Noname manuscript No.
(will be inserted by the editor)

An Augmented Lagrangian Proximal Alternating Method for Sparse Discrete Optimization Problems

Yue Teng · Li Yang · Bo Yu · Xiaoliang Song

Received: date / Accepted: date

Abstract In this paper, an augmented Lagrangian proximal alternating (ALPA) method is proposed for two class of large-scale sparse discrete constrained optimization problems in which a sequence of augmented Lagrangian subproblems are solved by utilizing proximal alternating linearized minimization framework and sparse projection techniques. Under the Mangasarian-Fromovitz and the basic constraint qualification , we show that any local minimizer is a Karush-Kuhn-Tuker (KKT) point of the problem. And under some suitable assumptions, any accumulation point of the sequence generated by the ALPA method is a KKT point or a local minimizer of the original 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 Discrete constrained optimization · l_0 minimization · Proximal alternating linearized minimization method · Augmented Lagrangian method · Sparse projection

The work was supported by the National Natural Science Foundation of China ()

Bo Yu · Yue Teng · Xiaoliang Song
School of Mathematical Sciences, Dalian University of Technology, Dalian, Liaoning 116024, P.R. China
E-mail: yubo@dlut.edu.cn

Li Yang
School of Science, Dalian University of Technology, Panjin, Liaoning 124221, P.R. China
E-mail: yangli96@dlut.edu.cn

Yue Teng
E-mail: tengyue86@mail.dlut.edu.cn

Xiaoliang Song
E-mail: songxiaoliang@mail.dlut.edu.cn

1 Introduction

In this paper, we consider the following two minimization problems:

$$\begin{aligned} \min_{x} \quad & f(x)+g(x) \\ \text{s.t.} \quad & c(x)=0, \\ & d(x)\le 0, \\ & x_j\in\mathcal{D}_j,\ j\in J, \\ & \|x_J\|_0\le K, \end{aligned} \tag{1}$$

and

$$\begin{aligned} \min_{x} \quad & f(x)+g(x)+\lambda\|x_J\|_0 \\ \text{s.t.} \quad & c(x)=0, \\ & d(x)\le 0, \\ & x_j\in\mathcal{D}_j,\ j\in J, \end{aligned} \tag{2}$$

with decision variable $x\in\mathrm{R}^n$, where $f:\mathrm{R}^n\to\mathrm{R}$, $c:\mathrm{R}^n\to\mathrm{R}^m$, $d:\mathrm{R}^n\to\mathrm{R}^p$, are continuously differentiable functions, $g:\mathrm{R}^n\to\mathrm{R}\cup\{+\infty\}$ is a proper and lower semi-continuous function, $\|x_J\|_0$ denotes the cardinality of the subvector x_J formed by the entries of x indexed by $J\subseteq\{1,\dots,n\}$ and for all $j\in J$, $\mathcal{D}_j$ is composed by the union of a finite number of closed intervals of R and a single point set $\{0\}$ as following:

$$\mathcal{D}_j=\left(\bigcup_{l=1}^{p_j}[a_{lj},b_{lj}]\right)\bigcup\{0\},\ j\in J. \tag{3}$$

Furthermore, $0<K<|J|$ is an integer number, $\lambda>0$ is a regular parameter, for all l and j, it has $-\infty<a_{lj}<b_{lj}<+\infty$ and $0\notin[a_{lj},b_{lj}]$, and for all $j\in J$, it has $[a_{ij},b_{ij}]\cap[a_{lj},b_{lj}]=\emptyset$, $i,l=1,\dots,p_j$.

In problems (1) and (2), function g can be a nonsmooth function possibly nonconvex or a indicator function of a feasible set of easy constraints, and we need these functions and constraints can be handled in proximal problems, furthermore g also can be the sum of the functions and the indicator functions described above. The feasible set of typical semi-continuous variable usually contains one closed interval. From the definition of $\mathcal{D}_j$, we can observe that for all $j\in J$, discrete variables x_j are the generalized semi-continuous variables, thus optimization problems with semi-continuous variables and l_0 minimization problems can be regarded as the special cases of problems (1) and (2).

In recent years the optimization problem with semi-continuous variables and the l_0 minimization problem have gained considerable attention. As the generalizations of these two problems, many problems can be put into the form of problems (1) and (2), such as the portfolio selection problems [1,2,3,4,5,7], the compressed sensing problems [8,10], the production planning and the facility location problems [11], and the unite commitment problems [12,

14]. Problems (1) and (2) are nonconvex and nonsmooth problems and have been proved to be NP-hard in generally [9]. The main known method for solving problems (1) and (2) is the integer programming method which transforms problems (1) and (2) to mixed integer programming problems and uses the branch-and-bound framework and the cutting plane technologies to solve these problems. In recent decades, several methods which include the branch-and-bound methods, the perspective reformulation methods, the relaxation methods, the heuristic methods and the local methods have been developed for solving special cases of problems (1) and (2).

An important application of problems (1) and (2) in financial optimization is the portfolio selection problem with cardinality and minimum threshold constraints. Bonami and Lejeune [2] proposed a branch-and-bound method based on continuous relaxation and special branching rules for solving an exact solution of a portfolio selection model under stochastic and integer constraints including cardinality and minimum threshold constraints. Bienstock [9] developed a branch-and-cut method for solving cardinality constrained quadratic programming problems by using a primal feasible method and a surrogate constraint approach. Based on a novel geometric approach, a branch-and-bound method is developed in Gao and Li [6] for solving cardinality constrained portfolio selection problems. A novel reformulation technique for optimization problems with semi-continuous variables called perspective reformulation and its tractable reformulations were proposed by [12,14,15,11,16]. These reformulation methods are based on the perspective function and the cone programming technologies and derive much more efficient mixed-integer program reformulations than the standard formulation. In addition, the perspective reformulation method also can be used to solve the optimization problems with semi-continuous variables and cardinality constraint. In recent decades, many efficient integer programming methods and perspective reformulation methods have been developed for optimization problems with semi-continuous variables or cardinality constraint. But for the discrete constrained optimization problems arising from the financial optimization and the operation research, it still lack efficient integer programming methods. The traditional heuristic methods for special cases of problems (1) and (2) generally base on genetic algorithm, tabu search and simulated annealing [17,18,19,20]. However, these heuristic methods do not guarantee to find the optimal or a suboptimal solution of the problems. And similarly with the integer programming methods, the heuristic methods also require a large amount of computation time for solving large-scale problems.

In recent years several relaxation methods and local methods have been developed for l_0 minimization problems. The l_1-norm approximation has been a popular method in compressed sensing and other sparse optimization problems [8]. However, the l_1-norm approximation does not often produce solutions with desired sparsity. Several nonconvex approximations to the l_0-norm have been studied recently. Gulpinar et al. proposed to reformulate a cardinality constrained optimization problem as a DC program and 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 was proposed in [21]. By utilizing a new NLP-reformulation, Burdakov et al. [22] presented a standard nonlinear programming with complementarity-type constraints reformulation and a regularization method. In recent years, a novel penalty decomposition (PD) method for solving a stationary point or a local optimal solution of a general l_0 minimization problem and a novel proximal alternating linearized minimization (PALM) method for general nonconvex and nonsmooth problems has been developed in [23] and [24], respectively. Inspired by these two work, a penalty PALM method has been developed in [25] for sparse portfolio selection problems. Moreover, by utilizing the frameworks of the splitting method and the PD method, Bai et al. [26] presented a splitting augmented Lagrangian method for optimization problems with a cardinality constraint and semi-continuous variables.

In this paper, base on [25], we derive the optimality conditions for sparse discrete constrained optimization problems (1) and (2) and develop an augmented Lagrangian proximal alternating (ALPA) method for solving large-scale problems (1) and (2), efficiently. And this method also can be used to solve the special cases of problems (1) and (2), for instance, the discrete constrained optimization problems and the l_0 minimization problems. For solving multiple constraints sparse discrete optimization problems, the ALPA method generates a sequence of augmented Lagrangian (Al) subproblems which are nonconvex and nonsmooth problems and utilizes the PALM framework and sparse projection techniques to solve these subproblems. Under some suitable assumptions, we establish that any accumulation point of the iterative sequence generated by the ALPA method satisfies the first-order optimality conditions of problems (1) and (2). In addition, from the property of the PALM method, any accumulation point of the sequence generated by the PALM framework is a critical point of the Al subproblem. Finally, we test the performance of the ALPA method by applying it to solve the sparse portfolio selection problems with discrete constraints which are generalizations of the sparse portfolio selection problems with minimum threshold constraints. The computational results demonstrate that the framework of our method is more easily in programming, 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 Subsection 1.1, we introduce the notation and terminology that is used throughout the paper. In Section 2, we establish the first-order optimality conditions for problems (1) and (2). In Section 3, we develop the ALPA method for the problems. The convergence result of the ALPA method is established in Section 4. In Section 5, we carry out numerical experiments to test the performance of our method for solving problems (1) and (2). Finally, we give some concluding remarks in Section 6.

1.1 Notation and terminology

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 $\mathcal{A} \setminus \mathcal{B} = \{x \in \mathcal{A} : x \notin \mathcal{B}\}$.

We recall from [27] that for a proper and lower semi-continuous function $h : \mathrm{R}^n \to (-\infty, +\infty]$ and a point $x \in \operatorname{dom} h$ where $\operatorname{dom} h = \{x \in \mathrm{R}^n : h(x) < +\infty\}$, the limiting and horizon subdifferential of h at x are defined respectively as

$$\partial h(x) = \{u \in \mathrm{R}^n : \exists\, x^k \xrightarrow{h} x,\ u^k \to u \text{ with } \liminf_{z \to x^k} \tfrac{h(z)-h(x^k)-\langle u^k, z-x^k\rangle}{\|z-x^k\|} \geq 0\ \forall k\},$$

$$\partial^\infty h(x) = \{u \in \mathrm{R}^n : \exists\, x^k \xrightarrow{h} x,\ t_k u^k \to u,\ t_k \downarrow 0 \text{ with } \liminf_{z \to x^k} \tfrac{h(z)-h(x^k)-\langle u^k, z-x^k\rangle}{\|z-x^k\|} \geq 0\ \forall k\},$$

where $x^k \xrightarrow{h} x$ means that $x^k \to x$ and $h(x^k) \to h(x)$ and $t_k \downarrow 0$ means $t_k > 0$ and $t_k \to 0$. If h is convex, the limiting subdifferential coincides with the classical subdifferential in convex analysis. Moreover, let h be a continuously differential function and g be a proper lower semi-continuous function with finite at x, we simply have that $\partial h(x) = \{\nabla h(x)\}$ and $\partial(h(x) + g(x)) = \nabla h(x) + \partial g(x)$. For a proper lower semi-continuous function h, a point x is a critical point of h if $0 \in \partial h(x)$. In addition, for the limiting subdifferential the Fermat's rule remains true, that is, if x is a local minimizer of a proper lower semi-continuous function h, then $0 \in \partial h(x)$. Recall from [27] that a set-valued mapping $\sigma : \mathrm{R}^n \rightrightarrows \mathrm{R}^m$ is said to be outer semi-continuous at $x \in \mathrm{R}^n$ if

$$\left\{u \in \mathrm{R}^m : \exists\, x^k \to x,\ u^k \to u,\ u^k \in \sigma(x^k)\right\} \subseteq \sigma(x).$$

Given a closed set $\mathcal{C} \subseteq \mathrm{R}^n$, let $I_{\mathcal{C}}(x)$ denote the indicator function of $\mathcal{C}$ that satisfies $I_{\mathcal{C}}(x) = 0$ if $x \in \mathcal{C}$ and $+\infty$ otherwise, and let $N_{\mathcal{C}}(x)$ denote the limiting normal cone of $\mathcal{C}$ at any $x \in \mathcal{C}$, then we have $\partial I_{\mathcal{C}}(x) = N_{\mathcal{C}}(x)$ at $\forall x \in \mathcal{C}$. It is well known that the limiting subdifferential mapping ∂h, the limiting normal cone mapping $N_{\mathcal{C}}$ and the horizon subdifferential mapping $\partial^\infty h$ are outer semi-continuous everywhere.

