A study on augmented Lagrangian-based splitting methods for separable convex programming

Wang, Kai

2017

Wang, K. (2017). A study on augmented Lagrangian-based splitting methods for separable convex programming. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/73066>

<https://doi.org/10.32657/10356/73066>

A STUDY ON AUGMENTED
LAGRANGIAN-BASED SPLITTING METHODS FOR
SEPARABLE CONVEX PROGRAMMING

NANYANG
TECHNOLOGICAL
UNIVERSITY

PhD Thesis

Wang Kai

School of Mechanical and Aerospace Engineering

Nanyang Technological University

Acknowledgments

First and foremost, I would like to express my special appreciation and deepest gratitude to my supervisor, Prof. Jitamitra Desai, who has been patient and meticulous in tutoring me during the past four years. Without his guidance and persistent help, this dissertation would not have been possible. Besides professional skills and theoretical knowledge, a rigorous and demanding attitude towards scientific research is the most precious treasure that I have learnt from him, and I hope to carry this with me throughout my career.

Second, I am deeply indebted to Prof. Deren Han and Dr. Hongjin He, who encouraged me and led me towards the area of interdisciplinary studies. I would also like to thank Dr. Xihong Yan, Xingju Cai, Lingling Xu, Xiaofeng Nie, and Wenxing Zhang, for their useful suggestions during the course of my doctoral studies.

Last but not the least, I am eternally grateful to my parents for their dedicated love and encouragement. Their constant confidence and presence have stood me in good stead, now and always.

Contents

Acknowledgments	i
Abstract	v
List of Tables	vii
List of Abbreviations	viii
1 Introduction	1
1.1 Preliminaries	2
1.1.1 Basic Definitions	2
1.1.2 Equivalent Form	5
1.1.3 An ϵ -optimal Solution of MVI	8
1.2 Some Applications	8
1.3 Augmented Lagrangian Methods	11
1.4 Research Scope and Objectives	14
1.5 Outline	21
2 Literature Review	22
2.1 ALM-based Gauss-Seidel Splitting Methods	23
2.2 ALM-based Jacobian Splitting Methods	31
2.3 ALM-based Hybrid Splitting methods	33
2.4 Proximal Point Algorithms	36
3 Convergence of Augmented Lagrangian-based Parallel Splitting Method	40
3.1 Motivation	40
3.2 Algorithm	40
3.3 Convergence Analysis	41
3.4 Conclusions	48

4	On the Convergence Rate of Flexible Augmented Lagrangian-based Parallel Splitting Method	49
4.1	Motivation	49
4.2	Algorithm	49
4.3	Sublinear Convergence Rate	50
4.3.1	Convergence Rate under Ergodic Sense	51
4.3.2	Convergence Rate under Non-ergodic Sense	57
4.4	Linear Convergence	63
4.5	Conclusions	67
5	Proximal Partially Parallel Splitting Algorithm (PPSA)	68
5.1	Motivation	68
5.2	Algorithm	69
5.2.1	Main Framework of Algorithm	69
5.2.2	Relationships to Existing Methods	71
5.3	Convergence Analysis	73
5.4	Convergence Rate	83
5.4.1	Convergence Rate under Ergodic Sense	83
5.4.2	Convergence Rate under Nonergodic Sense	89
5.5	Numerical Results	95
5.5.1	Robust PCA Problem	95
5.5.2	Linear Equation Example	99
5.6	Conclusions	102
6	Linearized Partially-Parallel Splitting Methods	104
6.1	Motivation	104
6.2	A Set of Linearized PPSA-based Methods and Related Properties	105
6.2.1	Linearized version A of PPSA (LPSM-A)	105
6.2.2	Linearized Version B of PPSA (LPSM-B)	112
6.2.3	Linearized Version C of PPSA (LPSM-C)	117
6.3	Convergence Rate under Ergodic Sense	122
6.4	Numerical Results	125
6.5	Conclusions	129

7	A primal-dual prediction-correction algorithm	130
7.1	Motivation	130
7.2	Algorithm	131
7.3	Convergence results	132
7.4	Conclusions	139
8	Conclusions and Future Work	141
8.1	Conclusions	141
8.2	Future Work	142
	References	143

Abstract

Convex programming has played an important role in studying a wide class of applications arising from computer science, statistics, industrial engineering, and management. With the development of interdisciplinary research, an ever-increasing set of problems have been characterized as separable convex optimization problems. Moreover, the advent of big-data analytics has resulted in very large-scale structural convex programming problems, thereby necessitating the research towards devising fast numerical algorithms for solving such problems.

In this thesis, we consider the linearly constrained separable convex minimization problem, whose objective function consists of the sum of m individual convex functions in the absence of any coupling variables. While augmented Lagrangian-based decomposition methods have been well developed in the literature for solving such problems, a noteworthy requirement of these methods is that an additional correction step is a must to guarantee their convergence. In our first contribution, we show that a straightforward full Jacobian decomposition of the augmented Lagrangian method is *globally convergent* without additional strategies if the involved functions are further assumed to be *strongly convex*. Hence, our result serves as a supplement to ALM-based splitting methods.

Then, we extend our first research contribution to propose a flexible augmented Lagrangian-based parallel splitting method and derive the worst-case $O(1/t)$ convergence rate of this method under both ergodic and nonergodic conditions, where t represents the number of iterations. Furthermore, we also demonstrate that the convergence rate can be improved from $O(1/t)$ to $o(1/t)$, and finally, we show that this method can achieve global *linear convergence*, when the involved functions satisfy some additional conditions.

In our third contribution, we put forward a *proximal partially-parallel splitting algorithm* (**PPSA**) for solving convex minimization problems, where the objective function is

separable into m individual operators without any coupled variables, and the structural constraint set comprises only linear functions. At each iteration of this algorithm, one selected subproblem is solved, and subsequently the remaining subproblems are solved in parallel, utilizing the new iterate information. Hence, the proposed algorithm is a *hybrid mechanism* that combines the nice features of parallel decomposition methods and alternating direction methods, while simultaneously adopting the predictor-corrector strategy to ensure convergence of the algorithm. Our algorithmic framework is also amenable to admitting linearized versions of the subproblems, which frequently have closed-form solutions, thereby making the proposed method more implementable in practice. Furthermore, the worst-case $O(1/t)$ convergence rate of the proposed method is obtained under both ergodic and nonergodic conditions. The efficiency of this type algorithm is also verified by solving several instances of the robust PCA problem.

Next, continuing in the same vein as the third research work, we present a portfolio of *linearized versions* of the PPSA, which fully utilize the separable structure and result in easy subproblems that have closed-form solutions, or are relatively easy to solve as compared to their original nonlinear versions. Global convergence of these linearized methods under the projection contraction algorithmic framework is proven, and furthermore, detailed remarks that serve to clarify the interconnections between these linearized variants are highlighted. Finally, the worst-case $O(1/t)$ convergence rate under ergodic conditions is also established.

Lastly, we introduce a new primal-dual prediction-correction algorithm for solving a saddle point optimization problem, which serves as a bridge between the algorithms proposed in [12] and [77]. An interesting byproduct of the proposed method is that we obtain an easily implementable projection-based primal-dual algorithm, when the primal and dual variables belong to simple convex sets. Moreover, we establish the worst-case $O(1/t)$ convergence rate result in an ergodic sense, where t represents the number of iterations.

List of Tables

5.1	Numerical comparisons based on synthetically generated data.	99
5.2	Performance of Algorithm 5.1 with different relaxation factors γ	101
6.1	Numerical comparisons based on synthetically generated data.	128

List of Abbreviations

VI	Variational Inequality
VIP	Variational Inequality Problems
MVI	Mixed Variational Inequality
GNEP	Generalized Nash Equilibrium Problems
QVI	Quasi Variational Inequality
ALM	Augmented Lagrangian Methods
ADMM	Alternating Direction Methods of Multipliers
EADM	Extended Alternating Direction Methods of Multipliers
PPA	Proximal Point Algorithm
FOPDA	First-Order Primal-Dual Algorithm
PPSA	Proximal Partially Parallel Splitting Algorithm
RPCA	Robust Principal Component Analysis

Chapter 1

Introduction

With the rapid rise of interdisciplinary research, and our increasing ability to cast problems arising in diverse disciplines as either convex programs or variational inequality problems, optimization continues to play an ever expanding and important role in determining the solutions to such problems.

Consider the *multi-block convex minimization problem with linear constraints*, whose objective function is separable into the sum of m individual convex functions:

$$\min_{\{x_1, \dots, x_m\}} \left\{ \sum_{i=1}^m \theta_i(x_i) \mid \sum_{i=1}^m A_i x_i = b, x_i \in \mathcal{X}_i, i = 1, 2, \dots, m \right\}. \quad (1.1)$$

Here, $\theta_i(\cdot) : \mathbb{R}^{n_i} \rightarrow \mathbb{R}$ are closed proper convex functions (not necessarily smooth); $A_i \in \mathbb{R}^{l \times n_i}$ and $b \in \mathbb{R}^l$ are given matrices and a vector of appropriate dimensions, respectively; $\mathcal{X}_i \subseteq \mathbb{R}^{n_i}$ are closed and convex nonempty sets and $\sum_{i=1}^m n_i = n$. Throughout this thesis, we assume that the solution set of (1.1) is nonempty and denote $\mathbf{x} := (x_1, x_2, \dots, x_m)$ and $\mathcal{X} := \prod_{i=1}^m \mathcal{X}_i$, for notational simplicity. (Note that, although this thesis focuses on the case of convex programming with vector variables, our analysis can be easily extended to the case involving matrix variables as well.) In recent years, the multi-block convex minimization problem (1.1) has received considerable attention in the optimization and statistics communities due to its widespread applications in the areas of signal/image processing [4, 9, 110, 118, 135], statistical learning [10, 121, 125], network (transportation) optimization [27, 33, 104], amongst others.

1.1 Preliminaries

In this section, we begin by summarizing some basic notations, related definitions, and well-known results that serve to facilitate our subsequent analysis.

1.1.1 Basic Definitions

Let $\mathbb{R}^n$ be an n -dimensional Euclidean space. The symbol $^\top$ represents the transpose. For any two vectors $a, b \in \mathbb{R}^n$, $\langle a, b \rangle = a^\top b$ denotes the standard inner product. Furthermore, for a symmetric, positive definite matrix H (denoted as $H \succ 0$, or $H \in \mathcal{S}_{++}^n$), $\|u\|_H = \sqrt{\langle u, Hu \rangle} \equiv \sqrt{u^\top H u}$ represents the H -norm, and we denote $\|\cdot\|$ to be the standard Euclidean norm (obtained by setting $H \equiv I$, the identity matrix). Finally, let Ω be a nonempty closed convex subset of $\mathbb{R}^n$.

Definition 1.1 *For a given vector $x \in \mathbb{R}^n$, the projection of x onto the set Ω , is defined as the nearest vector $y \in \Omega$ to x , i.e.,*

$$P_\Omega[x] := \arg \min_{y \in \Omega} \|x - y\|.$$

Furthermore, the projection under the Q -norm, denoted by $P_{\Omega, Q}[\cdot]$, is defined as

$$P_{\Omega, Q}[x] := \arg \min_{y \in \Omega} \|x - y\|_Q,$$

where $Q \succ 0$. (Clearly, if $Q = I$, then $P_{\Omega, Q}[x] := P_\Omega[x]$).

Since Ω is convex and closed, in general, the projection operator can be difficult to compute in some cases as we would have to solve an underlying quadratic minimization problem. However, determining the projection on a set Ω is relatively easy to obtain in many practical settings, with an explicit characterization being readily available in the following cases:

- (1) If Ω is a nonnegative orthant, i.e., $\Omega = \{x \in \mathbb{R}^n | x \geq 0\}$, then

$$(P_\Omega[x])_i = \max\{x_i, 0\}.$$

- (2) If Ω is a box constraint, i.e., $\Omega = \{x \in \mathbb{R}^n | l \leq x \leq u\}$, then

$$(P_\Omega[x])_i = \max\{l_i, \min\{x_i, u_i\}\}.$$

(3) If Ω is a ball, i.e., $\Omega = \{x \in \mathbb{R}^n \mid \|x - x_0\| \leq r\}$, then

$$P_\Omega[x] = x_0 + \min\{1, \frac{r}{\|x - x_0\|}\}(x - x_0).$$

(4) If Ω is a positive semidefinite cone, i.e., $\Omega = \{X \in \mathbb{R}^{n \times n} \mid X = X^\top, X \succeq 0\}$, then

$$P_\Omega[X] = \sum_{i=1}^n \max\{\delta_i, 0\} v_i v_i^\top,$$

where $X = \sum_{i=1}^n \delta_i v_i v_i^\top$ is the eigen-decomposition of X .

Lemma 1.1 *Let Ω be a nonempty closed convex set and $Q \succ 0$. Then, for any $x \in \mathbb{R}^n$ and $z \in \Omega$, the following properties hold:*

- (1) $x \in \Omega \Leftrightarrow P_{\Omega, Q}[x] = x$.
- (2) $(x - P_{\Omega, Q}[x])^\top Q(z - P_{\Omega, Q}[x]) \leq 0$.
- (3) $\|P_{\Omega, Q}[x] - z\|_Q^2 \leq \|x - z\|_Q^2 - \|x - P_{\Omega, Q}[x]\|_Q^2$.

Definition 1.2 *A function $\theta(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be convex, if*

$$\theta(tx + (1 - t)y) \leq t\theta(x) + (1 - t)\theta(y), \quad \forall x, y \in \mathbb{R}^n, t \in [0, 1].$$

Furthermore, $\theta(\cdot)$ is said to be μ -strongly convex if, for a given $\mu > 0$,

$$\theta(tx + (1 - t)y) \leq t\theta(x) + (1 - t)\theta(y) - t(1 - t)\frac{\mu}{2}\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n, t \in [0, 1].$$

Definition 1.3 *Let $\theta(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex function and denote by $\mathbf{dom} \theta$ the domain of function $\theta(\cdot)$. Then, the subdifferential of $\theta(\cdot)$ at a point $v \in \mathbf{dom} \theta$ is defined as*

$$\partial\theta(y) = \{\xi \mid \theta(x) \geq \theta(y) + \langle \xi, x - y \rangle, \quad \forall x \in \mathbf{dom} \theta\},$$

and the vector ξ is said to be a subgradient of $\theta(\cdot)$ at y .

Accordingly, if ξ represents a subgradient of a μ -strongly convex function $\theta(\cdot)$ at a point $y \in \mathbf{dom} \theta$, we get that:

$$\theta(x) \geq \theta(y) + \langle \xi, x - y \rangle + \frac{\mu}{2}\|x - y\|^2, \quad \forall x \in \mathbf{dom} \theta.$$

Definition 1.4 Let Ω be a nonempty closed convex subset of $\mathbb{R}^n$ and let f be a mapping from Ω into $\mathbb{R}^n$. Then,

(a) f is said to be monotone on Ω if

$$\langle x - y, f(x) - f(y) \rangle \geq 0, \forall x, y \in \Omega. \quad (1.2)$$

(b) f is said to be strongly monotone on Ω if there is a positive constant μ , such that

$$\langle x - y, f(x) - f(y) \rangle \geq \mu \|x - y\|^2, \forall x, y \in \Omega. \quad (1.3)$$

(c) f is said to be Lipschitz continuous on Ω if there exists a constant $L_f > 0$ such that

$$\|f(x) - f(y)\| \leq L_f \|x - y\|, \forall x, y \in \Omega. \quad (1.4)$$

Definition 1.5 Let Ω be a nonempty subset of $\mathbb{R}^n$ and let $\{x^n\}$ be a sequence in $\mathbb{R}^n$. Then, $\{x^n\}$ is Fejér monotone with respect to Ω if

$$\|x^{n+1} - x\| \leq \|x^n - x\|, \forall x \in \Omega.$$

Definition 1.6 Let $\Omega \subseteq \mathbb{R}^n$ be a convex set and let $\emptyset$ denote the null set. Then, the set

$$N_\Omega(x) := \begin{cases} \{w | (y - x)^\top w \leq 0, \forall y \in \Omega\}, & \text{if } x \in \Omega \\ \emptyset, & \text{otherwise} \end{cases}$$

is called the normal cone of x to the set Ω .

Assumption 1.1 For Problem (1.1), there exists

$$\mathbf{x}' = (x'_1, \dots, x'_m) \in \mathbf{ri}(\text{dom}(\theta_1) \times \dots \times \text{dom}(\theta_m)) \cap \Omega,$$

where $\mathbf{ri}$ is the relative interior of a set and

$$\Omega := \left\{ \mathbf{x} := (x_1, \dots, x_m) \mid \sum_{i=1}^m A_i x_i = b, x_i \in \mathcal{X}_i, i = 1, \dots, m \right\}.$$

Lemma 1.2 Let $H \in \mathbb{R}^{n \times n}$ be a symmetric and positive semi-definite matrix. Then,

$$\langle a, b \rangle \leq \frac{1}{2} (\|a\|^2 + \|b\|^2), \quad \forall a, b \in \mathbb{R}^n, \quad (1.5)$$

$$\|a\|_H^2 - \|b\|_H^2 = 2\langle a, H(a - b) \rangle - \|a - b\|_H^2, \quad \forall a, b \in \mathbb{R}^n, \quad (1.6)$$

and

$$\langle a - b, H(c - d) \rangle = \frac{1}{2} (\|a - d\|_H^2 - \|a - c\|_H^2) + \frac{1}{2} (\|c - b\|_H^2 - \|d - b\|_H^2), \quad \forall a, b, c, d \in \mathbb{R}^n. \quad (1.7)$$

1.1.2 Equivalent Form

The convex programming problem (1.1) under consideration can be equivalently rewritten as a generic convex optimization problem of the form:

$$\begin{aligned} \min \quad & \theta(\mathbf{x}) := \sum_{i=1}^m \theta_i(x_i) \\ \text{s.t.} \quad & \mathbf{x} \in \Omega, \end{aligned}$$

where

$$\Omega := \left\{ \mathbf{x} := (x_1, \dots, x_m) \mid \sum_{i=1}^m A_i x_i = b, x_i \in \mathcal{X}_i, i = 1, \dots, m \right\}.$$

It is well-known that this problem can be further modified to give rise to the following *unconstrained* optimization problem by introducing an indicator function I_Ω :

$$\min_{\mathbf{x} \in \Omega} \theta(\mathbf{x}) \quad \Leftrightarrow \quad \min_{\mathbf{x}} \theta(\mathbf{x}) + I_\Omega(\mathbf{x}),$$

where,

$$I_\Omega(\mathbf{x}) = \begin{cases} 0, & \text{if } \mathbf{x} \in \Omega, \\ \infty, & \text{if } \mathbf{x} \notin \Omega. \end{cases}$$

Notice that the subdifferential of the indicator function at $\mathbf{x}$ is in fact the normal cone $N_\Omega(\mathbf{x})$, i.e., $N_\Omega(\mathbf{x}) \equiv \partial I_\Omega(\mathbf{x})$, and consequently, the first-order optimality conditions of the unconstrained problem (see [5]) amount to finding a zero point $\mathbf{x}^*$ (i.e., an optimal solution of the original problem) such that

$$0 \in \partial \theta(\mathbf{x}^*) + N_\Omega(\mathbf{x}^*).$$

This is essentially equivalent to a *variational inequality problem* (**VIP**): Finding a point $\mathbf{x}^* \in \Omega$ and $\xi^* \in \partial \theta(\mathbf{x}^*)$ such that

$$\langle \mathbf{x} - \mathbf{x}^*, \xi^* \rangle \geq 0, \quad \forall \mathbf{x} \in \Omega,$$

where

$$\mathbf{x} := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}, \quad \xi^* := \begin{pmatrix} \xi_{x_1}^* \\ \xi_{x_2}^* \\ \vdots \\ \xi_{x_m}^* \end{pmatrix}, \quad \text{and} \quad \xi_{x_i}^* \in \partial \theta_i(x_i^*), \quad (i = 1, \dots, m).$$

As the convex set Ω involves linear constraints, as a natural consequence, we introduce a Lagrangian multiplier λ corresponding to these constraints, and then Problem (1.1)

can be equivalently represented as the following variational inequality problem under the Assumption 1.1 by referring to [2, 3, 5, 46, 72, 94]: Find $\mathbf{u}^* = (x_1^*, \dots, x_m^*, \lambda^*) \in \mathcal{U} = \prod_{i=1}^m \mathcal{X}_i \times \mathbb{R}^l$ such that

$$\begin{cases} (x_i - x_i^*)^\top (\xi_{x_i}^* - A_i^\top \lambda^*) \geq 0, & \forall x_i \in \mathcal{X}_i, i = 1, \dots, m, \\ (\lambda - \lambda^*)^\top \left(\sum_{i=1}^m A_i x_i^* - b \right) \geq 0, & \forall \lambda \in \mathbb{R}^l. \end{cases} \quad (1.8a)$$

$$\quad (1.8b)$$

Indeed, for the i -th part of (1.8a), we have

$$\langle x_i - x_i^*, \xi_{x_i}^* - A_i^\top \lambda^* \rangle = \langle x_i - x_i^*, \xi_{x_i}^* \rangle + \langle x_i - x_i^*, -A_i^\top \lambda^* \rangle \geq 0, \quad \forall x_i \in \mathcal{X}_i.$$

Since $\xi_i^* \in \partial \theta_i(x_i^*)$, the convexity of $\theta_i(\cdot)$ implies

$$\theta_i(x_i) - \theta_i(x_i^*) \geq \langle x_i - x_i^*, \xi_{x_i}^* \rangle.$$

Then, the problem (1.8) can be equivalently rewritten as the following *mixed variational inequality*: Find $\mathbf{u}^* = (x_1^*, \dots, x_m^*, \lambda^*) \in \mathcal{U} = \prod_{i=1}^m \mathcal{X}_i \times \mathbb{R}^l$ such that

$$\begin{cases} \theta_i(x_i) - \theta_i(x_i^*) + \langle x_i - x_i^*, -A_i^\top \lambda^* \rangle \geq 0, & \forall x_i \in \mathcal{X}_i, i = 1, \dots, m, \\ (\lambda - \lambda^*)^\top \left(\sum_{i=1}^m A_i x_i^* - b \right) \geq 0, & \forall \lambda \in \mathbb{R}^l, \end{cases} \quad (1.9a)$$

$$\quad (1.9b)$$

which in turn can be compactly written as:

$$\text{VI}(\mathcal{U}, F, \theta) : \quad \theta(\mathbf{x}) - \theta(\mathbf{x}^*) + \langle \mathbf{u} - \mathbf{u}^*, F(\mathbf{u}^*) \rangle \geq 0, \quad \forall \mathbf{u} \in \mathcal{U}, \quad (1.10a)$$

where

$$\mathbf{u} := \begin{pmatrix} \mathbf{x} \\ \lambda \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \\ \lambda \end{pmatrix}, \quad \theta(\mathbf{x}) = \sum_{i=1}^m \theta_i(x_i), \quad \text{and} \quad F(\mathbf{u}) := \begin{pmatrix} -A_1^\top \lambda \\ -A_2^\top \lambda \\ \vdots \\ -A_m^\top \lambda \\ \sum_{i=1}^m A_i x_i - b \end{pmatrix}. \quad (1.10b)$$

It is easily verified that the operator $F(\mathbf{u})$, defined in (1.10b), is monotone because of $(\mathbf{u}_1 - \mathbf{u}_2)^\top (F(\mathbf{u}_1) - F(\mathbf{u}_2)) = 0$, for any $\mathbf{u}_1, \mathbf{u}_2 \in \mathcal{U}$. Furthermore, strong convexity of $\theta_i(\cdot)$ implies that

$$\theta_i(x_i) - \theta_i(x_i^*) - \frac{\mu_i}{2} \|x_i - x_i^*\|^2 \geq \langle x_i - x_i^*, \xi_{x_i}^* \rangle.$$

Therefore, VIP (1.8) can be represented via its alternate compact form: Find a point $\mathbf{u}^* \in \mathcal{U}$ such that

$$\theta(\mathbf{x}) - \theta(\mathbf{x}^*) - \frac{1}{2} \|\mathbf{x} - \mathbf{x}^*\|_G^2 + \langle \mathbf{u} - \mathbf{u}^*, F(\mathbf{u}^*) \rangle \geq 0, \quad \forall \mathbf{u} \in \mathcal{U}, \quad (1.11)$$

where G is a block diagonal matrix given by $G := \text{diag}(\mu_1 I_1, \mu_2 I_2, \dots, \mu_m I_m)$, where I_i ($i = 1, \dots, m$) represents the identity matrix of dimensionality n_i corresponding to the i -th variable x_i . As the solution set of (1.1) is assumed to be nonempty, the solution set of $\text{VI}(\mathcal{U}, F, \theta)$, denoted by $\mathcal{U}^*$, is also obviously nonempty.

It is noteworthy that Problems (1.10) and (1.11) are equivalent forms of Problem (1.1) corresponding to the cases when $\theta_i(\cdot)$ is convex and strongly convex, respectively. Therefore, we can conveniently reformulate Problem (1.1) into its mixed variational inequality form, as this proves useful for establishing several of the results provided in this thesis.

Another closely related problem that shares several commonalities with the mixed variational inequality (1.10) is the *saddle point* optimization problem. Let $(x_1^*, x_2^*, \dots, x_m^*, \lambda^*)$ be a saddle point of the Lagrangian function

$$\mathcal{L}'(x_1, x_2, \dots, x_m, \lambda) = \sum_{i=1}^m \theta_i(x_i) - \lambda^\top \left(\sum_{i=1}^m A_i x_i - b \right).$$

Then, by virtue of saddle point optimality, we have

$$\mathcal{L}'_{\lambda \in \mathbb{R}^l}(x_1^*, x_2^*, \dots, x_m^*, \lambda) \leq \mathcal{L}'(x_1^*, x_2^*, \dots, x_m^*, \lambda^*) \leq \mathcal{L}'_{\mathbf{x} \in \mathcal{X}}(x_1, x_2, \dots, x_m, \lambda^*).$$

Therefore, the above saddle point optimization problem is equivalent to finding $x_i^* \in \mathcal{X}_i$, $i = 1, \dots, m$, and $\lambda^* \in \mathbb{R}^l$ such that

$$\begin{cases} \theta_i(x_i) - \theta_i(x_i^*) + (x_i - x_i^*)^\top (-A_i^\top \lambda^*) \geq 0, & \forall x_i \in \mathcal{X}_i, \quad i = 1, \dots, m, \\ (\lambda - \lambda^*)^\top (\sum_{i=1}^m A_i x_i^* - b) \geq 0, & \forall \lambda \in \mathbb{R}^l, \end{cases} \quad (1.12)$$

which in turn can be equivalently rewritten as the above mixed variational inequality (1.10).

1.1.3 An ϵ -optimal Solution of MVI

Consider the following characterization of an ϵ -solution, where the solution set Ω^* has been characterized below (as in Theorem 2.3.5 of [36] and [67]), which facilitates the worst case $O(1/t)$ convergence rate of some first-order algorithms for mixed variational inequality problems.

Theorem 1.1 *The solution set of $\text{VI}(\mathcal{U}, F, \theta)$, denoted by $\mathcal{U}^*$, is convex and can be characterized as*

$$\mathcal{U}^* = \bigcap_{\mathbf{u} \in \mathcal{U}} \{\mathbf{u}^* \in \mathcal{U} : \theta(\mathbf{x}) - \theta(\mathbf{x}^*) + \langle \mathbf{u} - \mathbf{u}^*, F(\mathbf{u}) \rangle \geq 0\}. \quad (1.13)$$

According to the above theorem, [67, 78] said that $\bar{\mathbf{u}} \in \Omega$ can be seen as an approximate solution of $\text{VI}(\mathcal{U}, F, \theta)$, with accuracy ϵ (refer to [107] for further details on ϵ -approximate solutions) if it satisfies:

$$\theta(\mathbf{x}) - \theta(\bar{\mathbf{x}}) + (\mathbf{u} - \bar{\mathbf{u}})^\top F(\mathbf{u}) \geq -\epsilon, \quad \forall \mathbf{u} \in \Omega.$$

It can be shown that, for a given $\epsilon > 0$, after $\lfloor O(1/\epsilon) \rfloor$ iterations, any one of the proposed first-order methods can find a $\bar{\mathbf{u}}$ such that

$$\bar{\mathbf{u}} \in \mathcal{U} \quad \text{and} \quad \sup_{\mathbf{u} \in D_{\mathcal{U}}(\bar{\mathbf{u}})} \{\theta(\bar{\mathbf{x}}) - \theta(\mathbf{x}) + \langle \bar{\mathbf{u}} - \mathbf{u}, F(\mathbf{u}) \rangle\} \leq \epsilon,$$

for any given compact set $D_{\mathcal{U}}(\bar{\mathbf{u}}) = \{\mathbf{u} \in \mathcal{U} \mid \|\mathbf{u} - \bar{\mathbf{u}}\| \leq 1\}$.

1.2 Some Applications

In this section, we introduce some convex programming models and variational inequality problems that arise in real-world applications, and which subsequently provide us with a test-bed of instances to be used in our numerical computations.

- **Basis Pursuit Model:** In [10], the Basis Pursuit (BP) problem is described as follows:

$$\min_{x \in \mathbb{R}^n} \{\|x\|_1 \mid Ax = b\}, \quad (1.14)$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $\|\cdot\|_1$ is the vector ℓ_1 -norm. The BP model is widely applied in compressive sensing, especially in cases where a sparse signal x is required to be recovered using a minimal number of observations ($m \ll n$).

- LASSO Model ([111]): Simply stated, the Lasso problem is an ℓ_1 -regularized linear regression problem, which can be expressed as:

$$\min_{x \in \mathbb{R}^n} \{ \mu \|x\|_1 + \frac{1}{2} \|Ax - b\|_2^2 \}, \quad (1.15)$$

where $\mu > 0$ is a scalar regularization parameter that is usually selected by cross-validation. This problem can be seen as finding a sparse solution to a least squares or a linear regression problem.

- Dantzig Selector Model: Dantzig Selector (DS) is another interesting statistical model, which also performs variable selection and model fitting within a linear regression framework. This mathematical model is given by:

$$\min_{x \in \mathbb{R}^n} \{ \|x\|_1 \mid \|A^\top(Ax - b)\|_\infty \leq \kappa \}, \quad (1.16)$$

where $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^m$, κ is a tuning (or penalty) parameter, and $\|\cdot\|_\infty$ denotes the vector ℓ_∞ -norm.

- Robust Principal Component Analysis (PCA): The task of recovering low-rank and sparse components of a given matrix has useful applications and arises in many practical settings such as model selection in statistics, matrix rigidity in computer science, and system identification in engineering. This problem can be expressed as the following optimization model:

$$\min_{X, Y} \{ \|X\|_* + \tau \|Y\|_1 \mid X + Y = D \}, \quad (1.17)$$

where $D \in \mathbb{R}^{p \times q}$ is a (given) data matrix; the nuclear norm (defined as the sum of all singular values of a matrix), denoted by $\|\cdot\|_*$, plays the role of inducing the low-rank component X of D ; the ℓ_1 -norm (defined as the sum of absolute values of all entries), denoted by $\|\cdot\|_1$, is used to induce the sparse component Y of D ; and $\tau > 0$ is a balancing parameter, that serves to impose the required balance between low rank and sparsity. In [121], model (1.17) was extended to the more general setting, where only a fraction of entries of the matrix are assumed to be observable, and furthermore, these observations are corrupted by noise. Let Ω be a subset of the index set of entries $\{1, 2, \dots, p\} \times \{1, 2, \dots, q\}$.

The observed incomplete information is summarized by the operator $P_\Omega : \mathbb{R}^{p \times q} \rightarrow \mathbb{R}^{p \times q}$, which is defined as:

$$[P_\Omega(X)]_{ij} = \begin{cases} X_{ij}, & \text{if } (i, j) \in \Omega \\ 0, & \text{otherwise} \end{cases}, \quad 1 \leq i \leq p, 1 \leq j \leq q.$$

Hence, Problem (1.17) with Gaussian noise can be expressed as

$$\min\{\|X\|_* + \tau\|Y\|_1 \mid \|P_\Omega(D - X - Y)\|_F \leq \kappa\}, \quad (1.18)$$

where $\kappa > 0$ is the level of Gaussian noise and $\|\cdot\|_F$ denotes the Frobenius norm. It follows from the definition of P_Ω that

$$\|P_\Omega(D - X - Y)\|_F = \|P_\Omega(P_\Omega(D) - X - Y)\|_F,$$

and therefore, assuming that $P_\Omega(D) = N$ and referring to [61, 62, 89, 121], Problem (1.18) can be equivalently reformulated as an unconstrained model:

$$\min\{\|X\|_* + \tau\|Y\|_1 + \frac{1}{2\mu}\|P_\Omega(N - X - Y)\|_F\}, \quad (1.19)$$

where $\mu > 0$ is a penalty parameter corresponding to the linear constraint in (1.18).

- **Generalized Nash Equilibrium Problem (GNEP):** The GNEP [35, 36] comprises a set of N players, where the variables $x^i \in \mathbb{R}^{n_i}$ ($i = 1, \dots, N$) are controlled by player i . Assume that the vector $\mathbf{x}$ is formed by collating these decision variables, i.e.,

$$\mathbf{x} := \begin{pmatrix} x^1 \\ \vdots \\ x^N \end{pmatrix},$$

where $n := \sum_{i=1}^N n_i$, and let $\mathbf{x}^{-i}$ denote the vector formed by all the players' decision variables except those of player i . In order to highlight the i -th player's variables within $\mathbf{x}$, we use $(x^i, \mathbf{x}^{-i})$ in lieu of $\mathbf{x}$. Each player i has a to-be-minimized function $\theta_i : \mathbb{R}^n \rightarrow \mathbb{R}$ that depends not only on that player's variables x^i , but also on the remaining variables $\mathbf{x}^{-i}$ belonging to all the other players. Moreover, the strategy x^i for each player must conform to $\mathcal{K}_i(\mathbf{x}^{-i}) \subseteq \mathbb{R}^{n_i}$, which relies on the opponents' strategies, and this is referred to as the *feasible set* or *strategy set* of player i .

Suppose that the other players have identified their strategies $\mathbf{x}^{-i}$, then the aim of player i is to choose a strategy x^i that solves the following minimization problem:

$$\min_{x^i} \{\theta_i(x^i, \mathbf{x}^{-i}) \mid x^i \in \mathcal{K}_i(\mathbf{x}^{-i})\}. \quad (1.20)$$

For any $\mathbf{x}^{-i}$, denote the solution set of (1.20) by $\mathcal{S}_i(\mathbf{x}^{-i})$. Then, the GNEP is to find a vector $\hat{\mathbf{x}}$ such that $\hat{x}^i \in \mathcal{S}_i(\hat{\mathbf{x}}^{-i})$, for all $i = 1, \dots, N$. Under the assumption that each $\mathcal{K}_i \subseteq \mathbb{R}^{n_i}$ is a closed convex set and $\theta_i(\cdot, \mathbf{x}^{-i})$ is a convex and continuously differentiable function, the set of minimization problems (1.20) can be equivalently reformulated as the following variational inequality problem for each $i = 1, \dots, N$:

$$(y_i - \hat{x}_i)^\top \nabla_{x_i} \theta_i(\hat{x}_i, \mathbf{x}^{-i}) \geq 0, \quad \forall y_i \in \mathcal{K}_i(\mathbf{x}^{-i}). \quad (1.21)$$

Denoting $\mathcal{K}(\mathbf{x}) = \prod_{i=1}^N \mathcal{K}_i(\mathbf{x}^{-i}) \subseteq \mathbb{R}^n$, a point $\hat{\mathbf{x}}$ is a solution of GNEP if and only if $\hat{\mathbf{x}}$ is a solution of the following Quasi Variational Inequality (QVI):

$$(y - \hat{\mathbf{x}})^\top F(\hat{\mathbf{x}}) \geq 0, \quad \forall y \in \mathcal{K}(\hat{\mathbf{x}}), \quad (1.22)$$

where $F(\hat{\mathbf{x}}) = (\nabla_{x_i} \theta_i(\hat{x}_i, \mathbf{x}^{-i}))_{i=1}^N \in \mathbb{R}^n$.

1.3 Augmented Lagrangian Methods

There are a number of iterative methods in the literature for solving convex programming and variational inequality problems, including first-order algorithms, e.g., proximal point algorithms ([101, 111, 116]), augmented Lagrangian methods ([10, 33, 45, 48, 87, 113, 117]); and second-order algorithms, e.g., Newton-type methods [14, 100, 114]. Given the focus of this thesis, in this section, we will only concentrate on introducing and reviewing first-order algorithms for solving the convex minimization problem (1.1), along with their related properties.

A benchmark algorithm for solving the linearly constrained convex minimization problem is the well-established *Augmented Lagrangian Method* (**ALM**) [87, 113]. To begin with, consider the convex programming problem (1.1) with $m = 1$, where the convex feasible set has the following structure:

$$\mathcal{U} = \{x \in \mathbb{R}^n \mid Ax = b, x \in \mathcal{X}\},$$

where $A \in \mathbb{R}^{l \times n}$, $b \in \mathbb{R}^l$, and $\mathcal{X}$ is a simple nonempty closed convex subset of $\mathbb{R}^n$. Under the ALM framework, we first introduce a Lagrangian multiplier vector $\lambda \in \mathbb{R}^l$ associated with the linear constraint $Ax = b$, and define the augmented Lagrangian function as:

$$\mathcal{L}(x, \lambda) := \theta(x) - \langle \lambda, Ax - b \rangle + \frac{\beta}{2} \|Ax - b\|^2,$$

where $\beta > 0$ is a penalty parameter. Applying the classical augmented Lagrangian method [87, 113], leads to the following iterative scheme:

$$\begin{cases} x^{k+1} = \arg \min \{ \mathcal{L}(x, \lambda^k) \mid x \in \mathcal{X} \} \\ \lambda^{k+1} = \lambda^k - \beta (Ax^{k+1} - b) \end{cases}. \quad (1.23)$$

Extending this to the more general case where $m \geq 1$, let

$$\mathcal{L}(x_1, \dots, x_m, \lambda) := \sum_{i=1}^m \theta_i(x_i) - \left\langle \lambda, \sum_{i=1}^m A_i x_i - b \right\rangle + \frac{\beta}{2} \left\| \sum_{i=1}^m A_i x_i - b \right\|^2,$$

be the augmented Lagrangian function corresponding to Problem (1.1), where $\lambda \in \mathbb{R}^l$ is the Lagrange multiplier associated with the linear constraints and $\beta > 0$ plays the role of a penalty parameter. Accordingly, applying the ALM to (1.1) yields the following iterative scheme:

$$\begin{cases} \mathbf{x}^{k+1} := (x_1^{k+1}, \dots, x_m^{k+1}) = \arg \min_{\mathbf{x} \in \mathcal{X}} \{ \mathcal{L}(x_1, x_2, \dots, x_m, \lambda^k) \}, \\ \lambda^{k+1} = \lambda^k - \beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right). \end{cases} \quad (1.24a)$$

$$(1.24b)$$

It is clear that a direct application of ALM (1.24) fails to be easily implemented for all cases where $m \geq 2$ as the quadratic term $\frac{\beta}{2} \left\| \sum_{i=1}^m A_i x_i - b \right\|^2$ results in *coupled variables*. To circumvent this difficulty, *decomposition methods*, which serve to solve Problem (1.1) by dissecting it into a series of *decoupled subproblems*, have attracted a lot of attention in the literature. In several recent works, the successful application of both Gauss-Seidel and Jacobian decomposition-based techniques within the ALM framework [55, 61, 67, 71–73, 129] has been witnessed and these are introduced next.

The *Alternating Direction Method of Multipliers* (**ADMM**) [45, 48] is one of the most efficient solvers for the *two block* ($m = 2$) case of Problem (1.1). At each iteration, the

ADMM decomposes the ALM subproblems in Gauss-Seidel order, and obtains the next iterate $(x_1^{k+1}, x_2^{k+1}, \lambda^{k+1})$ via the following iterative scheme:

$$\begin{cases} x_1^{k+1} = \arg \min_{x_1 \in \mathcal{X}_1} \{ \mathcal{L}(x_1, x_2^k, \lambda^k) \}, & (1.25a) \end{cases}$$

$$\begin{cases} x_2^{k+1} = \arg \min_{x_2 \in \mathcal{X}_2} \{ \mathcal{L}(x_1^{k+1}, x_2, \lambda^k) \}, & (1.25b) \end{cases}$$

$$\begin{cases} \lambda^{k+1} = \lambda^k - \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b). & (1.25c) \end{cases}$$

This scheme solves for the variables x_1 and x_2 separately in *sequential fashion* and fully utilizes the separable structure of the original model (1.1). Such a method has witnessed many successful applications in the areas of image processing and statistical learning [10, 33]. The exceptional performance of ADMM in scientific computing has led towards extending the ADMM for solving the separable convex minimization model (1.1), for generic cases with $m \geq 3$. Applying a straightforward extension of the ADMM scheme to Problem (1.1) yields the *Extended Alternating Direction Method of Multipliers* (**EADM**), and correspondingly the iterative scheme (1.25) is modified as:

$$\begin{cases} x_i^{k+1} := \arg \min_{x_i \in \mathcal{X}_i} \{ \mathcal{L}(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i, x_{i+1}^k, \dots, x_m^k, \lambda^k) \}, \quad i = 1, \dots, m; & (1.26a) \\ \lambda^{k+1} = \lambda^k - \beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right). & (1.26b) \end{cases}$$

Such an extension of ADMM indeed demonstrates the computational effectiveness of this method [72]. However, a counter example in [18] shows that the EADM scheme (1.26) is not necessarily convergent for $m \geq 3$, which in theory restricts its potential applicability. In order to obtain the desired convergence result, one of the most efficient strategies is a variant of (1.26), which requires the help of an additional substitution procedure for correcting the output of (1.26). Such variants, when using the Gauss-Seidel decomposition, have been well-studied in the literature [71, 72] and have been shown to achieve good numerical stability. Recently, in [13, 59, 93], the authors also proved certain global convergence results for these variants under the strong convexity assumption. Furthermore, by employing *proximal point algorithm*-based ideas, some convergent Gauss-Seidel variants of a majorized ADMM have also been developed for solving convex composite conic programming problems [21, 95, 119].

While these methods showcase sequential approaches, the subproblems in (1.24) can also be solved *simultaneously*. When applying the Jacobian decomposition to the quadratic term of ALM, the resulting method leads to the following iterative scheme:

$$\begin{cases} x_i^{k+1} = \arg \min_{x_i \in \mathcal{X}_i} \{ \mathcal{L}(x_1^k, \dots, x_{i-1}^k, x_i, x_{i+1}^k, \dots, x_m^k, \lambda^k) \}, & i = 1, \dots, m; \\ \lambda^{k+1} = \lambda^k - \beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right). \end{cases} \quad (1.27a) \quad (1.27b)$$

As compared to (1.24), the above augmented Lagrangian-based *splitting* method (1.27) enjoys *decoupled* subproblems, thereby making it more computationally attractive and furthermore, it is deemed amenable to *parallel computation* when addressing large-scale problem instances. Unfortunately, the recent work of He et al. [67] demonstrates that even this method (1.27) may also fail to converge, thereby suffering from the same drawbacks as its Gauss-Seidel variants. To guarantee convergence when using such methods, two additional types of strategies are implemented. The first is to correct the output (predictor) generated by (1.27) by using a further correction step [55, 61, 67], whereas the second strategy is to introduce a proximal term for each subproblem of (1.27) [75, 129]. These Jacobian-type splitting methods, augmented with such correction strategies, have been demonstrated to perform favorably when solving large-scale problems.

1.4 Research Scope and Objectives

Specifically, the goal of this research work is to propose efficient numerical methods for solving the convex programming problem (1.1), provide an in-depth study of the properties of these algorithms, and detail computational results establishing the efficiency of these methods in comparison to existing approaches. As aforementioned, most splitting methods for the convex minimization problem (1.1) are based on augmented Lagrangian methods. Armed with this context and background, we will now delineate our research objectives.

Research Objective 1: Establish convergence of the ALM-based parallel splitting method

When dealing with Problem (1.1) with $m \geq 2$, it has been shown in [67] that the full ALM-based Jacobian decomposition method (1.27) is not necessarily convergent, which

leads to the following question: Under what conditions can we ensure *global convergence* of the method (1.27)? Han and Yuan [59] proved that the global convergence of the full ALM-based Gauss-Seidel decomposition scheme (1.26) is valid under *strong convexity* assumptions of the involved functions. Inspired by their result, in the first phase of our work, we prove that the iterative scheme (1.27), without any additional strategies, is also globally convergent if the involved functions are further assumed to be *strongly convex*. This result therefore serves as a supplement to such augmented Lagrangian-based splitting methods.

Research Objective 2: Convergence rate of the ALM-based parallel splitting method

Recently, in [75, 120, 129], some Jacobian variants of the augmented Lagrangian method were presented, which introduce a *proximal term* for each subproblem of (1.27), leading to the following iterative scheme:

$$\begin{cases} x_i^{k+1} = \arg \min_{x_i \in \mathcal{X}_i} \left\{ \mathcal{L}(x_1^k, \dots, x_{i-1}^k, x_i, x_{i+1}^k, \dots, x_m^k, \lambda^k) + \frac{1}{2} \|x_i - x_i^k\|_{S_i}^2 \right\}, \\ i = 1, \dots, m; \end{cases} \quad (1.28a)$$

$$\lambda^{k+1} = \lambda^k - \beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right), \quad (1.28b)$$

where S_i are symmetric positive definite matrices, thereby guaranteeing global convergence. Furthermore, for such variants, the worst-case $O(1/k)$ convergence rate was analyzed in [67, 75, 120] and the worst-case $o(1/k)$ convergence rate was also established in [25].

Following up on our first research objective, in our second research thrust, we propose a flexible augmented Lagrangian-based parallel splitting method, and prove the worst-case $O(1/t)$ convergence rate of this method under both ergodic and nonergodic conditions, where t represents the number of iterations. Furthermore, we show that the convergence rate can be improved from $O(1/t)$ to $o(1/t)$, and finally, we also demonstrate that this method can achieve global *linear convergence*, when the involved functions satisfy some additional conditions. This is an important result as most ALM-based methods are often limited to *sublinear* convergence rates, but our result extends this and proves that a full Jacobian-based decomposition method enjoys a *linear convergence* rate when

the involved functions are strongly convex, and the gradient of at least one of the objective functions possesses the property of Lipschitz continuity.

Research Objective 3: A proximal partially-parallel splitting algorithm (**PPSA**)

Given the computational effectiveness of the ADMM in solving two-block separable convex programs, when solving a multi-block convex separable program, it is natural to cluster the involved objective functions and their corresponding variables into two (mutually exclusive) groups, and then directly apply the original ADMM. However, such a straightforward extension is not easily implemented because coupled variables may appear in the resulting subproblems. To circumvent this difficulty, He et al. [73] proposed a novel splitting method with the following scheme:

$$\begin{cases} x_1^{k+1} := \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \left\| A_1 x_1 + \sum_{i=2}^m A_i x_i^k - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, & (1.29a) \end{cases}$$

$$\begin{cases} \hat{\lambda}^k = \lambda^k - \beta \left(A_1 x_1^{k+1} + \sum_{i=2}^m A_i x_i^k - b \right), & (1.29b) \end{cases}$$

$$\begin{cases} x_i^{k+1} := \arg \min_{x_i \in \mathcal{X}_i} \left\{ \theta_i(x_i) + \frac{s\beta}{2} \left\| A_i x_i - A_i x_i^k - \frac{\hat{\lambda}^k}{s\beta} \right\|^2 \right\}, \quad i = 2, \dots, m; & (1.29c) \end{cases}$$

$$\begin{cases} \lambda^{k+1} = \lambda^k - \beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right), & (1.29d) \end{cases}$$

where $s > m - 1$ is a positive proximal parameter. This can be equivalently expressed as:

$$\begin{cases} x_1^{k+1} := \arg \min_{x_1 \in \mathcal{X}_1} \{ \mathcal{L}(x_1, x_2^k, \dots, x_m^k, \lambda^k) \}, & (1.30a) \end{cases}$$

$$\begin{cases} x_i^{k+1} := \arg \min_{x_i \in \mathcal{X}_i} \left\{ \mathcal{L}(x_1^{k+1}, x_2^k, \dots, x_{i-1}^k, x_i, x_{i+1}^k, \dots, x_m^k, \lambda^k) \right. \\ \quad \left. + \frac{(s-1)\beta}{2} \|A_i(x_i - x_i^k)\|^2 \right\}, \quad i = 2, \dots, m; & (1.30b) \end{cases}$$

$$\begin{cases} \lambda^{k+1} = \lambda^k - \beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right), & (1.30c) \end{cases}$$

and, in fact, the augmented Lagrangian-based splitting method proposed in [121] can be seen as a particular case of the method in [73]. It is worthwhile to note that the necessary global convergence for this category of splitting methods is guaranteed by employing

proximal techniques. (Without these proximal terms, it has been demonstrated in [80] that such approaches are not necessarily convergent.) By utilizing a similar decomposition philosophy, Hou et al. [89] presented a new partially-parallel splitting method with the following simpler iterative scheme:

$$\begin{cases} x_1^{k+1} = \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2s} \left\| A_1 x_1 + \sum_{i=2}^m A_i x_i^k - b - \frac{s}{\beta} \lambda^k \right\|^2 \right\}, & (1.31a) \\ \lambda^{k+1} = \lambda^k - \frac{\beta}{s} \left(A_1 x_1^{k+1} + \sum_{i=2}^m A_i x_i^k - b \right), & (1.31b) \\ x_i^{k+1} = \arg \min_{x_i \in \mathcal{X}_i} \left\{ \theta_i(x_i) + \frac{r\beta}{2} \left\| A_i x_i - A_i x_i^k - \frac{1}{r\beta} (2\lambda^{k+1} - \lambda^k) \right\|^2 \right\}, & (1.31c) \\ i = 2, \dots, m; \end{cases}$$

and the numerical results reported in their work further verified the efficacy of these splitting methods. Moreover, in [73, 89], the worst-case $O(1/k)$ convergence rate was derived under both ergodic and non-ergodic senses. As aforementioned, adopting an extra corrector step is a popular strategy of Gauss-Seidel and Jacobian splitting methods to ensure global convergence when solving the generic model (1.1) with $m \geq 3$. Such an extra corrector step requires minimal overhead computational cost, while achieving the necessary performance gains [73, 89].

