Journal of Computational Mathematics
Vol.xx, No.x, 200x, 1–49.

http://www.global-sci.org/jcm
doi:10.4208/jcm.2207-m2021-0349

AN INEXACT PROXIMAL DC ALGORITHM FOR THE LARGE-SCALE CARDINALITY CONSTRAINED MEAN-VARIANCE MODEL IN SPARSE PORTFOLIO SELECTION*

Mingcai Ding, Xiaoliang Song[1)] and Bo Yu
School of Mathematical Sciences, Dalian University of Technology, Dalian 116081, China
Email: Dingmc_dlut@mail.dlut.edu.cn, songxiaoliang@dlut.edu.cn, yubo@dlut.edu.cn

Abstract

Optimization problem of cardinality constrained mean-variance (CCMV) model for sparse portfolio selection is considered. To overcome the difficulties caused by cardinality constraint, an exact penalty approach is employed, then CCMV problem is transferred into a difference-of-convex-functions (DC) problem. By exploiting the DC structure of the gained problem and the superlinear convergence of semismooth Newton (ssN) method, an inexact proximal DC algorithm with sieving strategy based on a majorized ssN method (siPDCA-mssN) is proposed. For solving the inner problems of siPDCA-mssN from dual, the second-order information is wisely incorporated and an efficient mssN method is employed. The global convergence of the sequence generated by siPDCA-mssN is proved. To solve large-scale CCMV problem, a decomposed siPDCA-mssN (DsiPDCA-mssN) is introduced. To demonstrate the efficiency of proposed algorithms, siPDCA-mssN and DsiPDCA-mssN are compared with the penalty proximal alternating linearized minimization method and the CPLEX (12.9) solver by performing numerical experiments on real-word market data and large-scale simulated data. The numerical results demonstrate that siPDCA-mssN and DsiPDCA-mssN outperform the other methods from computation time and optimal value. The out-of-sample experiments results display that the solutions of CCMV model are better than those of other portfolio selection models in terms of Sharp ratio and sparsity.

Mathematics subject classification: 65K05, 90C06, 90C26, 91G80.
Key words: Sparse portfolio selection, Cardinality constrained mean-variance model, Inexact proximal difference-of-convex-functions algorithm, Sieving strategy, Decomposed strategy.

1. Introduction

The classical Markowitz model [1], also called mean-variance (MV) model, was proposed to find the optimal portfolio selection between different assets in a frictionless market. Based on the MV model, researchers have conducted a large number of studies on the out-of-sample performance and sparsity of the portfolios. On one hand, some researches [2–6] have been carried out to improve the out-of-sample performance of allocation. On the other hand, many researches [7–12] focused on constructing new MV models to find sparse portfolios, which can greatly reduce the administrative and transaction costs. One common class of the approaches for

* Received December 10, 2021 / Revised version received March 16, 2022 / Accepted July 11, 2022 / Published online xxxxxxx xx, 20xx /
1) Corresponding author

obtaining sparse solutions is introducing sparse regularization strategies in the MV model [7–9]. Another popular class of the methods for getting sparse portfolios is introducing cardinality constraint into the MV model [10–12]. The cardinality constrained MV (CCMV) model can be expressed as

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} & F(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \mathbf{Q}\boldsymbol{x} \\ \text{s.t. } & \boldsymbol{e}^\top \boldsymbol{x} - 1 = 0, \\ & r - \boldsymbol{R}^\top \boldsymbol{x} \le 0, \\ & \mathbf{0} \le \boldsymbol{x} \le \boldsymbol{b}, \\ & \|\boldsymbol{x}\|_0 \le K, \end{aligned} \tag{1.1}$$

where $\|\boldsymbol{x}\|_0$ denotes the number of the non-zero entries of $\boldsymbol{x} \in \mathbb{R}^n$ and K is the upper bound of the number of investments. $\mathbf{Q}$ is the covariance matrix of the n assets, which is symmetric positive semidefinite. $\boldsymbol{e} \in \mathbb{R}^n$ is the vector with all components one. $\boldsymbol{R} \in \mathbb{R}^n$ is the vector of the expected return of the n assets. $r \in \mathbb{R}$ is the minimum profit target. The box constraint $\mathbf{0} \le \boldsymbol{x} \le \boldsymbol{b}$ means that the short-selling is prohibited and the investment proportion of i-th asset has an upper bound $\boldsymbol{b}_i$.

The CCMV model in (1.1) belongs to the class of the cardinality constrained quadratic optimization problems, which have been widely studied, see [13–18]. Due to the combinatorial nature of cardinality function, the cardinality constrained optimization problems are NP-hard in most cases [13]. The common optimization methods for the cardinality constrained problem mainly fall into two main categories: integer programming methods [10, 13, 17, 19, 20] and heuristic methods [5, 12, 15, 21]. In the class of integer programming methods, the cardinality constrained problem was reformulated into a mixed integer programming, then the branch-and-bound framework was used to solve it. The heuristic algorithms for solving cardinality constrained problem mainly include genetic algorithm, simulated annealing algorithm and tabu search algorithm. However, both these two kinds of methods are computationally expansive and time consuming, especially when the scale of problem is large.

Recently, by using the alternating iteration method and sparse projection approach, some novel and competitive methods were proposed to solve the cardinality constrained optimization problems. In 2010, Lu et al. [22, 23] proposed a penalty decomposed (PD) method for solving the cardinality constrained problem, in which, the quadratic penalty subproblems were solved by a block coordinate descent method. By making full use of the advantages of PD method and a proximal alternating linearized minimization (PALM) method [24], Teng et al. [25] proposed a penalty proximal alternating linearized minimization (PPALM) method for large-scale sparse portfolio problems. As shown in [25], PPALM method can efficiently find the support set of the local optimal solution of CCMV problem, and PPALM outperforms PD method from computational time and the performance of solutions. However, the parameter associated with the proximal term of PPALM method need to be larger than the Lipschitz constant of penalty subproblem to guarantee its convergence, which would limit its convergence rate severely. In addition, the penalty parameter is difficult to set because the large penalty parameter usually leads to ill-conditioned penalty subproblem.

In addition to the above methods those directly dealing with the cardinality constraint, another common method is to transfer the cardinality constraint problems into the l_0-norm regularization problems. For the l_0-norm regularization optimization problem, many convex/nonconvex relaxation approaches are usually used to approximate the l_0-norm. On one

hand, the convex relaxation of the l_0-norm is one of the most commonly used methods in sparse optimization problems, which gives rise to a convex optimization problem. One of the most successful examples is the compressed sensing, in which, the l_0-norm usually is replaced by the l_1-norm. However, this may reduce the sparsity of the solution and lead to poor performance [7,26]. On the other hand, some nonconvex relaxations of the l_0-norm have been proposed, including the exponential concave function approximation [27], the Capped-l_1 approximation [28], the piecewise linear approximation [29, 30], the smoothly clipped absolute deviation (SCAD) approximation [31] and the logarithm approximation [32]. However, these convex/nonconvex relaxation methods only approximate the cardinality constraint, which cannot be completely equivalent to the original cardinality constrained problems. In order to more precisely deal with the cardinality constraint, a popular nonconvex approach [33] is used to equivalently transfer the cardinality constraint $\|\boldsymbol{x}\|_0 \leq K$ into a DC constraint $\|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_{(K)} = 0$. Hence, the CCMV problem in (1.1) can be equivalently reformulated as

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \; & F(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \mathbf{Q}\boldsymbol{x} \\ \text{s.t.} \; & \boldsymbol{e}^\top \boldsymbol{x} - 1 = 0, \\ & r - \boldsymbol{R}^\top \boldsymbol{x} \leq 0, \\ & \mathbf{0} \leq \boldsymbol{x} \leq \boldsymbol{b}, \\ & \|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_{(K)} = 0. \end{aligned} \tag{1.2}$$

It should be pointed out that such an equivalent transformation does not eliminate the difficulties caused by the cardinality constraint. Fortunately, based on the DC structure of this constraint, we can employ a penalty approach to penalize the DC constraint into the objective function of (1.2). Therefore, a quadratic DC problem can be obtained, which is formulated as

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \; & F_c(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \mathbf{Q}\boldsymbol{x} + c(\|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_{(K)}) \\ \text{s.t.} \; & \boldsymbol{e}^\top \boldsymbol{x} - 1 = 0, \\ & r - \boldsymbol{R}^\top \boldsymbol{x} \leq 0, \\ & \mathbf{0} \leq \boldsymbol{x} \leq \boldsymbol{b}. \end{aligned} \tag{1.3}$$

Theoretically, we can prove that such a penalty approach is an exact penalty method, which guarantees the equivalence of solving (1.3) with solving (1.2) or (1.1). Consequently, (1.3) can be solved under the framework of DC programming approach. For a sparse portfolio selection problem with a feasible set $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{e}^\top \boldsymbol{x} - 1 = 0, \boldsymbol{x} \geq \mathbf{0}, \|\boldsymbol{x}\|_0 \leq K\}$, Gotoh et al. [33] provided an exact penalty result and gave a proximal DC algorithm to solve the gained penalized problem. In their algorithm framework, the inner problems of the proximal DC algorithm need to be solved exactly.

For large-scale DC programming $\min_{\boldsymbol{x}} f_1(\boldsymbol{x}) - f_2(\boldsymbol{x})$, it may be impossible or unnecessary to solve the convex inner problems of a DC algorithm (DCA) exactly. In general, it takes a lot of computation to solve an optimization problem to a high precision by using an iterative algorithm. Hence, it is computationally expensive to solve all the inner problems of DCA to a high precision. In particular, it is unnecessary to solve the inner problems to a high precision at the initial iteration of DCA. To deal with this issue, a popular approach is to consider the inexactness of solution when solving the inner problem of DCA. The inner problem of proximal

DCA in step k can be expressed as

$$\min_{\boldsymbol{x}} G(\boldsymbol{x}) := f_1(\boldsymbol{x}) - \langle \boldsymbol{w}^k, \boldsymbol{x} \rangle + \frac{\alpha}{2}\|\boldsymbol{x} - \boldsymbol{x}^k\|^2,$$

where $\boldsymbol{w}^k \in \partial f_2(\boldsymbol{x}^k)$. Let $\boldsymbol{x}^{k+1}$ be an inexact solution of $\min_{\boldsymbol{x}\in\mathbb{R}^n} G(\boldsymbol{x})$, then there exists an inexact term $\boldsymbol{\Delta}^{k+1}$ such that $\boldsymbol{\Delta}^{k+1} \in \partial G(\boldsymbol{x}^{k+1})$. To ensure the convergence of the general inexact proximal DCA, the inexact solution $\boldsymbol{x}^{k+1}$ and the corresponding inexact term $\boldsymbol{\Delta}^{k+1}$ were required to satisfy the termination condition similar to $\|\boldsymbol{\Delta}^{k+1}\| \leq \eta_k\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\|$ with a given constant $\alpha \geq \eta_k \geq 0$, see [36]. Since $\boldsymbol{\Delta}^{k+1}$ is related to $\boldsymbol{x}^{k+1}$ implicitly, the algorithm for solving the inner problem may be unable to terminate even if the inner problem is solved to the higher precision. Hence, this inexact strategy is difficult to implement in practical application, especially when the problem is nonsmooth. Therefore, it remains a challenging work to design an inexact proximal DCA that can guarantee the theoretical convergence and has good numerical performance for solving the large-scale DC programming.

For solving (1.3), we propose an efficient inexact proximal DC algorithm with sieving strategy (siPDCA) as the outer iterative algorithm of the whole algorithm framework. To ensure that the whole inexact algorithm is convergent and the method for solving inner problems can be terminated, we first use a simple and numerically implementable inexact condition for solving the inner problems and then employ a "sieve" to perform a post-process on the inexact solution. It is essential to efficiently solve the inner problems of siPDCA for the efficiency of whole algorithm framework. By making full use of the semismooth properties in the dual inner problems and the superlinear convergence of the semismooth Newton (ssN) method, it is natural to adopt a majorized ssN (mssN) method to solve the inner problems from the dual. As a result, we introduce an efficient algorithm framework, called siPDCA based on a mssN method (siPDCA-mssN), to solve (1.3). Intuitively, it is computationally expensive to solve the Newton equations. However, we make full use of the second-order sparsity of problem and the sparsity of solutions during deigning mssN method, which makes the computational cost of the whole mssN method comparable to that of a first-order algorithm, or even lower. Moreover, in mssN method, a numerically efficient algorithm is used to compute the Euclidean projection on a set composed of an inequality constraint and n box constraints. We prove that the sequence generated by siPDCA-mssN globally converges to a stationary point of (1.3) and the stationary point is also the local minimizer of the CCMV problem if it satisfies the cardinality constraint.

Based on the violation of first-order optimality conditions, Wang et al. [50] proposed a decomposed strategy to reduce the large-scale sparse optimization problem into a much small problem. Inspired by Wang et al.'s work, we employ an efficient decomposed strategy into siPDCA-mssN for solving the large-scale CCMV problem, and the resulting algorithm is called decomposed siPDCA-mssN (DsiPDCA-mssN). At each iteration of DsiPDCA-mssN, only a small-scale CCMV problem is solved by siPDCA-mssN, which is efficient in terms of computation and storage. We prove that the solution sequence generated by DsiPDCA-mssN subsequentially converges to a local minimizer of the original CCMV problem under some mild assumptions.

The rest of this paper is organized as the follows. In Section 2, the exact penalty property of (1.3) is analyzed. In Section 3, the siPDCA-mssN is proposed for solving (1.3), whose inner problems are solved by the mssN method from the dual. In Section 4, the global convergence of the sequence generated by siPDCA-mssN is proved. In Section 5, the DsiPDCA-mssN is proposed to solve the large-scale CCMV problem. In Section 6, the numerical experiments are performed on some real-word market data sets and large-scale stimulated data sets to display

the numerical effectiveness of siPDCA-mssN and DsiPDCA-mssN for solving (1.1).

Below are some common notations to be used in this paper. We use $\emptyset$ to denote the empty set. $\|\cdot\|$ is used to represent the l_2-norm of vectors and matrices. Let $\|\boldsymbol{x}\|_{(K)}$ be the Ky Fan K-norm of $\boldsymbol{x}$, which is given as $\|\boldsymbol{x}\|_{(K)} = \sum_{i=1}^{K} |\boldsymbol{x}_{(i)}|$ with $|\boldsymbol{x}_{(1)}| \geq \cdots \geq |\boldsymbol{x}_{(n)}|$. We use $\Pi_{\mathcal{C}}(\boldsymbol{x})$ to denote an orthogonal projection of $\boldsymbol{x}$ on set $\mathcal{C}$. The $\mathrm{Prox}_p(\boldsymbol{x})$ is used to denote the proximal projection of function p at $\boldsymbol{x}$. Let $\delta_{\mathcal{C}}(\boldsymbol{x})$ be the indicate function of the set $\mathcal{C}$ at $\boldsymbol{x}$. We use $\lambda_1(\mathbf{Q})$ to denote the largest eigenvalue of positive semidefinite matrix $\mathbf{Q}$. For a given positive integer t, $\mathbf{I}_t$ is used to denotes the identity matrix of size t.

2. Exact Penalty Approach for the Cardinality Constraint Mean-variance Problem

Let $\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{e}^\top \boldsymbol{x} - 1 = 0, r - \boldsymbol{R}^\top \boldsymbol{x} \leq 0, \mathbf{0} \leq \boldsymbol{x} \leq \boldsymbol{b}\}$. In order to ensure that (1.1) is feasible and that all the portfolios of size K have a chance to be selected, we give the following assumption about the set Ω.

Assumption 2.1. *Suppose that* $\min_{1\leq i\leq n} \boldsymbol{b}_i \geq \frac{1}{K}$ *and* $\Omega \neq \emptyset$.

Let $p(\boldsymbol{x}) := \|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_{(K)}$ and $\mathcal{F} := \{\boldsymbol{x} \in \Omega : p(\boldsymbol{x}) = 0\}$. It should be noted that Ω is a bounded set and under Assumption 2.1, both Ω and $\mathcal{F}$ are nonempty compact sets. Since $p(\boldsymbol{x}) \geq 0$, when the value of c is large enough, the penalty term will give a heavy cost for constraint violation $p(\boldsymbol{x}) > 0$. Consequently, the solution of (1.1) can be obtained by iteratively solving (1.3) with increasing penalty parameter c.

Some recent works [33–35] tried to find the exact penalty parameter values for various cardinality constrained problems under some assumptions. Based on the feasible set $\mathcal{F}' = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{e}^\top \boldsymbol{x} - 1 = 0, \boldsymbol{x} \geq \mathbf{0}, \|\boldsymbol{x}\|_0 \leq K\}$ and the Lipschitz smoothness of objective function, an exact penalty result for the sparse portfolio selection problem was given in [33]. For the additional constraints in feasible set $\mathcal{F}$, the constraint $\boldsymbol{R}^\top \boldsymbol{x} \geq r$ ensures that the optimal portfolio satisfies a minimum profit target and the constraints $\boldsymbol{x} \leq \boldsymbol{b}$ require that the investment proportion of each asset is bounded above. Thus these constraints are meaningful for portfolio selection. However, these additional constraints make the theoretical analysis of the exact penalty of our model more difficult than that in [33].

Following from [34, Theorem 2 and Proposition 1], we give the following two properties about the error bound of $\mathrm{dist}(\boldsymbol{x}, \mathcal{F})$ for any $\boldsymbol{x} \in \Omega$ by using the compactness of $\mathcal{F}$ and Assumption 2.1.

Proposition 2.1. *For any* $\epsilon > 0$ *and* $\boldsymbol{z} \in \mathcal{F}$, *there exists a constant* $\rho(\boldsymbol{z}, \epsilon) > 0$ *such that*

$$\mathrm{dist}(\boldsymbol{x}, \mathcal{F}) \leq \rho(\boldsymbol{z}, \epsilon) p(\boldsymbol{x}), \ \forall\, \boldsymbol{x} \in \mathcal{B}(\boldsymbol{z}, \epsilon) \cap \Omega. \tag{2.1}$$

Proof. Let $\overline{\boldsymbol{x}}$ be any point in $\mathcal{B}(\boldsymbol{z}, \epsilon) \cap \Omega$. If $\overline{\boldsymbol{x}} \in \mathcal{F}$, then $\mathrm{dist}(\overline{\boldsymbol{x}}, \mathcal{F}) = 0$ and $p(\overline{\boldsymbol{x}}) = 0$, which implies that (2.1) holds. Next, we prove that there exists a constant $\rho(\boldsymbol{z}, \epsilon)$ such that $\mathrm{dist}(\overline{\boldsymbol{x}}, \mathcal{F}) \leq \rho(\boldsymbol{z}, \epsilon) p(\overline{\boldsymbol{x}})$, $\forall\, \overline{\boldsymbol{x}} \in [\mathcal{B}(\boldsymbol{z}, \epsilon) \cap \Omega] \setminus \mathcal{F}$.

Since $\overline{\boldsymbol{x}} \in [\mathcal{B}(\boldsymbol{z}, \epsilon) \cap \Omega] \setminus \mathcal{F}$, then $p(\overline{\boldsymbol{x}}) > 0$ and $\mathrm{dist}(\overline{\boldsymbol{x}}, \mathcal{F}) > 0$. For any $\tau > 1$, consider the following function:

$$h(\boldsymbol{x}) = p(\boldsymbol{x}) + \tau \frac{p(\overline{\boldsymbol{x}})}{\mathrm{dist}(\overline{\boldsymbol{x}}, \mathcal{F})} \|\boldsymbol{x} - \overline{\boldsymbol{x}}\|, \quad \boldsymbol{x} \in \Omega.$$

Notice that $h(\boldsymbol{x})$ is lower semicontinuous on Ω, lower bounded and coercive ($\lim_{\|\boldsymbol{x}\|\to+\infty} h(\boldsymbol{x}) = +\infty$). Hence, we can suppose that an optimal solution of $\min_{\boldsymbol{x}\in\Omega} h(\boldsymbol{x})$ is obtained at $\boldsymbol{y} \in \Omega$, i.e.,

$$\begin{aligned} h(\boldsymbol{y}) =& p(\boldsymbol{y}) + \tau\frac{p(\overline{\boldsymbol{x}})}{\mathrm{dist}(\overline{\boldsymbol{x}}, \mathcal{F})}\|\boldsymbol{y} - \overline{\boldsymbol{x}}\| \le h(\boldsymbol{x}) \\ =& p(\boldsymbol{x}) + \tau\frac{p(\overline{\boldsymbol{x}})}{\mathrm{dist}(\overline{\boldsymbol{x}}, \mathcal{F})}\|\boldsymbol{x} - \overline{\boldsymbol{x}}\|, \quad \forall\, \boldsymbol{x} \in \Omega. \end{aligned} \tag{2.2}$$

This implies that

$$p(\boldsymbol{y}) \le p(\boldsymbol{x}) + \tau\frac{p(\overline{\boldsymbol{x}})}{\mathrm{dist}(\overline{\boldsymbol{x}}, \mathcal{F})}\|\boldsymbol{x} - \boldsymbol{y}\|, \quad \forall\, \boldsymbol{x} \in \Omega. \tag{2.3}$$

Since $p(\boldsymbol{y}) \ge 0$, by setting $\boldsymbol{x} = \overline{\boldsymbol{x}}$ in (2.2), it holds that

$$\tau\frac{p(\overline{\boldsymbol{x}})}{\mathrm{dist}(\overline{\boldsymbol{x}}, \mathcal{F})}\|\boldsymbol{y} - \overline{\boldsymbol{x}}\| \le p(\overline{\boldsymbol{x}}).$$

This, together with $\tau > 1$, implies that $\|\boldsymbol{y} - \overline{\boldsymbol{x}}\| < \mathrm{dist}(\overline{\boldsymbol{x}}, \mathcal{F}) \le \epsilon$ and $\boldsymbol{y} \in [\mathcal{B}(\boldsymbol{z}, 2\epsilon) \cap \Omega] \setminus \mathcal{F}$. Consequently, it holds that $p(\boldsymbol{y}) = \|\boldsymbol{y}\|_1 - \|\boldsymbol{y}\|_{(K)} > 0$. Let $\bar{i}$ and $\bar{j}$ be the indices of the K-th and $K+1$-st largest elements of $\boldsymbol{y}$, respectively. From $p(\boldsymbol{y}) > 0$ and Assumption 2.1, we have

$$0 < \boldsymbol{y}_{\bar{j}} \le \boldsymbol{y}_{\bar{i}} < \frac{1}{K} \le \min(\boldsymbol{b}_{\bar{i}}, \boldsymbol{b}_{\bar{j}}), \tag{2.4}$$

where the third inequality is due to $\boldsymbol{e}^\top\boldsymbol{y} = 1, \boldsymbol{y} \ge 0$ and $\boldsymbol{y}_{\bar{j}} > 0$. Let $\boldsymbol{e}_{\bar{i}}$ and $\boldsymbol{e}_{\bar{j}}$ be the unit vectors in the corresponding coordinate directions. We shall distinguish the following two cases.

(1) When $\boldsymbol{R}_{\bar{i}} \ge \boldsymbol{R}_{\bar{j}}$, let $\widehat{\boldsymbol{y}} = \boldsymbol{y} - t(\boldsymbol{e}_{\bar{j}} - \boldsymbol{e}_{\bar{i}})$, where t is a sufficiently small positive number such that $0 < t \le \boldsymbol{y}_{\bar{j}}$ and $\boldsymbol{y}_{\bar{i}} + t \le \boldsymbol{b}_{\bar{i}}$, then we have $\widehat{\boldsymbol{y}} \in \Omega$ and

$$\|\widehat{\boldsymbol{y}} - \boldsymbol{y}\| = \sqrt{2}t,\ p(\boldsymbol{y}) - p(\widehat{\boldsymbol{y}}) = \|\widehat{\boldsymbol{y}}\|_{(K)} - \|\boldsymbol{y}\|_{(K)} = t > 0.$$

Since $\tau > 1$ is arbitrary, by setting $\boldsymbol{x} = \widehat{\boldsymbol{y}}$ in (2.3), it holds that $\mathrm{dist}(\overline{\boldsymbol{x}}, \mathcal{F}) \le \sqrt{2}p(\overline{\boldsymbol{x}})$.

(2) When $\boldsymbol{R}_{\bar{i}} < \boldsymbol{R}_{\bar{j}}$, let $\widehat{\boldsymbol{y}} = \boldsymbol{y} - (\boldsymbol{y}_{\bar{i}} - t)(\boldsymbol{e}_{\bar{i}} - \boldsymbol{e}_{\bar{j}})$, where t is a positive number such that $0 < t < \boldsymbol{y}_{\bar{j}}$ and $\boldsymbol{y}_{\bar{i}} + \boldsymbol{y}_{\bar{j}} - t \le \boldsymbol{b}_{\bar{j}}$, then we have $\widehat{\boldsymbol{y}} \in \Omega$ and

$$\|\widehat{\boldsymbol{y}} - \boldsymbol{y}\| = \sqrt{2}(\boldsymbol{y}_{\bar{i}} - t),\ p(\boldsymbol{y}) - p(\widehat{\boldsymbol{y}}) = \|\widehat{\boldsymbol{y}}\|_{(K)} - \|\boldsymbol{y}\|_{(K)} = \boldsymbol{y}_{\bar{j}} - t > 0.$$

Thus, by setting $\boldsymbol{x} = \widehat{\boldsymbol{y}}$ in (2.3), from the arbitrariness of $\tau > 1$, it holds that $\mathrm{dist}(\overline{\boldsymbol{x}}, \mathcal{F}) \le \sqrt{2}\frac{\boldsymbol{y}_{\bar{i}} - t}{\boldsymbol{y}_{\bar{j}} - t}p(\overline{\boldsymbol{x}})$. Therefore, by setting $\rho(\boldsymbol{z}, \epsilon) = \max(\sqrt{2}, \sqrt{2}\frac{\boldsymbol{y}_{\bar{i}} - t}{\boldsymbol{y}_{\bar{j}} - t})$, we obtain that (2.1) holds. This completes the proof. □

Proposition 2.2. *There exists a constant $\rho > 0$ such that for any $\boldsymbol{x} \in \Omega$,*

$$\mathrm{dist}(\boldsymbol{x}, \mathcal{F}) \le \rho p(\boldsymbol{x}). \tag{2.5}$$

Proof. By the compactness of $\mathcal{F}$, there exist finite points $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N$ in $\mathcal{F}$ and $\epsilon(\boldsymbol{z}_i) > 0, i = 1, \cdots, N$ such that

$$\mathcal{F} \subset \bigcup_{i=1}^{N} \mathcal{B}(\boldsymbol{z}_i, \epsilon(\boldsymbol{z}_i)/2).$$

From Proposition 2.1, it holds that for each $i \in \{1, \cdots, N\}$, there exists a constant $\rho(\boldsymbol{z}_i, \epsilon(\boldsymbol{z}_i))$ such that

$$\operatorname{dist}(\boldsymbol{x}, \mathcal{F}) \leq \rho(\boldsymbol{z}_i, \epsilon(\boldsymbol{z}_i))p(\boldsymbol{x}), \quad \forall\, \boldsymbol{x} \in \mathcal{B}(\boldsymbol{z}_i, \epsilon(\boldsymbol{z}_i)) \cap \Omega. \tag{2.6}$$

Let $\epsilon = \min_{1\leq i\leq N} \epsilon(\boldsymbol{z}_i)$ and $\widehat{\rho} = \max_{1\leq i\leq N} \rho(\boldsymbol{z}_i, \epsilon(\boldsymbol{z}_i))$. For any $\boldsymbol{x} \in \Omega$ satisfying $dist(\boldsymbol{x}, \mathcal{F}) \leq \epsilon/2$, there exists a $\boldsymbol{z} \in \mathcal{F}$ such that $\|\boldsymbol{x} - \boldsymbol{z}\| \leq \epsilon/2$. Then there is an index $i \in \{1, \cdots, N\}$ such that $\boldsymbol{z} \in \mathcal{B}(\boldsymbol{z}_i, \epsilon(\boldsymbol{z}_i)/2)$. Consequently, it holds that

$$\|\boldsymbol{x} - \boldsymbol{z}_i\| \leq \|\boldsymbol{x} - \boldsymbol{z}\| + \|\boldsymbol{z} - \boldsymbol{z}_i\| \leq \tfrac{\epsilon}{2} + \tfrac{\epsilon(\boldsymbol{z}_i)}{2} \leq \epsilon(\boldsymbol{z}_i).$$

This, together with (2.6), implies that $\operatorname{dist}(\boldsymbol{x}, \mathcal{F}) \leq \rho(\boldsymbol{z}_i, \epsilon(\boldsymbol{z}_i))p(\boldsymbol{x}) \leq \widehat{\rho}p(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Omega$ satisfying $\operatorname{dist}(\boldsymbol{x}, \mathcal{F}) \leq \epsilon/2$. Next, let $\boldsymbol{x} \in \Omega$ with $\operatorname{dist}(\boldsymbol{x}, \mathcal{F}) > \epsilon/2$. We say that there is a constant $\beta > 0$ such that

$$p(\boldsymbol{x}) \geq \beta, \quad \forall\, \boldsymbol{x} \in \Omega, \operatorname{dist}(\boldsymbol{x}, \mathcal{F}) > \frac{\epsilon}{2}. \tag{2.7}$$

If this is not the case, suppose for contradiction that there is a sequence $\{\boldsymbol{x}_k\}$ such that for each k, $p(\boldsymbol{x}_k) < \beta_k$ with $\beta_k > 0$ and $\lim_{k\to\infty} \beta_k = 0$. By the compactness of Ω, without loss of generality, assume that $\boldsymbol{x}_k$ converges to a $\overline{\boldsymbol{x}} \in \Omega$. From the lower semicontinuity of $p(\boldsymbol{x})$, it holds that $p(\overline{\boldsymbol{x}}) = 0$, i.e., $\overline{\boldsymbol{x}} \in \mathcal{F}$, which yields a contradiction that $\operatorname{dist}(\boldsymbol{x}_k, \mathcal{F}) \leq \|\boldsymbol{x}_k - \overline{\boldsymbol{x}}\| \to 0$. Since Ω is bounded, then $\operatorname{dist}(\cdot, \mathcal{F})$ is bounded above on Ω, i.e., there is a constant $s > 0$ such that $\operatorname{dist}(\boldsymbol{x}, \mathcal{F}) \leq s, \forall\, \boldsymbol{x} \in \Omega$. Hence, we have

$$\operatorname{dist}(\boldsymbol{x}, \mathcal{F}) \leq s \leq \frac{s}{\beta}p(\boldsymbol{x}), \quad \forall\, \boldsymbol{x} \in \Omega, \operatorname{dist}(\boldsymbol{x}, \mathcal{F}) > \frac{\epsilon}{2}. \tag{2.8}$$

By setting $\rho = \max(\widehat{\rho}, \frac{s}{\beta})$, it holds that for any $\boldsymbol{x} \in \Omega$, $\operatorname{dist}(\boldsymbol{x}, \mathcal{F}) \leq \rho p(\boldsymbol{x})$. This completes the proof. □

It should be noted that $F(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \mathbf{Q}\boldsymbol{x}$ is Lipschitz continuous on Ω, based on which, we can give the following global exact penalty property of (1.3) with respect to (1.1).

Theorem 2.1. *Let L be the Lipschitz constant of $F(\boldsymbol{x})$ on Ω. Then for any $c > \rho L$ with ρ same as the one in Proposition 2.2, $\boldsymbol{x}_c \in \Omega$ is a global minimizer of (1.3) with penalty parameter $c > \rho L$ if and only if $\boldsymbol{x}_c$ is also a global minimizer of (1.1).*

Proof. For sufficiency, if $\boldsymbol{x}_c$ is a global minimizer of (1.1), then $p(\boldsymbol{x}_c) = 0$ and $\boldsymbol{x}_c$ is a feasible solution of (1.3). From Proposition 2.2, it holds that $\forall\, \boldsymbol{x} \in \Omega$,

$$\begin{aligned} F_c(\boldsymbol{x}) &= F(\boldsymbol{x}) + cp(\boldsymbol{x}) \\ &\geq F(\boldsymbol{x}) + \rho L p(\boldsymbol{x}) \\ &\geq F(\boldsymbol{x}) + L\|\boldsymbol{x} - \Pi_{\mathcal{F}}(\boldsymbol{x})\| \\ &\geq F(\Pi_{\mathcal{F}}(\boldsymbol{x})) \geq F(\boldsymbol{x}_c) = F_c(\boldsymbol{x}_c), \end{aligned} \tag{2.9}$$

where the first inequality holds from $c > \rho L$ and $p(\boldsymbol{x}) \geq 0$, the second inequality holds from Proposition 2.2 and $\operatorname{dist}(\boldsymbol{x}, \mathcal{F}) = \|\boldsymbol{x} - \Pi_{\mathcal{F}}(\boldsymbol{x})\|$, the third inequality is due to the Lipschitz continuity of $F(\boldsymbol{x})$ on Ω, the last inequality follows from the optimality of $\boldsymbol{x}_c$ to (1.1). These inequalities implies that $\boldsymbol{x}_c$ is also a global minimizer of (1.3).

For necessary, if $\boldsymbol{x}_c \in \Omega$ is a global minimizer of (1.3) with $c > \rho L$. Notice that if $\|\boldsymbol{x}_c\|_0 \leq K$, then we have $\boldsymbol{x}_c \in \mathcal{F}$, $p(\boldsymbol{x}_c) = 0$ and $\forall\, \boldsymbol{x} \in \mathcal{F}$,

$$F(\boldsymbol{x}) = F_c(\boldsymbol{x}) \geq F_c(\boldsymbol{x}_c) = F(\boldsymbol{x}_c),$$

where the first equality follows from $p(\boldsymbol{x}) = 0$, the inequality is due to the optimality of $\boldsymbol{x}_c$ to (1.3). This implies that $\boldsymbol{x}_c$ is also a global minimizer of (1.1) if $\|\boldsymbol{x}_c\|_0 \leq K$. Thus, it is sufficient to prove that $\|\boldsymbol{x}_c\|_0 \leq K$. Suppose that $\|\boldsymbol{x}_c\|_0 > K$, then we have $p(\boldsymbol{x}_c) > 0$. Consequently, it holds that

$$\begin{aligned} F_c(\boldsymbol{x}_c) &= F(\boldsymbol{x}_c) + cp(\boldsymbol{x}_c) \\ &> F(\boldsymbol{x}_c) + \rho L p(\boldsymbol{x}_c) \\ &\geq F(\boldsymbol{x}_c) + L\|\boldsymbol{x}_c - \Pi_{\mathcal{F}}(\boldsymbol{x})\| \\ &\geq F(\Pi_{\mathcal{F}}(\boldsymbol{x})) = F_c(\Pi_{\mathcal{F}}(\boldsymbol{x})), \end{aligned}$$

where the first inequality holds from $c > \rho L$ and $p(\boldsymbol{x}_c) > 0$, the third inequality is due to the Lipschitz continuity of $F(\boldsymbol{x})$ on Ω, the last equality is due to $p(\Pi_{\mathcal{F}}(\boldsymbol{x}_c)) = 0$. These inequalities implies that $F_c(\boldsymbol{x}_c) > F_c(\Pi_{\mathcal{F}}(\boldsymbol{x}_c))$ with $\Pi_{\mathcal{F}}(\boldsymbol{x}_c) \in \mathcal{F} \subset \Omega$, which is contradict to the optimality of $\boldsymbol{x}_c$ to (1.3). Hence, $\|\boldsymbol{x}_c\|_0 \leq K$ and $\boldsymbol{x}_c$ is also a global minimizer of (1.1). This completes the proof. □

By using Proposition 2.2 and Lipschitz continuity of $F(\boldsymbol{x})$ on Ω, we display in the next proposition the local exact penalty property of (1.3) with respect to (1.1).