2 First-order optimality conditions

In this section, we study the constraint qualifications and the first-order optimality conditions for problems (1) and (2). From the definition of the discrete variable, if we restrict the discrete variable x_j, $j \in J$ to some constant interval or point of $\mathcal{D}_j$, problems (1) and (2) reduce to two tightened nonlinear

programming problems. By utilizing these tightened nonlinear programming problems, we give the constraint qualifications and establish the first-order necessary conditions for problems (1) and (2). Furthermore, if the l_0-norm and the discrete variables are the only nonconvex part of problems (1) and (2) we can establish the first-order sufficient conditions for problems (1) and (2).

Before establishing the first-order necessary conditions for problems (1) and (2), we need to consider the qualifications on the constraints of problems (1) and (2). The Mangasarian-Fromovitz constraint qualification (MFCQ) is a standard constraint qualification for nonliner programs. Next, we give the MFCQ for problems (1) and (2).

Definition 1 Let $\mathcal{F}_1$ be the feasible region of problem (1). For a given feasible point $x^* \in \mathcal{F}_1$, let $J^* \subseteq J$ be an index set with $|J^*| = K$ such that $x_j^* = 0$ for all $j \in \bar{J}^*$, where $\bar{J}^* = J \setminus J^*$. We say that the MFCQ holds at $x^* \in \mathcal{F}_1$ if the there exists a point $x \in \mathrm{R}^n$ such that

$$\begin{aligned} \langle \nabla c_i(x^*), x - x^* \rangle &= 0,\ i = 1, \ldots, m, \\ \langle \nabla d_i(x^*), x - x^* \rangle &< 0,\ i \in \mathcal{A}_d(x^*), \\ \langle -e_j, x - x^* \rangle &< 0,\ j \in \mathcal{A}(x^*), \\ \langle e_j, x - x^* \rangle &< 0,\ j \in \mathcal{B}(x^*), \\ \langle e_j, x - x^* \rangle &= 0,\ j \in \mathcal{Z}_{J^*}(x^*) \cup \bar{J}^*, \end{aligned} \tag{4}$$

and $\{\nabla c_i(x^*), e_j\}$, $i = 1, \ldots, m$, $j \in \mathcal{Z}_{J^*}(x^*) \cup \bar{J}^*$ are linearly independent where $\mathcal{A}(x^*) = \{j \in J^* : x_j^* = a_{l_j^* j}\}$, $\mathcal{B}(x^*) = \{j \in J^* : x_j^* = b_{l_j^* j}\}$, $\mathcal{Z}_{J^*}(x^*) = \{j \in J^* : x_j^* = 0\}$, l_j^* is the index of the interval of $\mathcal{D}_j$ which contains the x_j^*, $\mathcal{A}_d(x^*) = \{1 \le i \le p : d_i(x^*) = 0\}$ and e_j is a n-dimensional vector of which the the j-th component is equal to 1 and other components are equal to 0.

Definition 2 Let $\mathcal{F}_2$ be the feasible region of problem (2). For a given feasible point $x^* \in \mathcal{F}_2$, let $J^* = \{j \in J : x_j^* \neq 0\}$ and $\bar{J}^* = J \setminus J^*$. We say that the MFCQ condition holds at $x^* \in \mathcal{F}_2$ if the there exists a point $x \in \mathrm{R}^n$ such that

$$\begin{aligned} \langle \nabla c_i(x^*), x - x^* \rangle &= 0,\ i = 1, \ldots, m, \\ \langle \nabla d_i(x^*), x - x^* \rangle &< 0,\ i \in \mathcal{A}_d(x^*), \\ \langle -e_j, x - x^* \rangle &< 0,\ j \in \mathcal{A}(x^*), \\ \langle e_j, x - x^* \rangle &< 0,\ j \in \mathcal{B}(x^*), \\ \langle e_j, x - x^* \rangle &= 0,\ j \in \bar{J}^*, \end{aligned} \tag{5}$$

and $\{\nabla c_i(x^*), e_j\}$, $i = 1, \ldots, m$, $j \in \bar{J}^*$ are linearly independent where $\mathcal{A}(x^*)$, $\mathcal{B}(x^*)$, l_j^*, $\mathcal{A}_d(x^*)$ and e_j are defined as in Definition 1.

It is well known that the MFCQ is equivalent to the positively liner independence of the gradient vectors of the constraints. In addition, since the objective functions of problems (1) and (2) are lower semi-continuous, a local minimizer of problem (1) or (2) may not be a KKT point under only the MFCQ condition. Hence, like [28] (Definition 2.3), we use the following basic qualification (BQ) for general discussion.

Definition 3 Let $\mathcal{F}_1$, $\mathcal{F}_2$ be the feasible regions of problems (1) and (2), respectively. For problem (1), we say that the BQ holds at $x^* \in \mathcal{F}_1$ if

$$-\partial^\infty g(x^*) \cap N_{\mathcal{F}_1}(x^*) = 0. \tag{6}$$

Analogously, for problem (2), we say that the BQ holds at $x^* \in \mathcal{F}_2$ if

$$-(\partial^\infty g(x^*) + \partial^\infty \lambda \|x_J\|_0) \cap N_{\mathcal{F}_2}(x^*) = 0. \tag{7}$$

Under the constraint qualifications above, we establish the first-order necessary conditions for problems (1) and (2).

Theorem 1 *Assume that x^* is a local minimizer of problem (1). Let $J^* \subseteq J$ be an index set with $|J^*| = K$ such that $x_j^* = 0$ for all $j \in \bar{J}^*$, where $\bar{J}^* = J \setminus J^*$. Suppose that at x^* the BQ (6) and the MFCQ (4) condition hold. Then, there exists $\mu^* \in \mathrm{R}^m$, $\nu_1^* \in \mathrm{R}^p_+$, $\nu_2^*, \nu_3^* \in \mathrm{R}^n_+$ and $z^* \in \mathrm{R}^n$ together with x^* satisfying*

$$\begin{gathered} 0 \in \nabla f(x^*) + \partial g(x^*) + \nabla c(x^*)\mu^* + \nabla d(x^*)\nu_1^* - \nu_2^* + \nu_3^* + z, \\ (\nu_1^*)_i d_i(x^*) = 0,\ i = 1, \ldots, p \\ (\nu_2^*)_j (a_{l_j^* j} - x_j^*) = 0,\ (\nu_3^*)_j (x_j^* - b_{l_j^* j}) = 0,\ j \in J^* \setminus \mathcal{Z}_{J^*}(x^*), \\ (\nu_2^*)_j = 0,\ (\nu_3^*)_j = 0,\ j \in \bar{J}^* \cup \mathcal{Z}_{J^*}(x^*) \cup \bar{J}, \\ z_j^* = 0,\ j \in (J^* \setminus \mathcal{Z}_{J^*}(x^*)) \cup \bar{J}, \end{gathered} \tag{8}$$

where $\bar{J} = \{1, \ldots, n\} \setminus J$ and $\mathcal{Z}_{J^*}(x^*)$, l_j^* are defined as in Definition 1.

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

$$\begin{aligned} \min_x \quad & f(x) + g(x) \\ \text{s.t.} \quad & c(x) = 0, \\ & d(x) \le 0, \\ & a_{l_j^* j} \le x_j \le b_{l_j^* j},\ j \in J^* \setminus \mathcal{Z}_{J^*}(x^*), \\ & x_j = 0,\ j \in \bar{J}^* \cup \mathcal{Z}_{J^*}(x^*). \end{aligned} \tag{9}$$

Together with the BQ (6) and the MFCQ (4) condition and the corollaries of [27], the conclusion holds.

Theorem 2 *Assume that x^* is a local minimizer of problem (2). Let $J^* = \{j \in J : x_j^* \neq 0\}$ and $\bar{J}^* = J \setminus J^*$. Suppose that at x^* the BQ (7) and the MFCQ (5) condition hold. Then, there exists $\mu^* \in \mathrm{R}^m$, $\nu_1^* \in \mathrm{R}^p_+$, $\nu_2^*, \nu_3^* \in \mathrm{R}^n_+$ and $z^* \in \mathrm{R}^n$ together with x^* satisfying*

$$\begin{gathered} 0 \in \nabla f(x^*) + \partial g(x^*) + \nabla c(x^*)\mu^* + \nabla d(x^*)\nu_1^* - \nu_2^* + \nu_3^* + z, \\ (\nu_1^*)_i d_i(x^*) = 0,\ i = 1, \ldots, p \\ (\nu_2^*)_j (a_{l_j^* j} - x_j^*) = 0,\ (\nu_3^*)_j (x_j^* - b_{l_j^* j}) = 0,\ j \in J^*, \\ (\nu_2^*)_j = 0,\ (\nu_3^*)_j = 0,\ j \in \bar{J}^* \cup \bar{J};\ z_j^* = 0,\ j \in J^* \cup \bar{J}, \end{gathered} \tag{10}$$

where $\bar{J} = \{1, \ldots, n\} \setminus J$ and l_j^* are defined as in Definition 1.

Proof It is obvious that x^* is a local minimizer of problem (2) if and only if x^* is a local minimizer of the tightened problem:

$$\begin{aligned} \min_{x} \quad & f(x)+g(x) \\ \text{s.t.} \quad & c(x)=0, \\ & d(x)\le 0, \\ & a_{l_j^* j}\le x_j\le b_{l_j^* j},\ j\in J^*, \\ & x_j=0,\ j\in \bar{J}^*. \end{aligned} \tag{11}$$

Together with the BQ (7) and the MFCQ (5) condition and the corollaries of [27], the conclusion holds.

It is obvious that, for fixed x^*, J^* and $\bar{J}^*$, the MFCQ condition and the first-order necessary conditions of problems (1) and (2) described above are the MFCQ condition and the first-order necessary conditions of the tightened problems (9) and (11). Furthermore, the MFCQ condition and the first-order necessary conditions defined above are the generalizations of the CC-MFCQ condition and the M-stationarity conditions in [22], respectively.

If the l_0-norm and the discrete constraints are the only nonconvex parts of problems (1) and (2), the tightened problems (9) and (11) for x^* is convex. Using this observation and the conclusion in [23] we can establish the first-order sufficient optimality conditions for problems (1) and (2).

