SMAI-JCM
SMAI Journal of
Computational Mathematics

The augmented Lagrangian method with full Jacobian decomposition and logarithmic-quadratic proximal regularization for multiple-block separable convex programming

Min Li & Xiaoming Yuan

Volume 4 (2018), p. 81-120.
<http://smai-jcm.cedram.org/item?id=SMAI-JCM_2018__4__81_0>

SMAI Journal of Computational Mathematics
Vol. 4, 81-120 (2018)

The augmented Lagrangian method with full Jacobian decomposition and logarithmic-quadratic proximal regularization for multiple-block separable convex programming

Min Li [1]
Xiaoming Yuan [2]

[1] School of Management and Engineering, Nanjing University, China
E-mail address: liminnju@yahoo.com
[2] Department of Mathematics, The University of Hong Kong, Hong Kong, China
E-mail address: xmyuan@hku.hk.

Abstract. We consider a separable convex minimization model whose variables are coupled by linear constraints and they are subject to the positive orthant constraints, and its objective function is in form of m functions without coupled variables. It is well recognized that when the augmented Lagrangian method (ALM) is applied to solve some concrete applications, the resulting subproblem at each iteration should be decomposed to generate solvable subproblems. When the Gauss-Seidel decomposition is implemented, this idea has inspired the alternating direction method of multiplier (for $m = 2$) and its variants (for $m \geq 3$). When the Jacobian decomposition is considered, it has been shown that the ALM with Jacobian decomposition in its subproblem is not necessarily convergent even when $m = 2$ and it was suggested to regularize the decomposed subproblems with quadratic proximal terms to ensure the convergence. In this paper, we focus on the multiple-block case with $m \geq 3$. We consider implementing the full Jacobian decomposition to ALM's subproblems and using the logarithmic-quadratic proximal (LQP) terms to regularize the decomposed subproblems. The resulting subproblems are all unconstrained minimization problems because the positive orthant constraints are all inactive; and they are fully eligible for parallel computation. Accordingly, the ALM with full Jacobian decomposition and LQP regularization is proposed. We also consider its inexact version which allows the subproblems to be solved inexactly. For both the exact and inexact versions, we comprehensively discuss their convergence, including their global convergence, worst-case convergence rates measured by the iteration-complexity in both the ergodic and nonergodic senses, and linear convergence rates under additional assumptions. Some preliminary numerical results are reported to demonstrate the efficiency of the ALM with full Jacobian decomposition and LQP regularization.

2010 Mathematics Subject Classification. 90C25, 90C33, 65K05.

Keywords. convex programming, splitting methods, augmented Lagrangian method, logarithmic-quadratic proximal, parallel computation, convergence rate.

1. Introduction

We consider the following separable convex minimization problem whose variables are coupled by linear constraints and they are subject to the positive orthant constraints, and its objective function is the sum of more than one function without coupled variables:

$$\min\Big\{\sum_{i=1}^{m}\theta_i(x_i)\;\Big|\;\sum_{i=1}^{m}A_ix_i=b,\;\; x_i\in\Re^{n_i}_+,\;\; i=1,\ldots,m\Big\}, \tag{1.1}$$

where $\theta_i : \Re^{n_i} \to \Re$ $(i = 1,\ldots,m)$ are convex but not necessarily smooth functions; $A_i \in \Re^{l\times n_i}$ and $b \in \Re^l$. The solution set of (1.1) is assumed to be nonempty throughout our discussions.

The first author was supported by the National Natural Science Foundation of China under grant 11001053 and Program for New Century Excellent Talents in University under grant NCET-12-0111.
The second author was supported by the General Research Fund from Hong Kong Research Grants Council: 12313516.

Let $\lambda \in \Re^l$ be the Lagrange multiplier associated with the linear equality constraints in (1.1) and the Lagrangian function of (1.1) be

$$\mathcal{L}(x_1, \ldots, x_m, \lambda) := \sum_{i=1}^m \theta_i(x_i) - \lambda^T\Big(\sum_{i=1}^m A_i x_i - b\Big), \tag{1.2}$$

defined on $\Omega := \Re_+^{n_1} \times \cdots \times \Re_+^{n_m} \times \Re^l$. Then, the augmented Lagrangian function of (1.1) is

$$\mathcal{L}_\beta(x_1, \ldots, x_m, \lambda) := \sum_{i=1}^m \theta_i(x_i) - \lambda^T\Big(\sum_{i=1}^m A_i x_i - b\Big) + \frac{\beta}{2}\Big\|\sum_{i=1}^m A_i x_i - b\Big\|^2, \tag{1.3}$$

where $\beta > 0$ is a penalty parameter. If we treat the model (1.1) as a whole and apply directly the augmented Lagrangian method (ALM) in [21, 41], then the resulting scheme is

$$\begin{cases} (x_1^{k+1}, \ldots, x_m^{k+1}) := \operatorname{argmin}\{\mathcal{L}_\beta(x_1, \ldots, x_m, \lambda^k) \mid x_i \in \Re_+^{n_i},\ i = 1, \ldots, m\}, \\ \lambda^{k+1} := \lambda^k - \beta(\sum_{i=1}^m A_i x_i^{k+1} - b). \end{cases} \tag{1.4}$$

In general, the minimization subproblem in (1.4) is hard because it requires minimizing m functions with variables coupled by the quadratic term in (1.3). This difficulty has inspired a series of splitting methods whose common idea is decomposing the subproblem in (1.4) and thus generating easier subproblems. For example, for the special case of (1.1) with $m = 2$, if the minimization subproblem in (1.4) is decomposed in Gauss-Seidel order, the scheme is

$$\begin{cases} x_1^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1, x_2^k, \lambda^k) \mid x_1 \in \Re_+^{n_1}\}, \\ x_2^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1^{k+1}, x_2, \lambda^k) \mid x_2 \in \Re_+^{n_2}\}, \\ \lambda^{k+1} := \lambda^k - \beta(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b). \end{cases} \tag{1.5}$$

This is the so-called alternating direction method of multiplier (ADMM) in [17] and it has found many efficient applications in a broad spectrum of application domains such as image processing, statistical learning, computer vision, network optimization, and so on. We refer to [5, 13, 16] for some review papers on the ADMM. If we consider directly extending the scheme (1.5) to the generic case of (1.1) with $m \geq 3$, then the resulting direct extension of ADMM reads as

$$\begin{cases} x_1^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1, x_2^k, \ldots, x_m^k, \lambda^k) \mid x_1 \in \Re_+^{n_1}\}, \\ \cdots\cdots \\ x_i^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1^{k+1}, \ldots, x_{i-1}^{k+1}, x_i, x_{i+1}^k, \ldots, x_m^k, \lambda^k) \mid x_i \in \Re_+^{n_i}\}, \\ \cdots\cdots \\ x_m^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1^{k+1}, \ldots, x_{m-1}^{k+1}, x_m, \lambda^k) \mid x_m \in \Re_+^{n_m}\}, \\ \lambda^{k+1} := \lambda^k - \beta(\sum_{i=1}^m A_i x_i^{k+1} - b). \end{cases} \tag{1.6}$$

The direct extension of ADMM scheme (1.6) indeed works empirically for some applications, as shown in, e.g., [40, 43]. However, it was shown in [6] that the scheme (1.6) is not necessarily convergent. The convergence rate of ADMM and its extension are analysed in [9, 28, 30, 33, 32].

On the other hand, if we consider implementing the Jacobian decomposition to the ALM subproblem in (1.4), the resulting scheme reads as

$$\begin{cases} x_1^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1, x_2^k, \ldots, x_m^k, \lambda^k) \mid x_1 \in \Re_+^{n_1}\}, \\ \cdots\cdots \\ x_i^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1^k, \ldots, x_{i-1}^k, x_i, x_i^k, \ldots, x_m^k, \lambda^k) \mid x_i \in \Re_+^{n_i}\}, \\ \cdots\cdots \\ x_m^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1^k, \ldots, x_{m-1}^k, x_m, \lambda^k) \mid x_m \in \Re_+^{n_m}\}, \\ \lambda^{k+1} := \lambda^k - \beta(\sum_{i=1}^m A_i x_i^{k+1} - b). \end{cases} \tag{1.7}$$

The x_i-subproblems in (1.7), which usually dominate the computation time at each iteration, can be solved in parallel; and they can be implemented in a distributed-computing system. This feature is of particular interest for the big-data scenario and the circumstances where parallel computing infrastructures are available. Note that the subproblems in (1.7) are of the same level of difficulty as those in (1.6) — each of them requires minimizing one θ_i in the original objective of (1.1) plus a quadratic term with the positive orthant constraint $\Re_+^{n_i}$. The scheme (1.7), however, is not necessarily convergent even when $m = 2$, as shown in [22]. In the literature, it was suggested to correct the output of (1.7) by some correction steps to ensure the convergence; some prediction-correction methods based on the Jacobian decomposition of ALM (1.7) were thus presented in the literature, see, e.g., [20, 22]. Note that these prediction-correction methods usually converge fast for some applications arising in image processing and other areas. But their correction steps need the solutions of the x_i-subproblems in (1.7) as the input and thus they are of less degrees of parallel computation. In [25, 10], it was proved that the convergence is ensured if the subproblems in (1.7) are regularized by quadratic proximal terms with sufficiently large proximal coefficients. For example, it was analyzed in [25] that the following scheme

$$\begin{cases} x_1^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1, x_2^k, \ldots, x_m^k, \lambda^k) + \frac{s\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \Re_+^{n_1}\}, \\ \cdots\cdots \\ x_i^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1^k, \ldots, x_{i-1}^k, x_i, x_i^k, \ldots, x_m^k, \lambda^k) + \frac{s\beta}{2}\|A_i(x_i - x_i^k)\|^2 \mid x_i \in \Re_+^{n_i}\}, \\ \cdots\cdots \\ x_m^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1^k, \ldots, x_{m-1}^k, x_m, \lambda^k) + \frac{s\beta}{2}\|A_m(x_m - x_m^k)\|^2 \mid x_m \in \Re_+^{n_m}\}, \\ \lambda^{k+1} := \lambda^k - \beta(\sum_{i=1}^m A_i x_i^{k+1} - b). \end{cases} \tag{1.8}$$

is convergent as long as the proximal coefficient $s \geq m - 1$. A more general analysis can be found in [10]. The scheme (1.8) requires no correction step because of the regularization of the quadratic terms $\frac{s\beta}{2}\|A_i(x_i - x_i^k)\|^2$ for $i = 1, \ldots, m$; and the eligibility for parallel computation is remained. The iteration-complexity of a Jacobi-type non-Euclidean proximal ADMM for solving multi-block linearly constrained nonconvex programs is established in [35].

Note that the x_i-subproblems in (1.8) are constrained minimization problems subject to the positive orthants. To further simplify these subproblems, we can apply the logarithmic-quadratic proximal terms, which firstly appeared in [3], to the subproblems in (1.7). The key point is that the LQP regularization automatically excludes the points on the boundaries of the constraints in the feasible regions; thus the positive orthant constraints in (1.7) all become inactive and the decomposed subproblems in (1.7) with the LQP regularization are all unconstrained. In the literature, the research on the combination of the LQP regularization with ALM-based splitting methods focuses only on the special case of (1.1) with $m = 2$ and mainly on the Gauss-Seidel decomposition. For instance, the combination of the LQP regularization with the ADMM scheme (1.5) in [46, 1] is in the variational inequality context. We also refer to [29] and [31] for the combination of the LQP with the generalized ADMM proposed in [12] and the strictly contractive Peaceman-Rachford splitting method proposed in [23], respectively. Some other interesting applications of LQP can be found in, e.g., [2]. Finally, it is referred to [31, 44, 7] for the convergence rate analysis for the mentioned methods.

Our first purpose is proposing the scheme of ALM with Jacobian decomposition and LQP regularization for the multiple-block convex minimization model (1.1) with $m \geq 3$, see (3.2) for detail. Both the exact and inexact versions will be proposed. The inexact version allows the decomposed subproblems to be solved inexactly subject to certain inexactness criterion. To the best of our knowledge, it is the first work of combining the LQP regularization with the Jacobian decomposition of the ALM (1.7) for the generic case of (1.1) with $m \geq 3$. Note that using the LQP regularization, instead of the quadratic proximal terms, is particularly useful for the case where the functions θ_i's are generic and the subproblems in (1.7) are not simple enough to have closed-form solutions and thus the constrained subproblems in (1.7) need to be solved iteratively by a certain algorithm. In other words, the new scheme mainly differs from (1.7) in that only unconstrained subproblems are required to solve. The second purpose of this paper is comprehensively analyzing the convergence for both the exact and inexact versions of the new scheme. More specifically, we discuss their convergence, including the global convergence, the worst-case convergence rate measured by the iteration-complexity in both the ergodic and nonergodic senses, and the linear convergence rates under additional assumptions.

The rest of this paper is organized as follows. In Section 2, we summarize some useful results and introduce some notation for further analysis. Then, we present the exact version of the ALM with full Jacobian decomposition and LQP regularization in Section 3, followed by some remarks. The convergence of the exact version of this new scheme is proved in Section 4. Then, we establish its convergence rate in Section 5. In Section 6, we present the inexact version of the new scheme, and analyze its convergence in Section 7. In Section 8, we report some preliminary numerical results to show the efficiency of the new scheme. Finally, we make some conclusions in Section 9.

2. Preliminaries

We first summarize some useful preliminaries known in the literature and introduce some notations to be used in the analysis. Some simple conclusions are also proved in this section.

2.1. The Logarithmic-quadratic Proximal Regularization

We first review the LQP regularization. More details are provided in [3]. Let us define

$$\varphi(c) := \begin{cases} \frac{1}{2}(c-1)^2 + \mu(c - \log c - 1) & \text{if } c > 0, \\ +\infty & \text{otherwise,} \end{cases} \tag{2.1}$$

for a given scalar $\mu \in (0,1)$. Associated with φ, for any $z \in \Re^N_{++}$, we define

$$d(z', z) := \begin{cases} \sum_{j=1}^N \left[\frac{1}{2}(z'_j - z_j)^2 + \mu(z_j^2 \log \frac{z_j}{z'_j} + z'_j z_j - z_j^2)\right] & \text{if } z' \in \Re^N_{++}, \\ +\infty & \text{otherwise,} \end{cases} \tag{2.2}$$

and

$$\Phi'(z, z') := (z_1 \varphi'(z'_1/z_1), \cdots, z_N \varphi'(z'_N/z_N))^T \quad \forall z, z' \in \Re^N_{++}, \tag{2.3}$$

where

$$\varphi'(z'_j/z_j) = z'_j/z_j - 1 + \mu(1 - z_j/z'_j), \qquad j = 1, \cdots, N. \tag{2.4}$$

For any $z', z \in \Re^N_{++}$, we have $d(z', z) \geq \|z' - z\|^2/2$ and $d(z', z) = 0$ if and only if $z' = z$. Moreover, the function $d(\cdot,\cdot)$ defined in (2.2) can be rewritten as

$$d(z', z) = \sum_{j=1}^N z_j^2 \varphi(z'_j/z_j) \quad \forall z', z \in \Re^N_{++},$$

and then we have

$$\Phi'(z,z') = \nabla_{z'} d(z',z) = (z'-z) + \mu[z - Z^2(z')^{-1}],$$

where $Z := \mathrm{diag}(z_1, z_2, \ldots, z_N) \in \Re^{N\times N}$, $(z')^{-1} \in \Re^N$ is a vector whose j-th element is $1/z'_j$.

The following lemma was proved in [46] and it was inspired by Proposition 1 in [3]. We need this lemma to analyze the convergence for the new algorithms.

Lemma 2.1. *Let $P := \mathrm{diag}(p_1, \ldots, p_N) \in \Re^{N\times N}$ be a positive definite diagonal matrix, $q(z) \in \Re^N$ be a monotone mapping of z with respect to $\Re^N_+$, and $\vartheta : \Re^N \to \Re$. Let $\mu \in (0,1)$ be a constant. For given $\bar z, z \in \Re^N_{++}$, we define $\bar Z := \mathrm{diag}(\bar z_1, \ldots, \bar z_N)$, $z^{-1} := (1/z_1, \ldots, 1/z_N)^T$ and*

$$\Phi'(\bar z, z) = (z - \bar z) + \mu(\bar z - \bar Z^2 z^{-1}).$$

Then, the variational inequality

$$\vartheta(z') - \vartheta(z) + (z'-z)^T[q(z) + P\Phi'(\bar z, z)] \ge 0 \qquad \forall z' \in \Re^N_+, \tag{2.5}$$

has the unique positive solution z. In addition, for this positive solution $z \in \Re^N_{++}$ and any $z' \in \Re^N_+$, we have

$$\vartheta(z) - \vartheta(z') + (z - z')^T[q(z) + (1+\mu)P(z - \bar z)] \le \mu\|\bar z - z\|_P^2, \tag{2.6}$$

where $\|z\|_P^2 := z^T P z$.

2.2. Variational Reformulation of (1.1)

In our analysis, we need a variational reformulation of the convex minimization model (1.1). More specifically, let $(x_1^*, \ldots, x_m^*, \lambda^*)$ be a saddle point of the Lagrange function (1.2). Then, for any $(x_1, \ldots, x_m, \lambda) \in \Re^{n_1}_+ \times \cdots \times \Re^{n_m}_+ \times \Re^l$, we have the inequalities

$$\mathcal{L}(x_1^*, \ldots, x_m^*, \lambda) \ \le\ \mathcal{L}(x_1^*, \ldots, x_m^*, \lambda^*) \ \le\ \mathcal{L}(x_1, \ldots, x_m, \lambda^*). \tag{2.7}$$

Setting $(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \cdots, x_m, \lambda^*) = (x_1^*, \ldots, x_{i-1}^*, x_i, x_{i+1}^*, \ldots, x_m^*, \lambda^*)$ in the second inequality of (2.7) for $i = 1, \cdots, m$, we get

$$x_i^* \in \Re^{n_i}_+ \qquad \theta_i(x_i) - \theta_i(x_i^*) + (x_i - x_i^*)^T(-A_i^T\lambda^*) \ge 0 \quad \forall x_i \in \Re^{n_i}_+,\ i = 1, \ldots, m.$$

On the other hand, the first inequality in (2.7) means

$$\lambda^* \in \Re^l \qquad (\lambda - \lambda^*)^T\Big(\sum_{i=1}^m A_i x_i^* - b\Big) \ge 0 \qquad \forall \lambda \in \Re^l.$$

Recall that $\Omega = \Re^{n_1}_+ \times \cdots \times \Re^{n_m}_+ \times \Re^l$. Thus, finding a saddle point of $\mathcal{L}(x_1, \ldots, x_m, \lambda)$ is equivalent to finding a vector $w^* = (x_1^*, \ldots, x_m^*, \lambda^*) \in \Omega$ such that

$$\begin{cases} \theta_1(x_1) - \theta_1(x_1^*) + (x_1 - x_1^*)^T(-A_1^T\lambda^*) \ge 0 \qquad \forall x_1 \in \Re^{n_1}_+, \\ \quad \cdots \\ \theta_i(x_i) - \theta_i(x_i^*) + (x_i - x_i^*)^T(-A_i^T\lambda^*) \ge 0 \qquad \forall x_i \in \Re^{n_i}_+, \\ \quad \cdots \\ \theta_m(x_m) - \theta_m(x_m^*) + (x_m - x_m^*)^T(-A_m^T\lambda^*) \ge 0 \qquad \forall x_m \in \Re^{n_m}_+, \\ (\lambda - \lambda^*)^T\Big(\sum_{i=1}^m A_i x_i^* - b\Big) \ge 0 \qquad \forall \lambda \in \Re^l. \end{cases} \tag{2.8}$$

We can rewrite (2.8) in a compact way: solving (1.1) is equivalent to finding $w^* = (x_1^*, \ldots, x_m^*, \lambda^*) \in \Omega := \Re^{n_1}_+ \times \cdots \times \Re^{n_m}_+ \times \Re^l$ such that

$$\mathrm{VI}(\Omega, F, \theta): \qquad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0 \qquad \forall\, w \in \Omega, \tag{2.9a}$$

where

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}, \qquad \theta(x) = \sum_{i=1}^{m} \theta_i(x_i), \tag{2.9b}$$

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_m \\ \lambda \end{pmatrix} \quad \text{and} \quad F(w) = \begin{pmatrix} -A_1^T \lambda \\ \vdots \\ -A_m^T \lambda \\ \sum_{i=1}^m A_i x_i - b \end{pmatrix}. \tag{2.9c}$$

Because the mapping $F(w)$ defined in (2.9c) is affine with a skew-symmetric matrix, it is monotone. We denote by Ω^* the solution set of $\mathrm{VI}(\Omega, F, \theta)$, and it is not nonempty under the non-emptiness assumption of the solution set of (1.1).

Then, we recall the characterization of the solution set Ω^* whose proof can be found in [14, 26]:

$$\Omega^* := \bigcap_{w \in \Omega} \{\tilde{w} \in \Omega \mid \theta(x) - \theta(\tilde{x}) + (w - \tilde{w})^T F(w) \geq 0\}. \tag{2.10}$$

With the characterization (2.10) and following Definition 1 in [38], we define an ε-approximation solution of $\mathrm{VI}(\Omega, F, \theta)$ as follows.