Motivated by the results in [61, 67, 72, 73, 75], in our third research objective, a new *proximal partially-parallel splitting method* (**PPSA**), which uses a similar decomposition strategy to the one displayed in [73], is presented for the general convex minimization problem (1.1) in the predictor-corrector fashion. In the proposed PPSA scheme, the next iterate point $\mathbf{u}^{k+1} = (\mathbf{x}^{k+1}, \lambda^{k+1})$ is generated by a correction step whose descent direction depends on both $\mathbf{u}^k$ and $\tilde{\mathbf{u}}^k$. Hence, the proposed splitting method is a *hybrid mechanism* combining the nice features of Gauss-Seidel and Jacobian decompositions, and convergence of the proposed method can be obtained under mild assumptions. The biggest differentiator of our approach as compared to existing methods is the use of *dynamic profitable step sizes* during the correction phase of the algorithm (to update the output of the prediction step), so that each iterate generated by this scheme is projected to be closer to the solution set.

The computational experience reported in this thesis validates that these strategies can significantly improve the numerical performance by reducing the number of iterations and CPU time taken to converge to the optimal solution. Furthermore, the worst-case $O(1/k)$ convergence rate can also be established under both ergodic and nonergodic scenarios.

Research Objective 4: Linearized versions of the PPSA algorithm

Clearly, the efficiency of augmented Lagrangian-based splitting methods relies heavily on the solvability of the underlying subproblems in (1.26) (or in (1.27)). However, the resulting subproblems in most cases cannot be solved efficiently and are found to be a drain on computational resources. To simplify these subproblems and (possibly) admit closed-form solutions, one of the most commonly prescribed strategies is the *linearization* of the nonlinear terms in the subproblems [55, 86].

Linearization is one of the most popular strategies to simplify the subproblems arising in the context of ALM-based approaches. In [70], the authors presented a portfolio of linearized proximal alternating direction methods for solving the convex minimization problem (1.1) under the predictor-corrector framework, which substantially enlarged the applicability of alternating direction-based contraction methods. Recently, Han et al. [55] also proposed some linearized versions of proximal parallel splitting methods for convex optimization problems with separable structure, and showed the corresponding global convergence for these linearized methods.

A significant advantage of our proposed PPSA algorithmic framework is that it also retains this same flexibility in deriving and solving partially-parallel linearized versions of the subproblems by incorporating certain PPA-based ideas. Specifically, in our fourth research objective, we propose a set of linearized versions of the proximal partially-parallel splitting method for the convex minimization problem (1.1) in the predictor-corrector framework, and establish the necessary convergence results. Our linearization strategies are judiciously selected such that the resulting subproblems either possess a *closed-form* expression or are rendered simple enough to be solved relatively efficiently as compared to their nonlinear counterparts. Moreover, we clarify the interconnections between these linearized variants and derive the worst-case $O(1/t)$ convergence rate of these methods under ergodic conditions.

Research Objective 5: A primal-dual prediction-correction algorithm for saddle point optimization

Another important class of optimization problems that shares a close-knit relationship with convex programming is the following saddle-point optimization problem:

$$\min_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} \Phi(x, y) := \langle y, Ax \rangle + \frac{\nu}{2} \|By - b\|^2, \quad (1.32)$$

where $\mathcal{X}$ and $\mathcal{Y}$ are two nonempty, closed and convex subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively; $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{n \times n}$ are two given matrices; $b \in \mathbb{R}^n$ is a given (observation) vector; $\nu > 0$ serves as a tuning parameter; $\langle \cdot, \cdot \rangle$ denotes the standard inner product of vectors; and $\|\cdot\|$ is the Euclidean norm. It is well documented in the literature that the saddle point problem (1.32) has many applications in diverse areas such as constrained optimization duality, zero-sum games, and image processing (see [15, 22, 34, 51, 57, 59, 105]), and so forth.

Now, let (x^*, y^*) be a solution of the saddle point problem (1.32). Then, from saddle point optimality conditions, we get,

$$\begin{aligned} \max_{x \in \mathcal{X}} \left\{ \langle x, A^\top y^* \rangle + \frac{\nu}{2} \|By^* - b\|^2 \right\} &\leq \langle x^*, A^\top y^* \rangle + \frac{\nu}{2} \|By^* - b\|^2 \\ &\leq \min_{y \in \mathcal{Y}} \left\{ \langle x^*, A^\top y \rangle + \frac{\nu}{2} \|By - b\|^2 \right\}, \end{aligned}$$

which reduces to the following pair of mixed variational inequalities:

$$\begin{cases} \langle x - x^*, -A^\top y^* \rangle \geq 0, & \forall x \in \mathcal{X}, \\ \theta(y) - \theta(y^*) + \langle y - y^*, Ax^* \rangle \geq 0, & \forall y \in \mathcal{Y}. \end{cases}$$

where

$$\theta(y) = \frac{\nu}{2} \|By - b\|^2.$$

The above variational characterization can be compactly rewritten as a problem of finding $\mathbf{u}^* \in \Omega$, such that

$$\theta(y) - \theta(y^*) + \langle \mathbf{u} - \mathbf{u}^*, F(\mathbf{u}^*) \rangle \geq 0, \quad \forall \mathbf{u} \in \Omega, \quad (1.33a)$$

where

$$\mathbf{u} := \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(\mathbf{u}) := \begin{pmatrix} -A^\top y \\ Ax \end{pmatrix}, \quad \text{and } \Omega := \mathcal{X} \times \mathcal{Y}. \quad (1.33b)$$

It is easy to verify that the underlying mapping F defined in (1.33b) is monotone, i.e.,

$$\langle \mathbf{u}_1 - \mathbf{u}_2, F(\mathbf{u}_1) - F(\mathbf{u}_2) \rangle \geq 0, \quad \forall \mathbf{u}_1, \mathbf{u}_2 \in \Omega.$$

To solve the saddle point problem (1.32), Chambolle and Pock [15] introduced the so-called First-Order Primal-Dual Algorithm (FOPDA), which has been successfully applied to solve many image restoration problems. In recent years, primal-dual methods have been extensively studied for solving convex optimization and saddle point problems arising in signal/image processing, computer vision, and machine learning. (See [91] for a recent overview on such primal-dual algorithmic advancements.) As pointed out in [15], the Primal-Dual Hybrid Gradient (PDHG) method [84, 141] as well as the Arrow-Hurwicz-Uzawa method [1] are special cases of the FOPDA method. Although the conditions for global convergence of FOPDA are weaker than those in [1, 141], the numerical results reported in [12, 59] show that FOPDA [15] performs better in practice when its parameters are set beyond their theoretically guaranteed limits. Motivated by these results, He and Yuan [59] introduced a new class of primal-dual based prediction-correction algorithms where the range of the involved parameters in [15] was extended to further augment the performance of FOPDA. Subsequently, Cai et al. [12] presented an improved version of He and Yuan's algorithm with a new correction step, which further weakens the conditions assumed in [59] and also improves its computational efficiency.

In our final contribution, we propose another primal-dual algorithm, which aims to bridge the gap between the algorithms in [12] and [59]. As a byproduct, we also obtain a *partially linearized* primal-dual algorithm, which only requires two projections during the prediction stage. This resulting linearized version of our algorithm is relatively easier to implement as compared to the algorithms in [12, 15, 59] as long as the projections of $\mathcal{X}$ and $\mathcal{Y}$ are readily available as closed-form solutions or are simple enough to compute numerically. Recently, Chambolle and Pock [16] established the $O(1/t)$ convergence rate of the FOPDA method under ergodic sense. In a similar fashion, we also establish the $O(1/t)$ convergence rate for our proposed algorithm, which further serves to provide an interesting complement to the methods introduced in [12, 59].

1.5 Outline

The remainder of this thesis is organized as follows:

Chapter 2 provides a brief, yet structured, literature review on augmented Lagrangian-based splitting methods for the separable convex program (1.1). These methods are classified into Gauss-Seidel type splitting method, Jacobian type splitting methods and hybrid splitting (partially Jacobian) methods as different sections in this chapter.

Chapter 3 derives the theoretical result that the global convergence of the directly full Jacobian decomposition method of ALM for solving the problem under consideration is guaranteed under strong convexity assumptions of the involved functions.

Chapter 4 introduces a flexible augmented Lagrangian-based parallel splitting method and proves the corresponding sub-linear convergence rate under some mild assumptions. Furthermore, global linear convergence is also established under additional conditions.

In Chapter 5, we advocate a new proximal partially-parallel splitting method (PPSA) for solving separable convex programming problems, and analyze the convergence result and sub-linear convergence rate of the proposed method under ergodic and non-ergodic conditions.

In Chapter 6, we propose a set of linearized proximal partially-parallel splitting methods, and analyze the relevant convergence results and the worst-case $O(1/t)$ convergence rate under ergodic conditions.

In Chapter 7, we put forward a new primal-dual prediction-correction algorithm for solving a saddle point optimization problem, and prove the global convergence and the worst-case $O(1/t)$ convergence rate under ergodic condition.

Finally, Chapter 8 concludes this thesis and identifies the future research directions.

Chapter 2

Literature Review

The separable convex programming problem (1.1) has received lots of attention due to its wide-spread applications in the areas of signal/image processing [4, 9, 110, 118, 131, 135], statistical learning [10, 121, 125], network optimization [27, 33, 104], etc. There are several efficient methods for solving the convex minimization problem (1.1) for the case of $m = 1$, and the augmented Lagrangian method is one of the most popularly used algorithms for solving this model. As reported in [87, 113], this method is now considered a benchmark algorithm from both theoretical as well as algorithmic viewpoints.

Of particular interest is the special case of Problem (1.1) with $m = 2$, which has many underlying applications arising across diversified areas, and can be characterized as follows:

$$\min_{\{x_1, x_2\}} \left\{ \theta_1(x_1) + \theta_2(x_2) \mid A_1 x_1 + A_2 x_2 = b, x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2. \right\}. \quad (2.1)$$

Remark 2.1 *In some sparse and low-rank optimization models, we often come across the case that both θ_1 and θ_2 are simple, e.g., in the sense that the resolvent of $\partial\theta_i$ often has an closed-form solution and can be characterized as:*

$$\left(I + \frac{1}{\beta} \partial\theta_i \right)^{-1} (x_i^0) := \arg \min \left\{ \theta_i(x_i) + \frac{\beta}{2} \|x_i - x_i^0\|^2 \right\}.$$

As an example, a representative $\ell_2 - \ell_1$ norm model, given by

$$\min_{\{x_1, x_2\}} \left\{ \frac{1}{2} \|Kx - b\|^2 + \tau \|x\|_1 \right\}, \quad (2.2)$$

can be recast into the desired structure of Problem (2.1) by introducing an auxiliary variable y as shown below:

$$\min_{\{x,y\}} \left\{ \frac{1}{2} \|Kx - b\|^2 + \tau \|y\|_1 \mid x - y = 0 \right\}.$$

However, a direct application of the augmented Lagrangian method on Problem 2.1 totally ignores the special separable structure inherent to the problem and fails to be easily implemented in practice (see Section 1.3 for details). This situation provided an impetus for the development of a popular algorithm for solving such structured convex optimization problems with linear constraints, namely, the *alternating direction method of multipliers* (ADMM).

2.1 ALM-based Gauss-Seidel Splitting Methods

ADMM was originally presented in [45, 48] for Problem (2.1) involving two separable variables and functions, and its the corresponding iterative scheme is given by:

$$\begin{cases} x_1^{k+1} = \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \left\| A_1 x_1 + A_2 x_2^k - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, \\ x_2^{k+1} = \arg \min_{x_2 \in \mathcal{X}_2} \left\{ \theta_2(x_2) + \frac{\beta}{2} \left\| A_1 x_1^{k+1} + A_2 x_2 - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, \\ \lambda^{k+1} = \lambda^k - \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b). \end{cases}$$

This method has been extensively studied in the fields of convex programming and variational inequalities [20, 29, 43, 126], and it is widely regarded as a benchmark solver for the linearly constrained problem (1.1). In [44], the ADMM was explained as an application of the classical *Douglas-Rachford splitting method* (DRSM) developed in [98]. Fukushima [43] obtained a decomposition algorithm by applying ADMM to the *dual* of the convex optimization problem under consideration, which was proved in [32] to be an equivalent form of the primal DRSM [28]. There are also algorithms resembling a combination of ADMM and the proximal method of multipliers [116], rather than the standard method

of multipliers (see [20, 29]). In [31], a generalized ADMM (**GADMM**) is proposed as:

$$\begin{cases} x_1^{k+1} = \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \left\| A_1 x_1 + A_2 x_2^k - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, \\ x_2^{k+1} = \arg \min_{x_2 \in \mathcal{X}_2} \left\{ \theta_2(x_2) + \frac{\beta}{2} \left\| \rho A_1 x_1^{k+1} + (1 - \rho)(b - A_2 x_2^k) + A_2 x_2 - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, \\ \lambda^{k+1} = \lambda^k - \beta (\rho A_1 x_1^{k+1} + (1 - \rho)(b - A_2 x_2^k) + A_2 x_2^{k+1} - b), \end{cases}$$

where the parameter $\rho \in (0, 2)$ is a relaxation factor. (When $\rho = 1$, it is easy to see that the GADMM reduces to the classical ADMM.) The GADMM inherits the nice features of the ADMM, namely that it fully utilizes the favorable property of dealing with $\theta_1(\cdot)$ and $\theta_2(\cdot)$ separately. Recently, [11] proposed a primal application of PPA, and the corresponding scheme is defined as:

$$\begin{cases} x_1^{k+1} = \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \left\| A_1 x_1 + A_2 x_2^k - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, \end{cases} \quad (2.5a)$$

$$\begin{cases} \lambda^{k+1} = \lambda^k - \beta (A_1 x_1^{k+1} + A_2 x_2^k - b), \end{cases} \quad (2.5b)$$

$$\begin{cases} x_2^{k+1} = \arg \min_{x_2 \in \mathcal{X}_2} \left\{ \theta_2(x_2) + \frac{\beta}{2} \left\| A_1 x_1^{k+1} + A_2 x_2 - b - \frac{\lambda^{k+1}}{\beta} \right\|^2 \right\}, \end{cases} \quad (2.5c)$$

which can recover the GADMM proposed in [31] by an additional relaxation step. The worst-case $O(1/t)$ convergence rate in an ergodic scenario is also established for this method. Other representative publications include [30, 126], and we refer to [10] for a comprehensive review of the applications of ADMM in the areas of distributed optimization and statistical learning, such as constrained sparse regression [7], sparse signal recovery [37], image restoration and denoising [108], trace norm regularized least squares minimization [133], the Dantzig selector [99], amongst others.

It is clear that ADMM fully utilizes the separable structure inherent to the objective function, and a beneficial consequence is that all the underlying subproblems often possess analytical (closed form) solutions or can be solved efficiently to a high degree of numerical accuracy. Then, based on results featured in [105, 106, 127], He and Yuan firstly analyzed the worst-case $O(1/k)$ convergence rate of the ADMM under ergodic and non-ergodic senses in [78] and [82], respectively. When considering two-block convex quadratic programs, the authors of [8, 60] both analyzed the *local linear* convergence for

ADMM. Moreover, if the sub-differentials of the involved functions in Problem (1.1) are piecewise linear multi-functions, [136] derived the linear convergence rates of ADMM and proximal ADMM under some error bound conditions. Recently, it was displayed in [26] that a generalized ADMM can also achieve linear convergence under some certain conditions. Furthermore, some ADMM-based linearized methods have been proposed for a set of specific applications, such as the Dantzig selector problem, image reconstruction problem, nuclear norm minimization, and low-rank, sparse matrix decomposition problems (see [130, 134, 138, 140]). Then, in [39], the authors put forward the linearized GADMM (L-GADMM) by linearizing the x_1 -subproblem, which leads to the iterative scheme:

$$\begin{cases} x_1^{k+1} = \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \left\| A_1 x_1 + A_2 x_2^k - b - \frac{\lambda^k}{\beta} \right\|^2 + \frac{1}{2} \|x_1 - x_1^k\|_{S_1}^2 \right\} \\ x_2^{k+1} = \arg \min_{x_2 \in \mathcal{X}_2} \left\{ \theta_2(x_2) + \frac{\beta}{2} \left\| \rho A_1 x_1^{k+1} + (1 - \rho)(b - A_2 x_2^k) + A_2 x_2 - b - \frac{\lambda^k}{\beta} \right\|^2 \right\} \\ \lambda^{k+1} = \lambda^k - \beta (\rho A_1 x_1^{k+1} + (1 - \rho)(b - A_2 x_2^k) + A_2 x_2 - b), \end{cases}$$

where S_1 is a symmetric positive definite matrix. The worst-case $O(1/k)$ convergence rate of L-GADMM was also analyzed under both ergodic and nonergodic conditions, and the numerical behavior was further studied using some statistical learning applications [39].

In the realm of partial differential equations, the *Peaceman-Rachford Splitting Method* (**PRSM**) [98, 112] is a well-known method, and shares algorithmic similarities with the DRSM. As using the DRSM to solve Problem (1.1) gained sufficient traction, [58] and [90] in turn extended the Peaceman-Rachford splitting method to the structured convex optimization Problem (1.1). As mentioned in [6, 47], when convergent, the PRSM often outperforms the DRSM, and the main iterative scheme of the PRSM can be summarized as follows:

$$\begin{cases} x_1^{k+1} = \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \left\| A_1 x_1 + A_2 x_2^k - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, \end{cases} \quad (2.6a)$$

$$\begin{cases} \lambda^{k+\frac{1}{2}} = \lambda^k - \beta (A_1 x_1^{k+1} + A_2 x_2^k - b), \end{cases} \quad (2.6b)$$

$$\begin{cases} x_2^{k+1} = \arg \min_{x_2 \in \mathcal{X}_2} \left\{ \theta_2(x_2) + \frac{\beta}{2} \left\| A_1 x_1^{k+1} + A_2 x_2 - b - \frac{\lambda^{k+\frac{1}{2}}}{\beta} \right\|^2 \right\}, \end{cases} \quad (2.6c)$$

$$\begin{cases} \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b), \end{cases} \quad (2.6d)$$

which is different from the ADMM as the Lagrange multipliers are updated *twice* at each iteration. However, it was shown in [23] that the sequence generated by (2.6) may be divergent, and therefore, to obtain the necessary global convergence, a new *strictly contractive PRSM* (**SC-PRSM**) was proposed in [69] with the following recursion:

$$\begin{cases} x_1^{k+1} = \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \left\| A_1 x_1 + A_2 x_2^k - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, \end{cases} \quad (2.7a)$$

$$\begin{cases} \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta (A_1 x_1^{k+1} + A_2 x_2^k - b), \end{cases} \quad (2.7b)$$

$$\begin{cases} x_2^{k+1} = \arg \min_{x_2 \in \mathcal{X}_2} \left\{ \theta_2(x_2) + \frac{\beta}{2} \left\| A_1 x_1^{k+1} + A_2 x_2 - b - \frac{\lambda^{k+\frac{1}{2}}}{\beta} \right\|^2 \right\}, \end{cases} \quad (2.7c)$$

$$\begin{cases} \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - r\beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b), \end{cases} \quad (2.7d)$$

where the parameter $r \in (0, 1)$. The convergence of this scheme was also analyzed in [23], and the sub-linear convergence rate was also derived under both ergodic and nonergodic scenarios. The effectiveness of the SC-PRSM was demonstrated by a series of numerical experiments involving statistical learning image reconstruction problems, including the least absolute shrinkage and selection operator (LASSO) model, the group LASSO model, the sparse logistic regression model, the image deblurring problem, the image inpainting problem, and the magnetic resonance imaging (MRI) problem.

Then, [68] introduced a new symmetric ADMM (**SADMM**):

$$\begin{cases} x_1^{k+1} := \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \left\| A_1 x_1 + A_2 x_2^k - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, \end{cases} \quad (2.8a)$$

$$\begin{cases} \lambda^{k+\frac{1}{2}} = \lambda^k - r_1\beta (A_1 x_1^{k+1} + A_2 x_2^k - b), \end{cases} \quad (2.8b)$$

$$\begin{cases} x_2^{k+1} := \arg \min_{x_2 \in \mathcal{X}_2} \left\{ \theta_2(x_2) + \frac{\beta}{2} \left\| A_1 x_1^{k+1} + A_2 x_2 - b - \frac{\lambda^{k+\frac{1}{2}}}{\beta} \right\|^2 \right\}, \end{cases} \quad (2.8c)$$

$$\begin{cases} \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - r_2\beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b), \end{cases} \quad (2.8d)$$

where r_1 and r_2 are two independent parameters and belong to the following set:

$$\mathcal{D} = \left\{ (r_1, r_2) \mid r_1 \in \left(0, \frac{1+\sqrt{5}}{2} \right), r_2 \in (-1, 1) \ \& \ r_1 + r_2 > 0, |r_1| < 1 + r_2 - r_2^2 \right\}.$$

In a similar decomposition vein, [52] described a semi-proximal-based strictly contractive

Peaceman-Rachford splitting method (**sP-PRSM**):

$$x_1^{k+1} := \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \left\| A_1 x_1 + A_2 x_2^k - b - \frac{\lambda^k}{\beta} \right\|^2 + \frac{1}{2} \|x_1 - x_1^k\|_{S_1}^2 \right\}, \quad (2.9a)$$

$$\lambda^{k+\frac{1}{2}} = \lambda^k - r_1 \beta (A_1 x_1^{k+1} + A_2 x_2^k - b), \quad (2.9b)$$

$$x_2^{k+1} := \arg \min_{x_2 \in \mathcal{X}_2} \left\{ \theta_2(x_2) + \frac{\beta}{2} \left\| A_1 x_1^{k+1} + A_2 x_2 - b - \frac{\lambda^{k+\frac{1}{2}}}{\beta} \right\|^2 + \frac{1}{2} \|x_2 - x_2^k\|_{S_2}^2 \right\}, \quad (2.9c)$$

$$\lambda^{k+1} = \lambda^{k+\frac{1}{2}} - r_2 \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b), \quad (2.9d)$$

where S_1 and S_2 are two positive semi-definite matrices and the involved parameters satisfy the conditions that

$$r_1 \in [0, 1), \quad r_2 \in \left(0, \frac{1 - r_1 + \sqrt{(1 + r_1)^2 + 4(1 - r_1^2)}}{2} \right).$$

Moreover, the authors proved that sP-PRSM is $O(1/t)$ sublinearly convergent, and the nonergodic convergence rate is derived on the required assumption that $r_1 \in [0, 1)$ and $r_2 \in (0, 1]$. Following this, the sP-PRSM reduces to a set of existing methods, such as the ADMM, SC-PRSM in [69], based on the assumptions of the introduced matrices and the involved parameters.

As can be gauged from the above discussion, there has been a great deal of effort spent on developing efficient algorithms for the two-block case of Problem (1.1). However, in some real-world applications, situations where there are more than two-blocks of separable variables and functions may be encountered. While it is certainly plausible to rearrange the variables and functions into two (mutually exclusive) groups and apply the ADMM to solve the resulting problem, this simple scheme also violates the favorable separable structure of the original problem. Therefore, a better approach would be to reconstruct the ADMM idea by solving the subproblems in one variable sequentially, and subsequently update the corresponding multipliers. The corresponding iterative procedure, namely the Extended Alternating Directing method of Multipliers (**EADM**), is

defined as follows:

$$\left\{ \begin{array}{l} x_1^{k+1} := \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \left\| A_1 x_1 + \sum_{i=2}^m A_i x_i^k - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, \\ \dots \end{array} \right. \quad (2.10a)$$

$$\left\{ \begin{array}{l} x_i^{k+1} := \arg \min_{x_i \in \mathcal{X}_i} \left\{ \theta_i(x_i) + \frac{\beta}{2} \left\| A_i x_i + \sum_{j=1}^i A_j x_j^{k+1} + \sum_{j=i+1}^m A_j x_j^k - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, \\ \dots \end{array} \right. \quad (2.10b)$$

$$\left\{ \begin{array}{l} x_m^{k+1} := \arg \min_{x_m \in \mathcal{X}_m} \left\{ \theta_m(x_m) + \frac{\beta}{2} \left\| A_m x_m + \sum_{j=1}^{m-1} A_j x_j^{k+1} - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, \\ \dots \end{array} \right. \quad (2.10c)$$

$$\left\{ \begin{array}{l} \lambda^{k+1} = \lambda^k - \beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right). \end{array} \right. \quad (2.10d)$$

Although heuristic results indicate that such direct extensions of ADMM do indeed demonstrate the required levels of computational efficiency, it was recently shown in [18] that this scheme may fail to converge for $m \geq 3$, except for the special case where all the functions are strongly convex [59]. Then, in [96], the sub-linear convergence rate of EADM is analyzed under both ergodic and non-ergodic sense provided that $(m - 1)$ functions are strongly convex and β is restricted to lie within a certain bounded region. In [88], Hong and Luo suggest modifying the update step of the Lagrange multiplier λ^k as:

$$\lambda^{k+1} = \lambda^k - \gamma \beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right),$$

where $\gamma > 0$ represents a very small step length, and this variant of EADM was proved to be locally linearly convergent on the basis of certain error bound conditions. Then, using a similar Lagrange multiplier update step and introducing proximal terms for the involved subproblems in (2.10), [115] designed a new flexible ADMM for Problem (1.1). Under conditions of strongly convexity of the involved functions and Lipschitzian gradients, some convergence properties were derived, where the step length γ can be selected from the interval $(0, 2)$. Moreover, Lin et al. [96] derived the global linear convergence under some stringent conditions.

Recently, Cai et al. [13] and Li et al. [93] proved that the 3-block EADM globally converges when one of three objective functions is strongly convex and has a Lipschitz gradient, and the parameter β is smaller than a certain bound. They also analyzed the sub-linear convergence rate and linear convergence rate. Then, [123] extended these two research results from the special case of $m = 3$ to the general case of $m \geq 3$, and established the necessary global convergence result for EADM under the assumptions that $(m - 2)$ functions are strongly convex and the penalty parameter is appropriately restricted. The worst-case convergence rate as measured by iteration complexity and asymptotically linear convergence are also derived under some additional assumptions.

To ensure global convergence of the iterative scheme (1.26), while yet retaining the nice features of the ADMM, a set of ALM-based Gaussian-Seidel splitting algorithms were proposed by utilizing a simple “correction” strategy [24, 71, 72, 74]. In [71, 72], the authors proposed that, based on the iterative point generated by EADM, one can construct a *descent direction* of the implicit merit function to the solution set. Using this direction, a correction step which only involves easy-to-compute matrix-vector products was used to obtain the next iterate, and under some assumptions, the generated sequence was shown to be globally convergent. (A similar scheme is also adopted in [62].) In [72], the authors further showed that the resulting sequence is convergent if the output of (1.26) is further modified by a Gaussian back-substitution procedure, and the numerical results reported in [72, 109] indicate that these schemes are quite effective in practice. An algorithmic framework combining EADM with such a substitution procedure was also proposed in [74]. Moreover, the authors proved the necessary convergence result and estimated the worst-case convergence rate by iteration complexity under both ergodic and nonergodic conditions.

Furthermore, by employing PPA-based ideas, some convergent Gauss-Seidel variants of a majorized ADMM have also been developed for solving convex composite conic programming problems [21, 94, 95, 119]. Besides, a set of linearized proximal ADMM methods were proposed in [70], based on the predictor-corrector mechanism, which substantially enlarged the application range of ALM-based Gauss-Seidel splitting methods. In addition, under certain inexact criteria, two inexact alternating-direction-based contraction methods were presented for Problem (1.1) in [50], which vastly widened the

applicability of these algorithms. Finally, [81] introduced a prototype method for Problem (1.1), which spans ALM, ADMM, and the EADM as special cases, and a unified and easily verifiable condition to ensure the convergence of this prototype approach was derived.

In order to solve the multi-block version of Problem (1.1), [68] also proposed a direct application of the SC-PRSM by using a proximal technique, namely SC-PRSM-PR, which is given below:

$$\left\{ \begin{array}{l} x_1^{k+1} = \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \left\| A_1 x_1 + A_2 x_2^k + A_3 x_3^k - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, \quad (2.11a) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta (A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k - b), \quad (2.11b) \\ x_2^{k+1} = \arg \min_{x_2 \in \mathcal{X}_2} \left\{ \theta_2(x_2) + \frac{\beta}{2} \left\| A_1 x_1^{k+1} + A_2 x_2 + A_3 x_3^k - b - \frac{\lambda^{k+\frac{1}{2}}}{\beta} \right\|^2 \right. \\ \quad \left. + \frac{s\beta}{2} \|A_2(x_2 - x_2^k)\|^2 \right\}, \quad (2.11c) \\ x_3^{k+1} = \arg \min_{x_3 \in \mathcal{X}_3} \left\{ \theta_3(x_3) + \frac{\beta}{2} \left\| A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3 - b - \frac{\lambda^{k+\frac{1}{2}}}{\beta} \right\|^2 \right. \\ \quad \left. + \frac{s\beta}{2} \|A_3(x_3 - x_3^k)\|^2 \right\}, \quad (2.11d) \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - r\beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b), \quad (2.11e) \end{array} \right.$$

where $r \in (0, 1)$. Without these proximal terms, a counterexample presented in [68] shows that this direct application of SC-PRCM fails to converge. Under the condition that $s > r$, the necessary global convergence is proved, and the sub-linear convergence rate for SC-PRSM-PR is obtained under ergodic and nonergodic conditions. The SC-PRSM-PR was also applied on some practical instances, which includes image restoration with mixed impulsive and Gaussian noises, robust principal component analysis with missing and noisy data, and quadratic discriminant analysis [38, 41, 102]. A set of numerical results further demonstrated the effectiveness of this splitting method.

Furthermore, in [24], under the predictor-corrector framework, a sequential updating scheme of the Lagrange multiplier method was developed for solving Problem (1.1), which

generates the intermediate points as follows:

$$\begin{aligned}
 & \left\{ \begin{aligned} x_1^{k+1} &= \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \left\| A_1 x_1 + \sum_{j=2}^m A_j x_j^k - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, & (2.12a) \\ \lambda^{k+\frac{1}{m}} &= \lambda^k - \beta \left(A_1 x_1^{k+1} + \sum_{j=2}^m A_j x_j^k - b \right), & (2.12b) \\ \dots & \\ x_i^{k+1} &= \arg \min_{x_i \in \mathcal{X}_i} \left\{ \theta_i(x_i) + \frac{\beta}{2} \left\| \sum_{j=1}^{i-1} A_j x_j^{k+1} + A_i x_i + \sum_{j=i+1}^m A_j x_j^k - b - \frac{\lambda^{k+\frac{i-1}{m}}}{\beta} \right\|^2 \right\}, & (2.12c) \\ \lambda^{k+\frac{i}{m}} &= \lambda^{k+\frac{i-1}{m}} - \beta \left(\sum_{j=1}^i A_j x_j^{k+1} + \sum_{j=i+1}^m A_j x_j^k - b \right), & (2.12d) \\ \dots & \\ \lambda^{k+\frac{m-1}{m}} &= \lambda^{k+\frac{m-2}{m}} - \beta \left(\sum_{j=1}^{m-1} A_j x_j^{k+1} + A_m x_m^k - b \right), & (2.12e) \\ x_m^{k+1} &= \arg \min_{x_m \in \mathcal{X}_m} \left\{ \theta_m(x_m) + \frac{\beta}{2} \left\| \sum_{j=1}^{m-1} A_j x_j^{k+1} + A_m x_m - b - \frac{\lambda^{k+\frac{m-1}{m}}}{\beta} \right\|^2 \right\}, & (2.12f) \\ \lambda^{k+1} &= \lambda^{k+\frac{m-1}{m}} - \beta \left(\sum_{j=1}^m A_j x_j^{k+1} - b \right). & (2.12g) \end{aligned} \right.
 \end{aligned}$$

Note that the scheme (2.12) is a natural extension of (2.7) for Problem (1.1) for $m \geq 3$, and can be regarded as embedding a full Gauss-Seidel decomposition for both primal and dual variables. As mentioned in [68], the direct extension method (2.12) may also fail to converge; however, provable convergence was established by utilizing the predictor-corrector framework. Then, an inexact version of the method (2.12) was also investigated in this research work.

2.2 ALM-based Jacobian Splitting Methods

Another commonly employed set of splitting schemes for solving Problem (1.1) are the ALM-based Jacobian decomposition methods, which result in the corresponding iterative

scheme (1.27). Besides, this iterative scheme (1.27) can be equivalently rewritten as:

$$\left\{ \begin{array}{l} x_1^{k+1} := \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \left\| A_1 x_1 + \sum_{i=2}^m A_i x_i^k - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, \\ \dots \\ x_i^{k+1} := \arg \min_{x_i \in \mathcal{X}_i} \left\{ \theta_i(x_i) + \frac{\beta}{2} \left\| A_i x_i + \sum_{j \neq i}^m A_j x_j^k - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, \\ \dots \\ x_m^{k+1} := \arg \min_{x_m \in \mathcal{X}_m} \left\{ \theta_m(x_m) + \frac{\beta}{2} \left\| A_m x_m + \sum_{j=2}^m A_j x_j^k - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, \\ \lambda^{k+1} = \lambda^k - \beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right). \end{array} \right. \quad \begin{array}{l} (2.13a) \\ (2.13b) \\ (2.13c) \\ (2.13d) \end{array}$$

The most noteworthy feature of this splitting approach (1.27) or (2.13) is that the resulting subproblems can be solved in parallel (via a suitably selected distributed computing mechanism), and such an approach has been shown to demonstrate impressive computational effectiveness when solving large-scale problem instances. Unfortunately, for certain problems, the iterative scheme (1.27) also fails to converge [67], even for the two-block case, and therefore to obtain the desired convergence, some variants of (1.27) have been proposed in [61, 64, 67, 128, 139], which operate in the same vein as other predictor-corrector methods. Moreover, [139] introduced an inexact simultaneous method for solving the constrained multiple-sets split feasibility problem, which combines proximal techniques with decomposition methods. These Jacobian-type splitting methods often outperform Gauss-Seidel type splitting methods by virtue of consuming lesser computing time. (See [61, 85] for detailed reported data.)

Furthermore, in [25, 75, 120, 129], introducing a proximal term for each subproblem of (1.27) has also been considered as an alternative strategy to ensure the necessary

convergence, and the main procedure in this case is given as follows:

$$\begin{cases}
 x_1^{k+1} = \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \left\| A_1 x_1 + \sum_{i=2}^m A_i x_i^k - b - \frac{\lambda^k}{\beta} \right\|^2 + \frac{1}{2} \|x_1 - x_1^k\|_{S_1}^2 \right\}, & (2.14a) \\
 \dots \\
 x_i^{k+1} = \arg \min_{x_i \in \mathcal{X}_i} \left\{ \theta_i(x_i) + \frac{\beta}{2} \left\| A_i x_i + \sum_{j \neq i}^m A_j x_j^k - b - \frac{\lambda^k}{\beta} \right\|^2 + \frac{1}{2} \|x_i - x_i^k\|_{S_i}^2 \right\}, & (2.14b) \\
 \dots \\
 x_m^{k+1} = \arg \min_{x_m \in \mathcal{X}_m} \left\{ \theta_m(x_m) + \frac{\beta}{2} \left\| A_m x_m + \sum_{j=1}^{m-1} A_j x_j^k - b - \frac{\lambda^k}{\beta} \right\|^2 + \frac{1}{2} \|x_m - x_m^k\|_{S_m}^2 \right\}, & (2.14c) \\
 \lambda^{k+1} = \lambda^k - \beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right), & (2.14d)
 \end{cases}$$

where S_i are symmetric positive definite matrices, and the worst-case convergence rate has also been analyzed under ergodic and non-ergodic conditions. In particular, [75] clarified the interconnections between the ADMM and the full Jacobian decomposition ALM-based methods. Additionally, under the predictor-corrector framework, Han et al. [55] presented a set of linearized parallel splitting methods by incorporating proximal techniques. Moreover, in [122], Tao and Yuan introduced a set of inexact parallel splitting augmented Lagrangian method by using two different correction strategies, which were deemed suitable for solving variational inequalities and convex programs with separable structures.

2.3 ALM-based Hybrid Splitting methods

Recently, Tao and Yuan [121] introduced the first model to accurately recover low-rank and sparse components of matrices from incomplete and noisy observations, and provided an efficient augmented Lagrangian-based algorithm to numerically solve Problem (1.19),

which is summarized below:

$$x_1^{k+1} := \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \left\| A_1 x_1 + A_2 x_2^k + A_3 x_3^k - b - \frac{\lambda^k}{\beta} \right\|^2 \right\}, \quad (2.15a)$$

$$\hat{\lambda}^k = \lambda^k - \beta (A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k - b), \quad (2.15b)$$

$$x_2^{k+1} := \arg \min_{x_2 \in \mathcal{X}_2} \left\{ \theta_2(x_2) + \frac{s\beta}{2} \left\| A_2 x_2 - A_2 x_2^k - \frac{\hat{\lambda}^k}{s\beta} \right\|^2 \right\}, \quad (2.15c)$$

$$x_3^{k+1} := \arg \min_{x_3 \in \mathcal{X}_3} \left\{ \theta_3(x_3) + \frac{s\beta}{2} \left\| A_3 x_3 - A_3 x_3^k - \frac{\hat{\lambda}^k}{s\beta} \right\|^2 \right\}, \quad (2.15d)$$

$$\lambda^{k+1} = \lambda^k - \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b), \quad (2.15e)$$

where $s > 2$ is a proximal parameter. Then, He et al. [73] proposed a novel splitting method using a combination of Gauss-Seidel and Jacobian decompositions, and the augmented Lagrangian-based splitting method in [121] can be shown to be a special case of the method in [73]. Similarly, [89] also proposed a new partially-parallel splitting method with a simpler iterative scheme, which follows a similar decomposition philosophy to [73], and a series of numerical simulations further verified the efficiency of these splitting methods. These variants of splitting methods inherited the idea of partial PPA regularization [54, 92, 103], and furthermore, [73, 89] also analyzed the sub-linear convergence rate for the proposed algorithms under both ergodic and non-ergodic senses.

Recently, in [80], a new block-wise ADMM splitting scheme was introduced for problem (1.1):

$$x_i^{k+1} := \arg \min_{x_i \in \mathcal{X}_i} \left\{ \theta_i(x_i) + \frac{\beta}{2} \left\| A_i x_i + \sum_{t \neq i}^{m_1} A_t x_t^k + \mathcal{B} \mathbf{y}^k - b - \frac{\lambda^k}{\beta} \right\|^2 + \frac{s_i \beta}{2} \|A_i(x_i - x_i^k)\|^2 \right\}, \quad i = 1, \dots, m_1; \quad (2.16a)$$

$$y_j^{k+1} := \arg \min_{y_j \in \mathcal{Y}_j} \left\{ \theta_j(y_j) + \frac{\beta}{2} \left\| B_j y_j + \mathcal{A} \mathbf{x}^{k+1} + \sum_{t \neq j}^{m_2} B_t y_t^k - b - \frac{\lambda^k}{\beta} \right\|^2 + \frac{s_j \beta}{2} \|B_j(y_j - y_j^k)\|^2 \right\}, \quad j = 1, \dots, m_2; \quad (2.16b)$$

$$\lambda^{k+1} = \lambda^k - \beta \left(\sum_{i=1}^{m_1} A_i x_i^{k+1} + \sum_{j=1}^{m_2} B_j y_j^{k+1} - b \right), \quad (2.16c)$$

where $m_1 + m_2 = m$, $\mathbf{x} \in \prod_{i=1}^{m_1} \mathcal{X}_i$, $\mathbf{y} \in \prod_{j=1}^{m_2} \mathcal{Y}_j = \prod_{i=m_1+1}^m \mathcal{X}_i$, $\mathcal{A} = [A_1, \dots, A_{m_1}]$, $\mathcal{B} = [B_1, \dots, B_{m_2}] = [A_{m_1+1}, \dots, A_m]$,

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_{m_1} \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_{m_2} \end{pmatrix} = \begin{pmatrix} x_{m_1+1} \\ \vdots \\ x_m \end{pmatrix}.$$

It is worthy to note that the involved $\mathbf{x}$ - and $\mathbf{y}$ -subproblems are completely parallelizable. By setting $s_1 = 0$, $\mathbf{x} = x_1$, and $\mathbf{y} = (y_1^\top, \dots, y_{m_1}^\top)^\top = (x_2^\top, \dots, x_m^\top)^\top$, the method (2.16) then reduces to the methods (1.29) and (1.30). Moreover, by setting $s_1 = s_2 = 0$, $\mathbf{x} = x_1$ and $\mathbf{y} = x_2$, (2.16) reduces to the classical ADMM. Then, in [76], a new block-wise ADMM with a relaxation factor is proposed, given by:

$$\begin{cases} x_i^{k+1} := \arg \min_{x_i \in \mathcal{X}_i} \left\{ \theta_i(x_i) + \frac{\beta}{2} \left\| A_i x_i + \sum_{t \neq i}^{m_1} A_t x_t^k + \mathcal{B} \mathbf{y}^k - b - \frac{\lambda^k}{\beta} \right\|^2 \right. \\ \quad \left. + \frac{s_i \beta}{2} \|A_i(x_i - x_i^k)\|^2 \right\}, \quad i = 1, \dots, m_1; \end{cases} \quad (2.17a)$$

$$\begin{cases} y_j^{k+1} := \arg \min_{y_j \in \mathcal{Y}_j} \left\{ \theta_j(y_j) + \frac{\beta}{2} \left\| B_j y_j + \mathcal{A} \mathbf{x}^{k+1} + \sum_{t \neq j}^{m_2} B_t y_t^k - b - \frac{\lambda^k}{\beta} \right\|^2 \right. \\ \quad \left. + \frac{s_j \beta}{2} \|B_j(y_j - y_j^k)\|^2 \right\}, \quad j = 1, \dots, m_2; \end{cases} \quad (2.17b)$$

$$\lambda^{k+1} = \lambda^k - r\beta \left(\sum_{i=1}^{m_1} A_i x_i^{k+1} + \sum_{j=1}^{m_2} B_j y_j^{k+1} - b \right), \quad (2.17c)$$

where $r \in (0, \frac{1+\sqrt{5}}{2})$ is a relaxation factor. Furthermore, global convergence is ensured under the conditions $s_1 > m_1 - 1$ and $s_2 > m_2 - 1$, and the worst-case convergence rate estimated by the iterations is established in the ergodic sense. Recently, adopting some proximal techniques, [115] presented a hybrid ADMM for solving the multiple block-wise Problem (1.1). The necessary global convergence is ensured by imposing strongly convexity assumptions on the involved objective functions. Based on the predictor-corrector framework, [42] introduced a block-wise ADMM with Gaussian back-substitution for solving multiple-block case of model (1.1).

2.4 Proximal Point Algorithms

The *Proximal Point Algorithm* (**PPA**), originally introduced in [101], is also an efficient tool for solving convex programming problems, as judged from recent monographs [5, 111]. In a seminal work, Rockafellar [116] extended the proximal minimization to a general proximal point algorithm for solving the maximal monotone operator problem, and demonstrated that the ALM is indeed an application of PPA to the dual problem of Problem (1.1) with $m = 1$. This theoretical convergence result has been generalized in subsequent research works [40, 53]. Proximal methods are well-known to be related to multiplier methods [6] and augmented Lagrangian methods [117]. For the case of $m = 2$, in [31], the authors showed that the classical Douglas-Rachford splitting method is a particular form of PPA, and that the ADMM can also be seen as a specific application of DRSM to the dual problem of (1.1). Moreover, [20] proposed a proximal-based decomposition method for Problem (1.1) with $m = 2$, which is defined as:

$$\begin{cases} \hat{\lambda}^k = \lambda^k - \beta_k (A_1 x_1^k + A_2 x_2^k - b), \\ x_1^{k+1} = \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \langle \hat{\lambda}^k, A_1 x_1 \rangle + \frac{1}{2\beta_k} \|x_1 - x_1^k\|^2 \right\}, \\ x_2^{k+1} = \arg \min_{x_2 \in \mathcal{X}_2} \left\{ \theta_2(x_2) + \langle \hat{\lambda}^k, A_2 x_2 \rangle + \frac{1}{2\beta_k} \|x_2 - x_2^k\|^2 \right\}, \\ \lambda^{k+1} = \lambda^k - \beta_k (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b), \end{cases} \quad (2.18)$$

where $\{\beta_k\}$ is a sequence of positive scalars, and can be viewed as an ALM-based linearized parallel splitting method. Recently, an interesting result in [11] further demonstrated that a customized choice of the proximal term can yield a modified ADMM for (1.1) for the case of $m = 2$, which can also be viewed as an application of PPA to the primal problem. In order to solve the generic model (1.1), a series of structure-utilizing algorithms, as recently demonstrated in [51], can be obtained by customizing the applications of PPA. In [83], a *relaxed customized PPA* (**RC-PPA**) was presented for Problem (1.1) involving a single block of variables, which first generates the prediction points $(\tilde{x}^k, \tilde{\lambda}^k)$ by the following scheme:

$$\begin{cases} \tilde{\lambda}^k = \lambda^k - \frac{1}{s}(Ax^k - b), \\ \tilde{x}^k = \arg \min_{x \in \mathcal{X}} \left\{ \theta(x) + \frac{r}{2} \left\| (x - x^k) - \frac{1}{r} A^\top (2\tilde{\lambda}^k - \lambda^k) \right\|^2 \right\}, \end{cases} \quad (2.19a)$$

$$\quad (2.19b)$$

or

$$\begin{cases} \tilde{x}^k = \arg \min_{x \in \mathcal{X}} \left\{ \theta(x) + \frac{r}{2} \left\| x - \left[x^k - \frac{1}{r} A^\top \lambda^k \right] \right\|^2 \right\}, \\ \tilde{\lambda}^k = \lambda^k - \frac{1}{s} [A(2\tilde{x}^k - x^k) - b], \end{cases} \quad (2.20a)$$

$$(2.20b)$$

where the positive scalars r and s satisfy $rs > \|A^\top A\|$, and the new iteration points utilize a relaxation step with step-length $\gamma \in (0, 2)$, and can be obtained as:

$$\begin{pmatrix} x^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ \lambda^k \end{pmatrix} - \gamma \begin{pmatrix} x^k - \tilde{x}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}.$$

For some special applications in image processing, the resulting problems of this method are easier than those of the ALM, and a series of numerical results further verified this claim. Then, [51] proposed a *unified* approach of customized proximal point algorithms for the linearly constrained Problem (1.1) and saddle point optimization problems. The conceptual algorithmic framework of the relaxed PPA for Problem (1.1) consists of two steps:

- First, generate the intermediate points by solving the mixed variational inequality:

$$\tilde{\mathbf{u}}^k \in \mathcal{U}, \quad \theta(\mathbf{u}) - \theta(\tilde{\mathbf{u}}^k) + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, F(\tilde{\mathbf{u}}^k) + Q(\tilde{\mathbf{v}}^k - \mathbf{v}^k) \rangle \geq 0, \quad \forall \mathbf{u} \in \mathcal{U}.$$

- Second, generate the new iterations $\mathbf{v}^{k+1}$

$$\mathbf{v}^{k+1} = \mathbf{v}^k - \gamma(\mathbf{v}^k - \tilde{\mathbf{v}}^k),$$

where the partial proximal matrix Q is positive semi-definite, the relaxation factor $\gamma \in (0, 2)$ and $\mathbf{v}$ is an appropriately chosen sub-vector of $\mathbf{u}$. Some relaxed PPA-based approaches in [11, 77, 83] can be boiled down to this algorithmic framework. When considering the multi-block cases of Problem (1.1), [51] also introduced a customized PPA,

which generates the prediction points by the following procedure:

$$\begin{cases} \tilde{x}_1^k = \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2m} \left\| mA_1(x_1 - x_1^k) + \sum_{j=1}^m A_j x_j^k - \frac{1}{\beta} \lambda^k \right\|^2 \right\}, & (2.21a) \\ \dots \end{cases}$$

$$\begin{cases} \tilde{x}_i^k = \arg \min_{x_i \in \mathcal{X}_i} \left\{ \theta_i(x_i) + \frac{\beta}{2m} \left\| mA_i(x_i - x_i^k) + \sum_{j=1}^m A_j x_j^k - \frac{1}{\beta} \lambda^k \right\|^2 \right\}, & (2.21b) \\ \dots \end{cases}$$

$$\begin{cases} \tilde{x}_m^k = \arg \min_{x_m \in \mathcal{X}_m} \left\{ \theta_m(x_m) + \frac{\beta}{2m} \left\| mA_m(x_m - x_m^k) + \sum_{j=1}^m A_j x_j^k - \frac{1}{\beta} \lambda^k \right\|^2 \right\} & (2.21c) \end{cases}$$

$$\begin{cases} \tilde{\lambda}^k = \lambda^k - \beta \left(\sum_{i=1}^m A_i \tilde{x}_i^k - b \right), & (2.21d) \end{cases}$$

and subsequently, the new iterations are produced by the relaxation step. One thing which deserves mention is that all x_i -subproblems can onceagain be solved in parallel, which proves to be useful favorable property when addressing large-scale problem instances.