Theorem 3 *Assume that f, g and d's are convex functions, and c's are affine functions. Let x^* be a feasible point of problem (1), $J^*\subseteq J$ be an index set with $|J^*|=K$ such that $x_j^*=0$ for all $j\in\bar{J}^*$, where $\bar{J}^*=J\setminus J^*$. Suppose that for such J^*, there exists some $\mu^*\in \mathrm{R}^m$, $\nu_1^*\in \mathrm{R}^p_+$, $\nu_2^*,\nu_3^*\in \mathrm{R}^n_+$ and $z^*\in \mathrm{R}^n$ satisfying (8). Then, x^* is a local minimizer of problem (1).*

Proof : By the assumptions and the corollaries of [27], for constant J^*, x^* is a minimizer of the tightened problem (9). Then there exists $\varepsilon_1>0$ such that $f(x)+g(x)\ge f(x^*)+g(x^*)$ for all $x\in\mathcal{O}_{J^*}(x^*;\varepsilon_1)$ where

$$\begin{aligned} \mathcal{O}_{J^*}(x^*;\varepsilon_1)=\{&x\in \mathrm{R}^n : c(x)=0,\ d(x)\le 0, \\ & a_{l_j^* j}\le x_j\le b_{l_j^* j},\ j\in J^*\setminus \mathcal{Z}_{J^*}(x^*),\ x_{\bar{J}^*\cup\mathcal{Z}_{J^*}(x^*)}=0,\ \|x-x^*\|<\varepsilon_1\} \end{aligned}$$

with $\bar{J}^*=J\setminus J^*$ and $\mathcal{Z}_{J^*}(x^*)=\{j\in J^* : x_j^*=0\}$. Let $\mathcal{J}_1^*=\{J_1^*\subseteq J : |J_1^*|=K,\ x_j^*=0,\ \forall j\in J\setminus J_1^*\}$, then we can observe from (1) that there exists ε_2 such that for any $x\in\mathcal{O}(x^*;\varepsilon_2)$, where

$$\begin{aligned} \mathcal{O}(x^*;\varepsilon_2)=\{&x\in \mathrm{R}^n : c(x)=0,\ d(x)\le 0, \\ & x_j\in\mathcal{D}_j,\ j\in J,\ \|x_J\|_0\le K,\ \|x-x^*\|<\varepsilon_2\}, \end{aligned}$$

there exists $J_1^*\in\mathcal{J}^*$ satisfying $x\in\mathcal{O}_{J_1^*}(x^*;\varepsilon_2)$. Let $\bar{\varepsilon}=\min\{\varepsilon_1,\varepsilon_2\}$, from the definition of $\mathcal{O}_{J^*}(x^*;\bar{\varepsilon})$ and $\mathcal{O}_{J_1^*}(x^*;\bar{\varepsilon})$ we have that $\mathcal{O}_{J^*}(x^*;\bar{\varepsilon})=\mathcal{O}_{J_1^*}(x^*;\bar{\varepsilon})$. Hence, for any $x\in\mathcal{O}(x^*;\bar{\varepsilon})$, it has $f(x)+g(x)\ge f(x^*)+g(x^*)$. It then implies that x^* is a local minimizer of (1).

Theorem 4 *Assume that f, g and d's are convex functions, and c's are affine functions. Let x^* be a feasible point of problem (2), and let $J^* = \{j \in J : x_j^* \neq 0\}$, $\bar{J}^* = J \setminus J^*$. Suppose that for such J^*, there exists some $\mu^* \in \mathrm{R}^m$, $\nu_1^* \in \mathrm{R}_+^p$, $\nu_2^*, \nu_3^* \in \mathrm{R}_+^n$ and $z^* \in \mathrm{R}^n$ satisfies (10). Then, x^* is a local minimizer of problem (2).*

Proof : By the above assumptions and the corollaries of [27], x^* is a minimizer of the tightened problem (11) with $\bar{J}^*$. We can observe that any point is a local minimizer of problem (2) if and only if it is a minimizer of the tightened problem (11). Then, it implies that x^* is a local minimizer of problem (2).

From the conclusions of (3) and (4), we observe that if the l_0-norm and the discrete constraints are the only nonconvex parts of problems (1) and (2), then for any feasible point x of problem (1) or (2), which satisfies the first-order necessary conditions of problem (1) or (2), x is a local minimizer of problem (1) or (2).

3 Augmented Lagrangian Proximal Alternating Method

In this section, we propose an ALPA method for solving problems (1) and (2). And we introduce the projection techniques and the closed-form solutions of two types of special l_0 minimization problems which are utilized in solving the augmented Lagrangian (AL) subproblems in the ALPA Method.

3.1 Augmented Lagrangian proximal alternating method for problem (1)

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

$$\begin{aligned} \min_{x,y} \quad & f(x) + g(x) \\ \text{s.t.} \quad & c(x) = 0, \\ & d(x) \leq 0, \\ & x_J = y, \\ & y_i \in \mathcal{D}_{J(i)}, \ i = 1, \ldots, |J|, \\ & \|y\|_0 \leq K, \end{aligned} \tag{12}$$

where $x \in \mathrm{R}^n$, $y \in \mathrm{R}^{|J|}$. Let $\mathcal{E}_1 = \left\{y \in \mathrm{R}^{|J|} : y_i \in \mathcal{D}_{J(i)},\ i = 1, \ldots, |J|,\ \|y\|_0 \leq K\right\}$ denote the feasible set of the cardinality constraint and the discrete constraints.

For any given penalty parameter $\rho > 0$, we define the augmented Lagrangian (AL) function for (1) (or equivalently (12)) as following

$$\begin{aligned} \mathcal{L}_1(x, y, \mu, \nu, \eta, \rho) = f(x) + g(x) + \tfrac{1}{2\rho}(\|\mu + \rho c(x)\|^2 - \|\mu\|^2) \\ + \tfrac{1}{2\rho}(\|[\nu + \rho d(x)]_+\|^2 - \|\nu\|^2) + \tfrac{1}{2\rho}(\|\eta + \rho(x_J - y)\|^2 - \|\eta\|^2). \end{aligned}$$

Similarly with [28,29], at each outer iteration we approximately solve problem (1) by solving the following AL subproblem:

$$\begin{aligned} \min_{x,y} \quad & \mathcal{L}_1(x, y, \mu, \nu, \eta, \rho) \\ \text{s.t.} \quad & y \in \mathcal{E}_1, \end{aligned} \tag{13}$$

For convenience, we write the AL function as a sum of a smooth term and a nonsmooth term: $\mathcal{L}_1(x, y, \mu, \nu, \eta, \rho) = \varphi(x, y, \mu, \nu, \eta, \rho) + g(x)$ where

$$\begin{aligned} \varphi(x, y, \mu, \nu, \eta, \rho) := f(x) + \tfrac{1}{2\rho}(\|\mu + \rho c(x)\|^2 - \|\mu\|^2) \\ + \tfrac{1}{2\rho}(\|[\nu + \rho d(x)]_+\|^2 - \|\nu\|^2) + \tfrac{1}{2\rho}(\|\eta + \rho(x_J - y)\|^2 - \|\eta\|^2). \end{aligned} \tag{14}$$

The minimization problem (13) can be rewritten as the following unconstrained minimization problem:

$$\begin{aligned} \min_{x,y} \Psi_1(x, y, \mu, \nu, \eta, \rho) = g(x) + \varphi(x, y, \mu, \nu, \eta, \rho) + h_1(y) \\ \text{over all } (x, y) \in \mathrm{R}^n \times \mathrm{R}^{|J|}, \end{aligned} \tag{15}$$

where $h_1(y) = I_{\mathcal{E}_1}(y)$ and $I_{\mathcal{E}_1}(y)$ denotes the indicator function of $\mathcal{E}_1$. We can observe that problem (15) is a nonconvex and nonsmooth minimization problem. From Fermat's rule, if a point x is a local minimizer of a proper and lower semi-continuous function σ, then x is a critical point of σ.

Recently, Bolte et al. developed a novel PALM method [24] 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 [24] that for any proper and lower semi-continuous functions $g(x)$, $h(y)$ and smooth function $\varphi(x, y)$ in the form of

$$\min_{x,y} \Psi(x, y) = g(x) + \varphi(x, y) + h(y), \text{ over all } (x, y) \in \mathrm{R}^n \times \mathrm{R}^m \tag{16}$$

if $g(x)$, $h(x)$ and $\varphi(x, y)$ satisfy Assumption A and B in [24] and $\Psi(x, y)$ is a KL function then each bounded sequence generated by the PALM method globally converges to a critical point of problem (16). If a proper and lower semi-continuous function h satisfies the KL property at each point of $\operatorname{dom} \partial h$ then h is called a KL function. The KL property plays a central role in the convergence analysis of the PALM method. We can see [24] and [30,31] for the details of the definition of the KL property and more important results on the KL theory.

Hence, for ensuring the convergence of the PALM method for problem (15) we take the following as our assumption.

Assumption 1 *(i)* $\inf_{x \in \mathrm{R}^n} g(x) > -\infty$, *and for any constant* μ, ν, η *and* ρ *it has* $\inf_{(x,y) \in \mathrm{R}^n \times \mathrm{R}^{|J|}} \Psi_1(x, y, \mu, \nu, \eta, \rho) > -\infty$.

(ii) For any fixed y *and constant* μ, ν, η *and* ρ *the function* $x \to \varphi(x, y, \mu, \nu, \eta, \rho)$ *is* $C^{1,1}_{L_1(y)}$, *namely the partial gradient* $\nabla_x \varphi(x, y, \mu, \nu, \eta, \rho)$ *is globally Lipschitz with moduli* $L_1(y)$, *that is*

$$\|\nabla_x \varphi(x_1, y, \mu, \nu, \eta, \rho) - \nabla_x \varphi(x_2, y, \mu, \nu, \eta, \rho)\| \le L_1(y)\|x_1 - x_2\|, \ \forall x_1, x_2 \in \mathrm{R}^n.$$

(iii) There exists $L_1^-, L_1^+ > 0$ *such that*

$$\inf\{L_1(y^l) : l \in \mathrm{N}\} \geq L_1^- \text{ and } \sup\{L_1(y^l) : l \in \mathrm{N}\} \leq L_1^+,$$

where $\{y^l\}$ *is a sequence of variable* y *generated by the PALM method.*

(iv) For constant μ, ν, η *and* ρ, $\nabla\varphi(x, y, \mu, \nu, \eta, \rho)$ *is Lipschitz continuous on bounded subsets of* $\mathrm{R}^n \times \mathrm{R}^{|J|}$, *that is, for each bounded subsets* $B_1 \times B_2$ *of* $\mathrm{R}^n \times \mathrm{R}^{|J|}$ *there exists* $M > 0$ *such that for all* $(x_i, y_i) \in B_1 \times B_2$, $i = 1, 2$*:*

$$\begin{aligned}&\|(\nabla_x\varphi(x_1, y_1, \mu, \nu, \eta, \rho) - \nabla_x\varphi(x_2, y_2, \mu, \nu, \eta, \rho)\\ &\quad, \nabla_y\varphi(x_1, y_1, \mu, \nu, \eta, \rho) - \nabla_y\varphi(x_2, y_2, \mu, \nu, \eta, \rho))\| \leq M\|(x_1 - x_2, y_1 - y_2)\|.\end{aligned}$$

(v) For constant μ, ν, η *and* ρ, *function* $\Psi_1(x, y, \mu, \nu, \eta, \rho)$ *is a KL function and the level set of* $\Psi_1(x, y, \mu, \nu, \eta, \rho)$ *is bounded.*

Remark 1 We can obvious that, for any fixed x and constant μ, ν, η and ρ the function $y \to \varphi(x, y, \mu, \nu, \eta, \rho)$ is $C^{1,1}_{L_2}$, where $L_2 = \rho$. Function $h_1(y)$ is the indicator function of $\mathcal{E}_1$ then we have $h_1(y)$ is lower semi-continuous and $\inf_{y\in\mathrm{R}^{|J|}} h_1(y) > -\infty$. Hence functions $g(x)$, $\varphi(x, y, \mu, \nu, \eta, \rho)$ and $h_1(y)$ satisfy Assumption A and B in [24]. In addition, in the assumption (v) the condition of the level set of function $\Psi_1(x, y, \mu, \nu, \eta, \rho)$ is bounded can be replaced by the condition that the sequence generated by the PALM method is bounded or some other conditions which can guarantee the boundedness of the sequence generated by the PALM method.