Theorem 2.2. *Let L be the Lipschitz constant of $F(\boldsymbol{x})$ on Ω. For any $c > \rho L$ with ρ same as the one in Proposition 2.2, then any local minimizer $\overline{\boldsymbol{x}} \in \mathcal{F}$ of (1.1) is also a local minimizer of (1.3) with penalty parameter $c > \rho L$. Conversely, if $\overline{\boldsymbol{x}} \in \Omega$ is a local minimizer of (1.3) with penalty parameter $c > 0$ and $\|\overline{\boldsymbol{x}}\|_0 \leq K$, then $\overline{\boldsymbol{x}}$ is also a local minimizer of (1.1).*

Proof. For the first part, if $\overline{\boldsymbol{x}} \in \mathcal{F}$ is a local minimizer of (1.1), then $p(\overline{\boldsymbol{x}}) = 0$ and

$$F(\overline{\boldsymbol{x}}) \leq F(\boldsymbol{x}), \quad \forall\, \boldsymbol{x} \in \mathcal{B}(\overline{\boldsymbol{x}}, \varepsilon) \cap \mathcal{F}, \varepsilon > 0. \tag{2.10}$$

It should be noted that $p(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_{(K)}$ is continuous, then there must exist $\widehat{\varepsilon} > 0$ such that $p(\boldsymbol{x}) \leq \frac{\varepsilon}{2\rho}$ for all $\boldsymbol{x} \in \mathcal{B}(\overline{\boldsymbol{x}}, \widehat{\varepsilon})$. By setting $0 < \widetilde{\varepsilon} \leq \min(\frac{\varepsilon}{2}, \widehat{\varepsilon})$, it holds that for any $\boldsymbol{x} \in \mathcal{B}(\overline{\boldsymbol{x}}, \widetilde{\varepsilon}) \cap \Omega$, $\Pi_{\mathcal{F}}(\boldsymbol{x}) \in \mathcal{F}$ and

$$\operatorname{dist}(\boldsymbol{x}, \mathcal{F}) = \|\boldsymbol{x} - \Pi_{\mathcal{F}}(\boldsymbol{x})\| \leq \rho p(\boldsymbol{x}) \leq \frac{\varepsilon}{2}, \tag{2.11}$$

where the first inequality follows from Proposition 2.2. This implies that for any $\boldsymbol{x} \in \mathcal{B}(\overline{\boldsymbol{x}}, \widetilde{\varepsilon}) \cap \Omega$,

$$\|\overline{\boldsymbol{x}} - \Pi_{\mathcal{F}}(\boldsymbol{x})\| \leq \|\boldsymbol{x} - \Pi_{\mathcal{F}}(\boldsymbol{x})\| + \|\boldsymbol{x} - \overline{\boldsymbol{x}}\| \leq \frac{\varepsilon}{2} + \widetilde{\varepsilon} \leq \varepsilon,$$

i.e., $\Pi_{\mathcal{F}}(\boldsymbol{x}) \in \mathcal{B}(\overline{\boldsymbol{x}}, \varepsilon) \cap \mathcal{F}$. This, together with Proposition 2.2 and (2.10), yields that for any $\boldsymbol{x} \in \mathcal{B}(\overline{\boldsymbol{x}}, \widetilde{\varepsilon}) \cap \Omega$,

$$\begin{aligned} F_c(\overline{\boldsymbol{x}}) = F(\overline{\boldsymbol{x}}) &\leq F(\Pi_{\mathcal{F}}(\boldsymbol{x})) \\ &\leq F(\boldsymbol{x}) + L\|\boldsymbol{x} - \Pi_{\mathcal{F}}(\boldsymbol{x})\| \\ &\leq F(\boldsymbol{x}) + \rho L p(\boldsymbol{x}) \\ &\leq F(\boldsymbol{x}) + cp(\boldsymbol{x}) = F_c(\boldsymbol{x}), \end{aligned}$$

where the second inequality follows from the Lipschitz continuity of $F(\boldsymbol{x})$ on Ω, the last inequality is due to $c > \rho L$. This implies that $\overline{\boldsymbol{x}}$ is also a local minimizer of (1.3) with $c > \rho L$.

For the second part, if $\overline{\boldsymbol{x}} \in \Omega$ is a local minimizer of (1.3) and $\|\overline{\boldsymbol{x}}\|_0 \leq K$, then $p(\overline{\boldsymbol{x}}) = 0$ and $\overline{\boldsymbol{x}} \in \mathcal{F}$. Consequently, it holds that $\forall\, \boldsymbol{x} \in \mathcal{B}(\overline{\boldsymbol{x}}, \varepsilon) \cap \mathcal{F}$,

$$F(\overline{\boldsymbol{x}}) = F_c(\overline{\boldsymbol{x}}) \leq F_c(\boldsymbol{x}) = F(\boldsymbol{x}),$$

where the inequality follows from the local optimality of $\overline{\boldsymbol{x}}$ for (1.3) with penalty parameter $c > 0$, the last equality is due to $p(\boldsymbol{x}) = 0$. This implies that $\overline{\boldsymbol{x}}$ is also a local minimizer of (1.1). This completes the proof. □

Remark 2.1. Notice that the constant $\rho(\boldsymbol{z}, \epsilon) > 0$ in Proposition 2.1, the upper bound s of $\mathrm{dist}(\boldsymbol{x}, \mathcal{F})$ with $\boldsymbol{x} \in \Omega$ and the constant β in Proposition 2.2 are difficult to estimate, thus it is impractical to directly compute the constant ρ in practical numerical experiment. Based on Theorem 2.1, we employ a strategy of gradually increasing the penalty parameter c. When c is large enough, i.e., $c > \rho L$, the global optimal solution of (1.3) with penalty parameter c is also a global optimal solution of (1.1).

3. An Inexact Proximal DC Algorithm with Sieving Strategy Based on a Majorized Semismooth Newton Method

Since the variable $\boldsymbol{x}$ is required to satisfy $\mathbf{0} \leq \boldsymbol{x} \leq \boldsymbol{b}$, then $\|\boldsymbol{x}\|_1 = \boldsymbol{e}^\top \boldsymbol{x}$. Consequently, the DC problem in (1.3) can be equivalently reformulated as follows:

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \; & J_c(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \mathbf{Q}\boldsymbol{x} + c(\boldsymbol{e}^\top \boldsymbol{x} - \|\boldsymbol{x}\|_{(K)}) \\ \text{s.t.} \; & \boldsymbol{e}^\top \boldsymbol{x} - 1 = 0, \\ & r - \boldsymbol{R}^\top \boldsymbol{x} \leq 0, \\ & \mathbf{0} \leq \boldsymbol{x} \leq \boldsymbol{b}. \end{aligned} \tag{3.1}$$

Remark 3.1. Since $\boldsymbol{e}^\top \boldsymbol{x} = 1$, then it holds that (1.3) and (3.1) have the same optimal solution as the following problem:

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \; & \frac{1}{2}\boldsymbol{x}^\top \mathbf{Q}\boldsymbol{x} - c\|\boldsymbol{x}\|_{(K)} \\ \text{s.t.} \; & \boldsymbol{e}^\top \boldsymbol{x} - 1 = 0, \\ & r - \boldsymbol{R}^\top \boldsymbol{x} \leq 0, \\ & \mathbf{0} \leq \boldsymbol{x} \leq \boldsymbol{b}. \end{aligned} \tag{3.2}$$

Hence, from the perspective of numerical experiments, there is no essential difference between (3.1) and (3.2). However, from the perspective of the theoretical analysis, for the consistency of later convergence analysis, we retain the term $\boldsymbol{e}^\top \boldsymbol{x}$ in objective function of (3.1).

It should be noted that (3.1) can be easily rewritten as a standard unconstrained DC programming by adding the indicator function of the feasible set Ω into its objective function. However, the computational cost of the Euclidean projection $\Pi_\Omega(\boldsymbol{x})$ is expensive. To address this issue, we divided the feasible set Ω into two parts: $\mathcal{D} = \left\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{e}^\top \boldsymbol{x} - 1 = 0\right\}$ and $\mathcal{C} = \left\{\boldsymbol{x} \in \mathbb{R}^n : r - \boldsymbol{R}^\top \boldsymbol{x} \leq 0, \mathbf{0} \leq \boldsymbol{x} \leq \boldsymbol{b}\right\}$. As a result, the Euclidean projection on either the $\mathcal{C}$ or the $\mathcal{D}$ can be efficiently computed. Evidently, the DC problem in (3.1) can be reformulated into a standard DC programming:

$$\min_{\boldsymbol{x}} \; f_1(\boldsymbol{x}) - g_c(\boldsymbol{x}), \tag{3.3}$$

where

$$f_1(\boldsymbol{x}) = \tfrac{1}{2}\boldsymbol{x}^\top \mathbf{Q}\boldsymbol{x} + c\boldsymbol{e}^\top \boldsymbol{x} + \delta_{\mathcal{C}}(\boldsymbol{x}) + \delta_{\mathcal{D}}(\boldsymbol{x}), \quad g_c(\boldsymbol{x}) = c\|\boldsymbol{x}\|_{(K)}.$$

Let

$$f_c(\boldsymbol{x}) = \tfrac{1}{2}\boldsymbol{x}^\top \mathbf{Q}\boldsymbol{x} + c\boldsymbol{e}^\top \boldsymbol{x}.$$

The DC programming (3.3) can be efficiently solved by a proximal DC algorithm (PDCA), the details are shown in Algorithm 3.1.

Algorithm 3.1. Proximal DC algorithm for (3.3)

Step 0. Give $c > 0$, tolerance error $\varepsilon \geq 0$ and proximal parameter $\alpha > 0$. Initialize $\boldsymbol{x}^0 \in \mathbb{R}^n$. Choose $\boldsymbol{\xi}^0 \in \partial g_c(\boldsymbol{x}^0)$. Set $k = 0$.
Step 1. Compute $\boldsymbol{x}^{k+1}$ by solving the following inner problem:

$$\boldsymbol{x}^{k+1} = \arg\min_{\boldsymbol{x}\in\mathcal{D}} G_c^k(\boldsymbol{x}) = f_c(\boldsymbol{x}) + \delta_{\mathcal{C}}(\boldsymbol{x}) - \langle \boldsymbol{x}, \boldsymbol{\xi}^k\rangle + \frac{\alpha}{2}\|\boldsymbol{x} - \boldsymbol{x}^k\|^2, \tag{3.4}$$

Step 2. If $\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\| \leq \varepsilon$, stop and return $\boldsymbol{x}^{k+1}$.
Step 3. Choose $\boldsymbol{\xi}^{k+1} \in \partial g_c(\boldsymbol{x}^{k+1})$. Set $k := k + 1$ and go to **Step 1**.

Remark 3.2. The subdifferential of Ky Fan K-norm at $\boldsymbol{x}^{k+1}$ can be expressed as

$$\partial\|\boldsymbol{x}^{k+1}\|_{(K)} = \operatorname*{argmax}_{s}\left\{\sum_{i=1}^n \boldsymbol{x}_i^{k+1}\boldsymbol{s}_i : \sum_{i=1}^n |\boldsymbol{s}_i| = K, -1 \leq \boldsymbol{s}_i \leq 1, i = 1, \ldots, n\right\},$$

see [33, 37, 38]. Let the elements of $|\boldsymbol{x}^{k+1}|$ be sorted nonincreasing, i.e., $|\boldsymbol{x}_{(1)}^{k+1}| \geq \cdots \geq |\boldsymbol{x}_{(n)}^{k+1}|$, then the subgradient $\boldsymbol{\xi}^{k+1} \in \partial g_c(\boldsymbol{x}^{k+1})$ can be computed by

$$\boldsymbol{\xi}_{(i)}^{k+1} = \begin{cases} c\,\mathrm{sign}(\boldsymbol{x}_{(i)}^{k+1}), & \text{if } i = 1, \cdots, K \\ 0, & \text{if } i = K+1, \cdots, n \end{cases}$$

Since the convex inner problem in (3.4) has no closed-form solution, it is impossible and unnecessary to solve this inner problem exactly by an iterative algorithm in practice. Generally, solving an optimization problem to a high precision is computationally expensive and time-consuming. Therefore, it is unnecessary to solve all inner problems to a high precision. Especially in the initial stage of PDCA, we only need to solve the inner problem to a low or medium precision. In light of this, we take the inexactness of the solution of (3.4) into account. Let $\boldsymbol{x}^{k+1}$ be an inexact solution of (3.4), equivalently, we say $\boldsymbol{x}^{k+1}$ is an optimal solution of the following problem:

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n}\ & f_c(\boldsymbol{x}) + \delta_{\mathcal{C}}(\boldsymbol{x}) - \langle \boldsymbol{x}, \boldsymbol{\xi}^k\rangle + \frac{\alpha}{2}\|\boldsymbol{x} - \boldsymbol{x}^k\|^2 - \langle \Delta_1^{k+1}, \boldsymbol{x}\rangle \\ \text{s.t. }& \boldsymbol{e}^\top \boldsymbol{x} - 1 = \Delta_2^{k+1}, \end{aligned} \tag{3.5}$$

where Δ_1^{k+1} and Δ_2^{k+1} are the inexact terms.

Remark 3.3. The inexact terms $(\Delta_1^{k+1}, \Delta_2^{k+1})$ are unknown before computing the inexact solution $\boldsymbol{x}^{k+1}$.

To introduce an efficient inexact proximal DC algorithm (iPDCA), we need to deal with the following three issues: (1) designing an inexact strategy to ensure the theoretical convergence of the iPDCA; (2) giving an efficient algorithm to solve the inner problem of the iPDCA; (3) finding a strategy to compute or estimate $(\Delta_1^{k+1}, \Delta_2^{k+1})$ so that the inexact strategy can be numerically implemented.

3.1. An effective inexact strategy and algorithm framework

For a given $\epsilon_{k+1} > 0$, if $\max(\|\Delta_1^{k+1}\|, |\Delta_2^{k+1}|) \leq \epsilon_{k+1}$, we say the point $\boldsymbol{x}^{k+1}$ is an ϵ_{k+1}-inexact solution of (3.4). In common inexact algorithms [39, 40], the sequence $\{\epsilon_k\}$ is assumed to satisfy the condition $\sum_{k=0}^{\infty} \epsilon_k < \infty$. Although this inexact strategy is numerically implementable, it cannot guarantee the convergence of the corresponding iPDCA. Generally, to ensure the convergence of iPDCA for (3.3), an inexact condition similar to that of [36, 41] can be given as follows:

$$\max(\|\Delta_1^{k+1}\|, |\Delta_2^{k+1}|) \leq \eta_k \|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\|, \tag{3.6}$$

where $0 \leq \eta_k \leq \alpha$. The general iPDCA for (3.3) is shown in Algorithm 3.2.

It should be noted that the inexact terms $(\Delta_1^{k+1}, \Delta_2^{k+1})$ are implicitly related to the inexact solution $\boldsymbol{x}^{k+1}$. Hence, the inexact condition in (3.6) may be unable to satisfy even if one solves the inner problem in (3.4) to the higher precision. Consequently, the algorithm to solve (3.4) may not be terminated. Thus, the condition in (3.6) is not a good choice of termination criterion for inexactly solving (3.4).

By absorbing the merits of the above two inexact strategies, we first employ

$$\max(\|\Delta_1^{k+1}\|, |\Delta_2^{k+1}|) \leq \epsilon_{k+1} \tag{3.7}$$

as the termination condition for inexactly solving (3.4), where $\{\epsilon_k\}$ satisfies $\lim_{k\to\infty} \epsilon_k = 0$. Evidently, for a suitable $\epsilon_{k+1} > 0$, this termination condition can ensure that the algorithm for solving (3.4) is terminable. Notice that by only using this inexact strategy, it is difficult to guarantee the convergence of the obtained inexact solution sequence $\{\boldsymbol{x}^k\}$ and that the decrement of the corresponding objective function value sequence $\{J_c(\boldsymbol{x}^k)\}$, where $J_c(\boldsymbol{x}^k)$ is the objective function value of (3.1) at $\boldsymbol{x}^k$, computed by

$$J_c(\boldsymbol{x}^k) = \frac{1}{2}(\boldsymbol{x}^k)^\top \mathbf{Q}\boldsymbol{x}^k + c\boldsymbol{e}^\top \boldsymbol{x}^k - c\|\boldsymbol{x}^k\|_{(K)}.$$

To tackle this problem, we employ a 'sieve' to perform a post-processing on the inexact solution satisfying (3.7). It should be noted that a suitable 'sieve' is crucial to the convergence of the iPDCA. As discussed above, the condition in (3.6) can guarantee the convergence of the corresponding iPDCA, although it is not a good termination condition for solving (3.4). Therefore, we choose a sieving condition similar to (3.6) as the 'sieve' to perform a post-processing on the obtained inexact solution rather than the termination criterion for solving (3.4). By applying this post-processing technique, we retain the inexact solution satisfying the sieving conditions and take the inexact solution not satisfying the sieving conditions as the initial point of the next iteration. As a result, we present an inexact proximal DCA with sieving strategy (siPDCA) for solving (3.3). For efficiently solving the dual inner problem of siPDCA, we introduce a majorized semismooth Newton method. Consequently, we give an algorithm framework for solving (3.3), which is called siPDCA based on a mssN method (siPDCA-mssN), see Algorithm 3.3 for more details.

Algorithm 3.2. Inexact proximal DCA for solving (3.3)

Step 0. Give penalty parameter $c > 0$, tolerance error $\varepsilon \geq 0$, proximal parameter $\alpha > 0$, non-negative sequence $\{\eta_k\}$ with $0 \leq \eta_k \leq \alpha$. Initialize $\boldsymbol{x}^0 \in \mathbb{R}^n$. Choose $\boldsymbol{\xi}^0 \in \partial g_c(\boldsymbol{x}^0)$. Set $k = 0$.

Step 1. Compute $\boldsymbol{x}^{k+1}$ by inexactly solving (3.4),

$$\boldsymbol{x}^{k+1} \approx \arg\min_{\boldsymbol{x}\in\mathcal{D}} G_c^k(\boldsymbol{x})$$

such that the inexact solution $\boldsymbol{x}^{k+1}$ and the inexact terms $(\Delta_1^{k+1}, \Delta_2^{k+1})$ satisfy the inexact condition $\max(\|\Delta_1^{k+1}\|, |\Delta_2^{k+1}|) \leq \eta_k \|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\|$.

Step 2. If $\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\| \leq \varepsilon$ holds, stop and return $\boldsymbol{x}^{k+1}$.

Step 3. Choose $\boldsymbol{\xi}^{k+1} \in \partial g_c(\boldsymbol{x}^{k+1})$. Set $k := k + 1$ and go to **Step 1**.

Remark 3.4. In Algorithm 3.3, the inexactness of the equality constraint is considered. Let $\boldsymbol{y}^{k+1}$ be an ϵ_{k+1}-inexact solution. If $\boldsymbol{y}^{k+1}$ and $(\Delta_1^{k+1}, \Delta_2^{k+1})$ satisfy the sieving conditions in (5.1), we say a serious step is performed, otherwise, we say a null step is performed. We call the inexact solution satisfying sieving conditions the stability center. Notice that when the sieving conditions in (5.1) do not hold, only the iteration counter k and ϵ_{k+1} are changed, while the stability center $\boldsymbol{x}^{k+1}$ and $\boldsymbol{\xi}^{k+1}$ remain unchanged. The sieving parameter κ, as a tuning parameter, can balance the efficiency of siPDCA-mssN and the inexactness of the solution for (3.4).

Algorithm 3.3. Inexact proximal DCA with sieving strategy based on a majorized semismooth Newton method for (3.3)

Step 0. Give penalty parameter $c > 0$, tolerance error $\varepsilon \geq 0$, nonnegative nonincreasing sequence $\{\epsilon_k\}$, proximal parameter $\alpha > 0$ and sieving parameter $\kappa \in (0, 1)$. Initialize $\boldsymbol{x}^0 \in \mathbb{R}^n$. Choose $\boldsymbol{\xi}^0 \in \partial g_c(\boldsymbol{x}^0)$. Set $k = 0$.

Step 1. Compute $\boldsymbol{y}^{k+1}$ and the corresponding multiplier $\widetilde{\boldsymbol{u}}^{k+1}$ of the equality constraint by using a mssN method to inexactly solving the inner problem in (3.4) from the dual,

$$\boldsymbol{y}^{k+1} \approx \arg\min_{\boldsymbol{x}\in\mathcal{D}} G_c^k(\boldsymbol{x}),$$

such that inexact condition $\max(\|\Delta_1^{k+1}\|, |\Delta_2^{k+1}|) \leq \epsilon_{k+1}$ holds.
condition $\max(\|\Delta_1^{k+1}\|, |\Delta_2^{k+1}|) \leq \eta_k \|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\|$.

Step 2. If $\|\boldsymbol{y}^{k+1} - \boldsymbol{x}^k\| \leq \varepsilon$ and $\max(\|\Delta_1^{k+1}\|, |\Delta_2^{k+1}|) \leq \varepsilon$ hold, stop and return $\boldsymbol{y}^{k+1}$.

Step 3. If sieving conditions:

$$\begin{aligned} &\|\Delta_1^{k+1}\| \leq (1-\kappa)\frac{\alpha}{2}\|\boldsymbol{y}^{k+1} - \boldsymbol{x}^k\|,\ |\Delta_2^{k+1}| \leq \|\boldsymbol{y}^{k+1} - \boldsymbol{x}^k\| \\ &\text{and} \quad |(\widetilde{\boldsymbol{u}}^{k+1} - \boldsymbol{u}^k)(\boldsymbol{e}^\top \boldsymbol{x}^k - 1)| \leq (1-\kappa)\frac{\alpha}{2}\|\boldsymbol{y}^{k+1} - \boldsymbol{x}^k\|^2 \end{aligned} \tag{3.8}$$

hold, set $\boldsymbol{x}^{k+1} := \boldsymbol{y}^{k+1}$, $\boldsymbol{u}^{k+1} := \widetilde{\boldsymbol{u}}^{k+1}$ and choose $\boldsymbol{\xi}^{k+1} \in \partial g_c(\boldsymbol{x}^{k+1})$; otherwise, set $\boldsymbol{x}^{k+1} := \boldsymbol{x}^k$, $\boldsymbol{u}^{k+1} := \boldsymbol{u}^k$, $\boldsymbol{\xi}^{k+1} := \boldsymbol{\xi}^k$. Set $k := k + 1$ and go to **Step 1**.

Remark 3.5. When ε is set to 0, the following two situations would occur: (1) only finite serious steps are performed in Algorithm 3.3, then infinite null steps are performed or Algorithm 3.3 terminates in finite steps; (2) infinite serious steps are performed in Algorithm 3.3. When $\varepsilon > 0$, to ensure that as many serious steps as possible are performed in Algorithm 3.3, one can adjust the sequence $\{\epsilon_k\}$.

3.2. A majorized semismooth Newton method for (3.4)

The main computation of siPDCA-mssN is in solving the inner problem in (3.4). Thus, it is essential to employ an efficient method to solve (3.4) for the effectiveness of the whole algorithm framework. In CCMV model, the covariance matrix $\mathbf{Q}$ can be naturally rewritten as a factored form, i.e., $\mathbf{Q} = \mathbf{S}\mathbf{S}^\top$, where $\mathbf{S} = \frac{1}{\sqrt{m-1}}\bar{\mathbf{S}}$ and each column of $\bar{\mathbf{S}}$ denotes the yield of the n asserts during the same time period. Evidently, the inner problem in (3.4) can be reformulated as follows:

$$\begin{aligned} &\min_{\boldsymbol{x}\in\mathbb{R}^n} G_c^k(\boldsymbol{x}) = \frac{1}{2}\|\mathbf{S}^\top\boldsymbol{x}\|^2 + c\boldsymbol{e}^\top\boldsymbol{x} + \delta_{\mathcal{C}}(\boldsymbol{x}) - \langle \boldsymbol{x}, \boldsymbol{\xi}^k\rangle + \frac{\alpha}{2}\|\boldsymbol{x}-\boldsymbol{x}^k\|^2 \\ &\text{s.t. } \boldsymbol{e}^\top\boldsymbol{x} - 1 = 0. \end{aligned} \tag{3.9}$$

The equivalent minimization form of the dual problem of (3.9) can be formulated as

$$\min_{(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})\in\mathbb{R}\times\mathbb{R}^m\times\mathbb{R}^n} h_c^k(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}) := \frac{1}{2}\|\boldsymbol{v}\|^2 + \frac{1}{2\alpha}\|\boldsymbol{u}\boldsymbol{e} + \boldsymbol{w} + \mathbf{S}\boldsymbol{v} - \boldsymbol{\Phi}_c^k\|^2 + \delta_{\mathcal{C}}^*(\boldsymbol{w}) + \boldsymbol{u}, \tag{3.10}$$

where $\boldsymbol{\Phi}_c^k = \alpha\boldsymbol{x}^k + \boldsymbol{\xi}^k - c\boldsymbol{e}$. The KKT conditions for solving (3.10) can be given as

$$1 - \boldsymbol{e}^\top(\tfrac{1}{\alpha}(\boldsymbol{\Phi}_c^k - \boldsymbol{u}\boldsymbol{e} - \mathbf{S}\boldsymbol{v} - \boldsymbol{w})) = 0, \tag{3.11}$$

$$\boldsymbol{v} - \mathbf{S}^\top(\tfrac{1}{\alpha}(\boldsymbol{\Phi}_c^k - \boldsymbol{u}\boldsymbol{e} - \mathbf{S}\boldsymbol{v} - \boldsymbol{w})) = \mathbf{0}, \tag{3.12}$$

$$\mathbf{0} \in -\tfrac{1}{\alpha}(\boldsymbol{\Phi}_c^k - \boldsymbol{u}\boldsymbol{e} - \mathbf{S}\boldsymbol{v} - \boldsymbol{w}) + \partial\delta_{\mathcal{C}}^*(\boldsymbol{w}). \tag{3.13}$$

Let $\Theta_c^k(\boldsymbol{u},\boldsymbol{v}) := \inf_{\boldsymbol{w}} h_c^k(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})$ for any $(\boldsymbol{u},\boldsymbol{v}) \in \mathbb{R}\times\mathbb{R}^m$. Consequently, an optimal solution $(\boldsymbol{u}^*,\boldsymbol{v}^*,\boldsymbol{w}^*)$ of (3.10) can be computed simultaneously by

$$(\boldsymbol{u}^*,\boldsymbol{v}^*) = \arg\min_{(\boldsymbol{u},\boldsymbol{v})\in\mathbb{R}\times\mathbb{R}^m} \Theta_c^k(\boldsymbol{u},\boldsymbol{v}),\ \boldsymbol{w}^* = \mathrm{Prox}_{\alpha\delta_{\mathcal{C}}^*}(\boldsymbol{t}_c^k(\boldsymbol{u}^*,\boldsymbol{v}^*)),$$

where $\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{\Phi}_c^k - \boldsymbol{u}\boldsymbol{e} - \mathbf{S}\boldsymbol{v}$. The Moreau Envelope of $\delta_{\mathcal{C}}^*$ at $\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v})$ can be denoted as

$$M_{\delta_{\mathcal{C}}^*}^{\alpha}(\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v})) = \min_{\boldsymbol{w}\in\mathbb{R}^n} \frac{1}{2\alpha}\|\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v}) - \boldsymbol{w}\|^2 + \delta_{\mathcal{C}}^*(\boldsymbol{w}). \tag{3.14}$$

Evidently,

$$\Theta_c^k(\boldsymbol{u},\boldsymbol{v}) = \frac{1}{2}\|\boldsymbol{v}\|^2 + \boldsymbol{u} + M_{\delta_{\mathcal{C}}^*}^{\alpha}(\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v})). \tag{3.15}$$

From the extended Moreau decomposition $\boldsymbol{x} = \lambda\,\mathrm{Prox}_{\lambda^{-1}f^*}(\frac{\boldsymbol{x}}{\lambda}) + \mathrm{Prox}_{\lambda f}(\boldsymbol{x})$, it holds that

$$\mathrm{Prox}_{\alpha\delta_{\mathcal{C}}^*}(\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v})) = \boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v}) - \alpha\Pi_{\mathcal{C}}(\tfrac{1}{\alpha}\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v})).$$

From [42, Theorem 6.60], we obtain that the Moreau Envelope $M_{\delta_{\mathcal{C}}^*}^{\alpha}(\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v}))$ is a continuously differentiable function with gradient

$$\nabla M_{\delta_{\mathcal{C}}^*}^{\alpha}(\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v})) = \frac{1}{\alpha}(\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v}) - \mathrm{Prox}_{\alpha\delta_{\mathcal{C}}^*}(\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v}))) = \Pi_{\mathcal{C}}(\tfrac{1}{\alpha}\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v})).$$

Since $\Pi_{\mathcal{C}}$ is piecewise linear and strongly semismooth, then the KKT equations

$$\nabla\Theta_c^k(\boldsymbol{u},\boldsymbol{v}) = \left[\begin{array}{c} 1-\boldsymbol{e}^\top\Pi_{\mathcal{C}}(\frac{1}{\alpha}\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v})) \\ \boldsymbol{v}-\mathbf{S}^\top\Pi_{\mathcal{C}}(\frac{1}{\alpha}\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v})) \end{array}\right] = \mathbf{0} \tag{3.16}$$

is strongly semismooth. Therefore, we can solve (3.10) under the framework of semismooth Newton (ssN) method.

$\forall\,(\boldsymbol{u},\boldsymbol{v}) \in \mathbb{R}\times\mathbb{R}^m$, we define

$$\widehat{\partial}^2\Theta_c^k(\boldsymbol{u},\boldsymbol{v}) = \left[\begin{array}{cc} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{array}\right] + \frac{1}{\alpha}\mathbf{B}\partial\Pi_{\mathcal{C}}(\tfrac{1}{\alpha}\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v}))\mathbf{B}^\top, \tag{3.17}$$

where $\mathbf{B} = [\boldsymbol{e},\mathbf{S}]^\top$ and $\partial\Pi_{\mathcal{C}}(\frac{1}{\alpha}\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v}))$ is the Clarke subdifferential of $\Pi_{\mathcal{C}}$ at $\frac{1}{\alpha}\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v})$. From Hiriart-Urruty et al. [43], we have

$$\widehat{\partial}^2\Theta_c^k(\boldsymbol{u},\boldsymbol{v})(\boldsymbol{d_u},\boldsymbol{d_v}) = \partial^2\Theta_c^k(\boldsymbol{u},\boldsymbol{v})(\boldsymbol{d_u},\boldsymbol{d_v}), \quad \forall\,(\boldsymbol{d_u},\boldsymbol{d_v}) \in \mathbb{R}\times\mathbb{R}^m,$$

where $\partial^2\Theta_c^k(\boldsymbol{u},\boldsymbol{v})$ is a generalized Hessian of $\Theta_c^k(\boldsymbol{u},\boldsymbol{v})$ at $(\boldsymbol{u},\boldsymbol{v})$. By choosing $\mathbf{U} \in \partial\Pi_{\mathcal{C}}(\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v}))$, we can obtain a matrix

$$\mathbf{V} := \left[\begin{array}{cc} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{array}\right] + \frac{1}{\alpha}\mathbf{B}\mathbf{U}\mathbf{B}^\top. \tag{3.18}$$

Evidently, $\mathbf{V} \in \widehat{\partial}^2\Theta_c^k(\boldsymbol{u},\boldsymbol{v})$, then we have $\mathbf{V} \in \partial^2\Theta_c^k(\boldsymbol{u},\boldsymbol{v})$. Notice that the matrix $\mathbf{V}$ may not be positive definite, then the semismooth Newton equations

$$\mathbf{V}(\boldsymbol{d_u},\boldsymbol{d_v}) = -\nabla\Theta_c^k(\boldsymbol{u},\boldsymbol{v})$$

is not well defined. To address this issue, we consider the majorized optimization problem of (3.10):

$$\min_{(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})\in\mathbb{R}\times\mathbb{R}^m\times\mathbb{R}^n} h_{c,\epsilon}^k(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w};\boldsymbol{u}') := h_c^k(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}) + \frac{\epsilon}{2}|\boldsymbol{u}-\boldsymbol{u}'|^2. \tag{3.19}$$

Let

$$\begin{aligned}\Theta_{c,\epsilon}^k(\boldsymbol{u},\boldsymbol{v};\boldsymbol{u}') &:= \Theta_c^k(\boldsymbol{u},\boldsymbol{v}) + \frac{\epsilon}{2}|\boldsymbol{u}-\boldsymbol{u}'|^2 \\ &= \frac{1}{2}\|\boldsymbol{v}\|^2 + \boldsymbol{u} + M_{\delta_{\mathcal{C}}^*}^{\alpha}(\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v})) + \frac{\epsilon}{2}|\boldsymbol{u}-\boldsymbol{u}'|^2.\end{aligned} \tag{3.20}$$

Clearly, $\Theta_{c,\epsilon}^k(\boldsymbol{u},\boldsymbol{v};\boldsymbol{u}')$ is strongly convex with parameter $\gamma = \min(1,\epsilon)$ and its generalized Hessian is positive definite. Consequently, the corresponding semismooth Newton equations are well defined. It is noted that $\Theta_{c,\epsilon}^k(\boldsymbol{u},\boldsymbol{v}) \to \Theta_c^k(\boldsymbol{u},\boldsymbol{v})$ as $\epsilon \to 0$, then an approximate solution of (3.10) can be obtained by solving (3.19) with a sufficiently small $\epsilon > 0$. Thus we can solve (3.10) by a majorized semismooth Newton (mssN) method, see Algorithm 3.4 for more details.

Algorithm 3.4. Majorized semismooth Newton method for (3.10)

Step 0. Initialize $(\boldsymbol{u}^0, \boldsymbol{v}^0) \in \mathbb{R} \times \mathbb{R}^m$. Give tolerance error $\chi_{k+1} > 0$, majorized parameter $\tau_1, \tau_2 \in (0,1)$, line search parameter $\sigma \in (0,1)$, $\eta \in (0, 1/2)$. Set $j = 1$.

Step 1. Choose $\mathbf{U}^j \in \partial\Pi_{\mathcal{C}}(\frac{1}{\alpha}\boldsymbol{t}_c^k(\boldsymbol{u}^{j-1}, \boldsymbol{v}^{j-1}))$. Let $\mathbf{V}^j$ be the matrix computed by (3.18) and $\epsilon_j = \tau_1 \min(\tau_2, \|\nabla\Theta_c^k(\boldsymbol{u}^{j-1}, \boldsymbol{v}^{j-1})\|)$. Find the solution $(\boldsymbol{d}_{\boldsymbol{u}}^j, \boldsymbol{d}_{\boldsymbol{v}}^j) \in \mathbb{R} \times \mathbb{R}^m$ of the following linear system

$$(\boldsymbol{d}_{\boldsymbol{u}}^j, \mathbf{0})\epsilon_j + \mathbf{V}^j(\boldsymbol{d}_{\boldsymbol{u}}^j, \boldsymbol{d}_{\boldsymbol{v}}^j) + \nabla\Theta_{c,\epsilon_j}^k(\boldsymbol{u}^{j-1}, \boldsymbol{v}^{j-1}; \boldsymbol{u}^{j-1}) = 0. \tag{3.21}$$

Step 2. (Line search) Find the first nonnegative integer t such that the following inequality

$$\begin{aligned} &\Theta_{c,\epsilon_j}^k(\boldsymbol{u}^{j-1} + \sigma^t \boldsymbol{d}_{\boldsymbol{u}}^j, \boldsymbol{v}^{j-1} + \sigma^t \boldsymbol{d}_{\boldsymbol{v}}^j; \boldsymbol{u}^{j-1}) \\ \le &\Theta_{c,\epsilon_j}^k(\boldsymbol{u}^{j-1}, \boldsymbol{v}^{j-1}; \boldsymbol{u}^{j-1}) + \eta\sigma^t \langle \nabla\Theta_{c,\epsilon_j}^k(\boldsymbol{u}^{j-1}, \boldsymbol{v}^{j-1}; \boldsymbol{u}^{j-1}), (\boldsymbol{d}_{\boldsymbol{u}}^j, \boldsymbol{d}_{\boldsymbol{v}}^j)\rangle \end{aligned} \tag{3.22}$$

holds. Set the step size as $l_j = \sigma^t$. Set $\boldsymbol{u}^j = \boldsymbol{u}^{j-1} + l_j \boldsymbol{d}_{\boldsymbol{u}}^j$ and $\boldsymbol{v}^j = \boldsymbol{v}^{j-1} + l_j \boldsymbol{d}_{\boldsymbol{v}}^j$.