Furthermore, based on the predictor-corrector framework, [66] put forward a set of PPA-based methods for structured monotone variational inequalities, which is also applicable for solving the separable convex problem (1.1). Then, in [79], a unified framework of proximal-based decomposition methods was proposed for solving structured VIs with two individual operators, including both exact and inexact versions. Moreover, some concert numerical approaches are presented for this type of VIs. In particular, the inexact version of this unified framework gives rise to some applicable algorithms that allow the involved sub-VIs to be efficiently solved (using certain mechanisms developed in the PPA literature). In addition, [65] gave a review on the recently developed Q -norm based customized PPA (where Q is a positive definite matrix). Under the customized PPA framework, it is easy to construct contraction-type methods for convex optimization with different categories of linear constraints. At each iteration of the proposed methods, we only need to solve the proximal subproblems that either have the closed form solutions or can be efficiently solved up to a high precision. Using the variational inequality approach, the contractive convergence and convergence rate proofs of the customized framework can be obtained.

Recently, Chen et al. [19] presented a general inertial proximal point algorithm for mixed variational inequality problems, which utilize the following algorithmic framework:

- First, generate the new iterations $\tilde{\mathbf{u}}^k$:

$$\tilde{\mathbf{u}}^k = \mathbf{u}^k + \alpha_k(\mathbf{u}^k - \mathbf{u}^{k-1}).$$

- Second, generate the intermediate points by solving the mixed variational inequality:

$$\mathbf{u}^{k+1} \in \mathcal{U}, \quad \theta(\mathbf{u}) - \theta(\mathbf{u}^{k+1}) + \langle \mathbf{u} - \mathbf{u}^{k+1}, F(\mathbf{u}^{k+1}) + \rho_k^{-1}Q(\mathbf{u}^{k+1} - \tilde{\mathbf{u}}^k) \rangle \geq 0, \quad \forall \mathbf{u} \in \mathcal{U}.$$

where, $Q \succeq 0$, $\{\alpha_k\}$ and $\{\rho_k\}$ are two sequences of positive scalars. Under mild assumptions, the necessary convergence and non-asymptotic $O(1/k)$ convergence rate was proved for the suggested method. This method is also applicable to the separable convex programming problem (1.1). When dealing with Problem (1.1) with $m = 1$ and $m = 2$, the suggested inertial PPA directly leads to inertial versions of ALM and ADMM, and the corresponding convergence results were established. Furthermore, in [17], a set of inertial proximal ADMM-based methods was proposed by combining inertial techniques and the proximal ADMM, which is described as:

$$\begin{cases} (\tilde{x}_1^k, \tilde{x}_2^k, \tilde{\lambda}^k) = (x_1^k, x_2^k, \lambda^k) + \alpha_k(x_1^k - \tilde{x}_1^k, x_2^k - \tilde{x}_2^k, \lambda^k - \tilde{\lambda}^k), & (2.22a) \\ x_1^{k+1} = \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \left\| A_1 x_1 + A_2 \tilde{x}_2^k - b - \frac{\tilde{\lambda}^k}{\beta} \right\|^2 + \frac{1}{2} \|x_1 - \tilde{x}_1^k\|_{S_1}^2 \right\}, & (2.22b) \\ \lambda^{k+1} = \tilde{\lambda}^k - \beta (A_1 x_1^{k+1} + A_2 \tilde{x}_2^k - b), & (2.22c) \\ x_2^{k+1} = \arg \min_{x_2 \in \mathcal{X}_2} \left\{ \theta_2(x_2) + \frac{\beta}{2} \left\| A_1 x_1^{k+1} + A_2 x_2 - b - \frac{\lambda^{k+1}}{\beta} \right\|^2 + \frac{1}{2} \|x_2 - \tilde{x}_2^k\|_{S_2}^2 \right\}, & (2.22d) \end{cases}$$

where $S_1 \succeq 0$, $S_2 \succeq 0$, and $\{\alpha_k\}$ is a sequence of nonnegative scalars. Some convergence properties were analyzed under some reasonable conditions, which contains the necessary global convergence, the asymptotic $O(1/\sqrt{k})$ and nonasymptotic $o(1/\sqrt{k})$ convergence rates. Moreover, the inertial primal-dual algorithm proposed in [16] and the inertial linearized ADMM presented in [19] can be considered as special cases of the inertial method (2.22) in [17]. The computational effectiveness of these inertial methods was verified by a series of numerical experiments using the total variation-based image reconstruction and the compressive principal component pursuit problems, respectively.

Chapter 3

Convergence of Augmented Lagrangian-based Parallel Splitting Method

3.1 Motivation

As mentioned in Chapter 1, the global convergence of the full Gauss-Seidel decomposition technique applied to ALM, when solving the generic model (1.1), was proved under *strong convexity* assumptions of the involved functions [59]. Moreover, it has been demonstrated in [67] that a straightforward Jacobian decomposition of ALM is not necessarily convergent when solving two block cases of Problem (1.1). Inspired by these two works, in this chapter, we prove that the iterative scheme (1.27), without using any additional strategies, is also *globally convergent* under the same assumptions as in [59]. Hence, our research result serves as a supplement to splitting methods that are based on the augmented Lagrangian method.

3.2 Algorithm

In this section, we first present an intuitive form of the parallel splitting method (1.27) for the separable model (1.1), based on the directly full Jacobian decomposition of ALM (1.27) as described in Algorithm 3.1. Then, we prove its global convergence in the variational context under strong convexity assumptions on θ_i 's.

Algorithm 3.1 Augmented Lagrangian-based parallel splitting method

Step 0. Initial Step: Given $\epsilon > 0$ and the initial iterate $(\mathbf{x}^0, \lambda^0) \in \mathcal{X} \times \mathbb{R}^l$. Choose a parameter $\beta > 0$. Set $k := 0$.

Step 1. Parallel Splitting Step: Obtain the new iterate by solving the following minimization problems corresponding to $i = 1, \dots, m$ (simultaneously, if possible):

$$x_i^{k+1} = \arg \min_{x_i \in \mathcal{X}_i} \left\{ \theta_i(x_i) + \frac{\beta}{2} \left\| A_i x_i + \sum_{j \neq i}^m A_j x_j^k - b - \beta^{-1} \lambda^k \right\|^2 \right\}. \quad (3.1)$$

Step 2. Lagrangian Multiplier Updating Step:

$$\lambda^{k+1} = \lambda^k - \beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right). \quad (3.2)$$

Step 3. Termination Step: If

$$\max \left\{ \max_{1 \leq i \leq m} \{ \|A_i x_i^k - A_i x_i^{k+1}\| \}, \|\lambda^k - \lambda^{k+1}\| \right\} < \epsilon, \quad (3.3)$$

then terminate the algorithm with $\mathbf{u}^{k+1} = (\mathbf{x}^{k+1}, \lambda^{k+1})$ being an acceptable approximate solution of (1.1). Otherwise, set $k := k + 1$ and go to **Step 1**.

3.3 Convergence Analysis

Next, we begin the global convergence analysis of Algorithm 3.1 with the following lemma.

Lemma 3.1 *If $\lambda^k = \lambda^{k+1}$ and $A_i x_i^k = A_i x_i^{k+1}$, for $i = 1, \dots, m$. Then, $\mathbf{u}^{k+1} := (\mathbf{x}^{k+1}, \lambda^{k+1})$ is a solution of the problem $\text{VI}(\mathcal{U}, F, \theta)$.*

Proof: By invoking the first-order optimality condition of the x_i -related subproblem in (3.1), for all $x_i \in \mathcal{X}_i$, we have

$$\begin{aligned} & \theta_i(x_i) - \theta_i(x_i^{k+1}) - \frac{\mu_i}{2} \|x_i - x_i^{k+1}\|^2 \\ & + \left\langle x_i - x_i^{k+1}, -A_i^\top \left[\lambda^k - \beta \left(A_i x_i^{k+1} + \sum_{j \neq i}^m A_j x_j^k - b \right) \right] \right\rangle \geq 0. \end{aligned} \quad (3.4)$$

It then follows from (3.2) that the inequality (3.4) is equivalent to

$$\begin{aligned} & \theta_i(x_i) - \theta_i(x_i^{k+1}) - \frac{\mu_i}{2} \|x_i - x_i^{k+1}\|^2 \\ & + \left\langle x_i - x_i^{k+1}, -A_i^\top \lambda^{k+1} + \beta A_i^\top \left[\sum_{j \neq i}^m A_j (x_j^k - x_j^{k+1}) \right] \right\rangle \geq 0, \quad \forall x_i \in \mathcal{X}_i. \end{aligned} \quad (3.5)$$

Summing the above inequalities for $i = 1, \dots, m$ yields (for all $\mathbf{x} \in \mathcal{X}$):

$$\begin{aligned} & \theta(\mathbf{x}) - \theta(\mathbf{x}^{k+1}) - \sum_{i=1}^m \frac{\mu_i}{2} \|x_i - x_i^{k+1}\|^2 \\ & + \left(\begin{array}{c} x_1 - x_1^{k+1} \\ \vdots \\ x_i - x_i^{k+1} \\ \vdots \\ x_m - x_m^{k+1} \end{array} \right)^\top \left\{ \left(\begin{array}{c} -A_1^\top \lambda^{k+1} \\ \vdots \\ -A_i^\top \lambda^{k+1} \\ \vdots \\ -A_m^\top \lambda^{k+1} \end{array} \right) + \left(\begin{array}{c} \sum_{j=2}^m \beta A_1^\top A_j (x_j^k - x_j^{k+1}) \\ \vdots \\ \sum_{j \neq i}^m \beta A_i^\top A_j (x_j^k - x_j^{k+1}) \\ \vdots \\ \sum_{j=1}^{m-1} \beta A_m^\top A_j (x_j^k - x_j^{k+1}) \end{array} \right) \right\} \geq 0. \end{aligned} \quad (3.6)$$

On the other hand, the iterative scheme (3.2) can be rewritten as

$$\sum_{j=1}^m A_j x_j^{k+1} - b + \beta^{-1} (\lambda^{k+1} - \lambda^k) = 0.$$

Consequently, we have

$$\left\langle \lambda - \lambda^{k+1}, \sum_{j=1}^m A_j x_j^{k+1} - b + \beta^{-1} (\lambda^{k+1} - \lambda^k) \right\rangle \geq 0, \quad \forall \lambda \in \mathbb{R}^l. \quad (3.7)$$

Using (3.7), along with the notation in (1.10b), (3.6) can be compactly written as:

$$\begin{aligned} & \theta(\mathbf{x}) - \theta(\mathbf{x}^{k+1}) - \frac{1}{2} \|\mathbf{x} - \mathbf{x}^{k+1}\|_G^2 + \langle \mathbf{u} - \mathbf{u}^{k+1}, F(\mathbf{u}^{k+1}) \rangle \\ & + \left(\begin{array}{c} x_1 - x_1^{k+1} \\ \vdots \\ x_m - x_m^{k+1} \\ \lambda - \lambda^{k+1} \end{array} \right)^\top \left(\begin{array}{c} \sum_{j=2}^m \beta A_1^\top A_j (x_j^k - x_j^{k+1}) \\ \vdots \\ \sum_{j=1}^{m-1} \beta A_m^\top A_j (x_j^k - x_j^{k+1}) \\ \beta^{-1} (\lambda^{k+1} - \lambda^k) \end{array} \right) \geq 0, \quad \forall \mathbf{u} \in \mathcal{U}. \end{aligned}$$

Obviously, if $\lambda^k = \lambda^{k+1}$ and $A_i x_i^k = A_i x_i^{k+1}$, for $i = 1, \dots, m$, it immediately follows from the above inequality that

$$\theta(\mathbf{x}) - \theta(\mathbf{x}^{k+1}) - \frac{1}{2} \|\mathbf{x} - \mathbf{x}^{k+1}\|_G^2 + \langle \mathbf{u} - \mathbf{u}^{k+1}, F(\mathbf{u}^{k+1}) \rangle \geq 0, \quad \forall \mathbf{u} \in \mathcal{U},$$

which shows that $\mathbf{u}^{k+1}$ is a solution of $\text{VI}(\mathcal{U}, F, \theta)$. Hence, the lemma is proved. $\square$

Indeed, the result in Lemma 3.1 provides us with a stopping criterion condition to be used within the implementation of Algorithm 3.1. Specifically, we can set the termination criteria to be of the form

$$\max \left\{ \max_{1 \leq i \leq m} \{ \|A_i x_i^k - A_i x_i^{k+1}\| \}, \|\lambda^k - \lambda^{k+1}\| \right\} < \text{tol},$$

where ‘tol’ is a small preset tolerance. Since each θ_i in (3.1) is assumed to be strongly convex with modulus μ_i , it is easily to conclude that the subdifferential operator $\partial\theta_i(x_i)$ is also strongly monotone with the same modulus μ_i .

Lemma 3.2 *Let $\mathbf{u}^* = (\mathbf{x}^*, \lambda^*)$ be a solution of the Problem $\text{VI}(\mathcal{U}, F, \theta)$. Then, the sequence $\{\mathbf{u}^k\}$ generated by Algorithm 3.1 satisfies*

$$\begin{aligned} \langle \lambda^k - \lambda^*, \beta^{-1} (\lambda^k - \lambda^{k+1}) \rangle &\geq \sum_{i=1}^m \mu_i \|x_i^{k+1} - x_i^*\|^2 + \|\lambda^k - \lambda^{k+1}\|_{\beta^{-1}}^2 \\ &\quad + \sum_{i=1}^m \left\langle \beta A_i (x_i^{k+1} - x_i^*), \sum_{j \neq i}^m A_j (x_j^k - x_j^{k+1}) \right\rangle. \end{aligned}$$

Proof: Since $\mathbf{u}^*$ is a solution of $\text{VI}(\mathcal{U}, F, \theta)$, for $i = 1, 2, \dots, m$, we first obtain

$$\theta_i(x_i) - \theta_i(x_i^*) - \frac{\mu_i}{2} \|x_i - x_i^*\|^2 + \langle x_i - x_i^*, -A_i^\top \lambda^* \rangle \geq 0, \quad \forall x_i \in \mathcal{X}_i.$$

Consequently, by setting $x_i := x_i^{k+1}$ in the above inequality, we obtain

$$\theta_i(x_i^{k+1}) - \theta_i(x_i^*) - \frac{\mu_i}{2} \|x_i^{k+1} - x_i^*\|^2 + \langle x_i^{k+1} - x_i^*, -A_i^\top \lambda^* \rangle \geq 0. \quad (3.8)$$

On the other hand, setting $x_i := x_i^*$ in (3.4) yields

$$\begin{aligned} &\theta_i(x_i^*) - \theta_i(x_i^{k+1}) - \frac{\mu_i}{2} \|x_i^* - x_i^{k+1}\|^2 \\ &\quad + \left\langle x_i^* - x_i^{k+1}, -A_i^\top \left[\lambda^k - \beta \left(A_i x_i^{k+1} + \sum_{j \neq i}^m A_j x_j^k - b \right) \right] \right\rangle \geq 0. \end{aligned} \quad (3.9)$$

Adding (3.8) and (3.9), we get

$$\begin{aligned} &-\mu_i \|x_i^{k+1} - x_i^*\|^2 \\ &\quad + \left\langle x_i^{k+1} - x_i^*, A_i^\top (\lambda^k - \lambda^*) - \beta A_i^\top \left(A_i x_i^{k+1} + \sum_{j \neq i}^m A_j x_j^k - b \right) \right\rangle \geq 0. \end{aligned}$$

The above inequality can be further rewritten as

$$\begin{aligned} & \langle A_i (x_i^{k+1} - x_i^*), \lambda^k - \lambda^* \rangle \\ & \geq \mu_i \|x_i^{k+1} - x_i^*\|^2 + \left\langle x_i^{k+1} - x_i^*, \beta A_i^\top \left(A_i x_i^{k+1} + \sum_{j \neq i}^m A_j x_j^k - b \right) \right\rangle. \end{aligned} \quad (3.10)$$

Summing the above inequalities corresponding to $i = 1, \dots, m$, and using $\sum_{i=1}^m A_i x_i^* = b$ and

$$A_i x_i^{k+1} + \sum_{j=1, j \neq i}^m A_j x_j^k - b = \sum_{j=1}^m A_j x_j^{k+1} - b + \sum_{j=1, j \neq i}^m A_j (x_j^k - x_j^{k+1}),$$

we conclude that

$$\begin{aligned} \left\langle \lambda^k - \lambda^*, \sum_{i=1}^m A_i x_i^{k+1} - b \right\rangle & \geq \sum_{i=1}^m \mu_i \|x_i^{k+1} - x_i^*\|^2 + \beta \left\| \sum_{i=1}^m A_i x_i^{k+1} - b \right\|^2 \\ & \quad + \sum_{i=1}^m \left\langle \beta A_i (x_i^{k+1} - x_i^*), \sum_{j \neq i}^m A_j (x_j^k - x_j^{k+1}) \right\rangle. \end{aligned}$$

The statement of this lemma is thus proved by using (3.2) and the notation of G . $\square$

Hereafter, we utilize the following notation so that the remaining analysis can be simplified:

$$M_i := (m-1)\beta A_i^\top A_i \quad \text{and} \quad M := \text{diag}(M_1, M_2, \dots, M_m, \beta^{-1}I_l), \quad (3.11)$$

where M_i ($i = 1, \dots, m$) are symmetric, positive semidefinite matrices and M is a block diagonal matrix.

Lemma 3.3 *Let $\mathbf{u}^* = (\mathbf{x}^*, \lambda^*)$ be a solution of the Problem VI($\mathcal{U}, F, \theta$). Then, the sequence $\{\mathbf{u}^k\}$ generated by Algorithm 3.1 satisfies*

$$\begin{aligned} \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_M^2 & \leq \|\mathbf{u}^k - \mathbf{u}^*\|_M^2 - \beta^{-1} \|\lambda^k - \lambda^{k+1}\|^2 \\ & \quad - 2 \sum_{i=1}^m \mu_i \|x_i^{k+1} - x_i^*\|^2 + \sum_{i=1}^m 4 \|x_i^{k+1} - x_i^*\|_{M_i}^2, \end{aligned} \quad (3.12)$$

where M is defined in (3.11).

Proof: For one thing, it follows from the inequality in Lemma 3.2 that

$$\begin{aligned}
 \beta^{-1} \|\lambda^{k+1} - \lambda^*\|^2 &= \beta^{-1} \|\lambda^k - \lambda^* + \lambda^{k+1} - \lambda^k\|^2 \\
 &= \beta^{-1} \|\lambda^k - \lambda^*\|^2 + 2 \langle \lambda^k - \lambda^*, \beta^{-1} (\lambda^{k+1} - \lambda^k) \rangle \\
 &\quad + \beta^{-1} \|\lambda^{k+1} - \lambda^k\|^2 \\
 &\leq \beta^{-1} \|\lambda^k - \lambda^*\|^2 - \beta^{-1} \|\lambda^{k+1} - \lambda^k\|^2 - 2 \sum_{i=1}^m \mu_i \|x_i^{k+1} - x_i^*\|^2 \\
 &\quad - \sum_{i=1}^m 2 \left\langle \beta A_i (x_i^{k+1} - x_i^*), \sum_{j \neq i}^m A_j (x_j^k - x_j^{k+1}) \right\rangle. \tag{3.13}
 \end{aligned}$$

Now, to simplify the last term in (3.13), using the identity (1.5), we get:

$$\begin{aligned}
 &- 2\beta \left\langle A_i (x_i^{k+1} - x_i^*), \sum_{j \neq i}^m A_j (x_j^k - x_j^{k+1}) \right\rangle \\
 &= -2\beta \left\langle A_i (x_i^{k+1} - x_i^*), \sum_{j \neq i}^m A_j (x_j^k - x_j^*) \right\rangle \\
 &\quad + 2\beta \left\langle A_i (x_i^{k+1} - x_i^*), \sum_{j \neq i}^m A_j (x_j^{k+1} - x_j^*) \right\rangle \\
 &\leq 2(m-1)\beta \|x_i^{k+1} - x_i^*\|_{A_i^\top A_i}^2 + \sum_{j \neq i}^m \beta \|x_j^{k+1} - x_j^*\|_{A_j^\top A_j}^2 \\
 &\quad + \sum_{j \neq i}^m \beta \|x_j^k - x_j^*\|_{A_j^\top A_j}^2. \tag{3.14}
 \end{aligned}$$

Notice that

$$\sum_{i=1}^m \sum_{j \neq i}^m \beta \|x_j^{k+1} - x_j^*\|_{A_j^\top A_j}^2 = \sum_{i=1}^m (m-1)\beta \|x_i^{k+1} - x_i^*\|_{A_i^\top A_i}^2 = \sum_{i=1}^m \|x_i^{k+1} - x_i^*\|_{M_i}^2$$

and

$$\sum_{i=1}^m \sum_{j \neq i}^m \beta \|x_j^k - x_j^*\|_{A_j^\top A_j}^2 = \sum_{i=1}^m (m-1)\beta \|x_i^k - x_i^*\|_{A_i^\top A_i}^2 = \sum_{i=1}^m \|x_i^k - x_i^*\|_{M_i}^2.$$

Therefore, combining the above two equalities and (3.14) together yields:

$$\begin{aligned}
 & - \sum_{i=1}^m 2 \left\langle \beta A_i (x_i^{k+1} - x_i^*), \sum_{j \neq i}^m A_j (x_j^k - x_j^{k+1}) \right\rangle \\
 & \leq \sum_{i=1}^m 2 \|x_i^{k+1} - x_i^*\|_{M_i}^2 + \sum_{i=1}^m \|x_i^{k+1} - x_i^*\|_{M_i}^2 + \sum_{i=1}^m \|x_i^k - x_i^*\|_{M_i}^2 \\
 & = \sum_{i=1}^m 3 \|x_i^{k+1} - x_i^*\|_{M_i}^2 + \sum_{i=1}^m \|x_i^k - x_i^*\|_{M_i}^2.
 \end{aligned}$$

This, along with (3.13) implies that

$$\begin{aligned}
 \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_M^2 &= \sum_{i=1}^m \|x_i^{k+1} - x_i^*\|_{M_i}^2 + \beta^{-1} \|\lambda^{k+1} - \lambda^*\|^2 \\
 &\leq \sum_{i=1}^m \|x_i^k - x_i^*\|_{M_i}^2 + \beta^{-1} \|\lambda^k - \lambda^*\|^2 - \beta^{-1} \|\lambda^{k+1} - \lambda^k\|^2 \\
 &\quad - 2 \sum_{i=1}^m \mu_i \|x_i^{k+1} - x_i^*\|^2 + \sum_{i=1}^m 4 \|x_i^{k+1} - x_i^*\|_{M_i}^2 \\
 &= \|\mathbf{u}^k - \mathbf{u}^*\|_M^2 - \|\lambda^k - \lambda^{k+1}\|_{\beta^{-1}}^2 - 2 \sum_{i=1}^m \mu_i \|x_i^{k+1} - x_i^*\|^2 \\
 &\quad + \sum_{i=1}^m 4 \|x_i^{k+1} - x_i^*\|_{M_i}^2.
 \end{aligned}$$

Then, we obtain the inequality (3.12). $\square$

With Lemmas 3.2 and 3.3, we are now preparing for the statement of the convergence theorem of Algorithm 3.1. As a matter of convenience, the maximum eigenvalue of a given positive semidefinite matrix M is represented by the notation $\delta_{\max}(M)$.

Theorem 3.1 *Suppose that $\theta_i(\cdot)$, for all $i = 1, \dots, m$, are μ_i -strongly convex functions and*

$$0 < \beta < \frac{\mu_i}{2(m-1)\delta_{\max}(A_i^\top A_i)}, \quad (3.15)$$

then, the sequence $\{\mathbf{u}^k\}$ generated by Algorithm 3.1 globally converges to a solution of $\text{VI}(\mathcal{U}, F, \theta)$.

Proof: Let $\mathbf{u}^*$ be an arbitrary solution of $\text{VI}(\mathcal{U}, F, \theta)$. First, it is clear from the definition of M_i that for each i :

$$\|x_i^{k+1} - x_i^*\|_{M_i}^2 \leq \delta_{\max}(M_i) \|x_i^{k+1} - x_i^*\|^2 = (m-1)\beta\delta_{\max}(A_i^\top A_i) \|x_i^{k+1} - x_i^*\|^2.$$

Accordingly, the last two terms of (3.12) can be merged into the following form:

$$-2\mu_i \sum_{i=1}^m \|x_i^{k+1} - x_i^*\|^2 + \sum_{i=1}^m 4 \|x_i^{k+1} - x_i^*\|_{M_i}^2 \leq - \sum_{i=1}^m \rho_i \|x_i^{k+1} - x_i^*\|^2,$$

where,

$$\rho_i := (2\mu_i - 4(m-1)\beta\delta_{\max}(A_i^\top A_i)).$$

Thus, the inequality (3.12) can be further simplified as

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_M^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|_M^2 - \beta^{-1} \|\lambda^k - \lambda^{k+1}\|^2 - \sum_{i=1}^m \rho_i \|x_i^{k+1} - x_i^*\|^2, \quad (3.16)$$

where, ρ_i is obviously a positive constant under the assumption (3.15). Hence, (3.16) implies that

$$\sum_{i=1}^m \rho_i \|x_i^{k+1} - x_i^*\|^2 \leq \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_M^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|_M^2 \leq \dots \leq \|\mathbf{u}^0 - \mathbf{u}^*\|_M^2 < +\infty,$$

which essentially states that the sequences $\{\mathbf{x}^k\}$ and $\{\|\mathbf{u}^k - \mathbf{u}^*\|_M^2\}$ are bounded. Moreover, as

$$\|\mathbf{u}^k - \mathbf{u}^*\|_M^2 = \sum_{i=1}^m (m-1)\beta \|A_i(x_i^k - x_i^*)\|^2 + \beta^{-1} \|\lambda^k - \lambda^*\|^2$$

is bounded, it follows that $\{\|\lambda^k - \lambda^*\|^2\}$ must be bounded. Then, $\{\mathbf{x}^k\}$ and $\{\lambda^k\}$ are all bounded, and consequently, the sequence $\{\mathbf{u}^k\}$ is bounded.

On the other hand, rearranging the terms in (3.16) and summing both sides over all k yields:

$$\begin{aligned} & \sum_{k=0}^{\infty} \left(\beta^{-1} \|\lambda^k - \lambda^{k+1}\|^2 + \sum_{i=1}^m \rho_i \|x_i^{k+1} - x_i^*\|^2 \right) \\ & \leq \sum_{k=0}^{\infty} \left(\|\mathbf{u}^k - \mathbf{u}^*\|_M^2 - \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_M^2 \right) \\ & \leq \|\mathbf{u}^0 - \mathbf{u}^*\|_M^2 < +\infty. \end{aligned}$$

As an immediate consequence, for $i = 1, \dots, m$, we have

$$\lim_{k \rightarrow \infty} \|x_i^{k+1} - x_i^*\|^2 = 0 \quad \text{and} \quad \lim_{k \rightarrow \infty} \beta^{-1} \|\lambda^{k+1} - \lambda^k\|^2 = 0.$$

From (3.2), we notice that $\lambda^k - \lambda^{k+1} = \beta (\sum_{i=1}^m A_i x_i^{k+1} - b)$. Hence,

$$\lim_{k \rightarrow \infty} \beta^{-1} \|\lambda^{k+1} - \lambda^k\|^2 = \lim_{k \rightarrow \infty} \beta \left\| \sum_{i=1}^m A_i x_i^{k+1} - b \right\|^2 = 0. \quad (3.17)$$

Finally, the boundedness of $\{\mathbf{u}^k\}$ implies that the sequence $\{\lambda^k\}$ is also bounded. Then, it must have at least one cluster point. Without loss of generality, let λ^∞ be a cluster point and let $\{\lambda^{k_j}\}$ be the corresponding subsequence that converges to λ^∞ . Taking limits along this subsequence in (3.4) and (3.17), we have for all $i = 1, \dots, m$:

$$\begin{cases} \theta_i(x_i) - \theta_i(x_i^*) - \frac{\mu_i}{2} \|x_i - x_i^*\|^2 + \langle x_i - x_i^*, -A_i^\top \lambda^\infty \rangle \geq 0, & \forall x_i \in \mathcal{X}_i, \\ \sum_{i=1}^m A_i x_i^* - b = 0, \end{cases}$$

which shows that the cluster point λ^∞ is an optimal Lagrangian multiplier. Thanks to the arbitrariness of λ^* in above analysis, we can replace it with λ^∞ and conclude that the entire sequence globally converges to a solution of $\text{VI}(\mathcal{U}, F, \theta)$. Hence, this theorem is then proved. $\square$

3.4 Conclusions

In this chapter, we considered the full Jacobian decomposition-based ALM for solving the multiple-block separable convex minimization problem (1.1), and provided a theoretical result, which showed that the convergence of the directly full Jacobian decomposition method (Algorithm 3.1) is guaranteed under strong convexity assumptions of the involved functions.

Chapter 4

On the Convergence Rate of Flexible Augmented Lagrangian-based Parallel Splitting Method

4.1 Motivation

Recently, in [67], it was demonstrated that the straightforward Jacobian decomposition of ALM may be divergent when dealing with Problem (1.1) even when $m = 2$. In Chapter 3, we proved the global convergence of the augmented Lagrangian-based parallel splitting method under the assumption that all objective functions are strongly convex. In this chapter, we extend these results by developing a more flexible augmented Lagrangian-based parallel splitting method and deduce the worst-case $O(1/t)$ convergence rate of this algorithm under both ergodic and nonergodic conditions, where t represents the number of iterations. Furthermore, we establish that the convergence rate can be improved from $O(1/t)$ to $o(1/t)$, and finally, we also demonstrate that this method can achieve global *linear convergence*, when the involved functions satisfy some additional conditions.

4.2 Algorithm

In this section, we begin by designing more flexible parallel splitting method (1.27) for Problem (1.1), based on the directly full Jacobian decomposition of ALM (1.27), as described in Algorithm 3.1. Then, we highlight some properties of the proposed algorithm and its correlations with Algorithm 3.1, and finally, we infer its convergence rate in the variational context under strong convexity assumptions on θ_i 's.

Algorithm 4.1 Flexible augmented Lagrangian-based parallel splitting method

Step 0. Initial Step: Given $\epsilon > 0$ and the initial iterate $(\mathbf{x}^0, \lambda^0) \in \mathcal{X} \times \mathbb{R}^l$. Choose parameters $\beta > 0$ and $\gamma \in (0, 2)$. Set $k := 0$.

Step 1. Parallel Splitting Step: Obtain the new iterate by solving the following minimization problems corresponding to $i = 1, \dots, m$ (simultaneously, if possible):

$$x_i^{k+1} = \arg \min_{x_i \in \mathcal{X}_i} \left\{ \theta_i(x_i) + \frac{\beta}{2} \left\| A_i x_i + \sum_{j \neq i}^m A_j x_j^k - b - \beta^{-1} \lambda^k \right\|^2 \right\}. \quad (4.1)$$

Step 2. Lagrangian Multiplier Updating Step:

$$\lambda^{k+1} = \lambda^k - \gamma \beta \left(\sum_{i=1}^m A_i x_i^{k+1} - b \right). \quad (4.2)$$

Step 3. Termination Step: If

$$\max \left\{ \max_{1 \leq i \leq m} \{ \|A_i x_i^k - A_i x_i^{k+1}\| \}, \|\lambda^k - \lambda^{k+1}\| \right\} < \epsilon,$$

then terminate the algorithm with $\mathbf{u}^{k+1} = (\mathbf{x}^{k+1}, \lambda^{k+1})$ being an acceptable approximate solution of (1.1). Otherwise, set $k := k + 1$ and go to **Step 1**.

Remark 4.1 Notice that Algorithm 4.1 is a little different from Algorithm 3.1 proposed in Chapter 3 as the Lagrangian updating step now involves a relaxation factor γ , thereby rendering this version of the algorithm more flexible and amenable to algebraic manipulations. When $\gamma = 1$, Algorithm 3.1 is viewed as a special case of Algorithm 4.1.

4.3 Sublinear Convergence Rate

To begin with, define matrices G_x and G as follows:

$$G_x = \text{diag}(G_1, \dots, G_m), \quad G_i = \beta A_i^\top A_i \text{ and } G = \text{diag}((m+1)G_x, \frac{1}{\gamma\beta}I_l), \quad (4.3)$$

where I_l is an identity matrix, $\gamma \in (0, 2)$ and $\beta > 0$. Hence, G_x and G are both symmetric positive semi-definite. These matrices will be important for defining the regularization

matrices used in our subsequent algorithmic analysis. According to the proof framework and assumption in chapter 3, we can easily obtain the global convergence of Algorithm 4.1. In this chapter, we will focus on analyzing the convergence rate for Algorithm 4.1. Now, we are prepared to analyze the sub-linear convergence rate of the augmented lagrangian based parallel splitting method under both ergodic and non-ergodic conditions.

4.3.1 Convergence Rate under Ergodic Sense

First, we analyze the convergence rate of Algorithm 4.1 under an ergodic scenario. More specifically, our main objective is to show that Algorithm 4.1 can converge to a $\bar{\mathbf{u}} \in \mathcal{U}$ after at most $\lfloor O(1/\epsilon) \rfloor$ iterations. To begin with, define a sequence $\hat{\mathbf{u}}^{k+1} = (\hat{x}_1^{k+1}, \dots, \hat{x}_m^{k+1}, \hat{\lambda}^{k+1})$ as:

$$\mathbf{x}^{k+1} = \begin{pmatrix} \hat{x}_1^{k+1} \\ \hat{x}_2^{k+1} \\ \vdots \\ \hat{x}_m^{k+1} \end{pmatrix} = \begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \\ \vdots \\ x_m^{k+1} \end{pmatrix} \quad \text{and} \quad \hat{\lambda}^{k+1} = \lambda^k - \beta (\mathcal{A}\mathbf{x}^{k+1} - b), \quad (4.4)$$

where $\mathcal{A} = [A_1, \dots, A_m]$.

Assumption 4.1 *In Problem (1.1), we assume that $\theta_i(x_i)$ is strongly convex with modulus μ_i , $i = 1, \dots, m$, and*

$$0 < \beta \leq \min_{i=1, \dots, m} \left\{ \frac{\mu_i}{4m\delta_{\max}(A_i^\top A_i)} \right\}.$$

Lemma 4.1 *Suppose Assumption 4.1 holds. If $\mathbf{u}^{k+1}$ is generated by Algorithm 4.1 from a given vector $\mathbf{u}^k$, then it satisfies*

$$\begin{aligned} & \theta(\mathbf{x}) - \theta(\mathbf{x}^{k+1}) + \left\langle \mathbf{x} - \mathbf{x}^{k+1}, -A^\top \hat{\lambda}^{k+1} \right\rangle \\ & + \frac{(m+1)}{2} (\|\mathbf{x} - \mathbf{x}^k\|_{G_x}^2 - \|\mathbf{x} - \mathbf{x}^{k+1}\|_{G_x}^2) \geq \frac{1}{2} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_{G_x}^2, \forall \mathbf{x} \in \mathcal{X}. \end{aligned} \quad (4.5)$$

Proof: The first-order optimality condition for the x_i^{k+1} -related subproblem in Algorithm 4.1 is given by

$$\begin{aligned} & \theta_i(x_i) - \theta_i(x_i^{k+1}) - \frac{\mu_i}{2} \|x_i - x_i^{k+1}\|^2 + \langle x_i - x_i^{k+1}, -A_i^\top \lambda^k \rangle \\ & + \left\langle x_i - x_i^{k+1}, \beta A_i^\top \left(A_i x_i^{k+1} + \sum_{j \neq i}^m A_j x_j^k - b \right) \right\rangle \geq 0, \forall x_i \in \mathcal{X}_i. \end{aligned}$$

Then, it follows from $\hat{\lambda}^{k+1} = \lambda^k - \beta(\sum_{i=1}^m A_i x_i^{k+1} - b)$ that

$$\begin{aligned} & \theta_i(x_i) - \theta_i(x_i^{k+1}) - \frac{\mu_i}{2} \|x_i - x_i^{k+1}\|^2 - \left\langle x_i - x_i^{k+1}, A_i^\top \hat{\lambda}^{k+1} \right\rangle \\ & + \left\langle x_i - x_i^{k+1}, \sum_{j \neq i}^m \beta A_i^\top A_j (x_j^k - x_j^{k+1}) \right\rangle \geq 0, \forall x_i \in \mathcal{X}_i, \end{aligned} \quad (4.6)$$

Summing the above inequalities for $i = 1, \dots, m$ yields

$$\begin{aligned} & \theta(\mathbf{x}) - \theta(\mathbf{x}^{k+1}) - \sum_{i=1}^m \frac{\mu_i}{2} \|x_i - x_i^{k+1}\|^2 - \sum_{i=1}^m \left\langle x_i - x_i^{k+1}, A_i^\top \hat{\lambda}^{k+1} \right\rangle \\ & + \beta \sum_{i=1}^m \left\langle x_i - x_i^{k+1}, \sum_{j=1}^m A_i^\top A_j (x_j^k - x_j^{k+1}) \right\rangle \\ & - \beta \sum_{i=1}^m \left\langle x_i - x_i^{k+1}, A_i^\top A_i (x_i^k - x_i^{k+1}) \right\rangle \geq 0. \end{aligned} \quad (4.7)$$

Moreover, making use of the inequality (1.5) leads to

$$\begin{aligned} & \beta \left\langle A_i(x_i - x_i^{k+1}), \sum_{j=1}^m A_j(x_j^k - x_j^{k+1}) \right\rangle \\ & = \beta \left\langle A_i(x_i - x_i^{k+1}), \sum_{j=1}^m A_j(x_j^k - x_j) \right\rangle - \beta \left\langle A_i(x_i - x_i^{k+1}), \sum_{j=1}^m A_j(x_j^{k+1} - x_j) \right\rangle \\ & \leq m\beta \|A_i(x_i^{k+1} - x_i)\|^2 + \frac{\beta}{2} \sum_{j=1}^m \{ \|A_j(x_j^k - x_j)\|^2 + \|A_j(x_j^{k+1} - x_j)\|^2 \}. \end{aligned}$$

Then,

$$\begin{aligned} & \beta \sum_{i=1}^m \left\langle A_i(x_i - x_i^{k+1}), \sum_{j=1}^m A_j(x_j^k - x_j^{k+1}) \right\rangle \\ & \leq m\beta \sum_{i=1}^m \|A_i(x_i^{k+1} - x_i)\|^2 + \frac{m\beta}{2} \sum_{i=1}^m \|A_i(x_i^k - x_i)\|^2 + \frac{m\beta}{2} \sum_{i=1}^m \|A_i(x_i^{k+1} - x_i)\|^2 \\ & = \frac{3m\beta}{2} \sum_{i=1}^m \|A_i(x_i^{k+1} - x_i)\|^2 + \frac{m\beta}{2} \sum_{i=1}^m \|A_i(x_i^k - x_i)\|^2, \end{aligned}$$

and the inequality (4.7) can in turn be transformed into

$$\begin{aligned} \theta(\mathbf{x}) - \theta(\mathbf{x}^{k+1}) - \left\langle \mathbf{x} - \mathbf{x}^{k+1}, A^\top \hat{\lambda}^{k+1} \right\rangle - \sum_{i=1}^m \frac{\mu_i}{2} \|x_i^{k+1} - x_i\|^2 \\ + \sum_{i=1}^m 2m\beta \|A_i(x_i^{k+1} - x_i)\|^2 + \langle \mathbf{x}^{k+1} - \mathbf{x}, G_x(\mathbf{x}^k - \mathbf{x}^{k+1}) \rangle \\ + \sum_{i=1}^m \frac{m\beta}{2} [\|A_i(x_i^k - x_i)\|^2 - \|A_i(x_i^{k+1} - x_i)\|^2] \geq 0. \end{aligned}$$

Since

$$\langle a - b, G(c - a) \rangle = \frac{1}{2} \|b - c\|_G^2 - \frac{1}{2} \|b - a\|_G^2 - \frac{1}{2} \|a - c\|_G^2,$$

we get:

$$\begin{aligned} \langle \mathbf{x}^{k+1} - \mathbf{x}, G_x(\mathbf{x}^k - \mathbf{x}^{k+1}) \rangle \\ = \frac{1}{2} (\|\mathbf{x} - \mathbf{x}^k\|_{G_x}^2 - \|\mathbf{x} - \mathbf{x}^{k+1}\|_{G_x}^2) - \frac{1}{2} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_{G_x}^2. \end{aligned}$$

Therefore,

$$\begin{aligned} \theta(\mathbf{x}) - \theta(\mathbf{x}^{k+1}) \\ - \left\langle \mathbf{x} - \mathbf{x}^{k+1}, A^\top \hat{\lambda}^{k+1} \right\rangle + \frac{(m+1)}{2} (\|\mathbf{x} - \mathbf{x}^k\|_{G_x}^2 - \|\mathbf{x} - \mathbf{x}^{k+1}\|_{G_x}^2) \\ \geq \sum_{i=1}^m \frac{[\mu_i - 4m\beta\delta_{\max}(A_i^\top A_i)]}{2} \|x_i^{k+1} - x_i\|^2 + \frac{1}{2} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_{G_x}^2 \\ = \sum_{i=1}^m \frac{c_i}{2} \|x_i^{k+1} - x_i\|^2 + \frac{1}{2} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_{G_x}^2, \end{aligned} \tag{4.8}$$

where $c_i = \mu_i - 4m\beta\delta_{\max}(A_i^\top A_i)$. Then, the assertion of this lemma is obtained by Assumption 4.1. $\square$

Lemma 4.2 *Let $\mathbf{u}^{k+1}$ be generated by Algorithm 4.1 from the given vector $\mathbf{u}^k$. Then,*

$$\begin{aligned} \langle \lambda - \hat{\lambda}^{k+1}, \hat{\lambda}^{k+1} - \lambda^k \rangle \\ = \frac{1}{2\gamma} (\|\lambda - \lambda^k\|^2 - \|\lambda - \lambda^{k+1}\|^2) + \frac{(\gamma - 2)}{2\gamma^2} \|\lambda^k - \lambda^{k+1}\|^2. \end{aligned} \tag{4.9}$$

Proof: Combining (4.2) and (4.4) leads to

$$\lambda^k - \lambda^{k+1} = \gamma(\lambda^k - \hat{\lambda}^{k+1}), \quad (4.10)$$

and

$$\lambda^{k+1} - \hat{\lambda}^{k+1} = (1 - \gamma)\beta(\mathcal{A}\mathbf{x}^{k+1} - b). \quad (4.11)$$

In addition, it follows from (4.4) that

$$\beta(\mathcal{A}\mathbf{x}^{k+1} - b) = \lambda^k - \hat{\lambda}^{k+1}.$$

Substituting the above equality into (4.11) yields

$$\lambda^{k+1} - \hat{\lambda}^{k+1} = (1 - \gamma)(\lambda^k - \hat{\lambda}^{k+1}) = \frac{1 - \gamma}{\gamma}(\lambda^k - \lambda^{k+1}). \quad (4.12)$$

Therefore,

$$\langle \lambda - \hat{\lambda}^{k+1}, \hat{\lambda}^{k+1} - \lambda^k \rangle = \gamma^{-1} \langle \lambda - \hat{\lambda}^{k+1}, \lambda^{k+1} - \lambda^k \rangle. \quad (4.13)$$

Setting $a = \lambda$, $b = \hat{\lambda}^{k+1}$, $c = \lambda^{k+1}$ and $d = \lambda^k$ in the identity (1.7), we get:

$$\begin{aligned} & \langle \lambda - \hat{\lambda}^{k+1}, \lambda^{k+1} - \lambda^k \rangle \\ &= \frac{1}{2}(\|\lambda - \lambda^k\|^2 - \|\lambda - \lambda^{k+1}\|^2) + \frac{1}{2}(\|\lambda^{k+1} - \hat{\lambda}^{k+1}\|^2 - \|\lambda^k - \hat{\lambda}^{k+1}\|^2) \\ &= \frac{1}{2}(\|\lambda - \lambda^k\|^2 - \|\lambda - \lambda^{k+1}\|^2) + \frac{\gamma(\gamma - 2)}{2}\|\lambda^k - \hat{\lambda}^{k+1}\|^2, \end{aligned} \quad (4.14)$$

where the second equality is derived from (4.12). Then, combining (4.13) and (4.14) leads to

$$\begin{aligned} & \langle \lambda - \hat{\lambda}^{k+1}, \hat{\lambda}^{k+1} - \lambda^k \rangle \\ &= \frac{1}{2\gamma}(\|\lambda - \lambda^k\|^2 - \|\lambda - \lambda^{k+1}\|^2) + \frac{(\gamma - 2)}{2}\|\lambda^k - \hat{\lambda}^{k+1}\|^2, \\ &= \frac{1}{2\gamma}(\|\lambda - \lambda^k\|^2 - \|\lambda - \lambda^{k+1}\|^2) + \frac{(\gamma - 2)}{2\gamma^2}\|\lambda^k - \lambda^{k+1}\|^2, \end{aligned}$$

where the last equality is derived from (4.10), and the statement of this lemma is obtained.

□

Lemma 4.3 *Suppose Assumption 4.1 holds. Let $\mathbf{u}^{k+1}$ be generated by Algorithm 4.1 from a given vector $\mathbf{u}^k$. Then,*

$$\begin{aligned} & \langle \lambda - \hat{\lambda}^{k+1}, \mathcal{A}\mathbf{x}^{k+1} - b \rangle + \frac{1}{2\gamma\beta} (\|\lambda - \lambda^k\|^2 - \|\lambda - \lambda^{k+1}\|^2) \\ & \geq \frac{(2-\gamma)}{2\gamma^2\beta} \|\lambda^k - \lambda^{k+1}\|^2, \forall \lambda \in \mathbb{R}^l. \end{aligned} \quad (4.15)$$

Proof: The equality $\hat{\lambda}^{k+1} = \lambda^k - \beta(\mathcal{A}\mathbf{x}^{k+1} - b)$ can be rewritten as

$$\langle \lambda - \hat{\lambda}^{k+1}, \mathcal{A}\mathbf{x}^{k+1} - b + \beta^{-1}(\hat{\lambda}^{k+1} - \lambda^k) \rangle \geq 0, \forall \lambda \in \mathbb{R}^l.$$

Then, using (4.9) and the above inequality yields

$$\begin{aligned} & \langle \lambda - \hat{\lambda}^{k+1}, \mathcal{A}\mathbf{x}^{k+1} - b \rangle + \frac{1}{2\gamma\beta} (\|\lambda - \lambda^k\|^2 - \|\lambda - \lambda^{k+1}\|^2) \\ & \geq \frac{(2-\gamma)}{2\gamma^2\beta} \|\lambda^k - \lambda^{k+1}\|^2, \forall \lambda \in \mathbb{R}^l. \end{aligned}$$

Hence, the claim of this lemma is obtained. $\square$

Lemma 4.4 *Suppose Assumption 4.1 holds, the sequence $\{\mathbf{u}^{k+1}\}$ generated by Algorithm 4.1 satisfies*

$$\begin{aligned} & \theta(\mathbf{x}) - \theta(\hat{\mathbf{x}}^{k+1}) + \langle \mathbf{u} - \hat{\mathbf{u}}^{k+1}, F(\mathbf{u}) \rangle \\ & + \frac{1}{2} (\|\mathbf{u} - \mathbf{u}^k\|_G^2 - \|\mathbf{u} - \mathbf{u}^{k+1}\|_G^2) \geq 0, \forall \mathbf{u} \in \mathcal{U}. \end{aligned} \quad (4.16)$$

Proof: From (4.4), (4.5) and (4.15), we yield

$$\begin{aligned} & \theta(\mathbf{x}) - \theta(\hat{\mathbf{x}}^{k+1}) + \langle \mathbf{u} - \hat{\mathbf{u}}^{k+1}, F(\hat{\mathbf{u}}^{k+1}) \rangle \\ & + \frac{1}{2} (\|\mathbf{u} - \mathbf{u}^k\|_G^2 - \|\mathbf{u} - \mathbf{u}^{k+1}\|_G^2) \\ & \geq \frac{1}{2} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_{G_x}^2 + \frac{(2-\gamma)}{2\beta\gamma^2} \|\lambda^k - \lambda^{k+1}\|^2 + \sum_{i=1}^m c_i \|x_i - x_i^{k+1}\|^2, \forall \mathbf{u} \in \mathcal{U}, \end{aligned} \quad (4.17)$$

where $\hat{\mathbf{u}}^{k+1}$ is as defined in (4.4). Then, utilizing the monotonicity of the operator $F(\cdot)$ and dropping the nonnegative terms in the right hand side of (4.17), we obtain the assertion of this Lemma. $\square$

In the following, we set out to certify the worst-case $O(1/t)$ convergence rate of Algorithm 4.1 under an ergodic scenario .

