Penalty PALM Method for Sparse Portfolio Selection Problems

Yue Teng[a], Li Yang[b], Bo Yu[a*] and Xiaoliang Song[a]

[a]School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, P.R. China; [b]School of Science, Dalian University of Technology, Panjin 124221, P.R. China

In this paper, we propose a penalty proximal alternating linearized minimization method for the large-scale sparse portfolio problems in which a sequence of penalty subproblems are solved by utilizing proximal alternating linearized minimization frame and sparse projection techniques. For exploiting the structure of the problems and reducing the computation complexity, each penalty subproblem is solved by alternately solving two projection subproblems. The global convergence of the method to a KKT point or a local minimizer of the problem can be proved under the characteristic of the problem. The computational results with practical problems demonstrate that our method can find the suboptimal solutions of the problems efficiently and is competitive with some other local solution methods.

Keywords: sparse portfolio selection; proximal alternating linearized minimization method; l_0 minimization; cardinality constrained portfolio selection

AMS Subject Classification: 65K05; 90C26; 90C30; 91G10; 91G60

1. Introduction

The classical mean-variance (MV) portfolio selection formulation [1] has been constructed in a frictionless world. When the number of assets is typically large and the returns of these assets are highly correlated, the solution of MV problem is usually nonzero on almost all of the components. It means investors need to invest in a large number of assets. However, in real world for different reasons especially high administrative and transaction costs, almost all the investors can only invest in a limited number of assets. This gap motivates many researchers to study the sparse mean-variance portfolio selection problem, i.e., get a sparse asset allocation (solution). Another motivation of selecting a sparse portfolio is that if the returns of assets are highly correlated and the estimates of the mean and covariance of assets are not suitable, the MV problem usually performs poorly and the portfolio obtained from the problem is very sensitive to the perturbations in the parameters of the problem. Many statistical regularization methods have wide applications in constructing MV problems to find sparse optimal portfolios with better out-of-sample performances and to reduce the transaction costs [2–9].

One type of method for getting sparse portfolios is introducing the cardinality constraint in the original problem. A cardinality constrained mean-variance (CCMV) port-

---

*Corresponding author. Email: yubo@dlut.edu.cn

folio selection problem can be expressed as

$$\begin{aligned}
\min_{x} \quad & \frac{1}{2}x^{\top}Qx \\
\text{s.t.} \quad & e^{T}x = 1, \\
& R^{\top}x \geq r, \\
& 0 \leq x \leq u, \\
& \|x\|_0 \leq K,
\end{aligned} \tag{1}$$

where $\|x\|_0$ represents the number of the nonzero entries of x, $0 < K < n$ is an integer and represents the upper limit of the number of assets which are managed in the portfolio. Q is an $n \times n$ symmetric positive semidefinite covariance matrix, e is a n-dimensional vector with all components one, $R \in \mathrm{R}^n$ denotes the expected return of n assets and r is a minimum profit target. The constraint $0 \leq x \leq u$ indicates that short-selling is prohibited and the investment proportion of each asset has an upper bound, where u is a positive vector. Another type of method is utilizing the l_0-norm regularized model:

$$\begin{aligned}
\min_{x} \quad & \frac{1}{2}x^{\top}Qx + \nu\|x\|_0 \\
\text{s.t.} \quad & e^{T}x = 1, \\
& R^{\top}x \geq r, \\
& 0 \leq x \leq u,
\end{aligned} \tag{2}$$

where $\nu > 0$ is a regular parameter. Problems (1) and (2) are two special case of general l_0 minimization problems which are generally NP-hard. Many methods are proposed for solving the l_0-regularized problems arising in compressed sensing and image processing, but these methods cannot be applied to solve the l_0 minimization problems (1) and (2).

Problem (1) is a cardinality constrained quadratic optimization problems which have been studied in the literature [10–15]. The solution methods of problem (1) in the literature can be classified into two main categories: integer programming methods and heuristic or local methods. Most of the integer programming methods transform problem (1) to a mixed integer quadratic programming problem and use branch-and-bound framework to solve it. A surrogate constraint and a convex relaxation of the cardinality constraint relaxation are considered in [10, 16]. Several discrete features of the real trade and a class of factor models are discussed in [17] and [14, 18]. For quadratic programs with semi-continuous variables a perspective reformulation is proposed by [19, 20], which can also deal with the cardinality constraint. The heuristic methods for problem (1) generally based on genetic algorithm, tabu search and simulated annealing. Several heuristic methods for portfolio selection with discrete features are proposed in [12] and [15]. An efficient heuristic method for problem (1) by using the structure of factor-model of the asset returns and the clustering techniques are proposed in [21]. The integer programming methods and heuristic methods require a large amount of computation time for solving large-scale problems.

The l_1-norm approximation has been a popular method in compressed sensing and other sparse optimization problems [22]. We can obtain a convex approximation of problem (1) and (2) by replacing the l_0-norm by the l_1-norm. However, the l_1-norm does not often produce solutions with desired sparsity. It can be witnessed in many computational results and the literature [23]. Several nonconvex approximations to the l_0-norm have been

studied recently in the literature. In [24], the CCMV problem is formulated as a DC program to find a local optimal solution of the problem. Based on a new piecewise linear DC approximation of the cardinality function, a new DC approximation method for solving problem (1) is proposed in [25]. A novel penalty decomposition (PD) method for solving a stationary point or a local optimal solution of a general l_0-norm minimization problem is proposed in [26]. Recently, a novel proximal alternating linearized minimization (PALM) method for general nonconvex and nonsmooth problems has been developed in [27].

The PD method and PALM method can be grouped with the alternating minimization methods. In this paper, inspired by these two work, we develop a penalty proximal alternating linearized minimization (penalty PALM) method to solve problems (1) and (2). In order to exploit the structure of the problems, the penalty PALM method generates a sequence of penalty subproblems which are nonconvex and nonsmooth problems and utilizes the PALM framework to solve these subproblems. Utilizing the characteristic of problems (1) and (2), we establish that any accumulation point of the iterative sequence generated by the penalty PALM method satisfies the first-order optimality conditions of the problems. Furthermore, under a suitable assumption we show that such an accumulation point is also a local minimizer of the problems. In addition, from the property of PALM method, any accumulation point of the sequence generated by the PALM method is a critical point of the penalty subproblem. Finally, we test the performance of the penalty PALM method by applying it to solve practical problems. The computational results demonstrate that the framework of our method is more easily in programming and the solutions generated by our method have better quality and consume lesser amount of calculation than the solutions generated by the PD method.

The rest of this paper is organized as follows. In Section 2, we establish the first-order optimality conditions for problems (1) and (2). In Section 3, we develop the penalty PALM method for the problems. The convergence result of the penalty PALM method is established in Section 4. In Section 5, we carry out numerical experiments to test the performance of our penalty PALM method for solving problems (1) and (2). Finally, we give some concluding remarks in Section 6.

Notation. In this paper, the symbols R^n denote the n-dimensional Euclidean space. For any real vector, $\|\cdot\|_0$ and $\|\cdot\|$ denote the l_0-norm (i.e., the cardinality or the number of the nonzero entries) and the l_2-norm, respectively. Given an index set $L \subseteq \{1,\ldots,n\}$, $|L|$ denotes the size of L, and the elements of L are always arranged in ascending order. x_L denotes the subvector formed by the entries of x indexed by L. For any two sets $\mathcal{A}$ and $\mathcal{B}$, the subtraction of $\mathcal{A}$ and $\mathcal{B}$ is given by $\mathcal{A} \setminus \mathcal{B} = \{x \in \mathcal{A} : x \notin \mathcal{B}\}$. Given a closed set $\mathcal{C} \subseteq \mathrm{R}^n$, let $N_{\mathcal{C}}(x)$ denote the normal cone of $\mathcal{C}$ at any $x \in \mathcal{C}$, respectively.

2.   First-order optimality conditions

Lu and Zhang have discussed the first-order optimality conditions for general l_0 minimization problems in [26]. In this section, we give the first-order necessary conditions and the first-order sufficient conditions for problems (1) and (2), as special cases of the optimality conditions for general l_0 minimization problems.