Step 3. If $\|\nabla\Theta_c^k(\boldsymbol{u}^j, \boldsymbol{v}^j)\| < \chi_{k+1}$, stop and return $(\boldsymbol{u}^j, \boldsymbol{v}^j)$. Else, set $j := j + 1$ and go to **Step 1**

From primal-dual relation, we can obtain a feasible solution of (3.9) at the j-th step of Algorithm 3.4

$$\boldsymbol{x}^{k,j} = \frac{1}{\alpha}(\boldsymbol{\Phi}_c^k - \boldsymbol{u}^j\boldsymbol{e} - \mathbf{S}\boldsymbol{v}^j - \boldsymbol{w}^j) = \Pi_{\mathcal{C}}(\tfrac{1}{\alpha}\boldsymbol{t}_c^k(\boldsymbol{u}^j, \boldsymbol{v}^j)). \tag{3.23}$$

We measure the accuracy of an approximate optimal solution $(\boldsymbol{u}^j, \boldsymbol{v}^j, \boldsymbol{w}^j)$ for (3.10) by using the residuals of the KKT conditions in (3.11)-(3.13):

$$\boldsymbol{\beta}_1^j := 1 - \boldsymbol{e}^\top(\tfrac{1}{\alpha}(\boldsymbol{\Phi}_c^k - \boldsymbol{u}^j\boldsymbol{e} - \mathbf{S}\boldsymbol{v}^j - \boldsymbol{w}^j)), \tag{3.24}$$

$$\boldsymbol{\beta}_2^j := \boldsymbol{v}^j - \mathbf{S}^\top(\tfrac{1}{\alpha}(\boldsymbol{\Phi}_c^k - \boldsymbol{u}^j\boldsymbol{e} - \mathbf{S}\boldsymbol{v}^j - \boldsymbol{w}^j)), \tag{3.25}$$

$$\boldsymbol{\beta}_3^j := \boldsymbol{w}^j - \mathrm{Prox}_{\alpha\delta_{\mathcal{C}}^*}(\boldsymbol{t}_c^k(\boldsymbol{u}^j, \boldsymbol{v}^j)).$$

For a given accuracy tolerance $\chi_{k+1} > 0$, we terminate Algorithm 3.4 when

$$\max(|\boldsymbol{\beta}_1^j|, \|\boldsymbol{\beta}_2^j\|, \|\boldsymbol{\beta}_3^j\|) \le \chi_{k+1}.$$

By noting that for any $j > 0$,

$$\boldsymbol{w}^j = \mathrm{Prox}_{\alpha\delta_{\mathcal{C}}^*}(\boldsymbol{t}_c^k(\boldsymbol{u}^j, \boldsymbol{v}^j)) = \boldsymbol{t}_c^k(\boldsymbol{u}^j, \boldsymbol{v}^j) - \alpha\Pi_{\mathcal{C}}(\tfrac{1}{\alpha}\boldsymbol{t}_c^k(\boldsymbol{u}^j, \boldsymbol{v}^j)),$$

then it follows that for any $j > 0$, $\boldsymbol{\beta}_3^j = \mathbf{0}$ and $\nabla\Theta_c^k(\boldsymbol{u}^j, \boldsymbol{v}^j) = (\boldsymbol{\beta}_1^j, \boldsymbol{\beta}_2^j)$. Evidently,

$$\max(|\boldsymbol{\beta}_1^j|, \|\boldsymbol{\beta}_2^j\|) \le \sqrt{|\boldsymbol{\beta}_1^j|^2 + \|\boldsymbol{\beta}_2^j\|^2} = \|\nabla\Theta_c^k(\boldsymbol{u}^j, \boldsymbol{v}^j)\|. \tag{3.26}$$

Hence, we use $\|\nabla\Theta_c^k(\boldsymbol{u}^j, \boldsymbol{v}^j)\| < \chi_{k+1}$ as the termination condition of Algorithm 3.4.

Notice that mssN method is used to inexactly solve the inner problems of siPDCA-mssN from the dual. Thus, we should find the tolerance error χ_{k+1} that can ensure $\max(\|\Delta_1^{k+1}\|, |\Delta_2^{k+1}|) \le \epsilon_{k+1}$. In addition, in order to ensure that mssN method for (3.10) is numerically efficient, the following issues need to be addressed: (1) finding an efficient algorithm to compute the Euclidean projection on feasible set $\mathcal{C}$; (2) giving a method to compute the HS-Jacobian matrix of the Euclidean projection $\Pi_{\mathcal{C}}$ at a point $\boldsymbol{t} \in \mathbb{R}^n$; (3) giving an efficient method to solve the linear system in (3.21) by making full use of the second-order sparsity of problem and the sparsity of the solution.

3.3. An efficient strategy for estimating the inexact terms

It should be noted that Algorithm 3.3 is a type of inexact method. Then we need to check the inexact condition in (3.7) and the sieving conditions in (5.1) at each iteration of Algorithm 3.3. Thus it is important to give an efficient strategy to compute or estimate $(\Delta_1^{k+1}, \Delta_2^{k+1})$ for the whole algorithm framework.

From (3.9), it follows that the smooth part of $G_c^k(\boldsymbol{x})$ can be denoted as

$$F_c^k(\boldsymbol{x}) = \frac{1}{2}\|\mathbf{S}^\top \boldsymbol{x}\|^2 + c\boldsymbol{e}^\top \boldsymbol{x} - \langle \boldsymbol{x}, \boldsymbol{\xi}^k \rangle + \frac{\alpha}{2}\|\boldsymbol{x} - \boldsymbol{x}^k\|^2.$$

Evidently, the gradient of $F_c^k(\boldsymbol{x})$ can be computed by

$$\nabla F_c^k(\boldsymbol{x}) = \mathbf{S}\mathbf{S}^\top \boldsymbol{x} + c\boldsymbol{e} - \boldsymbol{\xi}^k + \alpha(\boldsymbol{x} - \boldsymbol{x}^k) = \mathbf{S}\mathbf{S}^\top \boldsymbol{x} + \alpha\boldsymbol{x} - \boldsymbol{\Phi}_c^k.$$

Hence, (3.9) can be formulated as

$$\begin{aligned} &\min_{\boldsymbol{x}\in\mathbb{R}^n} G_c^k(\boldsymbol{x}) = F_c^k(\boldsymbol{x}) + \delta_{\mathcal{C}}(\boldsymbol{x}) \\ &\text{s.t. } \ \boldsymbol{e}^\top \boldsymbol{x} - 1 = 0. \end{aligned} \tag{3.27}$$

If an inexact solution $\boldsymbol{x}^{\text{in}} \in \mathcal{C}$ and a multiplier $\boldsymbol{u}$ of the equality constraint for (3.27) are obtained, then there must exist inexact terms (Δ_1, Δ_2) satisfying the following KKT conditions:

$$\Delta_1 \in \nabla F_c^k(\boldsymbol{x}^{\text{in}}) + \partial\delta_{\mathcal{C}}(\boldsymbol{x}^{\text{in}}) + \boldsymbol{u}\boldsymbol{e}, \tag{3.28}$$

$$\boldsymbol{e}^\top \boldsymbol{x}^{\text{in}} - 1 = \Delta_2. \tag{3.29}$$

According to the Second Prox Theorem [42, Theorem 6.39], it holds that (3.28) is equivalent to

$$\boldsymbol{x}^{\text{in}} = \Pi_{\mathcal{C}}(\boldsymbol{x}^{\text{in}} - \tfrac{1}{\alpha}(\nabla F_c^k(\boldsymbol{x}^{\text{in}}) + \boldsymbol{u}\boldsymbol{e}) + \tfrac{1}{\alpha}\Delta_1). \tag{3.30}$$

Due to the nonsmoothness of $\delta_{\mathcal{C}}$, it is impossible to directly obtain the inexact term Δ_1 from (3.28) and (3.30). To deal with this problem, we introduce an auxiliary variable $\widetilde{\boldsymbol{x}}^{\text{in}} \in \mathcal{C}$,

$$\widetilde{\boldsymbol{x}}^{\text{in}} = \Pi_{\mathcal{C}}(\boldsymbol{x}^{\text{in}} - \tfrac{1}{\alpha}(\nabla F_c^k(\boldsymbol{x}^{\text{in}}) + \boldsymbol{u}\boldsymbol{e})). \tag{3.31}$$

By rearranging the terms, (3.31) can be rewritten as

$$\widetilde{\boldsymbol{x}}^{\text{in}} = \Pi_{\mathcal{C}}(\widetilde{\boldsymbol{x}}^{\text{in}} - \tfrac{1}{\alpha}(\nabla F_c^k(\widetilde{\boldsymbol{x}}^{\text{in}}) + \boldsymbol{u}\boldsymbol{e}) + \tfrac{1}{\alpha}\widetilde{\Delta}_1), \tag{3.32}$$

where $\widetilde{\Delta}_1$ is computed by

$$\widetilde{\Delta}_1 = \alpha(\boldsymbol{x}^{\text{in}} - \widetilde{\boldsymbol{x}}^{\text{in}}) + \nabla F_c^k(\widetilde{\boldsymbol{x}}^{\text{in}}) - \nabla F_c^k(\boldsymbol{x}^{\text{in}}). \tag{3.33}$$

Clearly, the inexact term for the equality constraint can be computed by

$$\widetilde{\Delta}_2 = \boldsymbol{e}^\top \widetilde{\boldsymbol{x}}^{\text{in}} - 1. \tag{3.34}$$

According to the Second Prox Theorem, we can obtain that (3.32) is equivalent to

$$\widetilde{\Delta}_1 \in \nabla F_c^k(\widetilde{\boldsymbol{x}}^{\text{in}}) + \partial\delta_{\mathcal{C}}(\widetilde{\boldsymbol{x}}^{\text{in}}) + u\boldsymbol{e}.$$

This, together with (3.34), implies that $\widetilde{\boldsymbol{x}}^{\text{in}}$ is an inexact solution of (3.27) with inexact terms $(\widetilde{\Delta}_1, \widetilde{\Delta}_2)$.

Thus, for a given solution $\boldsymbol{x}^{\rm in} \in \mathcal{C}$, the inexact solution $\widetilde{\boldsymbol{x}}^{\rm in}$ and the inexact terms $(\widetilde{\Delta}_1, \widetilde{\Delta}_2)$ can be obtained by using the expressions in (3.31), (3.33) and (3.34), respectively. Based on this approach, we can use the following two strategies to compute or estimate $(\Delta_1^{k+1}, \Delta_2^{k+1})$.

(1) At the j-th step of Algorithm 3.4, a feasible solution $\boldsymbol{x}^{k,j}$ defined in (3.23) can be obtained. Evidently, we can compute the inexact solution $\widetilde{\boldsymbol{x}}^{k,j}$ and inexact terms $(\widetilde{\Delta}_1^{k,j}, \widetilde{\Delta}_2^{k,j})$ by using the expressions in (3.31), (3.33) and (3.34), respectively. When $\max(\|\widetilde{\Delta}_1^{k,j}\|, |\widetilde{\Delta}_2^{k,j}|) \leq \epsilon_{k+1}$, we set $\boldsymbol{y}^{k+1} := \widetilde{\boldsymbol{x}}^{k,j}$ and $(\Delta_1^{k+1}, \Delta_2^{k+1}) = (\widetilde{\Delta}_1^{k,j}, \widetilde{\Delta}_2^{k,j})$. This means that $\boldsymbol{y}^{k+1}$ is an ϵ_{k+1}-inexact solution of (3.27). It should be noted that if this strategy is used to compute $\boldsymbol{y}^{k+1}$ and $(\Delta_1^{k+1}, \Delta_2^{k+1})$, we need to compute the additional $\widetilde{\boldsymbol{x}}^{k,j}$ and $(\widetilde{\Delta}_1^{k,j}, \widetilde{\Delta}_2^{k,j})$ at each step of Algorithm 3.4. As a result, this strategy is not very efficient in practical numerical experiments.

(2) Notice that Algorithm 3.4 is a dual algorithm and the KKT residuals $(\boldsymbol{\beta}_1^j, \boldsymbol{\beta}_2^j)$ of the dual problem is used to terminate Algorithm 3.4 in practical numerical experiments. Thus, if an upper bound estimation of $(\|\widetilde{\Delta}_1^{k,j}\|, |\widetilde{\Delta}_2^{k,j}|)$ from $(|\boldsymbol{\beta}_1^j|, \|\boldsymbol{\beta}_2^j\|)$ according to the relationship between $(\widetilde{\Delta}_1^{k,j}, \widetilde{\Delta}_2^{k,j})$ and $(\boldsymbol{\beta}_1^j, \boldsymbol{\beta}_2^j)$ is obtained, we just need to check that this upper bound is smaller than ϵ_{k+1}, instead of directly computing $(\|\widetilde{\Delta}_1^{k,j}\|, |\widetilde{\Delta}_2^{k,j}|)$ at each iteration of Algorithm 3.4. This will further reduce the computational cost of the whole algorithm framework.

From the definitions of $\boldsymbol{x}^{k,j}$, $\boldsymbol{\beta}_1^j$ and $\boldsymbol{\beta}_2^j$ in (3.23), (3.24) and (3.25), respectively, it holds that

$$\boldsymbol{\beta}_1^j := 1 - \boldsymbol{e}^\top \boldsymbol{x}^{k,j}, \quad \boldsymbol{\beta}_2^j := \boldsymbol{v}^j - \mathbf{S}^\top \boldsymbol{x}^{k,j}. \tag{3.35}$$

Since $\widetilde{\boldsymbol{x}}^{k,j}$ is computed by using the expression in (3.31), then we have

$$\begin{aligned} \widetilde{\boldsymbol{x}}^{k,j} &= \Pi_{\mathcal{C}}(\boldsymbol{x}^{k,j} - \tfrac{1}{\alpha}(\nabla F_c^k(\boldsymbol{x}^{k,j}) + \boldsymbol{u}^j \boldsymbol{e})) \\ &= \Pi_{\mathcal{C}}(\boldsymbol{x}^{k,j} - \tfrac{1}{\alpha}(\mathbf{S}\mathbf{S}^\top \boldsymbol{x}^{k,j} + \alpha \boldsymbol{x}^{k,j} - \boldsymbol{\Phi}_c^k + \boldsymbol{u}^j \boldsymbol{e})) \\ &= \Pi_{\mathcal{C}}(\tfrac{1}{\alpha}(\boldsymbol{\Phi}_c^k + \mathbf{S}\boldsymbol{\beta}_2^j - \mathbf{S}\boldsymbol{v}^j + \boldsymbol{u}^j \boldsymbol{e})) \\ &= \Pi_{\mathcal{C}}(\tfrac{1}{\alpha}(\boldsymbol{t}^k(\boldsymbol{u}^j, \boldsymbol{v}^j) + \mathbf{S}\boldsymbol{\beta}_2^j)), \end{aligned} \tag{3.36}$$

where the third equality follows from (3.35), the last equality holds from the definition of $\boldsymbol{t}_c^k(\boldsymbol{u}^j, \boldsymbol{v}^j)$. This, together with the non-expansiveness of the projection operator, implies that

$$\|\widetilde{\boldsymbol{x}}^{k,j} - \boldsymbol{x}^{k,j}\| = \|\Pi_{\mathcal{C}}(\tfrac{1}{\alpha}\boldsymbol{t}^k(\boldsymbol{u}^j, \boldsymbol{v}^j) + \tfrac{1}{\alpha}\mathbf{S}\boldsymbol{\beta}_2^j) - \Pi_{\mathcal{C}}(\tfrac{1}{\alpha}\boldsymbol{t}^k(\boldsymbol{u}^j, \boldsymbol{v}^j))\| \leq \frac{1}{\alpha}\|\mathbf{S}\boldsymbol{\beta}_2^j\|. \tag{3.37}$$

Let

$$\chi_{k+1} = \tfrac{\epsilon_{k+1}}{p}, \quad p = \max(\tfrac{1}{\alpha}\|\mathbf{S}\mathbf{S}^\top\|\|\mathbf{S}\|, 1 + \tfrac{1}{\alpha}\sqrt{n}\|\mathbf{S}\|).$$

From (3.26) and (3.36), it holds that if $\max(|\boldsymbol{\beta}_1^j|, \|\boldsymbol{\beta}_2^j\|) \leq \|\nabla\Theta_c^k(\boldsymbol{u}^j, \boldsymbol{v}^j)\| \leq \chi_{k+1}$,

$$\|\widetilde{\Delta}_1^{k,j}\| = \|\mathbf{S}\mathbf{S}^\top(\widetilde{\boldsymbol{x}}^{k,j} - \boldsymbol{x}^{k,j})\| \leq \frac{1}{\alpha}\|\mathbf{S}\mathbf{S}^\top\|\|\mathbf{S}\|\|\boldsymbol{\beta}_2^j\| \leq \epsilon_{k+1}$$

and

$$\begin{aligned} |\widetilde{\Delta}_2^{k,j}| &= \|\boldsymbol{e}^\top \widetilde{\boldsymbol{x}}^{k,j} - 1\| \leq \|\boldsymbol{e}^\top \boldsymbol{x}^{k,j} - 1\| + \|\boldsymbol{e}^\top(\boldsymbol{x}^{k,j} - \widetilde{\boldsymbol{x}^{k,j}})\| \\ &\leq |\boldsymbol{\beta}_1^j| + \tfrac{1}{\alpha}\sqrt{n}\|\mathbf{S}\|\|\boldsymbol{\beta}_2^j\| \leq \epsilon_{k+1}. \end{aligned}$$

This implies that we only need to check if $\|\nabla\Theta_c^k(\boldsymbol{u}^j, \boldsymbol{v}^j)\| \leq \chi_{k+1}$ holds at each iteration of Algorithm 3.4. If $\|\nabla\Theta_c^k(\boldsymbol{u}^j, \boldsymbol{v}^j)\| \leq \chi_{k+1}$ holds, which yields that $\max(\|\widetilde{\Delta}^{k,j}\|, |\widetilde{\Delta}_2^{k,j}|) \leq \epsilon_{k+1}$, then we set $\boldsymbol{y}^{k+1} := \widetilde{\boldsymbol{x}}^{k,j}$ and $(\Delta_1^{k+1}, \Delta_2^{k+1}) := (\widetilde{\Delta}_1^{k,j}, \widetilde{\Delta}_2^{k,j})$. In this strategy, we only need to compute $\widetilde{\boldsymbol{x}}^{k,j}$ and $(\widetilde{\Delta}_1^{k,j}, \widetilde{\Delta}_2^{k,j})$ once, which further reduces computational cost. Thus, it is more efficient than the first strategy in practical numerical experiments, although it may overestimate the inexactness of $\boldsymbol{y}^{k+1}$. Therefore, we adopt this strategy to compute the inexact solution $\boldsymbol{y}^{k+1}$ and to estimate the inexact terms $(\Delta_1^{k+1}, \Delta_2^{k+1})$.

3.4. Euclidean projection on the set $\mathcal{C}$

At each iteration of mssN method, we need to compute the Euclidean projection $\Pi_{\mathcal{C}}$. Thus, it is important to efficiently compute $\Pi_{\mathcal{C}}$ for the efficiency of the whole algorithm framework. The Euclidean projection of a point $\boldsymbol{t}$ on the set $\mathcal{C}$ can be formulated as the following optimization problem:

$$\begin{aligned} \min_{\boldsymbol{z}\in\mathbb{R}^n} \ & \frac{1}{2}\|\boldsymbol{z}-\boldsymbol{t}\|^2 \\ \text{s.t. } & r-\boldsymbol{R}^\top\boldsymbol{z}\le 0, \\ & \boldsymbol{0}\le\boldsymbol{z}\le\boldsymbol{b}. \end{aligned} \tag{3.38}$$

Evidently, (3.38) can be equivalently rewritten as

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} \quad \frac{1}{2}\|\boldsymbol{z}-\boldsymbol{t}\|^2 \tag{3.39}$$

$$\text{s.t.} \quad r-\boldsymbol{R}^\top\boldsymbol{z}\le 0 \qquad \longrightarrow \lambda, \tag{3.40}$$

$$-\boldsymbol{z}\le\boldsymbol{0} \qquad \longrightarrow \boldsymbol{\nu}, \tag{3.41}$$

$$\boldsymbol{z}-\boldsymbol{b}\le\boldsymbol{0} \qquad \longrightarrow \boldsymbol{\mu}, \tag{3.42}$$

where λ, $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ are the Lagrange multipliers of (3.40), (3.41) and (3.42), respectively. Then the KKT conditions of (3.39)-(3.42) can be given as

$$\boldsymbol{z}-\boldsymbol{t}-\lambda\boldsymbol{R}-\boldsymbol{\nu}+\boldsymbol{\mu}=0, \tag{3.43}$$

$$\lambda(r-\boldsymbol{R}^\top\boldsymbol{z})=0, \tag{3.44}$$

$$-\boldsymbol{\nu}^\top\boldsymbol{z}=\boldsymbol{0}, \tag{3.45}$$

$$\boldsymbol{\mu}^\top(\boldsymbol{z}-\boldsymbol{b})=\boldsymbol{0}, \tag{3.46}$$

$$\boldsymbol{\mu}_i\ge 0, \ \ \boldsymbol{\nu}_i\ge 0, \quad i=1,\cdots,n, \tag{3.47}$$

$$\lambda\ge 0. \tag{3.48}$$

To find the optimal solutions $(\boldsymbol{z}^*,\lambda^*,\boldsymbol{\mu}^*,\boldsymbol{\nu}^*)$ satisfying (3.40)-(3.48), we consider the following equivalent minimization form of the dual of (3.38):

$$\min_{(\lambda,\boldsymbol{\eta})\in\mathbb{R}\times\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{t}+\lambda\boldsymbol{R}-\boldsymbol{\eta}\|^2-r\lambda+\delta^*_{[\boldsymbol{0},\boldsymbol{b}]}(\boldsymbol{\eta})+\delta_{\mathbb{R}_+}(\lambda). \tag{3.49}$$

Let $\varphi(\lambda)$ be defined as

$$\begin{aligned} \varphi(\lambda) &= \inf_{\boldsymbol{\eta}} \frac{1}{2}\|\boldsymbol{t}+\lambda\boldsymbol{R}-\boldsymbol{\eta}\|^2-r\lambda+\delta^*_{[\boldsymbol{0},\boldsymbol{b}]}(\boldsymbol{\eta})+\delta_{\mathbb{R}_+}(\lambda) \\ &= \frac{1}{2}\|\boldsymbol{t}+\lambda\boldsymbol{R}-\mathrm{Prox}_{\delta^*_{[\boldsymbol{0},\boldsymbol{b}]}}(\boldsymbol{t}+\lambda^*\boldsymbol{R})\|^2+\delta^*_{[\boldsymbol{0},\boldsymbol{b}]}(\mathrm{Prox}_{\delta^*_{[\boldsymbol{0},\boldsymbol{b}]}}(\boldsymbol{t}+\lambda^*\boldsymbol{R}))-r\lambda+\delta_{\mathbb{R}_+}(\lambda). \end{aligned} \tag{3.50}$$

Then the optimal solutions $(\lambda^*,\boldsymbol{\eta}^*)$ of (3.49) can be simultaneously computed by

$$\lambda^*=\arg\min_{\lambda\in\mathbb{R}}\varphi(\lambda), \tag{3.51}$$

$$\boldsymbol{\eta}^*=\mathrm{Prox}_{\delta^*_{[\boldsymbol{0},\boldsymbol{b}]}}(\boldsymbol{t}+\lambda^*\boldsymbol{R})=\boldsymbol{t}+\lambda^*\boldsymbol{R}-\Pi_{[\boldsymbol{0},\boldsymbol{b}]}(\boldsymbol{t}+\lambda^*\boldsymbol{R}), \tag{3.52}$$

where the projection in (3.52) can be computed by $\Pi_{[\boldsymbol{0},\boldsymbol{b}]}(\boldsymbol{x})=\max(\min(\boldsymbol{x},\boldsymbol{b}),\boldsymbol{0})$. The optimal solution λ^* of (3.51) satisfies the following KKT condition:

$$0\in\boldsymbol{R}^\top\Pi_{[\boldsymbol{0},\boldsymbol{b}]}(\boldsymbol{t}+\lambda^*\boldsymbol{R})-r+\partial\delta_{\mathbb{R}_+}(\lambda^*),$$

which is equivalent to

$$\begin{cases} \lambda^* \geq 0, \\ r - \boldsymbol{R}^\top \Pi_{[\boldsymbol{0},\boldsymbol{b}]}(\boldsymbol{t} + \lambda^* \boldsymbol{R}) \leq 0, \\ \lambda^*(r - \boldsymbol{R}^\top \Pi_{[\boldsymbol{0},\boldsymbol{b}]}(\boldsymbol{t} + \lambda^* \boldsymbol{R})) = 0. \end{cases} \tag{3.53}$$

Let $\omega(\lambda) = r - \boldsymbol{R}^\top \Pi_{[\boldsymbol{0},\boldsymbol{b}]}(\boldsymbol{t} + \lambda \boldsymbol{R})$. Since for any $i \in \{1, \cdots, n\}$, $\max(\min(\boldsymbol{t}_i + \lambda \boldsymbol{R}_i, \boldsymbol{b}_i), 0)$ is a piecewise linear function and $-\boldsymbol{R}_i \max(\min(\boldsymbol{t}_i + \lambda \boldsymbol{R}_i, \boldsymbol{b}_i), 0)$ is nonincreasing, then it follows that $\omega(\lambda)$ is a piecewise linear nonincreasing function with breakpoints in the following set:

$$\mathcal{H} = \left\{ \tfrac{\boldsymbol{t}_i - \boldsymbol{b}_i}{\boldsymbol{R}_i}, -\tfrac{\boldsymbol{t}_i}{\boldsymbol{R}_i} \middle| \boldsymbol{R}_i \neq 0, i = 1, \ldots, n \right\}.$$

In addition, the range of $\omega(\lambda)$ is a closed and bounded interval $\left[r - [\boldsymbol{R}]_+^\top \boldsymbol{b}, r - [\boldsymbol{R}]_-^\top \boldsymbol{b}\right]$. Thus, we compute the optimal solution λ^* satisfying the conditions in (3.53) by binary search among the breakpoints in $\mathcal{H} \cup \{0\}$, see Algorithm 3.5 for more details.

Algorithm 3.5. Euclidean projection on set $\mathcal{C}$

Step 0. Give r, $\boldsymbol{R}$, $\boldsymbol{b}$, $\boldsymbol{t}$. If $r - [\boldsymbol{R}]_+^\top \boldsymbol{b} > 0$, stop with no feasible solution.
Step 1. If $\omega(0) = r - \boldsymbol{R}^\top \Pi_{[\boldsymbol{0},\boldsymbol{b}]}(\boldsymbol{t}) \leq 0$, return $\lambda^* = 0$ and $\boldsymbol{z}^* = \Pi_{[\boldsymbol{0},\boldsymbol{b}]}(\boldsymbol{t})$.
Step 2. Let $\mathcal{H}_+ = \{h_i \in \mathcal{H} \cup \{0\} : h_i \geq 0\}$ and $s = |\mathcal{H}_+|$. Sort all break points in $\mathcal{H}_+$ in ascending order, i.e., $0 = h_{(1)} \leq h_{(2)} \leq \cdots \leq h_{(s)}$. If $\omega(h_{(s)}) = 0$, stop and return $\lambda^* = h_{(s)}$ and $\boldsymbol{z}^* = \Pi_{[\boldsymbol{0},\boldsymbol{b}]}(\boldsymbol{t} + \lambda^* \boldsymbol{R})$.
Step 3. Let $I_l = 1$, $I_r = s$, $\omega_l = \omega(0)$, $\omega_r = \omega(h_{(I_r)})$.
while $I_r - I_l > 1$ *and* $h_{(I_r)} > h_{(I_l)}$ **do**
  $I_m = \lfloor \frac{I_l + I_r}{2} \rfloor$, $\omega_m = \omega(h_{(I_m)})$.
  if $\omega_m > 0$ **then**
    $I_r = I_m$, $\omega_r = \omega_m$.
  else if $\omega_m = 0$ **then**
    Break.
  else
    $I_l = I_m$, $\omega_l = \omega_m$.
  end if
end while
If $\omega_m = 0$, set $\lambda^* = h_{(I_m)}$. Else, set $\lambda^* = h_{(I_l)} - \frac{\omega_l}{\omega_r - \omega_l}(h_{(I_r)} - h_{(I_l)})$. Return λ^* and $\boldsymbol{z}^* = \Pi_{[\boldsymbol{0},\boldsymbol{b}]}(\boldsymbol{t} + \lambda^* \boldsymbol{R})$.

From primal-dual relation, we can obtain the optimal solution $\boldsymbol{z}^*$ of (3.38):

$$\boldsymbol{z}^* = \boldsymbol{t} + \lambda^* \boldsymbol{R} - \boldsymbol{\eta}^* = \Pi_{[\boldsymbol{0},\boldsymbol{b}]}(\boldsymbol{t} + \lambda^* \boldsymbol{R}). \tag{3.54}$$

By using (3.43), (3.45), (3.47) and (3.54), it holds that for each $i \in \{1, \cdots, n\}$, if $(\boldsymbol{t} + \lambda^* \boldsymbol{R})_i \leq 0$, $\boldsymbol{\nu}_i^* = -\boldsymbol{t}_i - \lambda^* \boldsymbol{R}_i$, otherwise, $\boldsymbol{\nu}_i = 0$. Thus, it holds that

$$\boldsymbol{\nu}^* = \max(-\boldsymbol{t} - \lambda^* \boldsymbol{R}, \boldsymbol{0}). \tag{3.55}$$

From (3.43), (3.46), (3.47) and (3.54), it follows that for each $i \in \{1, \cdots, n\}$, $(\boldsymbol{t} + \lambda^* \boldsymbol{R} - \boldsymbol{b})_i \geq 0$, $\boldsymbol{\mu}_i^* = (\boldsymbol{t} + \lambda^* \boldsymbol{R} - \boldsymbol{b})_i$, otherwise, $\boldsymbol{\mu}_i^* = 0$. This implies that

$$\boldsymbol{\mu}^* = \max(\boldsymbol{t} + \lambda^* \boldsymbol{R} - \boldsymbol{b}, \boldsymbol{0}). \tag{3.56}$$

Therefore, we obtain the optimal solutions $(\boldsymbol{z}^*, \lambda^*, \boldsymbol{\mu}^*, \boldsymbol{\nu}^*)$ satisfying KKT conditions (3.39)-(3.48).

3.5. HS-Jacobian matrix of $\Pi_{\mathcal{C}}$

At each iteration of mssN method, we need to compute $\mathbf{U} \in \partial\Pi_{\mathcal{C}}(\frac{1}{\alpha}\boldsymbol{t}_c^k(\boldsymbol{u},\boldsymbol{v}))$. Hence, it is crucial to efficiently compute the HS-Jacobian matrix of $\Pi_{\mathcal{C}}$ for the effectiveness of mssN method. The optimization problem in (3.38) can be reformulated as

$$\begin{aligned} \min_{\boldsymbol{z}\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{z}-\boldsymbol{t}\|^2 \\ \text{s.t. } \boldsymbol{R}^\top \boldsymbol{z} \geq r, \\ \mathbf{A}\boldsymbol{z} \geq \boldsymbol{g}, \end{aligned}$$

where $\mathbf{A} = [\mathbf{I}_n, -\mathbf{I}_n]^\top \in \mathbb{R}^{2n\times n}$, $\boldsymbol{g} = [\mathbf{0}^\top, -\boldsymbol{b}^\top]^\top \in \mathbb{R}^{2n}$. Let $\mathcal{I}(\boldsymbol{t})$ be the index set defined as

$$\mathcal{I}(\boldsymbol{t}) = \{i | \mathbf{A}_i \Pi_{\mathcal{C}}(\boldsymbol{t}) = \boldsymbol{g}_i, i = 1, \cdots, 2n\}, \tag{3.57}$$

where $\mathbf{A}_i$ is the i-th raw of matrix $\mathbf{A}$. Let $s = |\mathcal{I}(\boldsymbol{t})|$ be the cardinality of $\mathcal{I}(\boldsymbol{t})$. We compute the HS-Jacobian matrix $\mathbf{U} \in \partial\Pi_{\mathcal{C}}(\boldsymbol{t})$ by the following results.