Theorem 4.1 *Suppose Assumption 4.1 holds. Let the sequence $\{\mathbf{u}^k\}$ be generated by Algorithm 4.1. For a positive integer t , we yield*

$$\theta(\bar{\mathbf{x}}_t) - \theta(\mathbf{x}) + \langle \bar{\mathbf{u}}_t - \mathbf{u}, F(\mathbf{u}) \rangle \leq \frac{1}{2(t+1)} \|\mathbf{u} - \mathbf{u}^0\|_G^2, \forall \mathbf{u} \in \mathcal{W}, \quad (4.18)$$

where

$$\bar{\mathbf{u}}_t := \frac{1}{t+1} \sum_{k=0}^t \hat{\mathbf{u}}^{k+1}.$$

Proof: Summing the inequality (4.16) corresponding to $k = 0, 1, \dots, t$, we derive that

$$\begin{aligned} & \left((t+1)\theta(\mathbf{x}) - \sum_{k=0}^t \theta(\hat{\mathbf{x}}^{k+1}) \right) + \left\langle (t+1)\mathbf{u} - \sum_{k=0}^t \hat{\mathbf{u}}^{k+1}, F(\mathbf{u}) \right\rangle \\ & + \frac{1}{2} \|\mathbf{u} - \mathbf{u}^0\|_G^2 \geq 0, \forall \mathbf{u} \in \mathcal{W}. \end{aligned} \quad (4.19)$$

Making use of the definition of $\bar{\mathbf{u}}_t$, we reformulate the inequality (4.19) into

$$\frac{1}{t+1} \sum_{k=0}^t \theta(\hat{\mathbf{x}}^{k+1}) - \theta(\mathbf{x}) + \langle \bar{\mathbf{u}}_t - \mathbf{u}, F(\mathbf{u}) \rangle \leq \frac{1}{2(t+1)} \|\mathbf{u} - \mathbf{u}^0\|_G^2, \forall \mathbf{u} \in \mathcal{W}.$$

Since t is a positive integer, we can conclude that $\hat{\mathbf{x}}^k \in \mathcal{X}$ for $k = 0, 1, \dots, t$, and then, $\bar{\mathbf{x}}_t = \frac{1}{t+1} (\sum_{k=0}^t \hat{\mathbf{x}}^{k+1})$ is a convex combination of the $\hat{\mathbf{x}}^{k+1}$'s. As $\theta(\mathbf{x})$ is a convex function, we obtain $\bar{\mathbf{x}}_t := \frac{1}{t+1} \sum_{k=0}^t \hat{\mathbf{x}}^{k+1} \in \mathcal{U}$ and

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

Then, combining the above inequalities leads to the inequality (4.18), which completes the proof. $\square$

For any given compact set $\mathcal{D}_{\mathcal{U}}(\bar{\mathbf{u}}_t) \subset \mathcal{U}$, define

$$\bar{d} = \sup\{\|\mathbf{u} - \mathbf{u}^0\|_G^2 | \mathbf{u} \in \mathcal{D}_{\mathcal{U}}(\bar{\mathbf{u}}_t)\},$$

where $\mathbf{u}^0$ is an initialization point. Hence, based on Assumption 4.1, Theorem 4.1 implies that we can find a $\bar{\mathbf{u}}_t \in \mathcal{U}$ such that

$$\sup_{\mathbf{u} \in \mathcal{D}_{\mathcal{U}}(\bar{\mathbf{u}}_t)} \{\theta(\bar{\mathbf{x}}_t) - \theta(\mathbf{x}) + \langle \bar{\mathbf{u}}_t - \mathbf{u}, F(\mathbf{u}) \rangle\} \leq \frac{\bar{d}}{2(t+1)},$$

and the worse case $O(1/t)$ convergence rate of Algorithm 4.1 is thus obtained under an ergodic senario.

4.3.2 Convergence Rate under Non-ergodic Sense

In this subsection, we are prepared to deduce the worst-case $O(1/k)$ convergence rate under a non-ergodic scenario for Algorithm 4.1 by making the following assumption.

Assumption 4.2 *In Problem (1.1), we assume that $\theta_i(x_i)$ is strongly convex with modulus μ_i , $i = 1, \dots, m$, and the β in Algorithm 4.1 is defined as*

$$0 < \beta \leq \min_{i=1, \dots, m} \left\{ \frac{\mu_i}{2m\delta_{\max}(A_i^\top A_i)} \right\}.$$

Lemma 4.5 *Suppose Assumption 4.2 holds, and let $\mathbf{u}^{k+1}$ be generated by Algorithm 4.1 from a given vector $\mathbf{u}^k$. Then, we get*

$$\begin{aligned} & \frac{1}{\gamma\beta} \|\lambda^{k+1} - \lambda^k\| + (m-1) \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_{G_x}^2 \\ & \leq \frac{1}{\gamma\beta} \|\lambda^k - \lambda^{k-1}\| + (m-1) \|\mathbf{x}^k - \mathbf{x}^{k-1}\|_{G_x}. \end{aligned} \quad (4.20)$$

Proof: Setting $x_i = x_i^k$ in (4.6) yields

$$\begin{aligned} & \theta_i(x_i^k) - \theta_i(x_i^{k+1}) - \frac{\mu_i}{2} \|x_i^k - x_i^{k+1}\|^2 - \left\langle x_i^k - x_i^{k+1}, A_i^\top \hat{\lambda}^{k+1} \right\rangle \\ & + \left\langle x_i^k - x_i^{k+1}, \sum_{j \neq i}^n \beta A_i^\top A_j (x_j^k - x_j^{k+1}) \right\rangle \geq 0. \end{aligned} \quad (4.21)$$

Similarly, setting $x_i = x_i^{k+1}$ for the k -th iteration yields

$$\begin{aligned} & \theta_i(x_i^{k+1}) - \theta_i(x_i^k) - \frac{\mu_i}{2} \|x_i^{k+1} - x_i^k\|^2 - \left\langle x_i^{k+1} - x_i^k, A_i^\top \hat{\lambda}^k \right\rangle \\ & + \left\langle x_i^{k+1} - x_i^k, \sum_{j \neq i}^n \beta A_i^\top A_j (x_j^{k-1} - x_j^k) \right\rangle \geq 0. \end{aligned} \quad (4.22)$$

Adding (4.21) and (4.22) leads to

$$\begin{aligned} & \left\langle x_i^{k+1} - x_i^k, A_i^\top (\hat{\lambda}^{k+1} - \hat{\lambda}^k) \right\rangle + \left\langle A_i(x_i^k - x_i^{k+1}), \sum_{j \neq i}^n \beta A_j(x_j^k - x_j^{k+1}) \right\rangle \\ & - \left\langle A_i(x_i^k - x_i^{k+1}), \sum_{j \neq i}^n \beta A_j(x_j^{k-1} - x_j^k) \right\rangle \geq \mu_i \|x_i^{k+1} - x_i^k\|^2. \end{aligned}$$

Summing the above inequalities for $i = 1, \dots, m$, yields

$$\begin{aligned} & \sum_{i=1}^m \left\langle A_i(x_i^{k+1} - x_i^k), \hat{\lambda}^{k+1} - \hat{\lambda}^k \right\rangle + \sum_{i=1}^m \left\langle A_i(x_i^k - x_i^{k+1}), \sum_{j \neq i}^m \beta A_j(x_j^k - x_j^{k+1}) \right\rangle \\ & - \sum_{i=1}^m \left\langle A_i(x_i^k - x_i^{k+1}), \sum_{j \neq i}^m \beta A_j(x_j^{k-1} - x_j^k) \right\rangle \geq \sum_{i=1}^m \mu_i \|x_i^{k+1} - x_i^k\|^2. \end{aligned} \quad (4.23)$$

Furthermore, it follows from (4.2) and (4.4) that

$$\sum_{i=1}^m A_i(x_i^{k+1} - x_i^k) = \frac{1}{\gamma\beta} ((\lambda^k - \lambda^{k+1}) - (\lambda^{k-1} - \lambda^k)), \quad (4.24)$$

and

$$\hat{\lambda}^{k+1} - \hat{\lambda}^k = \lambda^k - \lambda^{k-1} - \sum_{i=1}^m \beta A_i(x_i^{k+1} - x_i^k).$$

The first summation term in the left-hand side of (4.23) can be reformulated into

$$\begin{aligned} & \left\langle \sum_{i=1}^m A_i(x_i^{k+1} - x_i^k), \hat{\lambda}^{k+1} - \hat{\lambda}^k \right\rangle \\ & = \frac{1}{\gamma\beta} \langle ((\lambda^k - \lambda^{k+1}) - (\lambda^{k-1} - \lambda^k)), \lambda^k - \lambda^{k-1} \rangle - \beta \|\mathcal{A}\mathbf{x}^{k+1} - \mathcal{A}\mathbf{x}^k\|^2 \\ & = \frac{1}{2\gamma\beta} \{ \|\lambda^{k-1} - \lambda^k\|^2 - \|\lambda^k - \lambda^{k+1}\|^2 + \|(\lambda^k - \lambda^{k+1}) - (\lambda^{k-1} - \lambda^k)\|^2 \} \\ & \quad - \beta \|\mathcal{A}\mathbf{x}^{k+1} - \mathcal{A}\mathbf{x}^k\|^2 \\ & = \frac{1}{2\gamma\beta} \{ \|\lambda^{k-1} - \lambda^k\|^2 - \|\lambda^k - \lambda^{k+1}\|^2 \} - \frac{(2-\gamma)\beta}{2} \|\mathcal{A}\mathbf{x}^{k+1} - \mathcal{A}\mathbf{x}^k\|^2. \end{aligned} \quad (4.25)$$

Additionally, the second summation term in the left-hand side of (4.23) implies

$$\begin{aligned} & \sum_{i=1}^m \beta \left\langle A_i(x_i^{k+1} - x_i^k), \sum_{j \neq i}^m A_j(x_j^k - x_j^{k+1}) \right\rangle \\ & = \sum_{i=1}^m \beta \left\langle A_i(x_i^{k+1} - x_i^k), \sum_{j=1}^m A_j(x_j^k - x_j^{k+1}) \right\rangle - \sum_{i=1}^m \beta \langle A_i(x_i^{k+1} - x_i^k), A_i(x_i^k - x_i^{k+1}) \rangle \\ & = \sum_{i=1}^m \beta \|A_i(x_i^k - x_i^{k+1})\|^2 - \beta \left\| \sum_{j=1}^m A_j(x_j^k - x_j^{k+1}) \right\|^2 \\ & \geq \sum_{i=1}^m \beta \|A_i(x_i^k - x_i^{k+1})\|^2 - m\beta \sum_{j=1}^m \|A_j(x_j^k - x_j^{k+1})\|^2. \end{aligned} \quad (4.26)$$

Furthermore, utilizing the inequality (1.5) leads to

$$\begin{aligned}
& \sum_{i=1}^m \beta \left\langle A_i(x_i^k - x_i^{k-1}), \sum_{j \neq i}^m A_j(x_j^k - x_j^{k+1}) \right\rangle \\
& \geq -\frac{\beta}{2} \sum_{i=1}^m \left\{ (m-1) \|A_i(x_i^{k-1} - x_i^k)\|^2 + \sum_{j \neq i}^m \|A_j(x_j^k - x_j^{k+1})\|^2 \right\} \\
& = -\frac{\beta(m-1)}{2} \sum_{i=1}^m \{ \|A_i(x_i^{k-1} - x_i^k)\|^2 + \|A_i(x_i^k - x_i^{k+1})\|^2 \}. \tag{4.27}
\end{aligned}$$

Combining (4.26) and (4.27) leads to

$$\begin{aligned}
& \sum_{i=1}^m \left\langle A_i(x_i^{k+1} - x_i^k), \sum_{j \neq i}^m \beta A_j(x_j^k - x_j^{k+1}) \right\rangle \\
& - \sum_{i=1}^m \left\langle A_i(x_i^{k+1} - x_i^k), \sum_{j \neq i}^m \beta A_j(x_j^{k-1} - x_j^k) \right\rangle \\
& \geq \sum_{i=1}^m (1-m) \beta \|A_i(x_i^k - x_i^{k+1})\|^2 \\
& - \frac{\beta(m-1)}{2} \sum_{i=1}^m \{ \|A_i(x_i^{k-1} - x_i^k)\|^2 + \|A_i(x_i^k - x_i^{k+1})\|^2 \} \\
& = \frac{\beta(m-1)}{2} \sum_{i=1}^m \{ \|A_i(x_i^k - x_i^{k+1})\|^2 - \|A_i(x_i^{k-1} - x_i^k)\|^2 \} \\
& - 2(m-1) \beta \sum_{j=1}^m \|A_j(x_j^k - x_j^{k+1})\|^2. \tag{4.28}
\end{aligned}$$

Finally, combining (4.23), (4.25), and (4.28) yields

$$\begin{aligned}
& \frac{1}{2\gamma\beta} \{ \|\lambda^{k-1} - \lambda^k\|^2 - \|\lambda^k - \lambda^{k+1}\|^2 \} \\
& \geq \frac{\beta(m-1)}{2} \sum_{i=1}^m \{ \|A_i(x_i^k - x_i^{k+1})\|^2 - \|A_i(x_i^{k-1} - x_i^k)\|^2 \} \\
& + \sum_{i=1}^m \{ \mu_i \|x_i^k - x_i^{k+1}\|^2 - 2(m-1) \beta \|A_i(x_i^k - x_i^{k+1})\|^2 \} \\
& + \frac{(2-\gamma)\beta}{2} \|\mathcal{A}\mathbf{x}^k - \mathcal{A}\mathbf{x}^{k+1}\|^2.
\end{aligned}$$

Since $\gamma \in (0, 2)$, dropping some non-negative terms in the right hand side, the above inequality can be rewritten as

$$\begin{aligned} & \frac{1}{2\gamma\beta} \|\lambda^{k-1} - \lambda^k\|^2 + \frac{(m-1)}{2} \|\mathbf{x}^{k-1} - \mathbf{x}^k\|_{G_x}^2 \\ & \geq \frac{1}{2\gamma\beta} \|\lambda^k - \lambda^{k+1}\|^2 + \frac{(m-1)}{2} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_{G_x}^2 \\ & \quad + \sum_{i=1}^m (\mu_i - 2(m-1)\beta\delta_{\max}(A_i^\top A_i)) \|x_i^k - x_i^{k+1}\|^2. \end{aligned} \quad (4.29)$$

Hence, under the Assumption 4.2, we obtain

$$\mu_i - 2(m-1)\beta\delta_{\max}(A_i^\top A_i) \geq 0$$

and

$$\begin{aligned} & \frac{1}{\gamma\beta} \|\lambda^{k-1} - \lambda^k\|^2 + (m-1) \|\mathbf{x}^{k-1} - \mathbf{x}^k\|_{G_x}^2 \\ & \geq \frac{1}{\gamma\beta} \|\lambda^k - \lambda^{k+1}\|^2 + (m-1) \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_{G_x}^2. \end{aligned}$$

Hence, the statement of this lemma is obtained. $\square$

Lemma 4.6 *Suppose Assumption 4.2 holds, the sequence $\{\mathbf{u}^{k+1}\}$ be generated by Algorithm 4.1. satisfies*

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_G^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|_G^2 - \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_{G_x}^2 - \frac{2-\gamma}{\beta\gamma^2} \|\lambda^k - \lambda^{k+1}\|^2, \quad (4.30)$$

where $\mathbf{u}^*$ is a solution point of problem (1.1).

Proof: Setting $\mathbf{x} = \mathbf{x}^*$ in (4.17) and using $\mathbf{x}^{k+1} = \hat{\mathbf{x}}^{k+1}$ yields

$$\begin{aligned} & \theta(\mathbf{x}^*) - \theta(\mathbf{x}^{k+1}) \\ & - \left\langle \mathbf{x}^* - \mathbf{x}^{k+1}, \mathcal{A}^\top \hat{\lambda}^{k+1} \right\rangle + \frac{(m+1)}{2} (\|\mathbf{x}^* - \mathbf{x}^k\|_{G_x}^2 - \|\mathbf{x}^* - \mathbf{x}^{k+1}\|_{G_x}^2) \\ & \geq \sum_{i=1}^m \frac{[\mu_i - 4m\beta\delta_{\max}(A_i^\top A_i)]}{2} \|x_i^{k+1} - x_i^*\|^2 + \frac{1}{2} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_{G_x}^2. \end{aligned}$$

As $\mathbf{u}^*$ is a solution point, it follows from (1.11) that

$$\theta(\mathbf{x}^{k+1}) - \theta(\mathbf{x}^*) - \langle \mathbf{x}^{k+1} - \mathbf{x}^*, \mathcal{A}^\top \lambda^* \rangle \geq \sum_{i=1}^m \frac{\mu_i}{2} \|x_i^* - x_i^{k+1}\|^2.$$

Combining the above two inequalities yields:

$$\begin{aligned} & \frac{(m+1)}{2} (\|\mathbf{x}^* - \mathbf{x}^k\|_{G_x}^2 - \|\mathbf{x}^* - \mathbf{x}^{k+1}\|_{G_x}^2) + \left\langle \mathbf{x}^{k+1} - \mathbf{x}^*, \mathcal{A}^\top (\hat{\lambda}^{k+1} - \lambda^*) \right\rangle \\ & \geq \sum_{i=1}^m [\mu_i - 2m\beta\delta_{\max}(A_i^\top A_i)] \|x_i^{k+1} - x_i^*\|^2 + \frac{1}{2} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_{G_x}^2. \end{aligned}$$

Then, setting $\lambda = \lambda^*$ in (4.15) yields

$$\begin{aligned} & \langle \lambda^* - \hat{\lambda}^{k+1}, \mathcal{A}\mathbf{x}^{k+1} - b \rangle + \frac{1}{2\gamma\beta} (\|\lambda^* - \lambda^k\|^2 - \|\lambda^* - \lambda^{k+1}\|^2) \\ & \geq \frac{(2-\gamma)}{2\gamma^2\beta} \|\lambda^k - \lambda^{k+1}\|^2. \end{aligned}$$

Adding the above two inequalities leads to

$$\begin{aligned} \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_G^2 & \leq \|\mathbf{u}^k - \mathbf{u}^*\|_G^2 - \sum_{i=1}^m 2h_i \|x_i^{k+1} - x_i^*\|^2 \\ & \quad - \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_{G_x}^2 - \frac{2-\gamma}{\beta\gamma^2} \|\lambda^k - \lambda^{k+1}\|^2, \end{aligned} \quad (4.31)$$

where $h_i = \mu_i - 2m\beta\delta_{\max}(A_i^\top A_i)$. Using the bounding value of β (from Assumption 4.2) in (4.31) leads to (4.30), and this lemma is proved. $\square$

Theorem 4.2 *Suppose Assumption 4.2 holds. Let the sequence $\{\mathbf{u}^k\}$ be generated by Algorithm 4.1. For a positive integer t , we have*

$$\frac{1}{\gamma\beta} \|\lambda^t - \lambda^{t+1}\|^2 + (m-1) \|\mathbf{x}^t - \mathbf{x}^{t+1}\|_{G_x}^2 = O(1/t). \quad (4.32)$$

Proof: To begin with, it follows from (4.30) that

$$\begin{aligned} & \sigma \left\{ \frac{1}{\gamma\beta} \|\lambda^{k+1} - \lambda^k\|^2 + (m-1) \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_{G_x}^2 \right\} \\ & \leq \frac{(2-\gamma)}{\gamma^2\beta} \|\lambda^k - \lambda^{k+1}\|^2 + \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_{G_x}^2 \\ & \leq \|\mathbf{u}^k - \mathbf{u}^*\|_G^2 - \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_G^2, \end{aligned} \quad (4.33)$$

where

$$\sigma_1 = \max \left\{ \frac{1}{\gamma\beta}, m-1 \right\}, \sigma_2 = \min \left\{ \frac{2-\gamma}{\gamma^2\beta}, 1 \right\}, \sigma = \frac{\sigma_2}{\sigma_1} > 0.$$

Summing the inequality (4.33) associated with $k = 0, \dots, t$, obtains

$$\begin{aligned} & \sum_{k=0}^t \sigma \left\{ \frac{1}{\gamma\beta} \|\lambda^{k+1} - \lambda^k\|^2 + (m-1) \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_{G_x}^2 \right\} \\ & \leq \|\mathbf{u}^0 - \mathbf{u}^*\|_G^2. \end{aligned} \quad (4.34)$$

Then, using Lemma 4.5 and (4.34) leads to

$$\frac{1}{\gamma\beta} \|\lambda^t - \lambda^{t+1}\|^2 + (m-1) \|\mathbf{x}^t - \mathbf{x}^{t+1}\|_{G_x}^2 \leq \frac{1}{\sigma(t+1)} \|\mathbf{u}^0 - \mathbf{u}^*\|_G^2, \quad (4.35)$$

which proves this lemma. $\square$

The above theorem demonstrates that the worst-case $O(1/k)$ convergence rate is proved under a nonergodic scenario for Algorithm 4.1. Furthermore, following the works [13] and [97], we can easily refine this convergence rate by an order of magnitude from $O(1/k)$ to $o(1/k)$, and the following theorem validates this claim under a nonergodic scenario.

Theorem 4.3 *Suppose Assumption 4.2 holds. Let the sequence $\{\mathbf{u}^k\}$ be generated by Algorithm 4.1. For a positive integer t , we yield*

$$\frac{1}{\gamma\beta} \|\lambda^t - \lambda^{t+1}\|^2 + (m-1) \|\mathbf{x}^t - \mathbf{x}^{t+1}\|_{G_x}^2 = o(1/t). \quad (4.36)$$

Proof: Setting $R_k = \frac{1}{\gamma\beta} \|\lambda^k - \lambda^{k+1}\|^2 + (m-1) \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_{G_x}^2$. It follows from Lemma 4.5 and (4.33) that the sequence $\{R_k\}$ is monotonically non-increasing and $\sum_{k=0}^{\infty} R_k \leq +\infty$, respectively. Then,

$$tR_{2t} \leq R_{t+1} + R_{t+2} + \dots + R_{2t} \rightarrow 0, \text{ as } t \rightarrow \infty.$$

Then, we conclude that

$$\frac{1}{\gamma\beta} \|\lambda^t - \lambda^{t+1}\|^2 + (m-1) \|\mathbf{x}^t - \mathbf{x}^{t+1}\|_{G_x}^2 = o(1/t).$$

This assertion of this theorem is obtained. $\square$

4.4 Linear Convergence

Next, we place emphasis on analyzing the *global linear convergence* of Algorithm 4.1 under the following assumption.

Assumption 4.3 *In Problem (1.1), we assume that $\theta_i(x_i)$ is strongly convex with modulus μ_i , $i = 1, \dots, m$, A_1 is full row rank, $\mathcal{X}_1 = \mathbb{R}^{n_1}$, θ_1 is differentiable and $\nabla\theta_1$ is Lipschitz continuous with modulus l_1 , and β in Algorithm 4.1 is defined as*

$$0 < \beta < \min_{i=1, \dots, m} \left\{ \frac{\mu_i}{2m\delta_{\max}(A_i^\top A_i)} \right\}.$$

As all subproblems can be solved in parallel, our analysis is clearly suitable for the case when each of $\nabla\theta_i(x_i)$ is Lipschitz continuous and A_i is full row rank, $\forall i = 1, \dots, m$. Therefore, in the following, we only discuss the case that $\nabla\theta_1(x_1)$ is Lipschitz continuous and A_1 is full row rank.

Lemma 4.7 *Suppose Assumption 4.3 holds. Let $\mathbf{u}^{k+1}$ be generated by Algorithm 4.1 from a given vector $\mathbf{u}^k$. Thus, we yield*

$$\begin{aligned} \|\lambda^{k+1} - \lambda^*\|^2 &\leq \frac{2(1-\gamma)^2}{\gamma^2} \|\lambda^{k+1} - \lambda^k\|^2 + 4\kappa l_1^2 \|x_1^{k+1} - x_1^*\|^2 \\ &\quad + 4(m-1)\kappa\beta^2\delta_{\max}(A_1 A_1^\top) \sum_{j=2}^m \|A_j(x_j^{k+1} - x_j^k)\|^2. \end{aligned} \quad (4.37)$$

Proof: Under Assumption 4.3, the optimality condition of x_1 -subproblem can also be rewritten as

$$\nabla\theta_1(x_1^{k+1}) - A_1^\top \hat{\lambda}^{k+1} + \sum_{j=2}^m \beta A_1^\top A_j (x_j^{k+1} - x_j^k) = 0,$$

and

$$\nabla\theta_1(x_1^*) - A_1^\top \lambda^* = 0.$$

Combining the above two equalities leads to

$$\begin{aligned}
& \|A_1^\top(\hat{\lambda}^{k+1} - \lambda^*)\|^2 \\
&= \left\| \nabla\theta_1(x_1^{k+1}) - \nabla\theta_1(x_1^*) + \sum_{j=2}^m \beta A_1^\top A_j(x_j^{k+1} - x_j^k) \right\|^2 \\
&\leq 2 \left\| \nabla\theta_1(x_1^{k+1}) - \nabla\theta_1(x_1^*) \right\|^2 + 2 \left\| \sum_{j=2}^m \beta A_1^\top A_j(x_j^{k+1} - x_j^k) \right\|^2 \\
&\leq 2l_1^2 \|x_1^{k+1} - x_1^*\|^2 + 2\beta^2 \delta_{\max}(A_1 A_1^\top) \left\| \sum_{j=2}^m A_j(x_j^{k+1} - x_j^k) \right\|^2 \\
&\leq 2l_1^2 \|x_1^{k+1} - x_1^*\|^2 + 2(m-1)\beta^2 \delta_{\max}(A_1 A_1^\top) \sum_{j=2}^m \|A_j(x_j^{k+1} - x_j^k)\|^2 \tag{4.38}
\end{aligned}$$

where the second inequality is inferred by the Lipschitzian property of the gradient of the first objective function. It follows from the inequalities

$$\delta_{\min}(A_1 A_1^\top) \|\hat{\lambda}^{k+1} - \lambda^*\|^2 \leq \|A_1^\top(\hat{\lambda}^{k+1} - \lambda^*)\|^2 \leq \delta_{\max}(A_1 A_1^\top) \|\hat{\lambda}^{k+1} - \lambda^*\|^2$$

that

$$\|\hat{\lambda}^{k+1} - \lambda^*\|^2 \leq \kappa \|A_1^\top(\hat{\lambda}^{k+1} - \lambda^*)\|^2,$$

where $\kappa = \delta_{\min}(A_1 A_1^\top)^{-1}$ and A_1 is full row rank. Hence,

$$\begin{aligned}
\|\lambda^{k+1} - \lambda^*\|^2 &\leq 2\|\lambda^{k+1} - \hat{\lambda}^{k+1}\|^2 + 2\|\hat{\lambda}^{k+1} - \lambda^*\|^2 \\
&\leq \frac{2(1-\gamma)^2}{\gamma^2} \|\lambda^{k+1} - \lambda^k\|^2 + 2\kappa \|A_1^\top(\hat{\lambda}^{k+1} - \lambda^*)\|^2 \\
&\leq \frac{2(1-\gamma)^2}{\gamma^2} \|\lambda^{k+1} - \lambda^k\|^2 + 4\kappa l_1^2 \|x_1^{k+1} - x_1^*\|^2 \\
&\quad + 4(m-1)\kappa\beta^2 \delta_{\max}(A_1 A_1^\top) \sum_{j=2}^m \|A_j(x_j^{k+1} - x_j^k)\|^2,
\end{aligned}$$

where the second inequality and the third inequality are derived from (4.12) and (4.38), respectively. $\square$

Lemma 4.8 *Suppose Assumption 4.3 holds. Let $\mathbf{u}^{k+1}$ be generated by Algorithm 4.1 from a given vector $\mathbf{u}^k$. Then, given $\delta > 0$, we have*

$$\begin{aligned}
& \sigma \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_G^2 \\
& \leq \sum_{j=2}^m (m+1) \sigma \beta \|A_j(x_j^{k+1} - x_j^*)\|^2 + \frac{2\sigma(1-\gamma)^2}{\gamma^3 \beta} \|\lambda^{k+1} - \lambda^k\|^2 \\
& \quad + \sum_{j=2}^m \sigma \left[\frac{4(m-1)\kappa\beta\delta_{\max}(A_1 A_1^\top)}{\gamma} \right] \|A_j(x_j^{k+1} - x_j^k)\|^2 \\
& \quad + \sigma \left[(m+1)\beta\delta_{\max}(A_1^\top A_1) + \frac{4\kappa l_1^2}{\gamma\beta} \right] \|x_1^{k+1} - x_1^*\|^2. \tag{4.39}
\end{aligned}$$

Proof: By the definition of G , we yield

$$\begin{aligned}
& \sigma \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_G^2 \\
& = \sigma \left[\sum_{i=1}^m (m+1) \beta \|A_i(x_i^{k+1} - x_i^*)\|^2 + \frac{1}{\gamma\beta} \|\lambda^{k+1} - \lambda^*\|^2 \right].
\end{aligned}$$

Then,

$$\begin{aligned}
& \sigma \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_G^2 \\
& \leq \sum_{i=1}^m (m+1) \sigma \beta \|A_i(x_i^{k+1} - x_i^*)\|^2 + \frac{2\sigma(1-\gamma)^2}{\gamma^3 \beta} \|\lambda^{k+1} - \lambda^k\|^2 \\
& \quad + \frac{4\sigma\kappa l_1^2}{\gamma\beta} \|x_1^{k+1} - x_1^*\|^2 + \sum_{j=2}^m \sigma \left[\frac{4(m-1)\kappa\beta\delta_{\max}(A_1 A_1^\top)}{\gamma} \right] \|A_j(x_j^{k+1} - x_j^k)\|^2 \\
& \leq \sum_{j=2}^m (m+1) \sigma \beta \|A_j(x_j^{k+1} - x_j^*)\|^2 \\
& \quad + \sum_{j=2}^m \sigma \left[\frac{4(m-1)\kappa\beta\delta_{\max}(A_1 A_1^\top)}{\gamma} \right] \|A_j(x_j^{k+1} - x_j^k)\|^2 \\
& \quad + \sigma \left[(m+1)\beta\lambda_{\max}(A_1^\top A_1) + \frac{4\kappa l_1^2}{\gamma\beta} \right] \|x_1^{k+1} - x_1^*\|^2 + \frac{2\sigma(1-\gamma)^2}{\gamma^3 \beta} \|\lambda^{k+1} - \lambda^k\|^2,
\end{aligned}$$

where the first inequality is derived from the inequality (4.37). $\square$

Theorem 4.4 *Suppose Assumption 4.3 holds. Let $\mathbf{u}^{k+1}$ be generated by Algorithm 4.1 from a given vector $\mathbf{u}^k$. Then,*

$$(1 + \sigma) \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_G^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|_G^2, \quad (4.40)$$

where $\sigma := \min\{\sigma_1, \sigma_2, \sigma_3, \sigma_4\} > 0$ and

$$\begin{cases} \sigma_1 := \frac{2\gamma\beta[\mu_1 - 2m\beta\delta_{\max}(A_1^\top A_1)]}{\gamma\beta^2(m+1)\delta_{\max}(A_1^\top A_1) + 4\kappa l_1^2}, \\ \sigma_2 := \min_{j=2, \dots, m} \left\{ \frac{2[\mu_j - 2m\beta\delta_{\max}(A_j^\top A_j)]}{(m+1)\beta\delta_{\max}(A_j^\top A_j)} \right\}, \\ \sigma_3 := \frac{\gamma}{4(m-1)\kappa\delta_{\max}(A_1 A_1^\top)}, \\ \sigma_4 := \frac{(2-\gamma)\gamma}{2(1-\gamma)^2}. \end{cases}$$

Proof: For one thing, utilizing the (4.39) leads to

$$\begin{aligned} & \sum_{i=1}^m 2h_i \|x_i^{k+1} - x_i^*\|^2 \\ & + \sum_{i=1}^m \beta \|A_i(x_i^{k+1} - x_i^k)\|^2 + \frac{2-\gamma}{\beta\gamma^2} \|\lambda^k - \lambda^{k+1}\|^2 - \sigma \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_G^2 \\ & \geq \sum_{i=1}^m 2h_i \|x_i^{k+1} - x_i^*\|^2 + \sum_{i=1}^m \beta \|A_i(x_i^{k+1} - x_i^k)\|^2 + \frac{2-\gamma}{\beta\gamma^2} \|\lambda^k - \lambda^{k+1}\|^2 \\ & - \sum_{j=2}^m (m+1)\sigma\beta \|A_j(x_j^{k+1} - x_j^*)\|^2 - \sum_{j=2}^m \frac{4\sigma(m-1)\kappa\beta\delta_{\max}(A_1 A_1^\top)}{\gamma} \|A_j(x_j^{k+1} - x_j^k)\|^2 \\ & - \sigma \left[(m+1)\beta\delta_{\max}(A_1^\top A_1) + \frac{4\kappa l_1^2}{\gamma\beta} \right] \|x_1^{k+1} - x_1^*\|^2 - \frac{2\sigma(1-\gamma)^2}{\gamma^3\beta} \|\lambda^{k+1} - \lambda^k\|^2 \\ & \geq \left\{ 2h_1 - \sigma \left[(m+1)\beta\delta_{\max}(A_1^\top A_1) + \frac{4\kappa l_1^2}{\gamma\beta} \right] \right\} \|x_1^{k+1} - x_1^*\|^2 \\ & + \sum_{j=2}^m [2h_j - \sigma(m+1)\beta\delta_{\max}(A_j^\top A_j)] \|x_j^{k+1} - x_j^*\|^2 \\ & + \sum_{j=2}^m \left[\beta - \frac{4\sigma(m-1)\kappa\beta\delta_{\max}(A_1 A_1^\top)}{\gamma} \right] \|A_j(x_j^{k+1} - x_j^k)\|^2 \\ & + \left[\frac{2-\gamma}{\beta\gamma^2} - \frac{2\sigma(1-\gamma)^2}{\gamma^3\beta} \right] \|\lambda^{k+1} - \lambda^k\|^2. \end{aligned}$$

Then, using the definition of σ yields

$$\begin{aligned} & \sum_{i=1}^m 2h_i \|x_i^{k+1} - x_i^*\|^2 \\ & + \sum_{i=1}^m \beta \|A_i(x_i^{k+1} - x_i^k)\|^2 + \frac{2-\gamma}{\beta\gamma^2} \|\lambda^k - \lambda^{k+1}\|^2 \geq \sigma \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_G^2. \end{aligned} \quad (4.41)$$

Hence, adding (4.31) and (4.41) proves the assertion of this theorem. $\square$

Under Assumption 4.3, Theorem 4.4 implies that Algorithm 4.1 converges globally linearly to a solution point. It is worthwhile to note that if we set $\gamma = 1$, the $\sigma_4 \rightarrow +\infty$. Then, σ_4 can be eliminated and σ can be simplified to: $\sigma := \min\{\sigma_1, \sigma_2, \sigma_3\} > 0$.

4.5 Conclusions

In this chapter, we put forward a flexible parallel splitting method for solving the considered multiple-block convex minimization problem. The worst-case $O(1/t)$ convergence rate of the suggested algorithm was also established under both ergodic and non-ergodic conditions, and the worst-case $O(1/t)$ convergence rate was further improved to the order $o(1/k)$. Additionally, we deduced the global linear convergence of the suggested algorithm under the additional assumption that the gradient of one of objective functions is Lipschitz continuous.

Chapter 5

Proximal Partially Parallel Splitting Algorithm (PPSA)

5.1 Motivation

As stated in our research scope and objectives, for solving the convex programming problem (1.1), where $m \geq 3$, a predictor-corrector strategy is often adopted by Gauss-Seidel and Jacobian splitting methods to ensure global convergence. In [67, 72, 86], it has been demonstrated that splitting methods equipped with an additional corrector step perform reliably in practice, while requiring only an incrementally greater computational effort. Recently, some novel splitting methods combining Gauss-Seidel and Jacobian decompositions have been proposed, and the efficiency of these methods was further verified in a series of numerical experiments (see [56, 73, 89, 121]).

Motivated by these results, in this chapter, we put forward a new *proximal partially-parallel splitting algorithm* for the convex minimization problem (1.1). The proposed algorithm firstly generates the prediction points by utilizing a similar decomposition strategy to the one displayed in [73], and then the new iterative points is produced by another corrector step. Therefore, the proposed splitting method is a *hybrid mechanism* and inherits the nice features of Gauss-Seidel and Jacobian decompositions, and the global convergence of the suggested method can be proved under some restrictions on the proximal matrices. As compared to existing methods, the biggest differentiator of our approach is the use of *dynamic profitable step sizes* during the correction phase of the algorithm (to update the output of the prediction step). Thus, each iterate generated by

this scheme is expected to be closer to the solution set. Furthermore, we will analyze the worst-case $O(1/k)$ convergence rate under both ergodic and nonergodic scenarios, and verify the feasibility and efficiency of the proposed algorithm by solving some instances of the robust RCA problem.

5.2 Algorithm

In this section, we begin by describing a new proximal partially-parallel splitting algorithm for the linearly constrained problem (1.1). Define a matrix M as follows:

$$M = \begin{pmatrix} S_1 & 0 & \dots & 0 & 0 \\ 0 & S_2 + \beta A_2^\top A_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & S_m + \beta A_m^\top A_m & 0 \\ 0 & 0 & \dots & 0 & \beta^{-1} I_l \end{pmatrix}, \quad (5.1)$$

where S_i ($i = 1, \dots, m$) are symmetric positive definite matrices (If the A_i 's are of full column rank, then S_i can be taken as $s\beta A_i^\top A_i$.) Hence, M is also positive definite. (Note that, in certain applications, S_i and M can be relaxed to be positive semi-definite matrices.) We are now ready to put forward a new proximal partially-parallel splitting algorithm for the linearly constrained problem (1.1).

5.2.1 Main Framework of Algorithm

Algorithm 5.1 A proximal partially-parallel splitting algorithm

Step 0. Initialization: Given $\epsilon > 0$ and the initial iterate $(\mathbf{x}^0, \lambda^0) \in \mathcal{X} \times \mathbb{R}^l$. Choose $\beta > 0$ and $Q \succ 0$ ($Q \in \mathcal{S}_{++}^{n+l}$). Set $k := 0$.

Step 1. Proximal Partially Parallel Splitting Step: Determine

$$\begin{cases} \tilde{x}_1^k := \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \left\| A_1 x_1 + \sum_{i=2}^m A_i x_i^k - b - \beta^{-1} \lambda^k \right\|^2 + \frac{1}{2} \|x_1 - x_1^k\|_{S_1}^2 \right\} \\ \tilde{x}_i^k := \arg \min_{x_i \in \mathcal{X}_i} \left\{ \theta_i(x_i) + \frac{\beta}{2} \left\| A_i x_i + A_1 \tilde{x}_1 + \sum_{j=2, j \neq i}^m A_j x_j^k - b - \beta^{-1} \lambda^k \right\|^2 \right. \\ \left. + \frac{1}{2} \|x_i - x_i^k\|_{S_i}^2 \right\}, \forall i \geq 2, \end{cases} \quad (5.2)$$

where $S_i (i = 1, \dots, m)$ are appropriately defined symmetric positive definite matrices.

Step 2. Lagrangian Multiplier Update Step: Compute

$$\tilde{\lambda}^k = \lambda^k - \beta \left(\sum_{i=1}^m A_i \tilde{x}_i^k - b \right). \quad (5.3)$$

Step 3. Termination Step: If

$$\max \left\{ \max_{1 \leq i \leq m} \{ \|x_i^k - \tilde{x}_i^k\| \}, \|\lambda^k - \tilde{\lambda}^k\| \right\} < \epsilon,$$

then terminate the algorithm with $\tilde{\mathbf{u}}^k = (\tilde{\mathbf{x}}^k, \tilde{\lambda}^k)$ being an acceptable approximate solution of (1.1). Otherwise, go to **Step 4**.

Step 4. Corrector Step: Generate the new iterate by

$$\mathbf{u}^{k+1} = \mathbf{u}^k - \alpha_k Q^{-1} M(\mathbf{u}^k - \tilde{\mathbf{u}}^k), \quad (5.4)$$

where

$$\alpha_k = \gamma \alpha_k^*, \quad \alpha_k^* = \frac{\varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k)}{\|Q^{-1} M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|_Q^2}, \quad \gamma \in (0, 2). \quad (5.5)$$

$$\text{and} \quad \varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k) = \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2 + \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k) \right\rangle. \quad (5.6)$$

Set $k := k + 1$ and go to **Step 1**.

In Algorithm 3.1, on the basis of the convergence proof framework for contraction methods [63, 70], we have chosen $\varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ as the *error measure*, which plays a part in defining the step-size and characterizing the optimality tolerance. The step-size α_k , defined in (5.5), is *dynamically updated* at each iteration. An advantage of such an update step is that an optimal step-size can be found at each iteration so as to maximize the contraction of the iterative sequence. Moreover, this chosen step-size often requires only an incrementally greater computational effort. While this update might be computationally demanding when dealing with very large-scale instances, for these cases, the step-length in the correction step can be fixed as a constant throughout the iteration, i.e.,

$$\mathbf{u}^{k+1} = \mathbf{u}^k - \alpha Q^{-1} M(\mathbf{u}^k - \tilde{\mathbf{u}}^k), \quad (5.7)$$

where

$$\alpha = \gamma \left(1 - \sqrt{\frac{m-1}{m}} \right), \quad \gamma \in (0, 2). \quad (5.8)$$

It is worthwhile to note that $-Q^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)$ is essentially a *descent direction* of the distance function $\|\mathbf{u} - \mathbf{u}^*\|_Q^2$ at $\mathbf{u} = \mathbf{u}^k$, where $\mathbf{u}^*$ is a solution to (1.10a). In practice, we can choose the matrix Q in accordance with the form of matrix M , thereby simplifying the correction scheme (5.4) greatly. For example, when each A_i is of full column rank, $\beta A_i^\top A_i$ is a positive definite matrix. Then, we can set $S_1 \succ 0$ and $S_i = s\beta A_i^\top A_i (i \geq 2)$ in Algorithm 3.1. In this case, taking $Q = M \succ 0$ leads to the new iterate by the following scheme

$$\mathbf{u}^{k+1} = \mathbf{u}^k - \alpha_k(\mathbf{u}^k - \tilde{\mathbf{u}}^k).$$

Clearly, the iterative schemes of Algorithm 5.1 can be greatly simplified by an appropriate choice of Q .

Remark 5.1 *If we set $S_1 = 0$, taking*

$$Q_0 = M_0 = \text{diag}(S_2 + \beta A_2^\top A_2, \dots, S_m + \beta A_m^\top A_m, \beta^{-1} I_l) \succ 0 \quad (5.9)$$

and $\mathbf{v}^\top = (x_2^\top, \dots, x_m^\top, \lambda^\top)$, the new iterate can be recovered as

$$\begin{cases} x_1^{k+1} = \tilde{x}_1^k \\ \mathbf{v}^{k+1} = \mathbf{v}^k - \alpha_k Q_0^{-1} M_0(\mathbf{v}^k - \tilde{\mathbf{v}}^k) = \mathbf{v}^k - \alpha_k(\mathbf{v}^k - \tilde{\mathbf{v}}^k) \end{cases}, \quad (5.10)$$

and the related step-size is defined as

$$\alpha_k = \frac{\|\mathbf{v}^k - \tilde{\mathbf{v}}^k\|_{M_0}^2 + \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle}{\|\mathbf{v}^k - \tilde{\mathbf{v}}^k\|_{M_0}^2}. \quad (5.11)$$

5.2.2 Relationships to Existing Methods

The partially-parallel splitting algorithm proposed above shares several interesting properties with existing methods and, in some special cases, reduces to well-known algorithms. Here, we highlight some of these interconnections next.

Let $m = 2$. When $S_i = 0$, the proposed algorithm reduces to the method in [137]. Specifically, this method first generates prediction points via

$$\begin{cases} \tilde{x}_1^k = \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \frac{\beta}{2} \|A_1 x_1 + A_2 x_2^k - b - \beta^{-1} \lambda^k\|^2 \right\} \\ \tilde{x}_2^k = \arg \min_{x_2 \in \mathcal{X}_2} \left\{ \theta_2(x_2) + \frac{\beta}{2} \|A_1 \tilde{x}_1^k + A_2 x_2 - b - \beta^{-1} \lambda^k\|^2 \right\} \\ \tilde{\lambda}^k = \lambda^k - \beta (A_1 \tilde{x}_1^k + A_2 \tilde{x}_2^k - b), \end{cases}$$

and the new iterate points are generated by the following correction step:

$$\begin{cases} x_1^{k+1} = \tilde{x}_1^k, \\ x_2^{k+1} = x_2^k - \alpha_k (x_2^k - \tilde{x}_2^k), \\ \lambda^{k+1} = \lambda^k - \alpha_k (\lambda^k - \tilde{\lambda}^k), \end{cases}$$

where α_k is defined in (5.5). In addition, when $\alpha_k = 1$, this method coincides with the ADMM [45, 48].

When $\alpha_k = 1$, $S_1 = \frac{\tau_1 \beta A_1^\top A_1}{2}$ and $S_i = \frac{\tau_2 \beta A_i^\top A_i}{2} (i \geq 2, A_i \text{ is of full column rank})$, the suggested algorithm is considered as a particular case of the method in [80]:

$$\begin{cases} x_1^{k+1} = \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \mathcal{L}(x_1, x_2^k, \dots, x_m^k, \lambda^k) + \frac{\tau_1 \beta}{2} \|A_1(x_1 - x_1^k)\|^2 \right\} \\ y_j^{k+1} = \arg \min_{y_j \in \mathcal{Y}_j} \left\{ \mathcal{L}(x_1^{k+1}, y_1^k, \dots, y_{j-1}^k, y_j, y_{j+1}^k, \dots, y_{m-1}^k, \lambda^k) + \frac{\tau_2 \beta}{2} \|B_j(y_j - y_j^k)\|^2 \right\} \\ \lambda^{k+1} = \lambda^k - \beta (A x_1^{k+1} + \mathbf{B} \mathbf{y}^{k+1} - b), \end{cases}$$

where $\mathbf{B} = (A_2, \dots, A_m)$ and $\mathbf{y} = (y_1^\top, \dots, y_{m-1}^\top)^\top = (x_2^\top, \dots, x_m^\top)^\top$. Furthermore, when $S_1 = 0$, the proposed method reduces to the method in [73]. However, our algorithm is essentially different from these existing methods because the step-size α_k in (5.5) is dynamically updated in the correction step, which in turn leads to a dynamic generation of the subsequent iterates.

Note that some splitting methods proposed in [25, 75, 120] are full Jacobian decomposition methods, wherein all subproblems are solved simultaneously by introducing proximal terms. However, our proposed method is only a partially-parallel splitting method where, at each iteration of this algorithm, one selected subproblem is solved, and the remaining subproblems are solved in parallel by absorbing the new iterate information obtained from the first subproblem. Therefore, our proposed method can be categorized under the class of *partial Jacobian methods*.

5.3 Convergence Analysis

Next, we are prepared to prove that the suggested method is *globally convergent* under the assumption that $S_1 \succeq 0$ and $S_j \succeq s\beta A_j^\top A_j$ ($j \geq 2$), where $s = \max\{\rho_1, 0\}$, $\rho_1 = \frac{m-1+\sigma}{4} - 1$, and $\sigma \geq 1$. We begin our analysis by inferring some useful inequalities with respect to the iterates $\mathbf{u}^{k+1} = (x_1^{k+1}, \dots, x_m^{k+1}, \lambda^{k+1})$ and $\tilde{\mathbf{u}}^k = (\tilde{x}_1^k, \dots, \tilde{x}_m^k, \tilde{\lambda}^k)$ generated by Algorithm 5.1.

Lemma 5.1 *Let $\tilde{\mathbf{u}}^k$ be generated by Algorithm 5.1 from the given vector $\mathbf{u}^k$. Then,*

$$\begin{aligned} \tilde{\mathbf{u}}^k &\in \mathcal{W}, \text{ and} \\ \theta(\mathbf{x}) - \theta(\tilde{\mathbf{x}}^k) + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, F(\tilde{\mathbf{u}}^k) + d_2(\mathbf{u}^k, \tilde{\mathbf{u}}^k) - d_1(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle &\geq 0, \forall \mathbf{u} \in \mathcal{W}, \end{aligned} \quad (5.12)$$

where,

$$d_1(\mathbf{u}^k, \tilde{\mathbf{u}}^k) = M(\mathbf{u}^k - \tilde{\mathbf{u}}^k), \quad (5.13)$$

$$d_2(\mathbf{u}^k, \tilde{\mathbf{u}}^k) = \beta (A_1, \dots, A_m, 0_{l \times 1})^\top \left(\sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k) \right), \quad (5.14)$$

and M is as defined in (5.1).