THEOREM 2.1   *Assume x^* is a local minimizer of problem (1). Let $I^* \subseteq \{1,\cdots,n\}$ be an index set with $|I^*| = K$ such that $x_i^* = 0$ for all $i \in \bar{I}^*$, where $\bar{I}^* = \{1,\cdots,n\} \setminus I^*$. Suppose that at x^* the following Mangasarian-Fromovitz constraint qualification (MFCQ)*

condition holds, that is, there exists a point $x \in \mathrm{R}^n$ such that

$$
\begin{aligned}
-R^\top (x - x^*) < 0, & \quad \text{if } R^\top x^* = r, \\
e_i^\top (x - x^*) < 0, & \quad \text{if } x_i^* = u_i, \\
-e_i^\top (x - x^*) < 0, & \quad \text{if } x_i^* = 0, i \in I^*, \\
e_i^\top (x - x^*) = 0, & \quad \text{if } i \in \bar{I}^*, \\
e^\top (x - x^*) = 0, &
\end{aligned}
\tag{3}
$$

is satisfied, where e_i is an n-dimensional vector of which the i-th component is equal to 1 and other components are equal to 0. Then, there exists $(\lambda_1^*, \lambda_2^*, \lambda_3^*, \mu^*, z^*) \in \mathrm{R}^{3n+2}$ together with x^* satisfying

$$
\begin{aligned}
Qx^* - \lambda_1^* R + \lambda_2^* - \lambda_3^* + \mu^* e + z^* = 0, \\
\lambda_1^*, \lambda_2^*, \lambda_3^* \geq 0, \ \lambda_1^* (r - R^\top x^*) = 0, \\
(\lambda_2^*)_i (x_i^* - u_i) = 0, \ (\lambda_3^*)_i x_i^* = 0, \ i = 1, \cdots, n, \\
R^\top x^* \geq r, \ e^\top x^* = 1, \ 0 \leq x \leq u; \\
(\lambda_3^*)_i = 0, \ i \in \bar{I}^*; \ z_i^* = 0, \ i \in I^*.
\end{aligned}
\tag{4}
$$

Proof. It is obvious that if x^* is a local minimizer of problem (1), x^* is also a minimizer of the problem:

$$
\min_x \left\{ \frac{1}{2} x^\top Q x : \ R^\top x \geq r, \ e^\top x = 1, \ 0 \leq x_{I^*} \leq u, \ x_{\bar{I}^*} = 0 \right\}.
\tag{5}
$$

Together with the MFCQ condition and Theorem 3.25 of [28], the conclusion holds. ∎

THEOREM 2.2 *Assume x^* is a local minimizer of problem (2). Let $I^* = \{i \in \{1, \ldots, n\} : x_i^* \neq 0\}$, $\bar{I}^* = \{1, \cdots, n\} \backslash I^*$. Suppose that at x^* the MFCQ condition holds, that is, there exists a point $x \in \mathrm{R}^n$ such that (3) is satisfied. Then, there exists $(\lambda_1^*, \lambda_2^*, \lambda_3^*, \mu^*, z^*) \in \mathrm{R}^{3n+2}$ together with x^* satisfying (4).*

Proof. It is obvious that x^* is a local minimizer of problem (2) if and only if x^* is a minimizer of problem (5). Together with the MFCQ condition and Theorem 3.25 of [28], the conclusion holds. ∎

The l_0-norm is the only nonconvex part of problems (1) and (2). Using this observation and the conclusion in [26] we can establish the first-order sufficient optimality conditions for problems (1) and (2).

THEOREM 2.3 *The objective function of problem (1) is convex and the constraints of problem (1) except the cardinality constraint are affine. Let x^* be a feasible point of problem (1), and let $\mathcal{I}^* = \{I^* \subseteq \{1, \cdots, n\} : |I^*| = K, \ x_i^* = 0, \ \forall i \in \{1, \cdots, n\} \backslash I^*\}$. Suppose that for any $I^* \in \mathcal{I}^*$, there exists some $(\lambda_1^*, \lambda_2^*, \lambda_3^*, \mu^*, z^*) \in \mathrm{R}^{3n+2}$ satisfying (4). Then, x^* is a local minimizer of problem (1).*

Proof. : By the assumptions and Theorem 3.27 of [28], x^* is a minimizer of problem (5) for all $\bar{I}^* \in \{\{1, \cdots, n\} \backslash I^* : I^* \in \mathcal{I}^*\}$. Let $q(x) = (1/2) x^\top Q x$, then there exists $\varepsilon > 0$ such that $q(x) \geq q(x^*)$ for all $x \in \cup_{I^* \in \mathcal{I}^*} \mathcal{O}_{I^*}(x^*; \varepsilon)$ where

$$
\mathcal{O}_{I^*}(x^*; \varepsilon) = \left\{ x \in \mathrm{R}^n : e^\top x = 1, R^\top x \geq r, 0 \leq x_{I^*} \leq u, x_{\bar{I}^*} = 0, \|x - x^*\| < \varepsilon \right\}
$$

with $\bar{I}^* = \{1, \cdots, n\} \setminus I^*$. We can observe from (1) that for any $x \in \mathcal{O}(x^*; \varepsilon)$, where

$$\mathcal{O}(x^*; \varepsilon) = \left\{ x \in \mathrm{R}^n : e^\top x = 1, R^\top x \geq r, 0 \leq x \leq u, \|x\|_0 \leq K, \|x - x^*\| < \varepsilon \right\},$$

there exists $I^* \in \mathcal{I}^*$ such that $x \in \mathcal{O}_{I^*}(x^*; \varepsilon)$ and hence $q(x) \geq q(x^*)$. It then implies that x^* is a local minimizer of (1). ∎

From Theorem 2.3 we observe that for any point x if $\|x\|_0 = K$ and x satisfies the first-order necessary conditions of problem (1), then x is a local minimizer of problem (1).

THEOREM 2.4 *The objective function of problem (2) except the l_0-norm part is convex and the constraints of problem (2) are affine. Let x^* be a feasible point of problem (2), and let $I^* = \{i \in \{1, \ldots, n\} : x_i^* \neq 0\}$, $\bar{I}^* = \{1, \cdots, n\} \setminus I^*$. Suppose that for such I^*, there exists some $(\lambda_1^*, \lambda_2^*, \lambda_3^*, \mu^*, z^*) \in \mathrm{R}^{3n+2}$ satisfying (4). Then, x^* is a local minimizer of problem (2).*

Proof. : By the assumptions and Theorem 3.27 of [28], x^* is a minimizer of problem (5) with $\bar{I}^* = \{1, \cdots, n\} \setminus I^*$. Any point is a local minimizer of problem (2) if and only if it is a minimizer of problem (5). Then, it implies that x^* is a local minimizer of (2). ∎

3. Penalty proximal alternating linearized minimization method

In this subsection, we propose the penalty proximal alternating linearized minimization (penalty PALM) method for solving problems (1) and (2). And introduce the projection techniques and closed-form solutions of two type of special l_0 minimization problems which are utilized in solving the penalty subproblems in the penalty PALM method.

3.1 *Penalty PALM method for problem (1)*

In this subsection we propose a penalty PALM method for solving problem (1). Inspired by PD method, we reformulated problem (1) as

$$\begin{aligned}
\min_{x,y} \quad & \frac{1}{2} x^\top Q x \\
\text{s.t.} \quad & R^\top x \geq r, \\
& e^\top x = 1, \\
& x = y, \\
& 0 \leq y \leq u, \\
& \|y\|_0 \leq K.
\end{aligned} \tag{6}$$

Let $\mathcal{C}_1 = \left\{ x \in \mathrm{R}^n : R^\top x \geq r, e^\top x = 1 \right\}$ denote the feasible set of the linear constraints, and $\mathcal{D}_1 = \left\{ y \in \mathrm{R}^n : \|y\|_0 \leq K, 0 \leq y \leq u \right\}$ denote the feasible set of other constraints.

We define the quadratic penalty problem of (1) as following:

$$
\begin{aligned}
\min_{x,y} \quad & \frac{1}{2} x^\top Q x + \frac{\rho}{2} \|x - y\|^2 \\
\text{s.t.} \quad & x \in \mathcal{C}_1, \\
& y \in \mathcal{D}_1,
\end{aligned}
\tag{7}
$$

where $\rho > 0$ is a penalty parameter.

The minimization problem (7) can be rewritten as the following unconstrained minimization problem:

$$
\min_{x,y} \Psi_\rho^1(x,y) = f_1(x) + q_\rho(x,y) + g_1(y) \quad \text{over all} \quad (x,y) \in \mathrm{R}^n \times \mathrm{R}^n.
\tag{8}
$$

by setting

$$
\begin{cases}
f_1(x) = I_{\mathcal{C}_1}(x); \\
q_\rho(x,y) = \frac{1}{2} x^\top Q x + \frac{\rho}{2} \|x - y\|^2; \\
g_1(y) = I_{\mathcal{D}_1}(y),
\end{cases}
\tag{9}
$$

where $I_{\mathcal{C}_1}(x)$ denotes the indicator function of $\mathcal{C}_1$ that satisfies $I_{\mathcal{C}_1}(x) = 0$ if $x \in \mathcal{C}_1$ and $+\infty$ otherwise as well as $I_{\mathcal{D}_1}(x)$. We can observe that problem (8) is a nonconvex and nonsmooth minimization problem.