Proposition 3.1. *For any $\boldsymbol{t} \in \mathbb{R}^n$, let*

$$\boldsymbol{\Sigma} := \mathbf{I}_n - \mathrm{Diag}(\boldsymbol{\theta}) \in \mathbb{R}^{n\times n},$$

where $\boldsymbol{\theta} \in \mathbb{R}^n$ can be computed by

$$\boldsymbol{\theta}_j = \begin{cases} 1, & j \in \mathcal{I}(\boldsymbol{t}) \text{ or } 2j \in \mathcal{I}(\boldsymbol{t}), \\ 0, & \text{otherwise}, \end{cases} \quad j = 1, \cdots, n,$$

then the HS-Jacobian matrix $\mathbf{U} \in \partial\Pi_{\mathcal{C}}(\boldsymbol{t})$ can be computed as the follows:

(1) *If $\boldsymbol{R}^\top \Pi_{\mathcal{C}}(\boldsymbol{t}) > r$, $\mathbf{U} = \boldsymbol{\Sigma}$.*

(2) *If $\boldsymbol{R}^\top \Pi_{\mathcal{C}}(\boldsymbol{t}) = r$,*

$$\mathbf{U} = \begin{cases} \boldsymbol{\Sigma}(\mathbf{I}_n - \frac{1}{\boldsymbol{R}^\top\boldsymbol{\Sigma}\boldsymbol{R}}\boldsymbol{R}\boldsymbol{R}^\top)\boldsymbol{\Sigma}, & \text{if } \boldsymbol{R}^\top\boldsymbol{\Sigma}\boldsymbol{R} \neq 0, \\ \boldsymbol{\Sigma}, & \text{otherwise.} \end{cases} \tag{3.58}$$

Proof. For statement (1), according to [44], the HS-Jacobian of $\Pi_{\mathcal{C}}$ at $\boldsymbol{t}$ can be computed by

$$\mathbf{U} = \mathbf{I}_n - \mathbf{A}_{\mathcal{I}(\boldsymbol{t})}^\top (\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}^\top)^\dagger \mathbf{A}_{\mathcal{I}(\boldsymbol{t})},$$

where † denotes the matrix Moore-Penrose pseudoinverse. From the definition of $\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}$, it holds that $\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}^\top = \mathbf{I}_s$, $\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}^\top\mathbf{A}_{\mathcal{I}(\boldsymbol{t})} = \mathbf{I}_n - \boldsymbol{\Sigma}$ and

$$\mathbf{U} = \mathbf{I}_n - \mathbf{A}_{\mathcal{I}(\boldsymbol{t})}^\top \mathbf{I}_s \mathbf{A}_{\mathcal{I}(\boldsymbol{t})} = \mathbf{I}_n - \mathbf{I}_n + \boldsymbol{\Sigma} = \boldsymbol{\Sigma}.$$

For statement (2), the HS-Jacobian matrix of $\Pi_{\mathcal{C}}$ at $\boldsymbol{t}$ can be computed by

$$\mathbf{U} = \mathbf{I}_n - [\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}^\top, \boldsymbol{R}]\mathbf{M}^\dagger \begin{bmatrix} \mathbf{A}_{\mathcal{I}(\boldsymbol{t})} \\ \boldsymbol{R}^\top \end{bmatrix}, \tag{3.59}$$

where $\mathbf{M} := \begin{bmatrix} \mathbf{A}_{\mathcal{I}(\boldsymbol{t})} \\ \boldsymbol{R}^\top \end{bmatrix} \left[\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}^\top, \boldsymbol{R}\right]$. Notice that the determinant of $\mathbf{M}$ can be computed by

$$\det(\mathbf{M}) = \det\left(\begin{bmatrix} \mathbf{A}_{\mathcal{I}(\boldsymbol{t})} \\ \boldsymbol{R}^\top \end{bmatrix} \left[\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}^\top, \boldsymbol{R}\right]\right) = \boldsymbol{R}^\top\boldsymbol{\Sigma}\boldsymbol{R}. \tag{3.60}$$

When $\boldsymbol{R}^\top\boldsymbol{\Sigma}\boldsymbol{R}\neq 0$, it holds that $\mathbf{M}^\dagger=\mathbf{M}^{-1}$ and

$$\begin{aligned}\mathbf{U}&=\mathbf{I}_n-[\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}^\top,\boldsymbol{R}]\mathbf{M}^{-1}\begin{bmatrix}\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}\\ \boldsymbol{R}^\top\end{bmatrix}\\&=\mathbf{I}_n-\left[\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}^\top,\boldsymbol{R}\right]\begin{bmatrix}\mathbf{I}_s+\frac{1}{\boldsymbol{R}^\top\boldsymbol{\Sigma}\boldsymbol{R}}\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}\mathbf{R}\mathbf{R}^\top\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}^\top & -\frac{1}{\boldsymbol{R}^\top\boldsymbol{\Sigma}\boldsymbol{R}}\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}\boldsymbol{R}\\ -\frac{1}{\boldsymbol{R}^\top\boldsymbol{\Sigma}\boldsymbol{R}}\boldsymbol{R}^\top\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}^\top & \frac{1}{\boldsymbol{R}^\top\boldsymbol{\Sigma}\boldsymbol{R}}\end{bmatrix}\begin{bmatrix}\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}\\ \boldsymbol{R}^\top\end{bmatrix}\\&=\boldsymbol{\Sigma}\left(\mathbf{I}_n-\frac{1}{\boldsymbol{R}^\top\boldsymbol{\Sigma}\boldsymbol{R}}\boldsymbol{R}\boldsymbol{R}^\top\right)\boldsymbol{\Sigma}.\end{aligned}$$

When $\boldsymbol{R}^\top\boldsymbol{\Sigma}\boldsymbol{R}=(\boldsymbol{R}\boldsymbol{\Sigma})^\top\boldsymbol{\Sigma}\boldsymbol{R}=0$, we can compute a full rank decomposition $\mathbf{M}=\mathbf{F}\mathbf{G}$, where

$$\mathbf{F}=\begin{bmatrix}\mathbf{I}_s\\ \boldsymbol{R}^\top\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}^\top\end{bmatrix},\quad \mathbf{G}=\left[\mathbf{I}_s,\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}\boldsymbol{R}\right].$$

Consequently, the Moore-Penrose pseudoinverse of $\mathbf{M}$ can be computed by

$$\mathbf{M}^\dagger=\mathbf{G}^\top(\mathbf{G}\mathbf{G}^\top)^{-1}(\mathbf{F}^\top\mathbf{F})^{-1}\mathbf{F}^\top.$$

Thus the HS-Jacobian matrix $\mathbf{U}\in\partial\Pi_{\mathcal{C}}(\boldsymbol{t})$ can be computed by

$$\mathbf{U}=\mathbf{I}_n-\left[\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}^\top,\boldsymbol{R}\right]\begin{bmatrix}\mathbf{I}_s-\frac{(\|\boldsymbol{R}\|^2+2)\mathbf{A}_{\mathcal{I}(t)}\mathbf{R}\mathbf{R}^\top\mathbf{A}_{\mathcal{I}(t)}^\top}{(1+\|\boldsymbol{R}\|^2)^2} & \frac{\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}\boldsymbol{R}}{(1+\|\boldsymbol{R}\|^2)^2}\\ \frac{\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}^\top\boldsymbol{R}^\top}{(1+\|\boldsymbol{R}\|^2)^2} & \frac{\|\boldsymbol{R}\|^2}{(1+\|\boldsymbol{R}\|^2)^2}\end{bmatrix}\begin{bmatrix}\mathbf{A}_{\mathcal{I}(\boldsymbol{t})}\\ \boldsymbol{R}^\top\end{bmatrix}=\boldsymbol{\Sigma}.$$

This completes the proof. □

3.6. Efficiently solving the linear system (3.21)

It should be noted that the main computation of mssN method is in solving linear system (3.21). Hence, it is important to reduce the computational cost of solving (3.21) by making full use of the second-order sparsity of the problem in (3.19) for whole algorithm framework. The linear system (3.21) can be reformulated as

$$(\mathbf{H}^j+\tfrac{1}{\alpha}\mathbf{B}\mathbf{U}^j\mathbf{B}^\top)\boldsymbol{d}=-\nabla\Theta_{c,\epsilon_j}^k(\boldsymbol{u}^{j-1},\boldsymbol{v}^{j-1};\boldsymbol{u}^{j-1}),\tag{3.61}$$

where $\mathbf{H}^j$ is a positive diagonal matrix, given as

$$\mathbf{H}^j=\begin{bmatrix}\epsilon^j & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_m\end{bmatrix}.$$

Notice that the Cholesky decomposition $\mathbf{H}^j=\mathbf{L}\mathbf{L}^\top$ can be easily computed. Consequently, (3.61) can be reformulated as

$$(\mathbf{I}_{m+1}+\tfrac{1}{\alpha}(\mathbf{L}^{-1}\mathbf{B})\mathbf{U}^j(\mathbf{L}^{-1}\mathbf{B})^\top)\mathbf{L}^\top\boldsymbol{d}=-\mathbf{L}^{-1}\nabla\Theta_{c,\epsilon_j}^k(\boldsymbol{u}^{j-1},\boldsymbol{v}^{j-1};\boldsymbol{u}^{j-1}).\tag{3.62}$$

Without loss of generality, we consider simplifying the linear system as the following formulation:

$$(\mathbf{I}_{m+1}+\tfrac{1}{\alpha}\mathbf{B}\mathbf{U}^j\mathbf{B}^\top)\boldsymbol{d}=-\nabla\Theta_{c,\epsilon_j}^k(\boldsymbol{u}^{j-1},\boldsymbol{v}^{j-1};\boldsymbol{u}^{j-1}).\tag{3.63}$$

Let $\mathcal{J}=\{j\in\{1,\cdots,n\}:j\notin\mathcal{I}(\boldsymbol{t}),2j\notin\mathcal{I}(\boldsymbol{t})\}$ with $\mathcal{I}(\boldsymbol{t})$ defined in (3.57). From Proposition 3.1, we solve (3.63) by distinguishing the following two cases.

(1) When $\mathbf{U}^j = \mathbf{\Sigma}$, notice that $\mathbf{\Sigma}$ is a diagonal matrix with the element 0 or 1, then it holds that $\mathbf{\Sigma} = \mathbf{\Sigma}\mathbf{\Sigma}$. If $|\mathcal{J}|$ is small, the inverse of $\mathbf{I}_{m+1} + \frac{1}{\alpha}\mathbf{B}\mathbf{U}^j\mathbf{B}^\top$ can be efficiently computed by the SMW formulation, i.e.,

$$(\mathbf{I}_{m+1} + \tfrac{1}{\alpha}\mathbf{B}\mathbf{U}^j\mathbf{B}^\top)^{-1} = (\mathbf{I}_{m+1} + \tfrac{1}{\alpha}\mathbf{B}_{\mathcal{J}}\mathbf{B}_{\mathcal{J}}^\top)^{-1} = \mathbf{I}_{m+1} - \mathbf{B}_{\mathcal{J}}(\alpha\mathbf{I}_{|\mathcal{J}|} + \mathbf{B}_{\mathcal{J}}^\top\mathbf{B}_{\mathcal{J}})^{-1}\mathbf{B}_{\mathcal{J}}^\top,$$

where $\mathbf{B}_{\mathcal{J}}$ is the submatrix of $\mathbf{B}$ indexed by $\mathcal{J}$. Hence, (3.63) can be solved by efficient matrix-vector multiplication.

(2) When $\mathbf{U}^j = \mathbf{\Sigma}(\mathbf{I}_n - \frac{1}{\boldsymbol{R}^\top\mathbf{\Sigma}\boldsymbol{R}}\boldsymbol{R}\boldsymbol{R}^\top)\mathbf{\Sigma}$, it holds that

$$\begin{aligned}\mathbf{B}\mathbf{\Sigma}(\mathbf{I}_n - \tfrac{1}{\boldsymbol{R}^\top\mathbf{\Sigma}\boldsymbol{R}}\boldsymbol{R}\boldsymbol{R}^\top)\mathbf{\Sigma}\mathbf{B}^\top &= \mathbf{B}\mathbf{\Sigma}\mathbf{\Sigma}(\mathbf{I}_n - \tfrac{1}{\boldsymbol{R}^\top\mathbf{\Sigma}\boldsymbol{R}}\boldsymbol{R}\boldsymbol{R}^\top)\mathbf{\Sigma}\mathbf{\Sigma}\mathbf{B}^\top \\ &= \mathbf{B}\mathbf{\Sigma}(\mathbf{\Sigma} - \tfrac{1}{\boldsymbol{R}^\top\mathbf{\Sigma}\boldsymbol{R}}\mathbf{\Sigma}\boldsymbol{R}(\mathbf{\Sigma}\boldsymbol{R})^\top)(\mathbf{B}\mathbf{\Sigma})^\top \\ &= \mathbf{B}_{\mathcal{J}}(\mathbf{I}_{|\mathcal{J}|} - \tfrac{1}{\boldsymbol{R}^\top\mathbf{\Sigma}\boldsymbol{R}}\boldsymbol{R}_{\mathcal{J}}\boldsymbol{R}_{\mathcal{J}}^\top)\mathbf{B}_{\mathcal{J}}^\top,\end{aligned}$$

where $\boldsymbol{R}_{\mathcal{J}}$ denotes the subvector of $\boldsymbol{R}$ with respect to $\mathcal{J}$. We consider the following eigenvalue decomposition: $\mathbf{I}_{|\mathcal{J}|} - \frac{1}{\boldsymbol{R}^\top\mathbf{\Sigma}\boldsymbol{R}}\boldsymbol{R}_{\mathcal{J}}\boldsymbol{R}_{\mathcal{J}}^\top = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$ with the orthogonal matrix $\mathbf{Q}$ and the diagonal matrix $\mathbf{\Lambda}$. Notice that the diagonal elements of matrix $\mathbf{\Lambda}$ are all 1 except for one 0. Moreover, the orthogonal matrix $\mathbf{Q}$ can be efficiently computed by performing QR-decomposition on $\frac{\boldsymbol{R}_{\mathcal{J}}}{\|\boldsymbol{R}_{\mathcal{J}}\|}$, which takes quite low computational cost, especially when $|\mathcal{J}| \ll m$. By setting $\bar{\mathbf{Q}} = \mathbf{Q}\sqrt{\mathbf{\Lambda}}$, it holds that $\mathbf{I}_{|\mathcal{J}|} - \frac{1}{\boldsymbol{R}^\top\mathbf{\Sigma}\boldsymbol{R}}\boldsymbol{R}_{\mathcal{J}}\boldsymbol{R}_{\mathcal{J}}^\top = \bar{\mathbf{Q}}\bar{\mathbf{Q}}^\top$. Consequently, the SMW formulation can be employed to compute the inverse of $\mathbf{I}_{m+1} + \frac{1}{\alpha}\mathbf{B}\bar{\mathbf{Q}}\bar{\mathbf{Q}}^\top\mathbf{B}^\top$, i.e.,

$$\begin{aligned}(\mathbf{I}_{m+1} + \tfrac{1}{\alpha}\mathbf{B}\mathbf{U}^j\mathbf{B}^\top)^{-1} &= (\mathbf{I}_{m+1} + \tfrac{1}{\alpha}\mathbf{B}_{\mathcal{J}}\bar{\mathbf{Q}}(\mathbf{B}_{\mathcal{J}}\bar{\mathbf{Q}})^\top)^{-1} \\ &= \mathbf{I}_{m+1} - \mathbf{B}_{\mathcal{J}}\bar{\mathbf{Q}}(\alpha\mathbf{I}_{|\mathcal{J}|} + (\mathbf{B}_{\mathcal{J}}\bar{\mathbf{Q}})^\top\mathbf{B}_{\mathcal{J}}\bar{\mathbf{Q}})^{-1}(\mathbf{B}_{\mathcal{J}}\bar{\mathbf{Q}})^\top.\end{aligned}$$

Remark 3.6. When $|\mathcal{J}| \ll m$, the inverse of matrix of size $|\mathcal{J}|$ can be efficiently computed and (3.63) can be solved directly by matrix-vector multiplication. For the practical numerical experiments of sparse optimization problems, $|\mathcal{J}| \ll m$ holds in most cases.

4. The Convergence of siPDCA-mssN

In this section, we prove the convergence of siPDCA-mssN for solving (3.3). A feasible point $\boldsymbol{x} \in \mathcal{D} \cap \mathcal{C}$ is said to be a stationary point of the DC problem (3.3) if

$$(\nabla f_c(\boldsymbol{x}) + \partial\delta_{\mathcal{C}}(\boldsymbol{x}) + \mathcal{N}_{\mathcal{D}}(\boldsymbol{x})) \cap \partial g_c(\boldsymbol{x}) \neq \emptyset, \tag{4.1}$$

where $\mathcal{N}_{\mathcal{D}}(\boldsymbol{x})$ is the normal cone of the convex set $\mathcal{D}$ at $\boldsymbol{x}$. The following results on the convergence of siPDCA-mssN for solving (3.3) follows from the basic convergence theorem of the classical DCA [45] and that of proximal DCA with extrapolation [49].

Lemma 4.1. *For any $\epsilon > 0$, the function $\mathbf{\Theta}_{c,\epsilon}^k(u, \boldsymbol{v}; \boldsymbol{u}')$ in (3.20) is level-bounded.*

Proof. Since $\Theta^k_{c,\epsilon}(\boldsymbol{u},\boldsymbol{v};\boldsymbol{u}')$ is a strongly convex function with parameter $\gamma=\min(1,\epsilon)$, then we have $\forall\,(\boldsymbol{u},\boldsymbol{v}),(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})\in\mathbb{R}\times\mathbb{R}^m$ with $\|(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})\|<\infty$

$$\begin{aligned}\Theta^k_{c,\epsilon}(\boldsymbol{u},\boldsymbol{v};\boldsymbol{u}') &\geq \Theta^k_{c,\epsilon}(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}};\boldsymbol{u}')+\nabla\Theta^k_{c,\epsilon}(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}};\boldsymbol{u}')^\top((\boldsymbol{u},\boldsymbol{v})-(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}}))+\frac{\gamma}{2}\|(\boldsymbol{u},\boldsymbol{v})-(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})\|^2\\ &\geq \Theta^k_{c,\epsilon}(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}};\boldsymbol{u}')-\nabla\Theta^k_{c,\epsilon}(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}};\boldsymbol{u}')^\top(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})+\frac{\gamma}{2}\|(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})\|^2\\ &\quad+(\nabla\Theta^k_{c,\epsilon}(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}};\boldsymbol{u}')-\gamma(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}}))^\top(\boldsymbol{u},\boldsymbol{v})+\frac{\gamma}{2}\|(\boldsymbol{u},\boldsymbol{v})\|^2\\ &\geq \Theta^k_{c,\epsilon}(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}};\boldsymbol{u}')-\nabla\Theta^k_{c,\epsilon}(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}};\boldsymbol{u}')^\top(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})+\frac{\gamma}{2}\|(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})\|^2\\ &\quad+\frac{\gamma}{2}\|(\boldsymbol{u},\boldsymbol{v})\|^2-\|\nabla\Theta^k_{c,\epsilon}(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}};\boldsymbol{u}')-\gamma(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})\|\|(\boldsymbol{u},\boldsymbol{v})\|.\end{aligned}\tag{4.2}$$

From these inequalities, it holds that for any $\rho\in\mathbb{R}$, $\{(\boldsymbol{u},\boldsymbol{v})\in\mathbb{R}\times\mathbb{R}^m:\Theta^k_{c,\epsilon}(\boldsymbol{u},\boldsymbol{v};\boldsymbol{u}')\leq\rho\}$ is bounded, which implies that $\Theta^k_{c,\epsilon}(\boldsymbol{u},\boldsymbol{v};\boldsymbol{u}')$ is level-bounded. This completes the proof. □

Proposition 4.1. *Let $\{\boldsymbol{x}^k\}$ be the sequence of stability centers generated by siPDCA-mssN for solving* (3.3) *and $\{(\widetilde{\boldsymbol{u}}^k,\boldsymbol{v}^k)\}$ be the sequence of the corresponding multipliers generated by Algorithm* 3.4*, then $\{\boldsymbol{x}^k\}$ and $\{(\widetilde{\boldsymbol{u}}^k,\boldsymbol{v}^k)\}$ are bounded.*

Proof. Since $\boldsymbol{x}^{k+1}$ is an ϵ_{k+1}-inexact solution of (3.4), then $\|\boldsymbol{e}^\top\boldsymbol{x}^{k+1}-1\|\leq\epsilon_{k+1}\leq\epsilon_1$ and $\boldsymbol{0}\leq\boldsymbol{x}^{k+1}\leq\boldsymbol{b}$. Evidently, $1-\epsilon_1\leq\boldsymbol{e}^\top\boldsymbol{x}^{k+1}\leq1+\epsilon_1$ holds, which implies that $\|\boldsymbol{x}^{k+1}\|\leq\|\boldsymbol{x}^{k+1}\|_1=\boldsymbol{e}^\top\boldsymbol{x}^{k+1}\leq1+\epsilon_1$. Thus $\{\boldsymbol{x}^k\}$ is bounded.

To display the boundedness of $\{(\widetilde{\boldsymbol{u}}^k,\boldsymbol{v}^k)\}$, we suppose that $\boldsymbol{x}^{k+1}$ is computed at the j^*-th step of Algorithm 3.4, i.e., $\boldsymbol{x}^{k+1}:=\widetilde{\boldsymbol{x}}^{k,j^*}$, $\widetilde{\boldsymbol{u}}^{k+1}:=u^{j^*}$ and $\boldsymbol{v}^{k+1}:=\boldsymbol{v}^{j^*}$. From Algorithm 3.4, it holds that $\forall\,j\in\{1,\cdots,j^*\}$

$$\Theta^k_{c,\epsilon_j}(\boldsymbol{u}^j,\boldsymbol{v}^j;\boldsymbol{u}^{j-1})\leq\Theta^k_{c,\epsilon_j}(\boldsymbol{u}^{j-1},\boldsymbol{v}^{j-1};\boldsymbol{u}^{j-1})+\rho l_j\langle\nabla\Theta^k_{c,\epsilon_j}(\boldsymbol{u}^{j-1},\boldsymbol{v}^{j-1};\boldsymbol{u}^{j-1}),(\boldsymbol{d}^j_{\boldsymbol{u}},\boldsymbol{d}^j_{\boldsymbol{v}})\rangle.$$

Since $\Theta^k_{c,\epsilon_j}(\boldsymbol{u},\boldsymbol{v};\boldsymbol{u}')$ is strongly convex, then it follows that $\mathbf{H}\in\partial^2\Theta^k_{c,\epsilon_j}(\boldsymbol{u}^{j-1},\boldsymbol{v}^{j-1};\boldsymbol{u}^{j-1})$ is positive definite and

$$\begin{aligned}&\langle\nabla\Theta^k_{c,\epsilon_j}(\boldsymbol{u}^{j-1},\boldsymbol{v}^{j-1};\boldsymbol{u}^{j-1}),(\boldsymbol{d}^j_{\boldsymbol{u}},\boldsymbol{d}^j_{\boldsymbol{v}})\rangle\\ =&-\nabla\Theta^k_{c,\epsilon_j}(\boldsymbol{u}^{j-1},\boldsymbol{v}^{j-1};\boldsymbol{u}^{j-1})^\top\mathbf{H}^{-1}\nabla\Theta^k_{c,\epsilon_j}(\boldsymbol{u}^{j-1},\boldsymbol{v}^{j-1};\boldsymbol{u}^{j-1})\leq0,\end{aligned}$$

which implies $\Theta^k_{c,\epsilon_j}(\boldsymbol{u}^j,\boldsymbol{v}^j;\boldsymbol{u}^{j-1})\leq\Theta^k_{c,\epsilon_j}(\boldsymbol{u}^{j-1},\boldsymbol{v}^{j-1};\boldsymbol{u}^{j-1})$, $\forall\,j\in\{1,\cdots,j^*\}$. This, together with

$$\Theta^k_{c,\epsilon_{j+1}}(\boldsymbol{u}^j,\boldsymbol{v}^j;\boldsymbol{u}^j)=\Theta^k_c(\boldsymbol{u}^j,\boldsymbol{v}^j)\leq\Theta^k_{c,\epsilon_j}(\boldsymbol{u}^j,\boldsymbol{v}^j;\boldsymbol{u}^{j-1}),$$

implies that

$$\Theta^k_{c,\epsilon_{j^*}}(\boldsymbol{u}^{j^*},\boldsymbol{v}^{j^*};\boldsymbol{u}^{j^*-1})\leq\cdots\leq\Theta^k_{c,\epsilon_1}(\boldsymbol{u}^1,\boldsymbol{v}^1;\boldsymbol{u}^0)\leq\Theta^k_c(\boldsymbol{u}^0,\boldsymbol{v}^0)\leq+\infty.$$

This, together with the level-boundedness of $\boldsymbol{\Theta}^k_{c,\epsilon^{j^*}}(\boldsymbol{u},\boldsymbol{v};\boldsymbol{u}')$ in Lemma 4.1, implies that $(\boldsymbol{u}^{j*},\boldsymbol{v}^{j*})$ is bounded, i.e., $\{(\widetilde{\boldsymbol{u}}^k,\boldsymbol{v}^k)\}$ is bounded. This completes the proof. □

x When the tolerance error is set as $\varepsilon=0$, our convergence analysis involves two cases. In the first case, only finite serious steps are performed in siPDCA-mssN for solving (3.3). In the second case, infinite serious steps are performed in siPDCA-mssN for solving (3.3).

4.1. Finite serious steps in siPDCA-mssN

For the first case that only finite serious steps are performed in siPDCA-mssN for solving (3.3), we have the following convergence results.

Theorem 4.1. *Set the tolerance error $\varepsilon = 0$. Suppose that only finite serious steps are performed in siPDCA-mssN for solving* (3.3)*. Then the following statements hold.*
(1) *If siPDCA-mssN terminates at $\bar{k}$-th step, then $\boldsymbol{x}^{\bar{k}}$ is a stationary point of* (3.3).
(2) *If only null step is performed in siPDCA-mssN after $\widehat{k}$-th iteration, i.e., for any $k > \widehat{k}$, $\boldsymbol{x}^k = \boldsymbol{x}^{\widehat{k}+1}$ and $\boldsymbol{\xi}^k = \boldsymbol{\xi}^{\widehat{k}+1}$, then the stability center $\boldsymbol{x}^{\widehat{k}+1}$ generated in the last serious step is a stationary point of* (3.3).

Proof. For statement (1), since $\boldsymbol{y}^{\bar{k}+1}$ satisfies the termination condition of siPDCA-mssN with tolerance error $\varepsilon = 0$, then it holds that $\boldsymbol{y}^{\bar{k}+1} = \boldsymbol{x}^{\bar{k}}$, $\Delta_1^{\bar{k}+1} = 0$ and $\Delta_2^{\bar{k}+1} = 0$. This, together with the optimality of $\boldsymbol{y}^{\bar{k}+1}$ and $\widetilde{\boldsymbol{u}}^{\bar{k}+1}$ for solving $\min_{\boldsymbol{x}\in\mathcal{D}} G_c^{\bar{k}}(\boldsymbol{x})$, yields that $\boldsymbol{x}^{\bar{k}}$ solves

$$\min_{\boldsymbol{x}\in\mathcal{D}} G_c^{\bar{k}}(\boldsymbol{x}) = f_c(\boldsymbol{x}) + \delta_{\mathcal{C}}(\boldsymbol{x}) - \langle \boldsymbol{x}, \boldsymbol{\xi}^{\bar{k}}\rangle + \frac{\alpha}{2}\|\boldsymbol{x} - \boldsymbol{x}^{\bar{k}}\|^2$$

and $\widetilde{\boldsymbol{u}}^{\bar{k}+1}$ is the multiplier of $\boldsymbol{e}^\top \boldsymbol{x}^{\bar{k}} = 1$. Consequently, it follows that:

$$\begin{aligned} &\mathbf{0} \in \nabla f_c(\boldsymbol{x}^{\bar{k}}) + \partial\delta_{\mathcal{C}}(\boldsymbol{x}^{\bar{k}}) - \boldsymbol{\xi}^{\bar{k}} + \widetilde{\boldsymbol{u}}^{\bar{k}+1}\boldsymbol{e},\\ &\boldsymbol{e}^\top \boldsymbol{x}^{\bar{k}} - 1 = 0.\end{aligned}$$

Evidently, we have $\boldsymbol{x}^{\bar{k}} \in \mathcal{D}$ and $\widetilde{\boldsymbol{u}}^{\bar{k}+1}\boldsymbol{e} \in \mathcal{N}_{\mathcal{D}}(\boldsymbol{x}^{\bar{k}})$. Thus, it holds that

$$\mathbf{0} \in \nabla f_c(\boldsymbol{x}^{\bar{k}}) + \partial\delta_{\mathcal{C}}(\boldsymbol{x}^{\bar{k}}) - \partial g_c(\boldsymbol{x}^{\bar{k}}) + \mathcal{N}_{\mathcal{D}}(\boldsymbol{x}^{\bar{k}})$$

and $\boldsymbol{x}^{\bar{k}}$ is a stationary point of (3.3).

For statement (2), we first prove $\lim_{k\to\infty} \boldsymbol{y}^{k+1} = \boldsymbol{x}^{\widehat{k}+1}$. Since only null step is performed after $\widehat{k}$-th iteration of siPDCA-mssN, i.e., $\forall\, k > \widehat{k}$, the sieving conditions in (5.1) do not hold. We just prove the case that $\forall\, k > \widehat{k}$, the test $\|\Delta_1^{k+1}\| \le (1-\kappa)\frac{\alpha}{2}\|\boldsymbol{y}^{k+1} - \boldsymbol{x}^k\|$ does not hold, i.e., $\forall\, k > \widehat{k}$,

$$(1-\kappa)\frac{\alpha}{2}\|\boldsymbol{y}^{k+1} - \boldsymbol{x}^{\widehat{k}+1}\| \le \|\Delta_1^{k+1}\| \le \epsilon_{k+1}. \tag{4.3}$$

Since $\lim_{k\to\infty} \epsilon_{k+1} = 0$, by taking limit on both sides of inequality (4.3), we have

$$\lim_{k\to\infty} (1-\kappa)\frac{\alpha}{2}\|\boldsymbol{y}^{k+1} - \boldsymbol{x}^{\widehat{k}+1}\| = 0,$$

which implies that $\lim_{k\to\infty} \boldsymbol{y}^{k+1} = \boldsymbol{x}^{\widehat{k}+1}$. The proof for the cases that the other sieving conditions do not hold is similar, then we omit it. Next, we prove that $\boldsymbol{x}^{\widehat{k}+1}$ is a stationary point of (3.3). From $|\boldsymbol{e}^\top \boldsymbol{y}^{k+1} - 1| < \epsilon_{k+1}$, we have

$$|\boldsymbol{e}^\top \boldsymbol{x}^{\widehat{k}+1} - 1| \le |\boldsymbol{e}^\top \boldsymbol{y}^{k+1} - 1| + \sqrt{n}\|\boldsymbol{y}^{k+1} - \boldsymbol{x}^{\widehat{k}+1}\| \le \epsilon_{k+1} + \sqrt{n}\|\boldsymbol{y}^{k+1} - \boldsymbol{x}^{\widehat{k}+1}\| \tag{4.4}$$

From the fact that the right-hand side of (4.4) goes to 0 as $k \to \infty$, it holds that

$$\boldsymbol{e}^\top \boldsymbol{x}^{\widehat{k}+1} - 1 = 0. \tag{4.5}$$

Since $\boldsymbol{x}^{\widehat{k}+1}$ is the stability center generated in the last serious step of Algorithm 3.3, then $\forall\, k > \widehat{k}$, $\boldsymbol{x}^k = \boldsymbol{x}^{\widehat{k}+1}$ and $\boldsymbol{\xi}^k = \boldsymbol{\xi}^{\widehat{k}+1}$ hold. This, together with the optimality of $\boldsymbol{y}^{k+1}$ for solving (3.5), yields that $\forall\, k > \widehat{k}$,

$$0 \in \nabla f_c(\boldsymbol{y}^{k+1}) + \partial\delta_{\mathcal{C}}(\boldsymbol{y}^{k+1}) - \boldsymbol{\xi}^{\widehat{k}+1} + \alpha(\boldsymbol{y}^{k+1} - \boldsymbol{x}^{\widehat{k}+1}) + \widetilde{\boldsymbol{u}}^{k+1}\boldsymbol{e} - \Delta_1^{k+1}.$$

Consequently, $\forall\, k > \widehat{k}$, there exists $\boldsymbol{\varsigma}^{k+1} \in \partial\delta_{\mathcal{C}}(\boldsymbol{y}^{k+1})$ such that

$$\nabla f_c(\boldsymbol{y}^{k+1}) + \boldsymbol{\varsigma}^{k+1} - \boldsymbol{\xi}^{\widehat{k}+1} + \alpha(\boldsymbol{y}^{k+1} - \boldsymbol{x}^{\widehat{k}+1}) + \widetilde{\boldsymbol{u}}^{k+1}\boldsymbol{e} - \Delta_1^{k+1} = 0.$$

Then it follows that $\forall\, k > \widehat{k}$,

$$\begin{aligned}&\|\nabla f_c(\boldsymbol{y}^{k+1}) + \boldsymbol{\varsigma}^{k+1} - \boldsymbol{\xi}^{\widehat{k}+1} + \widetilde{\boldsymbol{u}}^{k+1}\boldsymbol{e}\| \\ \leq&\alpha\|\boldsymbol{y}^{k+1} - \boldsymbol{x}^{\widehat{k}+1}\| + \|\Delta_1^{k+1}\| \leq \alpha\|\boldsymbol{y}^{k+1} - \boldsymbol{x}^{\widehat{k}+1}\| + \epsilon_{k+1}. \end{aligned} \tag{4.6}$$

According to proposition 4.1, we obtain that $\{\boldsymbol{x}^k\}$ and $\{\widetilde{\boldsymbol{u}}^k\}$ are bounded. Since $g_c(\boldsymbol{x})$ is a finite-valued convex function, then $\{\boldsymbol{\xi}^k\}$ is also bounded. Then the boundedness of $\{\boldsymbol{\varsigma}^k\}$ follows from (4.6). Thus, without loss of generality, we can suppose that there exists a subset $\mathcal{K}' \subset \mathcal{K} = \{0, 1, \cdots\}$ such that $\lim_{k\in\mathcal{K}'} \widetilde{\boldsymbol{u}}^{k+1} = \widehat{\boldsymbol{u}}$ and $\lim_{k\in\mathcal{K}'} \boldsymbol{\varsigma}^{k+1} = \widehat{\boldsymbol{\varsigma}}$. Since $\delta_{\mathcal{C}}(\boldsymbol{x})$ is lower semicontinuous, from [46, Proposition 4.1.1], it holds that $\widehat{\boldsymbol{\varsigma}} \in \partial\delta_{\mathcal{C}}(\boldsymbol{x}^{\widehat{k}+1})$. From $\lim_{k\in\mathcal{K}'} \epsilon_{k+1} = 0$, taking limit on both sides of inequality (4.6), we have

$$\lim_{k\in\mathcal{K}'} \|\nabla f_c(\boldsymbol{y}^{k+1}) + \boldsymbol{\varsigma}^{k+1} - \boldsymbol{\xi}^{\widehat{k}+1} + \widetilde{\boldsymbol{u}}^{k+1}\boldsymbol{e}\| = \|\nabla f_c(\boldsymbol{x}^{\widehat{k}+1}) + \widehat{\boldsymbol{\varsigma}} - \boldsymbol{\xi}^{\widehat{k}+1} + \widehat{u}\boldsymbol{e}\| = 0,$$

which implies that

$$0 \in \nabla f_c(\boldsymbol{x}^{\widehat{k}+1}) + \partial\delta_{\mathcal{C}}(\boldsymbol{x}^{\widehat{k}+1}) - \partial g_c(\boldsymbol{x}^{\widehat{k}+1}) + \widehat{u}\boldsymbol{e}.$$

From (4.5), we have that $\boldsymbol{x}^{\widehat{k}+1} \in \mathcal{D}$ and $\widehat{u}\boldsymbol{e} \in \mathcal{N}_{\mathcal{D}}(\boldsymbol{x}^{\widehat{k}+1})$. This implies that

$$0 \in \nabla f_c(\boldsymbol{x}^{\widehat{k}+1}) + \partial\delta_{\mathcal{C}}(\boldsymbol{x}^{\widehat{k}+1}) - \partial g_c(\boldsymbol{x}^{\widehat{k}+1}) + \mathcal{N}_{\mathcal{D}}(\boldsymbol{x}^{\widehat{k}+1})$$

and $\boldsymbol{x}^{\widehat{k}+1}$ is a stationary point of (3.3). This completes the proof. □

4.2. Infinite serious steps in siPDCA-mssN

In this subsection, we consider the case that infinite serious steps are performed in siPDCA-mssN for solving (3.3) when tolerance error ε is set to 0. Consequently, $\{\boldsymbol{x}^k\}$ can be expressed as

$$\left\{\cdots, \underbrace{\boldsymbol{x}^{k-N_l}}_{\boldsymbol{x}^{k_l}}, \underbrace{\boldsymbol{x}^{k-N_l+1}, \cdots, \boldsymbol{x}^{k}}_{N_l \text{ null steps}}, \underbrace{\boldsymbol{x}^{k+1}}_{\boldsymbol{x}^{k_{l+1}}}, \underbrace{\boldsymbol{x}^{k+2}, \cdots, \boldsymbol{x}^{k+N_{l+1}+1}}_{N_{l+1} \text{ null steps}}, \underbrace{\boldsymbol{x}^{k+N_{l+1}+2}}_{\boldsymbol{x}^{k_{l+2}}}, \cdots\right\},$$

where $\boldsymbol{x}^{k_l}$ denotes the stability center generated in the l-th serious step. The subsequence

$$\{\boldsymbol{x}^{k-N_l+1}, \cdots, \boldsymbol{x}^k\}$$

is the collection of the stability centers in null steps between the l-th serious step and the $(l+1)$-st serious step, then $\boldsymbol{x}^{k_l} = \boldsymbol{x}^{k-N_l} = \boldsymbol{x}^{k-N_l+1} = \cdots = \boldsymbol{x}^k$. From the assumption that infinite serious steps are performed in siPDCA-mssN for solving (3.3), we obtain that the stability centers in null steps between two adjacent serious steps are the finite repetitions of that generated in the previous serious step. By removing the $\boldsymbol{x}^k$ generated in null steps from $\{\boldsymbol{x}^k\}$, a subsequence $\{\boldsymbol{x}^{k_l}\}$ is obtained. Consequently, we can obtain subsequences $\{\boldsymbol{\xi}^{k_l}\}$, $\{(\Delta_1^{k_l}, \Delta_2^{k_l})\}$ and $\{\boldsymbol{u}^{k_l}\}$. For $\{\boldsymbol{x}^{k_l}\}$, we have the following global subsequential convergence results.