Definition 2.2. The vector $\tilde{w} \in \Omega$ is called an ε-approximation solution of $\mathrm{VI}(\Omega, F, \theta)$ if it satisfies

$$\sup_{w \in \mathcal{B}_\Omega(\tilde{w})} \{\theta(\tilde{x}) - \theta(x) + (\tilde{w} - w)^T F(w)\} \leq \varepsilon, \tag{2.11}$$

where

$$\mathcal{B}_\Omega(\tilde{w}) := \{w \in \Omega \mid \|w - \tilde{w}\| \leq 1\}.$$

Based on this definition, for an algorithm, if after t iterations, we can find $\tilde{w} \in \Omega$ such that

$$\theta(\tilde{x}) - \theta(x) + (\tilde{w} - w)^T F(w) \leq \varepsilon \quad \forall w \in \mathcal{B}_\Omega(\tilde{w}),$$

with $\varepsilon = O(1/t)$, then we say this algorithm has a worst-case $O(1/t)$ convergence rate measured by the iteration complexity. See, e.g., [36, 37] for more details.

The following lemma is useful for establishing a worst-case $o(1/t)$ convergence rate in Sections 5.3 and 7.4. It is similar as Lemma 1.2 in [10].

Lemma 2.3. *If a sequence $\{a_t\} \subseteq \Re$ obeys: (1) $a_t \geq 0$; (2) $\sum_{t=0}^{\infty} a_t < +\infty$; (3) $a_t \leq a_{t-1} + \sigma_{t-1}$ for any integer $t \geq 1$, where the sequence $\{\sigma_t\}$ satisfies $\sum_{t=1}^{\infty} t\sigma_t < +\infty$ with $\sigma_t \geq 0$ for any integer $t \geq 0$, then we have $a_t = o(1/t)$.*

Proof. Since $a_t \leq a_{t-1} + \sigma_{t-1}$, we get

$$a_t \leq a_k + \sum_{j=k}^{t-1} \sigma_j \qquad \forall k \leq t-1.$$

By assumptions (1)-(3) in this lemma, we have

$$\begin{aligned} 0 \leq \frac{t}{2} \cdot a_t \quad &\leq \sum_{k=\lfloor \frac{t}{2} \rfloor + 1}^{t} a_k + \sum_{k=\lfloor \frac{t}{2} \rfloor + 1}^{t-1} \sum_{j=k}^{t-1} \sigma_j \\ &\leq \sum_{k=\lfloor \frac{t}{2} \rfloor + 1}^{t} a_k + \sum_{k=\lfloor \frac{t}{2} \rfloor + 1}^{t-1} k\sigma_k \to 0, \end{aligned}$$

as $t \to \infty$. Therefore, we get $a_t = o(1/t)$. The proof is complete. ■

2.3. Some Notations

With the given positive scalars β and $\mu \in (0,1)$, we define the scalars

$$r_i > \frac{(m-1)\beta}{1-\mu}\lambda_{\max}(A_i^T A_i) \qquad \forall\, i = 1, \ldots, m,$$

where A_i is the coefficient matrix given in the model (1.1). We also define the matrices G, H, M, N_x and N as following:

$$G := \begin{pmatrix} (1+\mu)r_1 I_{n_1} & \cdots & -\beta A_1^T A_m & 0 \\ \vdots & \ddots & \vdots & \vdots \\ -\beta A_m^T A_1 & \cdots & (1+\mu)r_m I_{n_m} & 0 \\ 0 & \cdots & 0 & \frac{1}{\beta} I_l \end{pmatrix}, \tag{2.12}$$

$$H := \begin{pmatrix} (1+\mu)r_1 I_{n_1} & \cdots & -\beta A_1^T A_m & 0 \\ \vdots & \ddots & \vdots & \vdots \\ -\beta A_m^T A_1 & \cdots & (1+\mu)r_m I_{n_m} & 0 \\ 0 & \cdots & 0 & \frac{1}{\gamma\beta} I_l \end{pmatrix}, \tag{2.13}$$

$$M := \operatorname{diag}(I_{n_1}, \ldots I_{n_m}, \gamma I_l), \quad N_x := \mu \cdot \operatorname{diag}(r_1 I_{n_1}, \ldots, r_m I_{n_m}) \tag{2.14}$$

and

$$N := \operatorname{diag}(N_x, 0), \tag{2.15}$$

where $\gamma \in (0,2)$.

Below we prove three assertions regarding the matrices just defined. These assertions make it possible to present our convergence analysis for the new algorithms compactly with alleviated notation.

Lemma 2.4. *Let $\beta > 0$; $\mu \in (0,1)$; $\gamma \in (0,2)$ and $r_i > (m-1)\beta\lambda_{\max}(A_i^T A_i)/(1-\mu)$, $i = 1, \ldots, m$. The matrices G, H, M and N defined respectively in (2.12)-(2.15) have the following relationships:*

$$HM = G, \qquad \tilde{H} := G^T + G - M^T H M - 2N \succ 0 \quad \text{and} \quad H \succ 0. \tag{2.16}$$

Proof. Using the definitions of the matrices H, M and G, by a simple manipulation, we obtain

$$\begin{aligned} HM &= \begin{pmatrix} (1+\mu)r_1 I_{n_1} & \cdots & -\beta A_1^T A_m & 0 \\ \vdots & \ddots & \vdots & \vdots \\ -\beta A_m^T A_1 & \cdots & (1+\mu)r_m I_{n_m} & 0 \\ 0 & \cdots & 0 & \frac{1}{\gamma\beta} I_l \end{pmatrix} \begin{pmatrix} I_{n_1} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I_{n_m} & 0 \\ 0 & \cdots & 0 & \gamma I_l \end{pmatrix} \\ &= \begin{pmatrix} (1+\mu)r_1 I_{n_1} & \cdots & -\beta A_1^T A_m & 0 \\ \vdots & \ddots & \vdots & \vdots \\ -\beta A_m^T A_1 & \cdots & (1+\mu)r_m I_{n_m} & 0 \\ 0 & \cdots & 0 & \frac{1}{\beta} I_l \end{pmatrix} = G. \end{aligned}$$

The first assertion $HM = G$ is proved.

Consequently, we get

$$\begin{array}{rcl} M^T H M & = & M^T G \\ & = & \begin{pmatrix} I_{n_1} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I_{n_m} & 0 \\ 0 & \cdots & 0 & \gamma I_l \end{pmatrix} \begin{pmatrix} (1+\mu)r_1 I_{n_1} & \cdots & -\beta A_1^T A_m & 0 \\ \vdots & \ddots & \vdots & \vdots \\ -\beta A_m^T A_1 & \cdots & (1+\mu)r_m I_{n_m} & 0 \\ 0 & \cdots & 0 & \frac{1}{\beta} I_l \end{pmatrix} \\ & = & \begin{pmatrix} (1+\mu)r_1 I_{n_1} & \cdots & -\beta A_1^T A_m & 0 \\ \vdots & \ddots & \vdots & \vdots \\ -\beta A_m^T A_1 & \cdots & (1+\mu)r_m I_{n_m} & 0 \\ 0 & \cdots & 0 & \frac{\gamma}{\beta} I_l \end{pmatrix}. \end{array}$$

Using (2.12)-(2.15) and the above equation, we have

$$\begin{array}{rcl} \tilde{H} & = & G^T + G - M^T H M - 2N \\ & = & \begin{pmatrix} (1-\mu)r_1 I_{n_1} & \cdots & -\beta A_1^T A_m & 0 \\ \vdots & \ddots & \vdots & \vdots \\ -\beta A_m^T A_1 & \cdots & (1-\mu)r_m I_{n_m} & 0 \\ 0 & \cdots & 0 & \frac{2-\gamma}{\beta} I_l \end{pmatrix} \\ & = & \begin{pmatrix} (m-1)\beta A_1^T A_1 & \cdots & -\beta A_1^T A_m & 0 \\ \vdots & \ddots & \vdots & \vdots \\ -\beta A_m^T A_1 & \cdots & (m-1)\beta A_m^T A_m & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix} \\ & & + \begin{pmatrix} (1-\mu)r_1 I_{n_1} - (m-1)\beta A_1^T A_1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & (1-\mu)r_m I_{n_m} - (m-1)\beta A_m^T A_m & 0 \\ 0 & \cdots & 0 & \frac{2-\gamma}{\beta} I_l \end{pmatrix} \\ & = & P^T \begin{pmatrix} (m-1)I_l & \cdots & -I_l & 0 \\ \vdots & \ddots & \vdots & \vdots \\ -I_l & \cdots & (m-1)I_l & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix} P \\ & & + \begin{pmatrix} (1-\mu)r_1 I_{n_1} - (m-1)\beta A_1^T A_1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & (1-\mu)r_m I_{n_m} - (m-1)\beta A_m^T A_m & 0 \\ 0 & \cdots & 0 & \frac{2-\gamma}{\beta} I_l \end{pmatrix}, \end{array} \quad (2.17)$$

with

$$P = \text{diag}(\sqrt{\beta}A_1, \ldots, \sqrt{\beta}A_m, I_l).$$

Therefore, the matrix $\tilde{H}$ is positive definite if $\beta > 0$; $\mu \in (0,1)$; $\gamma \in (0,2)$ and

$$r_i > \frac{(m-1)\beta}{1-\mu}\lambda_{\max}(A_i^T A_i) \qquad \forall\, i = 1, \ldots, m.$$

The second assertion is proved.

Similar as (2.17), the matrix H can be written as

$$H = P^T \begin{pmatrix} (m-1)I_l & \cdots & -I_l & 0 \\ \vdots & \ddots & \vdots & \vdots \\ -I_l & \cdots & (m-1)I_l & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix} P$$
$$+ \begin{pmatrix} (1+\mu)r_1 I_{n_1} - (m-1)\beta A_1^T A_1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & (1+\mu)r_m I_{n_m} - (m-1)\beta A_m^T A_m & 0 \\ 0 & \cdots & 0 & \frac{1}{\gamma\beta} I_l \end{pmatrix}.$$

Therefore the matrix H is also positive definite. The proof is complete. ■

3. The ALM with Full Jacobian Decomposition and LQP Regularization — Exact Version

Now, we present the exact version of the ALM with full Jacobian decomposition and LQP regularization for solving the model (1.1). Some remarks will also be proved. Based on our previous introduction and motivation, the ALM with full Jacobian decomposition and LQP regularization can be summarized as follows:

$$\begin{cases} x_1^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1, x_2^k, \ldots, x_m^k, \lambda^k) + r_1 d(x_1, x_1^k) \mid x_1 \in \Re_+^{n_1}\}, \\ \quad \cdots \\ x_i^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1^k, \ldots, x_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_m^k, \lambda^k) + r_i d(x_i, x_i^k) \mid x_i \in \Re_+^{n_i}\}, \\ \quad \cdots \\ x_m^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1^k, \ldots, x_{m-1}^k, x_m, \lambda^k) + r_m d(x_m, x_m^k) \mid x_m \in \Re_+^{n_m}\}, \\ \lambda^{k+1} := \lambda^k - \gamma\beta(\sum_{j=1}^m A_j x_j^{k+1} - b), \end{cases} \tag{3.1}$$

where the LQP regularizer $d(\cdot,\cdot)$ is defined in (2.2), $\beta > 0$, $\mu \in (0,1)$, $\gamma \in (0,2)$ and

$$r_i > \frac{(m-1)\beta}{1-\mu} \lambda_{\max}(A_i^T A_i) \qquad \forall\, i = 1, \ldots, m.$$

Recall the analysis in [3]. The LQP regularization terms $r_i d(x_i, x_i^k)$ force the solution of the x_i-subproblem in (3.1) to stay strictly in the interior of $\Re_+^{n_i}$. Hence, the constraints $\Re_+^{n_i}$ are not active and the iterative scheme (3.1) can be further specified as

$$\begin{cases} x_1^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1, x_2^k, \ldots, x_m^k, \lambda^k) + r_1 d(x_1, x_1^k)\}, \\ \quad \cdots \\ x_i^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1^k, \ldots, x_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_m^k, \lambda^k) + r_i d(x_i, x_i^k)\}, \\ \quad \cdots \\ x_m^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1^k, \ldots, x_{m-1}^k, x_m, \lambda^k) + r_m d(x_m, x_m^k)\}, \\ \lambda^{k+1} := \lambda^k - \gamma\beta(\sum_{j=1}^m A_j x_j^{k+1} - b), \end{cases} \tag{3.2}$$

in which only m unconstrained minimization subproblems are involved.

Algorithm 1.

Step 0. *Let* $\varepsilon > 0$, $\beta > 0$, $\mu \in (0,1)$, $\gamma \in (0,2)$ *and* $r_i > (m-1)\beta\lambda_{\max}(A_i^T A_i)/(1-\mu)$, $i = 1, \ldots, m$. *Choose* $(x_1^0, \ldots, x_m^0, \lambda^0) \in \Re_{++}^{n_1} \times \cdots \times \Re_{++}^{n_m} \times \Re^l$. *Set* $k := 0$.

Step 1. *Find* $x_i^{k+1} \in \Re^{n_i}_{++}$, $i = 1, \ldots, m$, *in parallel, such that*

$$x_i^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1^k, \ldots, x_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_m^k, \lambda^k) + r_i d(x_i, x_i^k)\}. \tag{3.3}$$

Step 2. *Update the Lagrange multiplier*

$$\lambda^{k+1} := \lambda^k - \gamma\beta\Big(\sum_{j=1}^m A_j x_j^{k+1} - b\Big). \tag{3.4}$$

Step 3. *Set* $w^{k+1} := (x_1^{k+1}, \ldots, x_m^{k+1}, \lambda^{k+1})$. *If* $\|w^{k+1} - w^k\| \le \varepsilon$, *stop; otherwise set* $k := k+1$ *and goto Step 1.*

Remark 3.1. Note that the unconstrained minimization problems in (3.2) are generally easier than the constrained ones appearing in (3.1). In particular, for the special case of (1.1) where $\theta_i(x_i)$, $i = 1, \ldots, m$ are differentiable, the scheme (3.2) reduces to

$$\begin{cases} \nabla\theta_1(x_1^{k+1}) - A_1^T[\lambda^k - \beta(\sum_{j=1}^m A_j x_j^k - b)] + \beta A_1^T A_1(x_1^{k+1} - x_1^k) + r_1 \Phi'(x_1^k, x_1^{k+1}) = 0, \\ \quad\cdots \\ \nabla\theta_i(x_i^{k+1}) - A_i^T[\lambda^k - \beta(\sum_{j=1}^m A_j x_j^k - b)] + \beta A_i^T A_i(x_i^{k+1} - x_i^k) + r_i \Phi'(x_i^k, x_i^{k+1}) = 0, \\ \quad\cdots \\ \nabla\theta_m(x_m^{k+1}) - A_m^T[\lambda^k - \beta(\sum_{j=1}^m A_j x_j^k - b)] + \beta A_m^T A_m(x_m^{k+1} - x_m^k) + r_m \Phi'(x_m^k, x_m^{k+1}) = 0, \\ \lambda^{k+1} := \lambda^k - \gamma\beta(\sum_{j=1}^m A_j x_j^{k+1} - b), \end{cases}$$

from which we can see that the main computation for generating a new iterate

$$w^{k+1} := (x_1^{k+1}, \ldots, x_m^{k+1}, \lambda^{k+1}) \in \Re^{n_1}_{++} \times \cdots \times \Re^{n_m}_{++} \times \Re^l$$

is solving m systems of equations.

Remark 3.2. In Algorithm 1, we use different proximal coefficients r_i for different subproblems. We can choose a unified coefficient r for all the x_i-subproblems and show the same theoretical assertions under the condition

$$r > \frac{(m-1)\beta}{1-\mu} \max_{i=1,\ldots,m} \Big\{\lambda_{\max}(A_i^T A_i)\Big\}.$$

Since the coefficient r is chosen to ensure the sufficient control of the proximity for all the x_i-subproblems, the condition is in the most conservative way. Choosing different values of r_i for different subproblems, however, makes it possible to choose r_i that is independent of all the other coefficient matrices A_j's with $j \neq i$. As is well known, a larger proximal coefficient means the proximal term plays a heavier weight in the objective and thus the solution of the proximally regularized subproblem is forced to be closer the previous iterate. Thus, a smaller value is preferred for the proximal coefficient provided that it is sufficient to ensure the convergence. Because of this reason, in (3.2), we consider choosing different values r_i's for the subproblems, instead of choosing a unified value despite that the corresponding theoretical analysis is simpler.

Remark 3.3. Note that Algorithm 1 involves a relaxation parameter γ in the Lagrange-multiplier updating step. This relaxation parameter is indeed very important to ALM-based splitting algorithms. In fact, for the original ALM (1.4), it has been demonstrated in [4, 12] that the convergence can still be ensured if we attach a relaxation factor $\gamma \in (0, 2)$ to the Lagrange-multiplier updating step. The key point is the fact elucidated in [42] that the ALM is indeed an application of the proximal point algorithm in [34]; and thus the relaxation idea in [19] is applicable. When the model (1.1) with two blocks of variables and functions is considered, it was proved in [15, 18] that the convergence can be ensured if a relaxation factor $\gamma \in (0, \frac{\sqrt{5}+1}{2})$ is attached to the Lagrange-multiplier updating step in the ADMM scheme (1.5). Indeed, as well demonstrated in the literature (e.g., [4]), ALM-based splitting schemes usually can be accelerated with a relaxation factor $\gamma > 1$ in their Lagrange-multiplier updating steps. However, as shown in [6], even for the direct extension of ADMM (1.6) without proximal regularization in its subproblems, it seems that there is no such a problem-data-independent range for γ in the Lagrange-multiplier updating step even for $m = 3$. Interestingly, for Algorithm 1 which is originated from splitting the ALM with full Jacobian decomposition, attaching any $\gamma \in (0, 2)$ can still ensure the convergence. Intuitively, it can be explained that because the decomposed subproblems are regularized by the LQP regularization terms, the proximity to the last iterate is well controlled. Accordingly, all the decomposed subproblems together constitute a good approximation to the minimization subproblem in the ALM (1.4) and hence the range $(0, 2)$ for the relaxation parameter in the ALM is preserved in Algorithm 1. We regard it as an important feature of Algorithm 1 because of the LQP regularization.

4. Convergence

In this section, we prove the global convergence for Algorithm 1. In order to further alleviate the notation in our analysis, we define an auxiliary sequence $\{\tilde{w}^k\}$ as

$$\tilde{w}^k := \begin{pmatrix} \tilde{x}_1^k \\ \vdots \\ \tilde{x}_m^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x_1^{k+1} \\ \vdots \\ x_m^{k+1} \\ \lambda^k - \beta(\sum_{j=1}^m A_j x_j^{k+1} - b) \end{pmatrix}, \tag{4.1}$$

where $(x_1^{k+1}, \ldots, x_m^{k+1})$ is generated by Algorithm 1. Then, based on (3.4) and (4.1), we immediately have

$$x_i^{k+1} = \tilde{x}_i^k, \;\; i = 1, \ldots, m \quad \text{and} \quad \lambda^{k+1} = \lambda^k - \gamma(\lambda^k - \tilde{\lambda}^k). \tag{4.2}$$

Moreover, we have the following relationship

$$\begin{pmatrix} x_1^{k+1} \\ \vdots \\ x_m^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} x_1^k \\ \vdots \\ x_m^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I_{n_1} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I_{n_m} & 0 \\ 0 & \cdots & 0 & \gamma I_l \end{pmatrix} \begin{pmatrix} x_1^k - \tilde{x}_1^k \\ \vdots \\ x_m^k - \tilde{x}_m^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix},$$

which can be rewritten into a compact form by using the notation of w^k and $\tilde{w}^k$:

$$w^{k+1} = w^k - M(w^k - \tilde{w}^k), \tag{4.3}$$

where M is defined in (2.14).

Now, we start to prove some properties for the sequence $\{w^k\}$ generated by Algorithm 1. Since we will analyze the convergence rate for Algorithm 1 based on the solution characterization (2.10), and the accuracy of an approximate solution $\tilde{w} \in \Omega$ is measured by an upper bound of the quantity of $\theta(\tilde{x}) - \theta(x) + (\tilde{w} - w)^T F(w)$ for all $w \in \Omega$ (see (2.11)), we are interested in estimating how accurate the point w^{k+1} generated by (3.1) is to a solution point of $\mathrm{VI}(\Omega, F, \theta)$. The main result is proved in Theorem 4.3. To prove this main result, we first show two lemmas. The first lemma presents an upper bound of $\theta(\tilde{x}^k) - \theta(x) + (\tilde{w}^k - w)^T F(\tilde{w}^k)$ for all $w \in \Omega$ in term of a quadratic term involving the matrices G and N defined in (2.12) and (2.15).