Proof: From the first-order optimality conditions for the $\tilde{x}_i^k$ -related subproblems in Algorithm 5.1, we have that

$$\begin{cases} \theta_1(x_1) - \theta_1(\tilde{x}_1^k) + \langle x_1 - \tilde{x}_1^k, -A_1^\top \lambda^k \rangle \\ + \left\langle x_1 - \tilde{x}_1^k, \beta A_1^\top \left(\sum_{j=2}^m A_j x_j^k + A_1 \tilde{x}_1^k - b \right) + S_1(\tilde{x}_1^k - x_1^k) \right\rangle \geq 0, \forall x_1 \in \mathcal{X}_1. \\ \theta_i(x_i) - \theta_i(\tilde{x}_i^k) + \langle x_i - \tilde{x}_i^k, -A_i^\top \lambda^k \rangle \\ + \left\langle x_i - \tilde{x}_i^k, \beta A_i^\top \left(A_1 \tilde{x}_1^k + \sum_{j=2, j \neq i}^m A_j x_j^k + A_i \tilde{x}_i^k - b \right) + S_i(\tilde{x}_i^k - x_i^k) \right\rangle \geq 0, \\ \forall x_i \in \mathcal{X}_i, i \geq 2. \end{cases}$$

Then, it follows from $\tilde{\lambda}^k = \lambda^k - \beta (\sum_{i=1}^m A_i \tilde{x}_i^k - b)$ that

$$\begin{cases} \theta_1(x_1) - \theta_1(\tilde{x}_1^k) + \left\langle x_1 - \tilde{x}_1^k, -A_1^\top \tilde{\lambda}^k \right\rangle \\ + \left\langle x_1 - \tilde{x}_1^k, \sum_{j=2}^m \beta A_1^\top A_j (x_j^k - \tilde{x}_j^k) + S_1(\tilde{x}_1^k - x_1^k) \right\rangle \geq 0, \forall x_1 \in \mathcal{X}_1. \\ \theta_i(x_i) - \theta_i(\tilde{x}_i^k) + \left\langle x_i - \tilde{x}_i^k, -A_i^\top \tilde{\lambda}^k \right\rangle \\ + \left\langle x_i - \tilde{x}_i^k, \sum_{j=2, j \neq i}^m \beta A_i^\top A_j (x_j^k - \tilde{x}_j^k) + S_i(\tilde{x}_i^k - x_i^k) \right\rangle \geq 0, \forall x_i \in \mathcal{X}_i, i \geq 2. \end{cases}$$

The sum of the above inequalities corresponding to $i = 1, \dots, m$ is:

$$\begin{aligned} & \theta(\mathbf{x}) - \theta(\tilde{\mathbf{x}}^k) + \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_i - \tilde{x}_i^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^\top \begin{pmatrix} -A_1^\top \tilde{\lambda}^k \\ -A_2^\top \tilde{\lambda}^k \\ \vdots \\ -A_i^\top \tilde{\lambda}^k \\ \vdots \\ -A_m^\top \tilde{\lambda}^k \end{pmatrix} \\ & + \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_i - \tilde{x}_i^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^\top \begin{pmatrix} \sum_{j=2}^m \beta A_1^\top A_j (x_j^k - \tilde{x}_j^k) + S_1(\tilde{x}_1^k - x_1^k) \\ \sum_{j=3}^m \beta A_2^\top A_j (x_j^k - \tilde{x}_j^k) + S_2(\tilde{x}_2^k - x_2^k) \\ \vdots \\ \sum_{j=2, j \neq i}^m \beta A_i^\top A_j (x_j^k - \tilde{x}_j^k) + S_i(\tilde{x}_i^k - x_i^k) \\ \vdots \\ \sum_{j=2}^{m-1} \beta A_m^\top A_j (x_j^k - \tilde{x}_j^k) + S_m(\tilde{x}_m^k - x_m^k) \end{pmatrix} \geq 0, \quad \forall \mathbf{x} \in \mathcal{X}. \quad (5.15) \end{aligned}$$

Now, Equation (5.3), which is the Lagrange multiplier update step in Algorithm 5.1, can be rewritten as

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

and therefore,

$$\left\langle \lambda - \tilde{\lambda}^k, \left(\sum_{j=1}^m A_j \tilde{x}_j^k - b \right) + \beta^{-1}(\tilde{\lambda}^k - \lambda^k) \right\rangle \geq 0, \quad \forall \lambda \in \mathbb{R}^l. \quad (5.16)$$

Adding (5.15) and (5.16) together yields

$$\begin{aligned}
 & \theta(\mathbf{x}) - \theta(\tilde{\mathbf{x}}^k) \\
 & + \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^\top \left\{ \begin{pmatrix} -A_1^\top \tilde{\lambda}^k \\ -A_2^\top \tilde{\lambda}^k \\ \vdots \\ -A_m^\top \tilde{\lambda}^k \\ \sum_{j=1}^m A_j \tilde{x}_j^k - b \end{pmatrix} + \beta \begin{pmatrix} A_1^\top \\ A_2^\top \\ \vdots \\ A_m^\top \\ 0_{1 \times l} \end{pmatrix} \left(\sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k) \right) \right\} \\
 & \geq \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^\top \begin{pmatrix} S_1(x_1^k - \tilde{x}_1^k) \\ (\beta A_2^\top A_2 + S_2)(x_2^k - \tilde{x}_2^k) \\ \vdots \\ (\beta A_m^\top A_m + S_m)(x_m^k - \tilde{x}_m^k) \\ \beta^{-1}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}, \forall \mathbf{u} \in \mathcal{W}. \tag{5.17}
 \end{aligned}$$

Using the definitions of $d_1(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ and $d_2(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$, given by (5.13) and (5.14), respectively, (5.17) is able to be equivalently expressed as

$$\theta(\mathbf{x}) - \theta(\tilde{\mathbf{x}}^k) + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, F(\tilde{\mathbf{u}}^k) + d_2(\mathbf{u}^k, \tilde{\mathbf{u}}^k) - d_1(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle \geq 0, \forall \mathbf{u} \in \mathcal{W},$$

and the statement (5.12) is then obtained. $\square$

Lemma 5.2 *Let $\mathbf{u}^* = (x_1^*, \dots, x_m^*, \lambda^*)$ be a solution of $\text{VI}(\mathcal{U}, F, \theta)$ and the sequence $\{\tilde{\mathbf{u}}^k = (\tilde{x}_1^k, \dots, \tilde{x}_m^k, \tilde{\lambda}^k)\}$ be generated by Algorithm 5.1 from the given vector $\mathbf{u}^k$. Then,*

$$\langle \mathbf{u}^k - \mathbf{u}^*, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle \geq \varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k), \tag{5.18}$$

where $\varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ is defined in (5.6).

Proof: Making use of the definition of $d_2(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ leads to

$$\langle \tilde{\mathbf{u}}^k - \mathbf{u}^*, d_2(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle = \left\langle \sum_{j=1}^m A_j (\tilde{x}_j^k - x_j^*), \sum_{j=2}^m \beta A_j (x_j^k - \tilde{x}_j^k) \right\rangle. \tag{5.19}$$

Since $\sum_{j=1}^m A_j x_j^* = b$ and $\beta(\sum_{j=1}^m A_j \tilde{x}_j^k - b) = \lambda^k - \tilde{\lambda}^k$, we get

$$\left\langle \sum_{j=1}^m A_j (\tilde{x}_j^k - x_j^*), \sum_{j=2}^m \beta A_j (x_j^k - \tilde{x}_j^k) \right\rangle = \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k) \right\rangle.$$

Substituting it into (5.19) yields,

$$\langle \tilde{\mathbf{u}}^k - \mathbf{u}^*, d_2(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle = \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle. \quad (5.20)$$

Now that $\tilde{\mathbf{u}}^k \in \mathcal{U}$ and $\mathbf{u}^* \in \mathcal{U}^*$, utilizing the monotonicity of $F(\cdot)$ and (1.10a) leads to

$$\theta(\tilde{\mathbf{x}}^k) - \theta(\mathbf{x}^*) + \langle \tilde{\mathbf{u}}^k - \mathbf{u}^*, F(\tilde{\mathbf{u}}^k) \rangle \geq \theta(\tilde{\mathbf{x}}^k) - \theta(\mathbf{x}^*) + \langle \tilde{\mathbf{u}}^k - \mathbf{u}^*, F(\mathbf{u}^*) \rangle \geq 0.$$

Setting $\mathbf{u} = \mathbf{u}^*$ in (5.12) deduces,

$$\langle \tilde{\mathbf{u}}^k - \mathbf{u}^*, d_1(\mathbf{u}^k, \tilde{\mathbf{u}}^k) - d_2(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle \geq \theta(\tilde{\mathbf{x}}^k) - \theta(\mathbf{x}^*) + \langle \tilde{\mathbf{u}}^k - \mathbf{u}^*, F(\tilde{\mathbf{u}}^k) \rangle \geq 0,$$

where $d_1(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ and $d_2(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ are given by (5.13) and (5.14), respectively. Then,

$$\langle \tilde{\mathbf{u}}^k - \mathbf{u}^*, d_1(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle \geq \langle \tilde{\mathbf{u}}^k - \mathbf{u}^*, d_2(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle. \quad (5.21)$$

It immediately follows from (5.20) and (5.21) that

$$\langle \mathbf{u}^k - \mathbf{u}^*, d_1(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle \geq \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2 + \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle,$$

and the assertion of this lemma follows from the above inequality and the definition of $\varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$. $\square$

Next, following the analytic framework of convergence analysis for contraction methods in [67], we intend to show in the following lemma and its subsequent discussion that

$$\varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \geq c_0 \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2,$$

for some $c_0 > 0$, which is only dependent on m .

Lemma 5.3 *Assume that $S_1 \succ 0$ and $S_j \succeq s\beta A_j^\top A_j$ ($j \geq 2$), where $s = \max\{\rho_1, 0\}$, $\rho_1 = \frac{m-1+\sigma}{4} - 1$ and $\sigma \geq 1$. Let $\tilde{\mathbf{u}}^k$ be generated by Algorithm 5.1 from a given vector $\mathbf{u}^k$. Then,*

$$\varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \geq c_0 \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2, \quad (5.22)$$

where $\varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ is defined as (5.6) and

$$c_0 = 1 - \sqrt{\frac{m-1}{m}} \quad (5.23)$$

is a positive constant.

Proof: Given that $S_1 \succeq 0$ and $S_j \succeq \rho_1 \beta A_j^\top A_j$ ($j \geq 2$), where $\rho_1 = \frac{m-1+\sigma}{4} - 1$ and $\sigma \geq 1$. Then, it follows that

$$\sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j + \beta A_j^\top A_j}^2 \geq \rho \sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{\beta A_j^\top A_j}^2,$$

and

$$\sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j + \beta A_j^\top A_j}^2 \geq (1 - \frac{q}{t}) \sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j + \beta A_j^\top A_j}^2 + qt \sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{\beta A_j^\top A_j}^2, \quad (5.24)$$

where $\rho = \rho_1 + 1 = \frac{m-1+\sigma}{4} > 0$, $q = \frac{\sqrt{m-1}}{2}$ and $t = \sqrt{\rho} = \frac{\sqrt{m-1+\sigma}}{2}$. In addition, we have:

$$\beta^{-1} \|\lambda^k - \tilde{\lambda}^k\|^2 = \left(1 - \frac{q}{t}\right) \beta^{-1} \|\lambda^k - \tilde{\lambda}^k\|^2 + \left(\frac{q}{t}\right) \beta^{-1} \|\lambda^k - \tilde{\lambda}^k\|^2,$$

and,

$$\begin{aligned} & q \left(t\beta \sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{A_j^\top A_j}^2 + \frac{1}{t\beta} \|\lambda^k - \tilde{\lambda}^k\|^2 \right) + \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle \\ &= \frac{\beta}{4qt} \left(\sum_{j=2}^m 4q^2 t^2 \|A_j(x_j^k - \tilde{x}_j^k)\|^2 + \frac{4q^2}{\beta^2} \|\lambda^k - \tilde{\lambda}^k\|^2 \right) + \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle \\ &= \frac{\beta}{4qt} \left(\sum_{j=2}^m \left(4q^2 t^2 \|A_j(x_j^k - \tilde{x}_j^k)\|^2 + \frac{4qt}{\beta} \left\langle \lambda^k - \tilde{\lambda}^k, A_j(x_j^k - \tilde{x}_j^k) \right\rangle \right) + \frac{m-1}{\beta^2} \|\lambda^k - \tilde{\lambda}^k\|^2 \right) \\ &= \frac{\beta}{4qt} \sum_{j=2}^m \|2qt A_j(x_j^k - \tilde{x}_j^k) + \beta^{-1}(\lambda^k - \tilde{\lambda}^k)\|^2. \end{aligned}$$

Combining (5.24) and the above two equalities yields

$$\begin{aligned} & \sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j + \beta A_j^\top A_j}^2 + \beta^{-1} \|\lambda^k - \tilde{\lambda}^k\|^2 + \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle \\ & \geq \left(1 - \frac{q}{t}\right) \left(\sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j + \beta A_j^\top A_j}^2 + \beta^{-1} \|\lambda^k - \tilde{\lambda}^k\|^2 \right) \\ & \quad + \frac{\beta}{4qt} \sum_{j=2}^m \|2qt A_j(x_j^k - \tilde{x}_j^k) + \beta^{-1}(\lambda^k - \tilde{\lambda}^k)\|^2 \\ & \geq \left(1 - \frac{q}{t}\right) \left(\sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j + \beta A_j^\top A_j}^2 + \beta^{-1} \|\lambda^k - \tilde{\lambda}^k\|^2 \right). \end{aligned}$$

Utilizing the definition of $\varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ leads to

$$\begin{aligned}
 & \varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \\
 &= \|x_1^k - \tilde{x}_1^k\|_{S_1}^2 + \sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j + \beta A_j^\top A_j}^2 + \beta^{-1} \|\lambda^k - \tilde{\lambda}^k\|^2 \\
 &+ \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle \\
 &\geq \|x_1^k - \tilde{x}_1^k\|_{S_1}^2 + \left(1 - \frac{q}{t}\right) \left(\sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j + \beta A_j^\top A_j}^2 + \beta^{-1} \|\lambda^k - \tilde{\lambda}^k\|^2 \right). \quad (5.25)
 \end{aligned}$$

Then, we get

$$\begin{aligned}
 & \frac{\varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k)}{\|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2} \\
 &\geq \frac{\|x_1^k - \tilde{x}_1^k\|_{S_1}^2 + \left(1 - \frac{q}{t}\right) \left(\sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j + \beta A_j^\top A_j}^2 + \beta^{-1} \|\lambda^k - \tilde{\lambda}^k\|^2 \right)}{\|x_1^k - \tilde{x}_1^k\|_{S_1}^2 + \sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j + \beta A_j^\top A_j}^2 + \beta^{-1} \|\lambda^k - \tilde{\lambda}^k\|^2} \\
 &\geq \frac{\left(1 - \frac{q}{t}\right) \left(\sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j + \beta A_j^\top A_j}^2 + \beta^{-1} \|\lambda^k - \tilde{\lambda}^k\|^2 \right)}{\sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j + \beta A_j^\top A_j}^2 + \beta^{-1} \|\lambda^k - \tilde{\lambda}^k\|^2} \\
 &= 1 - \frac{q}{t} \\
 &= 1 - \sqrt{\frac{m-1}{m-1+\sigma}} \\
 &\geq 1 - \sqrt{\frac{m-1}{m}} \quad (\text{as } \sigma \geq 1), \quad (5.26)
 \end{aligned}$$

where the first inequality is derived from (5.25) and the second inequality is inferred by using

$$\frac{a+b}{a+c} \geq \frac{b}{c}, \text{ for any } a \geq 0, c \geq b > 0.$$

Hence, when $Q = M \succ 0$, we obtain a lower bound c_0 on the step-size α_k^* given by (5.23).

Then, combining (5.26) and (5.23) yields the assertion of this lemma. $\square$

Remark 5.2 When the proximal matrix $S_1 = 0$ and $S_i = s\beta A_i^\top A_i$ ($i \geq 2, s \geq \max\{\rho_1, 0\}$), we can assume that $Q_0 = M_0 = \text{diag}(S_2 + \beta A_2^\top A_2, \dots, S_m + \beta A_m^\top A_m, \beta^{-1} I_l)$. Then, the step size α_k^* defined as (5.11) has the same lower bound (given by c_0) in this case.

The inequalities (5.18) and (5.22) imply that $-Q^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)$ is a *descent direction* of the distance function $\|\mathbf{u} - \mathbf{u}^*\|_Q^2$ at $\mathbf{u} = \mathbf{u}^k$. Therefore, the new iterate $\mathbf{u}^{k+1}$ is determined along this direction, in concert with a suitable step size α_k , as follows:

$$\mathbf{u}^{k+1}(\alpha_k) = \mathbf{u}^k - \alpha_k Q^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k). \quad (5.27)$$

Therefore,

$$\vartheta(\alpha_k) = \|\mathbf{u}^k - \mathbf{u}^*\|_Q^2 - \|\mathbf{u}^{k+1}(\alpha_k) - \mathbf{u}^*\|_Q^2 \quad (5.28)$$

is the *distance decrease function* at the k -th iteration when using the updating form (5.27). However, $\vartheta(\alpha_k)$ cannot be directly maximized because $\mathbf{u}^* \in \mathcal{U}^*$ is unknown. Hence, we resort to derive a tight lower bound of $\vartheta(\alpha_k)$, denoted by $\psi(\alpha_k)$, which is independent of the unknown vector $\mathbf{u}^*$, and this is summarized in Lemma 5.4.

Lemma 5.4 *For any $\mathbf{u}^* \in \mathcal{U}^*$ and $\alpha_k \geq 0$, we yield*

$$\vartheta(\alpha_k) \geq \psi(\alpha_k), \quad (5.29)$$

where

$$\psi(\alpha_k) = 2\alpha_k \varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k) - \alpha_k^2 \|Q^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|_Q^2. \quad (5.30)$$

Proof: From (5.27) and (5.28), we get

$$\begin{aligned} \vartheta(\alpha_k) &= \|\mathbf{u}^k - \mathbf{u}^*\|_Q^2 - \|\mathbf{u}^k - \mathbf{u}^* - \alpha_k Q^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|_Q^2 \\ &= 2\alpha_k \langle \mathbf{u}^k - \mathbf{u}^*, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle - \alpha_k^2 \|Q^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|_Q^2 \\ &\geq 2\alpha_k \varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k) - \alpha_k^2 \|Q^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|_Q^2, \end{aligned} \quad (5.31)$$

where the last inequality follows from (5.18), and the right hand side of the above inequality is $\psi(\alpha_k)$. $\square$

It follows from Lemma 5.4 that $\vartheta(\alpha_k) \geq \psi(\alpha_k)$. Since $\psi(\alpha_k)$ is a quadratic function of α_k , it reaches its maximum value at

$$\alpha_k^* = \frac{\varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k)}{\|Q^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|_Q^2},$$

and this is just the same as defined in (5.5). Moreover, as a result of (5.22), for a more general $Q \succ 0$, $\alpha_k^* > 0$ is bounded away from zero, i.e., there exists a constant $\bar{c}_0 > 0$ such that

$$\alpha_k^* \geq \frac{c_0 \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2}{\|Q^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|_Q^2} \geq \bar{c}_0 > 0, \quad \forall k \geq 0, \quad (5.32)$$

where

$$\bar{c}_0 = \min \left\{ \frac{c_0 \lambda_{\min}(M)}{\lambda_{\max}(M^\top Q^{-\top} M)}, c_0 \right\}. \quad (5.33)$$

Thus, combining (5.5), (5.29), and (5.30) leads to

$$\vartheta(\alpha_k^*) \geq \psi(\alpha_k^*) = \alpha_k^* \varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \geq 0. \quad (5.34)$$

For computational efficiency, taking a relaxation factor $\gamma \in (0, 2)$ is beneficial for accelerating convergence, and the following theorem validates this claim by proving that the sequence $\{\mathbf{u}^k\}$ generated by the suggested algorithm is Fejér monotone.

Theorem 5.1 *Assume that $S_1 \succ 0$ and $S_j \succeq s\beta A_j^\top A_j$ ($j \geq 2$), where $s = \max\{\rho_1, 0\}$, $\rho_1 = \frac{m-1+\sigma}{4} - 1$ and $\sigma \geq 1$. Then, the sequence $\{\mathbf{u}^k\}$ generated by Algorithm 5.1 satisfies*

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_Q^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|_Q^2 - \gamma(2 - \gamma)\bar{c}_0^2 \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2, \quad \forall \mathbf{u}^* \in \mathcal{U}^*. \quad (5.35)$$

Proof: Combining (5.29) and (5.30) leads to

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_Q^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|_Q^2 - \psi(\alpha_k), \quad \forall \mathbf{u}^* \in \mathcal{U}^*.$$

Using (5.30) and (5.34) yields

$$\begin{aligned} \psi(\gamma\alpha_k^*) &= 2\gamma\alpha_k^* \varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k) - (\gamma\alpha_k^*)^2 \|Q^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|_Q^2 \\ &= \gamma(2 - \gamma)\alpha_k^* \varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k). \end{aligned}$$

Therefore,

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_Q^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|_Q^2 - \gamma(2 - \gamma)\alpha_k^* \varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k). \quad (5.36)$$

Then, the statement of this theorem is then derived from (5.22), (5.33) and (5.32) directly.

□

Remark 5.3 Consider the case that $\alpha = \gamma c_0$ and $Q = M$, which is a commonly used step-size in practice. Substituting this into (5.30) yields

$$\psi(\alpha) = 2\gamma c_0 \varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k) - \gamma^2 c_0^2 \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2.$$

Moreover, using (5.22) leads to

$$\psi(\alpha) \geq \gamma(2 - \gamma)c_0^2 \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2,$$

and the assertion (5.35) is also proved in this case.

Remark 5.4 If $S_1 = 0$, $S_j \succeq s\beta A_j^\top A_j$ ($j \geq 2$), for Algorithm 5.1 with correction scheme (5.10), we can conclude that $M_0 \succ 0$ and

$$\|\mathbf{v}^{k+1} - \mathbf{v}^*\|_{M_0}^2 \leq \|\mathbf{v}^k - \mathbf{v}^*\|_{M_0}^2 - \gamma(2 - \gamma)c_0^2 \|\tilde{\mathbf{v}}^k - \mathbf{v}^k\|_{M_0}^2, \forall \mathbf{v}^* \in \mathcal{V}^*$$

where $\mathcal{V}^* = \{\mathbf{v}^* | (x_1^*, \mathbf{v}^*) \in \mathcal{W}^*\}$.

Theorem 5.2 Assume that $S_1 \succ 0$ and $S_j \succeq s\beta A_j^\top A_j$ ($j \geq 2$), where $s = \max\{\rho_1, 0\}$, $\rho_1 = \frac{m-1+\sigma}{4} - 1$ and $\sigma \geq 1$. Then, the sequence $\{\mathbf{u}^k\}$ generated by Algorithm 5.1 globally converges to a solution $\mathbf{u}^* \in \mathcal{U}^*$.

Proof: For (5.35), we yield

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_Q^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|_Q^2 \leq \dots \leq \|\mathbf{u}^0 - \mathbf{u}^*\|_Q^2 \leq +\infty,$$

which implies that the generated sequence $\{\mathbf{u}^k\}$ is bounded and possesses at least one accumulation point. Furthermore, it follows from (5.35) that

$$\sum_{j=1}^m \gamma(2 - \gamma)\bar{c}_0^2 \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2 \leq \sum_{k=0}^{\infty} (\|\mathbf{u}^k - \mathbf{u}^*\|_Q^2 - \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_Q^2) \leq +\infty, \quad (5.37)$$

which further implies that

$$\lim_{k \rightarrow \infty} \|x_i^k - \tilde{x}_i^k\|^2 = 0, \quad \forall i = 1, \dots, m,$$

and

$$\lim_{k \rightarrow \infty} \|\lambda^k - \tilde{\lambda}^k\|^2 = 0.$$

Therefore, $\lim_{k \rightarrow \infty} \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|^2 = 0$, and for any $\epsilon > 0$, $\exists N_1 > 0$, then there exists a $k \geq N_1$ such that $\|\tilde{\mathbf{u}}^k - \mathbf{u}^k\| \leq \epsilon/2$. Assume that $\mathbf{u}^*$ is a cluster point of the sequence $\{\mathbf{u}^k\}$. Then, we obtain that $\forall \epsilon > 0$, $\exists N_2 > 0$, then there exists a $k \geq N_2$ such that $\|\mathbf{u}^k - \mathbf{u}^*\| \leq \epsilon/2$. Therefore, for any $\epsilon > 0$, there exists a $k \geq \max\{N_1, N_2\}$ such that

$$\|\tilde{\mathbf{u}}^k - \mathbf{u}^*\| \leq \|\tilde{\mathbf{u}}^k - \mathbf{u}^k\| + \|\mathbf{u}^k - \mathbf{u}^*\| = \epsilon/2 + \epsilon/2 = \epsilon,$$

which indicates that the sequence $\{\tilde{\mathbf{u}}^k\}$ also converges to the cluster point $\mathbf{u}^*$.

Additionally, we have

$$\lim_{k \rightarrow \infty} \left(\sum_{i=1}^m A_i \tilde{x}_i^k - b \right) = \lim_{k \rightarrow \infty} \beta^{-1}(\lambda^k - \tilde{\lambda}^k) = 0.$$

Substituting this in (5.17) leads to

$$\lim_{k \rightarrow \infty} \theta(\mathbf{x}) - \theta(\tilde{\mathbf{x}}^k) + (\mathbf{u} - \tilde{\mathbf{u}}^k)^\top F(\tilde{\mathbf{u}}^k) \geq 0, \forall \mathbf{u} \in \mathcal{U},$$

and thus any cluster point of $\{\tilde{\mathbf{u}}^k\}$ is a solution point of $\text{VI}(\mathcal{U}, F, \theta)$.

Let $\tilde{\mathbf{u}}^\infty$ be a cluster point of $\{\tilde{\mathbf{u}}^k\}$ and $\{\tilde{\mathbf{u}}^{k_j}\}$ be the subsequence converging to $\tilde{\mathbf{u}}^\infty$. It follows from the above inequality that $\tilde{\mathbf{u}}^{k_j} \in \mathcal{U}$ and

$$\lim_{k \rightarrow \infty} \theta(\mathbf{x}) - \theta(\tilde{\mathbf{x}}^{k_j}) + (\mathbf{u} - \tilde{\mathbf{u}}^{k_j})^\top F(\tilde{\mathbf{u}}^{k_j}) \geq 0, \forall \mathbf{u} \in \mathcal{U},$$

and consequently, $\tilde{\mathbf{u}}^\infty \in \mathcal{U}$ and

$$\theta(\mathbf{x}) - \theta(\tilde{\mathbf{x}}^\infty) + (\mathbf{u} - \tilde{\mathbf{u}}^\infty)^\top F(\tilde{\mathbf{u}}^\infty) \geq 0, \forall \mathbf{u} \in \mathcal{U}.$$

The above inequality proves that $\tilde{\mathbf{u}}^\infty \in \mathcal{U}^*$. As $\{\mathbf{u}^k\}$ is Fejér monotone and $\lim_{k \rightarrow \infty} \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|^2 = 0$, the sequence $\{\mathbf{u}^k\}$ cannot have any other cluster point, and hence, $\lim_{k \rightarrow \infty} \mathbf{u}^k = \tilde{\mathbf{u}}^\infty$. Then, the assertion of this theorem is obtained. $\square$

Remark 5.5 When $S_1 = 0$ and $S_j \succeq s\beta A_j^\top A_j$ ($j \geq 2$), it follows from Remark 5.4 that $\{\mathbf{v}^k\}$ is bounded and

$$\lim_{k \rightarrow \infty} \|\lambda^k - \tilde{\lambda}^k\|^2 = 0 \quad \text{and} \quad \lim_{k \rightarrow \infty} \|x_i^k - \tilde{x}_i^k\|^2 = 0, \quad \text{for all } i = 2, \dots, m.$$

Then, $\{\tilde{\mathbf{v}}^k\}$ is also bounded. Since A_1 is full column rank and

$$A_1 \tilde{x}_1^k = \beta^{-1}(\lambda^k - \tilde{\lambda}^k) + b - \sum_{i=2}^m A_i \tilde{x}_i^k,$$

we have

$$\|x_1^k\| = \|\tilde{x}_1^k\| \leq \|(A_1^\top A_1)^{-1}\| \cdot \|A_1^\top\| \cdot \left\{ \beta^{-1} \|\lambda^k - \tilde{\lambda}^k\| + \|b\| + \sum_{i=2}^m \|A_i\| \|\tilde{x}_i^k\| \right\}.$$

Then, $\{x_1^k\}$ and $\{\tilde{x}_1^k\}$ is also bounded. Hence, we can easily conclude that $\{\mathbf{u}^k\}$ converges to a solution of $\text{VI}(\mathcal{U}, F, \theta)$ by referring to the proof of Theorem 5.2.

5.4 Convergence Rate

In this section, we analyze the worst-case convergence rate for the proposed algorithm.

5.4.1 Convergence Rate under Ergodic Sense

Here, we derive a worst-case $O(1/k)$ convergence rate under ergodic condition for Algorithm 5.1 under the assumption that $S_1 \succ 0$ and $S_j \succeq s\beta A_j^\top A_j$ ($j \geq 2$) ($s \geq \max\{\frac{m-1+\sigma}{4} - 1, 0\}$, $\sigma \geq 1$). Notice that we can also prove the $O(1/k)$ convergence rate under ergodic sense for Algorithm 5.1, when equipped with the correction step (5.10), under the assumptions that $S_1 = 0$, and $S_i \succeq s\beta A_i^\top A_i$ ($i \geq 2$).

To begin with, define a sequence $\bar{\mathbf{u}} = (\bar{x}_1^k, \bar{x}_2^k, \dots, \bar{x}_m^k, \bar{\lambda}^k)$ as

$$\begin{pmatrix} \bar{x}_1^k \\ \bar{x}_2^k \\ \vdots \\ \bar{x}_m^k \end{pmatrix} = \begin{pmatrix} \tilde{x}_1^k \\ \tilde{x}_2^k \\ \vdots \\ \tilde{x}_m^k \end{pmatrix} \quad \text{and} \quad \bar{\lambda}^k = \tilde{\lambda}^k + \sum_{j=2}^k \beta A_j (\tilde{x}_j^k - x_j^k), \quad (5.38)$$

Additionally, define the following two matrices:

$$L = \begin{pmatrix} I_{n_1} & 0 & \dots & 0 & 0 \\ 0 & I_{n_2} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_{n_m} & 0 \\ 0 & \beta A_2 & \dots & \beta A_m & I_l \end{pmatrix} \quad (5.39)$$

and

$$\tilde{G} = \begin{pmatrix} S_1 & 0 & \dots & 0 & 0 \\ 0 & S_2 + \beta A_2^\top A_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & S_m + \beta A_m^\top A_m & 0 \\ 0 & -A_2 & \dots & -A_m & \beta^{-1} I_l \end{pmatrix}. \quad (5.40)$$

It can be easily obtain that,

$$\tilde{G}L = M,$$

and

$$\mathbf{u}^k - \bar{\mathbf{u}}^k = L(\mathbf{u}^k - \tilde{\mathbf{u}}^k), \quad (5.41)$$

where M and L are given by (5.1) and (5.39), respectively.

In the following, we translate the assertion (5.12) proved in Lemma 5.1 in terms of $\bar{\mathbf{u}}^k$.

Lemma 5.5 *Let the sequence $\{\tilde{\mathbf{u}}^k = (\tilde{x}_1^k, \dots, \tilde{x}_m^k, \tilde{\lambda}^k)\}$ be generated by Algorithm 5.1 with a given $\mathbf{u}^k$, and let the corresponding sequence $\{\bar{\mathbf{u}}^k\}$ be given by (5.38). Then,*

$$\bar{\mathbf{u}}^k \in \mathcal{U}, \quad \theta(\mathbf{x}) - \theta(\bar{\mathbf{x}}) + \langle \mathbf{u} - \bar{\mathbf{u}}^k, F(\bar{\mathbf{u}}^k) + \tilde{G}(\bar{\mathbf{u}}^k - \mathbf{u}^k) \rangle \geq 0, \quad \forall \mathbf{u} \in \mathcal{U}, \quad (5.42)$$

where $\tilde{G}$ is defined as (5.40).

Proof: By using (5.38), we have $\bar{x}_i^k = \tilde{x}_i^k$, $i = 1, \dots, m$, and

$$-\tilde{\lambda}^k = -\bar{\lambda}^k + \sum_{j=2}^m \beta A_j(\tilde{x}_j^k - x_j^k).$$

Substituting it into the variational inequality (5.15), we get

$$\begin{aligned} & \theta(\mathbf{x}) - \theta(\bar{\mathbf{x}}^k) \\ & + \begin{pmatrix} x_1 - \bar{x}_1^k \\ x_2 - \bar{x}_2^k \\ \vdots \\ x_m - \bar{x}_m^k \end{pmatrix}^\top \left\{ \begin{pmatrix} -A_1^\top \bar{\lambda}^k \\ -A_2^\top \bar{\lambda}^k \\ \vdots \\ -A_m^\top \bar{\lambda}^k \end{pmatrix} + \begin{pmatrix} S_1(\bar{x}_1^k - x_1^k) \\ (S_2 + \beta A_2^\top A_2)(\bar{x}_2^k - x_2^k) \\ \vdots \\ (S_m + \beta A_m^\top A_m)(\bar{x}_m^k - x_m^k) \end{pmatrix} \right\} \geq 0, \end{aligned}$$

for all $\mathbf{x} \in \mathcal{X}$.

Besides, it follows from (5.3) and (5.38) that $\bar{\lambda}^k = \lambda^k - \beta(A_1\bar{x}_1^k + \sum_{j=2}^m A_j\bar{x}_j^k - b)$ and

$$\left\langle \lambda - \bar{\lambda}^k, \sum_{j=1}^m (A_j\bar{x}_j^k - b) - \sum_{j=2}^m A_j(\bar{x}_j^k - x_j^k) + \beta^{-1}(\bar{\lambda}^k - \lambda^k) \right\rangle \geq 0.$$

Combining the above two inequalities yields

$$\begin{aligned} & \theta(\mathbf{x}) - \theta(\bar{\mathbf{x}}^k) + \begin{pmatrix} x_1 - \bar{x}_1^k \\ x_2 - \bar{x}_2^k \\ \vdots \\ x_m - \bar{x}_m^k \\ \lambda - \bar{\lambda}^k \end{pmatrix}^\top \begin{pmatrix} -A_1^\top \bar{\lambda}^k \\ -A_2^\top \bar{\lambda}^k \\ \vdots \\ -A_m^\top \bar{\lambda}^k \\ \sum_{j=1}^m A_j\bar{x}_j^k - b \end{pmatrix} \\ & + \begin{pmatrix} x_1 - \bar{x}_1^k \\ x_2 - \bar{x}_2^k \\ \vdots \\ x_m - \bar{x}_m^k \\ \lambda - \bar{\lambda}^k \end{pmatrix}^\top \begin{pmatrix} S_1(\bar{x}_1^k - x_1^k) \\ (S_2 + \beta A_2^\top A_2)(\bar{x}_2^k - x_2^k) \\ \vdots \\ (S_m + \beta A_m^\top A_m)(\bar{x}_m^k - x_m^k) \\ -\sum_{j=2}^m A_j(\bar{x}_j^k - x_j^k) + \beta^{-1}(\bar{\lambda}^k - \lambda^k) \end{pmatrix} \geq 0, \forall \mathbf{u} \in \mathcal{U}, \end{aligned}$$

which, using the notations for F (see (1.10b)) and $\tilde{G}$ in (5.40), can be compactly written as

$$\bar{\mathbf{u}}^k \in \mathcal{U}, \theta(\mathbf{x}) - \theta(\bar{\mathbf{x}}^k) + \langle \mathbf{u} - \bar{\mathbf{u}}^k, F(\bar{\mathbf{u}}^k) + \tilde{G}(\bar{\mathbf{u}}^k - \mathbf{u}^k) \rangle \geq 0, \forall \mathbf{u} \in \mathcal{U}.$$

The assertion of this lemma is thus obtained. $\square$

Lemma 5.6 *Let the sequence $\{\tilde{\mathbf{u}}^k\}$ be generated by Algorithm 5.1 with a given $\mathbf{u}^k$, and let the vector $\bar{\mathbf{u}}^k$ be as defined in (5.38). Then, we obtain*

$$\|\mathbf{u}^k - \bar{\mathbf{u}}^k\|_Q^2 - \|\mathbf{u}^{k+1} - \bar{\mathbf{u}}^k\|_Q^2 \geq 0, \quad (5.43)$$

where $\mathbf{u}^{k+1}$ is defined in (5.4).

Proof:

$$\begin{aligned} & \|\mathbf{u}^k - \bar{\mathbf{u}}^k\|_Q^2 - \|\mathbf{u}^{k+1} - \bar{\mathbf{u}}^k\|_Q^2 \\ & = \|\mathbf{u}^k - \bar{\mathbf{u}}^k\|_Q^2 - \|\mathbf{u}^k - \bar{\mathbf{u}}^k - \alpha_k Q^{-1} M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|_Q^2 \\ & = 2\alpha_k \langle \mathbf{u}^k - \bar{\mathbf{u}}^k, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle - \alpha_k^2 \|Q^{-1} M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|_Q^2. \end{aligned} \quad (5.44)$$

By using the definitions of $\bar{\mathbf{u}}^k$ in (5.38) and M in (5.1), we get

$$2\alpha_k \langle \mathbf{u}^k - \bar{\mathbf{u}}^k, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle = 2\alpha_k \langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, L^\top M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle.$$

Substituting this into the inequality (5.44) yields

$$\begin{aligned} & \|\mathbf{u}^k - \bar{\mathbf{u}}^k\|_Q^2 - \|\mathbf{u}^{k+1} - \bar{\mathbf{u}}^k\|_Q^2 \\ &= 2\alpha_k \langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, L^\top M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle - \alpha_k^2 \|Q^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|_Q^2. \end{aligned}$$

From (5.6), we already have

$$\begin{aligned} \langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, L^\top M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle &= \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2 + \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle \\ &= \varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k). \end{aligned}$$

Therefore,

$$\|\mathbf{u}^k - \bar{\mathbf{u}}^k\|_Q^2 - \|\mathbf{u}^{k+1} - \bar{\mathbf{u}}^k\|_Q^2 = 2\alpha_k \varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k) - \alpha_k^2 \|Q^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|_Q^2 = \psi(\alpha_k),$$

and at each update step, we will always have $\psi(\alpha_k) \geq 0$, which completes the proof. $\square$

Lemma 5.7 *Let the sequences $\{\mathbf{u}^k\}$ and $\{\tilde{\mathbf{u}}^k\}$ be generated by Algorithm 5.1 with the correction scheme (5.4), and let the vector $\bar{\mathbf{u}}^k$ be defined as (5.38). Then, we obtain*

$$\alpha_k(\theta(\mathbf{x}) - \theta(\bar{\mathbf{x}}^k) + \langle \mathbf{u} - \bar{\mathbf{u}}^k, F(\mathbf{u}) \rangle) \geq \frac{1}{2}(\|\mathbf{u} - \mathbf{u}^{k+1}\|_Q^2 - \|\mathbf{u} - \mathbf{u}^k\|_Q^2), \forall \mathbf{u} \in \mathcal{U}. \quad (5.45)$$

Proof: Using Lemma 5.5 and $\langle \mathbf{u} - \bar{\mathbf{u}}^k, F(\mathbf{u}) \rangle = \langle \mathbf{u} - \bar{\mathbf{u}}^k, F(\bar{\mathbf{u}}^k) \rangle$, we get

$$\alpha_k(\theta(\mathbf{x}) - \theta(\bar{\mathbf{x}}) + \langle \mathbf{u} - \bar{\mathbf{u}}^k, F(\mathbf{u}) \rangle) \geq \alpha_k \langle \mathbf{u} - \bar{\mathbf{u}}^k, \tilde{G}(\mathbf{u}^k - \bar{\mathbf{u}}^k) \rangle, \forall \mathbf{u} \in \mathcal{U}. \quad (5.46)$$

For the right-hand side of (5.46), we have

$$\mathbf{u}^k - \bar{\mathbf{u}}^k = L(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \quad \text{and} \quad \tilde{G}L = M.$$

Then, we obtain

$$\alpha_k \langle \mathbf{u} - \bar{\mathbf{u}}^k, \tilde{G}(\mathbf{u}^k - \bar{\mathbf{u}}^k) \rangle = \langle \mathbf{u} - \bar{\mathbf{u}}^k, Q(\mathbf{u}^k - \mathbf{u}^{k+1}) \rangle \quad (5.47)$$

from

$$\alpha_k Q^{-1} M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) = \mathbf{u}^k - \mathbf{u}^{k+1}.$$

Setting $a = \mathbf{u}$, $b = \bar{\mathbf{u}}^k$, $c = \mathbf{u}^k$ and $d = \mathbf{u}^{k+1}$ in the identity (1.7) we get:

$$\begin{aligned} & \langle \mathbf{u} - \bar{\mathbf{u}}^k, Q(\mathbf{u}^k - \mathbf{u}^{k+1}) \rangle \\ &= \frac{1}{2} (\|\mathbf{u} - \mathbf{u}^{k+1}\|_Q^2 - \|\mathbf{u} - \mathbf{u}^k\|_Q^2) + \frac{1}{2} (\|\mathbf{u}^k - \bar{\mathbf{u}}^k\|_Q^2 - \|\mathbf{u}^{k+1} - \bar{\mathbf{u}}^k\|_Q^2). \end{aligned} \quad (5.48)$$

Combining (5.46), (5.47), and (5.48) leads to,

$$\begin{aligned} & \alpha_k (\theta(\mathbf{x}) - \theta(\bar{\mathbf{x}}) + \langle \mathbf{u} - \bar{\mathbf{u}}^k, F(\mathbf{u}) \rangle) \\ & \geq \frac{1}{2} (\|\mathbf{u} - \mathbf{u}^{k+1}\|_Q^2 - \|\mathbf{u} - \mathbf{u}^k\|_Q^2) + \frac{1}{2} (\|\mathbf{u}^k - \bar{\mathbf{u}}^k\|_Q^2 - \|\mathbf{u}^{k+1} - \bar{\mathbf{u}}^k\|_Q^2), \forall \mathbf{u} \in \mathcal{U}, \end{aligned} \quad (5.49)$$

and the assertion (5.45) follows from (5.43) and (5.49). $\square$

Next, we are now ready to prove a worst-case $O(1/k)$ convergence rate under an ergodic scenario for Algorithm 5.1.

Theorem 5.3 *Let the sequence $\{\mathbf{u}^k\}$ be generated by Algorithm 5.1, and the vector $\bar{\mathbf{u}}^k$ be defined as (5.38). For a positive integer k , let*

$$\bar{\mathbf{u}}_k := \frac{1}{\Upsilon_k} \sum_{j=0}^k \alpha_j \bar{\mathbf{u}}^j \quad \text{with} \quad \Upsilon_k = \sum_{j=0}^k \alpha_j, \quad (5.50)$$

then $\bar{\mathbf{u}}_k \in \mathcal{U}$ and

$$\theta(\bar{\mathbf{x}}_k) - \theta(\mathbf{x}) + \langle \bar{\mathbf{u}}_k - \mathbf{u}, F(\mathbf{u}) \rangle \leq \frac{1}{2(k+1)\gamma\bar{c}_0} \|\mathbf{u} - \mathbf{u}^0\|_Q^2, \forall \mathbf{u} \in \mathcal{U},$$

where $\bar{c}_0$ (defined in (5.33)) is a lower bound on the step-size α_k^* .

Proof: Summing the inequality (5.45) corresponding to $j = 0, 1, \dots, k$ leads to

$$\begin{aligned} & \left(\Upsilon_k \theta(\mathbf{x}) - \sum_{j=0}^k \alpha_j \bar{\mathbf{x}}^j \right) \\ & + \left\langle \Upsilon_k \mathbf{u} - \sum_{j=0}^k \alpha_j \bar{\mathbf{u}}^j, F(\mathbf{u}) \right\rangle + \frac{1}{2} \|\mathbf{u} - \mathbf{u}^0\|_Q^2 - \frac{1}{2} \|\mathbf{u} - \mathbf{u}^{k+1}\|_Q^2 \geq 0, \forall \mathbf{u} \in \mathcal{U}. \end{aligned}$$

Dropping the negative term from the above inequality, we yield

$$\left(\Upsilon_k \theta(\mathbf{x}) - \sum_{j=0}^k \alpha_j \theta(\bar{\mathbf{x}}^j) \right) + \left\langle \Upsilon_k \mathbf{u} - \sum_{j=0}^k \alpha_j \bar{\mathbf{u}}^j, F(\mathbf{u}) \right\rangle + \frac{1}{2} \|\mathbf{u} - \mathbf{u}^0\|_Q^2 \geq 0, \forall \mathbf{u} \in \mathcal{U},$$

which is equivalent to

$$\begin{aligned} & \left(\frac{1}{\Upsilon_k} \sum_{j=0}^k \alpha_j \theta(\bar{\mathbf{x}}^j) - \theta(\mathbf{x}) \right) + \left\langle \frac{1}{\Upsilon_k} \sum_{j=0}^k \alpha_j \bar{\mathbf{u}}^j - \mathbf{u}, F(\mathbf{u}) \right\rangle \\ & \leq \frac{1}{2\Upsilon_k} \|\mathbf{u} - \mathbf{u}^0\|_Q^2, \forall \mathbf{u} \in \mathcal{U}. \end{aligned} \quad (5.51)$$

For a positive integer k , we obtain $\bar{\mathbf{x}}^j \in \mathcal{X}$ for $j = 0, 1, \dots, k$. Therefore,

$$\bar{\mathbf{x}}^k = \frac{1}{\Upsilon_k} \sum_{j=0}^k \alpha_j \bar{\mathbf{x}}^j$$

can be deemed as a convex combination of $\bar{\mathbf{x}}^j$'s. As $\theta(\mathbf{x})$ is convex, we obtain $\bar{\mathbf{x}}_k \in \mathcal{X}$ and

$$\theta(\bar{\mathbf{x}}_k) \leq \frac{1}{\Upsilon_k} \sum_{j=0}^k \alpha_j \theta(\bar{\mathbf{x}}^j).$$

Substituting the above inequality into (5.51) and using the definition of $\bar{\mathbf{u}}_k$, we yield:

$$\theta(\bar{\mathbf{x}}_k) - \theta(\mathbf{x}) + \langle \bar{\mathbf{u}}_k - \mathbf{u}, F(\mathbf{u}) \rangle \leq \frac{1}{2\Upsilon_k} \|\mathbf{u} - \mathbf{u}^0\|_Q^2, \forall \mathbf{u} \in \mathcal{U}. \quad (5.52)$$

Recall that $\alpha_k \geq \gamma \bar{c}_0$ (see (5.33)) and $\Upsilon_k \geq (k+1)\gamma \bar{c}_0$, and thus,

$$\frac{1}{\Upsilon_k} \leq \frac{1}{(k+1)\gamma \bar{c}_0}.$$

Then,

$$\theta(\bar{\mathbf{x}}_k) - \theta(\mathbf{x}) + \langle \bar{\mathbf{u}}_k - \mathbf{u}, F(\mathbf{u}) \rangle \leq \frac{1}{2(k+1)\gamma \bar{c}_0} \|\mathbf{u} - \mathbf{u}^0\|_Q^2, \forall \mathbf{u} \in \mathcal{U},$$

which completes the proof. $\square$

Hence, for any given compact set $\mathcal{D}_{\mathcal{U}}(\bar{\mathbf{u}}_k) \subset \Omega$, define

$$\bar{d} = \sup\{\|\mathbf{u} - \mathbf{u}^0\|_Q^2 \mid \mathbf{u} \in \mathcal{D}_{\mathcal{U}}(\bar{\mathbf{u}}_k)\},$$

where $\mathbf{u}^0$ is an initialization point. According to Theorem 5.3, we can find $\bar{\mathbf{u}}_k \in \mathcal{U}$ such that

$$\sup_{\mathbf{u} \in \mathcal{D}_{\mathcal{U}}(\bar{\mathbf{u}}_k)} \{\theta(\bar{\mathbf{x}}_k) - \theta(\mathbf{x}) + \langle \bar{\mathbf{u}}_k - \mathbf{u}, F(\mathbf{u}) \rangle\} \leq \frac{\bar{d}}{2(k+1)\gamma \bar{c}_0},$$

and therefore the $O(1/k)$ convergence rate of the proposed Algorithm 5.1 is established in an ergodic sense.

Remark 5.6 When $S_1 = 0$, and $S_i \succeq s\beta A_i^\top A_i$ ($i \geq 2$), we define

$$\bar{d} = \sup\{\|\mathbf{v} - \mathbf{v}^0\|_{Q_0}^2 \mid \mathbf{v} \in \mathcal{D}_{\mathbf{v}}(\bar{\mathbf{v}}_k)\},$$

and can also establish the $O(1/k)$ convergence rate under ergodic sense for Algorithm 5.1, when equipped with the correction step (5.10), by referring the above proofs.

5.4.2 Convergence Rate under Nonergodic Sense

Here, we prove an equivalent result as in the previous subsection for the nonergodic case for Algorithm 5.1, when equipped with the correction step (5.4), under the assumptions that $Q = M$, $S_1 \succ 0$, and $S_i \succeq s\beta A_i^\top A_i$ ($i \geq 2$) ($s \geq \max\{\frac{m-1+\sigma}{4} - 1, 0\}$, $\sigma \geq 1$). The technique used for this analysis is once again motivated by the work in [67]. It is worthwhile to note that we can also establish the $O(1/k)$ convergence rate under nonergodic sense for Algorithm 5.1, when equipped with the correction step (5.10), under the assumptions that $S_1 = 0$, and $S_i \succeq s\beta A_i^\top A_i$ ($i \geq 2$).

Recall that we have shown that the sequence generated by Algorithm 5.1 is strictly contractive with respect to the set $\mathcal{U}^*$ (see (5.35)):

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_M^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|_M^2 - \tau \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2, \quad (5.53)$$

where

$$\tau = \gamma(2 - \gamma)c_0^2 > 0.$$

In order to obtain the worst-case $O(1/k)$ convergence rate under a nonergodic scenario, our proof is composed of the following steps:

- First, we show that the sequence $\{\|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2\}$ is monotonically non-increasing, i.e.,

$$\|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2 \leq \|\mathbf{u}^{k-1} - \tilde{\mathbf{u}}^{k-1}\|_M^2, \quad \forall k \geq 1. \quad (5.54)$$

- Second, we infer a worst-case $O(1/k)$ convergence rate under a nonergodic scenario on the basis of (5.53) and (5.54), i.e.,

$$\|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2 \leq \frac{1}{\tau(k+1)} \|\mathbf{u}^0 - \mathbf{u}^*\|_M^2, \quad \forall \mathbf{u}^* \in \mathcal{U}^*. \quad (5.55)$$

Our rationale for measuring the accuracy of iterates using $\|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2$ is explained in the subsequent analysis.

Lemma 5.8 *Let the sequence $\{\mathbf{u}^k\}$ be generated by Algorithm 5.1, and the accompanying sequence $\{\bar{\mathbf{u}}^k\}$ be defined by (5.38). Then we have*

$$\begin{aligned} & \langle \mathbf{u}^k - \mathbf{u}^{k+1}, \tilde{G}((\mathbf{u}^k - \bar{\mathbf{u}}^k) - (\mathbf{u}^{k+1} - \bar{\mathbf{u}}^{k+1})) \rangle \\ & \geq \frac{1}{2} \|(\mathbf{u}^k - \bar{\mathbf{u}}^k) - (\mathbf{u}^{k+1} - \bar{\mathbf{u}}^{k+1})\|_{(\tilde{G}^\top + \tilde{G})}^2, \end{aligned} \quad (5.56)$$

where $\tilde{G}$ is given by (5.40).