Theorem 4.2 (Global subsequential convergence of siPDCA-mssN). *Set the tolerance error $\varepsilon = 0$. Suppose that infinite serious steps are performed in siPDCA-mssN for solving* (3.3)*. Let $\{\boldsymbol{x}^{k_l}\}$ be the stability center sequence generated in serious steps of siPDCA-mssN for solving* (3.3)*. Then the following statements hold:*
(1) $\lim_{l\to\infty} \|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\| = 0$.
(2) *Any accumulation point $\overline{\boldsymbol{x}} \in \left\{\boldsymbol{x}^{k_l}\right\}$ is a stationary point of* (3.3).

Proof. We display the proof of statement (1)-(2) in Appendix A.1. □

In order to prove that when infinite serious steps are performed in Algorithm 3.3, the sequence $\{\boldsymbol{x}^k\}$ actually converges to a stationary point of (3.3), we construct the following auxiliary function:

$$\begin{aligned} &E(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{u}, \mathbf{v}) \\ =&f_c(\boldsymbol{x}) - \langle \boldsymbol{x}, \boldsymbol{y}\rangle + g_c^*(\boldsymbol{y}) + \delta_{\mathcal{C}}(\boldsymbol{x}) + \alpha\|\boldsymbol{x} - \boldsymbol{z}\|^2 + \boldsymbol{u}(\boldsymbol{e}^\top \boldsymbol{x} - 1) - \langle \mathbf{v}, \boldsymbol{x} - \boldsymbol{z}\rangle, \end{aligned} \tag{4.7}$$

where $g_c^*(\boldsymbol{y})$ is convex conjugate of $g_c(\boldsymbol{x})$, given as

$$g_c^*(\boldsymbol{y}) = \sup_{\boldsymbol{x}\in\mathbb{R}^n} \left\{\langle \boldsymbol{y}, \boldsymbol{x}\rangle - g_c(\boldsymbol{x})\right\}.$$

Then it holds that $f_c(\boldsymbol{x}) - g_c(\boldsymbol{x}) \le f_c(\boldsymbol{x}) - \langle \boldsymbol{x}, \boldsymbol{y}\rangle + g_c^*(\boldsymbol{y})$. Since $g_c(\boldsymbol{x}) = c\|\boldsymbol{x}\|_{(K)}$ is a proper closed convex function, it follows that $g_c^*(\boldsymbol{\xi})$ is also a proper closed convex function and the Youngs inequality holds:

$$g_c^*(\boldsymbol{y}) + g_c(\boldsymbol{x}) \ge \langle \boldsymbol{x}, \boldsymbol{y}\rangle,$$

where the equality holds if and only if $\boldsymbol{y} \in \partial g_c(\boldsymbol{x})$. Moreover, for any $\boldsymbol{x}$ and $\boldsymbol{y}$, $\boldsymbol{y} \in \partial g_c(\boldsymbol{x})$ if and only if $\boldsymbol{x} \in \partial g_c^*(\boldsymbol{y})$. According to [24,47,48], we obtain that semialgebraic functions satisfy the Kurdyka-Łojaziewicz (KŁ) property. Notice that E is a semialgebraic function, then E satisfies KŁ property.

Based on the auxiliary function E defined in (4.7), we have the following conclusions on the sequence $\{E(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}})\}$.

Proposition 4.2. *Let E be defined in* (4.7)*. Suppose that infinite serious steps are performed in siPDCA-mssN for solving* (3.3)*. Let $\left\{\boldsymbol{x}^{k_l}\right\}$, $\left\{\Delta_1^{k_l}\right\}$, $\left\{\boldsymbol{u}^{k_l}\right\}$ and $\left\{\boldsymbol{\xi}^{k_l}\right\}$ be the subsequences generated in the serious steps of siPDCA-mssN. Then the following statements hold.*

(1) *For any $l \ge 0$,*

$$f_c(\boldsymbol{x}^{k_{l+1}}) - g_c(\boldsymbol{x}^{k_{l+1}}) + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \boldsymbol{x}^{k_{l+1}} - 1) \le E(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}}). \tag{4.8}$$

(2) *For any $l \ge 1$,*

$$E(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}}) - E(\boldsymbol{x}^{k_l}, \boldsymbol{\xi}^{k_{l-1}}, \boldsymbol{x}^{k_{l-1}}, \boldsymbol{u}^{k_l}, \Delta_1^{k_l}) \le -\frac{\kappa\alpha}{2}\|\boldsymbol{x}^{k_l} - \boldsymbol{x}^{k_{l-1}}\|^2. \tag{4.9}$$

(3) *The set of accumulation points of the sequence $\left\{(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}})\right\}$, denoted by $\boldsymbol{\Gamma}$, is a nonempty compact set.*

(4) *The limit $\Upsilon = \lim_{l\to\infty} E(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}})$ exists and $E \equiv \Upsilon$ on $\boldsymbol{\Gamma}$.*

(5) *There exists a constant $\rho > 0$ such that for any $l \geq 0$, we have*

$$\operatorname{dist}(0, \partial E(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}})) \leq \rho \|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|. \tag{4.10}$$

Proof. We present the proof of statements (1)-(5) in Appendix A.2. □

For simplicity of notation, we set $E^{k_{l+1}} = E(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}})$ for each $l \geq 0$. Based on the KŁ property of E, we give the global convergence of $\{\boldsymbol{x}^{k_l}\}$ as follows.

Theorem 4.3. *Set the tolerance error $\varepsilon = 0$. Suppose that infinite serious steps are performed in siPDCA-mssN for solving* (3.3)*. Let $\{\boldsymbol{x}^{k_l}\}$ be the stability center sequence generated in the serious steps of siPDCA-mssN for solving* (3.3)*. Then $\{\boldsymbol{x}^{k_l}\}$ converges to a stationary point of* (3.3)*. Moreover, $\sum_{l=0}^{\infty} \|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\| < \infty$.*

Proof. From Proposition 4.2, we have that $\left\{E^{k_l}\right\}$ is nonincreasing and its limitation Υ exists. Consequently, it holds that $E^{k_l} \geq \Upsilon$, $\forall\, l > 0$. Next, we display that $E^{k_l} > \Upsilon$, $\forall\, l > 0$. To this end, we suppose that $\exists\, \widehat{L} > 0$ such that $E^{k_{\widehat{L}}} = \Upsilon$. Then we can get that $E^{k_l} = \Upsilon$ for all $l > \widehat{L}$. From (4.9), we have $\boldsymbol{x}^{k_l} = \boldsymbol{x}^{k_L}$, $\forall\, l \geq \widehat{L}$. This implies that only finite steps are performed in siPDCA-mssN, which is contrary to the assumption that infinite serious steps are performed in siPDCA-mssN.

Since E satisfies the KŁ property at each point in the compact set $\boldsymbol{\Gamma} \subset \operatorname{dom} E$ and $E \equiv \Upsilon$ on $\boldsymbol{\Gamma}$, then it satisfies the uniform KŁ property [24]. Hence, there exist $\epsilon > 0$ and a continuous concave function $\varphi : [0, a) \to \mathbb{R}_+$ being continuously differentiable and monotonically increasing on $(0, a)$ and satisfying $\varphi(0) = 0$ with $a > 0$ such that $\forall\, (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) \in \boldsymbol{\Theta}$,

$$\varphi'(E(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) - \Upsilon) \cdot \operatorname{dist}(\boldsymbol{0}, \partial E(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v})) \geq 1,$$

where

$$\begin{aligned}\boldsymbol{\Theta} = &\{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) : \operatorname{dist}((\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}), \boldsymbol{\Gamma}) < \epsilon\} \\ &\cap \{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) : \Upsilon < E(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{u}, \boldsymbol{v}) < \Upsilon + a\}.\end{aligned}$$

Since $\boldsymbol{\Gamma}$ is the set of accumulation points of the $\left\{(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}})\right\}$, then

$$\lim_{l \to \infty} \operatorname{dist}((\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}}), \boldsymbol{\Gamma}) = 0.$$

Consequently, there exists a $\bar{L} > 0$ such that

$$\operatorname{dist}((\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}}), \boldsymbol{\Gamma}) < \epsilon, \quad \forall\, l > \bar{L} - 2.$$

From Proposition 4.2, it holds that $\left\{E^{k_{l+1}}\right\}$ converges to Υ. Hence, there exists a $\bar{\bar{L}} > 0$ such that

$$\Upsilon < E^{k_{l+1}} < \Upsilon + a, \quad \forall\, l > \bar{\bar{L}} - 2.$$

Let $\widetilde{L} = \max(\bar{L}, \bar{\bar{L}})$, then we have $\forall\, l > \widetilde{L}$, $(\boldsymbol{x}^{k_{l-1}}, \boldsymbol{\xi}^{k_{l-2}}, \boldsymbol{x}^{k_{l-2}}, u^{k_{l-1}}, \Delta_1^{k_{l-1}}) \in \boldsymbol{\Theta}$ and

$$\varphi'(E^{k_{l-1}} - \Upsilon) \cdot \operatorname{dist}(\boldsymbol{0}, \partial E^{k_{l-1}}) \geq 1. \tag{4.11}$$

From the concavity of φ and the fact that $\{E^{k_l}\}$ is nonincreasing, it holds that $\forall\, l > \widetilde{L}$,

$$\begin{aligned}&\left[\varphi(E^{k_{l-1}} - \Upsilon) - \varphi(E^{k_{l+1}} - \Upsilon)\right] \cdot \operatorname{dist}(\boldsymbol{0}, \partial E^{k_{l-1}}) \\ \geq &\varphi'(E^{k_{l-1}} - \Upsilon) \cdot \operatorname{dist}(\boldsymbol{0}, \partial E^{k_{l-1}}) \cdot (E^{k_{l-1}} - E^{k_{l+1}}) \geq E^{k_{l-1}} - E^{k_{l+1}},\end{aligned} \tag{4.12}$$

where the last inequality is due to (4.11). Let $\pi^{k_l} = \varphi(E^{k_l} - \Upsilon)$, $\forall\, l > \widetilde{L}$. Since φ is monotone increasing on $(0, a)$ and $\{E^{k_l}\}$ is nonincreasing, then $\{\pi^{k_l}\}$ is nonincreasing. This, together with (4.9)–(4.12), yields that there exists a constant $\rho > 0$ such that for any $l > \widetilde{L}$

$$\|\boldsymbol{x}^{k_l} - \boldsymbol{x}^{k_{l-1}}\|^2 + \|\boldsymbol{x}^{k_{l-1}} - \boldsymbol{x}^{k_{l-2}}\|^2 \leq \frac{2\rho}{\kappa\alpha}(\pi^{k_{l-1}} - \pi^{k_{l+1}})\|\boldsymbol{x}^{k_{l-1}} - \boldsymbol{x}^{k_{l-2}}\|. \tag{4.13}$$

By using the arithmetic mean-geometric mean inequality, we have

$$\begin{aligned}\|\boldsymbol{x}^{k_l} - \boldsymbol{x}^{k_{l-1}}\| &\leq \sqrt{\frac{\rho}{\kappa\alpha}(\pi^{k_{l-1}} - \pi^{k_{l+1}}) - \frac{1}{2}\|\boldsymbol{x}^{k_{l-1}} - \boldsymbol{x}^{k_{l-2}}\|} \cdot \sqrt{2\|\boldsymbol{x}^{k_{l-1}} - \boldsymbol{x}^{k_{l-2}}\|} \\ &\leq \frac{\rho}{2\kappa\alpha}(\pi^{k_{l-1}} - \pi^{k_{l+1}}) - \frac{1}{4}\|\boldsymbol{x}^{k_{l-1}} - \boldsymbol{x}^{k_{l-2}}\| + \|\boldsymbol{x}^{k_{l-1}} - \boldsymbol{x}^{k_{l-2}}\|.\end{aligned}$$

This implies that

$$\frac{1}{4}\|\boldsymbol{x}^{k_l} - \boldsymbol{x}^{k_{l-1}}\| \leq \frac{\rho}{2\kappa\alpha}(\pi^{k_{l-1}} - \pi^{k_{l+1}}) + \frac{3}{4}(\|\boldsymbol{x}^{k_{l-1}} - \boldsymbol{x}^{k_{l-2}}\| - \|\boldsymbol{x}^{k_l} - \boldsymbol{x}^{k_{l-1}}\|). \tag{4.14}$$

Summing both sides of (4.14) from $l = \widetilde{L}$ to ∞, we obtain that

$$\begin{aligned}\frac{1}{4}\sum_{l=\widetilde{L}}^{\infty}\|\boldsymbol{x}^{k_l} - \boldsymbol{x}^{k_{l-1}}\| \leq \frac{\rho}{2\kappa\alpha}(\pi^{k_{\widetilde{L}-1}} + \pi^{k_{\widetilde{L}}}) - \lim_{l\to\infty}\frac{\rho}{2\kappa\alpha}(\pi^{k_l} + \pi^{k_{l+1}}) \\ + \frac{3}{4}(\|\boldsymbol{x}^{k_{\widetilde{L}-1}} - \boldsymbol{x}^{k_{\widetilde{L}-2}}\| - \lim_{l\to\infty}\|\boldsymbol{x}^{k_l} - \boldsymbol{x}^{k_{l-1}}\|).\end{aligned}$$

From $\lim_{l\to\infty}\frac{\rho}{2\kappa\alpha}(\pi^{k_l} + \pi^{k_{l+1}}) = 0$ and $\lim_{l\to\infty}\|\boldsymbol{x}^{k_l} - \boldsymbol{x}^{k_{l-1}}\| = 0$, it holds that

$$\frac{1}{4}\sum_{l=\widetilde{L}}^{\infty}\|\boldsymbol{x}^{k_l} - \boldsymbol{x}^{k_{l-1}}\| \leq \frac{\rho}{2\kappa\alpha}(\pi^{k_{\widetilde{L}-1}} + \pi^{k_{\widetilde{L}}}) + \frac{3}{4}\|\boldsymbol{x}^{k_{\widetilde{L}-1}} - \boldsymbol{x}^{k_{\widetilde{L}-2}}\| < \infty, \tag{4.15}$$

which implies that $\left\{\boldsymbol{x}^{k_l}\right\}$ is convergent and $\sum_{l=0}^{\infty}\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\| < \infty$. This, together with the sieving conditions in (5.1), yields that $\lim_{l\to\infty}|\boldsymbol{e}^{\top}\boldsymbol{x}^{k_l} - 1| = 0$. Combining this with the results of Theorem 4.2, we obtain that the sequence $\left\{\boldsymbol{x}^{k_l}\right\}$ converges to a stationary point of (3.3). This completes the proof. □

By using Theorem 4.3, we give the following global convergence of $\{\boldsymbol{x}^k\}$.

Theorem 4.4 (Global sequential convergence of siPDCA-mssN). *Set the tolerance error $\varepsilon = 0$. Suppose that infinite serious steps are performed in siPDCA-mssN for solving* (3.3). *Let $\{\boldsymbol{x}^k\}$ be the stability center sequence generated by siPDCA-mssN for solving* (3.3). *Then $\{\boldsymbol{x}^k\}$ converges to a stationary point of* (3.3). *Moreover, $\sum_{k=0}^{\infty}\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\| < \infty$.*

Proof. From assumption of infinite serious steps in this subsection, we obtain that $\{\boldsymbol{x}^{k_l}\}$ is the subsequence of $\{\boldsymbol{x}^k\}$ removing the finite repeated points in null steps. This, together with Theorem 4.3, implies that the sequence $\left\{\boldsymbol{x}^k\right\}$ converges to a stationary point of (3.3) and

$$\sum_{k=0}^{\infty}\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\| = \sum_{l=0}^{\infty}\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\| < \infty.$$

This completes the proof. □

4.3. Local convergence of siPDCA-mssN

Proposition 4.3. *If $\boldsymbol{x}^*$ is a local minimizer of (1.1). Suppose that $\mathcal{I}^* \subset \{1, \cdots, n\}$ is an index set satisfying $|\mathcal{I}^*| = K$ and $\forall\, i \in \overline{\mathcal{I}^*}$, the $\boldsymbol{x}_i^* = 0$ holds, where $\overline{\mathcal{I}^*} = \{1, \cdots, n\} \setminus \mathcal{I}^*$. Suppose that the Mangasarian-Fromowitz Constraint Qualification (MFCQ) conditions hold at $\boldsymbol{x}^*$, i.e., there exists a point $\boldsymbol{x} \in \mathbb{R}^n$ such that*

$$\begin{aligned}
&-\boldsymbol{R}^\top(\boldsymbol{x} - \boldsymbol{x}^*) < 0, && \text{if } \boldsymbol{R}^\top \boldsymbol{x}^* = r,\\
&\boldsymbol{e}_i^\top(\boldsymbol{x} - \boldsymbol{x}^*) < 0, && \text{if } \boldsymbol{x}_i^* = \boldsymbol{b}_i,\\
&-\boldsymbol{e}_i^\top(\boldsymbol{x} - \boldsymbol{x}^*) < 0, && \text{if } \boldsymbol{x}_i^* = 0, i \in \mathcal{I}^*,\\
&\boldsymbol{e}_i^\top(\boldsymbol{x} - \boldsymbol{x}^*) = 0, && \text{if } i \in \mathcal{I}^*,\\
&\boldsymbol{e}^\top(\boldsymbol{x} - \boldsymbol{x}^*) = 0
\end{aligned} \tag{4.16}$$

hold. Then there exist $(\lambda^*, \boldsymbol{u}^*, \boldsymbol{\eta}^*, \boldsymbol{\zeta}^*, \boldsymbol{\gamma}^*) \in \mathbb{R}^{3n+2}$ together with $\boldsymbol{x}^*$ satisfy the following optimality conditions

$$\begin{aligned}
&\mathbf{Q}\boldsymbol{x}^* - \lambda^*\boldsymbol{R} + \boldsymbol{u}^*\boldsymbol{e} + \boldsymbol{\eta}^* - \boldsymbol{\zeta}^* + \boldsymbol{\gamma}^* = \mathbf{0},\\
&\lambda^* \le 0,\ \boldsymbol{\eta}^*, \quad \boldsymbol{\zeta}^* \ge \mathbf{0}, \quad \lambda^*(r - \boldsymbol{R}^\top \boldsymbol{x}^*) = 0,\\
&\boldsymbol{\eta}_i^*(\boldsymbol{x}_i^* - \boldsymbol{b}_i^*) = 0, \quad \boldsymbol{\zeta}_i^* \boldsymbol{x}_i^* = 0, \quad i = 1, \cdots, n,\\
&\boldsymbol{R}^\top \boldsymbol{x}^* \ge r, \quad \boldsymbol{e}^\top \boldsymbol{x}^* = 1, \quad \mathbf{0} \le \boldsymbol{x}^* \le \boldsymbol{b},\\
&(\boldsymbol{\zeta}^*)_i = 0, \quad i \in \overline{\mathcal{I}^*}, \quad (\boldsymbol{\gamma}^*)_i = 0, \quad i \in \mathcal{I}^*.
\end{aligned} \tag{4.17}$$

Proof. Notice that if $\boldsymbol{x}^*$ is a local minimizer of problem (1.1), then $\boldsymbol{x}^*$ is a global minimizer of the problem:

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ \frac{1}{2}\boldsymbol{x}^\top \mathbf{Q}\boldsymbol{x} : \boldsymbol{R}^\top \boldsymbol{x} \ge r, \boldsymbol{e}^\top \boldsymbol{x} = 1, 0 \le \boldsymbol{x}_{\mathcal{I}^*} \le \boldsymbol{b}_{\mathcal{I}^*}, \boldsymbol{x}_{\overline{\mathcal{I}^*}} = 0 \right\} \tag{4.18}$$

This, together with the MFCQ condition, implies that the optimality conditions in (4.16) hold. □

Proposition 4.4. *Let $\boldsymbol{x}^*$ be a feasible point of (1.1). Let*

$$\mathcal{J}^* = \{\mathcal{I}^* \subset \{1, \cdots, n\} : \boldsymbol{x}_i^* = 0, |\mathcal{I}^*| = K, \quad \forall\, i \in \{1, \cdots, n\} \setminus \mathcal{I}^*\}.$$

If $\forall\, \mathcal{I}^* \in \mathcal{J}^*$, there exist $(\lambda^*, \boldsymbol{u}^*, \boldsymbol{\eta}^*, \boldsymbol{\zeta}^*, \boldsymbol{\gamma}^*) \in \mathbb{R}^{3n+2}$ together with $\boldsymbol{x}^*$ satisfying (4.17), then $\boldsymbol{x}^*$ is a local minimizer of (1.1).

Proof. For the details of proof, one can refer to [25, Theorem 2.3]. □

Theorem 4.5. *Set the tolerance error $\varepsilon = 0$. Let $\{\boldsymbol{x}^k\}$ be the sequence generated by Algorithm 3.3 for solving (3.3) with penalty parameter $\bar{c} > 0$. Let $\boldsymbol{x}^*$ be an accumulation point of $\{\boldsymbol{x}^k\}$ satisfying $\|\boldsymbol{x}^*\|_0 \le K$ and*

$$\mathcal{J}^* = \left\{\mathcal{I}^* \subset \{1, \cdots, n\} : |\mathcal{I}^*| = K, \boldsymbol{x}_i^* = 0, \forall\, i \in \overline{\mathcal{I}^*} = \{1, \cdots, n\} \setminus \mathcal{I}^*\right\}.$$

Then $\forall\, \mathcal{I}^* \in \mathcal{J}^*$, there exist $(\lambda^*, \boldsymbol{u}^*, \boldsymbol{\eta}^*, \boldsymbol{\zeta}^*, \boldsymbol{\gamma}^*) \in \mathbb{R}^{3n+2}$ together with $\boldsymbol{x}^*$ satisfying (4.17), and $\boldsymbol{x}^*$ is a local minimizer of (1.1).

Proof. From Theorems 4.1 and x4.4, we obtain that $\boldsymbol{x}^*$ is a stationary point of (3.3), then there exists $\boldsymbol{u}^*$ such that the following KKT conditions hold:

$$\begin{aligned} &\boldsymbol{x}^* = \Pi_{\mathcal{C}}(\boldsymbol{x}^* + \boldsymbol{\xi}^* - \bar{c}\boldsymbol{e} - \mathbf{S}\mathbf{S}^\top \boldsymbol{x}^* - \boldsymbol{u}^*\boldsymbol{e}), \\ &\boldsymbol{e}^\top \boldsymbol{x}^* - 1 = 0, \end{aligned} \tag{4.19}$$

where $\boldsymbol{\xi}^* \in \partial g_{\bar{c}}(\boldsymbol{x}^*)$. By using (4.19) and the results in Subsection 3.4, we obtain that there exist $(\lambda^*, \boldsymbol{\nu}^*, \boldsymbol{\mu}^*)$ together with $\boldsymbol{x}^*$ and $\boldsymbol{t} = \boldsymbol{x}^* + \boldsymbol{\xi}^* - \bar{c}\boldsymbol{e} - \mathbf{S}\mathbf{S}^\top \boldsymbol{x}^* - \boldsymbol{u}^*\boldsymbol{e}$ satisfy (3.43), i.e.,

$$\boldsymbol{x}^* - (\boldsymbol{x}^* + \boldsymbol{\xi}^* - \bar{c}\boldsymbol{e} - \mathbf{S}\mathbf{S}^\top \boldsymbol{x}^* - \boldsymbol{u}^*\boldsymbol{e}) - \lambda^*\boldsymbol{R} - \boldsymbol{\nu}^* + \boldsymbol{\mu}^* = \mathbf{0}.$$

Consequently, it follows that

$$\mathbf{Q}\boldsymbol{x}^* - \lambda^*\boldsymbol{R} + \boldsymbol{u}^*\boldsymbol{e} + \bar{c}\boldsymbol{e} - \boldsymbol{\xi}^* - \boldsymbol{\nu}^* + \boldsymbol{\mu}^* = 0,$$

where $\mathbf{Q} = \mathbf{S}\mathbf{S}^\top$. For any $i \in \{1, \cdots, n\}$, let $\boldsymbol{\zeta}_i^*$, $\boldsymbol{\eta}_i^*$ and $\boldsymbol{\gamma}_i^*$ be computed by

$$\boldsymbol{\zeta}_i^* = \begin{cases} \boldsymbol{\nu}_i^*, & \text{if } i \in \mathcal{I}^*, \\ 0, & \text{if } i \in \overline{\mathcal{I}^*}, \end{cases} \qquad \boldsymbol{\eta}_i^* = \begin{cases} \boldsymbol{\mu}_i^*, & \text{if } i \in \mathcal{I}^*, \\ 0, & \text{if } i \in \overline{\mathcal{I}^*}, \end{cases}$$

and

$$\boldsymbol{\gamma}_i^* = \begin{cases} (\bar{c}\boldsymbol{e} - \boldsymbol{\xi}^* + \boldsymbol{\mu}^* - \boldsymbol{\nu}^*)_i, & \text{if } i \in \overline{\mathcal{I}^*}, \\ (\bar{c}\boldsymbol{e} - \boldsymbol{\xi}^*)_i, & \text{if } i \in \mathcal{I}^*. \end{cases}$$

Consequently, it holds that

$$\mathbf{Q}\boldsymbol{x}^* - \lambda^*\boldsymbol{R} + \boldsymbol{u}^*\boldsymbol{e} + \boldsymbol{\eta}^* - \boldsymbol{\zeta}^* + \boldsymbol{\gamma}^* = 0.$$

Due to $\|\boldsymbol{x}^*\|_0 \leq K$, it follows that $\forall \mathcal{I}^* \in \mathcal{J}^*$, $(\lambda^*, \boldsymbol{u}^*, \boldsymbol{\eta}^*, \boldsymbol{\zeta}^*, \boldsymbol{\gamma}^*) \in \mathbb{R}^{3n+2}$ together with $\boldsymbol{x}^*$ satisfy the conditions in (4.17). This, together with Proposition 4.4, yields that $\boldsymbol{x}^*$ is a local minimizer of (1.1). This completes the proof. □

5. A Decomposed siPDCA-mssN for Large-Scale CCMV Problem

5.1. The algorithm framework of decomposed siPDCA-mssN

The computational cost and storage cost of siPDCA-mssN for (1.1) will be expensive when the scale of the CCMV problem is large. As discussed in Section 1, Wang et al. [50] proposed a decomposed strategy based on the violation of the first-order optimality conditions to reduce a large-scale problem into a small problem. Inspired by their work, we use the violation of first-order optimality conditions defined in (4.17) as the decomposition criterion to reduce (1.1) into a small-scale problem.

Let $\boldsymbol{x}^q \in \mathcal{C} \cap \mathcal{D}$ be an inexact solution of (3.3) with penalty $c_{q-1} > 0$ and $\mathcal{I}^q$ be the index set of the nonzero elements of $\boldsymbol{x}^q$. By using $\mathbf{Q} = \mathbf{S}\mathbf{S}^\top$, it follows that there exists an inexact term $\Delta^q \in \mathbb{R}^n$ and a multiplier $\boldsymbol{u}^q \in \mathbb{R}$ of the equality constraint such that the following optimality conditions hold:

$$\begin{aligned} &\Delta^q \in \mathbf{S}\mathbf{S}^\top \boldsymbol{x}^q + c_{q-1}\boldsymbol{e} + \partial\delta_{\mathcal{C}}(\boldsymbol{x}^q) + \boldsymbol{u}^q\boldsymbol{e} - \partial g_c(\boldsymbol{x}^q), \\ &\boldsymbol{e}^\top \boldsymbol{x}^q = 1. \end{aligned} \tag{5.1}$$

Let $\boldsymbol{\xi}^q \in \partial g_c(\boldsymbol{x}^q)$, then it holds that

$$\boldsymbol{x}^q = \Pi_{\mathcal{C}}(\boldsymbol{x}^q + \boldsymbol{\xi}^q - \mathbf{S}\mathbf{S}^\top \boldsymbol{x}^q - c_{q-1}\boldsymbol{e} - \boldsymbol{u}^q\boldsymbol{e} + \Delta^q),$$

which means that $\boldsymbol{x}^q$ is an optimal solution of (3.38) with $\boldsymbol{t} = \boldsymbol{x}^q+\boldsymbol{\xi}^q-\mathbf{S}\mathbf{S}^\top\boldsymbol{x}^q-c_{q-1}\boldsymbol{e}-\boldsymbol{u}^q\boldsymbol{e}+\Delta^q$. This implies that there exist $(\lambda^q, \boldsymbol{\mu}^q, \boldsymbol{\nu}^q)$ together with $\boldsymbol{x}^q$ satisfying the KKT conditions in (3.40)-(3.48). Consequently, it holds that

$$\boldsymbol{x}^q - (\boldsymbol{x}^q + \boldsymbol{\xi}^q - \mathbf{S}\mathbf{S}^\top\boldsymbol{x}^q - c_{q-1}\boldsymbol{e} - \boldsymbol{u}^q\boldsymbol{e} + \Delta^q) - \lambda^q\boldsymbol{R} - \boldsymbol{\nu}^q + \boldsymbol{\mu}^q = \mathbf{0},$$

which can be simplified as

$$\mathbf{S}\mathbf{S}^\top\boldsymbol{x}^q + c_{q-1}\boldsymbol{e} + \boldsymbol{u}^q\boldsymbol{e} - \boldsymbol{\xi}^q - \lambda^q\boldsymbol{R} - \boldsymbol{\nu}^q + \boldsymbol{\mu}^q = \Delta^q, \tag{5.2}$$

Let $(\boldsymbol{\eta}^q, \boldsymbol{\zeta}^q, \boldsymbol{\gamma}^q)$ be defined as follows:

$$\boldsymbol{\eta}_i^q := \begin{cases} \boldsymbol{\mu}_i^q, & \text{if } i \in \mathcal{I}^q \\ 0, & \text{if } i \in \overline{\mathcal{I}^q} \end{cases} \quad i \in \{1, \cdots, n\}, \tag{5.3}$$

$$\boldsymbol{\zeta}_i^q := \begin{cases} \boldsymbol{\nu}_i^q, & \text{if } i \in \mathcal{I}^q \\ 0, & \text{if } i \in \overline{\mathcal{I}^q} \end{cases} \quad i \in \{1, \cdots, n\}, \tag{5.4}$$

$$\boldsymbol{\gamma}_i^q := \begin{cases} (c_{q-1}\boldsymbol{e} - \boldsymbol{\xi}^q + \boldsymbol{\mu}^q - \boldsymbol{\nu}^q)_i, & \text{if } i \in \overline{\mathcal{I}^q} \\ (c_{q-1}\boldsymbol{e} - \boldsymbol{\xi}^q)_i, & \text{if } i \in \mathcal{I}^q \end{cases} \quad i \in \{1, \cdots, n\}. \tag{5.5}$$

As a result, (5.2) can be reformulated as

$$\mathbf{S}\mathbf{S}^\top\boldsymbol{x}^q - \boldsymbol{\xi}^q - \lambda^q\boldsymbol{R} + \boldsymbol{\eta}^q - \boldsymbol{\zeta}^q + \boldsymbol{\gamma}^q = \Delta^q. \tag{5.6}$$

This implies that when $\|\boldsymbol{x}^q\|_0 \leq K$, $(\boldsymbol{x}^q, \lambda^q, \boldsymbol{u}^q, \boldsymbol{\eta}^q, \boldsymbol{\zeta}^q, \boldsymbol{\gamma}^q)$ satisfy the KKT conditions in (4.17) except the first one. Therefore, we define the violation of the first-order optimality conditions of $\boldsymbol{x}^q$ as follows:

$$V(\boldsymbol{x}^q; \lambda^q, \boldsymbol{u}^q, \boldsymbol{\eta}^q, \boldsymbol{\zeta}^q, \boldsymbol{\gamma}^q)_i := |\mathbf{S}\mathbf{S}^\top\boldsymbol{x}^q - \lambda^q\boldsymbol{R} + \boldsymbol{u}^q\boldsymbol{e} + \boldsymbol{\eta}^q - \boldsymbol{\zeta}^q + \boldsymbol{\gamma}^q|_i, \quad xi \in \{1, \cdots, n\}. \tag{5.7}$$

Evidently, for each $i \in \{1, \cdots, n\}$,

$$V(\boldsymbol{x}^q; \lambda^q, \boldsymbol{u}^q, \boldsymbol{\eta}^q, \boldsymbol{\zeta}^q, \boldsymbol{\gamma}^q)_i = |\mathbf{S}\mathbf{S}^\top\boldsymbol{x}^q + c_{q-1}\boldsymbol{e} + \boldsymbol{u}^q\boldsymbol{e} - \boldsymbol{\xi}^q - \lambda^q\boldsymbol{R} - \boldsymbol{\nu}^q + \boldsymbol{\mu}^q|_i. \tag{5.8}$$

From Proposition 4.4, it follows that when $\|V(\boldsymbol{x}^q; \lambda^q, \boldsymbol{u}^q, \boldsymbol{\eta}^q, \boldsymbol{\zeta}^q, \boldsymbol{\gamma}^q)\| = 0$ and $\|\boldsymbol{x}^q\|_0 \leq K$, $\boldsymbol{x}^q$ is a local minimizer of (1.1). Based on (5.7), the violation of the first-order optimality conditions of $\boldsymbol{x}^q$ can naturally be used as the decomposition criterion. As a result, we choose an index set $\mathcal{B}^q = \mathcal{I}^q \cup \mathcal{B}_1^q$, where $\mathcal{B}_1^q$ is the index set of the components those satisfying $V(\boldsymbol{x}^q; \lambda^q, \boldsymbol{u}^q, \boldsymbol{\eta}^q, \boldsymbol{\zeta}^q, \boldsymbol{\gamma}^q)_i > \bar{\epsilon}$ with $\bar{\epsilon} > 0$. Based on the index set $\mathcal{B}^q$, we decompose the original objective function in (1.1) as

$$\begin{aligned} \frac{1}{2}\boldsymbol{x}^\top\mathbf{Q}\boldsymbol{x} &= \frac{1}{2}\begin{bmatrix} \boldsymbol{x}_{\mathcal{B}^q} \\ \boldsymbol{x}_{\overline{\mathcal{B}^q}} \end{bmatrix}^\top \begin{bmatrix} \mathbf{Q}_{\mathcal{B}^q\mathcal{B}^q} & \mathbf{Q}_{\mathcal{B}^q\overline{\mathcal{B}^q}} \\ \mathbf{Q}_{\overline{\mathcal{B}^q}\mathcal{B}^q} & \mathbf{Q}_{\overline{\mathcal{B}^q}\overline{\mathcal{B}^q}} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_{\mathcal{B}^q} \\ \boldsymbol{x}_{\overline{\mathcal{B}^q}} \end{bmatrix} \\ &= \frac{1}{2}\boldsymbol{x}_{\mathcal{B}^q}^\top\mathbf{Q}_{\mathcal{B}^q\mathcal{B}^q}\boldsymbol{x}_{\mathcal{B}^q} + \boldsymbol{x}_{\mathcal{B}^q}^\top\mathbf{Q}_{\mathcal{B}^q\overline{\mathcal{B}^q}}\boldsymbol{x}_{\overline{\mathcal{B}^q}} + \frac{1}{2}\boldsymbol{x}_{\overline{\mathcal{B}^q}}^\top\mathbf{Q}_{\overline{\mathcal{B}^q}\mathcal{B}^q}\boldsymbol{x}_{\overline{\mathcal{B}^q}}. \end{aligned} \tag{5.9}$$

According to $\mathbf{Q} = \mathbf{S}\mathbf{S}^\top$, (5.9) can be rewritten as

$$\begin{aligned} \frac{1}{2}\|\mathbf{S}^\top\boldsymbol{x}\|^2 &= \frac{1}{2}\|\mathbf{S}_{\mathcal{B}^q}^\top\boldsymbol{x}_{\mathcal{B}^q} + \mathbf{S}_{\overline{\mathcal{B}^q}}^\top\boldsymbol{x}_{\overline{\mathcal{B}^q}}\|^2 \\ &= \frac{1}{2}\|\mathbf{S}_{\mathcal{B}^q}^\top\boldsymbol{x}_{\mathcal{B}^q}\|^2 + \frac{1}{2}\|\mathbf{S}_{\overline{\mathcal{B}^q}}^\top\boldsymbol{x}_{\overline{\mathcal{B}^q}}\|^2 + \boldsymbol{x}_{\mathcal{B}^q}^\top\mathbf{S}_{\mathcal{B}^q}\mathbf{S}_{\overline{\mathcal{B}^q}}^\top\boldsymbol{x}_{\overline{\mathcal{B}^q}}. \end{aligned} \tag{5.10}$$

Let

$$\mathcal{C}^q = \left\{ \boldsymbol{x}_{\mathcal{B}^q} \in \mathbb{R}^{|\mathcal{B}^q|} : \boldsymbol{R}_{\mathcal{B}^q}^\top \boldsymbol{x}_{\mathcal{B}^q} \geq r, \mathbf{0} \leq \boldsymbol{x}_{\mathcal{B}^q} \leq \boldsymbol{b}_{\mathcal{B}^q} \right\},$$
$$\mathcal{D}^q = \left\{ \boldsymbol{x}_{\mathcal{B}^q} \in \mathbb{R}^{|\mathcal{B}^q|} : \boldsymbol{e}_{\mathcal{B}^q}^\top \boldsymbol{x}_{\mathcal{B}^q} - 1 = 0 \right\}.$$

Based the decomposed form (5.3) of objective of (1.1), we fix the $\boldsymbol{x}_{\overline{\mathcal{B}^q}}$ and update the $\boldsymbol{x}_{\mathcal{B}^q}$ only. Consequently, we only need to solve the following problem of scale $|\mathcal{B}^q|$:

$$\begin{aligned} \min_{\boldsymbol{x}_{\mathcal{B}^q} \in \mathbb{R}^{|\mathcal{B}^q|}} & \frac{1}{2} \|\mathbf{S}_{\mathcal{B}^q}^\top \boldsymbol{x}_{\mathcal{B}^q}\|^2 + \delta_{\mathcal{C}^q}(\boldsymbol{x}) + \delta_{\mathcal{D}^q}(\boldsymbol{x}) \\ & \text{s.t.} \|\boldsymbol{x}_{\mathcal{B}^q}\|_0 \leq K. \end{aligned} \tag{5.11}$$

For a given penalty parameter $c_q > 0$, we transfer (5.11) into the following DC problem:

$$\min_{\boldsymbol{x}_{\mathcal{B}^q} \in \mathbb{R}^{|\mathcal{B}^q|}} \frac{1}{2} \|\mathbf{S}_{\mathcal{B}^q}^\top \boldsymbol{x}_{\mathcal{B}^q}\|^2 + \delta_{\mathcal{C}^q}(\boldsymbol{x}) + \delta_{\mathcal{D}^q}(\boldsymbol{x}) + c_q \boldsymbol{e}_{\mathcal{B}^q}^\top \boldsymbol{x}_{\mathcal{B}^q} - c_q \|\boldsymbol{x}_{\mathcal{B}^q}\|_{(K)}. \tag{5.12}$$

Evidently, (5.12) can be efficiently solved by siPDCA-mssN. Therefore, we give an efficient decomposed siPDCA-mssN (DsiPDCA-mssN) for solving (1.1), see Algorithm 5.1 for more details.