Lemma 4.1. *Let $\{w^k\}$ be generated by Algorithm 1 and $\{\tilde{w}^k\}$ be defined in (4.1). Then, for any $w \in \Omega$, we have*

$$\theta(\tilde{x}^k) - \theta(x) + (\tilde{w}^k - w)^T F(\tilde{w}^k) \le -(w - \tilde{w}^k)^T G(w^k - \tilde{w}^k) + \|w^k - \tilde{w}^k\|_N^2, \tag{4.4}$$

where the matrices G and N are defined in (2.12) *and* (2.15)*, respectively.*

Proof. We first observe the first-order optimality conditions of the minimization problems in (3.3). More specifically, the solution $x_i^{k+1} \in \Re_{++}^{n_i}$ of the x_i-subproblem in (3.3) can be expressed as

$$x_i^{k+1} := \operatorname{argmin}\Big\{\theta_i(x_i) - (\lambda^k)^T A_i x_i + \frac{\beta}{2}\Big\|A_i(x_i - x_i^k) + \Big(\sum_{j=1}^m A_j x_j^k - b\Big)\Big\|^2 + r_i d(x_i, x_i^k)\Big\},$$

and then the inequality

$$\theta_i(x_i) - \theta_i(x_i^{k+1}) + (x_i - x_i^{k+1})^T\Big\{ - A_i^T\Big[\lambda^k - \beta\Big(\sum_{j=1}^m A_j x_j^k - b\Big)\Big] \\ +\beta A_i^T A_i(x_i^{k+1} - x_i^k) + r_i\Phi'(x_i^k, x_i^{k+1})\Big\} \geq 0, \tag{4.5}$$

holds for any $x_i \in \Re_+^{n_i}$, $i = 1, \ldots, m$. Note that it follows from (4.1) that

$$\lambda^k = \tilde{\lambda}^k + \beta\Big(\sum_{j=1}^m A_j x_j^{k+1} - b\Big). \tag{4.6}$$

Substituting (4.6) into (4.5) and using $x_i^{k+1} = \tilde{x}_i^k$, $i = 1, \ldots, m$, we have

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\Big\{ - A_i^T\tilde{\lambda}^k + \beta A_i^T A_i(\tilde{x}_i^k - x_i^k) \\ -\beta A_i^T\Big[\sum_{j=1}^m A_j(\tilde{x}_j^k - x_j^k)\Big] + r_i\Phi'(x_i^k, \tilde{x}_i^k)\Big\} \geq 0 \qquad \forall x_i \in \Re_+^{n_i}. \tag{4.7}$$

Applying the assertion in Lemma 2.1 to (4.7) by setting $P = r_i I_{n_i}$, $\bar{z} = x_i^k$, $z = \tilde{x}_i^k$, $\vartheta(\cdot) = \theta_i(\cdot)$, $q(z) = -A_i^T\tilde{\lambda}^k + \beta A_i^T A_i(\tilde{x}_i^k - x_i^k) - \beta A_i^T[\sum_{j=1}^m A_j(\tilde{x}_j^k - x_j^k)]$ and $z' = x_i$ in (2.6), for any $x_i \in \Re_+^{n_i}$, $i = 1, \ldots, m$, we have

$$\theta_i(\tilde{x}_i^k) - \theta_i(x_i) + (\tilde{x}_i^k - x_i)^T\Big\{ - A_i^T\tilde{\lambda}^k + \beta A_i^T A_i(\tilde{x}_i^k - x_i^k) - \beta A_i^T \\ \times\Big[\sum_{j=1}^m A_j(\tilde{x}_j^k - x_j^k)\Big] + (1+\mu)r_i(\tilde{x}_i^k - x_i^k)\Big\} \leq \mu r_i\|x_i^k - \tilde{x}_i^k\|^2. \tag{4.8}$$

In addition, based on (4.1) we have

$$\Big(\sum_{j=1}^m A_j\tilde{x}_j^k - b\Big) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0,$$

or further as

$$\tilde{\lambda}^k \in \Re^l \qquad (\lambda - \tilde{\lambda}^k)^T\Big[\Big(\sum_{j=1}^m A_j\tilde{x}_j^k - b\Big) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\Big] \geq 0 \qquad \forall \lambda \in \Re^l. \tag{4.9}$$

Combining (4.8) and (4.9) together, and using the notation of θ and N_x, we get $\tilde{w}^k = (\tilde{x}_1^k, \ldots, \tilde{x}_m^k, \tilde{\lambda}^k) \in \Omega$; and for any $w = (x_1, \ldots, x_m, \lambda) \in \Omega$, it holds

$$\theta(\tilde{x}^k) - \theta(x) \\ + \begin{pmatrix} \tilde{x}_1^k - x_1 \\ \vdots \\ \tilde{x}_m^k - x_m \\ \tilde{\lambda}^k - \lambda \end{pmatrix}^T \left\{ \begin{pmatrix} -A_1^T\tilde{\lambda}^k \\ \vdots \\ -A_m^T\tilde{\lambda}^k \\ \sum_{j=1}^m A_j\tilde{x}_j^k - b \end{pmatrix} + \begin{pmatrix} -\beta A_1^T[\sum_{j=2}^m A_j(\tilde{x}_j^k - x_j^k)] + (1+\mu)r_1(\tilde{x}_1^k - x_1^k) \\ \vdots \\ -\beta A_m^T[\sum_{j=1}^{m-1} A_j(\tilde{x}_j^k - x_j^k)] + (1+\mu)r_m(\tilde{x}_m^k - x_m^k) \\ \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \end{pmatrix} \right\} \\ \leq \|x^k - \tilde{x}^k\|_{N_x}^2.$$

By using the notation of G and N in (2.12) and (2.15), and w and F in (2.9c), the compact form of the above inequality is exactly (4.4). ■

In the next lemma, we further analyze the right-hand side of the inequality (4.4) and reformulate it as the sum of some quadratic terms. This new form is more convenient for our further analysis, especially for the convergence rate analysis.

Lemma 4.2. *Let* $\{w^k\}$ *be generated by Algorithm* 1, $\{\tilde{w}^k\}$ *be defined in* (4.1), *and* G, H, N *and* $\tilde{H}$ *be defined in* (2.12)-(2.16). *Then for any* $w \in \Omega$, *we have*

$$(w-\tilde{w}^k)^T G(w^k-\tilde{w}^k) - \|w^k-\tilde{w}^k\|_N^2 = \frac{1}{2}(\|w-w^{k+1}\|_H^2 - \|w-w^k\|_H^2) + \frac{1}{2}\|w^k-\tilde{w}^k\|_{\tilde{H}}^2. \quad (4.10)$$

Proof. By using $G = HM$ and $M(w^k-\tilde{w}^k) = (w^k - w^{k+1})$ (see (4.3)), it follows that

$$\begin{aligned}(w-\tilde{w}^k)^T G(w^k-\tilde{w}^k) - \|w^k-\tilde{w}^k\|_N^2 &= (w-\tilde{w}^k)^T HM(w^k-\tilde{w}^k) - \|w^k-\tilde{w}^k\|_N^2 \\ &= (w-\tilde{w}^k)^T H(w^k-w^{k+1}) - \|w^k-\tilde{w}^k\|_N^2. \quad (4.11)\end{aligned}$$

For any vectors a, c, d, e in the same space and a matrix P with appropriate dimensionality, we have the identity

$$(a-c)^T P(d-e) = \frac{1}{2}(\|a-e\|_P^2 - \|a-d\|_P^2) + \frac{1}{2}(\|d-c\|_P^2 - \|e-c\|_P^2).$$

In this identity, we take

$$a = w, \quad c = \tilde{w}^k, \quad d = w^k, \quad e = w^{k+1} \quad \text{and} \quad P = H,$$

and submit it to the right-hand side of (4.11). The resulting equation is

$$\begin{aligned}&(w-\tilde{w}^k)^T G(w^k-\tilde{w}^k) - \|w^k-\tilde{w}^k\|_N^2 \\ &= \frac{1}{2}(\|w-w^{k+1}\|_H^2 - \|w-w^k\|_H^2) + \frac{1}{2}(\|w^k-\tilde{w}^k\|_H^2 - \|w^{k+1}-\tilde{w}^k\|_H^2 - 2\|w^k-\tilde{w}^k\|_N^2)(4.12)\end{aligned}$$

Now, we deal with the last term of the right-hand side of (4.12). By using (4.3) and (2.16), we get

$$\begin{aligned}&\|w^k-\tilde{w}^k\|_H^2 - \|w^{k+1}-\tilde{w}^k\|_H^2 - 2\|w^k-\tilde{w}^k\|_N^2 \\ &= \quad \|w^k-\tilde{w}^k\|_H^2 - \|(w^k-\tilde{w}^k) - (w^k-w^{k+1})\|_H^2 - 2\|w^k-\tilde{w}^k\|_N^2 \\ &\overset{(4.3)}{=} \quad \|w^k-\tilde{w}^k\|_H^2 - \|(w^k-\tilde{w}^k) - M(w^k-\tilde{w}^k)\|_H^2 - 2\|w^k-\tilde{w}^k\|_N^2 \\ &= \quad 2(w^k-\tilde{w}^k)^T HM(w^k-\tilde{w}^k) - (w^k-\tilde{w}^k)^T M^T HM(w^k-\tilde{w}^k) - 2\|w^k-\tilde{w}^k\|_N^2 \\ &\overset{(2.16)}{=} \quad (w^k-\tilde{w}^k)^T (G^T + G - M^T HM - 2N)(w^k-\tilde{w}^k) \\ &\overset{(2.16)}{=} \quad \|w^k-\tilde{w}^k\|_{\tilde{H}}^2.\end{aligned}$$

Substituting it in (4.12), we obtain the assertion (4.10). The proof is complete. ∎

Now we are ready to present an inequality where an upper bound of $\theta(\tilde{x}^k) - \theta(x) + (\tilde{w}^k - w)^T F(w)$ is found for all $w \in \Omega$. This inequality is also crucial for analyzing the contraction property and the convergence rate for the iterative sequence generated by Algorithm 1.

Theorem 4.3. *Let* $\{w^k\}$ *be generated by Algorithm* 1, $\{\tilde{w}^k\}$ *be defined in* (4.1), *and* H *and* $\tilde{H}$ *be defined in* (2.13) *and* (2.16), *respectively. Then for any* $w \in \Omega$, *we have*

$$\theta(\tilde{x}^k) - \theta(x) + (\tilde{w}^k - w)^T F(w) \le \frac{1}{2}(\|w-w^k\|_H^2 - \|w-w^{k+1}\|_H^2) - \frac{1}{2}\|w^k-\tilde{w}^k\|_{\tilde{H}}^2. \quad (4.13)$$

Proof. Note that F is monotone. We thus have

$$(\tilde{w}^k - w)^T F(w) \le (\tilde{w}^k - w)^T F(\tilde{w}^k).$$

It follows from the above inequality and (4.4) that

$$\theta(\tilde{x}^k) - \theta(x) + (\tilde{w}^k - w)^T F(w) \le -(w-\tilde{w}^k)^T G(w^k-\tilde{w}^k) + \|w^k-\tilde{w}^k\|_N^2 \quad (4.14)$$

for any $w \in \Omega$. The assertion (4.13) follows immediately from (4.10) and (4.14). ■

The assertion (4.13) also enables us to study the contraction property of the sequence $\{w^k\}$ generated by Algorithm 1.

Lemma 4.4. *Let $\{w^k\}$ be generated by Algorithm 1, $\{\tilde{w}^k\}$ be defined in (4.1), and H and $\tilde{H}$ be defined in (2.13) and (2.16), respectively. Then for any $w^* \in \Omega^*$, we have*

$$\|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 - \|w^k - \tilde{w}^k\|_{\tilde{H}}^2. \tag{4.15}$$

Proof. Setting $w = w^*$ in (4.13) where w^* being an arbitrary solution point in Ω^*, we get

$$\begin{aligned} &\|w^k - w^*\|_H^2 - \|w^{k+1} - w^*\|_H^2 \\ &\ge \|w^k - \tilde{w}^k\|_{\tilde{H}}^2 + 2[\theta(\tilde{x}^k) - \theta(x^*) + (\tilde{w}^k - w^*)^T F(w^*)] \\ &\ge \|w^k - \tilde{w}^k\|_{\tilde{H}}^2. \end{aligned}$$

The proof is complete. ■

Now, we are ready to prove the global convergence of Algorithm 1.

Theorem 4.5. *The sequence $\{w^k\}$ generated by Algorithm 1 converges to some w^∞ which is a solution of VI(Ω, F, θ).*

Proof. It follows from (4.15) that the sequence $\{w^k\}$ is bounded. Using (4.15), we have

$$\sum_{k=0}^{\infty} \|w^k - \tilde{w}^k\|_{\tilde{H}}^2 \le \|w^0 - w^*\|_H^2 < +\infty. \tag{4.16}$$

Therefore, we have

$$\lim_{k\to\infty} \|w^k - \tilde{w}^k\|_{\tilde{H}} = 0. \tag{4.17}$$

Thus the sequence $\{\tilde{w}^k\}$ is also bounded, and it has at least one cluster point. Let w^∞ be a cluster point of $\{\tilde{w}^k\}$ and the subsequence $\{\tilde{w}^{k_j}\}$ converges to w^∞. It follows from (4.4) and (4.17) that

$$\liminf_{j\to\infty} \Big\{\theta(x) - \theta(\tilde{x}^{k_j}) + (w - \tilde{w}^{k_j})^T F(\tilde{w}^{k_j})\Big\} \ge 0 \qquad \forall w \in \Omega,$$

and consequently

$$\theta(x) - \theta(x^\infty) + (w - w^\infty)^T F(w^\infty) \ge 0 \qquad \forall w \in \Omega.$$

This means that w^∞ is a solution of VI(Ω, F, θ). Note that the inequality (4.15) is true for all solution points of VI(Ω, F, θ), hence we have

$$\|w^k - w^\infty\|_H \le \|w^l - w^\infty\|_H \qquad \forall k \ge 0,\ \forall l \le k. \tag{4.18}$$

Since $\tilde{w}^{k_j} \to w^\infty$ $(j \to \infty)$, using (4.17) we have $w^{k_j} \to w^\infty$ $(j \to \infty)$. For any given $\varepsilon > 0$, there exists a $j_0 > 0$ such that

$$\|w^{k_{j_0}} - w^\infty\|_H \le \varepsilon. \tag{4.19}$$

Therefore, for any $k \ge k_{j_0}$, it follows from (4.18) and (4.19) that

$$\|w^k - w^\infty\|_H \le \|w^{k_{j_0}} - w^\infty\|_H \le \varepsilon.$$

This implies that the sequence $\{w^k\}$ converges to a point w^∞ in Ω^*. ■

5. Convergence Rate

In this section, we analyze the convergence rate for the scheme (3.2) from different perspectives. We divide the analysis into four subsections.

5.1. A Worst-case $O(1/t)$ Convergence Rate in the Ergodic Sense

First, we establish a worst-case $O(1/t)$ convergence rate measured by the iteration-complexity for Algorithm 1. Its proof is inspired by [26] for the original ADMM (1.5) and some contraction properties proved in the last subsection are useful.

Theorem 5.1. *Let $\{w^k\}$ be generated by Algorithm 1 and $\{\tilde{w}^k\}$ be defined by (4.1). Let $\tilde{w}_t$ be defined as*

$$\tilde{w}_t := \frac{1}{t+1}\sum_{k=0}^{t}\tilde{w}^k. \tag{5.1}$$

Then, for any integer $t > 0$, we have $\tilde{w}_t \in \Omega$ and

$$\theta(\tilde{x}_t) - \theta(x) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2(t+1)}\|w - w^0\|_H^2 \qquad \forall w \in \Omega, \tag{5.2}$$

where H is defined in (2.13).

Proof. First, because of (3.2) and (4.1), it holds that $\tilde{w}^k \in \Omega$ for all integer $k \geq 0$. Together with the convexity of Ω, (5.1) implies that $\tilde{w}_t \in \Omega$. Summing the inequality (4.13) over $k = 0, 1, \ldots, t$, we obtain

$$\begin{aligned}
&\sum_{k=0}^{t}\theta(\tilde{x}^k) - (t+1)\theta(x) + \Big[\sum_{k=0}^{t}\tilde{w}^k - (t+1)w\Big]^T F(w) \\
&\leq \frac{1}{2}(\|w - w^0\|_H^2 - \|w - w^{t+1}\|_H^2) - \frac{1}{2}\sum_{k=0}^{t}\|w^k - \tilde{w}^k\|_{\tilde{H}}^2 \\
&\leq \frac{1}{2}\|w - w^0\|_H^2 \quad \forall w \in \Omega.
\end{aligned}$$

Using the notation of $\tilde{w}_t$, it can be written as

$$\frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{x}^k) - \theta(x) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2(t+1)}\|w - w^0\|_H^2 \quad \forall w \in \Omega. \tag{5.3}$$

Since $\theta(x)$ is convex and

$$\tilde{x}_t = \frac{1}{t+1}\sum_{k=0}^{t}\tilde{x}^k,$$

we have that

$$\theta(\tilde{x}_t) \leq \frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{x}^k).$$

Substituting it in (5.3), the assertion of this theorem follows directly. ■

Note that it follows from the proof of Theorem 4.5 that the sequences $\{w^k\}$ and $\{\tilde{w}^k\}$ generated by Algorithm 1 are bounded. Therefore, there exists a constant $D > 0$ such that

$$\|w^k\|_H \leq D \qquad \text{and} \qquad \|\tilde{w}^k\|_H \leq D \qquad \forall k \geq 0.$$

Recall that $\tilde{w}_t$ is the average of $\{\tilde{w}^0, \tilde{w}^1, \ldots, \tilde{w}^t\}$. Thus, we have $\|\tilde{w}_t\|_H \leq D$. For any $w \in \mathcal{B}_\Omega(\tilde{w}_t) := \{w \in \Omega \mid \|w - \tilde{w}_t\|_H \leq 1\}$, we get

$$\begin{aligned}\theta(\tilde{x}_t) - \theta(x) + (\tilde{w}_t - w)^T F(w) &\leq \frac{1}{2(t+1)}\|w - w^0\|_H^2 \\ &\leq \frac{1}{2(t+1)}\big(\|w - \tilde{w}_t\|_H + \|\tilde{w}_t - w^0\|_H\big)^2 \\ &\leq \frac{1}{2(t+1)}\big(\|w - \tilde{w}_t\|_H + \|\tilde{w}_t\|_H + \|w^0\|_H\big)^2 \\ &\leq \frac{(1+2D)^2}{2(t+1)}.\end{aligned}$$

Thus, for any given $\varepsilon > 0$, after at most $t := \lceil \frac{(1+2D)^2}{2\varepsilon} - 1 \rceil$ iterations, we have

$$\theta(\tilde{x}_t) - \theta(x) + (\tilde{w}_t - w)^T F(w) \leq \varepsilon \quad \forall w \in \mathcal{B}_\Omega(\tilde{w}_t),$$

which means $\tilde{w}_t$ is an approximate solution of $\mathrm{VI}(\Omega, F, \theta)$ with an accuracy of $O(1/t)$. That is, a worst-case $O(1/t)$ convergence rate in the ergodic sense is established for Algorithm 1.

5.2. **A Worst-case $O(1/t)$ Convergence Rate in the Non-ergodic Sense**

In this subsection, we establish a worst-case $O(1/t)$ convergence rate in the non-ergodic sense for Algorithm 1. Its proof is mainly inspired by [27] for the original ADMM (1.5); Lemma 2.5 in [10] is also useful.

First, let us denote

$$\bar{H}_x := \mathrm{diag}\Big((1+\mu)r_1 I_{n_1} + \beta A_1^T A_1, \ldots, (1+\mu)r_m I_{n_m} + \beta A_m^T A_m\Big),$$

$$A := (A_1, \ldots, A_m), \quad \bar{H}'_x := \bar{H}_x - \beta A^T A + 2N_x, \quad \bar{H}' := \begin{pmatrix} \bar{H}'_x & 0 \\ 0 & \frac{1}{\gamma\beta} I_l \end{pmatrix}. \tag{5.4}$$

Then, we prove a lemma.