Proof: It follows from (5.42) that

$$\theta(\mathbf{x}) - \theta(\bar{\mathbf{x}}^k) + \langle \mathbf{u} - \bar{\mathbf{u}}^k, F(\bar{\mathbf{u}}^k) \rangle \geq \langle \mathbf{u} - \bar{\mathbf{u}}^k, \tilde{G}(\mathbf{u}^k - \bar{\mathbf{u}}^k) \rangle, \forall \mathbf{u} \in \mathcal{U}. \quad (5.57)$$

This inequality is also true for $k := k + 1$, and thus we get

$$\begin{aligned} & \theta(\mathbf{x}) - \theta(\bar{\mathbf{x}}^{k+1}) + \langle \mathbf{u} - \bar{\mathbf{u}}^{k+1}, F(\bar{\mathbf{u}}^{k+1}) \rangle \\ & \geq \langle \mathbf{u} - \bar{\mathbf{u}}^{k+1}, \tilde{G}(\mathbf{u}^{k+1} - \bar{\mathbf{u}}^{k+1}) \rangle, \forall \mathbf{u} \in \mathcal{U}. \end{aligned} \quad (5.58)$$

Setting $\mathbf{u} = \bar{\mathbf{u}}^{k+1}$ and $\mathbf{u} = \bar{\mathbf{u}}^k$ in (5.57) and (5.58), respectively, and then adding these two resulting inequalities, we have

$$\langle \bar{\mathbf{u}}^k - \bar{\mathbf{u}}^{k+1}, \tilde{G}((\mathbf{u}^k - \bar{\mathbf{u}}^k) - (\mathbf{u}^{k+1} - \bar{\mathbf{u}}^{k+1})) \rangle \geq \langle \bar{\mathbf{u}}^k - \bar{\mathbf{u}}^{k+1}, F(\bar{\mathbf{u}}^k) - F(\bar{\mathbf{u}}^{k+1}) \rangle.$$

Using the monotonicity of F leads to

$$\langle \bar{\mathbf{u}}^k - \bar{\mathbf{u}}^{k+1}, \tilde{G}((\mathbf{u}^k - \bar{\mathbf{u}}^k) - (\mathbf{u}^{k+1} - \bar{\mathbf{u}}^{k+1})) \rangle \geq 0. \quad (5.59)$$

Adding the identity

$$\begin{aligned} & \langle (\mathbf{u}^k - \bar{\mathbf{u}}^k) - (\mathbf{u}^{k+1} - \bar{\mathbf{u}}^{k+1}), \tilde{G}((\mathbf{u}^k - \bar{\mathbf{u}}^k) - (\mathbf{u}^{k+1} - \bar{\mathbf{u}}^{k+1})) \rangle \\ & = \frac{1}{2} \|(\mathbf{u}^k - \bar{\mathbf{u}}^k) - (\mathbf{u}^{k+1} - \bar{\mathbf{u}}^{k+1})\|_{(\tilde{G}^\top + \tilde{G})}^2 \end{aligned}$$

to both sides of (5.59), and with some simple algebraic manipulations, we obtain (5.56) and this completes the proof of this lemma. $\square$

Using the assertion (5.56) in Lemma 5.8 provides us with a useful estimate for measuring the difference $\|\mathbf{u}^k - \bar{\mathbf{u}}^k\|_M^2 - \|\mathbf{u}^{k+1} - \bar{\mathbf{u}}^{k+1}\|_M^2$, which is summarized in the following lemma.

Lemma 5.9 *Let the sequence $\{\mathbf{u}^k\}$ be generated by Algorithm 5.1, and the accompanying sequence $\{\tilde{\mathbf{u}}^k\}$ be defined by (5.38). Then we have*

$$\begin{aligned} \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2 - \|\mathbf{u}^{k+1} - \tilde{\mathbf{u}}^{k+1}\|_M^2 &\geq \frac{1}{\alpha_k} \|L[(\mathbf{u}^k - \tilde{\mathbf{u}}^k) - (\mathbf{u}^{k+1} - \tilde{\mathbf{u}}^{k+1})]\|_{(\tilde{G}^\top + \tilde{G})}^2 \\ &\quad - \|(\mathbf{u}^k - \tilde{\mathbf{u}}^k) - (\mathbf{u}^{k+1} - \tilde{\mathbf{u}}^{k+1})\|_M^2, \end{aligned} \quad (5.60)$$

where $\tilde{G}$ is given by (5.40).

Proof: By using $\mathbf{u}^k - \tilde{\mathbf{u}}^k = L(\mathbf{u}^k - \tilde{\mathbf{u}}^k)$ (see (5.41)) in (5.56), we have

$$\begin{aligned} &2\langle \mathbf{u}^k - \mathbf{u}^{k+1}, \tilde{G}L((\mathbf{u}^k - \tilde{\mathbf{u}}^k) - (\mathbf{u}^{k+1} - \tilde{\mathbf{u}}^{k+1})) \rangle \\ &\geq \|L[(\mathbf{u}^k - \tilde{\mathbf{u}}^k) - (\mathbf{u}^{k+1} - \tilde{\mathbf{u}}^{k+1})]\|_{(\tilde{G}^\top + \tilde{G})}^2. \end{aligned} \quad (5.61)$$

Using

$$\mathbf{u}^k - \mathbf{u}^{k+1} = \alpha_k(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \quad \text{and} \quad \tilde{G}L = M$$

in (5.61), we infer that

$$\begin{aligned} &2\langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, M((\mathbf{u}^k - \tilde{\mathbf{u}}^k) - (\mathbf{u}^{k+1} - \tilde{\mathbf{u}}^{k+1})) \rangle \\ &\geq \frac{1}{\alpha_k} \|L[(\mathbf{u}^k - \tilde{\mathbf{u}}^k) - (\mathbf{u}^{k+1} - \tilde{\mathbf{u}}^{k+1})]\|_{(\tilde{G}^\top + \tilde{G})}^2. \end{aligned} \quad (5.62)$$

On the other hand, by setting $a = (\mathbf{u}^k - \tilde{\mathbf{u}}^k)$ and $b = (\mathbf{u}^{k+1} - \tilde{\mathbf{u}}^{k+1})$ in the equality (1.6), we have

$$\begin{aligned} \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2 - \|\mathbf{u}^{k+1} - \tilde{\mathbf{u}}^{k+1}\|_M^2 &= 2\langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, M((\mathbf{u}^k - \tilde{\mathbf{u}}^k) - (\mathbf{u}^{k+1} - \tilde{\mathbf{u}}^{k+1})) \rangle \\ &\quad - \|(\mathbf{u}^k - \tilde{\mathbf{u}}^k) - (\mathbf{u}^{k+1} - \tilde{\mathbf{u}}^{k+1})\|_M^2. \end{aligned} \quad (5.63)$$

Substituting (5.62) into the right hand side of (5.63) leads to the assertion of this lemma.

□

In order to show the monotonicity of $\{\|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2\}$, it is sufficient to show that the right hand side of (5.60) is nonnegative, which is proved in the following lemma.

Lemma 5.10 *Given the matrices G , M , and L , we have*

$$L^\top(\tilde{G}^\top + \tilde{G})L - \alpha M \succeq 0. \quad (5.64)$$

where $0 < \alpha \leq 2 - \sqrt{\frac{m-1}{s+1}}$.

Proof: Since $\tilde{G}L = M$ and M is symmetric, we get

$$L^\top(\tilde{G}^\top + \tilde{G})L = ML + L^\top M.$$

From the above equation and the expansion of $L^\top M$, we obtain

$$\begin{aligned} & L^\top(\tilde{G}^\top + \tilde{G})L - \alpha M \\ &= \begin{pmatrix} (2-\alpha)S_1 & 0 & \dots & 0 & 0 \\ 0 & (2-\alpha)(S_2 + \beta A_2^\top A_2) & \dots & 0 & A_2^\top \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & (2-\alpha)(S_m + \beta A_m^\top A_m) & A_m^\top \\ 0 & A_2 & \dots & A_m & (2-\alpha)\beta^{-1}I_l \end{pmatrix}. \end{aligned}$$

Then, using the assumptions on $S_i \geq s\beta A_i^\top A_i (i \geq 2)$ yields

$$\begin{aligned} & L^\top(\tilde{G}^\top + \tilde{G})L - \alpha M \\ &\succeq \begin{pmatrix} (2-\alpha)S_1 & 0 & \dots & 0 & 0 \\ 0 & (2-\alpha)(s+1)\beta A_2^\top A_2 & \dots & 0 & A_2^\top \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & (2-\alpha)(s+1)\beta A_m^\top A_m & A_m^\top \\ 0 & A_2 & \dots & A_m & (2-\alpha)\beta^{-1}I_l \end{pmatrix}. \end{aligned}$$

In addition, making use of the notation

$$\mathcal{A} = \text{diag}(S_1^{1/2}, \beta^{1/2}A_2, \dots, \beta^{1/2}A_m, \beta^{-1/2}),$$

we have

$$\begin{aligned} & L^\top(\tilde{G}^\top + \tilde{G})L - \alpha M \\ &\succeq \mathcal{A}^\top \begin{pmatrix} (2-\alpha)I_l & 0 & \dots & 0 & 0 \\ 0 & (2-\alpha)(s+1)I_l & \dots & 0 & I_l \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & (2-\alpha)(s+1)I_l & I_l \\ 0 & I_l & \dots & I_l & (2-\alpha)I_l \end{pmatrix} \mathcal{A}, \end{aligned}$$

where $I_l \in \mathbb{R}^{l \times l}$ is an identity matrix. Continuing in this manner, the proof of (5.64) can be transformed to prove that the matrix $T \succeq 0$, where

$$T = \begin{pmatrix} (2-\alpha)(s+1)I_{m_1} & e \\ e^\top & (2-\alpha) \end{pmatrix}, \quad (5.65)$$

$m_1 = m - 1$, and $e \in \mathbb{R}^{m_1}$ is a vector of all ones. Let $Tz = \delta z$, where δ is an eigenvalue of matrix T and z is its corresponding eigenvector. Next, we will show that all the eigenvalues of T are non-negative. Without loss of generality, the eigenvectors of T are assumed with the following forms

$$z = \begin{pmatrix} y \\ 0 \end{pmatrix} \quad \text{or} \quad z = \begin{pmatrix} y \\ 1 \end{pmatrix},$$

where $y \in \mathbb{R}^{m_1}$. In the first instance, $z^\top = (y^\top, 0)$, combining $Tz = \delta z$ and (5.65) leads to

$$\begin{cases} (2 - \alpha)(s + 1)y = \delta y \\ e^\top y = 0 \end{cases}.$$

Thus, we deduce that $(m - 2)$ linearly independent vectors, y^i , associated with $i = 1, \dots, m - 2$, are the orthogonal vectors of e , and moreover,

$$z^i = \begin{pmatrix} y^i \\ 0 \end{pmatrix}, i = 1, \dots, m - 2,$$

are the eigenvectors of T with the corresponding eigenvalues

$$\delta_1 = \delta_2 = \dots = \delta_{m-2} = (2 - \alpha)(s + 1) > 0.$$

In the other case, $z^\top = (y^\top, 1)$, and from $Tz = \delta z$ and (5.65), we have

$$\begin{cases} (2 - \alpha)(s + 1)y + e = \delta y & (5.66a) \\ e^\top y + 2 - \alpha = \delta & (5.66b) \end{cases}$$

The equation (5.66a) is left-multiplied by $e^\top$, and then utilizing the equation (5.66b) and $e^\top e = m - 1$ yields:

$$\delta^2 - (2 - \alpha)(s + 2)\delta + [(2 - \alpha)^2(s + 1) - (m - 1)] = 0.$$

Thus, the minimum eigenvalue of T , denoted by $\delta_{\min}(T)$, can be evaluated as:

$$\delta_{\min}(T) = \frac{(2 - \alpha)(s + 2) - \sqrt{(2 - \alpha)^2(s + 2)^2 - 4[(2 - \alpha)^2(s + 1) - (m - 1)]}}{2}.$$

Since $\alpha \leq 2 - \sqrt{\frac{m-1}{s+1}}$, we have

$$0 \leq 4[(2 - \alpha)^2(s + 1) - (m - 1)] \leq (2 - \alpha)^2(s + 2)^2,$$

and thus $\delta_{\min}(T) \geq 0$, and the lemma is proved. $\square$

Under the condition that $0 < \alpha_k \leq 2 - \sqrt{\frac{m-1}{s+1}}$, the monotonicity of $\{\|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2\}$ is a straightforward extension of Lemma 5.9 and Lemma 5.10. Then, we are prepared to estimate a worst-case convergence rate $O(1/k)$ under a nonergodic scenario for Algorithm 5.1.

Theorem 5.4 *Let the sequence $\{\mathbf{u}^k\}$ be generated by Algorithm 5.1 with the following requirement on step size:*

$$\alpha_k = \begin{cases} \gamma\alpha_k^*, & \text{if } \gamma\alpha_k^* \leq 2 - \sqrt{\frac{m-1}{s+1}}, \\ \gamma(1 - \sqrt{\frac{m-1}{m}}), & \text{otherwise} \end{cases},$$

where $\gamma \in (0, 2)$ and $s \geq \max\{\frac{m-1+\sigma}{4} - 1, 0\}$, $\sigma \geq 1$. Then,

$$\|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2 \leq \frac{1}{\tau(k+1)} \|\mathbf{u}^0 - \mathbf{u}^*\|_M^2, \forall \mathbf{u}^* \in \mathcal{U}^*, \quad (5.67)$$

where τ is a constant.

Proof: If $\alpha_k = \gamma\alpha_k^*$, the inequality (5.53) can be obtained in a straightforward manner and we are done. Hence, we only ought to take into account the case where $\alpha_k = \gamma(1 - \sqrt{\frac{m-1}{m}})$. Then, it follows from inequality (5.35) and Remark 5.3 that

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_M^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|_M^2 - \gamma(2 - \gamma)c_0^2 \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2.$$

Therefore, the inequality (5.53) can also be established under the restriction placed on α_k . Then, it follows from Lemmas (5.9) and (5.10) that

$$\tau \sum_{i=0}^{\infty} \|\mathbf{u}^i - \tilde{\mathbf{u}}^i\|_M^2 \leq \|\mathbf{u}^0 - \mathbf{u}^*\|_M^2, \forall \mathbf{u}^* \in \mathcal{U}^*, \quad (5.68)$$

where $\tau = \gamma(2 - \gamma)c_0^2$. As $0 < \alpha_k \leq 2 - \sqrt{\frac{m-1}{s+1}}$, we then deduce that the sequence $\{\|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2\}$ is monotonically non-increasing. Thus, we yield

$$(k+1) \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2 \leq \sum_{i=0}^k \|\mathbf{u}^i - \tilde{\mathbf{u}}^i\|_M^2. \quad (5.69)$$

Then, the assertion (5.64) follows from (5.68) and (5.69). $\square$

Since $\mathcal{U}^*$ is nonempty, convex and closed set, let $d := \inf\{\|\mathbf{u}^0 - \mathbf{u}^*\|_M^2 \mid \mathbf{u}^* \in \mathcal{U}^*\}$. After $(k+1)$ iterations of Algorithm 5.1, we will then have $\|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2 \leq \frac{d}{\tau(k+1)} = O(1/k)$. Therefore, $\mathbf{u}^k$ is a solution of $\text{VI}(\mathcal{U}, F, \theta)$ if and only if $\|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_M^2 = 0$, and the worst-case $O(1/k)$ convergence rate of Algorithm 5.1 under a nonergodic sense is derived by Theorem 5.4.

5.5 Numerical Results

Having theoretically established the global convergence of the proposed method, we now turn our attention towards a computational study of this algorithm, and gauge its numerical performance by solving several synthetically generated instances of the robust principal component analysis (PCA) problem. All of our computational experiments were performed in MATLAB R2011b on a Windows 7 machine equipped with an Intel Core i5-3740 CPU, and 4GB of memory.

5.5.1 Robust PCA Problem

In this subsection, we apply Algorithm 5.1 for solving the unconstrained version of (1.18). Note that it is not easy to minimize (1.18) directly because of the coupling of X and Y and the non-smoothness of $\|X\|_*$ and $\|Y\|_1$. Fortunately, the unconstrained model (1.19) can be equivalently rewritten as:

$$\min \quad \|X\|_* + \tau\|Y\|_1 + \frac{1}{2\mu}\|P_\Omega(Z)\|_F^2 \quad (5.70a)$$

$$\text{s.t. } X + Y + Z = N, \quad (5.70b)$$

by introducing an auxiliary variable $Z \in \mathbb{R}^{p \times q}$. Now, the objective function (5.70a) has the required separable structure, and moreover Problem (5.70) is seen as a special case of (1.1) involving matrix variables with $m = 3$, where

$$(x_1, x_2, x_3) := (X, Y, Z), \quad A_i := I_i \ (i = 1, 2, 3), \quad b := N,$$

and

$$\theta_1(X) := \|X\|_*, \quad \theta_2(Y) := \tau\|Y\|_1, \quad \theta_3(Z) := \frac{1}{2\mu}\|P_\Omega(Z)\|_F^2.$$

Therefore, the proposed method is also applicable to Problem (5.70), with its resulting subproblems being simple enough to possess closed-form solutions.

In the following, we analyze the explicit forms of the resulting subproblems when Algorithm 5.1 is applied to solve Problem (5.70). First, two operators will be summarized for the deducing of the analytical solutions as in [62, 121]. For any $c > 0$ and $E \in \mathbb{R}^{p \times q}$, the shrinkage operator ‘shrink($\cdot, \cdot$)’ for matrix variables is given by

$$(\text{shrink}(E, c))_{ij} := \text{sign}(E_{ij}) \cdot \max\{|E_{ij}| - c, 0\}, 1 \leq i \leq p, 1 \leq j \leq q, \forall E \in \mathbb{R}^{p \times q},$$

and the operator $\mathcal{D}_c : \mathbb{R}^{p \times q} \rightarrow \mathbb{R}^{p \times q}$ is given by

$$\mathcal{D}_c(E) := U \text{diag}(\text{shrink}(\sum, c)) V^\top, \forall E \in \mathbb{R}^{p \times q},$$

where $U \sum V^\top$ is the singular value decomposition (SVD) of E . Let $\Lambda^k \in \mathbb{R}^{p \times q}$ be the Lagrangian multiplier associated to the linear constraints (5.70b), and set $S_1 = 0$, and $S_i = s\beta A_i^\top A_i$ ($i \geq 2, s > \max\{\frac{m}{4} - 1, 0\}$). Then, the explicit representations of the three subproblems arising from applying Algorithm (5.1) to Problem (5.70) are summarized as follows:

- The explicit solution of $\tilde{Y}^k$ is represented by the shrinkage operator

$$\begin{aligned} \tilde{Y}^k &= \arg \min_Y \left\{ \tau \|Y\|_1 + \frac{\beta}{2} \|Y + X^k + Z^k - N - \frac{1}{\beta} \Lambda^k\|_F^2 \right\} \\ &= \text{shrink}(N + \frac{1}{\beta} \Lambda^k - X^k - Z^k, \frac{\tau}{\beta}), \end{aligned}$$

- The solution $\tilde{X}^k$ can be given by

$$\begin{aligned} \tilde{X}^k &= \arg \min_X \left\{ \|X\|_* + \frac{\beta}{2} \|X + \tilde{Y}^k + Z^k - N - \frac{1}{\beta} \Lambda^k\|_F^2 + \frac{\beta}{2} \|X - X^k\|_F^2 \right\} \\ &= \arg \min_X \left\{ \|X\|_* + \frac{(s+1)\beta}{2} \|X - X^k - \frac{1}{(s+1)\beta} \hat{\Lambda}^k\|_F^2 \right\} \\ &= \mathcal{D}_{\frac{1}{(s+1)\beta}}(X^k + \frac{1}{(s+1)\beta} \hat{\Lambda}^k). \end{aligned}$$

where $\hat{\Lambda}^k = \Lambda^k - \beta(\tilde{Y}^k + X^k + Z^k - N)$.

- The Z -subproblem is given by the following form

$$\begin{aligned}\tilde{Z}^k &= \arg \min_Z \left\{ \frac{1}{2\mu} \|P_\Omega(Z)\|_F^2 + \frac{\beta}{2} \|\tilde{Y}^k + X^k + Z - N - \frac{1}{\beta} \Lambda^k\|_F^2 + \frac{s\beta}{2} \|Z - Z^k\|_F^2 \right\} \\ &= \arg \min_Z \left\{ \frac{1}{2\mu} \|P_\Omega(Z)\|_F^2 + \frac{(s+1)\beta}{2} \|Z - Z^k - \frac{1}{(s+1)\beta} \hat{\Lambda}^k\|_F^2 \right\}.\end{aligned}$$

Then, the closed form solution of Z is described as follows:

$$\tilde{Z}_{ij}^k = \begin{cases} \hat{Z}_{ij}^k, & \text{if } (i, j) \notin \Omega \\ \frac{(s+1)\mu\beta}{1+(s+1)\mu\beta} \hat{Z}_{ij}^k, & \text{if } (i, j) \in \Omega \end{cases},$$

where $\hat{Z}^k = Z^k + \frac{1}{(s+1)\beta} \hat{\Lambda}^k$.

- The dual variable $\tilde{\Lambda}^k$ can be updated by

$$\tilde{\Lambda}^k = \Lambda^k - \beta(\tilde{X}^k + \tilde{Y}^k + \tilde{Z}^k - N).$$

To demonstrate the feasibility of our new algorithm towards solving Problem (5.70), we synthetically generate the test data by following the method in [121]. More specifically, let $N = X^* + Y^*$ be an observation matrix, where X^* and Y^* are the low-rank and sparse components to be recovered, respectively. X^* is constructed as $X^* = L^\top R$, where L and R are independently of full row rank (with rank ϱ), whose entries are independent and identically distributed Gaussian random variables with zero mean and unit standard deviation. Therefore, the rank of X^* is ϱ . The index set Ω is determined randomly with a preset sample ratio (denoted by **sr**), which is defined as $\mathbf{sr} := \frac{100|\Omega|}{pq}\%$, where $|\Omega|$ is the cardinality of Ω . For the sparse part Y^* , its nonzero entries are generated uniformly in the interval $[-500, 500]$. Then, the observation data was corrupted by adding Gaussian noise with standard deviation $\sigma = 10^{-3}$. For the parameters involved in Problem (5.70), we set $\tau = 1/\sqrt{p}$ and $\mu = \sqrt{p + \sqrt{8p\sigma}}/10$, as used in [121]. Here, we mainly consider the case where $\mathbf{sr} := 80\%$ and $p = q$, where $p \in \{500; 1000; 1500\}$. Let **rr** be the low-rank ratio (i.e., $\mathbf{rr} := \varrho/p$), and let **spr** be the sparsity ratio (i.e., $\mathbf{spr} := \frac{\|Y^*\|_0}{pq}$, where $\|\cdot\|_0$ represents the number of nonzero entries). Accordingly, we test two scenarios with $\mathbf{rr} \in \{0.05; 0.10\}$ and three scenarios with $\mathbf{spr} \in \{0.05; 0.10; 0.15\}$, respectively.

In order to demonstrate its efficiency, we compare the proposed proximal partially-parallel splitting algorithm (namely PPSA) with the relaxed splitting method in [89] (namely PPSM), and VASALM [121]. Throughout, the stopping criterion used is

$$\max \left\{ \frac{\|\tilde{X}^k - X^k\|_F}{1 + \|X^k\|_F}, \frac{\|\tilde{Y}^k - Y^k\|_F}{1 + \|Y^k\|_F}, \frac{\|\tilde{\Lambda}^k - \Lambda^k\|_F}{1 + \|\Lambda^k\|_F} \right\} < \epsilon,$$

where $\epsilon > 0$ is a small preset tolerance. In our numerical experiments, all initial points are set as $(X^0, Y^0, \Lambda^0) = (\mathbf{0}, \mathbf{0}, \mathbf{0})$, and the tolerance $\epsilon = 10^{-5}$. Furthermore, we choose the same value of the penalty parameter $\beta = \hat{t}|\Omega|/\|P_\Omega(D)\|_1$ for all test scenarios. Specifically, we set $\hat{t} = 0.05, 0.10$, and 0.20 for the scenarios of **spr** = 0.05, 0.10, and 0.15, respectively. Finally, we set $r = 2.01$, $s = 1.0$, and $\gamma = 1.2$ for PPSM; $\eta = 2.01$ for VASALM; and $s = 0.8$ and $\gamma = 1.5$ for the proposed new method. The corresponding results are summarized in the tables below, where “Iter.”, “Time”, and “Tol” represent the number of iterations, the CPU time (in seconds), and the relative error of the recovered low-rank component (i.e., $\|X^k - X^*\|_F/\|X^*\|_F$), measured at optimality, respectively.

Finally, note that as these problem instances are synthetically generated with random data, 10 instances corresponding to each dimensional configuration are generated and the average performance for each scenario is reported. From the results displayed in Table 5.1, it can be seen that PPSA outperforms the other two methods by taking the fewest number of iterations to converge, as well as the least CPU time for most of instances. Note that VASALM is a relatively simpler scheme as compared to PPSA due to the additional step size computation involved in the proposed method. However, the good news is that the computational cost of calculating this step size is very inexpensive as compared to the singular value decomposition required by VASALM. Therefore, we can see that PPSA performs better than VASALM in terms of taking fewer iteration numbers to obtain an approximate solution, which results in a reduced computing time for the proposed method. Moreover, the PPSA also achieves better solutions than the other two methods by virtue of its smaller (low-rank) relative error at optimality for most of these instances.

Next, Fig 5.1 plots the evolution curves of the relative errors of the recovered low-rank and sparse components (i.e. $\|Y^k - Y^*\|_F/\|Y^*\|_F$), the rank of X^k , and the sparsity of Y^k as a function of the number of iterations, and Fig 5.2 plots the evolution curves of

Table 5.1: Numerical comparisons based on synthetically generated data.

p	spr	rr	PPSM			PPSA			VASALM		
			Iter.	Time	Tol	Iter.	Time	Tol	Iter.	Time	Tol
500	0.05	0.05	56.8	25.9	1.22e-4	54.3	28.0	9.82e-5	59.3	28.0	1.48e-4
	0.05	0.10	57.9	24.4	1.79e-4	50.0	24.1	1.70e-4	65.7	29.5	1.94e-4
500	0.10	0.05	61.2	26.3	1.42e-4	56.7	27.5	1.22e-4	68.0	30.7	1.57e-4
	0.10	0.10	71.8	33.8	1.87e-4	58.2	31.2	1.85e-4	79.2	40.9	2.22e-4
500	0.15	0.05	70.0	30.1	1.26e-4	63.3	29.7	1.11e-4	75.5	33.3	1.35e-4
	0.15	0.10	115.5	43.4	1.31e-3	87.8	38.3	1.35e-3	127.3	52.2	1.38e-3
1000	0.05	0.05	65.5	147.1	6.58e-5	60.0	149.0	6.19e-5	69.0	161.1	7.22e-5
	0.05	0.10	72.2	162.0	7.90e-5	62.4	153.8	7.45e-5	80.4	187.8	8.45e-5
1000	0.10	0.05	72.7	174.5	7.10e-5	65.2	166.5	6.83e-5	76.2	184.9	7.94e-5
	0.10	0.10	84.4	208.8	8.32e-5	70.0	189.7	8.06e-5	91.8	241.4	9.35e-5
1000	0.15	0.05	84.1	206.4	6.79e-5	76.4	207.4	6.62e-5	88.1	221.8	7.19e-5
	0.15	0.10	99.5	239.8	1.14e-4	82.8	218.0	1.12e-4	110.6	274.9	1.22e-4
1500	0.05	0.05	73.5	564.6	5.14e-5	68.0	546.4	5.05e-5	75.4	584.6	5.58e-5
	0.05	0.10	82.9	691.5	5.81e-5	72.1	633.1	5.65e-5	88.3	730.9	6.55e-5
1500	0.10	0.05	81.2	619.1	5.34e-5	74.9	603.5	5.25e-5	83.2	642.9	5.90e-5
	0.10	0.10	94.7	701.1	6.17e-5	81.3	641.2	6.04e-5	101.2	776.4	6.94e-5
1500	0.15	0.05	93.6	702.1	5.41e-5	87.0	691.2	5.39e-5	96.1	740.5	5.84e-5
	0.15	0.10	116.2	870.3	6.25e-5	98.1	782.1	6.13e-5	126.1	974.2	6.81e-5

the relative errors of linear constraint (i.e. $\|X^k + Y^k + Z^k - N\|_F / \|N\|_F$) as a function of the number of iterations, respectively, for $\text{spr} \in \{0.10, 0.15\}$ and $\text{rr} = 0.10$.

The graphs in Fig 5.1 and Fig 5.2 demonstrate that our proposed method (PPSA) outperforms VASALM and PPSM by converging in fewer number of iterations to the desired optimal value (within the prescribed tolerance level). We attribute this improved performance to our mechanism of adopting a dynamic optimal step size during the correction phase of the algorithm.

Furthermore, we studied the influence of the relaxation parameter γ on Algorithm 5.1. Table 5.2 demonstrates the numerical results obtained by solving the test instances using Algorithm 5.1 with three different values of γ . It is apparent that the numerical performance of Algorithm 5.1 with $\gamma \geq 1$ is better when compared to the case of $\gamma < 1$. (This is reflected in our choice of $\gamma = 1.5$ in our computational comparisons with other algorithms reported in Table 5.1.)

5.5.2 Linear Equation Example

Having demonstrated in Section 5.5.1 that the proposed algorithm is competitive with other state-of-the-art parallel splitting algorithms, we now turn our attention towards

Figure 5.1: Evolution of relative errors of the recovered low-rank and sparse components (the above four subplots), the rank of X^k and sparsity of Y^k (the below four subplots) with respect to the number of iterations.

studying the numerical behavior of PPSA for solving some specialized problems, for which the EADM has been shown to be divergent. In this context, we apply Algorithm 5.1 to solve a system of linear homogeneous equations in three variables proposed in [18], which takes the form of

$$A_1x_1 + A_2x_2 + A_3x_3 = 0, \quad (5.71)$$

Figure 5.2: Evolution of relative errors of the linear constraint with respect to the number of iterations.

Table 5.2: Performance of Algorithm 5.1 with different relaxation factors γ

p	$\mathbf{spr}$	$\mathbf{rr}$	$\gamma = 0.8$			$\gamma = 1.2$			$\gamma = 1.6$		
			Iter.	Time	Tol	Iter.	Time	Tol	Iter.	Time	Tol
500	0.05	0.05	56.6	23.3	1.18e-4	53.8	25.2	1.02e-4	60.6	26.7	1.29e-4
500	0.05	0.10	59.1	23.9	1.78e-4	51.9	24.3	1.61e-4	67.8	30.8	1.93e-4
500	0.10	0.05	62.4	26.3	1.32e-4	57.9	27.2	1.12e-4	67.3	31.1	1.55e-4
500	0.10	0.10	74.1	29.8	1.92e-4	58.8	28.3	2.05e-4	81.9	37.0	2.31e-4
500	0.15	0.05	69.1	27.6	1.25e-4	63.2	27.3	1.12e-4	74.9	33.4	1.36e-4
500	0.15	0.10	121.1	52.1	1.23e-3	89.7	45.2	1.30e-3	131.8	64.0	1.33e-3
1000	0.05	0.05	65.9	149.9	6.70e-5	60.0	155.9	6.40e-5	68.9	172.1	7.58e-5
1000	0.05	0.10	72.8	154.7	7.63e-5	61.7	142.1	7.58e-5	78.8	180.8	8.79e-5
1000	0.10	0.05	72.5	151.4	7.01e-5	65.5	153.7	6.70e-5	76.3	172.3	7.83e-5
1000	0.10	0.10	83.7	199.4	8.37e-5	70.3	203.9	8.01e-005	90.8	246.0	9.64e-5
1000	0.15	0.05	83.8	198.1	6.76e-5	76.2	195.1	6.59e-5	88.1	223.5	7.16e-5
1000	0.15	0.10	100.4	225.4	8.53e-5	83.4	202.9	8.12e-5	109.6	262.2	9.27e-5
1500	0.05	0.05	72.9	582.7	5.13e-5	68.0	572.2	4.97e-5	75.6	644.6	5.54e-5
1500	0.05	0.10	83.2	735.5	5.82e-5	72.1	675.0	5.63e-5	88.7	792.5	6.58e-5
1500	0.10	0.05	80.8	692.2	5.35e-5	74.9	672.1	5.27e-5	83.6	742.5	5.83e-5
1500	0.10	0.10	94.4	813.1	6.21e-5	81.3	749.9	6.03e-5	101.0	886.1	6.99e-5
1500	0.15	0.05	93.6	817.8	5.36e-5	87.0	812.9	5.31e-5	96.3	884.5	5.83e-5
1500	0.15	0.10	116.3	974.1	6.31e-5	97.9	872.9	6.24e-5	125.6	1091.9	6.99e-5

where,

$$A_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, A_2 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, A_3 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}.$$

Clearly, (5.71) is a special case of (1.1) with a null objective function. Correspondingly, setting $S_1 = 0$, $S_2 = s\beta A_2^\top A_2$ and $S_3 = s\beta A_3^\top A_3$, the iterative schemes of Algorithm 5.1 for (5.71) read as:

$$\begin{cases} -A_1^\top \lambda^k + \beta A_1^\top (A_1 \tilde{x}_1^k + A_2 x_2^k + A_3 x_3^k) = 0 \\ -A_2^\top \lambda^k + \beta A_2^\top (A_1 \tilde{x}_1^k + A_2 \tilde{x}_2^k + A_3 x_3^k) + s\beta A_2^\top A_2 (\tilde{x}_2^k - x_2^k) = 0 \\ -A_3^\top \lambda^k + \beta A_3^\top (A_1 \tilde{x}_1^k + A_2 x_2^k + A_3 \tilde{x}_3^k) + s\beta A_3^\top A_3 (\tilde{x}_3^k - x_3^k) = 0 \\ \tilde{\lambda}^k = \lambda^k - \beta (A_1 \tilde{x}_1^k + A_2 \tilde{x}_2^k + A_3 \tilde{x}_3^k - b), \end{cases}$$

which can also be rewritten explicitly as follows:

$$\begin{cases} \tilde{x}_1^k = (A_1^\top A_1)^{-1} [A_1^\top (\lambda^k / \beta - A_2 x_2^k - A_3 x_3^k)] \\ \tilde{x}_2^k = ((s+1)A_2^\top A_2)^{-1} [A_2^\top (\lambda^k / \beta + s x_2^k - A_1 \tilde{x}_1^k - A_3 x_3^k)] \\ \tilde{x}_3^k = ((s+1)A_3^\top A_3)^{-1} [A_3^\top (\lambda^k / \beta + s x_3^k - A_1 \tilde{x}_1^k - A_2 x_2^k)] \\ \tilde{\lambda}^k = \lambda^k - \beta (A_1 \tilde{x}_1^k + A_2 \tilde{x}_2^k + A_3 \tilde{x}_3^k - b). \end{cases}$$

We once again compare Algorithm 5.1 with PPSM and VASALM methods, where the various parameters for these methods are set as follows: We take $r = 4$, $s = 1.0$, and $\gamma = 1.2$ for PPSM, $\eta = 4$ for VASALM; and $s = 1$ and $\gamma = 1.9$ for PPSA (Algorithm 5.1). In Fig 5.3, the evolution of the primal variables and relative error, which is defined by

$$\text{RelError} := \log \left(\max \left\{ \max_{i=1,2,3} \{ \|A_i(x_i^k - x_i^{k+1})\|_\infty \}, \|\lambda^k - \lambda^{k+1}\|_\infty \right\} \right),$$

is displayed, and it can be easily seen from the plots in Fig 5.3 that the proposed PPSA (Algorithm 5.1) converges in the least number of iterations to the desired optimal value. More importantly, the proposed algorithm exhibits its convergent behavior even on problems that are capable of stalling other types of splitting algorithms.

5.6 Conclusions

For solving the separable convex minimization problem (1.1), we put forward a new parallel splitting method (Algorithm 5.1) based on proximal regularization techniques and some partially-parallel splitting schemes. In addition, we proved the global convergence of the proposed algorithm and the worst-case $O(1/k)$ convergence rate was also established under both ergodic and nonergodic scenarios. The efficiency of the PPSA was also demonstrated via illustrative numerical results.

Figure 5.3: Evolution of the values of primal variables and tolerance with respect to the number of iterations.

Chapter 6

Linearized Partially-Parallel Splitting Methods

6.1 Motivation

In [70] and [55], the authors, under the predictor-corrector framework, presented different *linearized versions* of proximal splitting methods by using Gauss-Seidel decomposition and Jacobian decomposition schemes, respectively, and proved the global convergence of these methods under some mild assumptions. In Chapter 5, a new *proximal partially-parallel splitting algorithm* (PPSA) was presented under the predictor-corrector framework, which is a hybrid mechanism that inherits the nice properties of both Gauss-Seidel and Jacobian substitution procedures for solving the multiple-block convex minimization problem, whose objective function is the sum of m individual functions without any coupled variables, subject to a linear coupling constraint. Clearly, the efficiency of such a splitting method relies heavily on the solvability of the underlying subproblems. However, the resulting subproblems in most cases cannot be solved efficiently.

To circumvent this issue, in this chapter, we resort to utilizing *linearization techniques* which are judiciously selected such that the resulting subproblems (5.2) either possess a *closed-form* expression or are rendered simple enough to be solved relatively efficiently as compared to their nonlinear counterparts. Note that these linearized subproblems are essentially approximations of the subproblems (5.2), whose solutions might not be the optimal choice for the next iterate. Therefore, with the help of an additional corrector step, we propose to reduce the approximation errors caused by the selected linearization

strategy. Although the computational cost is expected to slightly increase as a byproduct of prescribing this additional corrector step, we rationalize that the overall computational time consumed by the algorithm reduces as the number of iterations taken to converge decrease (as compared to directly implementing (5.2)).

Specifically, in this chapter, we design three linearized partially parallel-splitting methods and analyze their global convergence properties. We prove the worst-case $O(1/t)$ convergence rate under ergodic scenarios, and finally, the numerical performance of the proposed methods is also displayed on some example problems.

6.2 A Set of Linearized PPSA-based Methods and Related Properties

In this section, we will present a portfolio of *linearized* versions of the *proximal partially parallel splitting algorithm* (**PPSA**), recently introduced in Chapter 5, and prove the necessary global convergence results for these methods. It is worthwhile to note that the nonlinear terms in the augmented Lagrangian function may lead to cumbersome subproblems if $\mathcal{X}_i \neq \mathbb{R}_i$ or if the matrices A_i ($i = 1, \dots, m$) are not scalar matrices. Therefore, our motivation for this work stems from the fact that by linearizing selected nonlinear terms, we can potentially simplify the resulting subproblems and/or derive closed-form solutions, particularly when dealing with certain types of some sparse optimization problems.

6.2.1 Linearized version A of PPSA (LPSM-A)

In this subsection, we introduce the first of our linearized version, namely version A, of the proximal partially parallel splitting algorithm (Algorithm 5.1) by linearizing the quadratic terms in the augmented Lagrangian function. Towards this end, define a matrix M_1 as follows:

$$M_1 = \text{diag}(S_1 - \beta A_1^\top A_1, S_2, \dots, S_m, \beta^{-1} I_l), \quad (6.1)$$

where $S_1 \succ \beta A_1^\top A_1$ and S_i ($i = 1, \dots, m$) are positive definite matrices. Hence, M_1 is also positive definite. In addition, based on the convergence proof framework for contraction

methods [63, 70], we are required to define:

$$\varphi_A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) = \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_{M_1}^2 + \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle, \quad (6.2)$$

which in turn helps define the step size and characterize the optimality tolerance. Furthermore, we define

$$d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) = M_1(\mathbf{u}^k - \tilde{\mathbf{u}}^k), \quad (6.3)$$

and probe $Q^{-1}d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ as a *descent direction* of the distance function $\|\mathbf{u} - \mathbf{u}^*\|_Q^2$ at $\mathbf{u} = \mathbf{u}^k$. We are now ready to formally present the linearized version A of PPSA (LPSM-A).

Algorithm 6.1 Linearized version A of proximal partially parallel splitting algorithm.

Step 0. Initialization: Given $\epsilon > 0$ and the initial iterate $(\mathbf{x}^0, \lambda^0) \in \mathcal{X} \times \mathbb{R}^l$. Choose $\beta > 0$ and $Q \succ 0$ ($Q \in \mathcal{S}_{++}^{n+l}$). Set $k := 0$.

Step 1. Proximal Partially Parallel Splitting Step: Determine

$$\left\{ \begin{aligned} \tilde{x}_1^k &:= \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \theta_1(x_1) + \left\langle x_1 - x_1^k, \beta A_1^\top \left(\sum_{j=1}^m A_j x_j^k - b \right) - A_1^\top \lambda^k \right\rangle \right. \\ &\quad \left. + \frac{1}{2} \|x_1 - x_1^k\|_{S_1}^2 \right\}, \end{aligned} \right. \quad (6.4)$$

$$\left\{ \begin{aligned} \tilde{x}_i^k &:= \arg \min_{x_i \in \mathcal{X}_i} \left\{ \theta_i(x_i) + \left\langle x_i - x_i^k, \beta A_i^\top \left(A_1 \tilde{x}_1^k + \sum_{j=2}^m A_j x_j^k - b \right) - A_i^\top \lambda^k \right\rangle \right. \\ &\quad \left. + \frac{1}{2} \|x_i - x_i^k\|_{S_i}^2 \right\}, \forall i \geq 2. \end{aligned} \right. \quad (6.5)$$

where S_i ($i = 1, \dots, m$) are appropriately defined symmetric positive definite matrices.

Step 2. Lagrangian Multiplier Update Step: Compute

$$\tilde{\lambda}^k = \lambda^k - \beta \left(\sum_{i=1}^m A_i \tilde{x}_i^k - b \right). \quad (6.6)$$

Step 3. Termination Step: If

$$\max \left\{ \max_{1 \leq i \leq m} \{\|x_i^k - \tilde{x}_i^k\|\}, \|\lambda^k - \tilde{\lambda}^k\| \right\} < \epsilon,$$

then terminate the algorithm with $\tilde{\mathbf{u}}^k = (\tilde{\mathbf{x}}^k, \tilde{\lambda}^k)$ being an acceptable approximate solution of (1.1). Otherwise, go to **Step 4**.

Step 4. Corrector Step: Generate the new iterate by

$$\mathbf{u}^{k+1} = \mathbf{u}^k - \alpha_k Q^{-1} d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k), \quad (6.7)$$

where

$$\alpha_k = \gamma \alpha_k^*, \quad \alpha_k^A = \frac{\varphi_A(\mathbf{u}^k, \tilde{\mathbf{u}}^k)}{\|Q^{-1} d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k)\|_Q^2}, \quad \gamma \in (0, 2). \quad (6.8)$$

Set $k := k + 1$ and go to **Step 1**.

Remark 6.1 *As compared to the linearized method A seen in [132], we follow an identical primal variables updating scheme by setting $x = x_1$ and $y = (x_2^\top, \dots, x_m^\top)^\top$. However, the Lagrangian multiplier $\tilde{\lambda}^k$ in [132] is computed as*

$$\tilde{\lambda}^k = \lambda^k - \beta \left(A_1 \tilde{x}_1^k + \sum_{i=2}^m A_i x_i^k - b \right),$$

which is inherently different from the update scheme found in this work.

Remark 6.2 *If we set $S_i = s_i I_i$ in Algorithm 6.1 and $S_i = s_i I_i - \beta A_i^\top A_i$ in Algorithm 5.1, then, $d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) = d_1(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ and $\varphi_A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) = \varphi(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$, and Algorithm 6.1 reduces to Algorithm 5.1.*

In the following discussion, we begin by showcasing some lemmas, which set the stage for the convergence analysis of Algorithm 6.1.

Lemma 6.1 *Let $\tilde{\mathbf{u}}^k$ be generated by Algorithm 6.1 from a given vector $\mathbf{u}^k$. Then,*

$$\begin{aligned} & \tilde{\mathbf{u}}^k \in \mathcal{W}, \text{ and} \\ & \theta(\mathbf{x}) - \theta(\tilde{\mathbf{x}}^k) \\ & + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, F(\tilde{\mathbf{u}}^k) + d_2^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) - d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle \geq 0, \forall \mathbf{u} \in \mathcal{W}, \end{aligned} \quad (6.9)$$

where, $d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ is given by (6.3) and

$$d_2^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) = \beta (A_1, \dots, A_m, 0_{l \times 1})^\top \left(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right). \quad (6.10)$$

Proof: The first-order optimality conditions for the $\tilde{x}_1^k$ -related subproblem in Algorithm 6.1 yield:

$$\begin{aligned} & \theta_1(x_1) - \theta_1(\tilde{x}_1^k) + \langle x_1 - \tilde{x}_1^k, -A_1^\top \lambda^k \rangle \\ & + \left\langle x_1 - \tilde{x}_1^k, \beta A_1^\top \left(\sum_{j=1}^m A_j x_j^k - b \right) + S_1(\tilde{x}_1^k - x_1^k) \right\rangle \geq 0, \forall x_1 \in \mathcal{X}_1. \end{aligned}$$

Then, it follows from $\tilde{\lambda}^k = \lambda^k - \beta (\sum_{i=1}^m A_i \tilde{x}_i^k - b)$ that

$$\begin{aligned} & \theta_1(x_1) - \theta_1(\tilde{x}_1^k) + \langle x_1 - \tilde{x}_1^k, -A_1^\top \tilde{\lambda}^k \rangle \\ & + \left\langle x_1 - \tilde{x}_1^k, \sum_{j=1}^m \beta A_1^\top A_j(x_j^k - \tilde{x}_j^k) + S_1(\tilde{x}_1^k - x_1^k) \right\rangle \geq 0, \forall x_1 \in \mathcal{X}_1. \end{aligned} \quad (6.11)$$

Similarly, for $i = 2, \dots, m$, the first-order optimality conditions of (6.5) leads to

$$\begin{aligned} & \theta_i(x_i) - \theta_i(\tilde{x}_i^k) + \langle x_i - \tilde{x}_i^k, -A_i^\top \tilde{\lambda}^k \rangle \\ & + \left\langle x_i - \tilde{x}_i^k, \sum_{j=2}^m \beta A_i^\top A_j(x_j^k - \tilde{x}_j^k) + S_i(\tilde{x}_i^k - x_i^k) \right\rangle \geq 0, \forall x_i \in \mathcal{X}_i. \end{aligned}$$

Summing the above inequalities over $i = 2, \dots, m$ and combining with inequality (6.11), we get:

$$\begin{aligned} & \theta(\mathbf{x}) - \theta(\tilde{\mathbf{x}}^k) + \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^\top \begin{pmatrix} -A_1^\top \tilde{\lambda}^k \\ -A_2^\top \tilde{\lambda}^k \\ \vdots \\ -A_m^\top \tilde{\lambda}^k \end{pmatrix} \\ & + \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^\top \begin{pmatrix} \sum_{j=1}^m \beta A_1^\top A_j(x_j^k - \tilde{x}_j^k) + S_1(\tilde{x}_1^k - x_1^k) \\ \sum_{j=2}^m \beta A_2^\top A_j(x_j^k - \tilde{x}_j^k) + S_2(\tilde{x}_2^k - x_2^k) \\ \vdots \\ \sum_{j=2}^m \beta A_m^\top A_j(x_j^k - \tilde{x}_j^k) + S_m(\tilde{x}_m^k - x_m^k) \end{pmatrix} \geq 0, \forall \mathbf{x} \in \mathcal{X}. \end{aligned} \quad (6.12)$$

Now, Equation (6.6) can be equivalently expressed as

$$\left\langle \lambda - \tilde{\lambda}^k, \left(\sum_{j=1}^m A_j \tilde{x}_j^k - b \right) + \beta^{-1}(\tilde{\lambda}^k - \lambda^k) \right\rangle \geq 0, \quad \forall \lambda \in \mathbb{R}^l, \quad (6.13)$$

and summing (6.12) and (6.13) together yields

$$\begin{aligned}
 & \theta(\mathbf{x}) - \theta(\tilde{\mathbf{x}}^k) \\
 & + \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^\top \left\{ \begin{pmatrix} -A_1^\top \tilde{\lambda}^k \\ -A_2^\top \tilde{\lambda}^k \\ \vdots \\ -A_m^\top \tilde{\lambda}^k \\ \sum_{j=1}^m A_j \tilde{x}_j^k - b \end{pmatrix} + \beta \begin{pmatrix} A_1^\top \\ A_2^\top \\ \vdots \\ A_m^\top \\ 0 \end{pmatrix} \left(\sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k) \right) \right\} \\
 & \geq \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^\top \begin{pmatrix} (S_1 - \beta A_1^\top A_1)(x_1^k - \tilde{x}_1^k) \\ S_2(x_2^k - \tilde{x}_2^k) \\ \vdots \\ S_m(x_m^k - \tilde{x}_m^k) \\ \beta^{-1}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}, \forall \mathbf{u} \in \mathcal{U}. \tag{6.14}
 \end{aligned}$$

Simplifying, utilizing the definitions of $d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ and $d_2^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$, given by (6.3) and (6.10), the inequality (6.14) can be equivalently rewritten as:

$$\theta(\mathbf{x}) - \theta(\tilde{\mathbf{x}}^k) + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, F(\tilde{\mathbf{u}}^k) + d_2^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) - d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle \geq 0, \forall \mathbf{u} \in \mathcal{U}.$$

Thus, the statement of this lemma is obtained. $\square$

Lemma 6.2 *Let $\mathbf{u}^* = (x_1^*, \dots, x_m^*, \lambda^*)$ be a solution of $\text{VI}(\mathcal{U}, F, \theta)$ and the sequence $\{\tilde{\mathbf{u}}^k\}$ be generated by Algorithm 6.1 from a given vector $\mathbf{u}^k$. Then,*

$$\langle \mathbf{u}^k - \mathbf{u}^*, d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle \geq \varphi_A(\mathbf{u}^k, \tilde{\mathbf{u}}^k), \tag{6.15}$$

where $\varphi_A(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ is defined in (6.2).