When the solutions $(\boldsymbol{x}_{\mathcal{B}^q}^{q+1}, \boldsymbol{u}_{\mathcal{B}^q}^{q+1})$ are obtained by Algorithm 3.3 for solving inner problem (5.12) with penalty parameter c_q, $(\boldsymbol{x}^{q+1}, \boldsymbol{u}^{q+1}, \boldsymbol{\xi}^{q+1})$ can be computed as follows:

$$\boldsymbol{x}_i^{q+1} := \begin{cases} (\boldsymbol{x}_{\mathcal{B}^q}^{q+1})_i, & \text{if } i \in \mathcal{B}^q \\ 0, & \text{if } i \in \overline{\mathcal{B}^q} \end{cases} \quad i = 1, \cdots, n, \tag{5.13}$$

$$\boldsymbol{u}^{q+1} := \boldsymbol{u}_{\mathcal{B}^q}^{q+1}, \tag{5.14}$$

$$\boldsymbol{\xi}_i^{q+1} := \begin{cases} (\boldsymbol{\xi}_{\mathcal{B}^q}^{q+1})_i, & \text{if } i \in \mathcal{B}^q \\ 0, & \text{if } i \in \overline{\mathcal{B}^q} \end{cases} \quad i = 1, \cdots, n, \tag{5.15}$$

where $\boldsymbol{\xi}_{\mathcal{B}^q}^{q+1} \in \partial g_c(\boldsymbol{x}^{q+1})$ and $\overline{\mathcal{B}^q} = \mathcal{B}^q \setminus \{1, \cdots, n\}$. Let

$$\boldsymbol{t}^{q+1} := \boldsymbol{x}^{q+1} + \boldsymbol{\xi}^{q+1} - \mathbf{S}\mathbf{S}^\top \boldsymbol{x}^{q+1} - c_q \boldsymbol{e} - \boldsymbol{u}^{q+1} \boldsymbol{e} \tag{5.16}$$

Algorithm 5.1. Decomposed siPDCA-mssN for (1.1)

Step 0. Give $c_0 > 0$, $K > 0$, $M > 0$, $\sigma > 1$, $\boldsymbol{R} \in \mathbb{R}^n$, $\mathbf{S} \in \mathbb{R}^{n \times m}$, $\boldsymbol{b} \in \mathbb{R}^n$ and $\widehat{\epsilon} > \bar{\epsilon} > 0$. Initialize $\boldsymbol{x}^0 \in \mathbb{R}^n$. Choose $\mathcal{B}^0$ randomly without repetition from $\{1, \cdots, n\}$ such that $|\mathcal{B}^0| = M$. Let $q = 0$.

Step 1. Compute $\boldsymbol{x}_{\mathcal{B}^q}^{q+1}, \boldsymbol{u}_{\mathcal{B}^q}^{q+1}$ by solving the inner problem (5.12) with Algorithm 3.3. Update $\boldsymbol{x}^{q+1}$, λ^{q+1}, $\boldsymbol{u}^{q+1}$, $\boldsymbol{\eta}^{q+1}$, $\boldsymbol{\zeta}^{q+1}$ and $\boldsymbol{\gamma}^{q+1}$ by (5.13), Algorithm 3.5, (5.14), and (5.19)-(5.21), respectively.

Step 2. If $\|V(\boldsymbol{x}^{q+1}; \lambda^{q+1}, \boldsymbol{u}^{q+1}, \boldsymbol{\eta}^{q+1}, \boldsymbol{\zeta}^{q+1}, \boldsymbol{\gamma}^{q+1})\|_\infty \leq \widehat{\epsilon}$ and $\|\boldsymbol{x}^{q+1}\|_0 \leq K$ hold, stop and
x
return $\boldsymbol{x}^{q+1}$.

Step 3. Update $\mathcal{B}^{q+1}$ by $\mathcal{B}^{q+1} = \mathcal{I}^{q+1} \bigcup \mathcal{B}_1^{q+1}$, where $\mathcal{I}^{q+1} = \left\{ i \in \{1, \cdots, n\} : \boldsymbol{x}_i^{q+1} \neq 0 \right\}$ and

$$\mathcal{B}_1^{q+1} = \left\{ i \in \{1, \cdots, n\} : V(\boldsymbol{x}^{q+1}; \lambda^{q+1}, \boldsymbol{u}^{q+1}, \boldsymbol{\eta}^{q+1}, \boldsymbol{\zeta}^{q+1}, \boldsymbol{\gamma}^{q+1})_i > \bar{\epsilon} \right\}.$$

If $\|\boldsymbol{x}^{q+1}\|_0 > K$, set $c_{q+1} = \sigma c_q$. Else, set $c_{q+1} = c_q$. Let $q := q + 1$ and go to **Step 1**.

and λ^{q+1} be computed by using Algorithm (**??**) with $\boldsymbol{t} = \boldsymbol{t}^{q+1}$. Based on $\boldsymbol{x}^{q+1}$ and λ^{q+1}, $(\boldsymbol{\nu}^{q+1}, \boldsymbol{\mu}^{q+1}, \boldsymbol{\zeta}^{q+1}, \boldsymbol{\eta}^{q+1}, \boldsymbol{\gamma}^{q+1})$ can be computed by

$$\boldsymbol{\nu}^{q+1} := \max(-\boldsymbol{t}^{q+1} - \lambda^{q+1}\boldsymbol{R}, 0), \tag{5.17}$$

$$\boldsymbol{\mu}^{q+1} := \max(\boldsymbol{t}^{q+1} + \lambda^{q+1}\boldsymbol{R} - \boldsymbol{b}, 0), \tag{5.18}$$

$$\boldsymbol{\zeta}_i^{q+1} := \begin{cases} \boldsymbol{\nu}_i^{q+1}, & \text{if } i \in \mathcal{I}^{q+1}, \\ 0, & \text{if } i \in \overline{\mathcal{I}^{q+1}}, \end{cases} \quad i = 1, \cdots, n, \tag{5.19}$$

$$\boldsymbol{\eta}_i^{q+1} := \begin{cases} \boldsymbol{\mu}_i^{q+1}, & \text{if } i \in \mathcal{I}^{q+1}, \\ 0, & \text{if } i \in \overline{\mathcal{I}^{q+1}}, \end{cases} \quad i = 1, \cdots, n, \tag{5.20}$$

$$\boldsymbol{\gamma}_i^{q+1} := \begin{cases} (c_q\boldsymbol{e} - \boldsymbol{\xi}^{q+1} + \boldsymbol{\mu}^{q+1} - \boldsymbol{\nu}^{q+1})_i, & \text{if } i \in \overline{\mathcal{I}^{q+1}}, \\ (c_q\boldsymbol{e} - \boldsymbol{\xi}^{q+1})_i, & \text{if } i \in \mathcal{I}^{q+1}, \end{cases} \quad i = 1, \cdots, n. \tag{5.21}$$

As a result, the violation of the first-order optimality conditions of $\boldsymbol{x}^{q+1}$ can be computed by

$$\begin{aligned} &V(\boldsymbol{x}^{q+1}; \lambda^{q+1}, \boldsymbol{u}^{q+1}, \boldsymbol{\eta}^{q+1}, \boldsymbol{\zeta}^{q+1}, \boldsymbol{\gamma}^{q+1})_i \\ =&|\mathbf{S}\mathbf{S}^\top \boldsymbol{x}^{q+1} - \lambda^{q+1}\boldsymbol{R} + \boldsymbol{u}^{q+1}\boldsymbol{e} + \boldsymbol{\eta}^{q+1} - \boldsymbol{\zeta}^{q+1} + \boldsymbol{\gamma}^{q+1}|_i, \quad i \in \{1, \cdots, n\}. \end{aligned} \tag{5.22}$$

5.2. The convergence analysis of DsiPDCA-mssN

Let $\{\boldsymbol{x}^q\}$ be the sequence generated by Algorithm 5.1. Under a mild assumption about the subsequence of $\{\boldsymbol{x}^q\}$, we display that $\{\boldsymbol{x}^q\}$ subsequentially converges to a local minimizer of (1.1).

Theorem 5.1. *Let $\{(\boldsymbol{x}^q, \lambda^q, \boldsymbol{u}^q, \boldsymbol{\eta}^q, \boldsymbol{\zeta}^q, \boldsymbol{\gamma}^q)\}$ be the sequence generated by DsiPDCA-mssN. Then the following statements hold.*
(1) *The sequence $\{(\boldsymbol{x}^q, \lambda^q, \boldsymbol{u}^q, \boldsymbol{\eta}^q, \boldsymbol{\zeta}^q, \boldsymbol{\gamma}^q)\}$ is bounded and there exists a subsequence such that*

$$\lim_{j\to\infty} (\boldsymbol{x}^{q_j}, \lambda^{q_j}, \boldsymbol{u}^{q_j}, \boldsymbol{\eta}^{q_j}, \boldsymbol{\zeta}^{q_j}, \boldsymbol{\gamma}^{q_j}) = (\boldsymbol{x}^*, \lambda^*, \boldsymbol{u}^*, \boldsymbol{\eta}^*, \boldsymbol{\zeta}^*, \boldsymbol{\gamma}^*),$$

where $(\boldsymbol{x}^*, \lambda^*, \boldsymbol{u}^*, \boldsymbol{\eta}^*, \boldsymbol{\zeta}^*, \boldsymbol{\gamma}^*)$ is an accumulation point of $\{(\boldsymbol{x}^q, \lambda^q, \boldsymbol{u}^q, \boldsymbol{\eta}^q, \boldsymbol{\zeta}^q, \boldsymbol{\gamma}^q)\}$.
(2) *Let $\bar{\epsilon}$ be any given constant satisfying $0 < \bar{\epsilon} < \min(\boldsymbol{b})$. For any $\epsilon > 0$, let*

$$\widehat{\epsilon} = \bar{\epsilon} + (2 + 2\max_{1\le i\le n} \|\mathbf{S}_i\mathbf{S}\| + 2\|\boldsymbol{R}\|_\infty)\epsilon.$$

Then $\exists M > 0$, $\forall\, j > M$ such that $\|\boldsymbol{x}^{q_j} - \boldsymbol{x}^{q_{j+1}}\| \le \epsilon$, $|\lambda^{q_j} - \lambda^{q_{j+1}}| \le \epsilon$ and $|\boldsymbol{u}^{q_j} - \boldsymbol{u}^{q_{j+1}}| \le \epsilon$ hold. If exists a $\bar{j}$ such that $\bar{j} > M$, $q_{\bar{j}+1} = q_{\bar{j}} + 1$, $\|\boldsymbol{x}^{q_{\bar{j}+1}}\|_0 \le K$ and $c_{q_{\bar{j}}} = c_{q_{\bar{j}}-1}$, then $\boldsymbol{x}^{q_{\bar{j}+1}}$ is an $\widehat{\epsilon}$-inexact solution of (1.1) and $\boldsymbol{x}^*$ is a local minimizer of (1.1).

Proof. For statement (1), since $\boldsymbol{x}_{\mathcal{B}^q}^{q+1}$ is the solution generated by Algorithm 3.3 for solving the inner problem in (5.12), then $\boldsymbol{x}_{\mathcal{B}^q}^{q+1}$ satisfies $\mathbf{0} \le \boldsymbol{x}_{\mathcal{B}^q}^{q+1} \le \boldsymbol{b}_{\mathcal{B}^q}$ and $\boldsymbol{e}^\top \boldsymbol{x}_{\mathcal{B}^q}^{q+1} = 1$. This, together with the definition of $\boldsymbol{x}^{q+1}$ in (5.13), yields that

$$\|\boldsymbol{x}^{q+1}\| = \|\boldsymbol{x}_{\mathcal{B}^q}^{q+1}\| \le \|\boldsymbol{x}_{\mathcal{B}^q}^{q+1}\|_1 = \boldsymbol{e}_{\mathcal{B}^q}^\top \boldsymbol{x}_{\mathcal{B}^q}^{q+1} = 1.$$

This implies that $\{\boldsymbol{x}^q\}$ is bounded. Consequently, we can immediately obtain that $\{\boldsymbol{\xi}^q\}$ is also bounded because $g_c(\boldsymbol{x})$ is a finite-valued convex function. From Proposition 4.1, it follows that $\{\boldsymbol{u}^q\}$ and $\{\boldsymbol{v}^q\}$ are bounded. From Algorithm 3.5, we have $\lambda^{q+1} \in \big[0, \max(\mathcal{H}^{q+1})\big]$, where

$$\mathcal{H}^{q+1} = \Big\{ \tfrac{\boldsymbol{t}_i^{q+1} - \boldsymbol{b}_i}{\boldsymbol{R}_i}, -\tfrac{\boldsymbol{t}_i^{q+1}}{\boldsymbol{R}_i} \Big| \boldsymbol{R}_i \ne 0, i \in \{1, \cdots, n\} \Big\} \cup \{0\}$$

and $\boldsymbol{t}^{q+1}$ is defined in (5.16). This implies that $\{\lambda^q\}$ is bounded. As a result, the boundedness of $\{\boldsymbol{\mu}^q\}$ and $\{\boldsymbol{\nu}^q\}$ can be obtained from their definitions, see (5.17) and (5.18). Consequently, the boundedness of $\{\boldsymbol{\zeta}^q\}$, $\{\boldsymbol{\eta}^q\}$ and $\{\boldsymbol{\gamma}^q\}$ can be obtained from their definitions, see (5.19)-(5.21). Hence, it holds that $\{(\boldsymbol{x}^q, \lambda^q, \boldsymbol{u}^q, \boldsymbol{\eta}^q, \boldsymbol{\zeta}^q, \boldsymbol{\gamma}^q)\}_{q=1}^{\infty}$ is bounded and there exists a subsequence $\{(\boldsymbol{x}^{q_j}, \lambda^{q_j}, \boldsymbol{u}^{q_j}, \boldsymbol{\eta}^{q_j}, \boldsymbol{\zeta}^{q_j}, \boldsymbol{\gamma}^{q_j})\}_{j=1}^{\infty}$ such that

$$\lim_{j\to\infty}(\boldsymbol{x}^{q_j}, \lambda^{q_j}, \boldsymbol{u}^{q_j}, \boldsymbol{\eta}^{q_j}, \boldsymbol{\zeta}^{q_j}, \boldsymbol{\gamma}^{q_j}) = (\boldsymbol{x}^*, \lambda^*, \boldsymbol{u}^*, \boldsymbol{\eta}^*, \boldsymbol{\zeta}^*, \boldsymbol{\gamma}^*).$$

For statement (2), the result of statement (1) implies that $\forall\,\epsilon > 0$, $\exists M$ such that $\forall\, j > M$, $\|\boldsymbol{x}^{q_j} - \boldsymbol{x}^{q_{j+1}}\| \le \epsilon$, $|\lambda^{q_j} - \lambda^{q_{j+1}}| \le \epsilon$ and $|\boldsymbol{u}^{q_j} - \boldsymbol{u}^{q_{j+1}}| \le \epsilon$. Since $\bar{j} > M$ and $c_{q_{\bar{j}}} = c_{q_{\bar{j}}-1}$, $q_{\bar{j}+1} = q_{\bar{j}} + 1$, we have $\|\boldsymbol{x}^{q_{\bar{j}}} - \boldsymbol{x}^{q_{\bar{j}+1}}\| \le \epsilon$, $|\lambda^{q_{\bar{j}}} - \lambda^{q_{\bar{j}+1}}| \le \epsilon$ and $|\boldsymbol{u}^{q_{\bar{j}}} - \boldsymbol{u}^{q_{\bar{j}+1}}| \le \epsilon$. Let $\mathcal{B}^{q_{\bar{j}}}$ be the work set of Algorithm 5.1 at the $q_{\bar{j}}$-th step such that

$$\begin{aligned}&V(\boldsymbol{x}^{q_{\bar{j}}}; \lambda^{q_{\bar{j}}}, \boldsymbol{u}^{q_{\bar{j}}}, \boldsymbol{\eta}^{q_{\bar{j}}}, \boldsymbol{\zeta}^{q_{\bar{j}}}, \boldsymbol{\gamma}^{q_{\bar{j}}})_i > \bar{\epsilon}, \forall\, i \in \mathcal{B}^{q_{\bar{j}}} \setminus \mathcal{I}^{q_{\bar{j}}},\\ &V(\boldsymbol{x}^{q_{\bar{j}}}; \lambda^{q_{\bar{j}}}, \boldsymbol{u}^{q_{\bar{j}}}, \boldsymbol{\eta}^{q_{\bar{j}}}, \boldsymbol{\zeta}^{q_{\bar{j}}}, \boldsymbol{\gamma}^{q_{\bar{j}}})_i \le \bar{\epsilon}, \forall\, i \in \overline{\mathcal{B}^{q_{\bar{j}}}},\end{aligned}$$

where $\mathcal{I}^{q_{\bar{j}}}$ is the index set of nonzero elements of $\boldsymbol{x}^{q_{\bar{j}}}$. Due to $q_{\bar{j}+1} = q_{\bar{j}} + 1$, we can obtain that $\boldsymbol{x}^{q_{\bar{j}+1}}$ is generated at the $q_{\bar{j}}$-th step of Algorithm 5.1 and that the entries of $\boldsymbol{x}^{q_{\bar{j}}}_{\overline{\mathcal{B}^{q_{\bar{j}}}}}$, $\boldsymbol{x}^{q_{\bar{j}+1}}_{\overline{\mathcal{B}^{q_{\bar{j}}}}}$, $\boldsymbol{\xi}^{q_{\bar{j}}}_{\overline{\mathcal{B}^{q_{\bar{j}}}}}$ and $\boldsymbol{\xi}^{q_{\bar{j}+1}}_{\overline{\mathcal{B}^{q_{\bar{j}}}}}$ are all zero. Hence, it holds that

$$\mathbf{S}^\top \boldsymbol{x}^{q_{\bar{j}}} = \mathbf{S}^\top_{\mathcal{B}^{q_{\bar{j}}}} \boldsymbol{x}^{q_{\bar{j}}}_{\mathcal{B}^{q_{\bar{j}}}}, \quad \mathbf{S}^\top \boldsymbol{x}^{q_{\bar{j}+1}} = \mathbf{S}^\top_{\mathcal{B}^{q_{\bar{j}}}} \boldsymbol{x}^{q_{\bar{j}+1}}_{\mathcal{B}^{q_{\bar{j}}}}$$

and $\forall\, i \in \overline{\mathcal{B}^{q_{\bar{j}}}}$,

$$\boldsymbol{t}_i^{q_{\bar{j}}} = -c_{q_{\bar{j}}-1} - \boldsymbol{u}^{q_{\bar{j}}} - \mathbf{S}_i \mathbf{S}^\top_{\mathcal{B}^{q_{\bar{j}}}} \boldsymbol{x}^{q_{\bar{j}}}_{\mathcal{B}^{q_{\bar{j}}}}, \quad \boldsymbol{t}_i^{q_{\bar{j}+1}} = -c_{q_{\bar{j}}} - \boldsymbol{u}^{q_{\bar{j}+1}} - \mathbf{S}_i \mathbf{S}^\top_{\mathcal{B}^{q_{\bar{j}}}} \boldsymbol{x}^{q_{\bar{j}+1}}_{\mathcal{B}^{q_{\bar{j}}}},$$

where $\boldsymbol{t}^{q_{\bar{j}}}$ and $\boldsymbol{t}^{q_{\bar{j}+1}}$ can be computed by using the expression in (5.16). This, together with (5.8) and the definition of $\boldsymbol{\nu}^{q_{\bar{j}}}$ and $\boldsymbol{\mu}^{q_{\bar{j}}}$, yields that $\forall\, i \in \overline{\mathcal{B}^{q_{\bar{j}}}}$,

$$\begin{aligned}&V(\boldsymbol{x}^{q_{\bar{j}}}; \lambda^{q_{\bar{j}}}, \boldsymbol{u}^{q_{\bar{j}}}, \boldsymbol{\eta}^{q_{\bar{j}}}, \boldsymbol{\zeta}^{q_{\bar{j}}}, \boldsymbol{\gamma}^{q_{\bar{j}}}))_i\\ =&|\mathbf{S}\mathbf{S}^\top \boldsymbol{x}^{q_{\bar{j}}} - \lambda^{q_{\bar{j}}} \boldsymbol{R} + \boldsymbol{u}^{q_{\bar{j}}} \boldsymbol{e} + c_{q_{\bar{j}}-1} \boldsymbol{e} - \boldsymbol{\xi}^{q_{\bar{j}}} - \boldsymbol{\nu}^{q_{\bar{j}}} + \boldsymbol{\mu}^{q_{\bar{j}}}|_i\\ =&|-\boldsymbol{t}_i^{q_{\bar{j}}} - \lambda^{q_{\bar{j}}} \boldsymbol{R}_i + \max(\boldsymbol{t}_i^{q_{\bar{j}}} + \lambda^{q_{\bar{j}}} \boldsymbol{R}_i - \boldsymbol{b}_i, 0) - \max(-\boldsymbol{t}_i^{q_{\bar{j}}} - \lambda^{q_{\bar{j}}} \boldsymbol{R}_i, 0)|.\end{aligned}$$

and

$$\begin{aligned}&V(\boldsymbol{x}^{q_{\bar{j}+1}}; \lambda^{q_{\bar{j}+1}}, \boldsymbol{u}^{q_{\bar{j}+1}}, \boldsymbol{\mu}^{q_{\bar{j}+1}}, \boldsymbol{\zeta}^{q_{\bar{j}+1}}, \boldsymbol{\gamma}^{q_{\bar{j}+1}})_i\\ =&|-\boldsymbol{t}_i^{q_{\bar{j}+1}} - \lambda^{q_{\bar{j}+1}} \boldsymbol{R}_i + \max(\boldsymbol{t}_i^{q_{\bar{j}+1}} + \lambda^{q_{\bar{j}+1}} \boldsymbol{R}_i - \boldsymbol{b}_i, 0) - \max(-\boldsymbol{t}_i^{q_{\bar{j}+1}} - \lambda^{q_{\bar{j}+1}} \boldsymbol{R}_i, 0)|.\end{aligned}$$

Next, we prove that $\forall\, i \in \overline{\mathcal{B}^{q_{\bar{j}}}}$,

$$V(\boldsymbol{x}^{q_{\bar{j}+1}}; \lambda^{q_{\bar{j}+1}}, \boldsymbol{u}^{q_{\bar{j}+1}}, \boldsymbol{\eta}^{q_{\bar{j}+1}}, \boldsymbol{\zeta}^{q_{\bar{j}+1}}, \boldsymbol{\gamma}^{q_{\bar{j}+1}})_i \le \widehat{\epsilon}.$$

For simplicity of notation, we set

$$V^{q_{\bar{j}}} = V(\boldsymbol{x}^{q_{\bar{j}}}; \lambda^{q_{\bar{j}}}, \boldsymbol{u}^{q_{\bar{j}}}, \boldsymbol{\eta}^{q_{\bar{j}}}, \boldsymbol{\zeta}^{q_{\bar{j}}}, \boldsymbol{\gamma}^{q_{\bar{j}}}, \quad V^{q_{\bar{j}+1}} = V(\boldsymbol{x}^{q_{\bar{j}+1}}; \lambda^{q_{\bar{j}+1}}, \boldsymbol{u}^{q_{\bar{j}+1}}, \boldsymbol{\eta}^{q_{\bar{j}+1}}, \boldsymbol{\zeta}^{q_{\bar{j}+1}}, \boldsymbol{\gamma}^{q_{\bar{j}+1}}).$$

Suppose that $\exists\, i \in \overline{\mathcal{B}^{q_{\bar{j}}}}$, $\boldsymbol{t}_i^{q_{\bar{j}}} + \lambda^{q_{\bar{j}}} \boldsymbol{R}_i > \boldsymbol{b}_i$, then $V_i^{q_{\bar{j}}} = \boldsymbol{b}_i \ge min(\boldsymbol{b}) > \widehat{\epsilon} > \bar{\epsilon}$, which is contrary to $V_i^{q_{\bar{j}}} \le \bar{\epsilon}$, $\forall\, i \in \overline{\mathcal{B}^{q_{\bar{j}}}}$. Thus, we have $\boldsymbol{t}_i^{q_{\bar{j}}} + \lambda^{q_{\bar{j}}} \boldsymbol{R}_i \le \boldsymbol{b}_i$, $\forall\, i \in \overline{\mathcal{B}^{q_{\bar{j}}}}$. As a result, we divide the proof

into the following two cases.

Case 1. When $\boldsymbol{t}_i^{q_{\bar{j}}} + \lambda^{q_{\bar{j}}}\boldsymbol{R}_i \le 0$, it holds that

$$V_i^{q_{\bar{j}}} = -2\boldsymbol{t}_i^{q_{\bar{j}}} - 2\lambda^{q_{\bar{j}}}\boldsymbol{R}_i = 2(c_{q_{\bar{j}-1}} + \boldsymbol{u}^{q_{\bar{j}}} + \mathbf{S}_i\mathbf{S}_{\mathcal{B}^{q_{\bar{j}}}}^\top \boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}}}) - 2\lambda^{q_{\bar{j}}}\boldsymbol{R}_i \le 0.$$

Consequently,

$$\begin{aligned}
\boldsymbol{t}_i^{q_{\bar{j}+1}} + \lambda^{q_{\bar{j}+1}}\boldsymbol{R}_i &= (-c_{q_{\bar{j}}} - \boldsymbol{u}^{q_{\bar{j}+1}} - \mathbf{S}_i\mathbf{S}_{\mathcal{B}^{q_{\bar{j}}}}^\top \boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}+1}}) + \lambda^{q_{\bar{j}+1}}\boldsymbol{R}_i \\
&= (-c_{q_{\bar{j}-1}} - \boldsymbol{u}^{q_{\bar{j}}} - \mathbf{S}_i\mathbf{S}_{\mathcal{B}^{q_{\bar{j}}}}^\top \boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}}}) + \lambda^{q_{\bar{j}}}\boldsymbol{R}_i \\
&\quad + ((\boldsymbol{u}^{q_{\bar{j}}} - \boldsymbol{u}^{q_{\bar{j}+1}}) - \mathbf{S}_i\mathbf{S}_{\mathcal{B}^{q_{\bar{j}}}}^\top (\boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}+1}} - \boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}}})) + (\lambda^{q_{\bar{j}+1}} - \lambda^{q_{\bar{j}}})\boldsymbol{R}_i \\
&\le ((\boldsymbol{u}^{q_{\bar{j}}} - \boldsymbol{u}^{q_{\bar{j}+1}}) - \mathbf{S}_i\mathbf{S}_{\mathcal{B}^{q_{\bar{j}}}}^\top (\boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}+1}} - \boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}}}) + (\lambda^{q_{\bar{j}+1}} - \lambda^{q_{\bar{j}}})\boldsymbol{R}_i) \\
&\le (1 + \|\mathbf{S}_i\mathbf{S}\| + |\boldsymbol{R}_i|)\epsilon \le \frac{\widehat{\epsilon} - \bar{\epsilon}}{2},
\end{aligned}$$

where the first equality is due to $c_{q_{\bar{j}}} = c_{q_{\bar{j}}-1}$, the first inequality is due to $\boldsymbol{t}_i^{q_{\bar{j}}} + \lambda^{q_{\bar{j}}}\boldsymbol{R}_i \le 0$, the last inequality is due to the definition of $\widehat{\epsilon}$. This, together with $V_i^{q_{\bar{j}}} \le \bar{\epsilon}$, implies that when $\boldsymbol{t}_i^{q_{\bar{j}+1}} + \lambda^{q_{\bar{j}+1}}\boldsymbol{R}_i \le 0$,

$$\begin{aligned}
V_i^{q_{\bar{j}+1}} &= -2\boldsymbol{t}_i^{q_{\bar{j}+1}} - 2\lambda^{q_{\bar{j}+1}}\boldsymbol{R}_i \\
&= V_i^{q_{\bar{j}}} - 2((\boldsymbol{u}^{q_{\bar{j}}} - \boldsymbol{u}^{q_{\bar{j}+1}}) - \mathbf{S}_i\mathbf{S}_{\mathcal{B}^{q_{\bar{j}}}}^\top (\boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}+1}} - \boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}}}) + (\lambda^{q_{\bar{j}+1}} - \lambda^{q_{\bar{j}}})\boldsymbol{R}_i) \\
&\le \bar{\epsilon} - 2((\boldsymbol{u}^{q_{\bar{j}}} - \boldsymbol{u}^{q_{\bar{j}+1}}) - \mathbf{S}_i\mathbf{S}_{\mathcal{B}^{q_{\bar{j}}}}^\top (\boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}+1}} - \boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}}}) + (\lambda^{q_{\bar{j}+1}} - \lambda^{q_{\bar{j}}})\boldsymbol{R}_i) \\
&\le \bar{\epsilon} + 2(1 + \|\mathbf{S}_i\mathbf{S}\| + |\boldsymbol{R}_i|)\epsilon \le \widehat{\epsilon}
\end{aligned}$$

and when $0 < \boldsymbol{t}_i^{q_{\bar{j}+1}} + \lambda^{q_{\bar{j}+1}}\boldsymbol{R}_i < \frac{\widehat{\epsilon}-\bar{\epsilon}}{2}$, $V_i^{q_{\bar{j}+1}} = \boldsymbol{t}_i^{q_{\bar{j}+1}} + \lambda^{q_{\bar{j}+1}}\boldsymbol{R}_i \le \frac{\widehat{\epsilon}-\bar{\epsilon}}{2} < \widehat{\epsilon}$.

