.66 |
| 0.3 | 0.5050 | 2.1304 | 0.47 | 0.4555 | 1.0320 | 0 | 0.5829 | 2.4655 | 0.58 |
| 0.4 | 0.5180 | 2.0288 | 0.53 | 0.3760 | 0.8114 | 0 | 0.5796 | 2.2333 | 0.64 |
| 0.5 | 0.4865 | 1.7976 | 0.51 | 0.3003 | 0.6689 | 0 | 0.5374 | 1.9056 | 0.64 |
| 0.6 | 0.4237 | 1.5684 | 0.39 | 0.2646 | 0.5785 | 0 | 0.4675 | 1.6193 | 0.52 |
| 0.7 | 0.3677 | 1.4070 | 0.30 | 0.2223 | 0.5324 | 0 | 0.4655 | 1.4800 | 0.59 |
| 0.8 | 0.3581 | 1.3248 | 0.31 | 0.2057 | 0.4930 | 0 | 0.4521 | 1.3289 | 0.63 |
| 0.9 | 0.3787 | 1.2635 | 0.44 | 0.1750 | 0.4704 | 0 | 0.4182 | 1.2149 | 0.56 |
| 1 | 0.3455 | 1.1802 | 0.40 | 0.1560 | 0.4501 | 0 | 0.4014 | 1.1204 | 0.56 |

E.R. = Expected return, V.R. = Variance risk, Spar = Sparsity ratio.

a higher value of $\beta_1$ reflects our leverage on minimizing the variance risk over maximizing the expected return. At the same time, a higher expected return results means a higher risk as shown in Table 1. In general, the standard MVO model gives a lower measure for risk due to maximum diversification, as we can see from Table 1 and Fig. 1. The proposed SEPO-$\ell_0$, on the other hand, can lead to a higher expected return and a lower total transaction cost due to a sparse portfolio. This shows that SEPO-$\ell_0$ model is able to provide a good combination of portfolio selection under sparsity.

To illustrate the reliability of our model, we present the output of the proposed model for $r = 1$ using a scatterplot of the variables, as shown in Fig. 2, with $\beta_1$ as independent variable on the $x$-axis, the expected return and sparsity ratio on the left $y$-axis, while the risk scale is on the right of $y$-axis. We can observe a similar trend for the three lines, which clearly reflects the consistency of our model in obtaining an optimal portfolio selection.

The relationship between the independent variable $\beta_1$ and the response variables is further examined using the deterministic simple linear regression model as follows:

$$y_i = a_i + b_i\beta_1, \quad i = 1, 2, 3,$$

where $y_i$ are the response variables, $a_i$ the $y$-intercept, and $b_i$ the coefficients of $\beta_1 \in (0, 1]$. This model assigns weights to the independent variable $\beta_1$ to quantify its impact on the response variables. Here the response variables are expected return $y_1$, variance risk $y_2$, and sparsity ratio $y_3$. The relationship is presented in Table 2. As we can see from the table, the estimates of $b_i$ for response variables $y_i$ are all negative, which means their values decrease with the increase of $\beta_1$. Since the $p$-values of all response variables are approximately zero, it is clear that these three variables are significant. In particular, $\beta_1$ has a significant negative relationship with the expected return, risk, and sparsity ratio.

The significance of $\beta_1$ on these three dependent variables is supported by the values of R-squared of univariate regression, standing at 0.9076, 0.8748, and 0.5859 for the expected return, variance risk, and sparsity ratio, respectively. Since R-squared is the percentage of total variation contributed by a predictor variable, the high R-squared values, which are greater than 0.80 for the expected return and risk, mean that $\beta_1$ explains a high percentage of the variance in these two response variables. It is slightly lower for the sparsity ratio, however, any R-squared value greater than 0.50 can be considered as moderately high.