Lemma 5.2. *Let $\{w^k\}$ be generated by Algorithm 1. Assume that A is a matrix of full column rank,*

$$r_i > (m-1)\beta\lambda_{\max}(A_i^T A_i) \qquad \forall i = 1, \ldots, m \tag{5.5}$$

and

$$0 < \mu < \min\Big\{\min_{i=1,\ldots,m}\{1 - \frac{(m-1)\beta\lambda_{\max}(A_i^T A_i)}{r_i}\}, \frac{(2-\gamma)\beta\lambda_{\min}(A^T A)}{4\max_{i=1,\ldots,m}\{r_i\}}\Big\}. \tag{5.6}$$

Then, $\bar{H}' \succ 0$, where $\bar{H}'$ is defined by (5.4), and for any integer $k \geq 1$, we have

$$\|w^{k+1} - w^k\|_{\bar{H}'}^2 \leq \|w^k - w^{k-1}\|_{\bar{H}'}^2. \tag{5.7}$$

Proof. The conditions (5.5) and

$$0 < \mu < \min_{i=1,\ldots,m}\Big\{1 - \frac{(m-1)\beta\lambda_{\max}(A_i^T A_i)}{r_i}\Big\}$$

is equivalent to

$$r_i > \frac{(m-1)\beta}{1-\mu}\lambda_{\max}(A_i^T A_i) \qquad \forall\, i = 1, \ldots, m.$$

Note that $H \succ 0$. Using the definitions of H and $\bar{H}'$ in (2.13) and (5.4), we have

$$\bar{H}' = H + 2N \succ 0.$$

The sequence $\{w^k\}$ generated by Algorithm 1 under conditions (5.5) and (5.6) converges to a solution of VI(Ω, F, θ) by Theorem 4.5. It follows from (4.5) that

$$\theta_i(x_i) - \theta_i(x_i^{k+1}) + (x_i - x_i^{k+1})^T\Big\{ - A_i^T\lambda^k + \beta A_i^T\Big(\sum_{j=1}^m A_j x_j^k - b\Big)$$
$$+\beta A_i^T A_i(x_i^{k+1} - x_i^k) + r_i\Phi'(x_i^k, x_i^{k+1})\Big\} \ge 0 \tag{5.8}$$

for any $x_i \in \Re_+^{n_i}$, $i = 1, \dots, m$. Applying the assertion in Lemma 2.1 to (5.8) by setting $P = r_i I_{n_i}$, $\bar{z} = x_i^k$, $z = x_i^{k+1}$, $\vartheta(\cdot) = \theta_i(\cdot)$, $q(z) = -A_i^T\lambda^k + \beta A_i^T\Big(\sum_{j=1}^m A_j x_j^k - b\Big) + \beta A_i^T A_i(x_i^{k+1} - x_i^k)$ and $z' = x_i$ in (2.6), for any $x_i \in \Re_+^{n_i}$, $i = 1, \dots, m$, we have

$$\theta_i(x_i^{k+1}) - \theta_i(x_i) + (x_i^{k+1} - x_i)^T\Big\{ - A_i^T\lambda^k + \beta A_i^T\Big(\sum_{j=1}^m A_j x_j^k - b\Big)$$
$$+\beta A_i^T A_i(x_i^{k+1} - x_i^k) + (1+\mu)r_i(x_i^{k+1} - x_i^k)\Big\} \le \mu r_i\|x_i^k - x_i^{k+1}\|^2. \tag{5.9}$$

Setting $x_i = x_i^k$, $i = 1, \dots, m$ in (5.9), we have

$$\theta_i(x_i^{k+1}) - \theta_i(x_i^k) + (x_i^{k+1} - x_i^k)^T\Big\{ - A_i^T\lambda^k + \beta A_i^T\Big(\sum_{j=1}^m A_j x_j^k - b\Big)$$
$$+\beta A_i^T A_i(x_i^{k+1} - x_i^k) + (1+\mu)r_i(x_i^{k+1} - x_i^k)\Big\} \le \mu r_i\|x_i^k - x_i^{k+1}\|^2. \tag{5.10}$$

Note that (5.9) is also true for $k := k-1$ and thus we have

$$\theta_i(x_i^k) - \theta_i(x_i) + (x_i^k - x_i)^T\Big\{ - A_i^T\lambda^{k-1} + \beta A_i^T\Big(\sum_{j=1}^m A_j x_j^{k-1} - b\Big)$$
$$+\beta A_i^T A_i(x_i^k - x_i^{k-1}) + (1+\mu)r_i(x_i^k - x_i^{k-1})\Big\} \le \mu r_i\|x_i^{k-1} - x_i^k\|^2$$

for any $x_i \in \Re_+^{n_i}$, $i = 1, \dots, m$. Setting $x_i = x_i^{k+1}$, $i = 1, \dots, m$ in the above inequality, we obtain

$$\theta_i(x_i^k) - \theta_i(x_i^{k+1}) + (x_i^k - x_i^{k+1})^T\Big\{ - A_i^T\lambda^{k-1} + \beta A_i^T\Big(\sum_{j=1}^m A_j x_j^{k-1} - b\Big)$$
$$+\beta A_i^T A_i(x_i^k - x_i^{k-1}) + (1+\mu)r_i(x_i^k - x_i^{k-1})\Big\} \le \mu r_i\|x_i^{k-1} - x_i^k\|^2. \tag{5.11}$$

Adding (5.10) and (5.11), we get

$$(x_i^{k+1} - x_i^k)^T\Big\{ - A_i^T(\lambda^k - \lambda^{k-1}) + [\beta A_i^T A_i + (1+\mu)r_i I_{n_i}][(x_i^{k+1} - x_i^k)$$
$$-(x_i^k - x_i^{k-1})] + \beta A_i^T\Big[\sum_{j=1}^m A_j(x_j^k - x_j^{k-1})\Big]\Big\}$$
$$\le \mu r_i(\|x_i^k - x_i^{k+1}\|^2 + \|x_i^{k-1} - x_i^k\|^2) \quad \forall\, i = 1, \dots, m.$$

Denote $\Delta x_i^{k+1} := x_i^{k+1} - x_i^k$, $\Delta x_i^k := x_i^k - x_i^{k-1}$ and $\Delta\lambda^k := \lambda^k - \lambda^{k-1}$. From the above inequality, we obtain

$$(\Delta x_i^{k+1})^T\Big\{ - A_i^T\Delta\lambda^k + \beta A_i^T\Big(\sum_{j=1}^m A_j\Delta x_j^k\Big) + [\beta A_i^T A_i + (1+\mu)r_i I_{n_i}](\Delta x_i^{k+1} - \Delta x_i^k)\Big\}$$
$$\le \mu r_i(\|\Delta x_i^{k+1}\|^2 + \|\Delta x_i^k\|^2) \quad \forall\, i = 1, \dots, m.$$

Summing the above inequalities over $i = 1, \dots, m$, we have

$$-(\Delta x^{k+1})^T A^T\Delta\lambda^k + \beta(\Delta x^{k+1})^T A^T A\Delta x^k + (\Delta x^{k+1})^T\bar{H}_x(\Delta x^{k+1} - \Delta x^k)$$
$$\le \|\Delta x^{k+1}\|_{N_x}^2 + \|\Delta x^k\|_{N_x}^2.$$

It follows from the above inequality that

$$
\begin{aligned}
(\Delta x^{k+1})^T A^T \Delta\lambda^k &\geq \beta(\Delta x^{k+1})^T A^T A \Delta x^k + (\Delta x^{k+1})^T \bar{H}_x(\Delta x^{k+1} - \Delta x^k) - (\|\Delta x^{k+1}\|_{N_x}^2 + \|\Delta x^k\|_{N_x}^2) \\
&= \|\Delta x^{k+1}\|_{\bar{H}_x}^2 - (\Delta x^{k+1})^T(\bar{H}_x - \beta A^T A)\Delta x^k - \|\Delta x^{k+1}\|_{N_x}^2 - \|\Delta x^k\|_{N_x}^2. \quad (5.12)
\end{aligned}
$$

Since $H = \operatorname{diag}(\bar{H}_x - \beta A^T A, \frac{1}{\gamma\beta} I_l) \succ 0$, we have $\bar{H}_x - \beta A^T A \succ 0$. Then, using the Cauchy-Schwarz inequality, we obtain

$$
-2(\Delta x^{k+1})^T(\bar{H}_x - \beta A^T A)\Delta x^k \geq -\|\Delta x^{k+1}\|_{\bar{H}_x - \beta A^T A}^2 - \|\Delta x^k\|_{\bar{H}_x - \beta A^T A}^2. \quad (5.13)
$$

Substituting (5.13) into (5.12), we get

$$
\begin{aligned}
&2(\Delta x^{k+1})^T A^T \Delta\lambda^k \\
&\geq 2\|\Delta x^{k+1}\|_{\bar{H}_x}^2 - \|\Delta x^{k+1}\|_{\bar{H}_x - \beta A^T A}^2 - \|\Delta x^k\|_{\bar{H}_x - \beta A^T A}^2 - 2\|\Delta x^{k+1}\|_{N_x}^2 - 2\|\Delta x^k\|_{N_x}^2 \\
&= \|\Delta x^{k+1}\|_{\bar{H}_x + \beta A^T A - 2N_x}^2 - \|\Delta x^k\|_{\bar{H}_x - \beta A^T A + 2N_x}^2.
\end{aligned}
$$

Note that $\Delta\lambda^{k+1} = \Delta\lambda^k - \gamma\beta A \Delta x^{k+1}$. It follows from the above formula that

$$
\begin{aligned}
&\frac{1}{\gamma\beta}\|\Delta\lambda^k\|^2 - \frac{1}{\gamma\beta}\|\Delta\lambda^{k+1}\|^2 \\
&= \frac{1}{\gamma\beta}\|\Delta\lambda^k\|^2 - \frac{1}{\gamma\beta}\|\Delta\lambda^k - \gamma\beta A\Delta x^{k+1}\|^2 \\
&= 2(\Delta x^{k+1})^T A^T \Delta\lambda^k - \gamma\beta\|A\Delta x^{k+1}\|^2 \\
&\geq \|\Delta x^{k+1}\|_{\bar{H}_x + (1-\gamma)\beta A^T A - 2N_x}^2 - \|\Delta x^k\|_{\bar{H}_x - \beta A^T A + 2N_x}^2.
\end{aligned}
$$

Using this and the definition of $\bar{H}'_x$, we have

$$
\begin{aligned}
&(\|\Delta x^k\|_{\bar{H}'_x}^2 + \frac{1}{\gamma\beta}\|\Delta\lambda^k\|^2) - (\|\Delta x^{k+1}\|_{\bar{H}'_x}^2 + \frac{1}{\gamma\beta}\|\Delta\lambda^{k+1}\|^2) \\
&\geq \|\Delta x^{k+1}\|_{\bar{H}_x + (1-\gamma)\beta A^T A - 2N_x - \bar{H}'_x}^2 \\
&= \|\Delta x^{k+1}\|_{(2-\gamma)\beta A^T A - 4N_x}^2. \quad (5.14)
\end{aligned}
$$

It follows from

$$
\mu < \frac{(2-\gamma)\beta\lambda_{\min}(A^T A)}{4\max_{i=1,\ldots,m}\{r_i\}}
$$

that $(2-\gamma)\beta A^T A - 4N_x \succeq 0$. Using (5.14), we have

$$
\|\Delta x^{k+1}\|_{\bar{H}'_x}^2 + \frac{1}{\gamma\beta}\|\Delta\lambda^{k+1}\|^2 \leq \|\Delta x^k\|_{\bar{H}'_x}^2 + \frac{1}{\gamma\beta}\|\Delta\lambda^k\|^2.
$$

Then, by the definition of $\bar{H}'$, we get the assertion (5.7). ■

Now, using previous results, we can establish a worst-case $O(1/t)$ convergence rate in a non-ergodic sense for Algorithm 1.

Theorem 5.3. *Let $\{w^t\}$ be generated by Algorithm 1. Assume that A is a matrix of full column rank; $\beta > 0$; $\gamma \in (0,2)$; $r_i > 0$, $i = 1, \ldots, m$ and $\mu \in (0,1)$, where r_i and μ satisfy the conditions:*

$$
r_i > (m-1)\beta\lambda_{\max}(A_i^T A_i)
$$

and

$$
0 < \mu < \min\Big\{\min_{i=1,\ldots,m}\{1 - \frac{(m-1)\beta\lambda_{\max}(A_i^T A_i)}{r_i}\}, \frac{(2-\gamma)\beta\lambda_{\min}(A^T A)}{4\max_{i=1,\ldots,m}\{r_i\}}\Big\}.
$$

Then, there is a constant $c_0 > 0$ such that for any $w^* \in \Omega^*$ and any integer $t \geq 0$,

$$
\|w^{t+1} - w^t\|_{\bar{H}'}^2 \leq \frac{c_0}{t+1}\|w^0 - w^*\|_H^2, \quad (5.15)
$$

where H and $\bar{H}'$ are defined in (2.13) *and* (5.4)*, respectively.*

Proof. Since $\tilde{H} \succ 0$, it's easy to prove $M^{-T}\tilde{H}M^{-1} \succ 0$. Note that $\bar{H}' \succ 0$. There is a constant $c_0 > 0$, such that

$$c_0 M^{-T}\tilde{H}M^{-1} \succeq \bar{H}'. \tag{5.16}$$

And thus, it follows from $M(w^k - \tilde{w}^k) = (w^k - w^{k+1})$ (see (4.3)) and (4.16) that

$$\begin{aligned}\sum_{k=0}^{\infty} \|w^k - w^{k+1}\|_{\bar{H}'}^2 &\leq \sum_{k=0}^{\infty} c_0 \|w^k - w^{k+1}\|_{M^{-T}\tilde{H}M^{-1}}^2 \\ &= \sum_{k=0}^{\infty} c_0 \|w^k - \tilde{w}^k\|_{\tilde{H}}^2 \leq c_0 \|w^0 - w^*\|_H^2. \end{aligned} \tag{5.17}$$

It follows from (5.7) that

$$\|w^{t+1} - w^t\|_{\bar{H}'}^2 \leq \|w^{k+1} - w^k\|_{\bar{H}'}^2 \quad \forall\, 0 < k \leq t.$$

And thus we have

$$(t+1)\|w^{t+1} - w^t\|_{\bar{H}'}^2 \leq \sum_{k=0}^{t} \|w^{k+1} - w^k\|_{\bar{H}'}^2 \leq \sum_{k=0}^{\infty} \|w^{k+1} - w^k\|_{\bar{H}'}^2.$$

From the above inequality and (5.17), we get the assertion (5.15). The proof is complete. ∎

It follows from (4.3) and (4.4) that if $w^{t+1} = w^t$, we have $w^t = \tilde{w}^t$ and w^t is the solution of $\mathrm{VI}(\Omega, F, \theta)$. Therefore, $\|w^{t+1} - w^t\|_{\bar{H}'}^2$ can be viewed as an error measurement in term of the distance to the solution set of $\mathrm{VI}(\Omega, F, \theta)$ for the t-th iteration of Algorithm 1. Notice that Ω^* is convex and closed. Let $d := \inf\{c_0\|w^0 - w^*\|_H^2 \,|\, w^* \in \Omega^*\}$. Then, for any given $\varepsilon > 0$, the inequality (5.15) shows under the assumptions that A is a matrix of full column rank, (5.5) and (5.6), Algorithm 1 needs at most

$$\lceil \frac{d}{\varepsilon} - 1 \rceil$$

iterations to ensure that $\|w^{t+1} - w^t\|_{\bar{H}'}^2 \leq \varepsilon$. Therefore, a worst-case $O(1/t)$ convergence rate is established for Algorithm 1 in a non-ergodic sense.

5.3. A Worst-case $o(1/t)$ Convergence Rate in the Non-ergodic Sense

The worst-case $O(1/t)$ convergence rate of Algorithm 1 in a non-ergodic sense established in the last subsection can be easily refined as a worst-case $o(1/t)$ convergence rate. We summarize it in the following theorem.

Theorem 5.4. *Let $\{w^t\}$ be generated by Algorithm* 1*. Assume that A is a matrix of full column rank; $\beta > 0$; $\gamma \in (0,2)$; $r_i > 0$, $i = 1, \ldots, m$ and $\mu \in (0,1)$, where r_i and μ satisfy the conditions:*

$$r_i > (m-1)\beta\lambda_{\max}(A_i^T A_i)$$

and

$$0 < \mu < \min\Big\{ \min_{i=1,\ldots,m} \{1 - \frac{(m-1)\beta\lambda_{\max}(A_i^T A_i)}{r_i}\}, \frac{(2-\gamma)\beta\lambda_{\min}(A^T A)}{4\max_{i=1,\ldots,m}\{r_i\}} \Big\}.$$

Then we obtain

$$\|w^{t+1} - w^t\|_{\bar{H}'}^2 = o(1/t),$$

where $\bar{H}'$ is defined in (5.4)*.*

Proof. It follows from (5.17) that

$$\sum_{t=0}^{\infty} \|w^{t+1} - w^t\|_{\bar{H}'}^2 < +\infty.$$

On the other hand, Lemma 5.2 implies the monotone non-increasing of $\|w^{t+1} - w^t\|_{\bar{H}'}^2$. By Lemma 2.3, we have $\|w^{t+1} - w^t\|_{\bar{H}'}^2 = o(1/t)$, which completes the proof. ■

5.4. **Linear Convergence**

In this subsection, we investigate the linear convergence of the scheme (3.2) with additional assumptions on the model (1.1). Our discussion in this subsection is under the following assumptions.

Assumption 1. The functions θ_i are strongly convex with the parameters $\sigma_i > 0$, respectively, for $i = 1, \dots, m$. That is, the following inequality holds:

$$(x_i - y_i)^T (g_i(x_i) - g_i(y_i)) \geq \sigma_i \|x_i - y_i\|^2$$

for any $x_i, y_i \in \Re_+^{n_i}$, $g_i(x_i) \in \partial\theta_i(x_i)$, $g_i(y_i) \in \partial\theta_i(y_i)$, $i = 1, \dots, m$.

Assumption 2. The functions g_i are Lipschitz continuous with the parameters $L_{\theta_i} > 0$, respectively, for $i = 1, \dots, m$. That is, the following inequality holds:

$$\|g_i(x_i) - g_i(y_i)\| \leq L_{\theta_i} \|x_i - y_i\|$$

for any $x_i, y_i \in \Re_+^{n_i}$, $g_i(x_i) \in \partial\theta_i(x_i)$, $g_i(y_i) \in \partial\theta_i(y_i)$, $i = 1, \dots, m$.

Assumption 3. Assume for any $w^* = (x_1^*, \dots, x_m^*, \lambda^*) \in \Omega^*$, x_i^* is an interior point of $\Re_+^{n_i}$, that is $x_i^* \in \Re_{++}^{n_i}$ for any $i = 1, \dots, m$.

By Assumption 3, there is a constant $\varsigma > 0$ such that

$$X_i^k (\tilde{X}_i^k)^{-1} \preceq \varsigma I_{n_i} \qquad \forall i = 1, \dots, m, \tag{5.18}$$

since $\lim_{k\to\infty} \|x_i^k - x_i^*\| = 0$ and $\lim_{k\to\infty} \|x_i^k - \tilde{x}_i^k\| = 0$. For convenience, let

$$G_\beta := \beta \text{diag}(A_1^T A_1, \dots, A_m^T A_m) - \beta A^T A \quad \text{and} \quad \Psi(x^k, \tilde{x}^k) := \begin{pmatrix} r_1 \Phi'(x_1^k, \tilde{x}_1^k) \\ \vdots \\ r_m \Phi'(x_m^k, \tilde{x}_m^k) \end{pmatrix}. \tag{5.19}$$

Moreover, we denote

$$\sigma_\theta := \min_{i=1,\dots,m} \{\sigma_i\}, \quad L_\theta := \max_{i=1,\dots,m} \{L_{\theta_i}\} \quad \text{and} \quad g(x) := \begin{pmatrix} g_1(x_1) \\ \vdots \\ g_m(x_m) \end{pmatrix}, \tag{5.20}$$

where $g_i(x) \in \partial\theta_i(x)$, $i = 1, \dots, m$. Then we get $g(x) \in \partial\theta(x)$. By Assumptions 1 and 2, for any $x, y \in \Re_+^{n_1} \times \cdots \times \Re_+^{n_m}$, $g(x) \in \partial\theta(x)$, $g(y) \in \partial\theta(y)$, we have

$$(x - y)^T (g(x) - g(y)) \geq \sigma_\theta \|x - y\|^2 \tag{5.21}$$

and

$$\|g(x) - g(y)\| \leq L_\theta \|x - y\|. \tag{5.22}$$

We have the following assertion whose proof is similar as Lemma 4.1 and thus omitted.

Lemma 5.5. *Let $\{w^k\}$ be generated by Algorithm* 1 *and $\{\tilde{w}^k\}$ be defined in* (4.1). *Then, for any $w \in \Omega$, we have*

$$(\tilde{x}^k - x)^T g(\tilde{x}^k) + (\tilde{w}^k - w)^T F(\tilde{w}^k) \leq -(w - \tilde{w}^k)^T G(w^k - \tilde{w}^k) + \|w^k - \tilde{w}^k\|_N^2, \tag{5.23}$$

where $g(\cdot)$, the matrices G and N are defined in (5.20), (2.12) *and* (2.15).

Theorem 5.6. *Let $\{w^k\}$ be generated by Algorithm 1, $\{\tilde{w}^k\}$ be defined in (4.1), and H and $\tilde{H}$ be defined in (2.13) and (2.16), respectively. Assume that Assumption 1 holds. Then for any $w \in \Omega$, we have*

$$\begin{aligned}&(\tilde{x}^k - x)^T g(x) + (\tilde{w}^k - w)^T F(w)\\&\le \frac{1}{2}(\|w - w^k\|_H^2 - \|w - w^{k+1}\|_H^2) - \frac{1}{2}\|w^k - \tilde{w}^k\|_{\tilde{H}}^2 - \sigma_\theta\|\tilde{x}^k - x^k\|^2,\end{aligned} \tag{5.24}$$

where $g(\cdot)$ and σ_θ are defined in (5.20).

Proof. Note that F is monotone. And thus we have

$$(\tilde{w}^k - w)^T F(w) \le (\tilde{w}^k - w)^T F(\tilde{w}^k).$$

Since the functions θ_i, $i = 1, \ldots, m$ are strongly convex, using the notation of $g(x)$ and σ_θ, we obtain

$$(\tilde{x}^k - x)^T g(\tilde{x}^k) \ge (\tilde{x}^k - x)^T g(x) + \sigma_\theta\|\tilde{x}^k - x\|^2.$$

It follows from the above two inequalities and (5.23) that

$$\begin{aligned}&(\tilde{x}^k - x)^T g(x) + \sigma_\theta\|\tilde{x}^k - x\|^2 + (\tilde{w}^k - w)^T F(w)\\&\le -(w - \tilde{w}^k)^T G(w^k - \tilde{w}^k) + \|w^k - \tilde{w}^k\|_N^2 \qquad \forall\, w \in \Omega.\end{aligned} \tag{5.25}$$

The assertion (5.24) follows immediately from (5.25) and (4.10). ■

Lemma 5.7. *Let $\{w^k\}$ be generated by Algorithm 1, $\{\tilde{w}^k\}$ be defined in (4.1), and H and $\tilde{H}$ be defined in (2.13) and (2.16), respectively. Assume that Assumption 1 holds. Then for any $w^* \in \Omega^*$, we have*

$$\|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 - \|w^k - \tilde{w}^k\|_{\tilde{H}}^2 - 2\sigma_\theta\|x^{k+1} - x^*\|^2, \tag{5.26}$$

where σ_θ is defined in (5.20).