Remark 1 From (9) and the definition of (1), we know that for problem (8) the following statements hold: (a) f_1 and g_1 are proper and lower semicontinuous functions and $q_\rho : \mathrm{R}^n \times \mathrm{R}^n \to \mathrm{R}$ is a smooth function. (b) For some fixed $\rho > 0$, let $L_1 = \|Q\| + \rho$, $L_2 = \rho$, then for any fixed y the partial gradient $\nabla_x q_\rho(x,y)$ is globally Lipschitz with moduli L_1, likewise, for any fixed x the partial gradient $\nabla_y q_\rho(x,y)$ is globally Lipschitz with moduli L_2. And (c) ∇q_ρ is Lipschitz continuous on bounded subsets of $\mathrm{R}^n \times \mathrm{R}^n$.

From Fermat's rule, if a point x is a local minimizer of a proper and lower semicontinuous function σ, then x is a critical point of σ. Recently, Bolte et al. developed a novel PALM method [27] which build on the Kurdyka-Łojasiewicz (KL) property for solving a broad class of nonconvex and nonsmooth minimization problems. The PALM method can be viewed as alternating the steps of the proximal forward-backward scheme. Without the convexity assumption, it is shown in [27] that for any proper, lower semicontinuous and bounded below functions $f_1(x)$, $g_1(x)$ and smooth function $q_\rho(x,y)$ whose gradient is Lipschitz continuous on any bounded set, if $\Psi_\rho^1(x,y)$ is a KL function then each bounded sequence generated by the PALM method globally converges to a critical point of problem (8).

The definition of the critical point of a nonconvex and nonsmooth function and the KL function will be given in section 4 as well as the proof of the function $\Psi_\rho^1(x,y)$ is a KL function. Next, we use the framework of the PALM method to solve problem (8) as well as the quadratic penalty problem (7). The PALM method updates the iterative

sequence $\{(x_k, y_k)\}$ via solving the following two proximal problems:

$$x^{k+1} \in \arg\min_x f_1(x) + \langle x - x^k, \nabla_x q_\rho(x^k, y^k) \rangle + \frac{t_1}{2}\|x - x^k\|^2, \tag{10}$$

$$y^{k+1} \in \arg\min_y g_1(y) + \langle y - y^k, \nabla_y q_\rho(x^{k+1}, y^k) \rangle + \frac{t_2}{2}\|y - y^k\|^2, \tag{11}$$

where $t_1 = \gamma_1 L_1$, $\gamma_1 > 1$ and $t_2 = \gamma_2 L_2$, $\gamma_2 > 1$ are two appropriately chosen step sizes, L_1 and L_2 are the lipschitz constants in Remark 1. Using the proximal map notation in [29], we get the minimization problems (10), (11) are equivalent to the following proximal problems:

$$x^{k+1} \in \mathrm{Prox}_{t_1}^{f_1}(x^k - \frac{1}{t_1}\nabla_x q_\rho(x^k, y^k)), \tag{12}$$

$$y^{k+1} \in \mathrm{Prox}_{t_2}^{g_1}(y^k - \frac{1}{t_2}\nabla_y q_\rho(x^{k+1}, y^k)), \tag{13}$$

Note that $f_1(x) = I_{\mathcal{C}_1}(x)$, $g_1(y) = I_{\mathcal{D}_1}(y)$ are indicator functions of nonempty and closed sets. The proximal maps (12), (13) reduce to the projection maps onto $\mathcal{C}_1$ and $\mathcal{D}_1$, defined by

$$x^{k+1} \in P_{\mathcal{C}_1}(w_1) := \arg\min_x \{\|x - w_1\| : x \in \mathcal{C}_1\}, \tag{14}$$

$$y^{k+1} \in P_{\mathcal{D}_1}(w_2) := \arg\min_y \{\|y - w_2\| : y \in \mathcal{D}_1\}, \tag{15}$$

where $w_1 = x^k - (1/t_1)\nabla_x q_\rho(x^k, y^k)$, $w_2 = y^k - (1/t_2)\nabla_y q_\rho(x^{k+1}, y^k)$.

Set $\mathcal{C}_1 = \{x \in \mathrm{R}^n : R^\top x \geq r, e^\top x = 1\}$ is a nonempty closed convex set, hence $P_{\mathcal{C}_1}$ is guaranteed to be a single-valued map and can be computed by the following two situation:

$$P_{\mathcal{C}_1}(w_1) = \begin{cases} w_1 + \frac{1 - e^\top w_1}{n}e, & \text{if } R^\top(w_1 + \frac{1 - e^\top w_1}{n}e) \geq r; \\ \arg\min_x \{\|x - w_1\| : R^\top x = r, e^\top x = 1\}, & \text{otherwise}, \end{cases} \tag{16}$$

where the second situation has a closed-form solution. As opposed to $\mathcal{C}_1$, $\mathcal{D}_1 = \{y \in \mathrm{R}^n : \|y\|_0 \leq K, 0 \leq y \leq u\}$ is a nonempty closed set and $P_{\mathcal{D}_1}(w_2)$ is a projection onto the sparse feasible set. Hence, $P_{\mathcal{D}_1}$ defines in a general multi-valued map, for convenience, we can call $P_{\mathcal{D}_1}$ a sparse projection. Suppose $\bar{y}^*$ is a point of set $P_{\mathcal{D}_1}(w_2)$, then $\bar{y}^*$ can be computed as follows:

$$\bar{y}_i^* = \begin{cases} \bar{y}_i, & \text{if } i \in S^*; \\ 0, & \text{otherwise}, \end{cases} \quad i = 1, \ldots, n, \tag{17}$$

where $\bar{y}_i \in \arg\min\{(y_i - (w_2)_i)^2 : 0 \leq y_i \leq u_i\}$ and $S^* \subseteq \{1, \ldots, n\}$ be the index set corresponding to K largest values of $\{((w_2)_i)^2 - (\bar{y}_i - (w_2)_i)^2\}_{i=1}^n$.

Therefore, problem (8) can be solved by iteratively solving two projection subproblems. We rewrite the PALM method for solving problem (8) as following:

PALM method for (8):
Choose an initial point $(x^0, y^0) \in \mathrm{R}^n \times \mathrm{R}^n$. Set $k = 0$.

(1) Take $\gamma_1 > 1$, $t_1 = \gamma_1 L_1$ and compute

$$x^{k+1} \in \arg\min_x \left\{ \|x - (x^k - \frac{1}{t_1}\nabla_x q_\rho(x^k, y^k))\| : x \in \mathcal{C}_1 \right\}.$$

(2) Take $\gamma_2 > 1$, $t_2 = \gamma_2 L_2$ and compute

$$y^{k+1} \in \arg\min_y \left\{ \|y - (y^k - \frac{1}{t_2}\nabla_y q_\rho(x^{k+1}, y^k))\| : y \in \mathcal{D}_1 \right\}.$$

(3) Set $k \leftarrow k + 1$ and go to step (1).

Different from the penalty PALM method, the PD method for problem (1) rewrite Problem (1) as the following general cardinality constrained problem:

$$\min_x \left\{ \overline{f}(x) : \overline{g}(x) \leq 0, \ \overline{h}(x) = 0, \ \|x\|_0 \leq K \right\},$$

and reformulated the general problem above as

$$\min_{x,y} \left\{ \overline{f}(x) : \overline{g}(x) \leq 0, \ \overline{h}(x) = 0, \ x = y, \ \|y\|_0 \leq K \right\}. \tag{18}$$

The quadratic penalty function of the PD method associating with problem (18) is defined as

$$q_\rho^{\mathrm{PD}}(x, y) = \overline{f}(x) + \frac{\rho}{2}(\|[\overline{g}(x)]^+\|_2^2 + \|\overline{h}(x)\|_2^2 + \|x - y\|_2^2),$$

where $x \in \mathrm{R}^n$, $y \in \{y \in \mathrm{R}^n : \|y\|_0 \leq K\}$ and $\rho > 0$. The quadratic penalty problem

$$\min_{x,y} \left\{ q_\rho^{\mathrm{PD}}(x, y) : \ x \in \mathrm{R}^n, \ y \in \mathrm{R}^n, \ \|y\|_0 \leq K \right\},$$

is regarded as a good approximation to the original problem (1) for sufficiently large ρ. The PD method divides the cardinality constraint from other constraints of problem (1) by utilizing artificial variables and quadratic penalty functions, which make the cardinality constraint easier to solve. However, the PD method was designed for general l_0 minimization problems and did not fully exploit the structure of the constraints of problem (1). The PD method solves problem (1) by solving a sequence of penalty subproblems which are solved by a block coordinate descent (BCD) method. To solve the penalty subproblem, the BCD method needs to fix y and solve a unconstrained optimization problem, then fix x and solve a special l_0 minimization problem which has closed-form solution. Hence, we need to solve many unconstrained optimization problems during the entire solving process of the PD method. It may cost a lot of computations when we using the PD method to solve large-scale problems.

We are now ready to propose the penalty PALM method for (approximate) solving problem (6) (or equivalently problem (1)) in which each penalty subproblem is solved by the PALM method.