Case 2. When $0 < \boldsymbol{t}_i^{q_{\bar{j}}} + \lambda^{q_{\bar{j}}}\boldsymbol{R}_i < \boldsymbol{b}_i$, it holds that $V_i^{q_{\bar{j}}} = \boldsymbol{t}_i^{q_{\bar{j}}} + \lambda^{q_{\bar{j}}}\boldsymbol{R}_i \le \bar{\epsilon}$. This implies that

$$\begin{aligned}
\boldsymbol{t}_i^{q_{\bar{j}+1}} + \lambda^{q_{\bar{j}+1}}\boldsymbol{R}_i &= V_i^{q_{\bar{j}}} + (\boldsymbol{u}^{q_{\bar{j}}} - \boldsymbol{u}^{q_{\bar{j}+1}}) - \mathbf{S}_i\mathbf{S}_{\mathcal{B}^{q_{\bar{j}}}}^\top (\boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}+1}} - \boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}}}) + (\lambda^{q_{\bar{j}+1}} - \lambda^{q_{\bar{j}}})\boldsymbol{R}_i \\
&\le \bar{\epsilon} + ((\boldsymbol{u}^{q_{\bar{j}}} - \boldsymbol{u}^{q_{\bar{j}+1}}) - \mathbf{S}_i\mathbf{S}_{\mathcal{B}^{q_{\bar{j}}}}^\top (\boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}+1}} - \boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}}}) + (\lambda^{q_{\bar{j}+1}} - \lambda^{q_{\bar{j}}})\boldsymbol{R}_i) \\
&\le \bar{\epsilon} + (1 + \|\mathbf{S}_i\mathbf{S}\| + |\boldsymbol{R}_i|)\epsilon \le \frac{\widehat{\epsilon} + \bar{\epsilon}}{2} < \widehat{\epsilon}.
\end{aligned}$$

Consequently, it follows that when $\boldsymbol{t}_i^{q_{\bar{j}+1}} + \lambda^{q_{\bar{j}+1}}\boldsymbol{R}_i \le 0$,

$$\begin{aligned}
V_i^{q_{\bar{j}+1}} &= -2\boldsymbol{t}_i^{q_{\bar{j}+1}} - 2\lambda^{q_{\bar{j}+1}}\boldsymbol{R}_i \\
&= -2V_i^{q_{\bar{j}}} - 2((\boldsymbol{u}^{q_{\bar{j}}} - \boldsymbol{u}^{q_{\bar{j}+1}}) - \mathbf{S}_i\mathbf{S}_{\mathcal{B}^{q_{\bar{j}}}}^\top (\boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}+1}} - \boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}}}) + (\lambda^{q_{\bar{j}+1}} - \lambda^{q_{\bar{j}}})\boldsymbol{R}_i) \\
&\le -2((\boldsymbol{u}^{q_{\bar{j}}} - \boldsymbol{u}^{q_{\bar{j}+1}}) - \mathbf{S}_i\mathbf{S}_{\mathcal{B}^{q_{\bar{j}}}}^\top (\boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}+1}} - \boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}}}) + (\lambda^{q_{\bar{j}+1}} - \lambda^{q_{\bar{j}}})\boldsymbol{R}_i) \\
&\le 2(1 + \|\mathbf{S}_i\mathbf{S}\| + |\boldsymbol{R}_i|)\epsilon < \widehat{\epsilon}
\end{aligned}$$

and when $0 < \boldsymbol{t}_i^{q_{\bar{j}+1}} + \lambda^{q_{\bar{j}+1}}\boldsymbol{R}_i < \widehat{\epsilon}$, $V_i^{q_{\bar{j}+1}} = \boldsymbol{t}_i^{q_{\bar{j}+1}} + \lambda^{q_{\bar{j}+1}}\boldsymbol{R}_i < \widehat{\epsilon}$. Therefore, we have

$$V(\boldsymbol{x}^{q_{\bar{j}+1}}, \lambda^{q_{\bar{j}+1}}, \boldsymbol{u}^{q_{\bar{j}+1}}, \boldsymbol{\eta}^{q_{\bar{j}+1}}, \boldsymbol{\zeta}^{q_{\bar{j}+1}}, \boldsymbol{\gamma}^{q_{\bar{j}+1}})_i \le \widehat{\epsilon}, \quad \forall\, i \in \overline{\mathcal{B}^{q_{\bar{j}}}}.$$

The optimality of $\boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}+1}}$ for solving the $\mathcal{B}^{q_{\bar{j}}}$-subproblem in (5.12) yields that

$$V(\boldsymbol{x}_{\mathcal{B}^{q_{\bar{j}}}}^{q_{\bar{j}+1}}, \lambda_i^{q_{\bar{j}+1}}, \boldsymbol{u}_i^{q_{\bar{j}+1}}, \boldsymbol{\eta}_i^{q_{\bar{j}+1}}, \boldsymbol{\zeta}_i^{q_{\bar{j}+1}}, \boldsymbol{\gamma}_i^{q_{\bar{j}+1}})_i = 0, \quad \forall\, i \in \mathcal{B}^{q_{\bar{j}}}.$$

This, together with $\|\boldsymbol{x}^{q_{\bar{j}+1}}\|_0 \le K$ and definition of $(\boldsymbol{x}^{q_{\bar{j}+1}}, \lambda^{q_{\bar{j}+1}}, \boldsymbol{u}^{q_{\bar{j}+1}}, \boldsymbol{\eta}^{q_{\bar{j}+1}}, \boldsymbol{\zeta}^{q_{\bar{j}+1}}, \boldsymbol{\gamma}^{q_{\bar{j}+1}})$ in (5.13), (5.14) and (5.19)-(5.21), implies that $(\boldsymbol{x}^{q_{\bar{j}+1}}, \lambda^{q_{\bar{j}+1}}, \boldsymbol{u}^{q_{\bar{j}+1}}, \boldsymbol{\eta}^{q_{\bar{j}+1}}, \boldsymbol{\zeta}^{q_{\bar{j}+1}}, \boldsymbol{\gamma}^{q_{\bar{j}+1}})$ satisfies all

the conditions defined in (4.17) except the first one. Consequently, it follows that $\boldsymbol{x}^{q_{\bar{j}+1}}$ is an $\widehat{\epsilon}$-local minimizer of the CCMV problem (1.1). From the arbitrariness of $\epsilon > 0$ and $0 \leq \bar{\epsilon} \leq \min(\boldsymbol{b})$, we can obtain that

$$V(\boldsymbol{x}^*, \lambda^*, \boldsymbol{u}^*, \boldsymbol{\eta}^*, \boldsymbol{\zeta}^*, \boldsymbol{\gamma}^*)_i = 0, \quad i \in \{1, \cdots, n\}$$

and $\|\boldsymbol{x}^*\|_0 \leq K$. This implies that $(\boldsymbol{x}^*, \lambda^*, \boldsymbol{u}^*, \boldsymbol{\eta}^*, \boldsymbol{\zeta}^*, \boldsymbol{\gamma}^*)$ satisfy the first-order optimality conditions in (4.17) exactly. By noting that MFCQ conditions hold at $\boldsymbol{x}^*$, then it follows that $\boldsymbol{x}^*$ is a local minimizer of (1.1). This completes the proof. □

6. Numerical Experiments

In order to demonstrate the effectiveness of the proposed siPDCA-mssN and DsiPDCA-mssN for solving (1.1), we perform comparison numerical experiments with other methods on both real-world market data set and simulated data set. In addition, to illustrate the out-of-sample performance of the CCMV model, we compare the Sharp ratio of the solutions generated by siPDCA-mssN for solving the CCMV model in (1.1) with those of other portfolio selection models.

Computational environment. All experiments are performed in Matlab 2020a on a 64-bit PC with an Intel(R) Xeon(R) CPU E5-2609 v2 (2.50GHz) x(2 processor) and 56GB of RAM.

6.1. Numerical performance on real-world market data

Experimental data. The real-world market data used in this paper are the index tracking problem data selected from OR-Library, which is also described in [51, 52]. The selected data groups include weekly return of constituents on Nikkei 225 (Japan), Standard & Poors (S&P) 500 (US), Russell 2000 (US) and Russell 3000 (US) from the year 1992 to 1997 with the variable dimensions n = 225, 457, 1319 and 2152, respectively. In each group, we use the data between 1992 and 1996 as the training set ($m = 240$) to estimate mean return vector $\boldsymbol{R}$ and matrix $\mathbf{S}$, and use the data of 1997 as the test set. For more information about this database, we refer to the homepage of OR-Library [1)] .

Comparing algorithms. To display the numerical performance of siPDCA-mssN, we compare it with the standard CPLEX(12.9) solver and the penalty proximal alternating linearized minimization (PPALM) method for solving the CCMV problem in (1.1) by performing experiments on real-world market data. In order to use the CPLEX solver, we reformulate the CCMV problem (1.1) into a standard mixed-integer quadratic programming (MIQP) by introducing $0-1$ variables. To employ PPALM method [25], we divide the feasible set of (1.1) into the following two parts: $\mathcal{C}_1 = \left\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{e}^\top \boldsymbol{x} - 1 = 0, \boldsymbol{R}^\top \boldsymbol{x} \geq r\right\}$ and $\mathcal{D}_1 = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_0 \leq K, \mathbf{0} \leq \boldsymbol{x} \leq \boldsymbol{b}\}$. The PPALM method for solving (1.1) is presented in Algorithm 6.1.

[1)] http://people.brunel.ac.uk/ mastjjb/jeb/info.html

Algorithm 6.1. PPALM method [25] for solving (1.1)

Step 0. Give $\rho_0 > 0, K > 0, \sigma > 1, \gamma_1, \gamma_2 > 1, \boldsymbol{R} \in \mathbb{R}^n, \mathbf{Q} \in \mathbb{R}^{n\times n}, \boldsymbol{b} \in \mathbb{R}^n, \varepsilon_I \geq 0, \varepsilon_O \geq 0$. Initialize $(\boldsymbol{x}^0, \boldsymbol{y}^0) \in \mathbb{R}^n \times \mathbb{R}^n$. Set $k = 0$.

Step 1. Set $l = 0$, $L_1^k = \|\mathbf{Q}\| + \rho_k$, $L_2^k = \rho_k$, $t_1^k = \gamma_1 L_1^k$, $t_2^k = \gamma_2 L_2^k$ and apply the proximal alternating linearized minimization method [24] to find an approximate critical point $(\boldsymbol{x}^k, \boldsymbol{y}^k) \in \mathcal{C}_1 \times \mathcal{D}_1$ of the penalty subproblem

$$\min_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n} \Psi_{\rho_k}(\boldsymbol{x}, \boldsymbol{y}) = \delta_{\mathcal{C}_1}(\boldsymbol{x}) + \frac{1}{2}\boldsymbol{x}^\top \mathbf{Q}\boldsymbol{x} + \frac{\rho_k}{2}\|\boldsymbol{x} - \boldsymbol{y}\|^2 + \delta_{\mathcal{D}_1}(\boldsymbol{y}) \tag{6.1}$$

by performing steps **(1.1)**-**(1.3)**:

(1.1) Let $\boldsymbol{w}_1 = \boldsymbol{x}_l^k - 1/t_1^k(\mathbf{Q}\boldsymbol{x}_l^k + \rho_k(\boldsymbol{x}_l^k - \boldsymbol{y}_l^k))$, compute $\boldsymbol{x}_{l+1}^k = \Pi_{\mathcal{C}_1}(\boldsymbol{w}_1)$.

(1.2) Let $\boldsymbol{w}_2 = \boldsymbol{y}_l^k - 1/t_2^k \rho_k(\boldsymbol{y}_l^k - \boldsymbol{x}_{l+1}^k)$, compute $\boldsymbol{y}_{l+1}^k = \Pi_{\mathcal{D}_1}(\boldsymbol{w}_2)$.

(1.3) If $\max\left\{\frac{\|\boldsymbol{x}_{l+1}^k - \boldsymbol{x}_l^k\|_\infty}{\max(\|\boldsymbol{x}_l^k\|_\infty, 1)}, \frac{\|\boldsymbol{y}_{l+1}^k - \boldsymbol{y}_l^k\|_\infty}{\max(\|\boldsymbol{y}_l^k\|_\infty, 1)}\right\} \leq \varepsilon_I$ holds, set $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}) = (\boldsymbol{x}_{l+1}^k, \boldsymbol{y}_{l+1}^k)$ and go to **Step 2**. Else, set $l := l + 1$ and go to step **(1.1)**.

Step 2. Set $\rho_{k+1} = \sigma\rho_k$, $(\boldsymbol{x}_0^{k+1}, \boldsymbol{y}_0^{k+1}) = (\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1})$ and $k := k + 1$, go to **Step 1**.

Parameter setting for model. The upper limit on the number of stocks included in the portfolio is set as $K = 5, 10, 20, 30, 40$. The upper bound vector of investment proportion of assets is set as $\boldsymbol{b}_i = 0.3$, $1 \leq i \leq n$, which also can be set as different value for each asset under Assumption 2.1. The vector $\boldsymbol{R}$ is set as the expected return of n assets in the training set. The minimum profit target r is set as $r = (\sum_{i=1}^n \boldsymbol{R}_i - \max(\boldsymbol{R}))/n$.

Parameter setting for siPDCA-mssN. To guarantee that the solution satisfies the cardinality constraint and the penalized problem in (1.3) can be relatively easily solved, we set the penalty parameter c by a gradually increasing strategy. In addition, we start siPDCA-mssN with a small penalty parameter c_0 for each group data: $c_0 = 5 \times 10^{-6}(n = 225), c_0 = 10^{-5}(n = 457), c_0 = 5\times 10^{-5}(n = 1319), c_0 = 10^{-4}(n = 2152)$. For siPDCA-mssN, it is essential to choose the appropriate and effective proximal parameter for both theory analysis and numerical efficiency. The general rule is to choose the proximal parameter as small as possible so that the algorithm can take a large step and the inner problem still can be solved relatively easily. Then we choose the values of α through a 5-fold cross-validation procedure from the set of candidates

$$\left\{10^{-4}, 2 \times 10^{-4}, 4 \times 10^{-4}, 6 \times 10^{-4}, 8 \times 10^{-4}, 10^{-3}\right\}.$$

Notice that the sieving parameter $\kappa \in (0, 1)$ is used to balance the efficiency of siPDCA-mssN and the inexactness of solution. Large κ requires the new stability center to be more accurate, which results in fewer serious steps being performed. Hence, to obtain more serious steps, the sieving parameter should be set as small as possible. In this experiment, we set the sieving parameter as $\kappa = 0.01$. In addition, we use an adaptive strategy to set the sequence $\{\epsilon_k\}$: if the sieving conditions in (5.1) hold, set $\epsilon_{k+1} = \max(\frac{k}{1+k}, \rho_1)\epsilon_k$, otherwise, set $\epsilon_{k+1} = \max(\frac{k}{1+k}, \rho_2)\epsilon_k$. Then we set $\epsilon_1 = 10^{-1}, \rho_1 = 0.99$ and $\rho_2 = 0.9$.

Parameter setting for PPALM and CPLEX. The parameters for PPALM method are set as: $\rho_0 = 0.1$, $\sigma = \sqrt{10}$, $\varepsilon_I = 5 \times 10^{-5}$, $\gamma_1 = \gamma_2 = 1.01$. The parameters of CPLEX solver are set as the default value of software.

Initialization. The initial solutions of siPDCA-mssN and PPALM method are randomly generated from a uniform distribution with range $[0, 0.3]$. We use the default initial solution of the software for the CPLEX solver.

Termination criterion. The termination criterion of siPDCA-mssN is set as: $\frac{\|\boldsymbol{y}^{k+1}-\boldsymbol{x}^k\|}{\max(1,\|\boldsymbol{x}^k\|)} \leq \varepsilon$ and $\max(\|\Delta_1^{k+1}\|, |\Delta_2^{k+1}|) \leq \varepsilon$ with $\varepsilon = 10^{-3}$. We set the termination conditions of PPALM method as

$$\frac{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\|_\infty}{\max(\|\boldsymbol{x}^k\|, 1)} \leq \varepsilon_O \quad \text{and} \quad \|\boldsymbol{x}^{k+1}_{\mathcal{I}_\epsilon}\|_0 \leq K, \tag{6.2}$$

where $\mathcal{I}_\epsilon = \{i \in \{1, 2, \cdots n\} : |\boldsymbol{x}_i^{k+1}| > \epsilon\}$. Then we set $\epsilon = 10^{-5}$ and $\varepsilon_O = 10^{-3}$. For CPLEX solver, we use the default termination criterion of software.

As a comparison, numerical results for the solving time ($time/s$) and optimal value (F) of the objective function $F(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \mathbf{Q}\boldsymbol{x}$ across 50 run of tests of siPDCA-mssN, PPALM method and CPLEX solver are presented in Table 6.1. As a result from Table 6.1, one can see that our siPDCA-mssN outperforms PPALM method and the CPLEX solver from the solving time and optimal objective function value. The optimal value of siPDCA-mssN is smaller than that of other methods in most cases. The solving time of siPDCA-mssN is less than that of PPALM method and the CPLEX solver for all the situations. In most cases, for the same data group, the larger the values of K is, the less the solving time is taken and the small the optimal value is obtained.

6.2. Numerical performance on large-scale simulated data

To demonstrate the effectiveness of siPDCA-mssN and DsiPDCA-mssN for solving the large-scale CCMV portfolio selection problem, we compare siPDCA-mssN and DsiPDCA-mssN with PPALM method for solving (1.1) by performing numerical experiments on the large-scale simulated data set.

Experimental data. The simulated data sets with the variable dimensions n = 10000, 20000, 40000 and 80000 are used, which are obtained by adding Gaussian noise on the data set Russell 3000 mentioned above. The Gaussian noise vectors are generated from the zero-mean multivariate Gaussian distribution, whose standard deviation is 0.01.

Parameter setting for model. The upper limit on the number of stocks included in the portfolio is set as $K = 10, 20, 30, 40, 50, 60$. The upper bound vector of investment proportion of assets is set as $\boldsymbol{b}_i = 0.5, 1 \leq i \leq n$. We set $\boldsymbol{R}$ and r in the same way as in Subsection 6.1.

Parameter setting for Algorithms. For siPDCA-mssN and DsiPDCA-mssN, we set the same initial penalty parameter $c_0 = 6\times10^{-5}$. The other parameters for siPDCA-mssN are set as the same as in Subsection 6.1. For DsiPDCA-mssN, the proximal parameter, sieving parameter and the sequence $\{\epsilon_k\}$ are set the same as that of siPDCA-mssN. The other parameters for DsiPDCA-mssN are set as: $\bar{\epsilon} = 10^{-12}$, $M = 500$. The parameters for PPALM method are set as: $\rho_0 = 0.1$, $\sigma = \sqrt{10}$, $\varepsilon_I = 10^{-5}$ and $\gamma_1 = \gamma_2 = 1.01$.

Initialization. The initial solutions of siPDCA-mssN, DsiPDCA-mssN and PPALM method are randomly generated from a uniform distribution with range $[0, 0.5]$.

Termination criterion. The termination criterion of siPDCA-mssN and PPALM method is set the same as in Subsection 6.1. For DsiPDCA-mssN, the outer iteration termination condition is set as:

$$\|V(\boldsymbol{x}^{q+1}; \lambda^{q+1}, \boldsymbol{u}^{q+1}, \boldsymbol{\eta}^{q+1}, \boldsymbol{\zeta}^{q+1}, \boldsymbol{\gamma}^{q+1})\|_\infty < \widehat{\epsilon} \quad \text{and} \quad \|\boldsymbol{x}^{q+1}\|_0 \leq K$$

Table 6.1: The performance of siPDCA-mssN, PPALM method and the CPLEX solver for solving the CCMV problem on real-world market data.

n	K	siPDCA-mssN		PPALM		CPLEX	
		$time/s$	F	$time/s$	F	$time/s$	F
Nikkei 225 224	5	0.0349	1.6063e-04	0.0488	1.8473e-04	0.5725	1.6063e-04
	10	0.0730	1.6859e-04	0.0592	2.3819e-04	0.5462	1.5365e-04
	15	0.0287	1.5312e-04	0.0655	1.6671e-04	0.1729	1.5312e-04
	20	0.0280	1.5312e-04	0.0665	1.6062e-04	0.1813	1.5312e-04
	30	0.0290	1.5312e-04	0.0608	1.5950e-04	0.1754	1.5312e-04
	40	0.0274	1.5312e-04	0.0515	1.5971e-04	0.1602	1.5312e-04
S&P 500 457	5	0.1893	1.4332e-04	0.1989	1.8587e-04	1.1910	1.6716e-04
	10	0.1537	1.1309e-04	0.1544	1.4337e-04	1.1233	1.5065e-04
	15	0.0662	1.0520e-04	0.0922	1.4390e-04	1.1176	1.2812e-04
	20	0.0589	9.9833e-05	0.0923	1.4164e-04	1.1310	1.1335e-04
	30	0.0653	9.6891e-05	0.0786	1.0948e-04	2.0211	1.0323e-04
	40	0.0670	9.4487e-05	0.1014	1.0421e-04	2.1381	1.0323e-04
Russell 2000 1319	5	0.3497	7.0892e-05	1.3323	9.8059e-05	45.3173	7.2938e-05
	10	0.2058	4.1415e-05	0.7418	5.3995e-05	17.8707	3.6303e-05
	15	0.2371	3.4325e-05	0.3535	5.9514e-05	21.4397	2.9927e-05
	20	0.1889	2.9781e-05	0.3197	5.4707e-05	17.6860	2.6558e-05
	30	0.1759	2.5470e-05	0.2262	5.5459e-05	22.1888	2.7010e-05
	40	0.1353	2.4746e-05	0.2765	5.1579e-05	18.6160	2.7010e-05
Russell 3000 2152	5	0.6726	7.7102e-05	5.4425	1.2629e-04	76.8841	6.5720e-05
	10	0.4753	4.5418e-05	2.7726	8.7698e-05	25.4944	3.6787e-05
	15	0.3913	3.2468e-05	1.4684	7.2732e-05	20.6647	4.9539e-05
	20	0.2837	3.0435e-05	0.7577	6.5346e-05	20.4594	3.7563e-05
	30	0.2725	2.6626e-05	0.9214	5.9099e-05	20.3108	3.0495e-05
	40	0.2561	2.3898e-05	1.1170	5.8800e-05	20.2162	2.7310e-05

with $\widehat{\epsilon} = 10^{-7}$ and the inner iteration tolerance error is set as $\varepsilon = 10^{-5}$.

As a comparison, the numerical results for solving time ($time/s$) and optimal value (F) of objective function $F(x) = \frac{1}{2}\boldsymbol{x}^\top \mathbf{Q}\boldsymbol{x}$ of siPDCA-mssN, DsiPDCA-mssN and PPALM method for solving (1.1) are presented in Table 6.2. The solving time and optimal value of these three algorithms are obtained from the average of 20 run tests. As a result from Table 6.2, one can see that siPDCA-mssN and DsiPDCA-mssN outperform PPALM method from solving time and optimal value. With the dimension growing from n = 10000 to n = 80000, the computation time of siPDCA-mssN and PPALM method increases significantly, but that of DsiPDCA-mssN increases slowly. When n = 80 000, the time consumption of the PPALM method is longer than half an hour, but that of the DsiPDCA-mssN is less than 1 second.

6.3. Out-of-sample performance of the CCMV model

To illustrate the out-of-sample performance of the CCMV model, we compare the Sharp ratio of the solution generated by siPDCA-mssN for solving CCMV model in (1.1).

Comparing models. The shorting-prohibited Markowitz (Non-Shortsale mean-variance,

Table 6.2: The performance of siPDCA-mssN, DsiPDCA-mssN and PPALM method for solving the large-scale CCMV problem on simulated data.

n	K	siPDCA-mssN		DsiPDCA-mssN		PPALM	
		$time/s$	F	$time/s$	F	$time/s$	F
10000	10	2.1978	8.2918e-05	0.4110	3.8213e-05	73.6205	2.6319e-4
	20	2.7305	3.8103e-05	0.3089	2.6377e-05	44.7699	7.5173e-05
	30	2.1143	2.8244e-05	0.3028	2.2996e-05	33.9316	6.2616e-05
	40	3.0821	2.2531e-05	0.2903	2.1722e-05	36.3180	6.1816e-05
	50	5.0259	1.9253e-05	0.2758	2.0789e-05	38.0101	5.7869e-05
	60	9.3288	1.6421e-05	0.2509	2.0706e-05	39.3867	5.5792e-05
20000	10	9.1212	8.1761e-05	0.4838	4.1265e-05	350.0777	5.1797e-04
	20	6.0668	3.608e-05	0.3696	2.4762e-05	201.6345	1.5315e-04
	30	5.8608	2.8873e-05	0.3532	2.0181e-05	161.6800	1.2390e-04
	40	8.8727	2.1757e-05	0.3213	1.8388e-05	176.8079	1.0653e-04
	50	17.7093	1.7854e-05	0.2983	1.8606e-05	173.9438	9.9564e-05
	60	37.1613	1.5978e-05	0.2768	1.9695e-05	169.6886	9.2651e-05
40000	10	30.8615	7.7861e-05	0.5367	3.8369e-05	1062.8312	3.6329e-04
	20	17.9685	3.5676e-05	0.4016	2.3554e-05	590.5480	2.4785e-04
	30	13.1156	2.7229e-05	0.3999	2.0377e-05	532.5221	2.3005e-04
	40	49.7563	2.0232e-05	0.3865	1.7835e-05	682.3270	3.2856e-04
	50	77.4470	1.6606e-05	0.3540	1.7278e-05	726.4977	3.1768e-04
	60	171.0652	1.4098e-05	0.3148	1.7286e-05	719.4106	3.1519e-04
80000	10	179.5655	7.3305e-05	0.6617	3.9505e-05	2003.9333	2.2406e-4
	20	314.7783	3.7089e-05	0.5122	2.4819e-05	2294.6601	5.2259e-4
	30	349.6836	2.8916e-05	0.4872	2.1345e-05	2529.7857	7.8469e-4
	40	453.4198	2.1484e-05	0.4455	1.9437e-05	2616.7341	9.5967e-4
	50	439.1101	1.8558e-05	0.4213	1.7986e-05	2879.4293	1.0081e-3
	60	530.3158	1.4834e-05	0.4127	1.5731e-05	3083.2883	1.0973e-3

NSMV) model

$$\min_{\boldsymbol{x}\in\mathbb{R}^n}\left\{\frac{1}{2}\boldsymbol{x}^\top\mathbf{Q}\boldsymbol{x} : \boldsymbol{R}^\top\boldsymbol{x}\geq r, \boldsymbol{e}^\top\boldsymbol{x}=1, \boldsymbol{x}\geq\mathbf{0}\right\} \tag{6.3}$$

is a classical model in portfolio selection. In addition, as evaluated by DeMiguel et al. [53], the performance of naive $1/N$ (equal proportion of every asset) portfolio is often better than that of the standard Markowitz model. The l_1 regularized mean-variance (l_1-MV) model

$$\min_{\boldsymbol{x}\in\mathbb{R}^n}\left\{\frac{1}{2}\boldsymbol{x}^\top\mathbf{Q}\boldsymbol{x}+\lambda\|\boldsymbol{x}\|_1 : \boldsymbol{R}^\top\boldsymbol{x}\geq r, \boldsymbol{e}^\top\boldsymbol{x}=1\right\} \tag{6.4}$$

and l_0 regularized mean-variance (l_0-MV) model

$$\min_{\boldsymbol{x}\in\mathbb{R}^n}\left\{\frac{1}{2}\boldsymbol{x}^\top\mathbf{Q}\boldsymbol{x}+\lambda\|\boldsymbol{x}\|_0 : \boldsymbol{R}^\top\boldsymbol{x}\geq r, \boldsymbol{e}^\top\boldsymbol{x}=1, \mathbf{0}\leq\boldsymbol{x}\leq\boldsymbol{b}\right\} \tag{6.5}$$

are two popular sparse portfolio selection models, see [25]. The NSMV model, l_1-MV and l_0-MV are solved by the CPLEX solver.

Experimental data. This numerical experiment is also performed on the real-world market data described in Subsection 6.1.

Table 6.3: Comparison of the Sharpe ratio (Sr) and sparsity (Spa) of the solutions of problem (1.1) and that of other methods.

n	CCMV			$1/N$		$NSMV$		l_1-MV			l_0-MV		
	K	Sr	Spa	Sr	Spa	Sr	Spa	λ	Spa	Sr	λ	Spa	Sr
Nikkei 225 224	5	0.0315	5	-0.2362	224	0.0399	12	5.00e-4	12	0.404	5.00e-5	5	0.0315
	10	0.0392	10	-0.2362		0.0399		2.50e-4	12	0.0404	4.90e-7	8	0.0374
	15	0.0404	12	-0.2362		0.0399		1.50e-4	15	0.0990	4.00e-7	12	0.0404
	20	0.0404	12	-0.2362		0.0399		1.11e-4	20	0.1560	2.50e-7	12	0.0404
	30	0.0404	12	-0.2362		0.0399		1.96e-5	30	0.2186	1.50e-7	12	0.0404
	40	0.0404	12	-0.2362		0.0399		5.90e-5	40	0.2437	5.00e-8	12	0.0404
S&P 500 457	5	0.0841	5	-0.0507	457	-0.1094	52	5.00e-4	50	-0.1090	2.00e-5	5	0.0011
	10	0.0810	10	-0.0507		-0.1094		4.00e-4	50	-0.1090	2.65e-6	10	-0.0697
	15	0.0295	15	-0.0507		-0.1094		3.00e-4	50	-0.1090	1.25e-6	15	-0.0710
	20	-0.0484	20	-0.0507		-0.1094		2.00e-4	50	-0.1090	5.00e-7	20	-0.0889
	30	-0.0965	30	-0.0507		-0.1094		1.00e-4	61	-0.0998	1.53e-7	30	-0.1073
	40	-0.0981	40	-0.0507		-0.1094		5.00e-5	91	-0.0573	5.45e-8	40	-0.1146
Russell 2000 1319	5	0.1518	5	0.0568	1319	0.2547	83	5.00e-4	74	0.2552	5.75e-6	5	0.1317
	10	0.1732	10	0.0568		0.2547		4.00e-4	74	0.2552	2.23e-6	10	0.1448
	15	0.1968	15	0.0568		0.2547		3.00e-4	74	0.2552	1.00e-6	15	0.1922
	20	0.2239	20	0.0568		0.2547		2.00e-4	74	0.2552	5.00e-7	20	0.1991
	30	0.2654	30	0.0568		0.2547		1.00e-4	78	0.2645	2.50e-7	30	0.2462
	40	0.3187	40	0.0568		0.2547		5.00e-5	105	0.2691	1.53e-7	40	0.2556
Russell 3000 2152	5	0.1180	5	0.0355	2152	0.2432	92	5.00e-4	83	0.2422	5.00e-6	5	0.1889
	10	0.1784	10	0.0355		0.2432		4.00e-4	83	0.2422	2.50e-6	10	0.1965
	15	0.2151	15	0.0355		0.2432		3.00e-4	83	0.2422	1.20e-6	15	0.1650
	20	0.2483	20	0.0355		0.2432		2.00e-4	83	0.2424	6.25e-7	20	0.1787
	30	0.3113	30	0.0355		0.2432		1.00e-4	86	0.2552	2.75e-7	30	0.2421
	40	0.3021	40	0.0355		0.2432		5.00e-5	105	0.2662	1.50e-7	40	0.2758

Parameter setting. The parameters for siPDCA-mssN are set as the same as in Subsection 6.1. The parameters of CPLEX solver is set as the default of the software. For l_1-MV and l_0-MV, choosing a suitable regularization parameter λ is essential to sparsity of its solutions. We choose the values of λ for l_1-MV and l_0-MV through a 5-fold cross-validation procedure from the set of candidates

$$\{5\times10^{-7}, 10^{-6}, 5\times10^{-6}, 10^{-5}, 5\times10^{-5}, 10^{-4}, 5\times10^{-4}\}, \quad \{10^{-7}, 5\times10^{-7}, 10^{-6}, 5\times10^{-6}, 10^{-5}\},$$

respectively.

As a comparison, the Sharpe ratio (Sr) and sparsity (Spa) of the solutions of the CCMV problem in (1.1), the shorting-prohibited Markowitz model in (6.3), the naive $1/N$ portfolio model, the l_1-MV model in (6.4) and the l_0-MV model in (6.5) are presented in Table 6.3. As a result from Table 6.3, one can see that the Sharp ratio of the solution of siPDCA-mssN for solving (1.1) is larger than that of the naive $1/N$ portfolios model, NSMV model, l_1-MV model and l_0-MV model for most cases. In addition, the solutions of CCMV model and l_0-MV model are more sparse than those of the other portfolio selection models. The solution of the l_1-MV model cannot become more sparse even if the penalty parameter is increased. Therefore, when the transaction costs and other costs are taken into account, the solutions of CCMV model are

better than those of the other portfolio selection models.

7. Conclusions

In this paper, we considered the optimization problem of the cardinality constrained mean-variance (CCMV) model for sparse portfolio selection. To address the difficulties caused by the cardinality constraint, the cardinality constraint was equivalently transferred to a difference-of-convex functions (DC) constraint and a penalty approach was used to penalize to the DC constraint into objective function. We proved that there exists a penalty parameter such that the penalty approach is exact. To solving the gained DC problem, an efficient inexact proximal DC algorithm with sieving strategy (siPDCA) was proposed. By making full use of the semismooth properties of the dual inner problems of siPDCA and the superlinear convergence of semismooth Newton (ssN) method, an efficient majorized ssN (mssN) method was introduced to solve the inner problems of siPDCA from the dual. As a result, an algorithm framework, siPDCA based on a mssN method (siPDCA-mssN), was proposed to solve the DC problem. We proved that the sequence generated by siPDCA-mssN globally converges to a stationary point of the DC problem and the stationary point is also a local minimizer of the CCMV model when it satisfies the cardinality constraint.

For the large-scale CCMV problem, based on the violation of the first-order optimality conditions, a decomposed strategy was introduced into siPDCA-mssN, and the resulting algorithm was called decomposed siPDCA-mssN (DsiPDCA-mssN). In each iteration of DsiPDCA-mssN, siPDCA-mssN was used to solve a small-scale CCMV model. Under some mild assumption, we proved that the solution sequence generated by DsiPDCA-mssN subsequentially converges to a local minimizer of the original CCMV problem.

The results of the numerical experiments performed on the real-world market data demonstrated that siPDCA-mssN outperforms the standard CPLEX(12.9) solver and the penalty proximal alternating linearized minimization (PPALM) method for solving the CCMV model from the computation time and optimal value. In addition, by performing numerical experiments on large-scale simulated data, it was illustrated that DsiPDCA-mssN and siPDCA-mssN are more efficient than PPALM method for solving large-scale CCMV problem. Moreover, the out-of-sample performance experiments displayed that the solutions generated by siPDCA-mssN for solving CCMV model are better than those of native $1/N$ portfolio model, shorting-prohibited Markowitz model, l_1 regularization mean-variance model and l_0 regularization mean-variance model in terms of Sharp ratio and sparsity.

Appendix

A.1. The proof of Theorem 4.2

Proof. For statement (1), since $\boldsymbol{x}^{k_{l+1}} \in \mathcal{C}$ is the stability center generated in the serious step, then from the optimality of $\boldsymbol{x}^{k_{l+1}}$ for solving strongly convex problem in (3.5) and the feasibility of $\boldsymbol{x}^{k_l} \in \mathcal{C}$, it holds that

$$\begin{aligned} &f_c(\boldsymbol{x}^{k_{l+1}}) - \langle \boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l} \rangle + \frac{\alpha}{2}\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \boldsymbol{x}^{k_{l+1}} - 1) - \langle \boldsymbol{x}^{k_{l+1}}, \Delta_1^{k_{l+1}} \rangle \\ \leq & f_c(\boldsymbol{x}^{k_l}) - \langle \boldsymbol{x}^{k_l}, \boldsymbol{\xi}^{k_l} \rangle + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \boldsymbol{x}^{k_l} - 1) - \langle \boldsymbol{x}^{k_l}, \Delta_1^{k_{l+1}} \rangle - \frac{\alpha}{2}\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2. \end{aligned} \tag{A.1}$$

From the convexity of $g_c(\boldsymbol{x})$, we have

$$g_c(\boldsymbol{x}^{k_{l+1}}) \geq g_c(\boldsymbol{x}^{k_l}) + \langle \boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}, \boldsymbol{\xi}^{k_l} \rangle.$$

Combining this with (A.1), we have

$$\begin{aligned} &\alpha\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 - \langle \boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}, \Delta^{k_{l+1}} \rangle - (\boldsymbol{u}^{k_{l+1}} - \boldsymbol{u}^{k_l})(\boldsymbol{e}^\top \boldsymbol{x}^{k_l} - 1) \\ \leq &[f_c(\boldsymbol{x}^{k_l}) - g_c(\boldsymbol{x}^{k_l}) + \boldsymbol{u}^{k_l}(\boldsymbol{e}^\top \boldsymbol{x}^{k_l} - 1)] - [f_c(\boldsymbol{x}^{k_{l+1}}) - g_c(\boldsymbol{x}^{k_{l+1}}) + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \boldsymbol{x}^{k_{l+1}} - 1)]. \end{aligned} \tag{A.2}$$

Since $\boldsymbol{x}^{k_{l+1}}$ is generated in the serious step of siPDCA-mssN, then the sieving conditions in (5.1) holds, i.e.,

$$\|\Delta_1^{k+1}\| \leq (1-\kappa)\frac{\alpha}{2}\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|,$$

and

$$|(\boldsymbol{u}^{k_{l+1}} - \boldsymbol{u}^{k_l})(\boldsymbol{e}^\top \boldsymbol{x}^{k_l} - 1)| \leq (1-\kappa)\frac{\alpha}{2}\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2.$$

Consequently,

$$\begin{aligned} &\alpha\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 - \langle \boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}, \Delta_1^{k_{l+1}} \rangle - (\boldsymbol{u}^{k_{l+1}} - \boldsymbol{u}^{k_l})(\boldsymbol{e}^\top \boldsymbol{x}^{k_l} - 1) \\ \geq &\kappa\alpha\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2. \end{aligned} \tag{A.3}$$

By applying this to (A.2), we have

$$\begin{aligned} &\kappa\alpha\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 \\ \leq &[f_c(\boldsymbol{x}^{k_l}) - g_c(\boldsymbol{x}^{k_l}) + \boldsymbol{u}^{k_l}(\boldsymbol{e}^\top \boldsymbol{x}^{k_l} - 1)] \\ &- [f_c(\boldsymbol{x}^{k_{l+1}}) - g_c(\boldsymbol{x}^{k_{l+1}}) + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \boldsymbol{x}^{k_{l+1}} - 1)]. \end{aligned} \tag{A.4}$$

Thus the sequence $\left\{f(\boldsymbol{x}^{k_l}) - g_c(\boldsymbol{x}^{k_l}) + \boldsymbol{u}^{k_l}(\boldsymbol{e}^\top \boldsymbol{x}^{k_l} - 1)\right\}$ is nonincreasing. From Proposition 4.1, we get that $\left\{\boldsymbol{x}^k\right\}$ and $\left\{\widetilde{\boldsymbol{u}}^k\right\}$ are bounded, then $\left\{\boldsymbol{x}^{k_l}\right\}$ and $\left\{\boldsymbol{u}^{k_l}\right\}$ are also bounded. Since $f_c(\boldsymbol{x}) - g_c(\boldsymbol{x})$ is lower bounded, we can obtain that $\left\{f_c(\boldsymbol{x}^{k_l}) - g_c(\boldsymbol{x}^{k_l}) + \boldsymbol{u}^{k_l}(\boldsymbol{e}^\top \boldsymbol{x}^{k_l} - 1)\right\}$ is lower bounded. Consequently, the limit $s = \liminf_{l\to\infty}[f_c(\boldsymbol{x}^{k_{l+1}}) - g_c(\boldsymbol{x}^{k_{l+1}}) + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \boldsymbol{x}^{k_{l+1}} - 1)]$ exists. By summing both sides of (A.4) from $l = 0$ to ∞, we have

$$\sum_{l=0}^{\infty} \kappa\alpha\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 \leq [f_c(\boldsymbol{x}^0) - g_c(\boldsymbol{x}^0) + \boldsymbol{u}^0(\boldsymbol{e}^\top \boldsymbol{x}^0 - 1)] - s < \infty \tag{A.5}$$

and $\lim_{l\to\infty} \|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\| = 0$.