Proof. Similar as (2.9), if $w^* \in \Omega^*$ we have

$$(x - x^*)^T g(x^*) + (w - w^*)^T F(w^*) \ge 0 \quad \forall\, w \in \Omega. \tag{5.27}$$

Setting $w = w^*$ in (5.24) where w^* being an arbitrary solution point in Ω^*, using (5.27) we get

$$\begin{aligned}&\|w^k - w^*\|_H^2 - \|w^{k+1} - w^*\|_H^2\\&\ge \|w^k - \tilde{w}^k\|_{\tilde{H}}^2 + 2\sigma_\theta\|\tilde{x}^k - x^k\|^2 + 2[(\tilde{x}^k - x^*)^T g(x^*) + (\tilde{w}^k - w^*)^T F(w^*)]\\&\ge \|w^k - \tilde{w}^k\|_{\tilde{H}}^2 + 2\sigma_\theta\|\tilde{x}^k - x^k\|^2.\end{aligned}$$

Note that $x^{k+1} = \tilde{x}^k$. The assertion (5.26) is proved from the above inequality. ■

Lemma 5.8. *Let $\{w^k\}$ be generated by Algorithm 1 and $\{\tilde{w}^k\}$ be defined in (4.1). Assume that Assumptions 2 and 3 hold, and the matrix A has full row rank. Then for $w^* = (x^*, \lambda^*) \in \Omega^*$ and $\kappa > 1$, we have*

$$\|\tilde{\lambda}^k - \lambda^*\|^2 \le c_1\|x^{k+1} - x^k\|^2 + c_2\|x^{k+1} - x^*\|^2, \tag{5.28}$$

where

$$c_1 := 2\kappa\lambda_{\min}^{-1}(AA^T)(\|G_\beta^T G_\beta\| + \max_{i=1,\ldots,m}\{r_i^2\}(1 + \mu\varsigma)), \quad c_2 := \frac{\kappa L_\theta^2}{\kappa - 1}\lambda_{\min}^{-1}(AA^T),$$

ς, G_β and L_θ are defined in (5.18), (5.19) and (5.20), respectively.

Proof. Similar as (4.7), we get $\tilde{x}_i^k \in \Re_{++}^{n_i}$ and

$$\begin{aligned}&(x_i - \tilde{x}_i^k)^T\{g_i(\tilde{x}_i^k) - A_i^T\tilde{\lambda}^k + \beta A_i^T A_i(\tilde{x}_i^k - x_i^k) - \beta A_i^T[\sum_{j=1}^m A_j(\tilde{x}_j^k - x_j^k)]\\&+ r_i\Phi'(x_i^k, \tilde{x}_i^k)\} \ge 0 \qquad \forall x_i \in \Re_+^{n_i},\ i = 1, \ldots, m.\end{aligned} \tag{5.29}$$

Combining (5.29) from $i=1$ to m, we get $\tilde{x}^k=(\tilde{x}_1^k,\ldots,\tilde{x}_m^k)\in\Re_{++}^{n_1}\times\cdots\times\Re_{++}^{n_m}$; and for any $x=(x_1,\ldots,x_m)\in\Re_+^{n_1}\times\cdots\times\Re_+^{n_m}$, we have

$$\begin{pmatrix} x_1-\tilde{x}_1^k \\ \vdots \\ x_m-\tilde{x}_m^k \end{pmatrix}^T \left\{ \begin{pmatrix} g_1(\tilde{x}_1^k) \\ \vdots \\ g_m(\tilde{x}_m^k) \end{pmatrix} + \begin{pmatrix} -A_1^T\tilde{\lambda}^k-\beta A_1^T[\sum_{j=2}^m A_j(\tilde{x}_j^k-x_j^k)]+r_1\Phi'(x_1^k,\tilde{x}_1^k) \\ \vdots \\ -A_m^T\tilde{\lambda}^k-\beta A_m^T[\sum_{j=1}^{m-1} A_j(\tilde{x}_j^k-x_j^k)]+r_m\Phi'(x_m^k,\tilde{x}_m^k) \end{pmatrix} \right\} \geq 0.$$

Then we have $\tilde{x}^k\in\Re_{++}^{n_1}\times\cdots\times\Re_{++}^{n_m}$. For any $x\in\Re_+^{n_1}\times\cdots\times\Re_+^{n_m}$, using the notation of g, G_β and Ψ, we get

$$(x-\tilde{x}^k)^T\big[g(\tilde{x}^k)-A^T\tilde{\lambda}^k+G_\beta(\tilde{x}^k-x^k)+\Psi(x^k,\tilde{x}^k)\big]\geq 0.$$

Since $\tilde{x}^k\in\Re_{++}^{n_1}\times\cdots\times\Re_{++}^{n_m}$, from the above variational inequality, we get

$$g(\tilde{x}^k)=A^T\tilde{\lambda}^k-G_\beta(\tilde{x}^k-x^k)-\Psi(x^k,\tilde{x}^k). \tag{5.30}$$

Note that $(x^*,\lambda^*)\in\Omega^*$. It holds that

$$(x-x^*)^T\big(g(x^*)-A^T\lambda^*\big)\geq 0 \qquad \forall x\in\Re_+^{n_1}\times\cdots\times\Re_+^{n_m}.$$

Since $x^*\in\Re_{++}^{n_1}\times\cdots\times\Re_{++}^{n_m}$, from the above variational inequality, we obtain

$$g(x^*)=A^T\lambda^*. \tag{5.31}$$

Recall that Assumption 2 holds, that is, g_i is Lipschitz continuous with parameters L_{θ_i}, $i=1,\ldots,m$. Then, by the definition of L_θ in (5.20), we obtain

$$\|g(\tilde{x}^k)-g(x^*)\|^2\leq L_\theta^2\|\tilde{x}^k-x^*\|^2.$$

Substituting (5.30) and (5.31) into the above inequality, we have

$$\|A^T\tilde{\lambda}^k-A^T\lambda^*-G_\beta(\tilde{x}^k-x^k)-\Psi(x^k,\tilde{x}^k)\|^2\leq L_\theta^2\|\tilde{x}^k-x^*\|^2. \tag{5.32}$$

Applying the following basic inequality

$$\|u-v\|^2\geq(1-\frac{1}{\kappa})\|u\|^2+(1-\kappa)\|v\|^2 \qquad \forall\kappa>0,$$

to the left-hand side of (5.32) by setting $u=A^T\tilde{\lambda}^k-A^T\lambda^*$, $v=G_\beta(\tilde{x}^k-x^k)+\Psi(x^k,\tilde{x}^k)$ and $\kappa>1$, we have

$$(1-\frac{1}{\kappa})\|A^T\tilde{\lambda}^k-A^T\lambda^*\|^2+(1-\kappa)\|G_\beta(\tilde{x}^k-x^k)+\Psi(x^k,\tilde{x}^k)\|^2\leq L_\theta^2\|\tilde{x}^k-x^*\|^2.$$

Using the assumption that A has full row rank, we obtain $\lambda_{\min}(AA^T)>0$ and

$$\|A^T\tilde{\lambda}^k-A^T\lambda^*\|^2=(\tilde{\lambda}^k-\lambda^*)^TAA^T(\tilde{\lambda}^k-\lambda^*)\geq\lambda_{\min}(AA^T)\|\tilde{\lambda}^k-\lambda^*\|^2.$$

Substituting this into the above inequality, we obtain

$$\|\tilde{\lambda}^k-\lambda^*\|^2\leq\kappa\lambda_{\min}^{-1}(AA^T)\|G_\beta(\tilde{x}^k-x^k)+\Psi(x^k,\tilde{x}^k)\|^2+\frac{\kappa L_\theta^2}{\kappa-1}\lambda_{\min}^{-1}(AA^T)\|\tilde{x}^k-x^*\|^2. \tag{5.33}$$

Using the definitions of $\Psi(x^k,\tilde{x}^k)$ and $\Phi'(x_i^k,\tilde{x}_i^k)$, $i=1,\ldots,m$, we get

$$\begin{aligned}\|\Psi(x^k,\tilde{x}^k)\|^2 &= \sum_{i=1}^m r_i^2\|\Phi'(x_i^k,\tilde{x}_i^k)\|^2=\sum_{i=1}^m r_i^2\|(\tilde{x}_i^k-x_i^k)+\mu[x_i^k-(X_i^k)^2(\tilde{x}_i^k)^{-1}]\|^2 \\ &= \sum_{i=1}^m r_i^2\big\|[I_{n_i}+\mu X_i^k(\tilde{X}_i^k)^{-1}](\tilde{x}_i^k-x_i^k)\big\|^2.\end{aligned}$$

From the above inequality and the definition of ς in (5.18), it follows that

$$\|\Psi(x^k,\tilde{x}^k)\|^2\leq\max_{i=1,\ldots,m}\{r_i^2\}(1+\mu\varsigma)\|\tilde{x}^k-x^k\|^2.$$

Using the above inequality, we get

$$\begin{aligned}\|G_\beta(\tilde{x}^k - x^k) + \Psi(x^k, \tilde{x}^k)\|^2 &\leq 2\|G_\beta(\tilde{x}^k - x^k)\|^2 + 2\|\Psi(x^k, \tilde{x}^k)\|^2 \\ &\leq 2(\|G_\beta^T G_\beta\| + \max_{i=1,\ldots,m}\{r_i^2\}(1+\mu\varsigma))\|\tilde{x}^k - x^k\|^2.\end{aligned}$$

Substituting this into (5.33), we get the assertion (5.28). ∎

Theorem 5.9. *Let* $\{w^k\}$ *be generated by Algorithm* 1 *and* $\{\tilde{w}^k\}$ *be defined in* (4.1)*. Assume that Assumptions* 1-3 *hold, and the matrix* A *has full row rank. Then, there is a constant* $\delta > 0$*, such that*

$$\|w^{k+1} - w^*\|_H^2 \leq \frac{1}{1+\delta}\|w^k - w^*\|_H^2 \qquad \forall w^* \in \Omega^*, \tag{5.34}$$

where H *is defined in* (2.13)*.*

Proof. First, it follows from the definitions of λ^{k+1} and $\tilde{\lambda}^k$ that

$$\lambda^{k+1} - \tilde{\lambda}^k = (\gamma - 1)(\tilde{\lambda}^k - \lambda^k).$$

Using the Cauchy-Schwarz inequality, the above equation and (5.28), we have

$$\begin{aligned}\|\lambda^{k+1} - \lambda^*\|^2 &= \|\tilde{\lambda}^k - \lambda^* + \lambda^{k+1} - \tilde{\lambda}^k\|^2 \\ &\leq 2\|\tilde{\lambda}^k - \lambda^*\|^2 + 2\|\lambda^{k+1} - \tilde{\lambda}^k\|^2 \\ &= 2\|\tilde{\lambda}^k - \lambda^*\|^2 + 2(\gamma-1)^2\|\tilde{\lambda}^k - \lambda^k\|^2 \\ &\leq 2c_1\|x^{k+1} - x^k\|^2 + 2c_2\|x^{k+1} - x^*\|^2 + 2(\gamma-1)^2\|\tilde{\lambda}^k - \lambda^k\|^2.\end{aligned}$$

By this inequality and the definitions of H and $\tilde{H}$, there is a constant $\delta > 0$ such that

$$\delta\|w^{k+1} - w^*\|_H^2 \leq \|w^k - \tilde{w}^k\|_{\tilde{H}}^2 + 2\sigma_\theta\|x^{k+1} - x^*\|^2.$$

Substituting this into (5.26), we get the assertion (5.34). ∎

6. The ALM with Jacobian Decomposition and LQP Regularization — Inexact Version

In this section, we delineate an inexact version of (3.2). The motivation for considering inexact versions is that in general the subproblems in (3.2) still require iterations to pursue approximate solutions even though they are easier unconstrained minimization problems. We thus can only expect to execute the scheme (3.2) practically in the following sense:

$$\left\{\begin{array}{l} x_1^{k+1} :\approx \operatorname{argmin}\{\mathcal{L}_\beta(x_1, x_2^k, x_3^k, \ldots, x_{m-1}^k, x_m^k, \lambda^k) + r_1 d(x_1, x_1^k)\}, \\ \cdots \\ x_i^{k+1} :\approx \operatorname{argmin}\{\mathcal{L}_\beta(x_1^k, x_2^k, \ldots, x_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_m^k, \lambda^k) + r_i d(x_i, x_i^k)\}, \\ \cdots \\ x_m^{k+1} :\approx \operatorname{argmin}\{\mathcal{L}_\beta(x_1^k, x_2^k, \ldots, x_{m-1}^k, x_m, \lambda^k) + r_m d(x_m, x_m^k)\}, \\ \lambda^{k+1} := \lambda^k - \gamma\beta(\sum_{j=1}^m A_j x_j^{k+1} - b). \end{array}\right. \tag{6.1}$$

To make the approximation in (6.1) precise, we elaborate on its detail of implementation as follows. Note that we only give a prototype algorithm for the inexact version (6.1), as our emphasis is to show the possibility of designing inexact version for the scheme (3.2) when the generic case of (1.1) is considered where the functions in its objective are generic functions. For a specific scenario where the functions are specified, it is incremental to develop a concrete desirable algorithm based on the algorithmic framework given below.

Algorithm 2.

Step 0. *Let* $\varepsilon > 0$, $\beta > 0$, $\mu \in (0,1)$, $\gamma \in (0,2)$, $r_i > (m-1)\beta\lambda_{\max}(A_i^T A_i)/(1-\mu)$, $i = 1,\dots,m$, *and* $\{\nu_k\}$ *be a nonnegative sequence satisfying* $\sum_{k=0}^{\infty}\nu_k < +\infty$. *Choose* $(x_1^0,\dots,x_m^0,\lambda^0) \in \Re^{n_1}_{++}\times\cdots\times\Re^{n_m}_{++}\times\Re^l$. *Set* $k := 0$.

Step 1. *Find* $x_i^{k+1} \in \Re^{n_i}_{++}$, $i = 1,\dots,m$, *in parallel, such that*

$$\|x_i^{k+1} - x_{i*}^{k+1}\| \le \nu_k, \tag{6.2}$$

where

$$x_{i*}^{k+1} := \operatorname{argmin}\{\mathcal{L}_\beta(x_1^k,\dots,x_{i-1}^k,x_i,x_{i+1}^k,\dots,x_m^k,\lambda^k) + r_i d(x_i,x_i^k)\}. \tag{6.3}$$

Step 2. *Update the Lagrange multiplier*

$$\lambda^{k+1} := \lambda^k - \gamma\beta\Big(\sum_{j=1}^m A_j x_j^{k+1} - b\Big). \tag{6.4}$$

Step 3. *Set* $w^{k+1} := (x_1^{k+1},\dots,x_m^{k+1},\lambda^{k+1})$. *If* $\|w^{k+1} - w^k\| \le \varepsilon$, *stop; otherwise set* $k := k+1$ *and goto Step 1.*

Remark 6.1. In Step 1, we find $x_i^{k+1} = x_i^k$ only if $x_{i*}^{k+1} = x_i^k$.

7. Convergence Analysis

Now, we analyze the convergence of the inexact version (6.1) in the sense of (6.2)-(6.4). Similarly, we first prove its global convergence and then establish its worst-case convergence rate measured by the iteration-complexity. The analytic framework is analogous to that in the last section, but more sophisticated reasoning and analysis is needed.

7.1. Global Convergence

We first prove the global convergence for Algorithm 2 from the contraction perspective. Similar as (4.1) and (6.4), we define

$$\lambda_*^{k+1} := \lambda^k - \gamma\beta\Big(\sum_{j=1}^m A_j x_{j*}^{k+1} - b\Big), \tag{7.1}$$

and set

$$w_*^{k+1} := \begin{pmatrix} x_{1*}^{k+1} \\ \vdots \\ x_{m*}^{k+1} \\ \lambda_*^{k+1} \end{pmatrix}, \quad \tilde{w}^k := \begin{pmatrix} \tilde{x}_1^k \\ \vdots \\ \tilde{x}_m^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x_{1*}^{k+1} \\ \vdots \\ x_{m*}^{k+1} \\ \lambda^k - \beta(\sum_{j=1}^m A_j x_{j*}^{k+1} - b) \end{pmatrix}, \tag{7.2}$$

where $(x_{1*}^{k+1},\dots,x_{m*}^{k+1})$ are the exact solutions of (6.3). Then, based on (7.1) and (7.2), we immediately have

$$x_{i*}^{k+1} = \tilde{x}_i^k, \quad i = 1,\dots,m$$

and

$$\lambda_*^{k+1} = \lambda^k - \gamma\beta(\sum_{j=1}^m A_j x_{j*}^{k+1} - b) = \lambda^k - \gamma(\lambda^k - \tilde{\lambda}^k).$$

We also have the following relationship

$$\begin{pmatrix} x_{1*}^{k+1} \\ \vdots \\ x_{m*}^{k+1} \\ \lambda_*^{k+1} \end{pmatrix} = \begin{pmatrix} x_1^k \\ \vdots \\ x_m^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I_{n_1} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I_{n_m} & 0 \\ 0 & \cdots & 0 & \gamma I_l \end{pmatrix} \begin{pmatrix} x_1^k - \tilde{x}_1^k \\ \vdots \\ x_m^k - \tilde{x}_m^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix},$$

which can be rewritten into a compact form by using the notation of w^k and $\tilde{w}^k$:

$$w_*^{k+1} = w^k - M(w^k - \tilde{w}^k), \tag{7.3}$$

where M is defined in (2.14). We first prove a simple lemma.

Lemma 7.1. *Let $\{w^k\}$ be generated by Algorithm* 2 *and $\{w_*^k\}$ be defined by* (7.2)*. Then, there exists a positive constant ρ such that*

$$\|w_*^{k+1} - w^{k+1}\|_H \le \rho\nu_k \qquad \forall k \ge 0, \tag{7.4}$$

where H is defined by (2.13)*.*

Proof. It follows from (7.1) that

$$\lambda_*^{k+1} - \lambda^{k+1} = \gamma\beta \sum_{j=1}^{m} A_j(x_j^{k+1} - x_{j*}^{k+1}).$$

Together with (6.2) and Lemma 2.4, the above equation implies (7.4) immediately. The proof is complete. ∎

Similarly to Lemma 4.1, Theorem 4.3 and Lemma 4.4, we have the following assertions but their proofs are omitted.

Lemma 7.2. *Let $\{w^k\}$ be generated by Algorithm* 2 *and $\{\tilde{w}^k\}$ be defined by* (7.2)*. Then, we have $\tilde{w}^k \in \Omega$ and*

$$\theta(\tilde{x}^k) - \theta(x) + (\tilde{w}^k - w)^T F(\tilde{w}^k) \le -(w - \tilde{w}^k)^T G(w^k - \tilde{w}^k) + \|w^k - \tilde{w}^k\|_N^2 \tag{7.5}$$

for any $w \in \Omega$, where the matrices G and N are defined in (2.12) *and* (2.15)*, respectively.*

Theorem 7.3. *Let $\{w^k\}$ be generated by Algorithm* 2*, $\{w_*^{k+1}\}$ and $\{\tilde{w}^k\}$ be defined in* (7.2)*, and H and $\tilde{H}$ be defined in* (2.13) *and* (2.16)*, respectively. Then for any $w \in \Omega$, we have*

$$\theta(\tilde{x}^k) - \theta(x) + (\tilde{w}^k - w)^T F(w) \le \frac{1}{2}(\|w - w^k\|_H^2 - \|w - w_*^{k+1}\|_H^2) - \frac{1}{2}\|w^k - \tilde{w}^k\|_{\tilde{H}}^2. \tag{7.6}$$

Lemma 7.4. *Let $\{w^k\}$ be generated by Algorithm* 2*, $\{w_*^{k+1}\}$ and $\{\tilde{w}^k\}$ be defined in* (7.2)*, and H and $\tilde{H}$ be defined in* (2.13) *and* (2.16)*, respectively. Then for any $w^* \in \Omega^*$, we have*

$$\|w_*^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 - \|w^k - \tilde{w}^k\|_{\tilde{H}}^2. \tag{7.7}$$

The following result shows a contraction property of the sequence generated by Algorithm 2, based on which the convergence of Algorithm 2 can be established easily.

Lemma 7.5. *Let $\{w^k\}$ be generated by Algorithm* 2*. Then $\{w^k\}$ is bounded, i.e., for any $w^* \in \Omega^*$, there is a positive constant C_{w^*}, such that*

$$\|w^k - w^*\|_H \le C_{w^*} \qquad \forall k \ge 0, \tag{7.8}$$

and

$$\|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 + 2\rho C_{w^*}\nu_k + \rho^2\nu_k^2 - \|w^k - \tilde{w}^k\|_{\tilde{H}}^2, \tag{7.9}$$

where H and $\tilde{H}$ are defined in (2.13) *and* (2.16)*, respectively.*

Proof. According to (7.7), for any given $w^* \in \Omega^*$, we have

$$\|w_*^{k+1} - w^*\|_H \le \|w^k - w^*\|_H.$$

It follows from the above inequality and (7.4) that

$$\|w^{k+1} - w^*\|_H \le \|w_*^{k+1} - w^*\|_H + \|w^{k+1} - w_*^{k+1}\|_H \le \|w^k - w^*\|_H + \rho\nu_k. \tag{7.10}$$

And thus for any $l \le k$ we have

$$\|w^{k+1} - w^*\|_H \le \|w^l - w^*\|_H + \rho\sum_{i=l}^{k} \nu_i.$$

Since $\sum_{k=0}^{\infty} \nu_k < +\infty$, there is a constant $C_{w^*} > 0$, such that

$$\|w^k - w^*\|_H \le C_{w^*} < +\infty \qquad \forall k \ge 0. \tag{7.11}$$

Therefore, the sequence $\{w^k\}$ generated by Algorithm 2 is bounded. It follows from (7.4), (7.7) and (7.11) that

$$\begin{aligned}
\|w^{k+1} - w^*\|_H^2 &= \|(w_*^{k+1} - w^*) + (w^{k+1} - w_*^{k+1})\|_H^2 \\
&\le \|w_*^{k+1} - w^*\|_H^2 + 2\|w^{k+1} - w_*^{k+1}\|_H \cdot \|w_*^{k+1} - w^*\|_H + \|w^{k+1} - w_*^{k+1}\|_H^2 \\
&\le \|w^k - w^*\|_H^2 - \|w^k - \tilde{w}^k\|_{\tilde{H}}^2 + 2\rho\nu_k\|w^k - w^*\|_H + \rho^2\nu_k^2 \\
&\le \|w^k - w^*\|_H^2 + 2\rho C_{w^*}\nu_k + \rho^2\nu_k^2 - \|w^k - \tilde{w}^k\|_{\tilde{H}}^2.
\end{aligned}$$

The proof is complete. ■

Now, we are ready to prove the convergence of Algorithm 2.