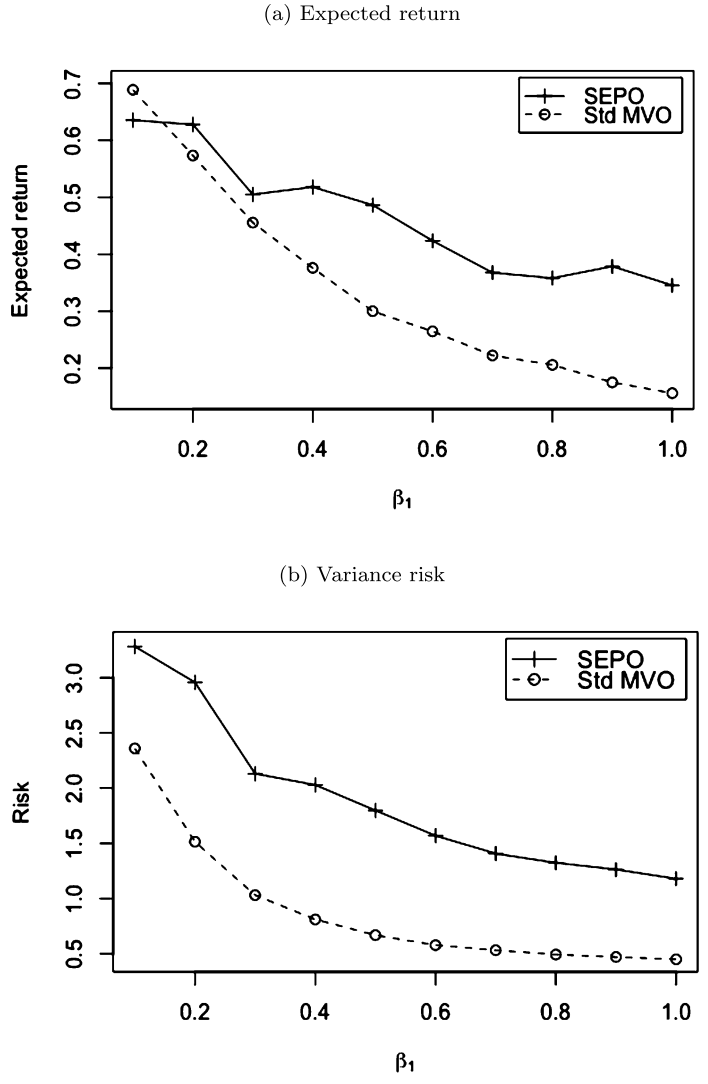


**Figure 1** SEPO-$\ell_0$ vs standard MVO with minimum guaranteed return ratio $r = 0.1$

**Table 2** Relationship between the independent variable $\beta_1 \in (0, 1]$ and the response variables using a simple linear regression model for SEPO-$\ell_0$ with $r = 0.1$

| Response variable | Estimate for intercept | Estimate for coefficient | Standard error for coefficient | *p*-value for coefficient | R-squared |
|---|---|---|---|---|---|
| Expected return, $y_1$ | 0.6514 | −0.3396 | 0.0383 | 2.0737e−0.5 | 0.9076 |
| Variance risk, $y_2$ | 3.1246 | −2.2371 | 0.2992 | 7.0894e−0.5 | 0.8748 |
| Sparsity ratio, $y_3$ | 0.6013 | −0.2679 | 0.0796 | 9.8657e−0.5 | 0.5859 |

## 5 Conclusion

The classical Markowitz portfolio scheme or mean–variance optimization (MVO) is one of the most successful frameworks due to the simplicity in implementation; in particular, it can be solved by quadratic programming which is widely available. However, it is very sensitive to input parameter and obtaining acceptable solutions requires the right weight constraints. Over the past decade, there has been renewed attention in considering non-