Penalty PALM method for problem (1):

Let $\{\varepsilon_k\}$ be a positive decreasing sequence. Let $\rho_0 > 0$, $\sigma > 1$, $\gamma_1 > 1$, $\gamma_2 > 1$ be given. Choose an arbitrary point $(x_0^0, y_0^0) \in \mathrm{R}^n \times \mathrm{R}^n$. Set $k = 0$.

(1) Set $l = 0$, $L_1^k = \|Q\| + \rho_k$, $L_2^k = \rho_k$, $t_1^k = \gamma_1 L_1^k$, $t_2^k = \gamma_2 L_2^k$ and apply the PALM method to find an approximate critical point $(x^k, y^k) \in \mathcal{C}_1 \times \mathcal{D}_1$ of the penalty subproblem

$$\min_{x,y} \ \Psi_{\rho_k}^1(x, y) = I_{\mathcal{C}_1}(x) + q_{\rho_k}(x, y) + I_{\mathcal{D}_1}(y), \quad (x, y) \in \mathrm{R}^n \times \mathrm{R}^n \quad (19)$$

by performing steps (1.1)-(1.4):

(1.1) $w_1 = x_l^k - \frac{1}{t_1^k}\nabla_x q_{\rho_k}(x_l^k, y_l^k)$, compute $x_{l+1}^k \in P_{\mathcal{C}_1}(w_1)$.

(1.2) $w_2 = y_l^k - \frac{1}{t_2^k}\nabla_y q_{\rho_k}(x_{l+1}^k, y_l^k)$, compute $y_{l+1}^k \in P_{\mathcal{D}_1}(w_2)$.

(1.3) If (x_{l+1}^k, y_{l+1}^k) satisfies

$$\mathrm{dist}(0, \partial\Psi_{\rho_k}^1(x_{l+1}^k, y_{l+1}^k)) < \varepsilon_k, \quad (20)$$

set $(x^k, y^k) := (x_{l+1}^k, y_{l+1}^k)$ and go to step (2).

(1.4) Set $l \leftarrow l + 1$ and go to step (1.1).

(2) Set $\rho_{k+1} = \sigma\rho_k$, and $(x_0^{k+1}, y_0^{k+1}) := (x^k, y^k)$.

(3) Set $k \leftarrow k + 1$ and go to step (1).

Remark 2 The condition (20) will be used in the analysis of the global convergence of the penalty PALM method, but it may not be easily confirmed. Just like the PD method, this condition can be replaced by another practical termination condition which is based on the relative change of the sequence $\{(x_l^k, y_l^k)\}$, that is,

$$\max\left\{\frac{\|x_{l+1}^k - x_l^k\|_\infty}{\max(\|x_l^k\|_\infty, 1)}, \frac{\|y_{l+1}^k - y_l^k\|_\infty}{\max(\|y_l^k\|_\infty, 1)}\right\} \leq \varepsilon_I \quad (21)$$

for some $\varepsilon_I > 0$. In addition, we can terminate the out iterations of the penalty PALM method by condition

$$\|x^k - y^k\|_\infty \leq \varepsilon_O \quad \text{or} \quad \frac{\|(x^{k+1}, y^{k+1}) - (x^k, y^k)\|_\infty}{\max(\|(x^k, y^k)\|_\infty, 1)} \leq \varepsilon_O, \quad (22)$$

for some $\varepsilon_O > 0$. For enhance the performance and convergence of the penalty PALM method, we may check whether the sparse subset of $\mathcal{D}_1$ which contain y^k and $\mathcal{C}_1$ have an intersection and recompute the penalty subproblem multiple times from a suitable perturbation of the current best approximate solution.

In numerical experiments, we found that for the CCMV problem the penalty PALM method can find the support set (index set of the nonzero components) of a local solution quickly, but need more iterations to improve the solution accuracy. For speeding up the computation, we can use an end game strategy which needs not necessary to set the parameters of termination conditions too strict (small) and find a very accurate solution. We can choose some moderate parameters and quickly find a support set of a local solution, then use a heuristic method that set the other variables which not belong

to the support set to zero, and solve a K-dimensional convex quadratic programming problem to find an accurate local solution of problem (1).

3.2 *Penalty PALM method for problem (2)*

In this subsection we propose a penalty PALM method for solving problem (2). Similar with the penalty PALM method for problem (1), we reformulated problem (2) as

$$
\begin{aligned}
\min_{x,y} \quad & \frac{1}{2}x^\top Q x + \nu\|x\|_0 \\
\text{s.t.} \quad & R^\top x \geq r, \\
& e^\top x = 1, \\
& x = y, \\
& 0 \leq y \leq u.
\end{aligned}
\tag{23}
$$

Let $\mathcal{C}_2 = \left\{x \in \mathrm{R}^n : R^\top x \geq r, e^\top x = 1\right\}$, and $\mathcal{D}_2 = \{y \in \mathrm{R}^n : 0 \leq y \leq u\}$. Similarly, we define the quadratic penalty problem of problem (2) as following:

$$
\begin{aligned}
\min_{x,y} \quad & \frac{1}{2}x^\top Q x + \frac{\rho}{2}\|x - y\|^2 + \nu\|y\|_0 \\
\text{s.t.} \quad & x \in \mathcal{C}_2, \\
& y \in \mathcal{D}_2,
\end{aligned}
\tag{24}
$$

where $\rho > 0$ is a penalty parameter.

By setting

$$
\begin{cases}
f_2(x) = I_{\mathcal{C}_2}(x); \\
q_\rho(x,y) = \dfrac{1}{2}x^\top Q x + \dfrac{\rho}{2}\|x - y\|^2; \\
g_2(y) = I_{\mathcal{D}_2}(y) + \nu\|y\|_0,
\end{cases}
\tag{25}
$$

the minimization problem (24) also can be rewritten as a unconstrained minimization problem:

$$
\min_{x,y} \Psi_\rho^2(x,y) = f_2(x) + q_\rho(x,y) + g_2(y) \quad \text{over all} \quad (x,y) \in \mathrm{R}^n \times \mathrm{R}^n.
\tag{26}
$$

Remark 3 From (25) and the definition of (2) we can observe that: (a) problem (26) is a nonconvex and nonsmooth minimization problem; (b) f_2 and g_2 are proper and lower semicontinuous functions; (c) q_ρ is same as the function defined in the penalty PALM method for problem (1) and has the same properties.

Hence, problem (26) can be solved by utilizing the PALM framework. The iterative sequence $\{(x_k, y_k)\}$ are updated by solving the following two subproblems:

$$
x^{k+1} \in P_{\mathcal{C}_2}(w_1) := \arg\min_x \left\{\|x - w_1\| : x \in \mathcal{C}_2\right\},
\tag{27}
$$

$$
y^{k+1} \in \mathrm{Prox}_{t_2}^{g_2}(w_2) := \arg\min_y \left\{\frac{t_2}{2}\|y - w_2\|^2 + \nu\|y\|_0 : y \in \mathcal{D}_2\right\},
\tag{28}
$$

10

where $w_1 = x^k - (1/t_1)\nabla_x q_\rho(x^k, y^k)$, $w_2 = y^k - (1/t_2)\nabla_y q_\rho(x^{k+1}, y^k)$.

It is obvious that $P_{\mathcal{C}_1} = P_{\mathcal{C}_2}$ and $\text{Prox}_{t_2}^{g_2}$ is a hard thresholding operator. Similar to $P_{\mathcal{D}_1}$, $\text{Prox}_{t_2}^{g_2}$ is a multi-valued map. Suppose $\bar{y}^* \in \text{Prox}_{t_2}^{g_2}(w_2)$, then $\bar{y}^*$ can be computed as follows:

$$\bar{y}_i^* = \left\{ \begin{array}{ll} \bar{y}_i, & \text{if } \bar{v}_i \geq 0; \\ 0, & \text{otherwise,} \end{array} \right. \quad i = 1, \ldots, n, \tag{29}$$

where $\bar{y}_i \in \arg\min\limits_{y_i}\{(y_i - (w_2)_i)^2 : 0 \leq y_i \leq u_i\}$ and $\bar{v}_i = (t_2/2)((w_2)_i)^2 - \nu - (t_2/2)(\bar{y}_i - (w_2)_i)^2$ for $i = 1, \ldots, n$.

We now present the penalty PALM method for problem (2) which is similar to the PALM method for problem (1).

Penalty PALM method for problem (2):
Let $\{\varepsilon_k\}$ be a positive decreasing sequence. Let $\rho_0 > 0$, $\sigma > 1$, $\gamma_1 > 1$, $\gamma_2 > 1$ be given. Choose an arbitrary point $(x_0^0, y_0^0) \in \mathrm{R}^n \times \mathrm{R}^n$. Set $k = 0$.