Theorem 7.6. *The sequence $\{w^k\}$ generated by Algorithm 2 converges to some w^∞ which is a solution of VI(Ω, F, θ).*

Proof. It follows from (7.9) that for any $l \le k$ and $w^* \in \Omega^*$, we have

$$\begin{aligned}
\|w^{k+1} - w^*\|_H^2 &\le \|w^k - w^*\|_H^2 + 2\rho C_{w^*}\nu_k + \rho^2\nu_k^2 \\
&\le \|w^l - w^*\|_H^2 + \sum_{i=l}^{k}(2\rho C_{w^*}\nu_i + \rho^2\nu_i^2).
\end{aligned} \tag{7.12}$$

Thus the sequence $\{w^k\}$ is bounded, since $\sum_{i=0}^{\infty}(2\rho C_{w^*}\nu_i + \rho^2\nu_i^2) < +\infty$. Summing the inequality (7.9) over $k = 0, 1, \ldots$, we get

$$\sum_{k=0}^{\infty} \|w^k - \tilde{w}^k\|_{\tilde{H}}^2 \le \|w^0 - w^*\|_H^2 + \sum_{k=0}^{\infty}(2\rho C_{w^*}\nu_k + \rho^2\nu_k^2) < +\infty. \tag{7.13}$$

Therefore, we have

$$\lim_{k\to\infty} \|w^k - \tilde{w}^k\|_{\tilde{H}} = 0. \tag{7.14}$$

Thus the sequence $\{\tilde{w}^k\}$ is also bounded, and it has at least one cluster point. Let w^∞ be a cluster point of $\{\tilde{w}^k\}$ and the subsequence $\{\tilde{w}^{k_j}\}$ converges to w^∞. It follows from (7.5) and (7.14) that

$$\liminf_{j\to\infty} \left\{\theta(x) - \theta(\tilde{x}^{k_j}) + (w - \tilde{w}^{k_j})^T F(\tilde{w}^{k_j})\right\} \ge 0 \qquad \forall w \in \Omega,$$

and consequently

$$\theta(x) - \theta(x^\infty) + (w - w^\infty)^T F(w^\infty) \ge 0 \qquad \forall w \in \Omega.$$

This means that w^∞ is a solution of VI(Ω, F, θ). Note that the inequality (7.12) is true for all solution points of VI(Ω, F, θ), hence we have

$$\|w^{k+1} - w^\infty\|_H^2 \le \|w^l - w^\infty\|_H^2 + \sum_{i=l}^{\infty}(2\rho C_{w^*}\nu_i + \rho^2\nu_i^2) \quad \forall k \ge 0,\ \forall l \le k. \tag{7.15}$$

Since $\tilde{w}^{k_j} \to w^\infty$ $(j \to \infty)$, using (7.14) we have $w^{k_j} \to w^\infty$ $(j \to \infty)$. For any given $\varepsilon > 0$, there exists a $j_0 > 0$ such that

$$\|w^{k_{j_0}} - w^\infty\|_H^2 \le \frac{\varepsilon^2}{2} \quad \text{and} \quad \sum_{i=k_{j_0}}^{\infty} (2\rho C_{w^*}\nu_i + \rho^2\nu_i^2) \le \frac{\varepsilon^2}{2}. \tag{7.16}$$

Therefore, for any $k \ge k_{j_0}$, it follows from (7.15) and (7.16) that

$$\|w^{k+1} - w^\infty\|_H \le \sqrt{\|w^{k_{j_0}} - w^\infty\|_H^2 + \sum_{i=k_{j_0}}^{\infty} (2\rho C_{w^*}\nu_i + \rho^2\nu_i^2)} \le \varepsilon.$$

This implies that the sequence $\{w^k\}$ converges to a point w^∞ in Ω^*. The proof is complete. ∎

7.2. **A Worst-case $O(1/t)$ Convergence Rate in the Ergodic Sense**

Now we establish a worst-case convergence rate measured by the iteration-complexity for the inexact version (6.1) in the sense of (6.2)-(6.4).

Theorem 7.7. *Let the sequence $\{w^k\}$ be generated by Algorithm 2. For any integer $t > 0$, there is a $\tilde{w}_t \in \Omega$ which is a convex combination of the iterates $\tilde{w}^0, \tilde{w}^1, \ldots, \tilde{w}^t$ defined by (7.2). Then for any $w \in \Omega$, we have*

$$\theta(\tilde{x}_t) - \theta(x) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{t+1}\Big(\frac{1}{2}\|w - w^0\|_H^2 + \rho\sum_{k=0}^{t} \nu_k \|w - w^{k+1}\|_H\Big), \tag{7.17}$$

where $\tilde{w}_t := (\sum_{k=0}^{t} \tilde{w}^k)/(t+1)$ and H is defined by (2.13).

Proof. From (7.6), we have

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2}\|w - w^k\|_H^2 \ge \frac{1}{2}\|w - w_*^{k+1}\|_H^2 \quad \forall w \in \Omega.$$

It follows from (7.4) that

$$\begin{aligned}\|w - w_*^{k+1}\|_H^2 &\ge \Big(\|w - w^{k+1}\|_H - \|w^{k+1} - w_*^{k+1}\|_H\Big)^2 \\ &= \|w - w^{k+1}\|_H^2 - 2\|w - w^{k+1}\|_H \cdot \|w^{k+1} - w_*^{k+1}\|_H + \|w^{k+1} - w_*^{k+1}\|_H^2 \\ &\ge \|w - w^{k+1}\|_H^2 - 2\rho\nu_k\|w - w^{k+1}\|_H \quad \forall w \in \Omega.\end{aligned}$$

From the above two inequalities, for any $w \in \Omega$ we have

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2}\|w - w^k\|_H^2 \ge \frac{1}{2}\|w - w^{k+1}\|_H^2 - \rho\nu_k\|w - w^{k+1}\|_H. \tag{7.18}$$

Summing the inequality (7.18) over $k = 0, 1, \ldots, t$, we obtain

$$\begin{aligned}&(t+1)\theta(x) - \sum_{k=0}^{t} \theta(\tilde{x}^k) + \Big[(t+1)w - \Big(\sum_{k=0}^{t} \tilde{w}^k\Big)\Big]^T F(w) + \frac{1}{2}\|w - w^0\|_H^2 \\ &\ge \frac{1}{2}\|w - w^{t+1}\|_H^2 - \rho\sum_{k=0}^{t} \nu_k\|w - w^{k+1}\|_H \\ &\ge -\rho\sum_{k=0}^{t} \nu_k\|w - w^{k+1}\|_H \qquad \forall w \in \Omega.\end{aligned}$$

Since $\sum_{k=0}^{t} 1/(t+1) = 1$, $\tilde{w}_t$ is a convex combination of $\tilde{w}^0, \tilde{w}^1, \ldots, \tilde{w}^t$ and thus $\tilde{w}_t \in \Omega$. Using the notation of $\tilde{w}_t$, we derive

$$\frac{1}{t+1}\Big(\sum_{k=0}^{t} \theta(\tilde{x}^k)\Big) - \theta(x) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{t+1}\Big(\frac{1}{2}\|w - w^0\|_H^2 + \rho \sum_{k=0}^{t} \nu_k \|w - w^{k+1}\|_H\Big) \quad (7.19)$$

for any $w \in \Omega$. Since $\theta(x)$ is convex and

$$\tilde{x}_t = \frac{1}{t+1}\sum_{k=0}^{t} \tilde{x}^k,$$

we have that

$$\theta(\tilde{x}_t) \le \frac{1}{t+1}\sum_{k=0}^{t} \theta(\tilde{x}^k).$$

Substituting it in (7.19), the assertion (7.17) follows directly. ■

It follows from the proof of Theorem 7.6 that the sequences $\{w^k\}$ and $\{\tilde{w}^k\}$ are bounded. Therefore, there exists a constant $D > 0$ such that

$$\|w^k\|_H \le D \qquad \text{and} \qquad \|\tilde{w}^k\|_H \le D \qquad \forall k \ge 0.$$

Recall that $\tilde{w}_t$ is the average of $\{\tilde{w}^0, \tilde{w}^1, \ldots, \tilde{w}^t\}$. Thus, we have $\|\tilde{w}_t\|_H \le D$. Denote

$$E_1 := \sum_{k=0}^{\infty} \nu_k < +\infty.$$

For any $w \in \mathcal{B}_\Omega(\tilde{w}_t) := \{w \in \Omega \mid \|w - \tilde{w}_t\|_H \le 1\}$, we get

$$\begin{aligned}
&\theta(\tilde{x}_t) - \theta(x) + (\tilde{w}_t - w)^T F(w) \\
&\le \frac{1}{t+1}\Big(\frac{1}{2}\|w - w^0\|_H^2 + \rho \sum_{k=0}^{t} \nu_k \|w - w^{k+1}\|_H\Big) \\
&\le \frac{1}{t+1}\Big[\frac{1}{2}\big(\|w - \tilde{w}_t\|_H + \|\tilde{w}_t\|_H + \|w^0\|_H\big)^2 + \rho \sum_{k=0}^{t} \nu_k \big(\|w - \tilde{w}_t\|_H + \|\tilde{w}_t\|_H + \|w^{k+1}\|_H\big)\Big] \\
&\le \frac{1}{t+1}\Big[\frac{1}{2}(1+2D)^2 + \rho E_1 (1+2D)\Big].
\end{aligned}$$

Thus, for any given $\varepsilon > 0$, after at most $t := \lceil \frac{(1+2D)(1+2D+2\rho E_1)}{2\varepsilon} - 1 \rceil$ iterations, we have

$$\theta(\tilde{x}_t) - \theta(x) + (\tilde{w}_t - w)^T F(w) \le \varepsilon \quad \forall w \in \mathcal{B}_\Omega(\tilde{w}_t),$$

which means $\tilde{w}_t$ is an approximate solution of $\mathrm{VI}(\Omega, F, \theta)$ with an accuracy of $O(1/t)$. That is, a worst-case $O(1/t)$ convergence rate measured by the iteration-complexity in the ergodic sense is established for the inexact version (6.1) in the sense of (6.2)-(6.4).

7.3. A Worst-case $O(1/t)$ Convergence Rate in the Non-ergodic Sense

Lemma 7.8. *Let $\{w^k\}$ be generated by Algorithm 2. Assume that A is a matrix of full column rank,*

$$r_i > (m-1)\beta\lambda_{\max}(A_i^T A_i) \quad (7.20)$$

and

$$0 < \mu < \min\Big\{\min_{i=1,\ldots,m}\{1 - \frac{(m-1)\beta\lambda_{\max}(A_i^T A_i)}{r_i}\}, \frac{(2-\gamma)\beta\lambda_{\min}(A^T A)}{4\max_{i=1,\ldots,m}\{r_i\}}\Big\}. \quad (7.21)$$

Then, $\bar{H}' \succ 0$, where $\bar{H}'$ is defined by (5.4), and there is a constant $c_3 > 0$, such that for any integer $k \geq 1$, we have

$$\|w^{k+1} - w^k\|_{\bar{H}'}^2 \leq \|w^k - w^{k-1}\|_{\bar{H}'}^2 + c_3(\nu_{k-1} + \nu_k). \tag{7.22}$$

Proof. As the proof in Lemma 5.2, we could prove $\bar{H}' \succ 0$, and the sequence $\{w^k\}$ generated by the scheme (6.2)-(6.4) under conditions (7.20) and (7.21) converges to a solution of VI(Ω, F, θ) by Theorem 7.6. Similar as (4.5), for (6.3), we have

$$\begin{aligned}&\theta_i(x_i) - \theta_i(x_{i*}^{k+1}) + (x_i - x_{i*}^{k+1})^T\Big\{ - A_i^T\lambda^k + \beta A_i^T\Big(\sum_{j=1}^m A_j x_j^k - b\Big)\\ &+\beta A_i^T A_i(x_{i*}^{k+1} - x_i^k) + r_i\Phi'(x_i^k, x_{i*}^{k+1})\Big\} \geq 0 \end{aligned}\tag{7.23}$$

for any $x_i \in \Re_+^{n_i}$, $i = 1, \ldots, m$. Applying the assertion in Lemma 2.1 to (7.23) by setting $P = r_i I_{n_i}$, $\bar{z} = x_i^k$, $z = x_{i*}^{k+1}$, $\vartheta(\cdot) = \theta_i(\cdot)$, $q(z) = -A_i^T\lambda^k + \beta A_i^T(\sum_{j=1}^m A_j x_j^k - b) + \beta A_i^T A_i(x_{i*}^{k+1} - x_i^k)$ and $z' = x_i$ in (2.6), for any $x_i \in \Re_+^{n_i}$, $i = 1, \ldots, m$, we have

$$\begin{aligned}&\theta_i(x_{i*}^{k+1}) - \theta_i(x_i) + (x_{i*}^{k+1} - x_i)^T\Big\{ - A_i^T\lambda^k + \beta A_i^T\Big(\sum_{j=1}^m A_j x_j^k - b\Big)\\ &+\beta A_i^T A_i(x_{i*}^{k+1} - x_i^k) + (1+\mu)r_i(x_{i*}^{k+1} - x_i^k)\Big\} \leq \mu r_i\|x_i^k - x_{i*}^{k+1}\|^2. \end{aligned}\tag{7.24}$$

Setting $x_i = x_{i*}^k$, $i = 1, \ldots, m$ in (7.24), we have

$$\begin{aligned}&\theta_i(x_{i*}^{k+1}) - \theta_i(x_{i*}^k) + (x_{i*}^{k+1} - x_{i*}^k)^T\Big\{ - A_i^T\lambda^k + \beta A_i^T\Big(\sum_{j=1}^m A_j x_j^k - b\Big)\\ &+\beta A_i^T A_i(x_{i*}^{k+1} - x_i^k) + (1+\mu)r_i(x_{i*}^{k+1} - x_i^k)\Big\} \leq \mu r_i\|x_i^k - x_{i*}^{k+1}\|^2. \end{aligned}\tag{7.25}$$

Note that (7.24) is also true for $k := k-1$ and thus we have

$$\begin{aligned}&\theta_i(x_{i*}^k) - \theta_i(x_i) + (x_{i*}^k - x_i)^T\Big\{ - A_i^T\lambda^{k-1} + \beta A_i^T\Big(\sum_{j=1}^m A_j x_j^{k-1} - b\Big)\\ &+\beta A_i^T A_i(x_{i*}^k - x_i^{k-1}) + (1+\mu)r_i(x_{i*}^k - x_i^{k-1})\Big\} \leq \mu r_i\|x_i^{k-1} - x_{i*}^k\|^2 \end{aligned}$$

for any $x_i \in \Re_+^{n_i}$, $i = 1, \ldots, m$. Setting $x_i = x_{i*}^{k+1}$, $i = 1, \ldots, m$ in the above inequality, we obtain

$$\begin{aligned}&\theta_i(x_{i*}^k) - \theta_i(x_{i*}^{k+1}) + (x_{i*}^k - x_{i*}^{k+1})^T\Big\{ - A_i^T\lambda^{k-1} + \beta A_i^T\Big(\sum_{j=1}^m A_j x_j^{k-1} - b\Big)\\ &+\beta A_i^T A_i(x_{i*}^k - x_i^{k-1}) + (1+\mu)r_i(x_{i*}^k - x_i^{k-1})\Big\} \leq \mu r_i\|x_i^{k-1} - x_{i*}^k\|^2. \end{aligned}\tag{7.26}$$

Adding (7.25) and (7.26), we get

$$\begin{aligned}&(x_{i*}^{k+1} - x_{i*}^k)^T\Big\{ - A_i^T(\lambda^k - \lambda^{k-1}) + [\beta A_i^T A_i + (1+\mu)r_i I_{n_i}][(x_{i*}^{k+1} - x_i^k)\\ &\quad -(x_{i*}^k - x_i^{k-1})] + \beta A_i^T\Big[\sum_{j=1}^m A_j(x_j^k - x_j^{k-1})\Big]\Big\}\\ &\leq \mu r_i(\|x_i^k - x_{i*}^{k+1}\|^2 + \|x_i^{k-1} - x_{i*}^k\|^2)\end{aligned}$$

for $i = 1, \ldots, m$. Denote $\Delta x_{i**}^{k+1} := x_{i*}^{k+1} - x_{i*}^k$, $\Delta x_{i*}^{k+1} := x_{i*}^{k+1} - x_i^k$, $\Delta x_i^k := x_i^k - x_i^{k-1}$, $\Delta x_{i*}^k := x_{i*}^k - x_i^{k-1}$ and $\Delta\lambda^k := \lambda^k - \lambda^{k-1}$. From the above inequality, we obtain

$$\begin{aligned}&(\Delta x_{i**}^{k+1})^T\Big\{ - A_i^T\Delta\lambda^k + \beta A_i^T\Big(\sum_{j=1}^m A_j\Delta x_j^k\Big) + [\beta A_i^T A_i + (1+\mu)r_i I_{n_i}](\Delta x_{i*}^{k+1} - \Delta x_{i*}^k)\Big\}\\ &\le \mu r_i(\|\Delta x_{i*}^{k+1}\|^2 + \|\Delta x_{i*}^k\|^2) \quad \forall\, i = 1, \ldots, m.\end{aligned}$$

Summing the above inequalities over $i = 1, \ldots, m$, we have

$$-(\Delta x_{**}^{k+1})^T A^T\Delta\lambda^k + \beta(\Delta x_{**}^{k+1})^T A^T A\Delta x^k + (\Delta x_{**}^{k+1})^T\bar{H}_x(\Delta x_*^{k+1} - \Delta x_*^k) \le \|\Delta x_*^{k+1}\|_{N_x}^2 + \|\Delta x_*^k\|_{N_x}^2.$$

It follows from the above inequality and $\Delta x_*^{k+1} - \Delta x_*^k = \Delta x_{**}^{k+1} - \Delta x^k$ that

$$\begin{aligned}&(\Delta x_{**}^{k+1})^T A^T\Delta\lambda^k\\ &\ge \beta(\Delta x_{**}^{k+1})^T A^T A\Delta x^k + (\Delta x_{**}^{k+1})^T\bar{H}_x(\Delta x_{**}^{k+1} - \Delta x^k) - (\|\Delta x_*^{k+1}\|_{N_x}^2 + \|\Delta x_*^k\|_{N_x}^2)\\ &= \|\Delta x_{**}^{k+1}\|_{\bar{H}_x}^2 - (\Delta x_{**}^{k+1})^T(\bar{H}_x - \beta A^T A)\Delta x^k - \|\Delta x_*^{k+1}\|_{N_x}^2 - \|\Delta x_*^k\|_{N_x}^2.\end{aligned} \tag{7.27}$$

Since $H = \mathrm{diag}(\bar{H}_x - \beta A^T A, \frac{1}{\gamma\beta}I_l) \succ 0$, we have $\bar{H}_x - \beta A^T A \succ 0$. Then, using the Cauchy-Schwarz inequality, we obtain

$$-2(\Delta x_{**}^{k+1})^T(\bar{H}_x - \beta A^T A)\Delta x^k \ge -\|\Delta x_{**}^{k+1}\|_{\bar{H}_x-\beta A^T A}^2 - \|\Delta x^k\|_{\bar{H}_x-\beta A^T A}^2. \tag{7.28}$$

Substituting (7.28) into (7.27), we get

$$\begin{aligned}&2(\Delta x_{**}^{k+1})^T A^T\Delta\lambda^k\\ &\ge 2\|\Delta x_{**}^{k+1}\|_{\bar{H}_x}^2 - \|\Delta x_{**}^{k+1}\|_{\bar{H}_x-\beta A^T A}^2 - \|\Delta x^k\|_{\bar{H}_x-\beta A^T A}^2 - 2\|\Delta x_*^{k+1}\|_{N_x}^2 - 2\|\Delta x_*^k\|_{N_x}^2\\ &= \|\Delta x_{**}^{k+1}\|_{\bar{H}_x+\beta A^T A}^2 - \|\Delta x^k\|_{\bar{H}_x-\beta A^T A}^2 - 2\|\Delta x_*^{k+1}\|_{N_x}^2 - 2\|\Delta x_*^k\|_{N_x}^2.\end{aligned}$$