Based on the theory for contraction methods developed in [63], we first prove in the following lemma that that

$$\varphi_A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \geq c \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|^2,$$

for some $c > 0$.

Lemma 6.3 *Assume that $S_1 \succ \beta A_1^\top A_1$ and $S_j \succ \frac{m\beta}{4} A_j^\top A_j$ ($j \geq 2$). Let $\tilde{\mathbf{u}}^k$ be generated by Algorithm 6.1 from a given vector $\mathbf{u}^k$. Then,*

$$\varphi_A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \geq \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_{M_A}^2, \tag{6.16}$$

where $\varphi_A(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ is defined in (6.2) and

$$M_A = \text{diag} \left(S_1 - \beta A_1^\top A_1, S_2 - \frac{m\beta}{4} A_2^\top A_2, \dots, S_m - \frac{m\beta}{4} A_m^\top A_m, \frac{1}{m\beta} I_l \right).$$

Proof: Making use of the definition of $d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ leads to

$$\begin{aligned} & \langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle \\ &= \|x_1^k - \tilde{x}_1^k\|_{S_1 - \beta A_1^\top A_1}^2 + \sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j}^2 + \beta^{-1} \|\lambda^k - \tilde{\lambda}^k\|^2. \end{aligned} \quad (6.17)$$

Furthermore,

$$\left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle \geq - \sum_{j=2}^m \left[\frac{m\beta}{4} \|A_j(x_j^k - \tilde{x}_j^k)\|^2 + \frac{1}{m\beta} \|\lambda^k - \tilde{\lambda}^k\|^2 \right], \quad (6.18)$$

is deduced from the inequality $a^\top b \geq -(\frac{\mu}{4}\|a\|^2 + \frac{1}{\mu}\|b\|^2)$, for any $\mu > 0$.

Combining the inequalities (6.17) and (6.18) leads to

$$\begin{aligned} & \varphi_A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \\ &= \langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle + \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle \\ &= \|x_1^k - \tilde{x}_1^k\|_{S_1 - \beta A_1^\top A_1}^2 + \sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j}^2 + \beta^{-1} \|\lambda^k - \tilde{\lambda}^k\|^2 \\ &\quad + \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle \\ &\geq \|x_1^k - \tilde{x}_1^k\|_{S_1 - \beta A_1^\top A_1}^2 + \sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j}^2 + \beta^{-1} \|\lambda^k - \tilde{\lambda}^k\|^2 \\ &\quad - \sum_{j=2}^m \frac{m\beta}{4} \|A_j(x_j^k - \tilde{x}_j^k)\|^2 - \sum_{j=2}^m \frac{1}{m\beta} \|\lambda^k - \tilde{\lambda}^k\|^2 \\ &= \|x_1^k - \tilde{x}_1^k\|_{S_1 - \beta A_1^\top A_1}^2 + \sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j - \frac{m\beta}{4} A_j^\top A_j}^2 + \frac{1}{m\beta} \|\lambda^k - \tilde{\lambda}^k\|^2. \end{aligned} \quad (6.19)$$

Under the assumption that $S_1 \succ \beta A_1^\top A_1$ and $S_j \succ \frac{m\beta}{4} A_j^\top A_j$ ($j \geq 2$), we have $\varphi_A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \geq 0$. Hence, the assertion (6.16) is derived from the definition of M_A . $\square$

If $\mathbf{u}^k \neq \tilde{\mathbf{u}}^k$, then Lemmas 6.2 and 6.3 indicate that $-Q^{-1}d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ is a *descent direction* of the distance function $\|\mathbf{u} - \mathbf{u}^*\|_Q^2$ at $\mathbf{u} = \mathbf{u}^k$. Therefore, the new iterate will

be determined along this direction in concert with an appropriate step size α_k . Then, we can conclude that the step size α_k is bounded away from zero, and validate this claim by the following theorem. Consequently, the sequence $\{\mathbf{u}^k\}$ generated by Algorithm 6.1 is proved to be Fejér monotone with respect to $\mathcal{U}^*$

Theorem 6.1 *Assume that $S_1 \succ \beta A_1^\top A_1$ and $S_j \succ \frac{m\beta}{4} A_j^\top A_j$ ($j \geq 2$), the sequence $\{\mathbf{u}^k\}$ generated by Algorithm 6.1 satisfies*

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_Q^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|_Q^2 - \gamma(2 - \gamma)c_A \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_{M_A}^2, \quad \forall \mathbf{u}^* \in \mathcal{U}^*, \quad (6.20)$$

where c_A is the lower bound of the step-length α_k^A .

Proof: In accordance with the definition of α_k^A , we obtain

$$\begin{aligned} \alpha_k^A &= \frac{\varphi_A(\mathbf{u}^k, \tilde{\mathbf{u}}^k)}{\|Q^{-1}d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k)\|_Q^2} \\ &\geq \frac{\delta_{\min}(M_A)\|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|^2}{\delta_{\max}(M_1^\top Q^{-1}M_1)\|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|^2} \\ &= c_A, \left(c_A = \frac{\delta_{\min}(M_A)}{\delta_{\max}(M_1^\top Q^{-1}M_1)} \right). \end{aligned} \quad (6.21)$$

where $\delta_{\min}(\cdot)$ and $\delta_{\max}(\cdot)$ denote the minimum and maximum eigenvalues of a matrix $(\cdot)$, respectively. Then, using (6.7) leads to

$$\begin{aligned} &\|\mathbf{u}^k - \mathbf{u}^*\|_Q^2 - \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_Q^2 \\ &= 2\alpha_k \langle \mathbf{u}^k - \mathbf{u}^*, d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle - \alpha_k^2 \|Q^{-1}d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k)\|_Q^2 \\ &\geq 2\gamma\alpha_k^A \varphi_A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) - (\gamma\alpha_k^A)^2 \|Q^{-1}d_1^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k)\|_Q^2 \\ &= \gamma(2 - \gamma)\alpha_k^A \varphi_A(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \\ &\geq \gamma(2 - \gamma)c_A \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_{M_A}^2. \end{aligned} \quad (6.22)$$

where the first inequality follows from (6.15). Thus, the statement of this theorem is proved. $\square$

Theorem 6.2 *Assume that $S_1 \succ \beta A_1^\top A_1$ and $S_j \succ \frac{m\beta}{4} A_j^\top A_j$ ($j \geq 2$), the sequence $\{\mathbf{u}^k\}$ generated by Algorithm 6.1 globally converges to a solution point of $\text{VI}(\mathcal{U}, F, \theta)$.*

Proof: The proof follows along the same lines as that of Theorem 5.2. $\square$

6.2.2 Linearized Version B of PPSA (LPSM-B)

In this subsection, we present the linearized version B of the proximal partially parallel splitting algorithm, where instead of linearizing the quadratic terms in the augmented Lagrangian function (as seen in the previous version), we resort to directly linearizing the functions $\theta_i(x_i)$. To begin with, define matrix M_2 as follows:

$$M_2 = \text{diag}(S_1, S_2 + \beta A_2^\top A_2, \dots, S_m + \beta A_m^\top A_m, \beta^{-1} I_l) \quad (6.23)$$

where S_i ($i = 1, \dots, m$) are positive definite matrices. Hence, M_2 is also positive definite. Furthermore, we set $\varphi_B(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ as:

$$\varphi_B(\mathbf{u}^k, \tilde{\mathbf{u}}^k) = \langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, d_1^B(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle + \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle, \quad (6.24)$$

where

$$d_1^B(\mathbf{u}^k, \tilde{\mathbf{u}}^k) = M_2(\mathbf{u}^k - \tilde{\mathbf{u}}^k) - \Phi(\mathbf{x}^k, \tilde{\mathbf{x}}^k) \quad (6.25)$$

and

$$\Psi(\mathbf{x}^k, \tilde{\mathbf{x}}^k) = \begin{pmatrix} f_1(x_1^k) - f_1(\tilde{x}_1^k) \\ \vdots \\ f_m(x_m^k) - f_m(\tilde{x}_m^k) \\ 0_{l \times 1} \end{pmatrix}. \quad (6.26)$$

For the remainder of this analysis, we assume that each $\theta_i(x_i)$ is differentiable and its gradient $f_i(x_i) \equiv \nabla \theta_i(x_i^k)$ is Lipschitzian. Then, $\theta_i(x_i)$ can be replaced by $\langle x_i - x_i^k, \nabla \theta_i(x_i^k) \rangle$, which leads to the linearized version B of PPSA (LPSM-B).

Algorithm 6.2 Linearized version B of proximal partially parallel splitting algorithm.

Step 0. Initialization: Given $\epsilon > 0$ and the initial iterate $(\mathbf{x}^0, \lambda^0) \in \mathcal{X} \times \mathbb{R}^l$. Choose $\beta > 0$ and $Q \succ 0$ ($Q \in \mathcal{S}_{++}^{n+l}$). Set $k := 0$.

Step 1. Proximal Partially Parallel Splitting Step: Determine

$$\left\{ \begin{aligned} \tilde{x}_1^k &:= \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \langle x_1 - x_1^k, \nabla \theta_1(x_1^k) \rangle + \frac{\beta}{2} \left\| A_1 x_1 + \sum_{i=2}^m A_i x_i^k - b - \beta^{-1} \lambda^k \right\|^2 \right. \\ &\quad \left. + \frac{1}{2} \|x_1 - x_1^k\|_{S_1}^2 \right\}, \end{aligned} \right. \quad (6.27)$$

$$\left\{ \begin{aligned} \tilde{x}_i^k &:= \arg \min_{x_i \in \mathcal{X}_i} \left\{ \langle x_i - x_i^k, \nabla \theta_i(x_i^k) \rangle + \frac{\beta}{2} \left\| A_i x_i + A_1 \tilde{x}_1^k + \sum_{j=2, j \neq i}^m A_j x_j^k - b - \beta^{-1} \lambda^k \right\|^2 \right. \\ &\quad \left. + \frac{1}{2} \|x_i - x_i^k\|_{S_i}^2 \right\}, \end{aligned} \right. \quad (6.28)$$

where S_i ($i = 1, \dots, m$) are appropriately defined symmetric positive definite matrices.

Step 2. Lagrangian Multiplier Update Step: Compute

$$\tilde{\lambda}^k = \lambda^k - \beta \left(\sum_{i=1}^m A_i \tilde{x}_i^k - b \right). \quad (6.29)$$

Step 3. Termination Step: If

$$\max \left\{ \max_{1 \leq i \leq m} \{\|x_i^k - \tilde{x}_i^k\|\}, \|\lambda^k - \tilde{\lambda}^k\| \right\} < \epsilon,$$

then terminate the algorithm with $\tilde{\mathbf{u}}^k = (\tilde{\mathbf{x}}^k, \tilde{\lambda}^k)$ being an acceptable approximate solution of (1.1). Otherwise, go to **Step 4**.

Step 4. Corrector Step: Generate the new iterate by

$$\mathbf{u}^{k+1} = \mathbf{u}^k - \alpha_k Q^{-1} d_1^B(\mathbf{u}^k, \tilde{\mathbf{u}}^k), \quad (6.30)$$

where

$$\alpha_k = \gamma \alpha_k^B, \quad \alpha_k^B = \frac{\varphi_B(\mathbf{u}^k, \tilde{\mathbf{u}}^k)}{\|Q^{-1} d_1^B(\mathbf{u}^k, \tilde{\mathbf{u}}^k)\|_Q^2}, \quad \gamma \in (0, 2). \quad (6.31)$$

Set $k := k + 1$ and go to **Step 1**.

Remark 6.3 If we set $S_i = s_i I_i - \beta A_i^\top A_i$, the $\tilde{x}_i$ -related subproblems can be equivalently expressed as the following projection steps, given by:

$$\begin{cases} \tilde{x}_1^k := P_{\mathcal{X}_1} \left[x_1^k - \frac{1}{s_1} \left(f_1(x_1^k) + \beta A_1^\top \left(\sum_{j=1}^m A_j x_j^k - b \right) - A_1^\top \lambda^k \right) \right], \\ \tilde{x}_i^k := P_{\mathcal{X}_i} \left[x_i^k - \frac{1}{s_i} \left(f_i(x_i^k) + \beta A_i^\top \left(A_1 \tilde{x}_1^k + \sum_{j=2}^m A_j x_j^k - b \right) - A_i^\top \lambda^k \right) \right]. \end{cases} \quad (6.32)$$

Similar to the analysis seen in Section 6.2.1, some useful lemmas that set the tone for proving the global convergence of Algorithm 6.2 are presented next.

Lemma 6.4 Let $\tilde{\mathbf{u}}^k$ be generated by Algorithm 6.2 from a given vector $\mathbf{u}^k$. Then,

$$\begin{aligned} & \tilde{\mathbf{u}}^k \in \mathcal{U}, \text{ and} \\ & \theta(\mathbf{x}) - \theta(\tilde{\mathbf{x}}^k) \\ & + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, F(\tilde{\mathbf{u}}^k) + d_2^B(\mathbf{u}^k, \tilde{\mathbf{u}}^k) - d_1^B(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle \geq 0, \forall \mathbf{u} \in \mathcal{U}, \end{aligned} \quad (6.33)$$

where $d_1^B(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ is as defined in (6.25) and

$$d_2^B(\mathbf{u}^k, \tilde{\mathbf{u}}^k) = d_2^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k). \quad (6.34)$$

Proof: The first-order optimality conditions for the $\tilde{x}_1^k$ -related subproblem in Algorithm 6.2 are given by

$$\begin{aligned} & \langle x_1 - \tilde{x}_1^k, f_1(x_1^k) - A_1^\top \lambda^k \rangle \\ & + \left\langle x_1 - \tilde{x}_1^k, \beta A_1^\top \left(A_1 \tilde{x}_1^k + \sum_{j=2}^m A_j x_j^k - b \right) + S_1(\tilde{x}_1^k - x_1^k) \right\rangle \geq 0, \forall x_1 \in \mathcal{X}_1. \end{aligned}$$

Then, it follows from $\tilde{\lambda}^k = \lambda^k - \beta (\sum_{i=1}^m A_i \tilde{x}_i^k - b)$ that

$$\left\langle x_1 - \tilde{x}_1^k, f_1(x_1^k) - A_1^\top \tilde{\lambda}^k + \sum_{j=2}^m \beta A_1^\top A_j (x_j^k - \tilde{x}_j^k) + S_1(\tilde{x}_1^k - x_1^k) \right\rangle \geq 0, \forall x_1 \in \mathcal{X}_1.$$

Similarly, for $i = 2, \dots, m$, the first-order optimality conditions for (6.28) yields:

$$\left\langle x_i - \tilde{x}_i^k, f_i(x_i^k) - A_i^\top \tilde{\lambda}^k + \sum_{j=2, j \neq i}^m \beta A_i^\top A_j (x_j^k - \tilde{x}_j^k) + S_i(\tilde{x}_i^k - x_i^k) \right\rangle \geq 0, \forall x_i \in \mathcal{X}_i.$$

Then, summing the above inequalities corresponding to $i = 1, \dots, m$ leads to

$$\begin{aligned}
 & \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^\top \begin{pmatrix} f_1(x_1^k) - A_1^\top \tilde{\lambda}^k \\ f_2(x_2^k) - A_2^\top \tilde{\lambda}^k \\ \vdots \\ f_m(x_m^k) - A_m^\top \tilde{\lambda}^k \end{pmatrix} \\
 & + \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^\top \begin{pmatrix} \sum_{j=2}^m \beta A_1^\top A_j (x_j^k - \tilde{x}_j^k) + S_1(\tilde{x}_1^k - x_1^k) \\ \sum_{j=2, j \neq i}^m \beta A_2^\top A_j (x_j^k - \tilde{x}_j^k) + S_2(\tilde{x}_2^k - x_2^k) \\ \vdots \\ \sum_{j=2, j \neq i}^m \beta A_m^\top A_j (x_j^k - \tilde{x}_j^k) + S_m(\tilde{x}_m^k - x_m^k) \end{pmatrix} \geq 0, \forall \mathbf{x} \in \mathcal{X}.
 \end{aligned} \tag{6.35}$$

Adding (6.13) and (6.35) together yields:

$$\begin{aligned}
 & \sum_{i=1}^m \langle x - \tilde{x}_i^k, f_i(\tilde{x}_i^k) \rangle \\
 & + \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^\top \left\{ \begin{pmatrix} -A_1^\top \tilde{\lambda}^k \\ -A_2^\top \tilde{\lambda}^k \\ \vdots \\ -A_m^\top \tilde{\lambda}^k \\ \sum_{j=1}^m A_j \tilde{x}_j^k - b \end{pmatrix} + \begin{pmatrix} A_1^\top \\ A_2^\top \\ \vdots \\ A_m^\top \\ 0 \end{pmatrix} \beta \left(\sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k) \right) \right\} \\
 & \geq \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^\top \begin{pmatrix} S_1(x_1^k - \tilde{x}_1^k) - (f_1(x_1^k) - f_1(\tilde{x}_1^k)) \\ (S_2 + \beta A_2^\top A_2)(x_2^k - \tilde{x}_2^k) - (f_2(x_2^k) - f_2(\tilde{x}_2^k)) \\ \vdots \\ (S_m + \beta A_m^\top A_m)(x_m^k - \tilde{x}_m^k) - (f_m(x_m^k) - f_m(\tilde{x}_m^k)) \\ \beta^{-1}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}, \forall \mathbf{u} \in \mathcal{U}.
 \end{aligned} \tag{6.36}$$

Therefore, making use of the convexity of $\theta_i(\cdot)$ leads to

$$\theta(x_i) - \theta(\tilde{x}_i^k) \geq \langle x - \tilde{x}_i^k, f(\tilde{x}_i^k) \rangle, \forall x_i \in \mathcal{X}_i.$$

Substituting the above inequality, and using the definitions of $d_1^B(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ and $d_2^B(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$, given by (6.25) and (6.34), the inequality (6.36) can be expressed as

$$\theta(\mathbf{x}) - \theta(\tilde{\mathbf{x}}^k) + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, F(\tilde{\mathbf{u}}^k) + d_2^B(\mathbf{u}^k, \tilde{\mathbf{u}}^k) - d_1^B(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle \geq 0, \forall \mathbf{u} \in \mathcal{U},$$

and the statement of this lemma is obtained. $\square$

Lemma 6.5 *Let $\mathbf{u}^* = (x_1^*, \dots, x_m^*, \lambda^*)$ be a solution of $\text{VI}(\mathcal{U}, F, \theta)$ and the sequence $\{\tilde{\mathbf{u}}^k\}$ be generated by Algorithm 6.2 from a given vector $\mathbf{u}^k$. Then,*

$$\langle \mathbf{u}^k - \mathbf{u}^*, d_1^B(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle \geq \varphi_B(\mathbf{u}^k, \tilde{\mathbf{u}}^k), \quad (6.37)$$

where $\varphi_B(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ is defined by (6.24).

Lemma 6.6 *Assume that $S_1 \succ l_1 I_1$ and $S_j \succ l_j I_j + \frac{(m-4)\beta}{4} A_j^\top A_j$ ($j \geq 2$). Let $\tilde{\mathbf{u}}^k$ be generated by Algorithm 6.2 from the given vector $\mathbf{u}^k$. Then,*

$$\varphi_B(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \geq \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_{M_B}^2, \quad (6.38)$$

where $\varphi_B(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ is defined in (6.24) and

$$M_B = \text{diag} \left\{ S_1 - l_1 I_1, S_2 - l_2 I_2 - \frac{(m-4)\beta}{4} A_2^\top A_2, \dots, S_m - l_m I_m - \frac{(m-4)\beta}{4} A_m^\top A_m, \frac{1}{m\beta} I_l \right\}.$$

Proof: Making use of the definition of $\varphi_B(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ leads to

$$\begin{aligned} & \varphi_B(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \\ &= \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_{M_2}^2 + \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle - \sum_{j=1}^m \langle x_j^k - \tilde{x}_j^k, f_j(x_j^k) - f_j(\tilde{x}_j^k) \rangle \\ &\geq \sum_{i=1}^m \|x_i^k - \tilde{x}_i^k\|_{S_i}^2 + \sum_{j=2}^m \beta \|A_j(x_j^k - \tilde{x}_j^k)\|^2 + \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle \\ &\quad - \sum_{i=1}^m l_i \|x_i^k - \tilde{x}_i^k\|^2 \\ &\geq \|x_1^k - \tilde{x}_1^k\|_{S_1 - l_1 I_1}^2 + \sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j - l_j I_j - \frac{(m-4)\beta}{4} A_j^\top A_j}^2 + \frac{1}{m\beta} \|\lambda^k - \tilde{\lambda}^k\|^2, \end{aligned} \quad (6.39)$$

where the first inequality follows from the property of Lipschitz continuity, and the second inequality is derived using (6.18). Assume that $S_1 \succ l_1 I_1$ and $S_j \succ l_j I_j + \frac{(m-4)\beta}{4} A_j^\top A_j$ ($j \geq 2$), we have $\varphi_B(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \geq 0$, and the assertion of this lemma immediately follows from the definition of M_B . $\square$

Theorem 6.3 Assume that $S_1 \succ l_1 I_1$ and $S_j \succ l_j I_j + \frac{(m-4)\beta}{4} A_j^\top A_j$ ($j \geq 2$), the sequence $\{\mathbf{u}^k\}$ generated by Algorithm 6.2 satisfies

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_Q^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|_Q^2 - \gamma(2 - \gamma)c_B \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_{M_B}^2, \quad \forall \mathbf{u}^* \in \mathcal{U}^*,$$

where c_B is the lower bound of the step-length α_k^B .

As the remainder of this convergence analysis is similar to that seen in Section 6.2.1, we omit the detailed proofs here, and proceed to presenting another linearized variant of the PPSA method.

6.2.3 Linearized Version C of PPSA (LPSM-C)

In the subsection, we introduce the last of our linearized versions of the proximal partially-parallel splitting algorithm by linearizing both the convex function $\theta_i(x_i)$ and the quadratic terms in the involved subproblems. In this case, we assume that each of the objective functions $\theta_i(x_i)$, $i = 1, \dots, m$, is differentiable and their gradient $f_i(x_i)$ possess the property of Lipschitz continuity. Furthermore, we define $\varphi_C(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ as:

$$\varphi_C(\mathbf{u}^k, \tilde{\mathbf{u}}^k) = \langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, d_1^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle + \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle, \quad (6.40)$$

where

$$d_1^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k) = M_1(\mathbf{u}^k - \tilde{\mathbf{u}}^k) - \Psi(\mathbf{x}^k, \tilde{\mathbf{x}}^k), \quad (6.41)$$

and M_1 and $\Psi(\mathbf{x}^k, \tilde{\mathbf{x}}^k)$ are as defined in (6.1) and (6.26), respectively.

Consequently, the linearized version C of PPSA (LPSM-C) can be described as follows.

Algorithm 6.3 Linearized version C of proximal partially parallel splitting algorithm.

Step 0. Initialization: Given $\epsilon > 0$ and the initial iterate $(\mathbf{x}^0, \lambda^0) \in \mathcal{X} \times \mathbb{R}^l$. Choose $\beta > 0$ and $Q \succ 0$ ($Q \in \mathcal{S}_{++}^{n+l}$). Set $k := 0$.

Step 1. Proximal Partially Parallel Splitting Step: Determine

$$\left\{ \begin{aligned} \tilde{x}_1^k &:= \arg \min_{x_1 \in \mathcal{X}_1} \left\{ \left\langle x_1 - x_1^k, \nabla \theta_1(x_1^k) + \beta A_1^\top \left(\sum_{j=1}^m A_j x_j^k - b \right) - A_1^\top \lambda^k \right\rangle \right. \\ &\quad \left. + \frac{1}{2} \|x_1 - x_1^k\|_{S_1}^2 \right\}, \end{aligned} \right. \quad (6.42)$$

$$\left\{ \begin{aligned} \tilde{x}_i^k &:= \arg \min_{x_i \in \mathcal{X}_i} \left\{ \left\langle x_i - x_i^k, \nabla \theta_i(x_i^k) + \beta A_i^\top \left(A_1 \tilde{x}_1^k + \sum_{j=2}^m A_j x_j^k - b \right) - A_i^\top \lambda^k \right\rangle \right. \\ &\quad \left. + \frac{1}{2} \|x_i - x_i^k\|_{S_i}^2 \right\}, \end{aligned} \right. \quad (6.43)$$

where $S_i (i = 1, \dots, m)$ are appropriately defined symmetric positive definite matrices.

Step 2. Lagrangian Multiplier Update Step: Compute

$$\tilde{\lambda}^k = \lambda^k - \beta \left(\sum_{i=1}^m A_i \tilde{x}_i^k - b \right). \quad (6.44)$$

Step 3. Termination Step: If

$$\max \left\{ \max_{1 \leq i \leq m} \{ \|x_i^k - \tilde{x}_i^k\| \}, \|\lambda^k - \tilde{\lambda}^k\| \right\} < \epsilon,$$

then terminate the algorithm with $\tilde{\mathbf{u}}^k = (\tilde{\mathbf{x}}^k, \tilde{\lambda}^k)$ being an acceptable approximate solution of (1.1). Otherwise, go to **Step 4**.

Step 4. Corrector Step: Generate the new iterate by

$$\mathbf{u}^{k+1} = \mathbf{u}^k - \alpha_k Q^{-1} d_1^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k), \quad (6.45)$$

where

$$\alpha_k = \gamma \alpha_k^C, \quad \alpha_k^C = \frac{\varphi_C(\mathbf{u}^k, \tilde{\mathbf{u}}^k)}{\|Q^{-1} d_1^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k)\|_Q^2}, \quad \gamma \in (0, 2). \quad (6.46)$$

Set $k := k + 1$ and go to **Step 1**.

Remark 6.4 If we set $S_i = s_i I_i$, the $\tilde{x}_i$ -related subproblems are equivalent to the projection steps given by (6.32). Then, $d_1^B(\mathbf{u}^k, \tilde{\mathbf{u}}^k) = d_1^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ and $\varphi_B(\mathbf{u}^k, \tilde{\mathbf{u}}^k) = \varphi_C(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$, and Algorithm 6.3 reduces to Algorithm 6.2 by setting $S_i = s_i I_i - \beta A_i^\top A_i$.

Once again, we start by inferring some useful lemmas for probing the global convergence of Algorithm 6.3.

Lemma 6.7 *Let $\tilde{\mathbf{u}}^k$ be generated by Algorithm 6.3 from a given vector $\mathbf{u}^k$. Then,*

$$\begin{aligned} \tilde{\mathbf{u}}^k &\in \mathcal{U}, \text{ and} \\ \theta(\mathbf{x}) - \theta(\tilde{\mathbf{x}}^k) \\ &+ \langle \mathbf{u} - \tilde{\mathbf{u}}^k, F(\tilde{\mathbf{u}}^k) + d_2^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k) - d_1^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle \geq 0, \forall \mathbf{u} \in \mathcal{U}, \end{aligned} \quad (6.47)$$

where, $d_1^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ is defined in (6.41) and

$$d_2^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k) = d_2^A(\mathbf{u}^k, \tilde{\mathbf{u}}^k). \quad (6.48)$$

Proof: The first-order optimality conditions for the $\tilde{x}_1^k$ -related subproblem in Algorithm 6.3 are given by

$$\left\langle x_1 - \tilde{x}_1^k, f_1(x_1^k) - A_1^\top \lambda^k + \beta A_1^\top \left(\sum_{j=1}^m A_j x_j^k - b \right) + S_1(\tilde{x}_1^k - x_1^k) \right\rangle \geq 0, \forall x_1 \in \mathcal{X}_1.$$

Then, it follows from $\tilde{\lambda}^k = \lambda^k - \beta (\sum_{i=1}^m A_i \tilde{x}_i^k - b)$ that

$$\left\langle x_1 - \tilde{x}_1^k, f_1(x_1^k) - A_1^\top \tilde{\lambda}^k + \sum_{j=1}^m \beta A_1^\top A_j (x_j^k - \tilde{x}_j^k) + S_1(\tilde{x}_1^k - x_1^k) \right\rangle \geq 0, \forall x_1 \in \mathcal{X}_1.$$

Similarly, for $i = 2, \dots, m$, the first-order optimality conditions of $\tilde{x}_i^k$ yields:

$$\left\langle x_i - \tilde{x}_i^k, f_i(x_i^k) - A_i^\top \tilde{\lambda}^k + \sum_{j=2}^m \beta A_i^\top A_j (x_j^k - \tilde{x}_j^k) + S_i(\tilde{x}_i^k - x_i^k) \right\rangle \geq 0, \forall x_i \in \mathcal{X}_i.$$

Then, summing the above inequalities corresponding to $i = 1, \dots, m$ leads to

$$\begin{aligned} &\begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^\top \begin{pmatrix} f_1(x_1^k) - A_1^\top \tilde{\lambda}^k \\ f_2(x_2^k) - A_2^\top \tilde{\lambda}^k \\ \vdots \\ f_m(x_m^k) - A_m^\top \tilde{\lambda}^k \end{pmatrix} \\ &+ \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^\top \begin{pmatrix} \sum_{j=1}^m \beta A_1^\top A_j (x_j^k - \tilde{x}_j^k) + S_1(\tilde{x}_1^k - x_1^k) \\ \sum_{j=2}^m \beta A_2^\top A_j (x_j^k - \tilde{x}_j^k) + S_2(\tilde{x}_2^k - x_2^k) \\ \vdots \\ \sum_{j=2}^m \beta A_m^\top A_j (x_j^k - \tilde{x}_j^k) + S_m(\tilde{x}_m^k - x_m^k) \end{pmatrix} \geq 0, \forall \mathbf{x} \in \mathcal{X}. \end{aligned} \quad (6.49)$$

Combining (6.49) and (6.13) together yields

$$\begin{aligned}
 & \sum_{i=1}^m \langle x - \tilde{x}_i^k, f_i(\tilde{x}_i^k) \rangle \\
 & + \left(\begin{array}{c} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \\ \lambda - \tilde{\lambda}^k \end{array} \right)^\top \left\{ \left(\begin{array}{c} -A_1^\top \tilde{\lambda}^k \\ -A_2^\top \tilde{\lambda}^k \\ \vdots \\ -A_m^\top \tilde{\lambda}^k \\ \sum_{j=1}^m A_j \tilde{x}_j^k - b \end{array} \right) + \beta \left(\begin{array}{c} A_1^\top \\ A_2^\top \\ \vdots \\ A_m^\top \\ 0 \end{array} \right) \left(\sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k) \right) \right\} \\
 & \geq \left(\begin{array}{c} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \\ \lambda - \tilde{\lambda}^k \end{array} \right)^\top \left(\begin{array}{c} (S_1 - \beta A_1^\top A_1)(x_1^k - \tilde{x}_1^k) - (f_1(x_1^k) - f_1(\tilde{x}_1^k)) \\ S_2(x_2^k - \tilde{x}_2^k) - (f_2(x_2^k) - f_2(\tilde{x}_2^k)) \\ \vdots \\ S_m(x_m^k - \tilde{x}_m^k) - (f_m(x_m^k) - f_m(\tilde{x}_m^k)) \\ \beta^{-1}(\lambda^k - \tilde{\lambda}^k) \end{array} \right), \forall \mathbf{u} \in \mathcal{U}. \quad (6.50)
 \end{aligned}$$

Furthermore, by the convexity of $\theta_i(\cdot)$, we have

$$\theta(x_i) - \theta(\tilde{x}_i^k) \geq \langle x - \tilde{x}_i^k, f(\tilde{x}_i^k) \rangle, \forall x_i \in \mathcal{X}_i.$$

Simplifying, substituting the above inequality into (6.50) and using the definitions of $d_1^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ and $d_2^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$, given by (6.41) and (6.47), respectively, the inequality (6.50) is expressed as

$$\theta(\mathbf{x}) - \theta(\tilde{\mathbf{x}}^k) + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, F(\tilde{\mathbf{u}}^k) + d_2^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k) - d_1^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle \geq 0, \forall \mathbf{u} \in \mathcal{U},$$

and the statement of this lemma is obtained. $\square$

Lemma 6.8 *Let $\mathbf{u}^* = (x_1^*, \dots, x_m^*, \lambda^*)$ be a solution of $\text{VI}(\mathcal{U}, F, \theta)$ and the sequence $\{\tilde{\mathbf{u}}^k\}$ be generated by Algorithm 6.3 from a given vector $\mathbf{u}^k$. Then,*

$$\langle \mathbf{u}^k - \mathbf{u}^*, d_1^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle \geq \varphi_C(\mathbf{u}^k, \tilde{\mathbf{u}}^k), \quad (6.51)$$

where $\varphi_C(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ is defined by (6.40).

Lemma 6.9 *Assume that $S_1 \succ l_1 I_1 + \beta A_1^\top A_1$ and $S_j \succ l_j I_j + \frac{m\beta}{4} A_j^\top A_j$ ($j \geq 2$). Let $\tilde{\mathbf{u}}^k$ be generated by Algorithm 6.3 from the given vector $\mathbf{u}^k$. Then, we have*

$$\varphi_C(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \geq \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_{M_C}^2, \quad (6.52)$$

where $\varphi_C(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ is defined in (6.40) and

$$M_C = \text{diag} \left(S_1 - l_1 I_1 - \beta A_1^\top A_1, S_2 - l_2 I_2 - \frac{m\beta}{4} A_2^\top A_2, \dots, S_m - l_m I_m - \frac{m\beta}{4} A_m^\top A_m, \frac{1}{m\beta} I_l \right).$$

Proof: Making use of the definition of $\varphi_C(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$ leads to

$$\begin{aligned} \varphi_C(\mathbf{u}^k, \tilde{\mathbf{u}}^k) &= \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_{M_1}^2 + \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle - \sum_{j=1}^m \langle x_j^k - \tilde{x}_j^k, f_j(x_j^k) - f_j(\tilde{x}_j^k) \rangle \\ &\geq \|x_1^k - \tilde{x}_1^k\|_{S_1 - \beta A_1^\top A_1}^2 + \sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j - \frac{m\beta}{4} A_j^\top A_j}^2 + \frac{1}{m\beta} \|\lambda^k - \tilde{\lambda}^k\|^2 \\ &\quad - \sum_{j=1}^m \langle x_j^k - \tilde{x}_j^k, f_j(x_j^k) - f_j(\tilde{x}_j^k) \rangle \\ &\geq \|x_1^k - \tilde{x}_1^k\|_{S_1 - l_1 I_1 - \beta A_1^\top A_1}^2 + \sum_{j=2}^m \|x_j^k - \tilde{x}_j^k\|_{S_j - l_j I_j - \frac{m\beta}{4} A_j^\top A_j}^2 + \frac{1}{m\beta} \|\lambda^k - \tilde{\lambda}^k\|^2, \end{aligned} \quad (6.53)$$

where the last inequality follows from the property of Lipschitz continuity. Under the assumptions that $S_1 \succ l_1 I_1 + \beta A_1^\top A_1$ and $S_j \succ l_j I_j + \frac{m\beta}{4} A_j^\top A_j$ ($j \geq 2$), we have $\varphi_C(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \geq 0$, and the inequality (6.52) immediately follows from the definition of M_C . $\square$

Theorem 6.4 Assume that $S_1 \succ l_1 I_1 + \beta A_1^\top A_1$ and $S_j \succ l_j I_j + \frac{m\beta}{4} A_j^\top A_j$ ($j \geq 2$), the sequence $\{\mathbf{u}^k\}$ generated by Algorithm 6.3 satisfies

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_Q^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|_Q^2 - \gamma(2 - \gamma)c_C \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_{M_C}^2, \quad \forall \mathbf{u}^* \in \mathcal{U}^*,$$

where c_C is the lower bound of the step-length α_k^C .

As the rest of this analysis is similar to that of Section 6.2.1, we omit the detailed proofs here, and proceed to deriving the convergence rates of the proposed linearized versions under ergodic conditions.

6.3 Convergence Rate under Ergodic Sense

In this section, we prove the worst-case $O(1/k)$ convergence rate for all of the proposed linearized methods. Note that, as we can utilize the same strategy given in [67] to arrive at the required convergence rate, we forego providing the details of the proofs for Algorithm 6.1 and Algorithm 6.2, and directly proceed to analyzing the worst-case $O(1/k)$ convergence rate for Algorithm 6.3, under the assumption that $S_1 \succ \beta A_1^\top A_1 + l_1 I_1$ and $S_j \succ l_j I_j + \frac{m\beta}{4} A_j^\top A_j$ ($j \geq 2$).

To begin with, define a sequence $\bar{\mathbf{u}} = (\bar{x}_1^k, \bar{x}_2^k, \dots, \bar{x}_m^k, \bar{\lambda}^k)$ as

$$\begin{pmatrix} \bar{x}_1^k \\ \bar{x}_2^k \\ \vdots \\ \bar{x}_m^k \end{pmatrix} = \begin{pmatrix} \tilde{x}_1^k \\ \tilde{x}_2^k \\ \vdots \\ \tilde{x}_m^k \end{pmatrix} \quad \text{and} \quad \bar{\lambda}^k = \tilde{\lambda}^k + \sum_{j=2}^m \beta A_j (\tilde{x}_j^k - x_j^k), \quad (6.54)$$

Additionally, define the following two matrices:

$$L_1 = \begin{pmatrix} I_{n_1} & 0 & \dots & 0 & 0 \\ 0 & I_{n_2} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_{n_m} & 0 \\ 0 & \beta A_2 & \dots & \beta A_m & I_l \end{pmatrix} \quad (6.55)$$

and

$$G_1 = \begin{pmatrix} S_1 - \beta A_1^\top A_1 & 0 & \dots & 0 & 0 \\ 0 & S_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & S_m & 0 \\ 0 & -A_2 & \dots & -A_m & \beta^{-1} I_l \end{pmatrix}. \quad (6.56)$$

It follows that

$$G_1 L_1 = M_1, \quad (6.57)$$

and

$$\mathbf{u}^k - \bar{\mathbf{u}}^k = L_1(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \quad (6.58)$$

where M_1 and L_1 are defined in (6.1) and (6.55), respectively.

Following on the heels of the above definitions, the assertion (6.47) can now be reformulated in terms of $\bar{\mathbf{u}}^k$ as shown below.

Lemma 6.10 *Let $\{\tilde{\mathbf{u}}^k\}$ be generated by Algorithm 6.3 with a given $\mathbf{u}^k$, and let the vector $\bar{\mathbf{u}}^k$ be defined by (6.54). Thus,*

$$\begin{aligned} \bar{\mathbf{u}}^k &\in \mathcal{U}, \theta(\mathbf{x}) - \theta(\bar{\mathbf{x}}) \\ &+ \langle \mathbf{u} - \bar{\mathbf{u}}^k, F(\bar{\mathbf{u}}^k) + G_1(\bar{\mathbf{u}}^k - \mathbf{u}^k) + \Psi(\mathbf{x}^k, \bar{\mathbf{x}}^k) \rangle \geq 0, \quad \forall \mathbf{u} \in \mathcal{U}, \end{aligned} \quad (6.59)$$

where G_1 is defined in (6.56) and $\Psi(\mathbf{x}^k, \bar{\mathbf{x}}^k)$ is defined in (6.26).

Lemma 6.11 *Let $\{\tilde{\mathbf{u}}^k\}$ be generated by Algorithm 6.3 with a given $\mathbf{u}^k$, and let the vector $\bar{\mathbf{u}}^k$ be as defined in (6.54). Thus,*

$$\|\mathbf{u}^k - \bar{\mathbf{u}}^k\|_Q^2 - \|\mathbf{u}^{k+1} - \bar{\mathbf{u}}^k\|_Q^2 \geq 0, \quad (6.60)$$

where $\mathbf{u}^{k+1}$ is defined in (6.45).

Proof:

$$\begin{aligned} &\|\mathbf{u}^k - \bar{\mathbf{u}}^k\|_Q^2 - \|\mathbf{u}^{k+1} - \bar{\mathbf{u}}^k\|_Q^2 \\ &= \|\mathbf{u}^k - \bar{\mathbf{u}}^k\|_Q^2 - \|\mathbf{u}^k - \bar{\mathbf{u}}^k - \alpha_k Q^{-1} d_1^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k)\|_Q^2 \\ &= 2\alpha_k \langle \mathbf{u}^k - \bar{\mathbf{u}}^k, d_1^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle - \alpha_k^2 \|Q^{-1} d_1^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k)\|_Q^2. \end{aligned} \quad (6.61)$$

By using the definitions of $\bar{\mathbf{u}}^k$ in (6.54) and M_1 in (6.1), we get

$$2\alpha_k \langle \mathbf{u}^k - \bar{\mathbf{u}}^k, d_1^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \rangle = 2\alpha_k \langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, L_1^\top M_1(\mathbf{u}^k - \tilde{\mathbf{u}}^k) - L_1^\top \Psi(\mathbf{x}^k, \tilde{\mathbf{x}}^k) \rangle.$$

Substituting this into the inequality (6.61) yields

$$\begin{aligned} &\|\mathbf{u}^k - \bar{\mathbf{u}}^k\|_Q^2 - \|\mathbf{u}^{k+1} - \bar{\mathbf{u}}^k\|_Q^2 \\ &= 2\alpha_k \langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, L_1^\top M_1(\mathbf{u}^k - \tilde{\mathbf{u}}^k) - L_1^\top \Psi(\mathbf{x}^k, \tilde{\mathbf{x}}^k) \rangle - \alpha_k^2 \|Q^{-1} d_1^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k)\|_Q^2. \end{aligned}$$

Using the definitions of L_1 , M_1 , and $\Psi(\mathbf{u}^k, \tilde{\mathbf{u}}^k)$, we get

$$\begin{aligned} &\langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, L_1^\top M_1(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle - \langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, L_1^\top \Psi(\mathbf{x}^k, \tilde{\mathbf{x}}^k) \rangle \\ &= \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_{M_1}^2 + \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) \right\rangle - \langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, \Psi(\mathbf{x}^k, \tilde{\mathbf{x}}^k) \rangle \\ &= \varphi_C(\mathbf{u}^k, \tilde{\mathbf{u}}^k). \end{aligned}$$

Therefore,

$$\begin{aligned}
 & \|\mathbf{u}^k - \bar{\mathbf{u}}^k\|_Q^2 - \|\mathbf{u}^{k+1} - \bar{\mathbf{u}}^k\|_Q^2 \\
 &= 2\alpha_k \varphi_C(\mathbf{u}^k, \tilde{\mathbf{u}}^k) - \alpha_k^2 \|Q^{-1}d_1^C(\mathbf{u}^k, \tilde{\mathbf{u}}^k)\|_Q^2 \\
 &= \alpha_k \varphi_C(\mathbf{u}^k, \tilde{\mathbf{u}}^k) \quad (\geq 0),
 \end{aligned}$$

and this completes the proof. $\square$

Lemma 6.12 *Let the sequences $\{\mathbf{u}^k\}$ and $\{\tilde{\mathbf{u}}^k\}$ be generated by Algorithm 6.3. Then,*

$$\alpha_k(\theta(\mathbf{x}) - \theta(\bar{\mathbf{x}}) + \langle \mathbf{u} - \bar{\mathbf{u}}^k, F(\mathbf{u}) \rangle) \geq \frac{1}{2}(\|\mathbf{u}^* - \mathbf{u}^{k+1}\|_Q^2 - \|\mathbf{u}^* - \mathbf{u}^k\|_Q^2), \forall \mathbf{u} \in \mathcal{U}. \quad (6.62)$$

Proof:

Using Lemma 6.10 and $\langle \mathbf{u} - \bar{\mathbf{u}}^k, F(\mathbf{u}) \rangle = \langle \mathbf{u} - \bar{\mathbf{u}}^k, F(\bar{\mathbf{u}}^k) \rangle$ yields

$$\begin{aligned}
 & \alpha_k(\theta(\mathbf{x}) - \theta(\bar{\mathbf{x}}) + \langle \mathbf{u} - \bar{\mathbf{u}}^k, F(\mathbf{u}) \rangle) \\
 & \geq \alpha_k \langle \mathbf{u} - \bar{\mathbf{u}}^k, G_1(\mathbf{u}^k - \bar{\mathbf{u}}^k) - \Psi(\mathbf{x}^k, \bar{\mathbf{x}}^k) \rangle, \forall \mathbf{u} \in \mathcal{U}.
 \end{aligned} \quad (6.63)$$

Now, for the right-hand side of (6.63), we have

$$\mathbf{u}^k - \bar{\mathbf{u}}^k = L_1(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \quad \text{and} \quad G_1 L_1 = M_1.$$

Thus, utilizing the

$$\alpha_k Q^{-1}[M_1(\mathbf{u}^k - \tilde{\mathbf{u}}^k) - \Psi(\mathbf{x}^k, \bar{\mathbf{x}}^k)] = \mathbf{u}^k - \mathbf{u}^{k+1}$$

yields

$$\alpha_k \langle \mathbf{u} - \bar{\mathbf{u}}^k, G_1 L_1(\mathbf{u}^k - \tilde{\mathbf{u}}^k) - \Psi(\mathbf{x}^k, \bar{\mathbf{x}}^k) \rangle = \langle \mathbf{u} - \bar{\mathbf{u}}^k, Q(\mathbf{u}^k - \mathbf{u}^{k+1}) \rangle. \quad (6.64)$$

Setting $a = \mathbf{u}$, $b = \bar{\mathbf{u}}^k$, $c = \mathbf{u}^k$ and $d = \mathbf{u}^{k+1}$ in the identity (1.7), we get:

$$\begin{aligned}
 & \langle \mathbf{u} - \bar{\mathbf{u}}^k, Q(\mathbf{u}^k - \mathbf{u}^{k+1}) \rangle \\
 &= \frac{1}{2}(\|\mathbf{u} - \mathbf{u}^{k+1}\|_Q^2 - \|\mathbf{u} - \mathbf{u}^k\|_Q^2) + \frac{1}{2}(\|\mathbf{u}^k - \bar{\mathbf{u}}^k\|_Q^2 - \|\mathbf{u}^{k+1} - \bar{\mathbf{u}}^k\|_Q^2).
 \end{aligned} \quad (6.65)$$

Combining (6.63), (6.64), and (6.65) leads to,

$$\begin{aligned}
 & \alpha_k(\theta(\mathbf{x}) - \theta(\bar{\mathbf{x}}) + \langle \mathbf{u} - \bar{\mathbf{u}}^k, F(\mathbf{u}) \rangle) \\
 & \geq \frac{1}{2}(\|\mathbf{u} - \mathbf{u}^{k+1}\|_Q^2 - \|\mathbf{u} - \mathbf{u}^k\|_Q^2) + \frac{1}{2}(\|\mathbf{u}^k - \bar{\mathbf{u}}^k\|_Q^2 - \|\mathbf{u}^{k+1} - \bar{\mathbf{u}}^k\|_Q^2), \forall \mathbf{u} \in \mathcal{U},
 \end{aligned} \quad (6.66)$$

and the assertion (6.62) follows from (6.60) and (6.66). $\square$

We are now finally ready to prove the worst-case $O(1/t)$ convergence rate result in an ergodic sense for Algorithm 6.3.

Theorem 6.5 *Let the sequence $\{\mathbf{u}^k\}$ be generated by Algorithm 6.3. For a positive integer t , let*

$$\bar{\mathbf{u}}_t := \frac{1}{\Upsilon_t} \sum_{k=0}^t \alpha_k \bar{\mathbf{u}}^k \quad \text{with} \quad \Upsilon_t = \sum_{k=0}^t \alpha_k,$$

then $\bar{\mathbf{u}}_t \in \mathcal{U}$ and

$$\theta(\bar{\mathbf{x}}_t) - \theta(\mathbf{x}) + \langle \bar{\mathbf{u}}_t - \mathbf{u}, F(\mathbf{u}) \rangle \leq \frac{1}{2(t+1)\gamma c_C} \|\mathbf{u} - \mathbf{u}^0\|_Q^2, \forall \mathbf{u} \in \mathcal{U}, \quad (6.67)$$

where c_C defined by (6.21), is a lower bound of the step length α_k^C .

Proof: The proof is similar to that of Theorem 5.3. $\square$

The foregoing theorem demonstrates that for a given arbitrary compact set $\mathcal{D}_{\mathcal{U}}(\bar{\mathbf{u}}_t) \subset \mathcal{U}$, after t iterations of Algorithm 6.3, we can obtain an approximate solution $\bar{\mathbf{u}}_t \in \mathcal{U}$, given by (4.15), which satisfies

$$\sup_{\mathbf{u} \in \mathcal{D}_{\mathcal{U}}(\bar{\mathbf{u}}_t)} \{\theta(\bar{\mathbf{x}}_t) - \theta(\mathbf{x}) + \langle \bar{\mathbf{u}}_t - \mathbf{u}, F(\mathbf{u}) \rangle\} \leq \frac{\bar{d}}{2(t+1)\gamma c_0},$$

where

$$\bar{d} = \sup\{\|\mathbf{u} - \mathbf{u}^0\|_Q^2 | \mathbf{u} \in \mathcal{D}_{\mathcal{U}}(\bar{\mathbf{u}}_t)\},$$

and $\mathbf{u}^0$ is an initialization point. Therefore, the convergence rate $O(1/t)$ of Algorithm 6.3 is obtained under the ergodic sense.

6.4 Numerical Results

In this section, we investigate the computational performance of the proposed algorithms by solving several synthetically generated convex quadratic programming problems, originally proposed in [55]. All of these numerical experiments were implemented in MATLAB R2011b on a Windows 7 machine equipped with an Intel Core i5-3740 CPU and 4GB of memory.