(1) Set $l = 0$, $L_1^k = \|Q\| + \rho_k$, $L_2^k = \rho_k$, $t_1^k = \gamma_1 L_1^k$, $t_2^k = \gamma_2 L_2^k$ and apply the PALM method to find an approximate critical point $(x^k, y^k) \in \mathcal{C}_2 \times \mathcal{D}_2$ of the penalty subproblem

$$\min_{x,y} \Psi_{\rho_k}^2(x, y) = I_{\mathcal{C}_2}(x) + q_{\rho_k}(x, y) + I_{\mathcal{D}_2}(y) + \nu\|y\|_0, \quad (x, y) \in \mathrm{R}^n \times \mathrm{R}^n \tag{30}$$

by performing steps (1.1)-(1.4):
(1.1) $w_1 = x_l^k - \frac{1}{t_1^k}\nabla_x q_{\rho_k}(x_l^k, y_l^k)$, compute $x_{l+1}^k \in P_{\mathcal{C}_2}(w_1)$.
(1.2) $w_2 = y_l^k - \frac{1}{t_2^k}\nabla_y q_{\rho_k}(x_{l+1}^k, y_l^k)$, compute $y_{l+1}^k \in \text{Prox}_{t_2}^{g_2}(w_2)$.
(1.3) If (x_{l+1}^k, y_{l+1}^k) satisfies

$$\text{dist}(0, \partial\Psi_{\rho_k}^2(x_{l+1}^k, y_{l+1}^k)) < \varepsilon_k, \tag{31}$$

set $(x^k, y^k) := (x_{l+1}^k, y_{l+1}^k)$ and go to step (2).
(1.4) Set $l \leftarrow l + 1$ and go to step (1.1).
(2) Set $\rho_{k+1} = \sigma\rho_k$, and $(x_0^{k+1}, y_0^{k+1}) := (x^k, y^k)$.
(3) Set $k \leftarrow k + 1$ and go to step (1).

Remark 4 The practical termination criterias used in Subsection 3.1 can also be applied to this method for problem (2).

Like the PD method for problem (1), the PD method for problem (2) divides the l_0-norm regular term from the other parts of penalty subproblem. However, this PD method still need to solve a series of unconstrained optimization problems during the entire solving process. It also cost a lot of computations when we using the PD method to solve large-scale problems. For speeding up the penalty PALM method, the end game strategy used in problem (1) can also be used in problem (2).

4. Convergence Analysis

In this section we establish the global convergence of the penalty PALM method for problems (1) and (2) in which each penalty subproblem is a nonconvex and nonsmooth

problem and approximately solved by PALM method. We first introduce the definition of a critical point of a nonconvex function.

DEFINITION 1   Let $\sigma : \mathrm{R}^n \to (-\infty, +\infty]$ be a proper and lower semicontinuous function and $\operatorname{dom} \sigma = \{x \in \mathrm{R}^n : \sigma(x) < +\infty\}$.
  (i) For a given $x \in \operatorname{dom} \sigma$, the Frséchet subdifferential of σ at x is defined as

$$\widehat{\partial}\sigma(x) = \left\{ u : \liminf_{y \to x, y \neq x} \frac{\sigma(y) - \sigma(x) - \langle u, y - x \rangle}{\|y - x\|} \geq 0 \right\}$$

and $\widehat{\partial}\sigma(x) = \emptyset$ if $x \notin \operatorname{dom}\sigma$.
  (ii) The Limiting Sub-differential of σ at x is defined as

$$\partial\sigma(x) = \left\{ u \in \mathrm{R}^n : \exists\, x_k \to x, \ \sigma(x_k) \to \sigma(x) \text{ and } u_k \in \widehat{\partial}\sigma(x_k) \to u \right\}.$$

  (iii) The point x is a critical point of σ if $0 \in \partial\sigma(x)$.

For a general proper and lower semicontinuous function σ, if a point x is a local minimizer of σ then x is a critical point of σ. For notational convenience, we omit the index k of the subproblems in the penalty PALM method for problems (1) and (2) and rewrite them as

$$\min_{x,y} \Psi^1_\rho(x,y) = I_{\mathcal{C}_1}(x) + q_\rho(x,y) + I_{\mathcal{D}_1}(y), \quad (x,y) \in \mathrm{R}^n \times \mathrm{R}^n, \tag{32}$$

and

$$\min_{x,y} \Psi^2_\rho(x,y) = I_{\mathcal{C}_2}(x) + q_\rho(x,y) + I_{\mathcal{D}_2}(y) + \nu\|y\|_0, \quad (x,y) \in \mathrm{R}^n \times \mathrm{R}^n. \tag{33}$$

Bolte et al. showed that for any nonconvex and nonsmooth problem which satisfies the assumptions A and B in [27] and the Kurdyka-Łojasiewicz (KL) property, any bounded sequence generated by PALM method globally converges to a critical point of the problem. From Remark 1 and 3 in section 3, we can confirm that functions $\Psi^1_\rho(x,y)$ and $\Psi^2_\rho(x,y)$ are proper and lower semicontinuous functions and satisfy the assumptions in [27]. Under these conditions and the conclusion of [27], we give the convergence theorem which ensure the convergency of the iterative sequences generated in solving the penalty subproblems.

THEOREM 4.1   *The sequence $\{(x_l, y_l)\}$ generated by the PALM method in solving the penalty subproblem (32) (or (33)) converges to a critical point of (32) (or (33)) , if the following conditions hold:*
  (i) $\Psi^1_\rho(x,y)$ (or $\Psi^2_\rho(x,y)$) is a KL function;
  (ii) $\{(x_l, y_l)\}$ is a bounded sequence.

The level set of the function $\Psi^1_\rho(x,y)$ (or $\Psi^2_\rho(x,y)$) is bounded, and it is shown in [27] that the function value sequence $\{\Psi^1_\rho(x_l, y_l)\}$ (or $\{\Psi^2_\rho(x_l, y_l)\}$) is nonincreasing. Therefore the sequence $\{(x_l, y_l)\}$ is bounded. If a proper and lower semicontinuous function σ satisfies the KL property at each point of $\operatorname{dom}\partial\sigma$ then σ is called a KL function. The KL property plays a central role in the convergence analysis of the PALM method. We can see the definition in [27] for more details of KL property. From [30] we know that

a semi-algebraic function is a KL function. In the next, we give the definition of the semi-algebraic set and function, and show that functions $\Psi_\rho^1(x,y)$ and $\Psi_\rho^2(x,y)$ are semi-algebraic functions.

DEFINITION 2 (Semi-algebraic sets and functions)
 (i) A set $S \in \mathrm{R}^n$ is a real semi-algebraic set if there exists a finite number of real polynomial functions g_{ij}, $h_{ij} : \mathrm{R}^n \to \mathrm{R}$ such that

$$S = \bigcup_j \bigcap_i \{u \in \mathrm{R}^n : g_{ij}(u) = 0 \text{ and } h_{ij}(u) < 0\}.$$

 (ii) A function $h : \mathrm{R}^n \to (-\infty, +\infty]$ is called semi-algebraic if its graph

$$\{(u,t) \in \mathrm{R}^{n+1} : h(u) = t\}$$

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

LEMMA 4.2 *Each term in functions $\Psi_\rho^1(x,y)$ and $\Psi_\rho^2(x,y)$ are semi-algebraic functions, thus functions $\Psi_\rho^1(x,y)$ and $\Psi_\rho^2(x,y)$ are semi-algebraic functions.*

Proof. As we know the graph of $\|x\|_0$ is a piecewise linear set, then $\|x\|_0$ is a semi-algebraic function and set $\{x \in \mathrm{R}^n : \|x\|_0 \le K\}$ is a semi-algebraic set. Finite unions and intersections of semi-algebraic sets are still semi-algebraic sets, thus, sets $\mathcal{C}_1$, $\mathcal{D}_1$, $\mathcal{C}_2$, $\mathcal{D}_2$ are semi-algebraic sets. Real polynomial functions and indicator functions of semi-algebraic sets are semi-algebraic functions. Hence, each term in functions $\Psi_\rho^1(x,y)$ and $\Psi_\rho^2(x,y)$ are semi-algebraic functions. Finite sums of semi-algebraic functions are semi-algebraic functions. Therefore, functions $\Psi_\rho^1(x,y)$ and $\Psi_\rho^2(x,y)$ are semi-algebraic functions. ∎

From the discussion above, the conditions of Theorem 4.1 are satisfied and the sequence generated by the PAlM method converges to a critical point of penalty subproblem (32) or (33).

Now we show that any accumulation point of the sequence generated by penalty PALM method for problem (1) satisfies the first-order necessary conditions of problem (1). Moreover we can show that under a suitable assumption, any accumulation point is a local minimizer of problem (1).