By a simple manipulation, we get

$$\begin{aligned}&2(\Delta x^{k+1})^T A^T\Delta\lambda^k\\ &\ge \|\Delta x^{k+1}\|_{\bar{H}_x+\beta A^T A}^2 - \|\Delta x^k\|_{\bar{H}_x-\beta A^T A}^2 - 2\|\Delta x^{k+1}\|_{N_x}^2 - 2\|\Delta x^k\|_{N_x}^2\\ &\quad +2(\Delta x^{k+1} - \Delta x_{**}^{k+1})^T A^T\Delta\lambda^k + (\|\Delta x_{**}^{k+1}\|_{\bar{H}_x+\beta A^T A}^2 - \|\Delta x^{k+1}\|_{\bar{H}_x+\beta A^T A}^2)\\ &\quad +2(\|\Delta x^{k+1}\|_{N_x}^2 - \|\Delta x_*^{k+1}\|_{N_x}^2) + 2(\|\Delta x^k\|_{N_x}^2 - \|\Delta x_*^k\|_{N_x}^2).\end{aligned} \tag{7.29}$$

From the definitions of Δx^{k+1} and Δx_{**}^{k+1}, we get

$$\Delta x^{k+1} - \Delta x_{**}^{k+1} = (x^{k+1} - x^k) - (x_*^{k+1} - x_*^k) = (x^{k+1} - x_*^{k+1}) - (x^k - x_*^k).$$

Together with (6.2) and the fact that $\{w^k\}$ is bounded, there is a positive constant c such that

$$\begin{aligned}&2(\Delta x^{k+1} - \Delta x_{**}^{k+1})^T A^T\Delta\lambda^k\\ &= 2(x^{k+1} - x_*^{k+1})^T A^T(\lambda^k - \lambda^{k-1}) - 2(x^k - x_*^k)^T A^T(\lambda^k - \lambda^{k-1})\\ &\ge -2\|x^{k+1} - x_*^{k+1}\|\|A^T(\lambda^k - \lambda^{k-1})\| - 2\|x^k - x_*^k\|\|A^T(\lambda^k - \lambda^{k-1})\|\\ &\ge -c(\nu_k + \nu_{k-1}).\end{aligned}$$

Similarly, it's easy to prove that there is a positive constant c_3 such that

$$\begin{aligned}&2(\Delta x^{k+1} - \Delta x_{**}^{k+1})^T A^T\Delta\lambda^k + (\|\Delta x_{**}^{k+1}\|_{\bar{H}_x+\beta A^T A}^2 - \|\Delta x^{k+1}\|_{\bar{H}_x+\beta A^T A}^2)\\ &\quad +2(\|\Delta x^{k+1}\|_{N_x}^2 - \|\Delta x_*^{k+1}\|_{N_x}^2) + 2(\|\Delta x^k\|_{N_x}^2 - \|\Delta x_*^k\|_{N_x}^2)\\ &\ge -c_3(\nu_{k-1} + \nu_k).\end{aligned}$$

Note that $\Delta\lambda^{k+1} = \Delta\lambda^k - \gamma\beta A\Delta x^{k+1}$. It follows from (7.29) and the above formula that

$$\begin{aligned}&\frac{1}{\gamma\beta}\|\Delta\lambda^k\|^2 - \frac{1}{\gamma\beta}\|\Delta\lambda^{k+1}\|^2\\ &= 2(\Delta x^{k+1})^T A^T\Delta\lambda^k - \gamma\beta\|A\Delta x^{k+1}\|^2\\ &\geq \|\Delta x^{k+1}\|^2_{\bar{H}_x+(1-\gamma)\beta A^TA-2N_x} - \|\Delta x^k\|^2_{\bar{H}_x-\beta A^TA+2N_x} - c_3(\nu_{k-1}+\nu_k).\end{aligned}$$

Using this and the definition of $\bar{H}'_x$, we have

$$\begin{aligned}&(\|\Delta x^k\|^2_{\bar{H}'_x} + \frac{1}{\gamma\beta}\|\Delta\lambda^k\|^2) - (\|\Delta x^{k+1}\|^2_{\bar{H}'_x} + \frac{1}{\gamma\beta}\|\Delta\lambda^{k+1}\|^2)\\ &\geq \|\Delta x^{k+1}\|^2_{\bar{H}_x+(1-\gamma)\beta A^TA-2N_x} - \|\Delta x^{k+1}\|^2_{\bar{H}'_x} - c_3(\nu_{k-1}+\nu_k)\\ &= \|\Delta x^{k+1}\|^2_{(2-\gamma)\beta A^TA-4N_x} - c_3(\nu_{k-1}+\nu_k).\end{aligned} \tag{7.30}$$

From

$$0 < \mu < \frac{(2-\gamma)\beta\lambda_{\min}(A^TA)}{4\max_{i=1,\ldots,m}\{r_i\}},$$

it follows that $(2-\gamma)\beta A^TA - 4N_x \succ 0$. Using (7.30), we have

$$\|\Delta x^{k+1}\|^2_{\bar{H}'_x} + \frac{1}{\gamma\beta}\|\Delta\lambda^{k+1}\|^2 \leq \|\Delta x^k\|^2_{\bar{H}'_x} + \frac{1}{\gamma\beta}\|\Delta\lambda^k\|^2 + c_3(\nu_{k-1}+\nu_k).$$

Then, by the definition of $\bar{H}'$, we get the assertion (7.22). ■

Theorem 7.9. *Let $\{w^t\}$ be generated by Algorithm 2. Assume that A is a matrix of full column rank; $\beta > 0$; $\gamma \in (0, 2)$; $r_i > 0$, $i = 1, \ldots, m$ and $\mu \in (0, 1)$, where r_i and μ satisfy the conditions:*

$$r_i > (m-1)\beta\lambda_{\max}(A_i^TA_i)$$

and

$$0 < \mu < \min\Big\{\min_{i=1,\ldots,m}\{1 - \frac{(m-1)\beta\lambda_{\max}(A_i^TA_i)}{r_i}\}, \frac{(2-\gamma)\beta\lambda_{\min}(A^TA)}{4\max_{i=1,\ldots,m}\{r_i\}}\Big\}.$$

Then for any $w^* \in \Omega^*$ and any integer $t \geq 0$, we obtain

$$\|w^{t+1} - w^t\|^2_{\bar{H}'} \leq \frac{1}{t+1}\Big\{c_0[\|w^0 - w^*\|^2_H + \sum_{k=0}^{\infty}(2\rho C_{w^*}\nu_k + \rho^2\nu_k^2)] + c_3\sum_{k=1}^{\infty}k(\nu_{k-1}+\nu_k)\Big\}, \tag{7.31}$$

where H and $\bar{H}'$ are defined in (2.13) and (5.4), and the positive constants c_0, C_{w^*} and c_3 are defined in (5.16), Lemmas 7.5 and 7.8.

Proof. Using $M(w^k - \tilde{w}^k) = (w^k - w^{k+1})$ (see (4.3)), (5.16) and (7.13), we obtain

$$\begin{aligned}\sum_{k=0}^{\infty}\|w^k - w^{k+1}\|^2_{\bar{H}'} &\leq \sum_{k=0}^{\infty}c_0\|w^k - w^{k+1}\|^2_{M^{-T}\tilde{H}M^{-1}}\\ &\leq \sum_{k=0}^{\infty}c_0\|w^k - \tilde{w}^k\|^2_{\tilde{H}}\\ &\leq c_0\Big[\|w^0 - w^*\|^2_H + \sum_{k=0}^{\infty}(2\rho C_{w^*}\nu_k + \rho^2\nu_k^2)\Big].\end{aligned} \tag{7.32}$$

It follows from (7.22) that

$$\|w^{t+1} - w^t\|^2_{\bar{H}'} \leq \|w^{k+1} - w^k\|^2_{\bar{H}'} + c_3\sum_{j=k+1}^{t}(\nu_{j-1}+\nu_j) \quad \forall\, 0 < k \leq t.$$

And thus we have

$$\begin{aligned}(t+1)\|w^{t+1}-w^t\|_{\bar{H}'}^2 &\le \sum_{k=0}^{t}\|w^{k+1}-w^k\|_{\bar{H}'}^2 + c_3\sum_{k=0}^{t-1}\sum_{j=k+1}^{t}(\nu_{j-1}+\nu_j)\\ &\le \sum_{k=0}^{t}\|w^{k+1}-w^k\|_{\bar{H}'}^2 + c_3\sum_{k=1}^{t}k(\nu_{k-1}+\nu_k).\end{aligned}$$

From the above inequality and (7.32), we get the assertion (7.31). The proof is complete. ∎

It follows from Remark 6.1, (7.3) and (7.5) that if $w^{t+1}=w^t$, we have $w_*^{t+1}=w^t=\tilde{w}^t$ and w^t is the solution of $\mathrm{VI}(\Omega,F,\theta)$. Therefore, $\|w^{t+1}-w^t\|_{\bar{H}'}^2$ can be viewed as an error measurement in term of the distance to the solution set of $\mathrm{VI}(\Omega,F,\theta)$ for the t-th iteration of Algorithm 2. Notice that Ω^* is convex and closed. Let

$$d := c_0\inf\{\|w^0-w^*\|_H^2 \mid w^*\in\Omega^*\} + c_0\sum_{k=0}^{\infty}(2\rho C_{w^*}\nu_k+\rho^2\nu_k^2) + c_3\sum_{k=1}^{\infty}k(\nu_{k-1}+\nu_k).$$

If $\sum_{k=1}^{\infty}k\nu_k<\infty$, we have $d<+\infty$. Then, for any given $\varepsilon>0$, the inequality (7.31) shows that under the assumptions that A is a matrix of full column rank, (7.20), (7.21) and $\sum_{k=1}^{\infty}k\nu_k<\infty$, Algorithm 2 needs at most

$$\lceil\frac{d}{\varepsilon}-1\rceil$$

iterations to ensure that $\|w^{t+1}-w^t\|_{\bar{H}'}^2\le\varepsilon$.

7.4. A Worst-case $o(1/t)$ Convergence Rate in the Non-ergodic Sense

With Lemma 2.3, we can refine the result in Theorem 7.9 and prove a worst-case $o(1/t)$ convergence rate for Algorithm 2. The result is summarized in the following theorem.

Theorem 7.10. *Let $\{w^t\}$ be generated by Algorithm 2. Assume that A is a matrix of full column rank; $\beta>0$; $\gamma\in(0,2)$; $r_i>0$, $i=1,\ldots,m$ and $\mu\in(0,1)$, where r_i and μ satisfy the conditions:*

$$r_i > (m-1)\beta\lambda_{\max}(A_i^TA_i)$$

and

$$0<\mu<\min\Big\{\min_{i=1,\ldots,m}\{1-\frac{(m-1)\beta\lambda_{\max}(A_i^TA_i)}{r_i}\},\frac{(2-\gamma)\beta\lambda_{\min}(A^TA)}{4\max_{i=1,\ldots,m}\{r_i\}}\Big\}.$$

If $\sum_{t=1}^{\infty}t\nu_t<+\infty$, then we obtain

$$\|w^{t+1}-w^t\|_{\bar{H}'}^2 = o(1/t),$$

where $\bar{H}'$ is defined in (5.4).

Proof. From (7.32), we have

$$\sum_{t=0}^{\infty}\|w^{t+1}-w^t\|_{\bar{H}'}^2<+\infty.$$

On the other hand, Lemma 7.8 implies $\|w^{t+1}-w^t\|_{\bar{H}'}^2\le\|w^t-w^{t-1}\|_{\bar{H}'}^2+c_3(\nu_{t-1}+\nu_t)$. If $\sum_{t=1}^{\infty}t\nu_t<+\infty$, then we obtain $\sum_{t=1}^{\infty}t(\nu_{t-1}+\nu_t)<+\infty$. By Lemma 2.3, we have $\|w^{t+1}-w^t\|_{\bar{H}'}^2=o(1/t)$, which completes the proof. ∎

7.5. **Linear Convergence**

Similarly as Theorem 5.9, we have the following assertion but the proofs are omitted.

Theorem 7.11. *Let $\{w^k\}$ be generated by Algorithm 2 and $\{w_*^{k+1}\}$ be defined in (7.2). Assume that Assumptions 1-3 hold, and the matrix A has full row rank. There is a constant $\delta > 0$, such that*

$$\|w_*^{k+1} - w^*\|_H^2 \leq \frac{1}{1+\delta}\|w^k - w^*\|_H^2 \quad \forall w^* \in \Omega^*, \tag{7.33}$$

where H is defined in (2.13).

Theorem 7.12. *Let $\{w^k\}$ be generated by Algorithm 2. Assume that Assumptions 1-3 hold, the matrix A has full row rank, and*

$$0 \leq \nu_k \leq \frac{\delta}{2\rho\sqrt{(1+\delta)(2+\delta)}}\|w^k - w^*\|_H \quad \forall w^* \in \Omega^*,$$

where H, ρ and δ are defined by (2.13), (7.4) and (7.33). Then we have

$$\|w^{k+1} - w^*\|_H^2 \leq \frac{4+3\delta}{4(1+\delta)}\|w^k - w^*\|_H^2.$$

Proof. Using the Cauchy-Schwarz inequality, (7.33) and (7.4), we have

$$\begin{aligned}
\|w^{k+1} - w^*\|_H^2 &= \|(w_*^{k+1} - w^*) + (w^{k+1} - w_*^{k+1})\|_H^2 \\
&\leq (1+\frac{\delta}{2})\|w_*^{k+1} - w^*\|_H^2 + (\frac{2}{\delta}+1)\|w^{k+1} - w_*^{k+1}\|_H^2 \\
&\leq \frac{1+\frac{\delta}{2}}{1+\delta}\|w^k - w^*\|_H^2 + (\frac{2}{\delta}+1)\rho^2\nu_k^2.
\end{aligned}$$

If $0 \leq \nu_k \leq \frac{\delta}{2\rho\sqrt{(1+\delta)(2+\delta)}}\|w^k - w^*\|_H$, then from the above inequality we get

$$\begin{aligned}
\|w^{k+1} - w^*\|_H^2 &\leq \frac{2+\delta}{2(1+\delta)}\|w^k - w^*\|_H^2 + \frac{(2+\delta)\rho^2\nu_k^2}{\delta} \\
&\leq \Big(\frac{2+\delta}{2(1+\delta)} + \frac{\delta}{4(1+\delta)}\Big)\|w^k - w^*\|_H^2 \\
&= \frac{4+3\delta}{4(1+\delta)}\|w^k - w^*\|_H^2.
\end{aligned}$$

The proof is complete. ■

8. **Numerical Experiments**

In this section, we apply the proposed ALM with full Jacobian decomposition and LQP regularization to an allocation problem arising in market mechanisms (see, e.g., [5, 39, 45]) and report some preliminary numerical results to verify its efficiency. For succinctness, we only focus on the exact version (3.2) and do not test the inexact version (6.1). We wrote our code by Matlab R2015a and all experiments were conducted on a personal computer with an Intel Core i5-3210M CPU (2.50GHz) and 8.00 GB of RAM.

We consider an economic system in which n resources are allocated by using m technological activities. The goal is to minimize the sum of cost functions of all the activities, denoted by θ_i $(i = 1, \ldots, m)$. The amount of each resource is denoted by $b_j \geq 0$ $(j = 1, \ldots, n)$, which justifies the nonnegativity and budget constraints. As in [5], the allocation problem can be modeled as

$$\min\left\{\sum_{i=1}^{m}\theta_i(x_i)\;\middle|\;\sum_{i=1}^{m}x_i=b,\ x_i\in\Re^n_+,\ i=1,\ldots,m\right\},\tag{8.1}$$

where $\theta_i:\Re^n\to\Re$ is the cost function of the i-th activity and the vector $b=(b_1,\ldots,b_n)\in\Re^n$ represents all the resources. The model (8.1) is a special case of (1.1) with $A_i:=I_n$ for $i=1,\ldots,m$.

To specify the cost functions in (8.1), we choose some stencil functions in Table 10.2 in [8] and list them in Table 8.1. For such a function $\phi(s)$, note that its logarithmic-quadratic proximity mapping

Table 8.1. The stencil function $\phi(s)$ for generating θ_i in (8.1).

No.	$\phi(s):\Re\to(-\infty,+\infty]$	$\text{argmin}_{s\geq 0}\phi(s)+\frac{\tau}{2}\|s-x\|^2-\eta\log s$
ii	$\begin{cases}\overline{\omega}s & \text{if } s\geq 0\\ \underline{\omega}s & \text{otherwise}\end{cases}$	$p=\left(\tau x-\overline{\omega}+\sqrt{(\tau x-\overline{\omega})^2+4\tau\eta}\right)/(2\tau)$
v	$\kappa\|s\|^q$	$p>0$, such that $q\kappa p^q+\tau p^2-\tau xp-\eta=0$
vii	$\omega\|s\|+\tilde{\tau}\|s\|^2+\kappa\|s\|^q$	$p>0$, such that $q\kappa p^q+(2\tilde{\tau}+\tau)p^2+(\omega-\tau x)p-\eta=0$
ix	$\begin{cases}\omega s & \text{if } s\geq 0\\ +\infty & \text{otherwise}\end{cases}$	$p=\left(\tau x-\omega+\sqrt{(\tau x-\omega)^2+4\tau\eta}\right)/(2\tau)$
x	$\begin{cases}-\omega s^{1/q} & \text{if } s\geq 0\\ +\infty & \text{otherwise}\end{cases}$	p^q, where $p>0$ and $\tau p^{2q}-\tau xp^q-\omega q^{-1}p-\eta=0$
xi	$\begin{cases}\omega s^{-q} & \text{if } s>0\\ +\infty & \text{otherwise}\end{cases}$	$p>0$, such that $\tau p^{q+2}-\tau xp^{q+1}-\eta p^q-\omega q=0$
xiv	$\begin{cases}-\kappa\log s+\tilde{\tau}s^2/2+\alpha s & \text{if } s>0\\ +\infty & \text{otherwise}\end{cases}$	$p=(\tau x-\alpha+\sqrt{(\tau x-\alpha)^2+4(\tau+\tilde{\tau})(\kappa+\eta)})/2(\tau+\tilde{\tau})$
xv	$\begin{cases}-\kappa\log s+\alpha s+\omega s^{-1} & \text{if } s>0\\ +\infty & \text{otherwise}\end{cases}$	$p>0$, such that $\tau p^3+(\alpha-\tau x)p^2-(\kappa+\eta)p-\omega=0$
xvi	$\begin{cases}-\kappa\log s+\omega s^q & \text{if } s>0\\ +\infty & \text{otherwise}\end{cases}$	$p>0$, such that $\omega qp^q+\tau p^2-\tau xp-(\kappa+\eta)=0$
xvii	$\begin{cases}-\underline{\kappa}\log(s-\underline{\omega})-\overline{\kappa}\log(\overline{\omega}-s) & \text{if } s\in(\underline{\omega},\overline{\omega})\\ +\infty & \text{otherwise}\end{cases}$	$p>0$, such that $\tau p^4-(x+\overline{\omega}+\underline{\omega})\tau p^3-(\overline{\kappa}+\underline{\kappa}+\eta-x\overline{\omega}\tau$ $-\underline{\omega}\overline{\omega}\tau-x\underline{\omega}\tau)p^2+(\underline{\kappa}\overline{\omega}+\overline{\kappa}\underline{\omega}+\eta\overline{\omega}+\eta\underline{\omega}$ $-x\tau\overline{\omega}\underline{\omega})p-\eta\overline{\omega}\underline{\omega}=0$

Table 8.2. Parameters in $\phi(s)$.

ii	v, vii, ix	x	xi	xiv, xv, xvi	xvii
$\overline{\omega}\sim U(10^2,10^3)$ $\underline{\omega}\sim U(-10^2,-10^3)$	$\omega\sim U(1,5)$ $\kappa\sim U(1,5)$ $\tilde{\tau}\sim U(1,5)$ $q\sim U(1,5)$	$\omega\sim U(-5,-1)$ $q\sim U(1,5)$	$\omega\sim U(10^7,10^8)$ $q\sim U(1,5)$	$\omega\sim U(1,5)$ $\kappa\sim U(1,5)$ $\tilde{\tau}\sim U(1,5)$ $q\sim U(1,5)$ $\alpha\sim U(1,5)$	$\overline{\kappa}\sim U(10^{-3},10^{-1})$ $\underline{\kappa}\sim U(10^{-3},10^{-1})$ $\overline{\omega}\sim U(10^3,10^4)$ $\underline{\omega}\sim U(-10^4,-10^1)$

$$\underset{s\in\Re}{\text{argmin}}\{\phi(s)+\frac{\tau}{2}\|s-x\|^2-\eta\log s\}$$

with $\eta > 0$ has a closed-form solution or can be efficiently computed by solving certain polynomial equations. Furthermore, for a one-dimensional stencil function $\phi(s)$ in Table 8.1, it can be easily extended to an n-dimensional function $\Phi(\mathbf{s})$ whose proximity function can also be easily computed. Let us take the stencil function $\phi(s)$ listed as Item (ii) in Table 8.1 (ii) as an illustrative example. Based on this $\phi(s)$, we can define

$$\Phi(\mathbf{s}) := \sum_{i=1}^{n} \phi(s_i) = \begin{cases} \sum_{i=1}^{n} \overline{\omega}_i s_i & \text{if } s_i \geq 0, \\ \sum_{i=1}^{n} \underline{\omega}_i s_i & \text{otherwise,} \end{cases} \tag{8.2}$$

where $\mathbf{s} = (s_1, \ldots, s_n)$, $\overline{\omega} = (\overline{\omega}_1, \ldots, \overline{\omega}_n)$ and $\underline{\omega} = (\underline{\omega}_1, \ldots, \underline{\omega}_n)$ are vectors in $\Re^n$. Particularly, if $\overline{\omega} = -\underline{\omega} = \mathbf{1}$, we have $\Phi(\mathbf{s}) = \|\mathbf{s}\|_1$ which corresponds to the standard l_1-norm; otherwise, we have $\Phi(\mathbf{s}) = \|W\mathbf{s}\|_1$ which corresponds to the weighted l_1-norm with $W = \mathrm{diag}(\overline{\omega})$. Similarly, we can extend all the other functions listed in Table 8.1 to n-dimensional functions for the use of the cost functions $\{\theta_i\}_{i=1}^{10}$ in the model (8.1). The parametric vectors, e.g., $\overline{\omega}$, $\underline{\omega}$, κ, etc, are chosen randomly by following some uniform distributions (see the right-column of Table 8.1 with $U(a, b)$ representing uniform distribution in the interval $[a, b]$). The resource amount vector b in the model (8.1) is set as $b := n\mathbf{1}$.