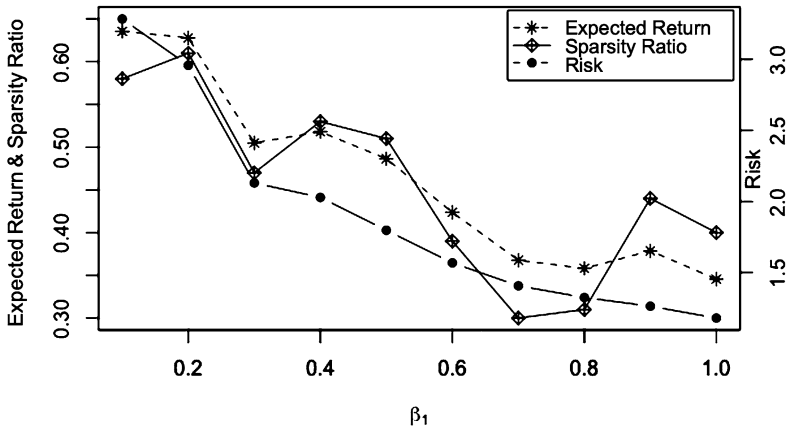


**Figure 2** Portfolio expected return, variance risk, and sparsity ratio subjected to minimum guaranteed return ratio $r = 0.1$ under SEPO-$\ell_0$ for different values of $\beta_1$

quadratic portfolio selection models due to the advancement in optimization algorithms for a more general class of functions. Here we proposed a new algorithmic framework that allows portfolio managers to strike a balance between diversifying investments and minimizing transaction costs, the latter of which is achieved by means of minimizing the $\ell_0$-norm, while being subjected to a minimum guaranteed return. This simply means that the model maximizes sparsity within the portfolio, since the weights $x_i$ are forced to be zero except for large ones. In practice, the regularization of $\ell_0$ results in a discontinuous and nonconvex problem. The inequality constraint, as a result of the minimum guaranteed return, also poses a similar problem.

In this study, we employed the proximal methods such that a function can be "smoothed" by means of linearizing part of the objective function at some given point and regularizing by a quadratic proximal term that acts as a measure for the "local error" in the approximation. Writing our problem in the form of augmented Lagrangian, the unconstrained problem can be divided into two parts, namely the smooth and nonsmooth terms. These are then handled separately through their proximal methods via the ADMM method. The global convergence of the proposed SEPO-$\ell_0$ algorithm for sparse equity portfolio has been established. The efficiency of our model in maximizing portfolio expected return while striking a balance between minimizing transaction cost and diversification has been analyzed using actual data of 100 companies. Empirically, the implementation of our model leads to a higher expected return and lower transaction costs. This shows that, despite its higher risk as compared to the standard MVO, the SEPO-$\ell_0$ model is promising in generating a good combination for an optimal investment portfolio.

**Acknowledgements**
The authors would like to thank the editor and the reviewers for their helpful comments and suggestions which have led to the improvement of the earlier version of this paper.

**Funding**
This project was supported by the Malaysian Ministry of Higher Education via Fundamental Research Grant Scheme (FRGS/1/2022/STG06/UTAR/02/2).

**Availability of data and materials**
The datasets analyzed during the current study are available from the corresponding author upon reasonable request.

## Declarations

**Competing interests**
The authors declare no competing interests.

**Author contributions**
All authors contributed to the study conception and design, material preparation, data collection and analysis. The first draft of the manuscript was written by C. Y. Chen and all authors commented on previous versions of the manuscript. All authors have read and approved the final manuscript.

**Author details**
[1]Centre for Mathematical Sciences, Universiti Tunku Abdul Rahman, 43000, Kajang, Selangor, Malaysia. [2]Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, 43400, UPM Serdang, Selangor, Malaysia.