THEOREM 4.3 *Assume that $\{\varepsilon_k\} \to 0$. Let $\{(x^k, y^k)\}$ be the sequence generated by the Penalty PALM method for problem (1), $I_k = \{i_1^k, \ldots, i_K^k\}$ be a set of K distinct indices in $\{1, \cdots, n\}$ such that $(y^k)_i = 0$ for any $i \notin I_k$. Then the following statements hold:*
 (i) The sequence $\{(x^k, y^k)\}$ is bounded.
 (ii) Suppose (x^*, y^*) is an accumulation point of $\{(x^k, y^k)\}$. Then $x^* = y^*$ is a feasible point of problem (1). Moreover, there exists a subsequence J such that $\{(x^k, y^k)\}_{k \in J} \to (x^*, y^*)$, $I_k = I^*$ for some index set $I^* \subseteq \{1, \cdots, n\}$ when $k \in J$ is sufficiently large.
 (iii) Let x^*, J, and I^* be defined as above, and let $\bar{I}^* = \{1, \cdots, n\} \setminus I^*$. Suppose that the MFCQ condition (3) holds at x^* for I^* and $\bar{I}^*$. Then x^* satisfies the first-order optimality conditions (4). Moreover, if $\|x^*\|_0 = K$, x^* is a local minimizer of (1).

Proof. By using the projection operator, we can observe that each y_k belong to feasible set $\mathcal{D}_1$, then $\{y^k\}$ is bounded. The sequence generated by the PALM method is nonincreasing,

13

we can obtain from it that $\Psi^1_{\rho_k}(x^k, y^k) \leq \Psi^1_{\rho_k}(x^{k-1}, y^{k-1})$ and

$$\frac{1}{2}(x^k)^\top Q x^k + \rho_k \|x^k - y^k\|^2 \leq \frac{1}{2}(x^{k-1})^\top Q x^{k-1} + \rho_k \|x^{k-1} - y^{k-1}\|^2.$$

Therefore, it implies that $(x^k)^\top Q x^k \leq (x^{k-1})^\top Q x^{k-1}$ or $\|x^k - y^k\|^2 \leq \|x^{k-1} - y^{k-1}\|^2$. Since the objective function $(1/2)x^\top Q x$ is bounded below, the level set of objective function is compact and $\{\rho_k\} \to +\infty$, there exists $\|x^k - y^k\|^2 \leq \|x^{k-1} - y^{k-1}\|^2$ when k is sufficiently large. Together with the boundedness of $\{y^k\}$, it yields that $\{x^k\}$ is bounded. Statement (i) holds.

Since (x^*, y^*) is an accumulation point of $\{(x^k, y^k)\}$, there exists a subsequence $\{(x^k, y^k)\}_{k \in \bar{J}} \to (x^*, y^*)$. Recall that I_k is an index set. It follows that $\{(i_1^k, \ldots, i_K^k)\}_{k \in \bar{J}}$ is bounded for all k. Thus there exists a subsequence $J \subseteq \bar{J}$ such that $\{(i_1^k, \ldots, i_K^k)\}_{k \in J} \to (i_1^*, \ldots, i_K^*)$ for some K distinct indices $i_1^*, \ldots, i_K^*$. Since $i_1^*, \ldots, i_K^*$ are K distinct integers, we can obtain that $(i_1^k, \ldots, i_K^k) = (i_1^*, \ldots, i_K^*)$ when $k \in J$ is sufficiently large and moreover $\{(x^k, y^k)\}_{k \in J} \to (x^*, y^*)$. From the termination condition of PALM method, there exist vectors s_x^k, s_y^k such that

$$\begin{aligned}
0 &\in \partial_x I_{\mathcal{C}_1}(x^k) + \partial_x q_{\rho_k}(x^k, y^k) + s_x^k, \\
0 &\in \partial_y I_{\mathcal{D}_1}(y^k) + \partial_y q_{\rho_k}(x^k, y^k) + s_y^k.
\end{aligned}$$

By the property of the indicator function, we have $\partial_x I_{\mathcal{C}_1}(x^k) = N_{\mathcal{C}_1}(x^k)$, $\partial_y I_{\mathcal{D}_1}(y^k) = N_{\mathcal{D}_1}(y^k)$. Hence, it has

$$\begin{aligned}
0 &\in N_{\mathcal{C}_1}(x^k) + Q x^k + \rho_k(x^k - y^k) + s_x^k, & (34) \\
0 &\in N_{\mathcal{D}_1}(y^k) + \rho_k(y^k - x^k) + s_y^k, & (35)
\end{aligned}$$

and

$$\begin{aligned}
0 &\in N_{\mathcal{C}_1}(x^k) + \frac{1}{\rho_k} Q x^k + (x^k - y^k) + \frac{1}{\rho_k} s_x^k, \\
0 &\in N_{\mathcal{D}_1}(y^k) + (y^k - x^k) + \frac{1}{\rho_k} s_y^k.
\end{aligned}$$

From (20) we observe that $\|s_x^k\| \leq \varepsilon_k$, $\|s_y^k\| \leq \varepsilon_k$ for all k, which together with $\lim_{k \to \infty} \varepsilon_k = 0$ implies $\lim_{k \to \infty} s_x^k = 0$, $\lim_{k \to \infty} s_y^k = 0$. Moreover, from $\{\rho_k\} \to +\infty$, taking limits on both sides of two expressions as $k \in J \to \infty$, we have

$$\begin{aligned}
0 &\in N_{\mathcal{C}_1}(x^*) + (x^* - y^*), \\
0 &\in N_{\mathcal{D}_1}(y^*) + (y^* - x^*).
\end{aligned}$$

By the properties of problem (1), we obtain that $x^* = y^* \in \mathcal{C}_1 \cap \mathcal{D}_1$. Therefore statement (ii) holds.

From (34) and (35), we obtain that

$$0 \in N_{\mathcal{C}_1}(x^k) + N_{\mathcal{D}_1}(y^k) + Q x^k + s_x^k + s_y^k. \tag{36}$$

Taking limits on both sides of (36) as $k \in J \to \infty$, by $\lim_{k \to \infty} s_x^k = 0$, $\lim_{k \to \infty} s_y^k = 0$, the

semicontinuity of $N_{\mathcal{C}_1}(x^k)$, $N_{\mathcal{D}_1}(y^k)$, and $x^* = y^*$, we can obtain that

$$0 \in N_{\mathcal{C}_1}(x^*) + N_{\mathcal{D}_1}(x^*) + Qx^*.$$

By the suppose, $x^* \in \mathcal{C}_1 \cap \mathcal{D}_1$ satisfies the MFCQ condition. Hence, we obtain that there exist $\lambda_1^*, \lambda_2^*, \lambda_3^*, \mu^*, z^*$ satisfying the first-order optimality conditions (4) for problem (1). Further, if $\|x^*\|_0 = K$, the conditions of Theorem 2.3 hold. It then follows from Theorem 2.3 that x^* is a local minimizer of problem (1). Statement (iii) holds. ∎

We next show that any accumulation point of the sequence generated by penalty PALM method for problem (2) satisfies the first-order necessary conditions of problem (2). Moreover, we show that any accumulation point is a local minimizer of problem (2).

THEOREM 4.4   *Assume that $\{\varepsilon_k\} \to 0$. Let $\{(x^k, y^k)\}$ be the sequence generated by the Penalty PALM method for problem (2). Then the following statements hold:*
(i) The sequence $\{(x^k, y^k)\}$ is bounded.
(ii) Suppose (x^*, y^*) is an accumulation point of $\{(x^k, y^k)\}$. Then $x^* = y^*$ is a feasible point of problem (2).
(iii) Let (x^*, y^*) be defined as above. Assume that $\{(x^k, y^k)\}_{k \in J} \to (x^*, y^*)$ for some subsequence J. Let $I^* = \{i \in \{1, \ldots, n\} : x_i^* \neq 0\}$, $\bar{I}^* = \{1, \cdots, n\} \setminus I^*$. Suppose that the MFCQ condition (3) holds at x^* for such I^* and $\bar{I}^*$. Then x^* satisfies the first-order optimality conditions (4). Further, x^* also is a local minimizer of (2).

Proof. The proofs of statement (i), (ii) and the first conclusion of statement (iii) are similar to that of Theorem 4.3. It follows from Theorem 2.4 and the first conclusion of statement (iii) that x^* is a local minimizer of (2). Statement (iii) holds. ∎

5.   Numerical experiments

In this section, we test the performance of the penalty PALM method for solving problems (1) and (2). The main purpose of our numerical experiments is to test the capability of the penalty PALM method for finding good quality suboptimal solutions and the computational efficiency of the method. We also compared our method with the PD method and the standard CPLEX(12.6) solver. Our method and PD method were coded in Matlab (R2014b), the penalty subproblems of the PD method were solved by Matlab optimization software package OptiToolbox, and all computations were performed on a PC (Intel core i7-4790 CPU, 3.6GHz, 16GB RAM). For constructing large-scale test problems, the data sets used in our experiments were selected from the index tracking problem data in ORlibrary [31]. We chose the data of the constituent stocks of Nikkei index (Japan), Standard and Poor's 500 (USA), Russell 2000 (USA) and Russell 3000 (USA) to structure four groups of test problems with the variable dimension $n = 225$, 457, 1319 and 2152. In each group, we used the weekly return data between January 1979 and December 2009 to estimate the mean and covariance matrixes. For convenience, the minimum profit target r in each instance was chosen to be the average of profits of all assets in each group.