For simplicity, if $n = 1$, the i-th subproblem is equivalent to solving

$$\min_{x_i \in \Re^n_+} \Big\{\theta_i(x_i) + \frac{\beta + r_i}{2}\Big[x_i - \frac{(1-\mu)r_i x_i^k + \lambda^k - \beta(\sum_{j=1, j\neq i}^{m} x_j^k - b)}{\beta + r_i}\Big] - \mu r_i (x_i^k)^2 \log x_i\Big\}. \tag{8.3}$$

This is just the problem

$$\min_{x_i \in \Re^n_+} \; \{\phi(x_i) + \frac{\tau}{2}\|x_i - x\|^2 - \eta \log x_i\},$$

where

$$\phi := \theta_i, \quad \tau := \beta + r_i, \quad x := \frac{(1-\mu)r_i x_i^k + \lambda^k - \beta(\sum_{j=1, j\neq i}^{m} x_j^k - b)}{\beta + r_i}$$

and

$$\eta := \mu r_i (x_i^k)^2.$$

It follows immediately from [11] that solving (1.1) is equivalent to finding a zero point of

$$e(w) := \begin{pmatrix} e_{x_1}(w) \\ \vdots \\ e_{x_m}(w) \\ e_\lambda(w) \end{pmatrix} = \begin{pmatrix} x_1 - P_{\Re^{n_1}_+}[x_1 - (g_1(x) - A_1^T\lambda)] \\ \vdots \\ x_m - P_{\Re^{n_m}_+}[x_m - (g_m(y) - A_m^T\lambda)] \\ \sum_{i=1}^{m} A_i x_i - b \end{pmatrix}, \tag{8.4}$$

where $g_i(x_i) \in \partial\theta_i(x_i)$, $i = 1, \ldots, m$, and $P_{\mathcal{V}}[v]$ denotes the projection of v onto $\mathcal{V}$ in the Euclidean norm. Therefore, we can use $\|e(w^k)\|/\|e(w^0)\|$ to evaluate the quality of the iterate w^k.

To see the efficiency of Algorithm 1 for solving the allocation problem (8.1), we compare it with the alternating direction method with Gaussian back substitution ("ADM-G" for short) proposed in [24] and the direct extension of ADMM (1.6) ("EADMM" for short). Note that the ADM-G is a competitive algorithm in the category of ADMM-based prediction-correction methods and its efficiency and stability have been well verified in the literature; and the EADMM usually performs very well despite it lacks of convergence. Thus, we choose these two algorithms to compare.

For the involved parameters of these iterative schemes, we chose $\beta = 1$ and $\alpha = 1$ for the ADM-G; $\beta = 1$ for the EADMM; $r = m/100$, $\mu = 0.1$, $\beta = 0.9(1-\mu)r/(m-1)$ and $\gamma = 1.9$ for Algorithm 1. Since $A_i = I$ for all i's, and for Algorithm 1, β should satisfy

$$0 < \beta < \frac{(1-\mu)r}{(m-1)\lambda_{\max}(A_i^T A_i)} \qquad \forall\, i = 1, \ldots, m,$$

we chose $\beta = 0.9(1-\mu)r/(m-1)$ in the numerical experiments. All initial iterates are chosen as **1**. Note that the subproblems of Algorithm 1 in (3.3) can be solved in parallel. But for comparison with ADM-G and EADMM whose subproblems can only be solved sequentially, we count the accumulated time for solving all the subproblems for Algorithm 1.

In Figures 8.1 - 8.4, we plot the evolutions of the objective function values (“Obj-Fun-Val” for short) and the values of $\|e(w^k)\|/\|e(w^0)\|$ (“|ew|/|ew0|-Val” for short) with respect to the computing time and iteration numbers for the cases of n where $n = 100, 1000, 2000, 3000, 5000$ and 8000. The plots in Figures 8.1 - 8.4 show that among the three methods under comparison, the proposed Algorithm 1 performs far better than the others.

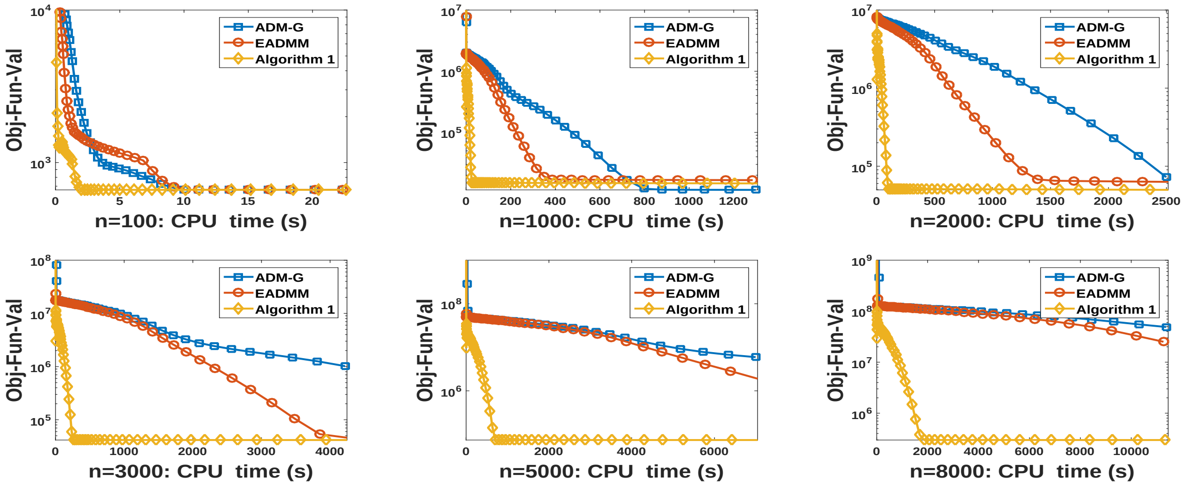

FIGURE 8.1. Evolutions of objective function values w.r.t. computing time for variant n's.

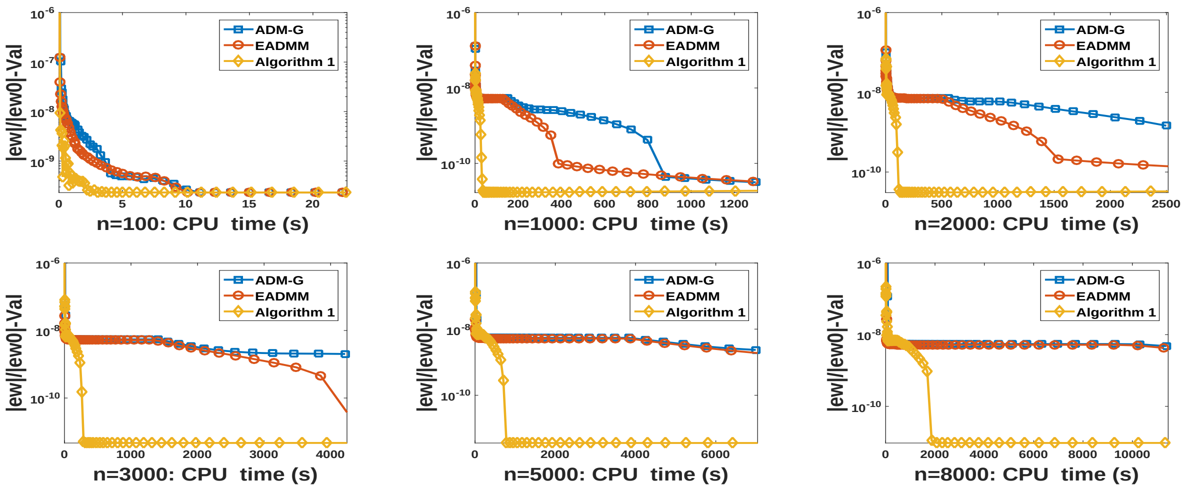

FIGURE 8.2. Evolutions of $\|e(w)\|/\|e(w^0)\|$ w.r.t. computing time for variant n's.

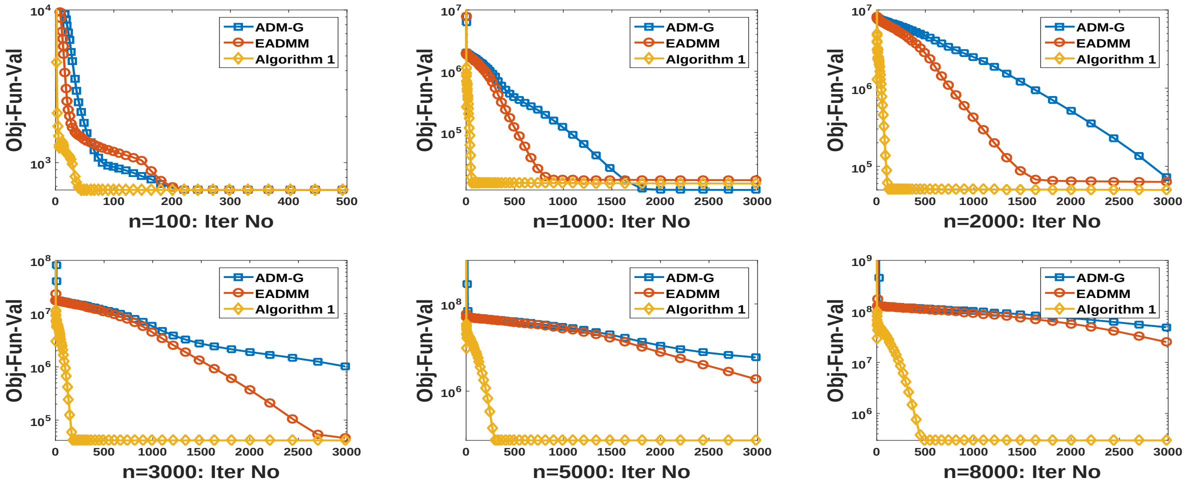

Figure 8.3. Evolutions of objective function values w.r.t. iteration No. for variant n's.

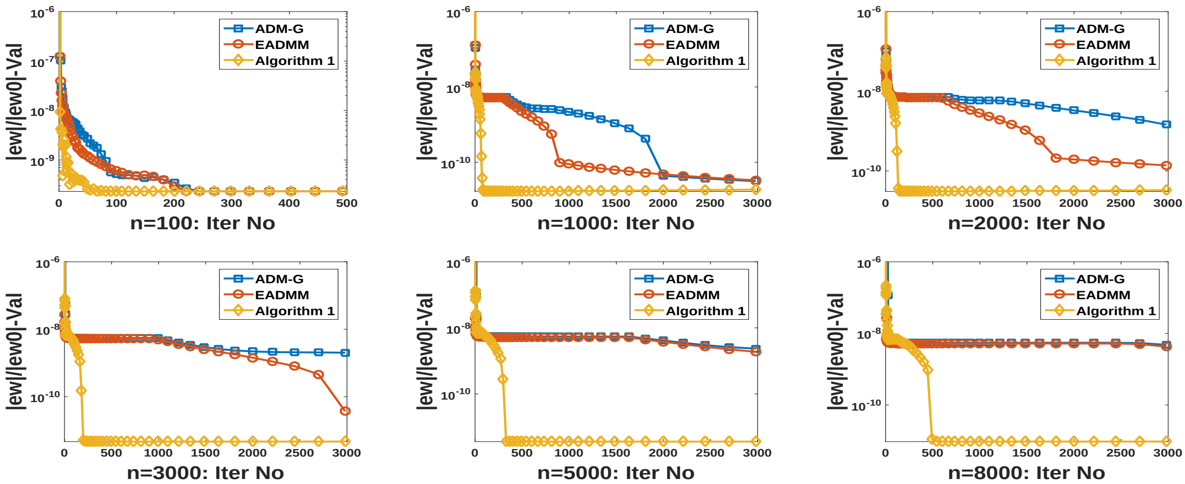

Figure 8.4. Evolutions of $\|e(w)\|/\|e(w^0)\|$ w.r.t. iteration No. for variant n's.

9. Conclusions

We consider a separable convex minimization model whose variables are coupled by linear constraints and they are subject to the positive orthant constraints, and its objective function is in form of m functions without coupled variables. We suggest applying the Jacobian decomposition to the subproblems obtained by the augmented Lagrangian method (ALM) at each iteration and regularizing the decomposed subproblems by the logarithmic-quadratic proximal (LQP) terms. The ALM with full Jacobian decomposition and LQP regularization is thus proposed for the generic case of the model under consideration with $m \geq 3$. The new scheme only requires solving some unconstrained subproblems at each iteration and these subproblems are eligible for fully parallel computation. The new scheme can be regarded as a further development of some existing work in combination of the LQP with Jacobian decomposition of the ALM for the general case of $m \geq 3$ which uses the standard quadratic proximal terms for regularization; or the extension of existing work in combination of LQP and Gauss-Seidel decomposition of the ALM for $m = 2$ to the generic case of $m \geq 3$. We analyze both the exact and inexact versions for the new scheme, and comprehensively investigate their convergence. The global convergence, worst-case convergence rates measured by the iteration-complexity and the linear convergence rates under additional assumptions are all derived. Note that the LQP term tends to be the regular quadratic proximal term when $\mu \to 0$, meaning the proposed Algorithm 1 asymptotically tends to the proximal version of the Jacobian decomposition of ALM (1.8) in [25] when $\mu \to 0$. Meanwhile, our convergence analysis for Algorithms 1 and 2 holds for any fixed value of μ. Therefore, our convergence analysis also implies that the same convergence rates can be derived for the proximal version of the Jacobian decomposition of ALM (1.8) and its inexact version.

References

[1] A. Auslender and M. Teboulle. Entropic proximal decomposition methods for convex programs and variational inequalities. *Math. Program.*, 91:33–47, 2001.

[2] A. Auslender and M. Teboulle. Interior projection-like methods for monotone variational inequalities. *Math. Program.*, 104:39–68, 2005.

[3] A. Auslender, M. Teboulle, and S. Ben-Tiba. A logarithmic-quadratic proximal method for variational inequalities. *Comput. Optim. Appl.*, 12:31–40, 1999.

[4] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, New York, 1982.

[5] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3:1–122, 2010.

[6] C. H. Chen, B. S. He, Ye Y. Y., and X. M. Yuan. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. *Math. Program., Ser A*, 155:57–79, 2016.

[7] C. H. Chen, M. Li, and X. M. Yuan. Further study on the convergence rate of alternating direction method of multipliers with logarithmic-quadratic proximal regularization. *J. Optim. Theory Appli.*, 166:906–929, 2015.

[8] P. L. Combettes and J. C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer, 2011.

[9] D. Davis and W. T. Yin. Convergence rate analysis of several splitting schemes. In R. Glowinski, S. J. Osher, and W. T. Yin, editors, *Splitting Methods in Communication, Imaging, Science, and Engineering*, pages 115–163. Springer, 2017.

[10] W. Deng, M. J. Lai, Z. M. Peng, and W. T. Yin. Parallel multi-block ADMM with $o(1/k)$ convergence. *J. Sci. Comput.*, 71:712–736, 2017.

[11] B. C. Eaves. On the basic theorem of complementarity. *Math. Program.*, 1:68–75, 1971.

[12] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.*, 55:293–318, 1992.

[13] J. Eckstein and W. Yao. Augmented Lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results. RUTCOR Research Report RRR 32-2012, 2012.

[14] F. Facchinei and J. S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems. Vol. I. Springer Series in Operations Research.* Springer-Verlag, New York, 2003.

[15] R. Glowinski. *Numerical Methods for Nonlinear Variational Problems.* Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.

[16] R. Glowinski. On alternating direction methods of multipliers: A historical perspective. In *Modeling, Simulation and Optimization for Science and Technology, Volume 34 of the series Computational Methods in Applied Sciences*, pages 59–82. Springer, 2014.

[17] R. Glowinski and A. Marrocco. Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problèmes non linéaires. *R.A.I.R.O.*, R2:41–76, 1975.

[18] R. Glowinski and P. Le Tallec. *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*. SIAM Studies in Applied Mathematics, Philadelphia, PA, 1989.

[19] E. G. Gol'shtein and N. V. Tret'yakov. Modified Lagrangian in convex programming and their generalizations. *Math. Program. Studies*, 10:86–97, 1979.

[20] D. R. Han, X. M. Yuan, and W. X. Zhang. An augmented-Lagrangian-based parallel splitting method for separable convex programming with applications to image processing. *Math. Comput.*, 83:2263–2291, 2014.

[21] P. C. Hansen, J. G. Nagy, and D. P. O'Leary. *Deblurring Images: Matrices, Spectra, and Filtering*. SIAM, Philadelphia, 2006.

[22] B. S. He, L. S. Hou, and X. M. Yuan. On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming. *SIAM J. Optim.*, 25:2274–2312, 2015.

[23] B. S. He, H. Liu, Wang Z. R., and X. M. Yuan. A strictly contractive Peaceman-Rachford splitting method for convex programming. *SIAM J. Optim.*, 24:1101–1140, 2014.

[24] B. S. He, M. Tao, and X. M. Yuan. Alternating direction method with Gaussian back substitution for separable convex programming. *SIAM J. Optim.*, 22:313–340, 2012.

[25] B. S. He, H. K. Xu, and X. M. Yuan. On the proximal Jacobian decomposition of ALM for multiple-block separable convex minimization problems and its relationship to ADMM. *J. Sci. Comput.*, 66:1204–1217, 2016.

[26] B. S. He and X. M. Yuan. On the $O(1/n)$ convergence rate of Douglas-Rachford alternating direction method. *SIAM J. Numer. Anal.*, 50:700–709, 2012.

[27] B. S. He and X. M. Yuan. On nonergodic convergence rate of Douglas-Rachford alternating direction method of multipliers. *Numer. Math.*, 130:567–577, 2015.

[28] M. Hong and Z. Q. Luo. On the linear convergence of the alternating direction method of multipliers. *Math. Program.*, 162:165–199, 2017.

[29] M. Li, L.-Z. Liao, and X. M. Yuan. Inexact alternating direction method of multipliers with logarithmic-quadratic proximal regularization. *J. Optim. Theory Appli.*, 159:412–436, 2013.

[30] M. Li, D. F. Sun, and K.-C. Toh. A majorized admm with indefinite proximal terms for linearly constrained convex composite optimization. *SIAM J. Optim.*, 26:922–950, 2016.

[31] M. Li and X. M. Yuan. A strictly contractive Peaceman-Rachford splitting method with logarithmic-quadratic proximal regularization for convex programming. *Math. Oper. Res.*, 40:842–858, 2015.

[32] T. Y. Lin, S. Q. Ma, and S. Z. Zhang. On the global linear convergence of the ADMM with multiblock variables. *SIAM J. Optim.*, 25:1478–1497, 2015.

[33] T. Y. Lin, S. Q. Ma, and S. Z. Zhang. On the sublinear convergence rate of multi-block ADMM. *J. Oper. Res. Soc. China*, 3:251–274, 2015.

[34] B. Martinet. Regularization d'inequations variationelles par approximations successives. *Revue Francaise d'Informatique et de Recherche Opérationelle*, 4:154–159, 1970.

[35] J. G. Melo and R. D. C. Monteiro. Iteration-complexity of a Jacobi-type non-Euclidean ADMM for multi-block linearly constrained nonconvex programs. arXiv:1705.07229v1, 2017.

[36] A. S. Nemirovsky and D. B. Yudin. *Problem Complexity and Method Efficiency in Optimization, Wiley-Interscience Series in Discrete Mathematics*. John Wiley & Sons, New York, 1983.

[37] Y. E. Nesterov. A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$. *Doklady AN SSSR*, 269:543–547, 1983.

[38] Y. E. Nesterov. Gradient methods for minimizing composite objective function. *Math. Program., Ser. B*, 140:125–161, 2013.

[39] M. Patriksson. A survey on the continuous nonlinear resource allocation problem. *European J. Oper. Res.*, 185:1–46, 2008.

[40] Y. G. Peng, A. Ganesh, J. Wright, W. L. Xu, and Y. Ma. Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Tran. Pattern Anal. Mach. Intel.*, 34:2233–2246, 2012.

[41] M. J. D. Powell. A method for nonlinear constraints in minimization problems. In R. Fletcher, editor, *Optimization*, pages 283–298. Academic Press, New York, 1969.

[42] R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.*, 1:97–116, 1976.

[43] M. Tao and X. M. Yuan. Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM J. Optim.*, 21:57–81, 2011.

[44] M. Tao and X. M. Yuan. On the $O(1/t)$ convergence rate of alternating direction method with logarithmic-quadratic proximal regularization. *SIAM J. Optim.*, 22:1431–1448, 2012.

[45] H. Uzawa. Market mechanisms and mathematical programming. *Econometrica*, 28:872–881, 1960.

[46] X. M. Yuan and M. Li. An LQP-based decomposition method for solving a class of variational inequalities. *SIAM J. Optim*, 21:1309–1318, 2011.