Next, we use the framework of the PALM method to solve problem (15) as well as the AL subproblem (13). For constant μ, ν, η and ρ, the PALM method updates the iterative sequence $\{(x^l, y^l)\}$ via solving the following two proximal problems:

$$x^{l+1} \in \arg\min_x\ g(x) + \langle x - x^l, \nabla_x\varphi(x^l, y^l, \mu, \nu, \eta, \rho)\rangle + \frac{t_l^1}{2}\|x - x^l\|^2, \quad (17)$$

$$y^{l+1} \in \arg\min_y\ h_1(y) + \langle y - y^l, \nabla_y\varphi(x^{l+1}, y^l, \mu, \nu, \eta, \rho)\rangle + \frac{t_l^2}{2}\|y - y^l\|^2, (18)$$

where $t_l^1 = \gamma_1 L_1(y^l)$, $t_l^2 = \gamma_2 L_2$, $\gamma_1, \gamma_2 > 1$ are two appropriately chosen step sizes, $L_1(y^l)$, L_2 are the lipschitz constants in Assumption 1 and Remark 1. Using the proximal map notation in [27], we get the minimization problems (17), (18) are equivalent to the following two proximal problems:

$$x^{l+1} \in \mathrm{Prox}^g_{t_l^1}(x^l - \frac{1}{t_l^1}\nabla_x\varphi(x^l, y^l, \mu, \nu, \eta, \rho)), \quad (19)$$

$$y^{l+1} \in \mathrm{Prox}^{h_1}_{t_l^2}(y^l - \frac{1}{t_l^2}\nabla_y\varphi(x^{l+1}, y^l, \mu, \nu, \eta, \rho)). \quad (20)$$

Here, we assume that proximal problem (19) can be handled easily. For convenience of presentation, set $w_1 = x^l - (1/t_l^1)\nabla_x\varphi(x^l, y^l, \mu, \nu, \eta, \rho)$ and $w_2 = y^l - (1/t_l^2)\nabla_y\varphi(x^{l+1}, y^l, \mu, \nu, \eta, \rho)$. Note that $h_1(y) = I_{\mathcal{E}_1}(y)$, it is a indicator

function of set $\mathcal{E}_1$. The proximal problem (20) reduce to a projection problem onto $\mathcal{E}_1$, defined by

$$y^{l+1} \in P_{\mathcal{E}_1}(w_2) := \arg\min_y \{\|y - w_2\| : y \in \mathcal{E}_1\}.$$

Set $\mathcal{E}_1$ is a nonempty closed set and $P_{\mathcal{E}_1}(w_2)$ is a projection onto a sparse and discrete feasible set. Hence, $P_{\mathcal{E}_1}$ defines in a general multi-valued map, for convenience, we also call $P_{\mathcal{E}_1}$ a sparse projection. Suppose $\bar{y}^* \in P_{\mathcal{E}_1}(w_2)$ then $\bar{y}^*$ can be computed as follows:

$$\bar{y}_i^* = \begin{cases} \bar{y}_i, & \text{if } i \in S^*, \\ 0, & \text{otherwise,} \end{cases} \qquad i = 1, \ldots, |J|,$$

where

$$\bar{y}_i \in \arg\min_{y_i} \left\{(y_i - (w_2)_i)^2 : y_i \in \mathcal{D}_{J(i)}\right\}, \ i = 1, \ldots, |J|, \tag{21}$$

and $S^* \subseteq \{1, \ldots, |J|\}$ is the index set corresponding to K largest value of $\{((w_2)_i)^2 - (\bar{y}_i - (w_2)_i)^2\}_{i=1}^{|J|}$. And from (3) and (21) we observe that for $i = 1, \ldots, |J|$, if there exists $0 \le \bar{l} \le p_{j(i)}$ such that $(w_2)_i \in [a_{\bar{l}J(i)}, b_{\bar{l}J(i)}]$, then $\bar{y}_i = (w_2)_i$, otherwise it has

$$\bar{y}_i \in \arg\min_{y_i} \left\{(y_i - (w_2)_i)^2 : y_i \in \{a_{lJ(i)}, b_{lJ(i)}\}_{l=1}^{p_{J(i)}} \cup \{0\}\right\}.$$

Therefore, problem (15) can be solved by iteratively solving a proximal problem and a sparse projection problem. We rewrite the PALM method for solving problem (15) as following:

PALM method for (15):
Choose an initial point $(x^0, y^0) \in \mathrm{R}^n \times \mathrm{R}^{|J|}$. Set $l = 0$.

(1) Take $\gamma_1 > 1$, $t_l^1 = \gamma_1 L_1(y^l)$ and compute

$$x^{l+1} \in \mathrm{Prox}_{t_l^1}^g(x^l - \frac{1}{t_l^1}\nabla_x \varphi(x^l, y^l, \mu, \nu, \eta, \rho)).$$

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

$$y^{l+1} \in \arg\min_y \left\{\|y - (y^l - \frac{1}{t_l^2}\nabla_y \varphi(x^{l+1}, y^l, \mu, \nu, \eta, \rho))\| : y \in \mathcal{E}_1\right\}.$$

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

We are now ready to propose the ALPA method for (approximate) solving problem (12) (or equivalently problem (1)) in which each AL subproblem is solved by the PALM method. Let x^{feas} be a feasible point of problem (1).

Augmented Lagrangian Proximal Alternating Method for (1):
Let $\{\varepsilon_k\}$ be a positive decreasing sequence. Let $\mu^0 \in \mathrm{R}^m$, $\nu^0 \in \mathrm{R}^p_+$, $\eta^0 \in \mathrm{R}^{|J|}$, $\rho_0 > 0$, $\sigma > 1$, $0 < \tau, \theta < 1$, $\gamma_1, \gamma_2 > 1$ be given. Choose an arbitrary point $(x_0^0, y_0^0) \in \mathrm{R}^n \times \mathrm{R}^{|J|}$ and a constant $\Upsilon \ge \max\{f(x^{\mathrm{feas}}) + g(x^{\mathrm{feas}}), \mathcal{L}(x_0^0, y_0^0, \mu^0, \nu^0, \eta^0, \rho_0)\}$. Set $k = 0$.

(1) Set $l = 0$ and apply the PALM method to find an approximate critical point $(x^k, y^k) \in \mathrm{R}^n \times \mathrm{R}^{|J|}$ of the AL subproblem

$$\min_{x,y} \Psi_1(x, y, \mu^k, \nu^k, \eta^k, \rho_k) = g(x) + \varphi(x, y, \mu^k, \nu^k, \eta^k, \rho_k) + h_1(y), \tag{22}$$

by performing steps (1.1)-(1.4):

(1.1) Set $t^k_{1,l} = \gamma_1 L^k_1(y^k_l)$, $w_1 = x^k_l - \frac{1}{t^k_{1,l}} \nabla_x \varphi(x^k_l, y^k_l, \mu^k, \nu^k, \eta^k, \rho_k)$, compute $x^k_{l+1} \in \mathrm{Prox}^g_{t^k_{1,l}}(w_1)$.

(1.2) Set $t^k_{2,l} = \gamma_2 L^k_2$, $w_2 = y^k_l - \frac{1}{t^k_{2,l}} \nabla_y \varphi(x^k_{l+1}, y^k_l, \mu^k, \nu^k, \eta^k, \rho_k)$, compute $y^k_{l+1} \in P_{\mathcal{E}_1}(w_2)$,

(1.3) If (x^k_{l+1}, y^k_{l+1}) satisfies

$$\mathrm{dist}(0, \partial \Psi_1(x^k_{l+1}, y^k_{l+1}, \mu^k, \nu^k, \eta^k, \rho_k) < \varepsilon_k, \tag{23}$$

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

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

(2) Set $\mu^{k+1} = \mu^k + \rho_k c(x^k)$, $\nu^{k+1} = [\nu^k + \rho_k d(x^k)]_+$, $\eta^{k+1} = \eta^k + \rho_k(x^k_J - y^k)$ and $\zeta^{k+1} = \min\{\nu^{k+1}, -d(x^k)\}$.

(3) If $k = 0$ or

$$\max\left\{\|c(x^k)\|, \|x^k_J - y^k\|, \|\zeta^{k+1}\|\right\} \le \theta \max\{\|c(x^{k-1})\|, \|x^{k-1}_J - y^{k-1}\|, \|\zeta^k\|\},$$

then set $\rho_{k+1} = \rho_k$. Otherwise, set

$$\rho_{k+1} = \max\{\sigma\rho_k, \|\mu^{k+1}\|^{1+\tau}, \|\nu^{k+1}\|^{1+\tau}, \|\eta^{k+1}\|^{1+\tau}\}$$

(4) If $\mathcal{L}(x^k, y^k, \mu^{k+1}, \nu^{k+1}, \eta^{k+1}, \rho_{k+1})\} > \Upsilon$, set $(x^{k+1}_0, y^{k+1}_0) = (x^{\mathrm{feas}}, x^{\mathrm{feas}}_J)$. Otherwise, set $(x^{k+1}_0, y^{k+1}_0) := (x^k, y^k)$.

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

Remark 2 The condition (23) will be used in the analysis of the global convergence of the ALPA 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 $\{(x^k_l, y^k_l)\}$, that is,

$$\max\left\{\frac{\|x^k_{l+1} - x^k_l\|_\infty}{\max(\|x^k_l\|_\infty, 1)}, \frac{\|y^k_{l+1} - y^k_l\|_\infty}{\max(\|y^k_l\|_\infty, 1)}\right\} \le \varepsilon_I \tag{24}$$

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

$$\|x^k - y^k\|_\infty \le \varepsilon_O \quad \text{or} \quad \frac{\|(x^{k+1}, y^{k+1}) - (x^k, y^k)\|_\infty}{\max(\|(x^k, y^k)\|_\infty, 1)} \le \varepsilon_O, \tag{25}$$

for some $\varepsilon_O > 0$. For enhancing the performance and convergence of the ALPA method, we may recompute the AL subproblem multiple times from a suitable perturbation of the current best approximate solution.

In the practical calculation, for speeding up the computation we can use an end game strategy which needs not necessary to set the parameters of the termination conditions too strict (small) and find a very accurate solution. We can choose some moderate parameters and find a accurate distribution of a solution in less time, this distribution indicates the intervals or point sets in which the components of the solution are contained. Then we use a heuristic method that restricts the discrete variables to constant intervals or point sets and solves a tightened problem by an efficient method to find an accurate solution of problem (1).

3.2 Augmented Lagrangian proximal alternating method for problem(2)

In this subsection, we propose an ALPA method for solving problem (2). Similarly with the ALPA method for problem (1), we reformulate problem (2) as

$$\begin{aligned} \min_{x,y} \quad & f(x)+g(x)+\lambda\|y\|_0 \\ \text{s.t.} \quad & c(x)=0, \\ & d(x)\le 0, \\ & x_J=y, \\ & y_i\in\mathcal{D}_{J(i)},\ i=1,\ldots,|J|. \end{aligned} \tag{26}$$

Let $\mathcal{E}_2=\left\{y\in \mathrm{R}^{|J|}: y_i\in\mathcal{D}_{J(i)}, i=1,\ldots,|J|\right\}$ denote the feasible set of the discrete constraints. Similarly, for any given penalty parameter $\rho>0$ we define the AL function of problem (2) (or equivalently (26)) as following:

$$\begin{aligned} \mathcal{L}_2(x,y,\mu,\nu,\eta,\rho) &= f(x)+g(x)+\tfrac{1}{2\rho}(\|\mu+\rho c(x)\|^2-\|\mu\|^2) \\ &+\tfrac{1}{2\rho}(\|[\nu+\rho d(x)]_+\|^2-\|\nu\|^2)+\tfrac{1}{2\rho}(\|\eta+\rho(x_J-y)\|^2-\|\eta\|^2)+\lambda\|y\|_0. \end{aligned}$$

By utilizing (14), the AL function of problem (2) can be rewritten as: $\mathcal{L}_2(x,y,\mu,\nu,\eta,\rho)=g(x)+\varphi(x,y,\mu,\nu,\eta,\rho)+\lambda\|y\|_0$ and the associated AL subproblem for problem (2) can be rewritten as the following unconstrained minimization problem:

$$\min_{x,y}\ \Psi_2(x,y,\mu,\nu,\eta,\rho)=g(x)+\varphi(x,y,\mu,\nu,\eta,\rho)+h_2(y) \qquad \text{over all}\ \ (x,y)\in \mathrm{R}^n\times \mathrm{R}^{|J|}, \tag{27}$$

where $h_2(y)=I_{\mathcal{E}_2}(y)+\lambda\|y\|_0$ and $I_{\mathcal{E}_2}(y)$ denotes the indicator function of $\mathcal{E}_2$.