For statement (2), since $\boldsymbol{x}^{k_{l+1}}$ is the stability center generated in the serious step of siPDCA-mssN, then the following optimality conditions for solving (3.5) hold:

$$\begin{aligned} &\Delta_1^{k_{l+1}} \in \nabla f_c(\boldsymbol{x}^{k_{l+1}}) + \partial\delta_{\mathcal{C}}(\boldsymbol{x}^{k_{l+1}}) - \boldsymbol{\xi}^{k_l} + \boldsymbol{u}^{k_{l+1}}\boldsymbol{e} + \alpha(\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}), \\ &\boldsymbol{e}^\top \boldsymbol{x}^{k_{l+1}} - 1 = \Delta_2^{k_{l+1}}. \end{aligned} \tag{A.6}$$

Then there exists $\boldsymbol{\varsigma}^{k_{l+1}} \in \partial\delta_{\mathcal{C}}(\boldsymbol{x}^{k_{l+1}})$ such that

$$\nabla f_c(\boldsymbol{x}^{k_{l+1}}) + \boldsymbol{\varsigma}^{k_{l+1}} - \boldsymbol{\xi}^{k_l} + \boldsymbol{u}^{k_{l+1}}\boldsymbol{e} + \alpha(\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}) - \Delta_1^{k_{l+1}} = 0.$$

Consequently, we have

$$\|\nabla f_c(\boldsymbol{x}^{k_{l+1}}) + \boldsymbol{\varsigma}^{k_{l+1}} - \boldsymbol{\xi}^{k_l} + \boldsymbol{u}^{k_{l+1}}\boldsymbol{e}\| \leq \alpha\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\| + \epsilon_{k_{l+1}}. \tag{A.7}$$

From Proposition 4.1, we have that $\{\boldsymbol{x}^{k_l}\}$ and $\{\boldsymbol{u}^{k_l}\}$ are bounded. Then there exists subsets $\mathcal{L}' \subset \mathcal{L} = \{0, 1, 2 \ldots\}$ such that $\{\boldsymbol{x}^{k_l}\}_{\mathcal{L}'}$ converges to an accumulation point $\overline{\boldsymbol{x}} \in \{\boldsymbol{x}^{k_l}\}_{\mathcal{L}}$. Combining the boundedness of $\{\boldsymbol{x}^{k_l}\}_{\mathcal{L}'}$ with the fact that $g_c(\boldsymbol{x})$ is a finite-valued convex function, it follows that the subsequence $\{\boldsymbol{\xi}^{k_l}\}_{\mathcal{L}'}$ is bounded. Consequently, it holds that $\{\boldsymbol{\varsigma}^{k_l}\}_{\mathcal{L}'}$ is bounded from (A.7). By the fact that the nonnegative sequence $\{\epsilon_k\}$ monotonically decreases to zero, we may assume that, without loss of generality, there exists a subset $\mathcal{L}'' \subset \mathcal{L}'$ such that $\lim_{l\in\mathcal{L}''} \boldsymbol{\xi}^{k_l} = \bar{\boldsymbol{\xi}}$, $\lim_{l\in\mathcal{L}''} \boldsymbol{\varsigma}^{k_l} = \bar{\varsigma}$, $\lim_{l\in\mathcal{L}''} \boldsymbol{u}^{k_l} = \bar{\boldsymbol{u}}$ and $\lim_{l\in\mathcal{L}''} \epsilon_{k_l} = 0$. Taking the limit on the two sides of inequality in (A.7) with $l \in \mathcal{L}''$, we have

$$\begin{aligned}\|\nabla f_c(\overline{\boldsymbol{x}}) + \bar{\boldsymbol{\varsigma}} - \bar{\boldsymbol{\xi}} + \bar{\boldsymbol{u}}\boldsymbol{e}\| &= \lim_{l\in\mathcal{L}''} \|\nabla f_c(\boldsymbol{x}^{k_{l+1}}) + \boldsymbol{\varsigma}^{k_{l+1}} - \boldsymbol{\xi}^{k_l} + \boldsymbol{u}^{k_{l+1}}\boldsymbol{e}\| \\ &\leq \lim_{l\in\mathcal{L}''} \alpha\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\| + \lim_{l\in\mathcal{L}''} \epsilon_{k_{l+1}} = 0,\end{aligned}$$

which implies that $\|\nabla f_c(\overline{\boldsymbol{x}}) + \bar{\boldsymbol{\varsigma}} - \bar{\boldsymbol{\xi}} + \bar{\boldsymbol{u}}\boldsymbol{e}\| = 0$. In addition,

$$|\boldsymbol{e}^\top \overline{\boldsymbol{x}} - 1| = \lim_{l\in\mathcal{L}''} |\boldsymbol{e}^\top \boldsymbol{x}^{k_{l+1}} - 1| \leq \lim_{l\in\mathcal{L}''} \epsilon_{k_{l+1}} = 0,$$

which implies that $\overline{\boldsymbol{x}} \in \mathcal{D}$ and $\bar{\boldsymbol{u}}\boldsymbol{e} \in \mathcal{N}_{\mathcal{D}}(\overline{\boldsymbol{x}})$. Since $\delta_{\mathcal{C}}(\boldsymbol{x})$ and $g_c(\boldsymbol{x})$ are all lower semicontinuous, then the accumulation points $\bar{\zeta}$ and $\bar{\boldsymbol{\xi}}$ satisfy $\bar{\zeta} \in \partial\delta_{\mathcal{C}}(\overline{\boldsymbol{x}})$ and $\bar{\boldsymbol{\xi}} \in \partial g_c(\overline{\boldsymbol{x}})$, respectively, as a consequence of [46, Proposition 4.1.1]. Therefore, it holds that

$$\mathbf{0} \in \partial f(\overline{\boldsymbol{x}}) - \partial g_c(\overline{\boldsymbol{x}}) + \mathcal{N}_{\mathcal{D}}(\overline{\boldsymbol{x}}). \tag{A.8}$$

This implies that any accumulation point $\overline{\boldsymbol{x}}$ of $\{\boldsymbol{x}^{k_l}\}$ is a stationary point of (3.3). This completes the proof. □

A.2. The proof of Proposition 4.2

Proof. For statement (1), since $\boldsymbol{x}^{k_{l+1}} = \boldsymbol{y}^{k+1} \in \mathcal{C}$ is the stability center generated in the serious step, then the sieving conditions in (5.1) hold. Consequently, it holds that

$$\|\Delta_1^{k_{l+1}}\| \leq (1-\kappa)\frac{\alpha}{2}\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\| \tag{A.9}$$

and

$$\frac{\kappa\alpha}{2}\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 \leq \frac{\alpha}{2}\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 - \langle\Delta_1^{k_{l+1}}, \boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\rangle. \tag{A.10}$$

Thus, we have

$$\begin{aligned}&E(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}}) \\ &= f_c(\boldsymbol{x}^{k_{l+1}}) - \langle\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}\rangle + g_c^*(\boldsymbol{\xi}^{k_l}) + \alpha\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 \\ &\quad + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top\boldsymbol{x}^{k_{l+1}} - 1) - \langle\Delta_1^{k_{l+1}}, \boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\rangle \\ &= f_c(\boldsymbol{x}^{k_{l+1}}) - \langle\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}, \boldsymbol{\xi}^{k_l}\rangle - g_c(\boldsymbol{x}^{k_l}) + \alpha\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 \\ &\quad + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top\boldsymbol{x}^{k_{l+1}} - 1) - \langle\Delta_1^{k_{l+1}}, \boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\rangle \\ &\geq f_c(\boldsymbol{x}^{k_{l+1}}) - g_c(\boldsymbol{x}^{k_{l+1}}) + \alpha\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 \\ &\quad + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top\boldsymbol{x}^{k_{l+1}} - 1) - \langle\Delta_1^{k_{l+1}}, \boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\rangle \\ &\geq f_c(\boldsymbol{x}^{k_{l+1}}) - g_c(\boldsymbol{x}^{k_{l+1}}) + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top\boldsymbol{x}^{k_{l+1}} - 1) + (1+\kappa)\frac{\alpha}{2}\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 \\ &\geq f_c(\boldsymbol{x}^{k_{l+1}}) - g_c(\boldsymbol{x}^{k_{l+1}}) + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top\boldsymbol{x}^{k_{l+1}} - 1),\end{aligned}$$

where the second equality is due to the convexity of g_c and the fact that $\boldsymbol{\xi}^{k_l} \in \partial g_c(\boldsymbol{x}^{k_l})$, the first inequality follows from the convexity of g_c.

For statement (2), since $\boldsymbol{x}^{k_{l+1}} = \boldsymbol{y}^{k+1} \in \mathcal{C}$ is the optimal solution of strongly convex inner problem (3.5), then the following inequality follows from the feasibility of $\boldsymbol{x}^{k_l} \in \mathcal{C}$:

$$\begin{aligned}
&f_c(\boldsymbol{x}^{k_{l+1}}) - \langle \boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l} \rangle + \delta_{\mathcal{C}}(\boldsymbol{x}^{k_{l+1}}) + \frac{\alpha}{2}\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \boldsymbol{x}^{k_{l+1}} - 1) - \langle \Delta_1^{k_{l+1}}, \boldsymbol{x}^{k_{l+1}} \rangle \\
\leq & f_c(\boldsymbol{x}^{k_l}) - \langle \boldsymbol{x}^{k_l}, \boldsymbol{\xi}^{k_l} \rangle + \delta_{\mathcal{C}}(\boldsymbol{x}^{k_l}) + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \boldsymbol{x}^{k_l} - 1) - \langle \Delta_1^{k_{l+1}}, \boldsymbol{x}^{k_l} \rangle - \frac{\alpha}{2}\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2. \qquad \text{(A.11)}
\end{aligned}$$

This impiles that

$$\begin{aligned}
&E(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}}) \\
=& f_c(\boldsymbol{x}^{k_{l+1}}) - \langle \boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l} \rangle + g_c^*(\boldsymbol{\xi}^{k_l}) + \delta_{\mathcal{C}}(\boldsymbol{x}^{k_{l+1}}) + \alpha\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 \\
&+ \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \boldsymbol{x}^{k_{l+1}} - 1) - \langle \Delta_1^{k_{l+1}}, \boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l} \rangle \\
\leq & f_c(\boldsymbol{x}^{k_l}) - \langle \boldsymbol{x}^{k_l}, \boldsymbol{\xi}^{k_l} \rangle + \delta_{\mathcal{C}}(\boldsymbol{x}^{k_l}) + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \boldsymbol{x}^{k_l} - 1) + g_c^*(\boldsymbol{\xi}^{k_l}).
\end{aligned}$$

From (A.3), we have

$$\begin{aligned}
&E(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}}) \\
\leq & f_c(\boldsymbol{x}^{k_l}) - \langle \boldsymbol{x}^{k_l}, \boldsymbol{\xi}^{k_l} \rangle + g_c^*(\boldsymbol{\xi}^{k_l}) + \delta_{\mathcal{C}}(\boldsymbol{x}^{k_l}) + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \boldsymbol{x}^{k_l} - 1) \\
= & f_c(\boldsymbol{x}^{k_l}) - g_c(\boldsymbol{x}^{k_l}) + \delta_{\mathcal{C}}(\boldsymbol{x}^{k_l}) + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \boldsymbol{x}^{k_l} - 1) \\
\leq & f_c(\boldsymbol{x}^{k_l}) - \langle \boldsymbol{x}^{k_l}, \boldsymbol{\xi}^{k_{l-1}} \rangle + g_c^*(\xi^{k_{l-1}}) + \delta_{\mathcal{C}}(\boldsymbol{x}^{k_l}) + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \boldsymbol{x}^{k_l} - 1) \\
= & E(\boldsymbol{x}^{k_l}, \boldsymbol{\xi}^{k_{l-1}}, \boldsymbol{x}^{k_{l-1}}, \boldsymbol{u}^{k_l}, \Delta_1^{k_l}) - \alpha\|\boldsymbol{x}^{k_l} - \boldsymbol{x}^{k_{l-1}}\|^2 \\
&+ \langle \Delta_1^{k_l}, \boldsymbol{x}^{k_l} - \boldsymbol{x}^{k_{l-1}} \rangle + (\boldsymbol{u}^{k_{l+1}} - \boldsymbol{u}^{k_l})(\boldsymbol{e}^\top \boldsymbol{x}^{k_l} - 1) \\
\leq & E(\boldsymbol{x}^{k_l}, \boldsymbol{\xi}^{k_{l-1}}, \boldsymbol{x}^{k_{l-1}}, \boldsymbol{u}^{k_l}, \Delta_1^{k_l}) - \kappa\alpha\|\boldsymbol{x}^{k_l} - \boldsymbol{x}^{k_{l-1}}\|^2,
\end{aligned}$$

where the first equality follows from the convexity of g_c and the fact that $\boldsymbol{\xi}^{k_l} \in \partial g_c(\boldsymbol{x}^{k_l})$, the second inequality follows from the convexity of g_c and the Youngs inequality applied to g_c.

For statement (3), from Proposition 4.1, we can obtain that $\{\boldsymbol{x}^{k_l}\}$ and $\{\boldsymbol{u}^{k_l}\}$ are bounded. Since $g_c(\boldsymbol{x})$ is a finite-valued convex function, then it holds that $\{\boldsymbol{\xi}^{k_l}\}$ is bounded. The boundedness of $\left\{\Delta_1^{k_l}\right\}$ follows from the fact that $\|\Delta_1^{k_l}\| \leq \epsilon_{k_l}$ and $\lim_{l\to\infty} \epsilon_{k_l} = 0$. Hence, $\left\{(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}})\right\}$ has nonempty accumulation point set $\boldsymbol{\Gamma}$.

For statement (4), the lower boundedness of $\{f_c(\boldsymbol{x}^{k_l}) - g_c(\boldsymbol{x}^{k_l}) + \boldsymbol{u}^{k_l}(\boldsymbol{e}^\top \boldsymbol{x}^{k_l} - 1)\}$, together with (4.8), yields that the sequence $\left\{E(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}})\right\}$ is bounded below . The inequality (4.9) implies that $\left\{E(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}})\right\}$ is nonincreasing and the limit

$$\Upsilon = \lim_{l\to\infty} E(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}})$$

exists. Next, we will prove that $E \equiv \Upsilon$ on $\boldsymbol{\Gamma}$. Take any $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{\xi}}, \widehat{\boldsymbol{x}}, \widehat{\boldsymbol{u}}, \widehat{\Delta}_1) \in \boldsymbol{\Gamma}$. Since the above limit exists, then there exists a subset $\mathcal{L}' \subset \mathcal{L} = \{1, 2, \cdots, \infty\}$ such that

$$\lim_{l\in\mathcal{L}'}(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}}) = (\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{\xi}}, \widehat{\boldsymbol{x}}, \widehat{\boldsymbol{u}}, \widehat{\Delta}_1).$$

Evidently, $\lim_{l\in\mathcal{L}'} \epsilon_{k_l} = 0$, which implies that $\boldsymbol{e}^\top \widehat{\boldsymbol{x}} - 1 = 0$. From the optimality of $\boldsymbol{x}^{k_{l+1}}$ and

$\boldsymbol{u}^{k_{l+1}}$ for solving (3.5), it follows that

$$\begin{aligned}
&f_c(\boldsymbol{x}^{k_{l+1}}) - \langle \boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}\rangle + \delta_{\mathcal{C}}(\boldsymbol{x}^{k_{l+1}}) + \frac{\alpha}{2}\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \boldsymbol{x}^{k_{l+1}} - 1) - \langle \boldsymbol{x}^{k_{l+1}}, \Delta_1^{k_{l+1}}\rangle \\
\leq & f_c(\widehat{\boldsymbol{x}}) - \langle \widehat{\boldsymbol{x}}, \boldsymbol{\xi}^{k_l}\rangle + \delta_{\mathcal{C}}(\widehat{\boldsymbol{x}}) + \frac{\alpha}{2}\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^{k_l}\|^2 + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \widehat{\boldsymbol{x}} - 1) - \langle \widehat{\boldsymbol{x}}, \Delta_1^{k_{l+1}}\rangle. \qquad \text{(A.12)}
\end{aligned}$$

Rearranging the terms in the above inequality, we can obtain that

$$\begin{aligned}
&f_c(\boldsymbol{x}^{k_{l+1}}) - \langle \boldsymbol{x}^{k_{l+1}} - \widehat{\boldsymbol{x}}, \boldsymbol{\xi}^{k_l}\rangle + \delta_{\mathcal{C}}(\boldsymbol{x}^{k_{l+1}}) + \frac{\alpha}{2}\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 \\
&\quad - \langle \Delta_1^{k_{l+1}}, \boldsymbol{x}^{k_{l+1}} - \widehat{\boldsymbol{x}}\rangle + \boldsymbol{u}^{k_{l+1}}\boldsymbol{e}^\top(\boldsymbol{x}^{k_{l+1}} - \widehat{\boldsymbol{x}}) \\
\leq & f_c(\widehat{\boldsymbol{x}}) + \delta_{\mathcal{C}}(\widehat{\boldsymbol{x}}) + \frac{\alpha}{2}\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^{k_l}\|^2. \qquad \text{(A.13)}
\end{aligned}$$

From the boundedness of $\{\boldsymbol{x}^{k_l}\}$, $\{\boldsymbol{\xi}^{k_l}\}$ and $\{\boldsymbol{\Delta}^{k_l}\}$ in statement (3), it holds that

$$\begin{aligned}
&\lim_{l\in\mathcal{L}'} \langle \boldsymbol{x}^{k_{l+1}} - \widehat{\boldsymbol{x}}, \boldsymbol{\xi}^{k_l}\rangle = 0, \\
&\lim_{l\in\mathcal{L}'} \langle \boldsymbol{x}^{k_{l+1}} - \widehat{\boldsymbol{x}}, \Delta_1^{k_{l+1}}\rangle = 0.
\end{aligned}$$

This, together with (A.13), yields that

$$\begin{aligned}
\Upsilon &= \lim_{l\in\mathcal{L}'} E(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}}) \\
&= \lim_{l\in\mathcal{L}'} f_c(\boldsymbol{x}^{k_{l+1}}) - \langle \boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}\rangle + g_c^*(\boldsymbol{\xi}^{k_l}) + \delta_{\mathcal{C}}(\boldsymbol{x}^{k_{l+1}}) + \alpha\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 \\
&\quad + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \boldsymbol{x}^{k_{l+1}} - 1) - \langle \Delta_1^{k_{l+1}}, \boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\rangle \\
&= \lim_{l\in\mathcal{L}'} f_c(\boldsymbol{x}^{k_{l+1}}) - \langle \boldsymbol{x}^{k_{l+1}} - \widehat{\boldsymbol{x}}, \boldsymbol{\xi}^{k_l}\rangle + \delta_{\mathcal{C}}(\boldsymbol{x}^{k_{l+1}}) + \alpha\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 \\
&\quad + \boldsymbol{u}^{k_{l+1}}\boldsymbol{e}^\top(\boldsymbol{x}^{k_{l+1}} - \widehat{\boldsymbol{x}}) - \langle \Delta_1^{k_{l+1}}, \boldsymbol{x}^{k_{l+1}} - \widehat{\boldsymbol{x}}\rangle - \langle \widehat{\boldsymbol{x}}, \boldsymbol{\xi}^{k_l}\rangle \\
&\quad + g_c^*(\boldsymbol{\xi}^{k_l}) - \langle \Delta_1^{k_{l+1}}, \widehat{\boldsymbol{x}} - \boldsymbol{x}^{k_l}\rangle + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \widehat{\boldsymbol{x}} - 1) \\
&\leq \limsup_{l\in\mathcal{L}'} f_c(\widehat{\boldsymbol{x}}) + \delta_{\mathcal{C}}(\widehat{\boldsymbol{x}}) + \frac{\alpha}{2}\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^{k_l}\|^2 + \frac{\alpha}{2}\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|^2 - \langle \widehat{\boldsymbol{x}}, \boldsymbol{\xi}^{k_l}\rangle \\
&\quad + g_c^*(\boldsymbol{\xi}^{k_l}) - \langle \Delta_1^{k_{l+1}}, \widehat{\boldsymbol{x}} - \boldsymbol{x}^{k_l}\rangle + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \widehat{\boldsymbol{x}} - 1) \\
&= \limsup_{l\in\mathcal{L}'} f_c(\widehat{\boldsymbol{x}}) + \delta_{\mathcal{C}}(\widehat{\boldsymbol{x}}) + \alpha\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^{k_l}\|^2 - \langle \widehat{\boldsymbol{x}} - \boldsymbol{x}^{k_l}, \boldsymbol{\xi}^{k_l}\rangle \\
&\quad - g_c(\boldsymbol{x}^{k_l}) - \langle \Delta_1^{k_{l+1}}, \widehat{\boldsymbol{x}} - \boldsymbol{x}^{k_l}\rangle + \boldsymbol{u}^{k_{l+1}}(\boldsymbol{e}^\top \widehat{\boldsymbol{x}} - 1) \\
&= f_c(\widehat{\boldsymbol{x}}) - g_c(\widehat{\boldsymbol{x}}) + \widehat{\boldsymbol{u}}(\boldsymbol{e}^\top \widehat{\boldsymbol{x}} - 1) \leq E(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{\xi}}, \widehat{\boldsymbol{x}}, \widehat{\boldsymbol{u}}, \widehat{\Delta}_1),
\end{aligned}$$

where the fourth equality follows from the convexity of g_c and the fact that $\boldsymbol{\xi}^{k_l} \in \partial g_c(\boldsymbol{x}^{k_l})$, the last equality is due to $\widehat{\boldsymbol{x}} \in \mathcal{C}$, the last inequality comes from (4.8) with l trending to infinity. Since E is lower semicontinuous, it follows that

$$E(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{\xi}}, \widehat{\boldsymbol{x}}, \widehat{\boldsymbol{u}}, \widehat{\Delta}_1) = \liminf_{l\in\mathcal{L}'} E(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}}) = \Upsilon$$

and $E \equiv \Upsilon$ on $\boldsymbol{\Gamma}$.

For statement (5), the subdifferential of E at $(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}})$ can be computed

by

$$\partial E(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}})$$
$$= \begin{bmatrix} \nabla f_c(\boldsymbol{x}^{k_{l+1}}) - \boldsymbol{\xi}^{k_l} + \partial\delta_{\mathcal{C}}(\boldsymbol{x}^{k_{l+1}}) + 2\alpha(\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}) + \boldsymbol{u}^{k_{l+1}}\boldsymbol{e} - \Delta_1^{k_{l+1}} \\ -\boldsymbol{x}^{k_{l+1}} + \partial g^*(\boldsymbol{\xi}^{k_l}) \\ -\alpha(\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}) + \Delta_1^{k_{l+1}} \\ \boldsymbol{x}^{k_l} - \boldsymbol{x}^{k_{l+1}} \\ \boldsymbol{e}^\top \boldsymbol{x}^{k_{l+1}} - 1 \end{bmatrix}.$$

Since $\boldsymbol{x}^{k_{l+1}} = \boldsymbol{y}^{k+1}$ is the optimal solution of (3.5), we have

$$\Delta_1^{k_{l+1}} \in \nabla f_c(\boldsymbol{x}^{k_{l+1}}) - \boldsymbol{\xi}^{k_l} + \partial\delta_{\mathcal{C}}(\boldsymbol{x}^{k_{l+1}}) + \alpha(\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}) + \boldsymbol{u}^{k_{l+1}}\boldsymbol{e}.$$

This, together with $\boldsymbol{x}^{k_l} \in \partial g_c^*(\boldsymbol{\xi}^{k_l})$, yields that

$$\begin{bmatrix} \alpha(\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}) \\ \boldsymbol{x}^{k_l} - \boldsymbol{x}^{k_{l+1}} \\ -\alpha(\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}) + \Delta_1^{k_{l+1}} \\ \boldsymbol{x}^{k_l} - \boldsymbol{x}^{k_{l+1}} \\ \boldsymbol{e}^\top \boldsymbol{x}^{k_{l+1}} - 1 \end{bmatrix} \in \partial E(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}}). \tag{A.14}$$

Since $\boldsymbol{x}^{k_{l+1}}$ satisfies the sieving conditions in (5.1), then we have

$$\|\Delta_1^{k_{l+1}}\| < (1-\kappa)\frac{\alpha}{2}\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\| \quad \text{and} \quad |\boldsymbol{e}^\top \boldsymbol{x}^{k_{l+1}} - 1| < \|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|.$$

Consequently, it holds that there exists a constant ρ such that

$$\operatorname{dist}(\boldsymbol{0}, \partial E(\boldsymbol{x}^{k_{l+1}}, \boldsymbol{\xi}^{k_l}, \boldsymbol{x}^{k_l}, \boldsymbol{u}^{k_{l+1}}, \Delta_1^{k_{l+1}})) \le \rho\|\boldsymbol{x}^{k_{l+1}} - \boldsymbol{x}^{k_l}\|. \tag{A.15}$$

This completes the proof. □

Acknowledgments. The authors would like to thank the Associate Editor and anonymous referees for their helpful suggestions. The work of Bo Yu and Mingcai Ding was supported by the National Natural Science Foundation of China (Grand No. 11971092). Xiaoliang Song's work was supported by the Fundamental Research Funds for the Central Universities (Grand No. DUT20RC(3)079).

References

[1] H. Markowitz, Portfolio Selection, *Journal of Finance*, **7**:1 (1952), 77–91.

[2] A. Bensoussan, S.R. Celine and Z.F. Yan, A mean-variance Approach to Capital Investment Optimization, *SIAM Journal on Financial Mathematics*, **10**:1 (2019), 156–180.

[3] A. Ismail and H. Pham, Robust Markowitz mean-variance portfolio selection under ambiguous covariance matrix, *Mathematical Finance*, **29**:1 (2016), 174–207.

[4] E.F. Fama, Efficient Capital Markets: A Review of Theory and Empirical Work, *Journal of Finance*, **25**:2 (1970), 383–417.

[5] M.E. Rubinstein, A mean-variance Synthesis of Corporate Financial Theory, *Journal of Finance*, **28**:1 (1973), 167–81.

[6] P. Boyle, L. Garlappi, R. Uppal and T. Wang, Keynes Meets Markowitz: The Tradeoff Between Familiarity and Diversification, *SSRN Electronic Journal*, **58**:2 (2009), 253–272.

[7] C.H. Chen, X.D. Li, C. Tolman, S.Y. Wang and Y.Y. Ye, Sparse portfolio selection via quasi-norm regularization, *ArXiv Preprint arXiv:1312.6350*, (2013).

[8] B. Bruder, N. Gaussel, J.C. Richard and T. Roncalli, Regularization of Portfolio Allocation, *Available at SSRN: https://ssrn.com/abstract=2767358*, (2013).

[9] Z. Dai and F. Wen, Some improved sparse and stable portfolio optimization problems, *Finance Research Letters*, **27** (2018), 46–52.

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

[11] F. Cesarone, A. Scozzari and F. Tardella, Efficient algorithms for mean-variance portfolio optimization with hard real-world constraints, *The 18th AFIR Colloquium: Financial Risk in a Changing World*, Italy, 2009, 37–56.

[12] D. Maringer and H. Kellerer, Optimization of cardinality constrained portfolios with a hybrid local search algorithm, *Or Spectrum*, **25**:4 (2003), 481–495.

[13] D. Bienstock, Computational Study of a Family of Mixed-Integer Quadratic Programming Problems, *Mathematical PXrogramming*, **74** (1995), 121–140.

[14] B. Blog, G. Hoek, A. Kan and G. Timmer, The Optimal Selection of Small Portfolios, *Management Science*, **29**:7 (1983), 792–798.

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

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

[17] D. Shaw, S.C. Liu and L. Kopman, Lagrangian relaxation procedure for cardinality-constrained portfolio optimization, *Optimization Methods & Software*, **23**:3 (2008), 411–420.

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

[19] 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 (2013), 1409–1423.

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

[21] J. Gao and D. Li, Optimal Cardinality Constrained Portfolio Selection, *Operations RXesearch*, **61**:3 (2013), 745–761.

[22] Z.S. Lu and Y. Zhang, Sparse Approximation via Penalty decomposed Methods, *SIAM Journal on Optimization*, **23**:4 (2012), 2248–2478.

[23] Z.S. Lu and Z. Yong, Penalty decomposed Methods for $L0$-Norm Minimization, *ArXiv PXreprint arXiv:1205.2334*, (2012).

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

[25] Y. Teng, L. Yang, B. Yu and X.L. Song, A penalty PALM method for sparse portfolio selection problems, *Optimization Methods and Software*, **32**:1 (2017), 126–147.

[26] J. Brodie, I. Daubechies, C.D. Mol, D. Giannone and I. Loris, Sparse and stable Markowitz portfolios, *Proceedings of the National Academy of SXciences*, US, 2009, 12267–12272.

[27] P.S. Bradley and O.L. Mangasarian, Feature selection via concave minimization and support vector machines, *Machine Learning Proceedings of the Fifteenth International Conferences*, US, 1998, 82–90.

[28] D. Peleg and R. Meir, A bilinear formulation for vector sparsity optimization. *Signal Processing*, **8**:2 (2008), 375–389.

[29] H.A. Le Thi, T. Pham Dinh, M. Le Hoai and X.T. Vo. DC approximation approaches for sparse optimization, *European Journal of Operational Research*, **244**:1 (2015), 26–46.

[30] X.J. Zheng, X.L. 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 (2014), 379–397.

[31] J. Fan and R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, **96** (2001), 1348–1360.

[32] E.J. Candès, M.B. Wakin and S.P. Boyd, Enhancing sparsity by reweighted l_1 minimization, *Journal of Fourier Analysis and Applications*, **14** (2008), 877–905.

[33] J. Gotoh, A. Takeda and K. Tono, DC formulations and algorithms for sparse optimization problems, *Mathematical Programming*, **169** (2018), 141–176.

[34] H.A. Le Thi, T. Pham Dinh and H.V.J. Ngai, Exact penalty and error bounds in DC programming. *Journal of Global Optimization*, **52**:3 (2012), 509–535.

[35] T. Pham Dinh and H.A. Le Thi, Recent advances in DC programming and DCA. *Transactions on Computational Intelligence XIII*, Springer, Berlin Heidelberg 2014, 1–37.

[36] J.C.O. Souza, P. R. Oliveira and A. Soubeyran, Global convergence of a proximal linearized algorithm for difference of convex functions, *Optimization Letters*, **10**:7 (2016), 1529–1539.

[37] G.A. Watson, Linear best approximation using a class of polyhedral norms, *Numerical Algorithms*, **2**:3 (1992), 321–335.

[38] B. Wu, C. Ding, D.F. Sun and K.C. Toh, On the Moreau–Yosida Regularization of the Vector k-Norm Related Functions, *SIAM Journal on Optimization*, **24**:2 (2014), 766–794.

[39] K.F. Jiang, D.F. Sun and K.C. Toh, An Inexact Accelerated Proximal Gradient Method for large-scale Linearly Constrained Convex SDP, *SIAM Journal on Optimization*, **22**:3 (2012), 1042–1064.

[40] D.F. Sun, K.C. Toh and L.Q. Yang, An Efficient Inexact ABCD Method for Least Squares Semidefinite Programming, *SIAM Journal on Optimization*, **26**:2 (2016), 1072–1100.

[41] Y.Y. Wang, R.S. Liu, L. Ma and X.L. Song, Task Embedded Coordinate Update: A Realizable Framework for Multivariate Non-Convex Optimization, *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 277–286.

[42] A. Beck, First-order methods in optimization, *SIAM*, New York, 2017.

[43] J.J. Hiriart-Urruty, J.J. Strodiot and V.H. Nguyen, Generalized Hessian matrix and second-order optimality conditions for problems with C1,1 data, *Applied Mathematics and Optimization*, **11**:1 (1984), 169–180.

[44] X.D. Li, D.F. Sun and K.C. Toh, On the efficient computation of a generalized Jacobian of the projector over the Birkhoff polytope, *Mathematical Programming*, **179**:1 (2020), 419–446.

[45] T. Pham Dinh and An, H.A. Le Thi, Convex analysis approach to DC programming: theory, algorithms and applications, *Acta Mathematica Vietnamica*, **22**:1 (1997), 289–355.

[46] J.B. Hiriart, Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods, *Springer-Verlag*, 1993.

[47] H. Attouch and J. Bolte, On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, *Mathematical Programming*, **116**:1 (2009), 5–16.

[48] J. Bolte and E. Pauwels, Majorization-minimization procedures and convergence of SQP methods for semi-algebraic and tame programs, *Mathematics of Operations Research*, **41**:2 (2016), 442–465.

[49] T.X. Liu, T.K. Pong and A. Takeda, A refined convergence analysis of pDCAe with applications to simultaneous sparse recovery and outlier detection, *Computational Optimization and Applications*, **73**:1 (2019), 69–100.

[50] G.Q. Wang, X.Y. Wei, B. Yu and L.J. Xu, An Efficient Proximal Block Coordinate Homotopy Method for Large-Scale Sparse Least Squares Problems, *SIAM Journal on Scientific Computing*, **42**:1 (2020), A395–A423.

[51] N.A. Canakgoz and J.E. Beasley, Mixed-integer programming approaches for index tracking and enhanced indexation, *European Journal of Operational Research*, **196**:1 (2009), 384–399.

[52] J.E. Beasley, OR-Library: Distributing Test Problems by Electronic Mail, *Journal of the Operational Research Society*, **41**:11 (1990), 1069–1072.

[53] 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 (2009), 798–812.