For problem (1), the upper bounds of the numbers of selected assets K are chosen as 5, 10, 15, 20, 30 and 40, respectively. We chose $\rho_0 = 0.1$, $\sigma = 1.1$, $\gamma_1 = 0.51$, $\gamma_2 = 0.51$ for the penalty PALM method, L_1 and L_2 were chosen by using the way in Remark 1. Although there is no proved convergence when $\gamma_1, \gamma_2 < 1$, these two parameters can help the method find the support set of the solution in few iteration for problem (1).

In addition, we used (21) and (22) as the inner and outer termination criterias for the penalty PALM method with $\varepsilon_I = 10^{-5}$, $\varepsilon_O = 10^{-6}$, respectively. For the PD method, we chose the initial penalty parameter $\rho_0 = 0.1$, the increasing rate of the penalty parameter $\sigma = \sqrt{10}$, and used the terminal conditions in [26] with $\varepsilon_I = 10^{-3}$, $\varepsilon_0 = 10^{-4}$, respectively. For the CPLEX solver, we used the default parameters of the software.

Table 1. Comparison of the methods of penalty PALM and PD for (1).

n	K	PPALM		PD		Imp
		obj	time	obj	time	
225	5	1.7842e-04	0.3745	2.3530e-04	148.4708	0.2002(7)
	10	1.5240e-04	0.5303	3.0953e-04	98.8947	0.5835(10)
	15	1.5232e-04	0.4236	3.0313e-04	73.0125	0.5746(10)
	20	1.5232e-04	0.4624	3.2824e-04	64.7721	0.5891(10)
	30	1.5232e-04	0.4266	3.0961e-04	44.9497	0.5614(10)
	40	1.5232e-04	0.4619	3.1291e-04	38.5848	0.5504(10)
457	5	2.6266e-04	0.2373	3.0176e-04	866.1126	0.1722(5)
	10	1.8599e-04	0.3190	2.4378e-04	591.6283	0.3806(9)
	15	1.3284e-04	0.4689	2.7324e-04	538.1945	0.4645(10)
	20	9.8891e-05	0.8257	2.5183e-04	501.9836	0.6339(10)
	30	9.1567e-05	0.7547	2.8593e-04	436.2076	0.7253(10)
	40	8.6726e-05	0.7446	2.9410e-04	387.5863	0.7384(10)
1319	5	1.8077e-04	3.4930	3.8429e-04	343.6675	0.2692(7)
	10	1.6438e-04	3.7824	2.4849e-04	333.7898	0.3040(10)
	15	1.1801e-04	5.4872	2.2851e-04	350.4437	0.5125(10)
	20	9.2538e-05	5.3349	2.4690e-04	335.1443	0.5907(10)
	30	4.2876e-05	6.0709	2.4782e-04	338.5734	0.7412(10)
	40	3.6967e-05	5.6627	1.9983e-04	344.6562	0.7846(10)
2152	5	1.8146e-04	12.2082	1.8909e-04	3049.7	0.1049(8)
	10	1.7579e-04	11.4686	3.0916e-04	1254.6	0.3314(10)
	15	1.3741e-04	9.0511	2.1240e-04	744.9663	0.4189(10)
	20	9.3411e-05	10.8509	1.7906e-04	543.8227	0.5359(10)
	30	4.5588e-05	12.6820	1.8286e-04	394.1302	0.6887(10)
	40	4.2670e-05	9.4432	1.7271e-04	331.7393	0.7706(10)

For problem (2), we chose six different values of ν, namely, $\nu_{\max}$, $0.5\nu_{\max}$, $0.25\nu_{\max}$, $0.1\nu_{\max}$, $0.01\nu_{\max}$, and $0.001\nu_{\max}$ where $\nu_{\max} = 0.5$. For the penalty PALM method, we chose $\rho_0 = 100$, $\sigma = 1.1$, $\gamma_1 = 1.01$, $\gamma_2 = 1.01$, other parameters were same to the parameters of problem (1). For the PD method, we chose $\rho_0 = 100$, $\sigma = 2$, other parameters were same to the parameters of problem (1).

Table 1 summarizes the numerical results of the penalty PALM and the PD method for problem (1). In each test problem of Table 1, we randomly generated ten initial points and solved the test problem by using the penalty PALM method and PD method ten times to find the best approximate solution. It is need to note that the PD method needs two feasible initial points, and the penalty PALM method just needs two arbitrary initial points. We chose the four initial points of the PD method and penalty PALM method to be the same one point. The method of random generating the initial points was that randomly choose K assets whose expected returns are no less than r, set the components of the initial point corresponding to these assets to $1/K$ and set other components to zero. If the number, m, of assets whose expected returns are no less than r is less than K, we just chose these m assets and set the components of the initial point corresponding to these assets to $1/m$ and set other components to zero. For measuring the quality of the suboptimal solution x_p^* found by the penalty PALM method and the suboptimal solution x_d^* found by the PD method in each time, we used the following relative improvement of

the objective function value of the solution x_p^* over the solution x_d^*:

$$\text{imp} = \frac{f_{\text{obj}}(x_d^*) - f_{\text{obj}}(x_p^*)}{|f_{\text{obj}}(x_d^*)|} \tag{37}$$

where $f_{\text{obj}}(x) = (1/2)x^\top Q x$.

Table 2 summarizes the numerical results of the penalty PALM method and the standard CPLEX(12.6) solver for problem (1). It shall be mentioned that for using the CPLEX solver we need to reformulate problem (1) to a standard mixed-integer quadratic programming (MIQP) by introducing 0-1 variables. In order to compare with the CPLEX solver, in each test problem, we solved the MIQP program once by using the CPLEX solver and randomly generated ten groups different initial points and solved the test problem by using the penalty PALM method ten times. Then, we compared the solution of the CPLEX solver and the best result of the penalty PALM method.

Table 2. Comparison of the penalty PALM method and C-PLEX solver for (1).

n	K	PPALM		Cplex	
		obj	time	obj	time
225	5	1.6833e-04	0.0059	1.5868e-04	0.7879
	10	1.5281e-04	0.0063	1.5240e-04	0.4326
	15	1.5232e-04	0.0066	1.5232e-04	0.3453
	20	1.5232e-04	0.0068	1.5232e-04	0.3760
	30	1.5232e-04	0.0081	1.5232e-04	0.3865
	40	1.5232e-04	0.0104	1.5232e-04	0.3648
457	5	1.7264e-04	0.1935	–	–
	10	1.4723e-04	0.1677	–	–
	15	1.2052e-04	0.1247	–	–
	20	9.3158e-05	0.1280	–	–
	30	8.7316e-05	0.0419	8.6804e-05	6.6837
	40	8.6581e-05	0.0286	8.6703e-05	2.1280
1319	5	1.3093e-04	1.0021	–	–
	10	7.1708e-05	0.4032	–	–
	15	5.9326e-05	0.4879	–	–
	20	4.6440e-05	0.1150	–	–
	30	4.1075e-05	0.2779	–	–
	40	3.6171e-05	0.2880	–	–
2152	5	7.5337e-05	2.9188	–	–
	10	6.7408e-05	3.6407	–	–
	15	5.4683e-05	1.1371	–	–
	20	4.8687e-05	2.0463	–	–
	30	3.9030e-05	2.7230	–	–
	40	3.5940e-05	1.6472	–	–

Table 3 summarizes the numerical results of the penalty PALM and the PD method for problem (2). As in Table 1, for each test problem we randomly generated ten initial points by the method in Table 1 and solved the test problem by using the penalty PALM method and the PD method to find the best approximate solution. We also used the relative improvement in (37) with $f_{\text{obj}}(x) = (1/2)x^\top Q x + \nu\|x\|_0$ to measure the quality of the suboptimal solutions found by two methods.

In Table 1, 2 and 3, "PPALM" stands for the penalty PALM method; "PD" stands for the PD method; "Cplex" stands for the standard CPLEX solver; "time" (in seconds) and "obj" are the average computation time and the best objective value, respectively; "Imp" stands for the average relative improvement of the objective function value of the penalty PALM method to the PD method in ten times tests, and the integer in

the brackets stands for the number of times the penalty PALM method finds a better suboptimal solution such that imp > 0; "-" stands that the CPLEX solver does not find a optimal solution in one hour.