Similarly with the PALM method for problem (15), for ensuring the convergence of the PALM method for problem (27) we take the following assumption.

Assumption 2 *Replace $\Psi_1(x,y,\mu,\nu,\eta,\rho)$ in the conditions of Assumption 1 by $\Psi_2(x,y,\mu,\nu,\eta,\rho)$ and assume that all the conditions are satisfied.*

Remark 3 Function $\|y\|_0$ is a lower semi-continuous and bounded function thus $h_2(y)$ is lower semi-continuous and $\inf_{y\in \mathrm{R}^{|J|}} h_2(y) > -\infty$. Similarly with Remark 1, we have that for any fixed x the function $y \to \varphi(x, y, \mu, \nu, \eta, \rho)$ is $C^{1,1}_{L_2}$ where $L_2 = \rho$ and functions $g(x)$, $\varphi(x, y, \mu, \nu, \eta, \rho)$ and $h_2(y)$ satisfy Assumption A and B in [24]. Like Assumption 1, in the assumption (v) the condition of the boundedness of the level set of function $\Psi_2(x, y, \mu, \nu, \eta, \rho)$ can be replaced by other conditions which can guarantee the boundedness of the sequence generated by the PALM method.

Hence, problem (27) can be solved by utilizing the PALM framework. For constant μ, ν, η and ρ, the iterative sequence $\{(x^l, y^l)\}$ are updated by solving the following two proximal problems:

$$x^{l+1} \in \mathrm{Prox}^{g}_{t^1_l}(w_1),$$
$$y^{l+1} \in \mathrm{Prox}^{h_2}_{t^2_l}(w_2) = \arg\min_y \left\{ \frac{t^2_l}{2}\|y - w_2\|^2 + \lambda\|y\|_0 : y \in \mathcal{E}_2 \right\},$$

where $w_1 = x^l - (1/t^1_l)\nabla_x \varphi(x^l, y^l, \mu, \nu, \eta, \rho)$, $w_2 = y^l - (1/t^2_l)\nabla_y \varphi(x^{l+1}, y^l, \mu, \nu, \eta, \rho)$, $t^1_l = \gamma_1 L_1(y^l)$, $t^2_l = \gamma_2 L_2$ $\gamma_1, \gamma_2 > 1$ are two appropriately chosen step sizes and $L_1(y^l)$, L_2 are the lipschitz constants in Assumption 1 and Remark 1.

It is obvious that $\mathrm{Prox}^{h_2}_{t^2_l}$ is a generalized hard thresholding operator. Similarly to $P_{\mathcal{E}_1}$, $\mathrm{Prox}^{h_2}_{t^2_l}$ is a multi-valued map. Suppose $\bar{y}^* \in \mathrm{Prox}^{h_2}_{t^2_l}(w_2)$, then $\bar{y}^*$ can be computed as follows:

$$\bar{y}^*_i = \begin{cases} \bar{y}_i, & \text{if } \bar{v}_i \geq 0; \\ 0, & \text{otherwise,} \end{cases} \quad i = 1, \ldots, |J|,$$

where $\bar{y}_i \in \arg\min_{y_i}\{(y_i - (w_2)_i)^2 : y_i \in \mathcal{D}_{J(i)}\}$ and $\bar{v}_i = (t^2_l/2)((w_2)_i)^2 - \lambda - (t^2_l/2)(\bar{y}_i - (w_2)_i)^2$ for $i = 1, \ldots, |J|$.

We now present the ALPA method for problem (2) which is similar to the ALPA method for problem (1). Let x^{feas} be a feasible point of problem (2).

Augmented Lagrangian Proximal Alternating Method for (2):
Let $\{\varepsilon_k\}$ be a positive decreasing sequence. Let $\mu^0 \in \mathrm{R}^m$, $\nu^0 \in \mathrm{R}^p_+$, $\eta^0 \in \mathrm{R}^{|J|}$, $\rho_0 > 0$, $\sigma > 1$, $0 < \tau, \theta < 1$, $\gamma_1 > 1$, $\gamma_2 > 1$ be given. Choose an arbitrary point $(x^0_0, y^0_0) \in \mathrm{R}^n \times \mathrm{R}^{|J|}$ and a constant $\Upsilon \geq \max\{f(x^{\mathrm{feas}}) + g(x^{\mathrm{feas}}) + \lambda\|x^{\mathrm{feas}}_J\|_0, \mathcal{L}(x^0_0, y^0_0, \mu^0, \nu^0, \eta^0, \rho_0)\}$. Set $k = 0$.

(1) Set $l = 0$ and apply the PALM method to find an approximate critical point $(x^k, y^k) \in \mathrm{R}^n \times \mathrm{R}^{|J|}$ of the AL subproblem

$$\min_{x,y} \Psi_2(x, y, \mu^k, \nu^k, \eta^k, \rho_k) = g(x) + \varphi(x, y, \mu^k, \nu^k, \eta^k, \rho_k) + h_2(y), \tag{28}$$

by performing steps (1.1)-(1.4):

(1.1) Set $t^k_{1,l} = \gamma_1 L^k_1(y^k_l)$, $w_1 = x^k_l - \frac{1}{t^k_{1,l}}\nabla_x \varphi(x^k_l, y^k_l, \mu^k, \nu^k, \eta^k, \rho_k)$, compute $x^k_{l+1} \in \mathrm{Prox}^{g}_{t^k_{1,l}}(w_1)$.

(1.2) Set $t^k_{2,l} = \gamma_2 L^k_2$, $w_2 = y^k_l - \frac{1}{t^k_{2,l}}\nabla_y \varphi(x^k_{l+1}, y^k_l, \mu^k, \nu^k, \eta^k, \rho_k)$, compute $y^k_{l+1} \in \mathrm{Prox}^{h_2}_{t^k_{2,l}}(w_2)$,

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

$$\mathrm{dist}(0, \partial\Psi_2(x_{l+1}^k, y_{l+1}^k, \mu^k, \nu^k, \eta^k, \rho_k) < \varepsilon_k, \tag{29}$$

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 $\mu^{k+1} = \mu^k + \rho_k c(x^k)$, $\nu^{k+1} = [\nu^k + \rho_k d(x^k)]_+$, $\eta^{k+1} = \eta^k + \rho_k(x_J^k - y^k)$ and $\zeta^{k+1} = \min\{\nu^{k+1}, -d(x^k)\}$.

(3) If $k = 0$ or

$$\max\left\{\|c(x^k)\|, \|x_J^k - y^k\|, \|\zeta^{k+1}\|\right\} \le \theta \max\{\|c(x^{k-1})\|, \|x_J^{k-1} - y^{k-1}\|, \|\zeta^k\|\},$$

then set $\rho_{k+1} = \rho_k$. Otherwise, set

$$\rho_{k+1} = \max\{\sigma\rho_k, \|\mu^{k+1}\|^{1+\tau}, \|\nu^{k+1}\|^{1+\tau}, \|\eta^{k+1}\|^{1+\tau}\}$$

(4) If $\mathcal{L}(x^k, y^k, \mu^{k+1}, \nu^{k+1}, \eta^{k+1}, \rho_{k+1})\} > \Upsilon$, set $(x_0^{k+1}, y_0^{k+1}) = (x^{\mathrm{feas}}, x_J^{\mathrm{feas}})$. Otherwise, set $(x_0^{k+1}, y_0^{k+1}) := (x^k, y^k)$.

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

Remark 4 The practical termination criteria (24) and (25) used in the ALPA method for problem (1) can also be applied to this method for problem (2).

For speeding up the ALPA 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 convergence of the ALPA method for problems (1) and (2) in which each AL subproblem is a nonconvex and nonsmooth problem and approximately solved by the PALM method. For notational convenience, we omit the index k of the parameters in the subproblem of the ALPA method for problems (1) and (2) and rewrite them as

$$\min_{x,y} \Psi_1(x, y, \mu, \nu, \eta, \rho) = g(x) + \varphi(x, y, \mu, \nu, \eta, \rho) + h_1(y),\ (x, y) \in \mathrm{R}^n \times \mathrm{R}^{|J|}. \tag{30}$$

and

$$\min_{x,y} \Psi_2(x, y, \mu, \nu, \eta, \rho) = g(x) + \varphi(x, y, \mu, \nu, \eta, \rho) + h_2(y),\ (x, y) \in \mathrm{R}^n \times \mathrm{R}^{|J|}. \tag{31}$$

Bolte et al. showed that for any nonconvex and nonsmooth problem, if the objective function of the problem satisfies the assumptions A and B in [24] and is a KL function, then any bounded sequence generated by the PALM method globally converges to a critical point of the problem. From Assumption 1 and 2, we can confirm that the lower semi-continuous functions $\Psi_1(x, y, \mu, \nu, \eta, \rho)$ and $\Psi_2(x, y, \mu, \nu, \eta, \rho)$ satisfy the assumptions in [24] and are KL functions. Hence, utilizing Assumption 1, 2 and the conclusions of [24,32], we give the convergence theorem which ensures the convergency of the iterative sequence generated in solving the AL subproblem.

Theorem 5 *Let μ, ν, η and ρ be given. Under Assumption 1 or 2, the sequence $\{(x^l, y^l)\}$ generated by the PALM method in solving the AL subproblem (30) or (31) converges to a critical point of (30) or (31). And for given $\varepsilon > 0$ there exists $\bar{L} > 0$ such that for all $l > \bar{L}$ it has*

$$dist(0, \partial\Psi_1(x^l, y^l, \mu, \nu, \eta, \rho) < \varepsilon \quad or \quad dist(0, \partial\Psi_2(x^l, y^l, \mu, \nu, \eta, \rho) < \varepsilon. \tag{32}$$

Proof The level set of the function $\Psi_1(x, y, \mu, \nu, \eta, \rho)$ is bounded, and it is shown in [24] that the function value sequence $\{\Psi_1(x^l, y^l, \mu, \nu, \eta, \rho)\}$ is nonincreasing. Therefore the sequence $\{(x^l, y^l)\}$ is bounded. By Assumption 1 and the conclusions in [24], we have that the sequence $\{(x^l, y^l)\}$ converges to a critical point of subproblem (30). It follows from Lemma 3.4, Lemma 3.5 and Theorem 3.1 of [24] that

$$\lim_{l\to\infty} \text{dist}(0, \partial\Psi_1(x^l, y^l, \mu, \nu, \eta, \rho) = 0.$$

Hence, the first part of (32) immediately follows. The proof of the PALM method for problem (31) is analogous to the proof of the PALM method for problem (30).

From the theorem above, the sequence generated by the PALM method converges to a critical point of the AL subproblem (30) or (31). Now we show that under some suitable assumption, any accumulation point of the sequence generated by the ALPA method for problem (1) satisfies the first-order necessary conditions of problem (1). Moreover, when the l_0-norm and the discrete constraints are the only nonconvex parts of problem (1), we can show that any accumulation point is a local minimizer of problem (1).