Received: 23 March 2023 Accepted: 23 October 2023 Published online: 21 November 2023

**References**

1. Bertsekas, D.P.: Nonlinear Programming, 3rd edn. Athena Scientific (2016)
2. Best, M.J., Grauer, R.R.: On the sensitivity of mean–variance-efficient portfolios to changes in asset means: some analytical and computational results. Rev. Financ. Stud. **4**(2), 315–342 (1991)
3. Bochnak, J., Coste, M., Roy, M.F.: Real Algebraic Geometry, vol. 36. Springer, Berlin (2013)
4. Bolte, J., Daniilidis, A., Lewis, A.: The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J. Optim. **17**(4), 1205–1223 (2007)
5. Bolte, J., Daniilidis, A., Lewis, A., Shiota, M.: Clarke subgradients of stratifiable functions. SIAM J. Optim. **18**(2), 556–572 (2007)
6. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math. Program. **146**(1), 459–494 (2014)
7. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**(1), 1–122 (2011)
8. Brodie, J., Daubechies, I., De Mol, C., Giannone, D., Loris, I.: Sparse and stable Markowitz portfolios. Proc. Natl. Acad. Sci. **106**(30), 12,267–12,272 (2009)
9. Chen, J., Dai, G., Zhang, N.: An application of sparse-group lasso regularization to equity portfolio optimization and sector selection. Ann. Oper. Res. **284**(1), 243–262 (2020)
10. Dai, Z., Wen, F.: A generalized approach to sparse and stable portfolio optimization problem. J. Ind. Manag. Optim. **14**(4), 1651–1666 (2018)
11. Daubechies, I., Defrise, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun. Pure Appl. Math. **57**(11), 1413–1457 (2004)
12. De Mol, C.: Financial Signal Processing and Machine Learning, chap. Sparse Markowitz Portfolios, pp. 11–22. Wiley Online Library (2016)
13. DeMiguel, V., Garlappi, L., Nogales, F.J., Uppal, R.: A generalized approach to portfolio optimization: improving performance by constraining portfolio norms. Manag. Sci. **55**(5), 798–812 (2009)
14. Fastrich, B., Paterlini, S., Winker, P.: Constructing optimal sparse portfolios using regularization methods. Comput. Manag. Sci. **12**(3), 417–434 (2015)
15. Fazel, M., Pong, T.K., Sun, D., Tseng, P.: Hankel matrix rank minimization with applications to system identification and realization. SIAM J. Matrix Anal. Appl. **34**(3), 946–977 (2013)
16. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. **2**(1), 17–40 (1976)
17. Huang, X.: Portfolio selection with a new definition of risk. Eur. J. Oper. Res. **186**(1), 351–357 (2008)
18. Konno, H., Suzuki, K.: A mean-variance-skewness portfolio optimization model. J. Oper. Res. Soc. Jpn. **38**(2), 173–187 (1995)
19. 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. **19**, 1–28 (2018)
20. Li, B., Li, X., Teo, K.L., Zheng, P.: A new uncertain random portfolio optimization model for complex systems with downside risks and diversification. Chaos Solitons Fractals **160**, 112,213 (2022)
21. Łojasiewicz, S.: Une propriété topologique des sous-ensembles analytiques réels. Les Équ. aux Dériv. Partielles **117**, 87–89 (1963)
22. Markowitz, H.: Portfolio selection. J. Finance **7**(1), 77–91 (1952)
23. Michaud, R.O.: The Markowitz optimization enigma: is 'optimized' optimal? Financ. Anal. J. **45**(1), 31–42 (1989)
24. Parikh, N., Boyd, S.: Proximal algorithms. Found. Trends Optim. **1**(3), 127–239 (2014)
25. Perrin, S., Roncalli, T.: Machine learning optimization algorithms & portfolio allocation. Mach. Learn. Asset Manag.: New Dev. Financ. Appl., 261–328 (2020)
26. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer, New York (1998)
27. Sun, Y., Aw, G., Teo, K.L., Zhou, G.: Portfolio optimization using a new probabilistic risk measure. J. Ind. Manag. Optim. **11**(4), 1275 (2015)

## Publisher's Note