Table 3. Comparison of the methods of penalty PALM and PD for (1).

n	$\lambda/\lambda_{\max}$	PPALM obj	PPALM time	PD obj	PD time	Imp
225	1	2.503	0.0374	112.5	0.8583	0.9644(10)
	0.5	1.2503	0.0086	56.25	0.4649	0.9733(10)
	0.25	0.5003	0.0085	28.13	0.4275	0.9730(10)
	0.1	0.4503	0.0069	11.12	0.3871	0.9511(10)
	0.01	0.0502	0.0063	0.2004	0.0653	0.8742(10)
	0.001	0.0027	0.0062	0.0204	0.0707	0.7648(10)
457	1	5.0002	0.0499	228.5	2.8382	0.9803(10)
	0.5	2.7502	0.0171	114.3	2.3163	0.9759(10)
	0.25	1.5002	0.0154	57.16	2.2105	0.9694(10)
	0.1	0.7002	0.0115	22.85	1.9113	0.9672(10)
	0.01	0.0902	0.0136	0.2004	0.2511	0.4502(10)
	0.001	0.0072	0.0136	0.0204	0.2497	0.5505(10)
1319	1	6.0001	0.7131	659.5	34.5151	0.9924(10)
	0.5	3.0001	0.3756	329.6	32.3718	0.9894(10)
	0.25	2.0001	0.3414	164.9	30.6479	0.9901(10)
	0.1	0.9001	0.3353	65.95	27.5679	0.9864(10)
	0.01	0.1251	0.2583	0.2004	3.1709	0.3856(10)
	0.001	0.0151	0.2486	0.0203	3.1520	0.3539(10)
2152	1	5.5001	2.3827	1076	170.1137	0.9949(10)
	0.5	3.5001	2.9617	538.3	156.7696	0.9935(10)
	0.25	1.8751	1.6782	269.1	150.8373	0.9930(10)
	0.1	0.9501	1.3986	107.6	141.9196	0.9912(10)
	0.01	0.1251	0.9554	0.2003	16.1478	0.5758(10)
	0.001	0.013	0.9291	0.0203	15.9728	0.5735(10)

We can see from Table 1 and 3 that the solutions obtained from the penalty PALM method is often better than the solutions from the PD method and the computation time of the penalty PALM method is less than that of the PD method especially for the large-scale problems. In order to compare with the PD method we choose the two initial points of the penalty PALM method to be same, in generally if we choose two different initial points the computation time of the penalty PALM method will be reduced.

We can see from Table 2 that for the first two groups of the test problems, the best objective value of the solutions obtained from the penalty PALM method is approximate to the objective value of the solution obtained from the CPLEX solver, which can be regarded as a global minimum of the test problem. The computation speed of the CPLEX solver is much quicker than Matlab, but the computation time of the penalty PALM method for the first two groups of the test problems is less than the time of the CPLEX solver. It can suggest that the penalty PALM method is a efficient computing method for problem (1).

6.   Conclusion

In this paper we propose a penalty PALM method for the sparse portfolio selection problems (1) and (2). In solving process, the penalty PALM method generates a sequence of penalty subproblems each of which is a nonconvex and nonsmooth problem, and utilizes the PALM method framework and the sparse projection techniques to solve these penalty subproblems. Based on the characteristic of the problem, we can establish

that any accumulation point of the sequence generated by the penalty PALM method for problems (1) and (2) satisfies the first-order optimality conditions of problems (1) and (2). Furthermore, under a suitable assumption, it is also a local minimizer of problems (1) and (2). The utilizing of the PALM method framework and the sparse projection can exploit the structure of problems (1) and (2) and reduce the computation complexity. The computational results demonstrate that for problems (1) and (2) our method generally has better solution quality and needs lesser computation time than the PD method especially for the large-scale problems. We shall remark that our method can be extended to solve the long-short cardinality constrained portfolio selection problems. The difference is that we need to use the different sparse projections to solve the subproblems in the PALM method framework.

7. Acknowledgements

References

[1] H. Markowitz, *Portfolio selection*, The journal of finance 7(1), 77-91 (1952).

[2] D. Di Lorenzo, G. Liuzzi, F. Rinaldi, F. Schoen and M.Sciandrone *A concave optimization-based approach for sparse portfolio selection*, Optimization Methods and Software 27(6), 983-1000 (2012).

[3] V. DeMiguel, L. Garlappi, F.J. Nogales and R. Uppal, *A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms*, Management Science 55(5), 798-812 (2009).

[4] J. Brodie, I. Daubechies, C. De Mol, D. Giannone and I. Loris, *Sparse and stable Markowitz portfolios*, Proceedings of the National Academy of Sciences 106(30), 12267-12272 (2009).

[5] J. Fan, J. Zhang and K. Yu, *Asset allocation and risk assessment with gross exposure constraints for vast portfolios* , Available at SSRN 1307423 (2008).

[6] Y.-M. Yen, *A note on sparse minimum variance portfolios and coordinate-wise descent algorithms*, Available at SSRN 1604093 (2010).

[7] M. Carrasco and N. Noumon, *Optimal portfolio selection using regulaization*. Available at `http://www.cireq.umontreal.ca/activites/100514/carrasco.pdf`, (2011).

[8] S. Still and I. Kondor, *Regularizing portfolio optimization*, New Journal of Physics 12(7), 075034 (2010).

[9] M.W. Brandt, P. Santa-Clara and R. Valkanov *Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns*, Review of Financial Studies, hhp003 (2009).

[10] D. Bienstock, *Computational study of a family of mixed-integer quadratic programming problems*, Mathematical programming 74(2), 121-140 (1996).

[11] B. Blog, G. Van Der Hoek, A. Rinnooy Kan and G. Timmer, *The optimal selection of small portfolios*, Management Science 29(7), 792-798 (1983).

[12] T.-J. Chang, N. Meade, J.E. Beasley and Y.M. Sharaiha, *Heuristics for cardinality constrained portfolio optimisation*, Computers & Operations Research 27(13), 1271-1302 (2000).

[13] N.L. Jacob, *A limited-diversification portfolio selection model for the small investor*, The Journal of Finance 29(3), 847-856 (1974).

[14] D.X. Shaw, S. Liu and L. Kopman, *Lagrangian relaxation procedure for cardinality-constrained portfolio optimization*, Optimisation Methods & Software 23(3), 411-420 (2008).

[15] M. Woodside-Oriakhi, C. Lucas and J.E. Beasley, *Heuristic algorithms for the cardinality constrained efficient frontier*, European Journal of Operational Research 213(3), 538-550 (2011).

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

[17] D. Li, X. Sun and J. Wang, *Optimal lot solution to cardinality constrained meanCvariance formulation for portfolio selection*, Mathematical Finance 16(1), 83-101 (2006).

[18] X. Cui, X. Zheng, S. Zhu and X. Sun, *Convex relaxations and MIQCQP reformulations for a class of cardinality-constrained portfolio selection problems*, Journal of Global Optimization 56(4), 1409-1423 (2013).

[19] A. Frangioni and C. Gentile, *Perspective cuts for a class of convex 0-1 mixed integer programs*, Mathematical Programming 106(2), 225-236 (2006).

[20] A. Frangioni and C. Gentile, *SDP diagonalizations and perspective cuts for a class of nonseparable MIQP*, Operations Research Letters 35(2), 181-185 (2007).

[21] W. Murray and H. Shek, *A local relaxation method for the cardinality constrained portfolio optimization problem*, Computational Optimization and Applications 53(3), 681-709 (2012).

[22] A.M. Bruckstein, D.L. Donoho and M. Elad, *From sparse solutions of systems of equations to sparse modeling of signals and images* , SIAM review 51(1), 34-81 (2009).

[23] C. Chen ,X. Li, C. Tolman, S. Wang and Y. Ye, *Sparse portfolio selection via quasi-Norm regularization* , arXiv preprint arXiv:1312.6350 (2013).

[24] N. Gulpinar, L.T.H. An and M. Moeini, *Robust investment strategies with discrete asset choice constraints using DC programming*, Optimization 59(1), 45-62 (2010).

[25] X. Zheng, X. Sun, D. Li and J. Sun, *Successive convex approximations to cardinality-constrained convex programs: a piecewise-linear DC approach*, Computational Optimization and Applications 59(1-2), 379-397 (2014).

[26] Z. Lu and Y. Zhang, *Sparse approximation via penalty decomposition methods*, SIAM Journal on Optimization 23(4), 2448-2478 (2013).

[27] J. Bolte, S. Sabach and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Mathematical Programming 146(1-2), 459-494 (2014).

[28] A.P. Ruszczyński, *Nonlinear optimization, vol. 13*, Princeton university press, (2006).

[29] R.T. Rockafellar and R. Wets, *Variational Analysis*, Grundlehren der Mathematis-chen Wissenschaften, 317, Springer, 1998.

[30] J. Bolte, A. Daniilidis and A. Lewis,: *The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM Journal on Optimization 17(4), 1205-1223 (2007).

[31] J.E. Beasley, *OR-Library: Distributing test problems by electronic mail*, Journal of the Operational Research Society, 41(11): 1069C1072, 1990.