Theorem 6 *Assume that $\{\varepsilon_k\} \to 0$. Let $\{(x^k, y^k)\}$ be the sequence generated by the ALPA method for problem (1), $I_k = \{i_1^k, \ldots, i_K^k\}$ be a set of K distinct indices in $\{1, \cdots, |J|\}$ such that $(y^k)_i = 0$ for any $i \notin I_k$ and let $J_k = \{J(i) : i \in I_k\}$. Then the following statements hold:*

(i) Suppose (x^*, y^*) is an accumulation point of $\{(x^k, y^k)\}$. Then $x^*_J = y^*$ and x^* is a feasible point of problem (1). Moreover, there exists a subsequence Q such that $\{(x^k, y^k)\}_{k\in Q} \to (x^*, y^*)$, $I_k = I^*$ and $J_k = J^* := \{J(i) : i \in I^*\}$ for some index set $I^* \subseteq \{1, \cdots, |J|\}$ when $k \in Q$ is sufficiently large.

(ii) Let x^*, Q, and J^* be defined as above, and let $\bar{J}^* = J \setminus J^*$. Suppose that the BQ (6) and the MFCQ condition (4) hold at x^* for such J^* and $\bar{J}^*$. Then x^* satisfies the first-order optimality conditions (8). Moreover, if f, g and d's are convex functions, and c's are affine functions, then x^* is a local minimizer of (1).

Proof Without loss of generality, we can assume that $(x^k, y^k) \to (x^*, y^*)$ as $k \to \infty$. Then we consider the following two separate cases:

case(a): $\{\rho_k\}$ is bounded. By the updating rule in Step 3) of the ALPA method, we can observe that for any sufficiently large k,

$$\max\left\{\|c(x^k)\|, \|x_J^k - y^k\|, \|\zeta^{k+1}\|\right\} \le \theta \max\{\|c(x^{k-1})\|, \|x_J^{k-1} - y^{k-1}\|, \|\zeta^k\|\},$$

which implies that $c(x^*) = 0$, $x_J^* = y^*$ and

$$-d(x^*) = \lim_{k\to\infty} -d(x^k) \geq \lim_{k\to\infty} \zeta^{k+1} = 0.$$

Case (b): $\{\rho_k\}$ is unbounded. Considering a subsequence if necessary, for convenience we assume that $\{\rho_k\}$ strictly increases and goes to infinity as $k \to \infty$. Hence, $\rho_{k+1} = \max\{\sigma\rho_k, \|\mu^{k+1}\|^{1+\tau}, \|\nu^{k+1}\|^{1+\tau}, \|\eta^{k+1}\|^{1+\tau}\}$ for all k, then as $k \to \infty$

$$0 \leq \frac{\|\mu^k\|}{\rho_k}, \ \frac{\|\nu^k\|}{\rho_k}, \ \frac{\|\eta^k\|}{\rho_k} \leq (\rho_k)^{-\frac{\tau}{1+\tau}} \to 0. \tag{33}$$

From Step 4) of the ALPA method and the nonincreasing property of the sequence generated by the PALM method, we have that

$$\begin{aligned} &f(x^k) + g(x^k) + \tfrac{1}{2\rho_k}(\|\mu^k + \rho_k c(x_k)\|^2 - \|\mu^k\|^2) \\ &+\tfrac{1}{2\rho_k}(\|[\nu^k + \rho_k d(x^k)]_+\|^2 - \|\nu^k\|^2) + \tfrac{1}{2\rho_k}(\|\eta^k + \rho_k(x_J^k - y^k)\|^2 - \|\eta^k\|^2) \leq \Upsilon. \end{aligned}$$

It follows that

$$\begin{aligned} \|c(x^k) + \tfrac{\mu^k}{\rho_k}\|^2 + \|[d(x^k) + \tfrac{\nu^k}{\rho_k}]_+\|^2 + \|x_J^k - y^k + \tfrac{\eta^k}{\rho_k}\|^2 \\ \leq \tfrac{2}{\rho_k}[\Upsilon - f(x^k) - g(x^k)] + \tfrac{\|\mu^k\|^2 + \|\nu^k\|^2 + \|\eta^k\|^2}{\rho_k^2}. \end{aligned}$$

Taking limits on both sides of this relation as $k \to \infty$, and using (33), we can obtain

$$\|c(x^*)\|^2 + \|[d(x^*)]_+\|^2 + \|x_J^* - y^*\|^2 = 0,$$

which implies that $c(x^*) = 0$, $x_J^* = y^*$ and $d(x^*) \leq 0$. Hence, we can obtain that $x_J^* = y^*$ and x^* is a feasible point of problem (1). Next we discuss the second part of statement (i). Since (x^*, y^*) is an accumulation point of $\{(x^k, y^k)\}$, there exists a subsequence $\{(x^k, y^k)\}_{k\in\bar{Q}} \to (x^*, y^*)$. Recall that I_k is an index set. It follows that $\{(i_1^k, \ldots, i_K^k)\}_{k\in\bar{Q}}$ is bounded for all k. Thus there exists a subsequence $Q \subseteq \bar{Q}$ such that $\{(i_1^k, \ldots, i_K^k)\}_{k\in Q} \to (i_1^*, \ldots, i_K^*)$ for some K distinct indices $i_1^*, \ldots, i_K^*$. Since $i_1^k, \ldots, i_K^k$ are K distinct integers, we can obtain that $(i_1^k, \ldots, i_K^k) = (i_1^*, \ldots, i_K^*)$ when $k \in Q$ is sufficiently large. Let $I^* = \{i_1^*, \ldots, i_K^*\}$. Then, we obtain that $\{(x^k, y^k)\}_{k\in Q} \to (x^*, y^*)$ and $I_k = I^*$ and $J_k = J^*$ for sufficiently large $k \in Q$. Therefore statement (i) holds.

From the termination condition of the PALM method, there exist vectors s_x^k, s_y^k with $\|(s_x^k, s_y^k)\| \leq \varepsilon_k$ such that for any k sufficiently large,

$$\begin{aligned} s_x^k &\in \partial g(x^k) + \nabla_x \varphi(x^k, y^k, \mu^k, \nu^k, \eta^k, \rho_k), \\ s_y^k &\in \partial h_1(y^k) + \nabla_y \varphi(x^k, y^k, \mu^k, \nu^k, \eta^k, \rho_k). \end{aligned}$$

From $h_1(y) = I_{\mathcal{E}_1}(y)$ and the property of the indicator function, we have $\partial h_1(y^k) = \partial I_{\mathcal{E}_1}(y^k) = N_{\mathcal{E}_1}(y^k)$. Since $\nu^{k+1} = [\nu^k + \rho_k d(x_k)]_+$ and $\|\nu^k\|/\rho_k \to 0$ as $k \to \infty$, it follows that for any k sufficiently large, it has $\nu_i^{k+1} = 0$,

$i \notin \mathcal{I}_d^* := \mathcal{I}_d(x^*)$ where $\mathcal{I}_d(z) := \{i : d_i(z) = 0\}$, $\forall z \in \mathrm{R}^n$. Furthermore, from Step 2) in the ALPA method we have

$$s_x^k \in \partial g(x^k) + \nabla f(x^k) + \sum_{i=1}^{m} \mu_i^{k+1} \nabla c_i(x^k) + \sum_{i \in \mathcal{I}_d^*} \nu_i^{k+1} \nabla d_i(x^k) + \bar{\eta}^{k+1}, \quad (34)$$

$$s_y^k \in N_{\mathcal{E}_1}(y^k) - \eta^{k+1}, \quad (35)$$

where $\bar{\eta}^{k+1}$ is the expansion of η^{k+1} in R^n with $\bar{\eta}_J^{k+1} = \eta^{k+1}$ and $\bar{\eta}_{\bar{J}}^{k+1} = 0$ for $\bar{J} = \{1, \ldots, n\} \setminus J$.

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

$$\begin{aligned} s_x^k + \bar{s}_y^k \in \partial g(x^k) + \nabla f(x^k) + \textstyle\sum_{i=1}^{m} \mu_i^{k+1} \nabla c_i(x^k) \\ + \textstyle\sum_{i \in \mathcal{I}_d^*} \nu_i^{k+1} \nabla d_i(x^k) + N_{\mathcal{E}_1}(y^k) \times \{0\}^{n-|J|}, \end{aligned} \quad (36)$$

where $(\bar{s}_y^k)_J = s_y^k$ and $(\bar{s}_y^k)_{\bar{J}} = 0$.

Let $t_k := \max\{|\mu_i^{k+1}|, |\nu_j^{k+1}| : i \in \{1, \ldots, m\}, j \in \mathcal{I}_d^*\}$. Next, we prove that $\{t_k\}$ is bounded. Suppose on the contrary that $\{t_k\}$ is unbounded. Without loss of generality, we assume that as $k \to \infty$, $t_k \to +\infty$ and

$$\frac{\mu_i^{k+1}}{t_k} \to \mu_i^* \quad i \in \{1, \ldots, m\}, \quad \frac{\nu_j^{k+1}}{t_k} \to \nu_j^* \quad j \in \mathcal{I}_d^*.$$

It is clear that

$$\max\{|\mu_i^*|, |\nu_j^*| : i \in \{1, \ldots, m\}, j \in \mathcal{I}_d^*\} = 1. \quad (37)$$

Moreover, since $\nu_j^k \geq 0$, $j \in \mathcal{I}_d^*$ for any sufficiently large k , we have $\nu_j^* \geq 0$, $j \in \mathcal{I}_d^*$. Dividing (36) by t_k and taking limits on both sides as $k \to \infty$, it from the definition of horizon subdifferential, the outer semi-continuity of limiting subdifferential and limiting normal cone that

$$0 \in \partial^\infty g(x^*) + \sum_{i=1}^{m} \mu_i^* \nabla c_i(x^*) + \sum_{i \in \mathcal{I}_d^*} \nu_i^* \nabla d_i(x^*) + N_{\mathcal{E}_1}(x^*) \times \{0\}^{n-|J|}.$$

Then, there exists $\beta^* \in N_{\mathcal{E}_1}(x^*) \times \{0\}^{n-|J|}$ such that

$$\sum_{i=1}^{m} \mu_i^* \nabla c_i(x^*) + \sum_{i \in \mathcal{I}_d^*} \nu_i^* \nabla d_i(x^*) + \beta^* \in -\partial^\infty g(x^*).$$

Moreover, we can observe that

$$\sum_{i=1}^{m} \mu_i^* \nabla c_i(x^*) + \sum_{i \in \mathcal{I}_d^*} \nu_i^* \nabla d_i(x^*) + \beta^* \in N_{\mathcal{F}_1}(x^*).$$

Then it follows from the BQ (6) and the last relations that

$$\sum_{i=1}^{m} \mu_i^* \nabla c_i(x^*) + \sum_{i \in \mathcal{I}_d^*} \nu_i^* \nabla d_i(x^*) + \beta^* = 0. \tag{38}$$

If $\beta^* = 0$ then (38) together with (37) and $\nu_j^* \geq 0$, $j \in \mathcal{I}_d^*$ implies that

$$\{\nabla c_i(x^*), \nabla d_i(x^*) : i \in \{1, \ldots, m\}, j \in \mathcal{I}_d^*\}$$

is positively linearly dependent. This contradicts with the MFCQ condition. If $\beta^* \neq 0$ then (38) together with (37) and $\nu_j^* \geq 0$, $j \in \mathcal{I}_d^*$ implies that

$$\{\nabla c_i(x^*), \nabla d_i(x^*), \beta^* : i \in \{1, \ldots, m\}, j \in \mathcal{I}_d^*\}$$

is positively linearly dependent. This also contradicts with the MFCQ condition. Hence, t_k is bounded and it follows that $\{\mu_i^{k+1}\}$, $\{\nu_j^{k+1}\}$ are bounded for all $i \in \{1, \ldots, m\}$, $j \in \mathcal{I}_d^*$. Without loss of generality, we assume that as $k \to \infty$,

$$\mu_i^{k+1} \to \mu_i^* \quad i \in \{1, \ldots, m\}, \quad \nu_j^{k+1} \to \nu_j^* \quad j \in \mathcal{I}_d^*.$$

Moreover, since $\nu_j^k \geq 0$, $j \in \mathcal{I}_d^*$ for any sufficiently large k , we have $\nu_j^* \geq 0$, $j \in \mathcal{I}_d^*$. Then taking limits on both sides of (36), we have

$$0 \in \partial g(x^*) + \nabla f(x^*) + \sum_{i=1}^{m} \mu_i^* \nabla c_i(x^*) + \sum_{i \in \mathcal{I}_d^*} \nu_i^* \nabla d_i(x^*) + N_{\mathcal{E}_1}(x^*) \times \{0\}^{n-|J|}. \tag{39}$$

By the assumption and (39), we have that x^* satisfies the first-order optimality conditions (8) for problem (1). Further, if f, g and d's are convex functions, and c's are affine functions, the assumptions of Theorem 3 hold. It then follows from Theorem 3 that x^* is a local minimizer of problem (1). Statement (ii) holds.

We next show that under some suitable assumption, any accumulation point of the sequence generated by the ALPA method for problem (2) satisfies the first-order necessary conditions of problem (2). Moreover, when the l_0-norm and the discrete constraints are the only nonconvex parts of problem (2), we can show that any accumulation point is a local minimizer of problem (2).

Theorem 7 *Assume that $\{\varepsilon_k\} \to 0$. Let $\{(x^k, y^k)\}$ be the sequence generated by the ALPA method for problem (2). Then the following statements hold:*

(i) Suppose (x^*, y^*) is an accumulation point of $\{(x^k, y^k)\}$. Then $x_J^* = y^*$ and x^* is a feasible point of problem (2).

(ii) Let (x^*, y^*) be defined as above, and assume that $\{(x^k, y^k)\}_{k \in Q} \to (x^*, y^*)$ for some subsequence Q. Let $J^* = \{j \in J : x_j^* \neq 0\}$, $\bar{J}^* = J \setminus J^*$. Suppose that the BQ (7) and the MFCQ condition (5) hold at x^* for such J^* and $\bar{J}^*$. Then x^* satisfies the first-order optimality conditions (10). Moreover, if f, g and d's are convex functions, and c's are affine functions, then x^* is a local minimizer of problem (2).

Proof The proofs of statement (i) and the first part of statement (ii) are similar to that of Theorem 6. It follows from Theorem 4 and the first conclusion of statement (ii) that x^* is a local minimizer of problem (2). Statement (ii) holds.

5 Numerical experiments

In this section, we provide numerical results that show the efficiency and advantage of the ALPA method for finding good quality suboptimal solutions of problems (1) and (2). We compared our method with the PD method and the standard CPLEX(12.6) solver by solving following sparse portfolio selection problems with generalized minimum transaction levels:

$$\begin{array}{ll} \min & \frac{1}{2}x^\top Qx \\ \text{s.t.} & e^\top x = 1, \\ & R^\top x \geq r, \\ & x \in \mathcal{D}, \\ & \|x\|_0 \leq K, \end{array} \tag{40}$$

and

$$\begin{array}{ll} \min & \frac{1}{2}x^\top Qx + \lambda\|x\|_0 \\ \text{s.t.} & e^\top x = 1, \\ & R^\top x \geq r, \\ & x \in \mathcal{D}, \end{array} \tag{41}$$

where $x \in \mathrm{R}^n$ denotes the proportion of the total funds invested on n assets, Q denotes the covariance matrix, R denotes the expected return of n assets, e is a n-dimensional vector with all components one, r is a minimum profit target, and $\mathcal{D} = \{x \in \mathrm{R}^n : (l_1)_i \leq x_i \leq (u_1)_i$ or $(l_2)_i \leq x_i \leq (u_2)_i$ or $x_i = 0, i = 1, \ldots, n\}$, $(l_1)_i < (u_1)_i < 0$, $0 < (l_2)_i < (u_2)_i$ for $i = 1, \ldots, n$ are bounds of the transaction levels, l_1, u_1, l_2, u_2 are transaction level vectors.

For constructing large-scale test problems, the data sets used in our experiments were selected from the index tracking problem data in ORlibrary [33]. We chose the weekly return 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 to estimate the expected return and the covariance matrix. For convenience, in each group the minimum profit target r was chosen to be the average of the expected return of all assets and all the components of the transaction level vectors l_1, u_1, l_2, u_2 were equal to -1, -0.01, 0.01, 1 respectively. The ALPA method and the PD method were coded in Matlab (R2014b), the penalty subproblems of the PD method were solved by Matlab optimization software

package OptiToolbox, and all the computations were performed on a PC (Intel core i7-4790 CPU, 3.6GHz, 16GB RAM).

For problem (40), the upper bounds of the numbers of selected assets (nonzero components) K were chosen as 5, 10, 15, 20, 30 and 40, respectively. We chose $\rho_0 = 10^{-6}$, $\sigma = 1.1$, $\gamma_1 = 1.001$, $\gamma_2 = 1.001$, $\theta = 0.9$, $\tau = 0.01$ for the ALPA method and set the initial Lagrangian multiplier to zero. For the PALM method, for any $x, y \in \mathrm{R}^n$, $L_1^k = \|Q\| + \rho_k$, $L_2^k = \rho_k$. In addition, we used (24) and (25) as the inner and outer termination criteria for the ALPA method with $\varepsilon_I = 10^{-5}$, $\varepsilon_O = 10^{-6}$, respectively. For the PD method, we chose the initial penalty parameter ρ_0 from 10^{-3} to 0.1, the increasing rate of the penalty parameter σ from 1.2 to 5, and used the termination criteria in [23] with $\varepsilon_I = 10^{-3}$, $\varepsilon_O = 10^{-4}$, respectively. For the CPLEX solver, we used the default parameters of the software.

Table 1 Comparison of the methods of ALPA and PD for (40)

		ALPA		PD		
n	K	obj	time	obj	time	imp
225	5	1.3825e-04	1.2712	1.6854e-04	18.2191	0.1449(10)
	10	1.0472e-04	1.3312	1.2230e-04	18.8968	0.1324(10)
	15	8.5015e-05	1.4364	1.0489e-04	14.9526	0.1697(10)
	20	7.2863e-05	1.3511	9.3096e-05	15.2011	0.1720(10)
	30	5.7799e-05	1.2338	7.1688e-05	14.7974	0.1779(10)
	40	4.7481e-05	1.2480	5.8551e-05	14.5048	0.1739(10)
457	5	1.2506e-04	2.9671	2.5258e-04	229.1533	0.5959(10)
	10	9.5219e-05	3.0476	1.3161e-04	133.7818	0.3443(10)
	15	8.0248e-05	3.0964	1.0431e-04	87.3111	0.3174(10)
	20	6.5320e-05	2.7954	9.3869e-05	49.6248	0.3333(10)
	30	5.0777e-05	2.6540	7.0128e-05	45.2400	0.3033(10)
	40	3.8746e-05	2.6300	5.8572e-05	40.8366	0.3818(10)
1319	5	1.0149e-04	46.3189	3.8429e-04	1335.8849	0.7243(10)
	10	6.3368e-05	45.3852	2.1794e-04	748.0053	0.7749(10)
	15	4.4908e-05	42.6478	1.1597e-04	447.8651	0.7087(10)
	20	3.6899e-05	42.1432	1.2465e-04	384.8168	0.7890(10)
	30	2.2996e-05	43.9572	1.0347e-04	307.5458	0.7924(10)
	40	1.5917e-05	39.8018	9.0856e-05	247.3906	0.8438(10)
2152	5	1.2987e-04	93.1092	2.7692e-04	3177.9061	0.2668(10)
	10	7.4435e-04	93.3688	2.2765e-04	1728.6431	0.4065(10)
	15	5.8224e-04	91.2005	1.6444e-04	1271.4945	0.6234(10)
	20	4.6336e-05	92.3795	1.5889e-04	915.3815	0.7165(10)
	30	2.7064e-05	92.1395	1.5146e-04	488.6003	0.8337(10)
	40	1.5757e-05	90.9637	1.4839e-04	411.9239	0.9046(10)

For problem (41), we chose six different values of the regular parameter λ, namely, $\lambda_{\max}$, $0.5\lambda_{\max}$, $0.25\lambda_{\max}$, $0.1\lambda_{\max}$, $0.01\lambda_{\max}$, $0.001\lambda_{\max}$ where $\lambda_{\max} = 10^{-6}$. For the ALPA method, we chose the same parameters as that of problem (40). For the PD method, we chose ρ_0 from 0.01 to 0.1, $\sigma = 2$ and other parameters were same to the parameters of problem (40).

Table 2 Comparison of the ALPA method and the CPLEX solver for (40)

		ALPA		Cplex	
n	K	obj	time	obj	time
225	5	1.3825e-04*	0.8260	1.3825e-04*	334.5133
	10	1.0274e-04	0.8503	1.0354e-04	3600.8769
	15	8.5015e-05	0.9197	8.5729e-05	3600.6741
	20	7.2188e-05	0.9500	7.2188e-05	740.5071
	30	5.7229e-05	0.9751	5.7744e-05	3600.3321
	40	4.6976e-05	1.0036	4.6897e-05	3488.6749
457	5	1.2306e-04	2.8900	1.3591e-04	3600.1434
	10	9.3597e-05	2.7391	1.1462e-04	3600.0935
	15	7.7035e-05	2.6996	1.1236e-04	3600.1051
	20	6.4775e-05	2.7571	8.9337e-05	3600.1126
	30	4.8152e-05	2.5401	-	-
	40	3.7776e-05	2.3478	6.9483e-05	3600.2012
1319	5	6.3658e-05	40.6254	-	-
	10	4.1685e-05	43.0943	-	-
	15	3.1370e-05	43.3311	-	-
	20	2.4146e-05	41.1389	-	-
	30	1.5849e-05	39.3120	-	-
	40	1.2770e-05	38.5303	-	-
2152	5	6.1421e-05	71.3342	-	-
	10	3.7015e-05	79.3973	-	-
	15	2.7164e-05	84.8183	-	-
	20	2.1251e-05	90.9375	-	-
	30	1.4785e-05	87.0291	-	-
	40	1.0978e-05	82.3379	-	-

Table 1 summarizes the numerical results of the ALPA method and the PD method for problem (40). In each test problem of Table 1, we randomly generated ten groups of initial points and solved the test problem by using the ALPA method and the PD method ten times to find the best approximate solution. The PD method needs a feasible initial point and a sparse initial point, while the ALPA method needs a feasible initial point and two arbitrary initial points. Hence, for comparing these two methods, we chose these initial points to be the same one point. The method of random generating the initial points was that randomly chose K assets whose expected returns were no less than r, and 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 were no less than r was 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_a^* found by the ALPA method and the suboptimal solution x_p^* found by the PD method in each time, we used the following relative improvement of the objective function value of the solution

x_a^* over the solution x_p^*:

$$\text{imp} = \frac{f_{\text{obj}}(x_p^*) - f_{\text{obj}}(x_a^*)}{|f_{\text{obj}}(x_p^*)|} \tag{42}$$

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

Table 2 summarizes the numerical results of the ALPA method and the standard CPLEX(12.6) solver for problem (40). It shall be mentioned that for using the CPLEX solver we need to reformulate problem (40) to a standard mixed-integer quadratic program (MIQP) problem. By replacing x by $x_1 + x_2$ and adding 0-1 variables z_1, z_2 problem (40) can be equivalently reformulated as following MIQP problem

$$\begin{aligned}
\min \quad & \frac{1}{2}(x_1 + x_2)^\top Q(x_1 + x_2) \\
\text{s.t.} \quad & e^\top(x_1 + x_2) = 1, \\
& R^\top(x_1 + x_2) \geq r, \\
& z_1 + z_2 \leq e, \\
& e^\top(z_1 + z_2) \leq K, \\
& L_1 z_1 \leq x_1 \leq U_1 z_1, \\
& L_2 z_2 \leq x_2 \leq U_2 z_2, \\
& (z_1)_i, (z_2)_i \in \{0, 1\}, \ i = 1, \ldots, n,
\end{aligned} \tag{43}$$

where $x_1, x_2 \in \mathrm{R}^n$, L_1, L_2, U_1, U_2 are diagonal matrixes whose main diagonal are made up by elements of vectors l_1, l_2, u_1, u_2 respectively. In order to compare with the CPLEX solver, in each test problem, we chose the variables with the $K + m$ largest absolute values of the solution of the relaxation of problem (43) in which the binary constraints were replaced by the box constraints $0 \leq z_1, z_2 \leq 1$ in the usual way, where m was from 10 to 20, to randomly generate ten groups of different initial points and solved the test problem by using the ALPA method ten times. Then, we solved the MIQP problem (43) by using the CPLEX solver and compared how much time the CPLEX solver needed to find a solution as good as the best solution found by the ALPA method. If the computing time of the CPLEX solver was more than one hour, we just stopped the CPLEX solver and output the optimal solution which the CPLEX solver had found.

Table 3 summarizes the numerical results of the ALPA method and the PD method for problem (41). 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 ALPA method and the PD method to find the best suboptimal solution. We also used the relative improvement in (42) with $f_{\text{obj}}(x) = (1/2)x^\top Qx + \lambda\|x\|_0$ to measure the quality of the suboptimal solutions found by the two methods.

In Tables 1, 2 and 3, "ALPA" stands for the ALPA method; "PD" stands for the PD method; "Cplex" stands for the standard CPLEX solver; "time" (in

Table 3 Comparison of the methods of ALPA and PD for (41)

		ALPA		PD		
n	$\lambda/\lambda_{\max}$	obj	time	obj	time	Imp
225	1	8.6561e-05	1.0951	2.2917e-04	7.0730	0.6994(10)
	0.5	6.5802e-05	1.0422	2.1266e-04	7.0397	0.7233(10)
	0.25	4.9246e-05	1.0979	2.1177e-04	7.0047	0.7815(10)
	0.1	3.4940e-05	1.2043	2.0509e-04	7.2560	0.8429(10)
	0.01	2.1131e-05	1.3248	1.9283e-04	7.3545	0.9042(10)
	0.001	1.9235e-05	1.3106	1.8453e-04	7.0163	0.9131(10)
457	1	8.0056e-05	2.1413	2.0385e-04	32.7012	0.6490(10)
	0.5	5.2868e-05	2.1001	1.9071e-04	34.5735	0.7457(10)
	0.25	3.5468e-05	2.0741	1.8765e-04	31.1474	0.8269(10)
	0.1	1.8057e-05	2.2807	1.8475e-04	31.2997	0.9128(10)
	0.01	3.1988e-06	1.5409	1.8164e-04	31.2353	0.9840(10)
	0.001	4.0979e-07	1.5607	1.8140e-04	29.4073	0.9979(10)
1319	1	5.3470e-05	30.4594	1.7935e-04	243.3121	0.7279(10)
	0.5	3.6469e-05	30.6705	1.4865e-04	218.8169	0.8040(10)
	0.25	2.3588e-05	26.2088	1.5329e-04	204.9859	0.8548(10)
	0.1	1.4241e-05	19.1274	1.2385e-04	181.4983	0.8991(10)
	0.01	5.7736e-06	8.4926	1.0916e-04	174.5865	0.9338(10)
	0.001	2.9876e-06	8.1691	1.0351e-04	164.4820	0.9618(10)
2152	1	6.0273e-05	92.1587	2.1755e-04	426.8685	0.7317(10)
	0.5	3.6670e-05	81.9204	1.8041e-04	405.2515	0.8129(10)
	0.25	2.6168e-05	69.2267	1.7287e-04	402.8543	0.8660(10)
	0.1	1.6373e-05	39.1844	1.7571e-04	401.7480	0.9005(10)
	0.01	5.4977e-06	24.4754	1.6928e-04	390.3046	0.9594(10)
	0.001	3.1789e-06	24.5883	1.6660e-04	386.9057	0.9743(10)

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 ALPA method to the PD method in ten times tests, and the integers in the brackets stand for the numbers of times the ALPA method finds a better suboptimal solution such that imp > 0; "*" stands that the CPLEX solver or the ALPA method finds the optimal solution of the test problem; "-" stands that the CPLEX solver does not find a effective solution and the operation is terminated by the CPLEX solver automatically.

From Table 1 and 3, we observe that the solutions obtained from the ALPA method are often better than the solutions from the PD method and the computation time of the ALPA 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 initial points of the ALPA method to be same, in generally if we choose two different initial points the computation time of the ALPA method will be reduced.

We can see from Table 2 that for the first group of the test problems, the best objective value of the solutions obtained from the ALPA method is approximate to the objective value of the solution obtained from the CPLEX solver and for the first example the ALPA method and the CPLEX solver find

the global minimizer of the test problem, but the computation time of the ALPA method is less than the time of the CPLEX solver. For the other test problems, the best objective value of the solutions and the computation speed of the ALPA method are better than that of the CPLEX solver. It can suggest that the ALPA method is a efficient computing method for problem (1).

6 Conclusion

In this paper we propose an ALPA method for two class of large-scale sparse discrete constrained optimization problems (1) and (2). In solving process, the ALPA method generates a sequence of AL subproblems, each of these subproblems is a nonconvex and nonsmooth problem, and utilizes the PALM method framework and the sparse projection techniques to solve these subproblems. Based on some suitable assumptions, we can establish that any accumulation point of the sequence generated by the ALPA method for problems (1) and (2) satisfies the first-order necessary optimality conditions of problems (1) and (2). In addition, under another assumption it is also a local minimizer of problems (1) and (2). The utilizing of the PALM method and the sparse projection techniques 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. And compare to the standard integer programming method our method can find a KKT point or a local minimizer with a good quality.

References

1. Bertsimas, D., Shioda, R.: *Algorithm for cardinality-constrained quadratic optimization.* Comput. Optim. Appl. 43, 1C22 (2009)
2. Bonami, P., Lejeune, M.: *An exact solution approach for portfolio optimization problems under stochastic and integer constraints.* Oper. Res. 57, 650C670 (2009)
3. Cesarone, F., Scozzari, A., Tardella, F.: *Efficient algorithms for mean-variance portfolio optimization with hard real-world constraints.* G. Ist. Ital. Attuari 72, 37C56 (2009)
4. Cui, X.T., Zheng, X.J., Zhu, S.S., Sun, X.L.: *Convex relaxations and MIQCQP reformulations for a class of cardinality-constrained portfolio selection problems.* J. Glob. Optim. (2012).
5. Li, D., Sun, X., Wang, J.: *Optimal lot solution to cardinality constrained mean-variance formulation for portfolio selection.* Math. Finance 16, 83C101 (2006)
6. Gao, J., Li, D.: *Optimal cardinality constrained portfolio selection.* Oper. Res. 61:745C761.
7. Shaw, D., Liu, S., Kopman, L.: *Lagrangian relaxation procedure for cardinality-constrained portfolio optimization.* Optim. Methods Softw. 23, 411C420 (2008)
8. Bruckstein, A.M., Donoho, D.L., Elad, M.: *From sparse solutions of systems of equations to sparse modeling of signals and images.* SIAM Rev. 51, 34C81 (2009)
9. Bienstock, D.: *Computational study of a family of mixed-integer quadratic programming problems.* Math. Program. 74(2, Ser. A) (1996), pp. 121C140.
10. Miller, A.: *Subset Selection in Regression, 2nd edn.* Chapman & Hall, London (2002)

11. Günlük, O., Linderoth, J.: *Perspective reformulations of mixed integer nonlinear programs with indicator variables*. Math. Program. 124, 183C205 (2010)
12. Frangioni, A., Gentile, C.: *Perspective cuts for a class of convex 0-1 mixed integer programs*. Math. Program. 106, 225C236 (2006)
13. Frangioni, A., Gentile, C.: *Tighter approximated MILP formulations for unit commitment problems*. IEEE Trans. Power Syst. 24, 105C113 (2009)
14. Frangioni, A., Gentile, C.: *SDP diagonalizations and perspective cuts for a class of nonseparable MIQP*. Oper. Res. Lett. 35, 181C185 (2007)
15. Frangioni, A., Gentile, C.: *A computational comparison of reformulations of the perspective relaxation: SOCP vs. cutting planes*. Oper. Res. Lett. 37, 206C210 (2009)
16. Tawarmalani, M., Sahinidis, N.: *Semidefinite relaxations of fractional programs via novel convexification techniques*. J. Glob. Optim. 20, 133C154 (2001)
17. Chang, T., Meade, N., Beasley, J., Sharaiha, Y.: *Heuristics for cardinality constrained portfolio optimisation*. Comput. Oper. Res. 27, 1271C1302 (2000)
18. Crama, Y., Schyns, M.: *Simulated annealing for complex portfolio selection problems*. Eur. J. Oper. Res. 150, 546C571 (2003)
19. Maringer, D., Kellerer, H.: *Optimization of cardinality constrained portfolios with a hybrid local search algorithm*. IEEE Spectr. 25, 481C495 (2003)
20. Schaerf, A.: *Local search techniques for constrained portfolio selection problems*. Comput. Econ. 20, 177C190 (2002)
21. Zheng, X., Sun, X., Li, D., Sun, J.: *Successive convex approximations to cardinality-constrained convex programs: a piecewise-linear DC approach*. Computational Optimization and Applications 59(1-2), 379-397 (2014).
22. Burdakov, O.P., Kanzow, C., Schwartz, A.: *Mathematical programs with cardinality constraints: Reformulation by complementarity-type conditions and a regularization method*. SIAM Journal on Optimization, 2016, 26(1): 397-425.
23. Lu, Z., Zhang, Y.: *Sparse approximation via penalty decomposition methods*. SIAM Journal on Optimization 23(4), 2448-2478 (2013).
24. Bolte, J., Sabach, S., Teboulle, M.: *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*. Mathematical Programming 146(1-2), 459-494 (2014).
25. Teng, Y., Yang, L., Yu, B., Song, X.L.: *A penalty PALM method for sparse portfolio selection problems*. http://dx.doi.org/10.1080/10556788.2016.1204299
26. Bai, Y., Liang, R., Yang, Z.: *Splitting augmented Lagrangian method for optimization problems with a cardinality constraint and semicontinuous variables*. Optimization Methods and Software, 31(5), 1089-1109
27. Rockafellar, R. T., Wets, R.: *Variational Analysis*. Grundlehren der Mathematischen Wissenschaften, 317, Springer, 1998.
28. Chen, X., Guo, L., Lu, Z., Ye J.: *A Feasible Augmented Lagrangian Method for Non-Lipschitz Nonconvex Programming*. SIAM Journal on Numerical Analysis, submitted, 2015.
29. Lu, Z., Zhang, Y.: *An augmented Lagrangian approach for sparse principal component analysis*. Math. Program. 135, 149-193, 2012.
30. Bolte, J., Daniilidis, A., Ley, O., Mazet, L.: *Characterizations of Łojasiewicz inequalities: Subgradient flows, talweg, convexity*. Transactions of the American Mathematical Society, 362 (2010).
31. Bolte, J., Daniilidis, A., Lewis, A.: *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*. SIAM Journal on Optimization 17(4), 1205-1223 (2007).
32. Xu, Y., Yin, W.: *A globally convergent algorithm for nonconvex optimization based on block coordinate update*. arXiv preprint arXiv:1410.1386
33. Beasley, J. E. : *OR-Library: Distributing test problems by electronic mail*. Journal of the Operational Research Society, 41(11): 1069C1072, 1990.