Specifically, the quadratic programming problem used in our computational tests has the following structure:

$$\min \quad \sum_{i=1}^3 \frac{1}{2} x_i^\top M_i x_i + q_i^\top x_i \quad (6.68a)$$

$$\text{s.t.} \quad A_1 x_1 + x_2 + A_3 x_3 = b, \quad x_1 \in \mathcal{B}, x_2 \in \odot_r, x_3 \in \mathbb{R}_+^{n_3}, \quad (6.68b)$$

where the set $\mathcal{B}$ is a box, i.e., $\mathcal{B} = [B_L, B_U]$, $\odot_r$ denotes a ball centered at the origin with radius r , and $\mathbb{R}_+^{n_3}$ denotes the nonnegative orthant of appropriate dimension. Additionally, as in [55], the matrix $M_1 = I_{n_1 \times n_1}$ and the matrices M_i ($i = 2, 3$) are generated using the following procedure:

$$t = 1000, \quad V_i = \text{rand}(0.2 * n_i, n_i), \\ \tau = \frac{\lambda_{\max}(V_i^\top V_i) - t \cdot \lambda_{\min}(V_i^\top V_i)}{t - 1}, \quad M_i = V_i^\top V_i + \tau I_{n_1 \times n_1}.$$

Moreover, the entries of $A_i \in \mathbb{R}^{l \times n_i}$ ($i = 1, 3$) in the linear constraint (6.68b) are uniformly distributed in $(0, 1)$ with density $\rho = 0.1$. In our experiments, we set the box $[B_L, B_U] = [0, 10]$ and the radius $r = 10$. Furthermore, we set $q_i = -M_i x_i^*$ ($i = 1, 2, 3$) and $b = A_1 x_1^* + x_2^* + A_3 x_3^*$, where $(x_1^*, x_2^*, x_3^*) \in \mathcal{B} \times \odot_r \times \mathbb{R}_+^{n_3}$ are given feasible points generated by $x_i^* = \text{sprand}(n_i, 1, 0.5)$, $\forall i = 1, 2, 3$. Note that such a data generation scheme ensures that all of the tested instances have feasible solutions.

As some of the subproblems arising from the methods PPSA, LPSM-A, do not result in closed-form solutions (for some problem instances), we resort to using an *inner iterative procedure* to obtain approximate solutions. The maximum number of iterations for this inner iterative loop is as 30, and the related stopping precision (tolerance limit) is taken to be $\epsilon = 10^{-4}$. Furthermore, we set the maximum number of outer iterations 5000, and the corresponding stopping criterion for the outer loop is defined as follows:

$$\text{tol} = \max \left\{ \max_{1 \leq i \leq 3} \{\|x_i^k - \tilde{x}_i^k\|\}, \|\lambda^k - \tilde{\lambda}^k\| \right\} \leq 10^{-4}.$$

In our computational implementation, some of the parameters involved in the proposed methods are specified as follows. Beginning with $\gamma = 1.9$, for PPSA and LPSM-A, we choose $\beta = 0.01$ and $S_i = 0.01 \|A_i^\top A_i\|_2 I_i$, ($i = 1, 2, 3$). For LPSM-B and LPSM-C, we set $\beta = 0.04$, $S_i = (\|M_i\|_2 + 0.08 \|A_i^\top A_i\|_2) I_i$ and $S_i = (\|M_i\|_2 + 0.08 \|A_i^\top A_i\|_2) I_i$, ($i = 1, 2, 3$), respectively. Finally, the initial iteration points are set as $x_i^0 = 0_{n_i \times 1}$ ($i = 1, 2, 3$) and $\lambda^0 = 0_{l \times 1}$.

Remark 6.5 *Under certain parameter settings, recognize that LPSM-B and LPSM-C can be reduced to be equivalent to one another. However, in our numerical study, to differentiate between the alternative linearization strategies, such specialized cases are not considered here.*

Table 6.1 displays the obtained numerical results, where “Iter.” and “Time” denote the number of iterations and the CPU time (in seconds), respectively. For PPSA, LPSM-A, LPSM-B, the number of inner iterations (“Iter.”) are also reported for solving the corresponding subproblems. Note that as these problem instances are synthetically generated with random data, ten instances corresponding to each dimensional configuration are generated and the average performance for each scenario is reported. From the results recorded in Table 6.1, it is evident that LPSM-C is computationally more efficient than all the other algorithms, by consuming only 35%, 37%, and 12% of the CPU time taken by PPSA, LPSM-A, and LPSM-B, respectively. We attribute this to the fact that the subproblems in LPSM-C are greatly simplified to possess a closed-form solution (refer Remark 6.5), which in turn leads to a reduced overall computational time. This is despite the fact that the number of iterations consumed by LPSM-C is greater than the number of iterations required by PPSA, LPSM-A, and LPSM-B.

Table 6.1: Numerical comparisons based on synthetically generated data.

n_1	n_2	n_3	PPSA			LPSM-A			LPSM-B			LPSM-C		
			Iter.	Init.	Time	Iter.	Init.	Time	Iter.	Init.	Time	Iter.	Time	Time
100	300	100	273.2	12614.5	9.71	277.2	12810.3	9.63	3638.0	27114.1	25.16	3639.00	4.03	
300	300	300	330.0	20460.0	18.44	328.7	20379.4	17.50	1491.0	16870.8	26.22	1700.00	2.93	
500	300	500	230.0	14260.0	15.56	229.2	14210.4	15.07	1067.0	12751.3	45.33	1090.00	4.17	
600	300	600	228.0	14136.0	23.75	227.1	14080.2	20.16	1125.0	13043.3	73.54	1191.00	6.21	
800	300	800	162.4	10068.8	38.79	161.0	9982.0	34.93	1095.0	12475.0	144.48	1195.00	9.39	
1000	300	1000	94.0	5828.0	36.11	96.0	5952.0	32.78	1062.0	12398.0	217.09	1176.00	12.52	
100	600	100	92.0	5639.0	8.76	92.0	5641.0	8.78	4060.1	32736.4	37.73	4062.00	9.73	
300	600	300	199.0	12338.0	22.16	197.0	12214.0	21.45	3944.0	40853.0	80.39	4000.00	16.44	
500	600	500	479.0	29698.0	57.12	485.7	30113.4	54.88	2355.0	28066.3	109.90	2711.00	16.79	
600	600	600	494.8	30677.6	71.68	496.0	30752.0	62.58	2030.0	24770.0	147.09	2344.00	17.19	
800	600	800	528.9	32791.8	133.26	535.0	33170.0	119.94	1311.0	17203.0	194.23	1351.00	13.59	
1000	600	1000	427.0	26474.0	162.24	431.9	26777.8	148.05	1172.0	15346.0	255.73	1224.00	16.01	

6.5 Conclusions

In this chapter, we introduced three linearized partially-parallel splitting methods based on proximal regularization techniques, which can be used for solving convex programming problems with separable structure in the presence of a linear coupling constraint. A significant advantage arising from our approach is that the subproblems can be resolved quickly as they likely possess closed-form solutions, or are relatively easy to solve as compared to their original (nonlinear) versions. Following the general framework of projection contraction methods, we also proved the global convergence of these linearized algorithms. As a byproduct, the worst-case $O(1/t)$ convergence rate was also established in an ergodic scenario. Some preliminary numerical results further illustrated the feasibility and efficiency of these linearized methods, as a viable choice for solving large-scale convex programming problems.

Chapter 7

A primal-dual prediction-correction algorithm

7.1 Motivation

In our final research contribution, we turn our attention to solving saddle point problems, which are loosely related to separable convex programs. As aforementioned in Chapter 1, several primal-dual algorithms have been presented for solving the saddle point problem (1.32). In [77], He and Yuan proposed a class of primal-dual algorithms but convergence was guaranteed only under restrictive assumptions. However, they showed that the set from which the involved proximal parameters need to be selected could be expanded by incorporating additional predictor-corrector strategies. Global convergence results are proved on the basis of the framework of contraction-type methods, and numerical results demonstrate the efficiency of these methods. Then, adopting a new correction strategy, Cai et al. [12] proposed an improved version of He and Yuan's algorithm. As compared to the algorithm in [77], global convergence can be established under more flexible assumptions, and the computational effectiveness is further verified by some numerical experiments.

Motivated by these research works, we develop a new primal-dual algorithm, also in the predictor-corrector fashion, which aims to bridge the gap between the approaches in [12, 77]. Furthermore, we also obtain a *partially linearized* primal-dual algorithm, which only requires two projections during the prediction stage. This resulting linearized version of our algorithm is relatively easier to implement as compared to the algorithms

in [12, 15, 59] as long as the projections of $\mathcal{X}$ and $\mathcal{Y}$ are readily available as closed-form solutions or are simple enough to compute numerically. Recently, Chambolle and Pock [16] established the $O(1/t)$ convergence rate of the FOPDA method in an ergodic sense. In a similar fashion, we also establish the $O(1/t)$ convergence rate for the proposed algorithm under ergodic condition.

7.2 Algorithm

To begin with, define:

$$M := \begin{pmatrix} \frac{1}{\tau}I & A^\top \\ \theta A & \frac{1}{\sigma}I + 2\rho B^\top B \end{pmatrix} \quad \text{and} \quad H := \begin{pmatrix} \frac{1}{\tau}I & 0 \\ 0 & \frac{1}{\sigma}I + 2\rho B^\top B \end{pmatrix}, \quad (7.1)$$

where $\tau, \theta, \sigma, \rho$ are given constants such that both M and H are positive definite, denoted using standard notation as $M \succ 0$ and $H \succ 0$.

We are now ready to formally present the algorithmic framework of the primal-dual prediction-correction algorithm where, for the sake of notational convenience, we have represented $\mathbf{u} := (x, y)$ and $\lambda_{\max}(B^\top B)$ denotes the maximum eigenvalue of $B^\top B$.

Algorithm 7.1 A primal-dual prediction-correction algorithm.

Step 0. Initialization: Select $\tau > 0$, $\sigma > 0$, $\theta \in \mathbb{R}$, and $\rho > -1/(2\sigma\lambda_{\max}(B^\top B))$.

Step 1. Generate a pair of predictors $\tilde{\mathbf{u}}^k := (\tilde{x}^k, \tilde{y}^k)$ via

$$\begin{cases} \tilde{x}^k = \arg \max_{x \in \mathcal{X}} \left\{ \tau \Phi(x, y^k) - \frac{1}{2} \|x - x^k\|^2 \right\}, & (7.2a) \end{cases}$$

$$\begin{cases} \bar{x}^k = \tilde{x}^k + \theta(\tilde{x}^k - x^k), & (7.2b) \end{cases}$$

$$\begin{cases} \tilde{y}^k = \arg \min_{y \in \mathcal{Y}} \left\{ \sigma \Phi(\bar{x}^k, y) + \frac{1}{2} \|y - y^k\|^2 + \sigma \rho \|B(y - y^k)\|^2 \right\}. & (7.2c) \end{cases}$$

Step 2. Termination Step: If

$$\max \{ \|x^k - \tilde{x}^k\|, \|y^k - \tilde{y}^k\| \} < \epsilon,$$

Step 3. Update the next iterate $\mathbf{u}^{k+1} := (x^{k+1}, y^{k+1})$ via

$$\mathbf{u}^{k+1} = \mathbf{u}^k - \gamma \alpha_k H^{-1} M(\mathbf{u}^k - \tilde{\mathbf{u}}^k), \quad \gamma \in (0, 2), \quad (7.3)$$

where H and M are defined by (7.1) and

$$\alpha_k := \frac{\langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle}{\|H^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|_H^2}. \quad (7.4)$$

Remark 7.1 When we set $\rho := 0$ in (7.2c), Algorithm 7.1 reduces to the first algorithm proposed in [59]. Besides, if we set $\rho := \nu$, while taking H as the identity matrix and

$$\alpha_k^* := \frac{\langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, \widetilde{M}(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle}{\|M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|^2}, \quad \text{where} \quad \widetilde{M} := \begin{pmatrix} \frac{1}{\theta} I & A^\top \\ \frac{1}{\theta} A & \frac{1}{\sigma} I + \nu B^\top B \end{pmatrix}, \quad (7.5)$$

as the step size (instead of α_k in (7.3)), the iterative scheme then coincides with the algorithm introduced in [12]. Clearly, Algorithm 7.1 can be regarded as an extension of the respective methods proposed in [12, 77].

Note that the y -related subproblem (7.2c) is essentially a constrained minimization problem if $\mathcal{Y}$ is a proper subset of $\mathbb{R}^n$. In this situation, Algorithm 7.1 fails to be easily implemented when $\rho := 0$ (as is the case in [12, 15, 77]). Indeed, observe that, when taking $\rho := -\nu/2$, the iterative schemes (7.2a) and (7.2c) amount to two projection steps, given by,

$$\begin{cases} \tilde{x}^k = P_{\mathcal{X}}(x^k + \tau A^\top y^k), \end{cases} \quad (7.6a)$$

$$\begin{cases} \tilde{y}^k = P_{\mathcal{Y}}(y^k - \sigma A \tilde{x}^k + \sigma \nu B^\top (B y^k - b)). \end{cases} \quad (7.6b)$$

Clearly, the projection steps (7.6a) and (7.6b) are computationally easier to implement than the original projection operators introduced in [12, 15, 77], as long as the sets $\mathcal{X}$ and $\mathcal{Y}$ are relatively simple. It is worthwhile to note that such a linearization strategy has appeared in a recent work [124], where the authors only considered the case of a fixed $\theta = 1$ in (7.2b), and showed that this variant yields significant computational benefits.

7.3 Convergence results

In this section, we first prove that Algorithm 7.1 is globally convergent to a solution of (1.32). Our secondary goal is to establish the worst-case $O(1/t)$ convergence rate result in an ergodic sense, which is not estimated in [12, 77]. In our analysis, we impose the same assumptions on the parameters τ , σ , and δ , as stated in [77].

Lemma 7.1 *Let $\{\mathbf{u}^k := (x^k, y^k)\}$ be generated by Algorithm 7.1. Then, we have*

$$\theta(y) - \theta(\tilde{y}^k) + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, F(\tilde{\mathbf{u}}^k) + M(\tilde{\mathbf{u}}^k - \mathbf{u}^k) \rangle \geq 0, \quad \forall \mathbf{u} \in \Omega, \quad (7.7)$$

where M is as given in (7.1).

Proof: By invoking the first-order optimality conditions of (7.2a) and (7.2c), we have

$$\begin{cases} \langle x - \tilde{x}^k, -A^\top y^k + \frac{1}{\tau}(\tilde{x}^k - x^k) \rangle \geq 0, \quad \forall x \in \mathcal{X}, \\ \theta(y) - \theta(\tilde{y}^k) + \langle y - \tilde{y}^k, A\tilde{x}^k + (\frac{1}{\sigma}I + 2\rho B^\top B)(\tilde{y}^k - y^k) \rangle \geq 0, \quad \forall y \in \mathcal{Y}. \end{cases}$$

Using the notation in (1.8), together with (7.2b), yields the desired inequality (7.7). $\square$

Lemma 7.2 *Let $\mathbf{u}^* := (x^*, y^*)$ be an arbitrary solution of (1.32). Then, the sequences $\{\mathbf{u}^k\}$ and $\{\tilde{\mathbf{u}}^k\}$ generated by Algorithm 7.1 satisfy the following property:*

$$\langle \mathbf{u}^k - \mathbf{u}^*, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle \geq \langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle. \quad (7.9)$$

Proof: Note that the inequality (7.7) holds for any $\mathbf{u} \in \Omega$. As a consequence, we have

$$\theta(y^*) - \theta(\tilde{y}^k) + \langle \mathbf{u}^* - \tilde{\mathbf{u}}^k, F(\tilde{\mathbf{u}}^k) + M(\tilde{\mathbf{u}}^k - \mathbf{u}^k) \rangle \geq 0, \quad \forall \mathbf{u}^* \in \Omega. \quad (7.10)$$

On the other hand, it follows from (1.33) that

$$\theta(\tilde{y}^k) - \theta(y^*) + \langle \tilde{\mathbf{u}}^k - \mathbf{u}^*, F(\mathbf{u}^*) \rangle \geq 0. \quad (7.11)$$

Adding (7.10) and (7.11) yields

$$\langle \tilde{\mathbf{u}}^k - \mathbf{u}^*, F(\mathbf{u}^*) - F(\tilde{\mathbf{u}}^k) \rangle \geq \langle \mathbf{u}^* - \tilde{\mathbf{u}}^k, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle,$$

which together with the monotonicity of F implies that

$$\langle \tilde{\mathbf{u}}^k - \mathbf{u}^*, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle \geq 0.$$

Thus, we conclude that

$$\langle \mathbf{u}^k - \mathbf{u}^*, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle \geq \langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle,$$

and the assertion of this lemma is obtained. $\square$

Remark 7.2 When $\rho = \nu > 0$, note that the matrix M can be expressed as the sum of two parts

$$M = \widetilde{M} + N, \quad \text{where} \quad N := \begin{pmatrix} 0 & 0 \\ 0 & \rho B^\top B \end{pmatrix},$$

and $\widetilde{M}$ is as previously defined in (7.5). Consequently, the term on the right hand side of (7.9) satisfies

$$\begin{aligned} \langle \mathbf{u}^k - \widetilde{\mathbf{u}}^k, M(\mathbf{u}^k - \widetilde{\mathbf{u}}^k) \rangle &= \langle \mathbf{u}^k - \widetilde{\mathbf{u}}^k, \widetilde{M}(\mathbf{u}^k - \widetilde{\mathbf{u}}^k) \rangle + \rho \|B(y^k - \widetilde{y}^k)\|^2 \\ &\geq \langle \mathbf{u}^k - \widetilde{\mathbf{u}}^k, \widetilde{M}(\mathbf{u}^k - \widetilde{\mathbf{u}}^k) \rangle, \end{aligned}$$

and this validates the reasonableness of the choice of step size used in [12].

It is easy to verify that the matrix H , defined in (7.1), is always positive definite under the condition that $\rho > -1/(2\sigma\lambda_{\max}(B^\top B))$. Therefore, it follows that

$$\langle \mathbf{u}^k - \widetilde{\mathbf{u}}^k, M(\mathbf{u}^k - \widetilde{\mathbf{u}}^k) \rangle \geq \frac{\delta}{1+\delta} \|\mathbf{u}^k - \widetilde{\mathbf{u}}^k\|_H^2 > 0, \quad \forall \mathbf{u}^k \neq \widetilde{\mathbf{u}}^k. \quad (7.12)$$

A formal proof of this result is presented below, following a similar line of reasoning as used in [77, 84].

Lemma 7.3 Let H and M be defined as (7.1). If the parameters τ and σ satisfy

$$\tau \|A^\top A\| \frac{(1+\theta)^2}{4} < \frac{1+2\rho\sigma\lambda_{\min}(B^\top B)}{\sigma},$$

we have

$$\langle \mathbf{u} - \widetilde{\mathbf{u}}, M(\mathbf{u} - \widetilde{\mathbf{u}}) \rangle \geq \frac{\delta}{1+\delta} \|\mathbf{u} - \widetilde{\mathbf{u}}\|_H^2 > 0, \quad \forall \mathbf{u} \neq \widetilde{\mathbf{u}},$$

where

$$\delta := \begin{cases} \frac{2}{|1+\theta|} \sqrt{\frac{1+2\rho\sigma\lambda_{\min}(B^\top B)}{\tau\sigma\|A^\top A\|}} - 1 > 0, & \text{if } \theta \neq -1. \\ \text{arbitray positive numbers,} & \text{if } \theta = -1. \end{cases}$$

Proof: When $\theta = -1$, we can easily get

$$\langle \mathbf{u} - \widetilde{\mathbf{u}}, M(\mathbf{u} - \widetilde{\mathbf{u}}) \rangle = \|\mathbf{u} - \widetilde{\mathbf{u}}\|_H^2 \geq \frac{\delta}{1+\delta} \|\mathbf{u} - \widetilde{\mathbf{u}}\|_H^2 > 0, \quad \forall \mathbf{u} \neq \widetilde{\mathbf{u}},$$

for any positive number $\delta > 0$. Next, we mainly discuss the case of $\theta \neq -1$. Under the assumed condition that $\delta > 0$, it holds that

$$\frac{\tau(1+\delta)\|A^\top A\|(1+\theta)^2}{4} = \frac{1+2\rho\sigma\lambda_{\min}(B^\top B)}{\sigma(1+\delta)}.$$

Using the Cauchy-Schwarz inequality yields

$$\begin{aligned}
 & (1 + \theta) \langle y - \tilde{y}, A(x - \tilde{x}) \rangle \\
 & \geq -(\tau(1 + \delta) \|A^\top A\|) \frac{(1 + \theta)^2}{4} \|y - \tilde{y}\|^2 - \frac{1}{\tau(1 + \delta) \|A^\top A\|} \|A(x - \tilde{x})\|^2 \\
 & = -\frac{(1 + 2\rho\sigma\lambda_{\min}(B^\top B))}{\sigma(1 + \delta)} \|y - \tilde{y}\|^2 - \frac{1}{\tau(1 + \delta) \|A^\top A\|} \|A(x - \tilde{x})\|^2 \\
 & \geq -\frac{1}{1 + \delta} \left[\frac{1}{\tau} \|x - \tilde{x}\|^2 + \left(\frac{1}{\sigma} + 2\rho\lambda_{\min}(B^\top B) \right) \|y - \tilde{y}\|^2 \right].
 \end{aligned}$$

On the other hand,

$$\begin{aligned}
 \|\mathbf{u} - \tilde{\mathbf{u}}\|_H^2 &= \frac{1}{\tau} \|x - \tilde{x}\|^2 + \frac{1}{\sigma} \|y - \tilde{y}\|^2 + 2\rho \|y - \tilde{y}\|_{B^\top B}^2 \\
 &\geq \frac{1}{\tau} \|x - \tilde{x}\|^2 + \left(\frac{1}{\sigma} + 2\rho\lambda_{\min}(B^\top B) \right) \|y - \tilde{y}\|^2 \\
 &\geq 0.
 \end{aligned}$$

Combining the above inequalities leads to

$$\frac{1}{1 + \delta} \|\mathbf{u} - \tilde{\mathbf{u}}\|_H^2 + (1 + \theta) \langle y - \tilde{y}, A(x - \tilde{x}) \rangle \geq 0.$$

For any $\mathbf{u} \neq \tilde{\mathbf{u}}$, we get

$$\begin{aligned}
 & \langle \mathbf{u} - \tilde{\mathbf{u}}, M(\mathbf{u} - \tilde{\mathbf{u}}) \rangle \\
 &= \|\mathbf{u} - \tilde{\mathbf{u}}\|_H^2 + (1 + \theta) \langle y - \tilde{y}, A(x - \tilde{x}) \rangle \\
 &= \frac{\delta}{1 + \delta} \|\mathbf{u} - \tilde{\mathbf{u}}\|_H^2 + \frac{1}{1 + \delta} \|\mathbf{u} - \tilde{\mathbf{u}}\|_H^2 + (1 + \theta) \langle y - \tilde{y}, A(x - \tilde{x}) \rangle \\
 &\geq \frac{\delta}{1 + \delta} \|\mathbf{u} - \tilde{\mathbf{u}}\|_H^2 > 0.
 \end{aligned}$$

Hence, the assertion of this lemma is proved. $\square$

Therefore, it follows from Lemma 7.3 that the inequality (7.12) holds. Then, an immediate consequence to (7.9) and (7.12) is that,

$$\langle H(\mathbf{u}^k - \mathbf{u}^*), H^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle \geq \frac{\delta}{1 + \delta} \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_H^2 > 0, \quad \forall \mathbf{u}^k \neq \tilde{\mathbf{u}}^k.$$

Obviously, the above inequality indicates that, at any iteration k of the algorithm, $-H^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)$ is a *descent direction* of the distance function $\frac{1}{2} \|\mathbf{u} - \mathbf{u}^*\|_H^2$, which also ensures that the correction step (7.3) is well defined.

Theorem 7.1 *Let $\mathbf{u}^*$ be an arbitrary solution of (1.32). Then, there exists a constant $c > 0$ such that the sequence $\{\mathbf{u}^k\}$ generated by Algorithm 7.1 satisfies*

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_H^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|_H^2 - c\|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_H^2. \quad (7.13)$$

Proof: Recalling the iterative scheme (7.3), we have

$$\begin{aligned} & \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_H^2 \\ &= \|\mathbf{u}^k - \gamma\alpha_k H^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) - \mathbf{u}^*\|_H^2 \\ &= \|\mathbf{u}^k - \mathbf{u}^*\|_H^2 - 2\gamma\alpha_k \langle \mathbf{u}^k - \mathbf{u}^*, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle + \gamma^2\alpha_k^2 \|H^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|_H^2 \\ &\leq \|\mathbf{u}^k - \mathbf{u}^*\|_H^2 - 2\gamma\alpha_k \langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle + \gamma^2\alpha_k^2 \|H^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|_H^2 \\ &= \|\mathbf{u}^k - \mathbf{u}^*\|_H^2 - \gamma(2 - \gamma)\alpha_k \langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle \\ &\leq \|\mathbf{u}^k - \mathbf{u}^*\|_H^2 - \gamma(2 - \gamma)\alpha_k \frac{\delta}{1 + \delta} \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_H^2. \end{aligned} \quad (7.14)$$

It follows from the definition of α_k , defined by (7.4), that

$$\begin{aligned} \alpha_k &:= \frac{(\mathbf{u}^k - \tilde{\mathbf{u}}^k)^\top M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)}{\|H^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k)\|_H^2} \\ &\geq \frac{\delta \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_H^2}{(1 + \delta) \|M^\top H^{-1}M\| \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|^2} \\ &\geq \frac{\delta \lambda_{\min}(H)}{(1 + \delta) \|M^\top H^{-1}M\|} \\ &= \frac{\delta \cdot \min\left\{\frac{1}{\tau}, \frac{1}{\sigma} + 2\rho\lambda_{\min}(B^\top B)\right\}}{(1 + \delta) \|M^\top H^{-1}M\|} \\ &:= \alpha_{\min} > 0, \end{aligned}$$

where $\lambda_{\min}(\bullet)$ denotes the minimum eigenvalue of a matrix $(\bullet)$. Therefore, we conclude from (7.14) that

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_H^2 \leq \|\mathbf{u}^k - \mathbf{u}^*\|_H^2 - \left(\frac{\gamma(2 - \gamma)\delta\alpha_{\min}}{1 + \delta} \right) \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_H^2.$$

Setting $c := (\gamma(2 - \gamma)\delta\alpha_{\min}) / (1 + \delta)$ completes the proof. $\square$

The above theorem proves that the sequence $\{\mathbf{u}^k\}$ generated by Algorithm 7.1 is Fejér monotone with respect to the solution set of (1.8). Following the line of reasoning presented in Theorem 3.7 [77] (see also [12, 15]), we can also prove that Algorithm 7.1 enjoys global convergence. Therefore, we only state the following theorem and refer the reader to the aforementioned references for a detailed proof of this result.

Theorem 7.2 *The sequence $\{\mathbf{u}^k\}$ generated by Algorithm 7.1 is globally convergent to a solution point of (1.32).*

Next, we prove that Algorithm 7.1 has a worst-case $O(1/t)$ convergence rate in an ergodic sense. We begin this analysis with a fundamental inequality proved in the following lemma.

Lemma 7.4 *Let the sequences $\{\mathbf{u}^k\}$ and $\{\tilde{\mathbf{u}}^k\}$ be generated by Algorithm 7.1. Then, the following inequality*

$$\theta(y) - \theta(\tilde{y}^k) + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, F(\mathbf{u}) \rangle + \frac{1}{2\gamma\alpha_k} (\|\mathbf{u} - \mathbf{u}^k\|_H^2 - \|\mathbf{u} - \mathbf{u}^{k+1}\|_H^2) \geq \frac{(2-\gamma)\delta}{2(1+\delta)} \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_H^2, \quad (7.15)$$

holds for all $\mathbf{u} \in \Omega$, where H is as defined in (7.1).

Proof: Rearranging the terms in (7.7), we get

$$\theta(y) - \theta(\tilde{y}^k) + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, F(\tilde{\mathbf{u}}^k) \rangle \geq \langle \mathbf{u} - \tilde{\mathbf{u}}^k, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle. \quad (7.16)$$

It follows from the monotonicity of F that

$$\langle \mathbf{u} - \tilde{\mathbf{u}}^k, F(\mathbf{u}) \rangle \geq \langle \mathbf{u} - \tilde{\mathbf{u}}^k, F(\tilde{\mathbf{u}}^k) \rangle, \quad \forall \mathbf{u} \in \Omega.$$

Clearly, from (7.16), we have

$$\theta(y) - \theta(\tilde{y}^k) + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, F(\mathbf{u}) \rangle \geq \langle \mathbf{u} - \tilde{\mathbf{u}}^k, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle, \quad \forall \mathbf{u} \in \Omega. \quad (7.17)$$

Note that the correction step (7.3) can be rewritten as

$$\mathbf{u}^k - \tilde{\mathbf{u}}^k = \frac{1}{\gamma\alpha_k} M^{-1} H(\mathbf{u}^k - \mathbf{u}^{k+1}),$$

and consequently, we have

$$\langle \mathbf{u} - \tilde{\mathbf{u}}^k, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle = \frac{1}{\gamma\alpha_k} \langle \mathbf{u} - \tilde{\mathbf{u}}^k, H(\mathbf{u}^k - \mathbf{u}^{k+1}) \rangle. \quad (7.18)$$

An application, on the right hand side of (7.18), of the following equality

$$\langle \mathbf{a} - \mathbf{b}, H(\mathbf{c} - \mathbf{d}) \rangle = \frac{1}{2} (\|\mathbf{a} - \mathbf{d}\|_H^2 - \|\mathbf{a} - \mathbf{c}\|_H^2) + \frac{1}{2} (\|\mathbf{c} - \mathbf{b}\|_H^2 - \|\mathbf{d} - \mathbf{b}\|_H^2)$$

with settings $\mathbf{a} := \mathbf{u}$, $\mathbf{b} := \tilde{\mathbf{u}}^k$, $\mathbf{c} := \mathbf{u}^k$, and $\mathbf{d} := \mathbf{u}^{k+1}$, yields

$$\begin{aligned} & \langle \mathbf{u} - \tilde{\mathbf{u}}^k, H(\mathbf{u}^k - \mathbf{u}^{k+1}) \rangle \\ &= \frac{1}{2} (\|\mathbf{u} - \mathbf{u}^{k+1}\|_H^2 - \|\mathbf{u} - \mathbf{u}^k\|_H^2) + \frac{1}{2} (\|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_H^2 - \|\mathbf{u}^{k+1} - \tilde{\mathbf{u}}^k\|_H^2). \end{aligned} \quad (7.19)$$

The last term of (7.19) satisfies

$$\begin{aligned} & \frac{1}{2} (\|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_H^2 - \|\mathbf{u}^{k+1} - \tilde{\mathbf{u}}^k\|_H^2) \\ &= \frac{1}{2} (\|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_H^2 - \|\mathbf{u}^k - \gamma\alpha_k H^{-1}M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) - \tilde{\mathbf{u}}^k\|_H^2) \\ &= \frac{1}{2} \gamma(2 - \gamma)\alpha_k \langle \mathbf{u}^k - \tilde{\mathbf{u}}^k, M(\mathbf{u}^k - \tilde{\mathbf{u}}^k) \rangle \\ &\geq \frac{\delta\gamma(2 - \gamma)\alpha_k}{2(1 + \delta)} \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_H^2, \end{aligned} \quad (7.20)$$

where the first equality comes from (7.3) and the second inequality follows from (7.12). Combining (7.17), (7.18), (7.19), and (7.20) leads to

$$\begin{aligned} & \theta(y) - \theta(\tilde{y}^k) + \\ & \langle \mathbf{u} - \tilde{\mathbf{u}}^k, F(\mathbf{u}) \rangle \geq \frac{1}{2\gamma\alpha_k} (\|\mathbf{u} - \mathbf{u}^{k+1}\|_H^2 - \|\mathbf{u} - \mathbf{u}^k\|_H^2) + \frac{(2 - \gamma)\delta}{2(1 + \delta)} \|\mathbf{u}^k - \tilde{\mathbf{u}}^k\|_H^2, \end{aligned}$$

and the desired result is obtained by rearranging the terms in the above inequality. $\square$

Theorem 7.3 *For any integer $t > 0$, we denote*

$$\bar{\mathbf{u}}_t = \frac{1}{\Upsilon_t} \sum_{k=0}^t \alpha_k \tilde{\mathbf{u}}^k, \quad \text{with} \quad \Upsilon_t = \sum_{k=0}^t \alpha_k, \quad (7.21)$$

where $\tilde{\mathbf{u}}^k$ ($k = 0, 1, \dots, t$) are generated by Algorithm 7.1 and α_k is given by (7.4). Then, we have $\bar{\mathbf{u}}_t \in \Omega$ and

$$\theta(\bar{y}_t) - \theta(y) + \langle \bar{\mathbf{u}}_t - \mathbf{u}, F(\mathbf{u}) \rangle \leq \frac{1}{2\gamma\alpha_{\min}(t+1)} \|\mathbf{u} - \mathbf{u}^0\|_H^2.$$

Proof: First, as a direct consequence of (7.15), we have

$$\alpha_k \{ \theta(y) - \theta(\tilde{y}^k) + \langle \mathbf{u} - \tilde{\mathbf{u}}^k, F(\mathbf{u}) \rangle \} + \frac{1}{2\gamma} \|\mathbf{u} - \mathbf{u}^k\|_H^2 \geq \frac{1}{2\gamma} \|\mathbf{u} - \mathbf{u}^{k+1}\|_H^2, \quad \forall \mathbf{u} \in \Omega.$$

Summing the above inequality over $k = 0, 1, \dots, t$ results in

$$\Upsilon_t \theta(y) - \sum_{k=0}^t \alpha_k \theta(\tilde{y}^k) + \left\langle \Upsilon_t \mathbf{u} - \sum_{k=0}^t \alpha_k \tilde{\mathbf{u}}^k, F(\mathbf{u}) \right\rangle + \frac{1}{2\gamma} \|\mathbf{u} - \mathbf{u}^0\|_H^2 \geq 0, \quad \forall \mathbf{u} \in \Omega.$$

It follows from the convexity of $\theta(\cdot)$ that

$$\Upsilon_t \theta(\bar{y}_t) \leq \sum_{k=0}^t \alpha_k \theta(\tilde{y}^k).$$

Then,

$$\Upsilon_t \theta(y) - \Upsilon_t \theta(\bar{y}_t) + \left\langle \Upsilon_t \mathbf{u} - \sum_{k=0}^t \alpha_k \tilde{\mathbf{u}}^k, F(\mathbf{u}) \right\rangle + \frac{1}{2\gamma} \|\mathbf{u} - \mathbf{u}^0\|_H^2 \geq 0, \quad \forall \mathbf{u} \in \Omega.$$

As $\alpha_k \geq \alpha_{\min}$, we have $\Upsilon_t \geq (t+1)\alpha_{\min}$ and

$$\frac{1}{\Upsilon_t} \leq \frac{1}{(t+1)\alpha_{\min}}.$$

Therefore,

$$\begin{aligned} & \theta(\bar{y}_t) - \theta(y) \\ & + \left\langle \frac{1}{\Upsilon_t} \sum_{k=0}^t \alpha_k \tilde{\mathbf{u}}^k - \mathbf{u}, F(\mathbf{u}) \right\rangle \leq \frac{1}{2\gamma \Upsilon_t} \|\mathbf{u} - \mathbf{u}^0\|_H^2 \leq \frac{1}{2\gamma \alpha_{\min}(t+1)} \|\mathbf{u} - \mathbf{u}^0\|_H^2, \end{aligned}$$

and the assertion of this theorem is obtained. $\square$

The foregoing theorem shows that for a given arbitrary compact set $\mathcal{D} \subset \Omega$, after t iterations of Algorithm 7.1, we obtain an approximate solution $\bar{\mathbf{u}}_t$, given by (7.21), for the mixed variational inequality (1.33) (and equivalently for the saddle point problem (1.32)) such that

$$\sup \{ \theta(\bar{y}_t) - \theta(y) + \langle \bar{\mathbf{u}}_t - \mathbf{u}, F(\mathbf{u}) \rangle \} \leq \epsilon,$$

where $\epsilon := \frac{d}{2\gamma \alpha_{\min}(t+1)}$, with $d := \sup \{ \|\mathbf{u} - \mathbf{u}^0\|_H^2 \mid \mathbf{u} \in \Omega \}$. Hence, the $O(1/t)$ convergence rate of Algorithm 7.1 is established in an ergodic sense.

7.4 Conclusions

In this chapter, we introduced a primal-dual prediction-correction algorithm for solving a saddle point optimization problem. The proposed algorithm enjoys an algorithmic

framework that bridges the gap between the approaches described in [12] and [77]. As a byproduct, we also obtain a projection-based primal-dual algorithm by choosing an appropriate proximal parameter, which is comparatively easier to implement than the methods described in [12, 15, 77], as long as the projections of both $\mathcal{X}$ and $\mathcal{Y}$ are simple enough to be computed. Furthermore, we also established that our proposed algorithm is globally convergent with a worst-case $O(1/t)$ convergence rate in an ergodic sense, which supports the numerical performance reported in [12, 77] from a theoretical perspective.

In conclusion, note that the computational experience reported in several previous works in the literature has indicated that the speed of convergence of the PDHG method is highly sensitive with respect to the choice of the values of τ and σ . Most recently, the authors [49] proposed an adaptive version of PDHG to accelerate its numerical performance. Taking a cue from this approach, we can also gainfully employ an adaptive strategy for the acceleration of our method, and the results of this research are forthcoming.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

In this dissertation, we presented four research results on augmented Lagrangian-based splitting methods for the separable convex minimization problem with linear constraints, whose objective function is separable into m individual convex functions without any coupling variables.

In our first contribution, we stated a theoretical result on the full Jacobian decomposition of ALM for solving the problem under consideration. The convergence of the directly full Jacobian decomposition method (Algorithm 3.1) was proved under the strong convexity assumptions of the involved functions.

In our second contribution, we put forward a more flexible full Jacobian decomposition method of ALM. Under some assumptions, we analyzed the sub-linear convergence rate under ergodic and nonergodic scenarios. Furthermore, we proved the global linear convergence of the new method.

In our third research thrust, a new proximal partially parallel splitting method (Algorithm 4.1) was proposed for the convex programming problem with separable structure. We established the convergence result for this new algorithm, and furthermore, the worst-case $O(1/t)$ convergence rate of the proposed method was obtained under both ergodic and nonergodic scenarios.

In our fourth research thrust, we proposed a portfolio of linearized methods for Algorithm 4.1 under the prediction-correction framework umbrella, and derived some relevant convergence results. Then, we clarified the relationships between these splitting methods

and established worst-case $O(1/t)$ convergence rate under ergodic scenario. Furthermore, we demonstrated some preliminary numerical results, which serve to demonstrate the viability of these linearized methods for solving separable convex programs.

Finally, we presented a primal-dual prediction-correction algorithm for the solving a saddle point optimization problem, which established some connections between the algorithms proposed in [12] and [77]. As a byproduct, the resulting subproblems can be simplified by selecting an appropriate proximal parameters and possess closed form solutions. Furthermore, we also analyzed the global convergence and estimated a worst-case $O(1/t)$ convergence rate under ergodic condition.

8.2 Future Work

In the future, we intend to consider some ALM-based parallel splitting methods for solving specially structured nonconvex minimization problems, where the objective function is separable into m individual nonconvex functions without any coupled variables, and the structural constraint set comprises only linear functions. One such extension would be to analyze the relevant convergence and convergence rate results when the involved functions satisfy the Kurdyka–Lojasiewicz properties. Furthermore, it is noteworthy that some nonconvex problems in the realms of matrix optimization and image processing can be solved by these splitting methods. As a special case, we can also consider some proximal point algorithms for solving *linear programming* (LP) problems. Because of the special structure of linear programs, they admit an easy-to-determine “closed-form” solution, when using some customized proximal point algorithms. Therefore, the primal and dual solutions of the LP can be easily obtained, and we can also prove the convergence results of these customized proximal point algorithms. We are currently investigating these research directions and results are forthcoming.

References

- [1] Arrow, K., Hurwicz, L., Uzawa, H.: With contributions by H.B. Chenery, S.M. Johnson, S. Karlin, T. Marschak, and R.M. Solow. Studies in Linear and Non-Linear Programming, *Stanford Mathematical Studies in the Social Science*, vol. II. Stanford University Press, Stanford, California (1958)
- [2] Auslender, A., Teboulle, M.: Lagrangian duality and related multiplier methods for variational inequality problems. *SIAM J. Optim.* **10**(4), 1097–1115 (2000)
- [3] Auslender, A., Teboulle, M.: Entropic proximal decomposition methods for convex programs and variational inequalities. *Math. Program.* **91**(1), 33–47 (2001)
- [4] Bardsley, J., Knepper, S., Nagy, J.: Structured linear algebra problems in adaptive optics imaging. *Adv. Comput. Math.* **35**, 103–117 (2011)
- [5] Bauschke, H., Combettes, P.: Convex analysis and monotone operator theory in Hilbert spaces. Springer (2011)
- [6] Bertsekas, D.: Constrained optimization and Lagrangian multiplier methods. Academic Press (1982)
- [7] Bioucas-Dias, J.M., Figueiredo, M.A.: Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing. In: *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. IEEE (2010)
- [8] Boley, D.: Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs. *SIAM J. Optim.* **23**(4), 2183–2207 (2013)

- [9] Bose, N., Boo, K.: High-resolution image reconstruction with multisensors. *Int. J. Imag. Syst. Tech.* **9**, 294–304 (1998)
- [10] Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**, 1–122 (2010)
- [11] Cai, X., Gu, G., He, B., Yuan, X.: A proximal point algorithm revisit on the alternating direction method of multipliers. *Sci. China Math.* **56(10)**, 2179–2186 (2013)
- [12] Cai, X., Han, D., Xu, L.: An improved first-order primal-dual algorithm with a new correction step. *J. Global Optim.* **57**, 1419–1428 (2013)
- [13] Cai, X., Han, D., Yuan, X.: On the convergence of the direct extension of ADMM for three-block separable convex minimization models with one strongly convex function. *Comput. Optim. Appl.* **66(1)**, 39–73 (2016)
- [14] Censor, Y., Iusem, A.N., Zenios, S.A.: An interior point method with bregman functions for the variational inequality problem with paramonotone operators. *Math. Program.* **81(3)**, 373–400 (1998)
- [15] Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**, 120–145 (2011)
- [16] Chambolle, A., Pock, T.: On the ergodic convergence rates of a first-order primal-dual algorithm. *Math. Program.* **159(1-2)**, 253–287 (2016)
- [17] Chen, C., Chan, R.H., Ma, S., Yang, J.: Inertial proximal admm for linearly constrained separable convex optimization. *SIAM J. Imaging Sci.* **8(4)**, 2239–2267 (2015)
- [18] Chen, C., He, B., Ye, Y., Yuan, X.: The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. *Math. Program.* **155**, 57–79 (2016)

- [19] Chen, C., Ma, S., Yang, J.: A general inertial proximal point algorithm for mixed variational inequality problem. *SIAM J. Optim.* **25(4)**, 2120–2142. (2015)
- [20] Chen, G., Teboulle, M.: A proximal-based decomposition method for convex minimization problems. *Math. Program.* **64**, 81–101 (1994)
- [21] Chen, L., Sun, D., Toh, K.C.: An efficient inexact symmetric Gauss-Seidel based majorized ADMM for high-dimensional convex composite conic programming. *Math. Program.* **161(1-2)**, 237–270 (2017)
- [22] Chen, Y., Lan, G., Ouyang, Y.: Optimal primal dual methods for a class of saddle point problems. *SIAM J. Optim.* **24**, 1779–1814 (2014)
- [23] Corman, E., Yuan, X.: A generalized proximal point algorithm and its convergence rate. *SIAM J. Optim.* **24(4)**, 1614–1638 (2014)
- [24] Dai, Y., Han, D., Yuan, X., Zhang, W.: A sequential updating scheme of the Lagrange multiplier for separable convex programming. *Math. Comput.* **86(303)**, 315–343 (2017)
- [25] Deng, W., Lai, M., Peng, Z., Yin, W.: Parallel multi-block ADMM with $o(1/k)$ convergence. Tech. rep., UCLA CAM 13-64 (2013)
- [26] Deng, W., Yin, W.: On the global and linear convergence of the generalized alternating direction method of multipliers. *J. Sci. Comput.* **66(3)**, 889–916 (2016)
- [27] Dinh, Q., Necoara, I., Savorgnan, C., Diehl, M.: An inexact perturbed path-following method for Lagrangian decomposition in large-scale separable convex optimization. *SIAM J. Optim.* **23**, 95–125 (2013)
- [28] Eckstein, J.: Splitting methods for monotone operators with applications to parallel optimization. Ph.D. thesis, MIT (1989)
- [29] Eckstein, J.: Some saddle-function splitting methods for convex programming. *Optim. Methods Softw.* **75–83**, 1994 (4(1))

REFERENCES

- [30] Eckstein, J., Bertsekas, D.: An alternating direction method for linear programming. Tech. rep., MIT (1990)
- [31] Eckstein, J., Bertsekas, D.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.* **55**, 293–318 (1992)
- [32] Eckstein, J., Fukushima, M.: Some reformulations and applications of the alternating direction method of multipliers. In: *Large scale optimization*, pp. 115–134. Springer US (1994)
- [33] Eckstein, J., Yao, W.: Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives. *Pac. J. Optim.* **11**, 619–644 (2015)
- [34] Esser, E., Zhang, X., Chan, T.: A general framework for a class of first-order primal-dual algorithms for convex optimization in imaging sciences. *SIAM J. Imaging Sci.* **3**, 1015–1046 (2010)
- [35] Facchinei, F., Kanzow, C.: Generalized Nash equilibrium problems. *Ann. Oper. Res.* **175(1)**, 177–211 (2010)
- [36] Facchinei, F., Pang, J.: *Finite-Dimensional Variational Inequalities and Complementarity Problems*, vol. I and II. Springer Verlag, New York (2003)
- [37] Fadili, M., Starck, J.: Monotone operator splitting for optimization problems in sparse recovery. *IEEE ICIP* pp. 1461–1464 (2009)
- [38] Fan, J., Ke, Z.T., Liu, H., Xia, L.: QUADRO: A supervised dimension reduction method via Rayleigh quotient optimization. *Ann. Stat.* **43(4)**, 1498 (2015)
- [39] Fang, E.X., He, B., Liu, H., Yuan, X.: Generalized alternating direction method of multipliers: new theoretical insights and applications. *Math. Prog. Comp.* **7(2)**, 149–187 (2015)
- [40] Ferris, M.: Finite termination of the proximal point algorithm. *Math. Program.* **50(1)**, 359–366 (1991)

- [41] Friedman, J.H.: Regularized discriminant analysis. *J. Am. Stat. Assoc.* **84(405)**, 165–175 (1989)
- [42] Fu, X., He, B., Wang, X., Yuan, X.: Block-wise alternating direction method of multipliers with Gaussian back substitution for multiple-block convex programming. Tech. rep., Department of Mathematics, Hong Kong Baptist University (2014)
- [43] Fukushima, M.: Application of the alternating direction method of multipliers to separable convex programming problems. *Comput. Optim. Appl.* **2**, 93–111 (1992)
- [44] Gabay, D.: Applications of the method of multipliers to variational inequalities. In: M. Fortin, R. Glowinski (eds.) *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, pp. 299–331. The Netherlands (1983)
- [45] Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximations. *Comput. Math. Appl.* **2**, 16–40 (1976)
- [46] Giannessi, F., Antonino, A.: *Variational inequalities and network equilibrium problems*. Plenum Press, New York (1995)
- [47] Glowinski, R., Le Tallec, P.: *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*, vol. 9. SIAM (1989)
- [48] Glowinski, R., Marrocco, A.: Approximation par éléments finis d’ordre un et résolution par pénalisation-dualité d’une classe de problèmes non linéaires. *Rev. Fr. Inform. Rech. Opér.* **9(2)**, 41–76 (1975)
- [49] Goldstein, T., Li, M., Yuan, X., Esser, E., Baraniuk, R.: Adaptive primal-dual hybrid gradient methods for saddle point problems (2015). ArXiv:1305.0546v2
- [50] Gu, G., He, B., Yang, J.: Inexact alternating-direction-based contraction methods for separable linearly constrained convex optimization. *J. Optim. Theory Appl.* **163**, 105–129 (2014)

- [51] Gu, G., He, B., Yuan, X.: Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach. *Comput. Optim. Appl.* **59(1-2)**, 135–161 (2014)
- [52] Gu, Y., Jiang, B., D., H.: A semi-proximal-based strictly contractive Peaceman-Rachford splitting method (2015). Manuscript.
- [53] Güler, O.: On the convergence of the proximal point algorithm for convex minimization. *SIAM J. Optim.* **29(2)**, 403–419 (1991)
- [54] Ha, C.D.: A generalization of the proximal point algorithm. *SIAM J. Control Optim.* **28(3)**, 503–512 (1990)
- [55] Han, D., He, H., Xu, L.: A proximal parallel splitting method for minimizing sum of convex functions with linear constraints. *J. Comput. Appl. Math.* **256**, 36–51 (2014)
- [56] Han, D., Kong, W., Zhang, W.: A partial splitting augmented Lagrangian method for low patch-rank image decomposition. *J. Math. Imaging Vis* **51(1)**, 145–160 (2015)
- [57] Han, D., Xu, W., Yang, H.: An operator splitting method for variational inequalities with partially unknown mappings. *Numer. Math.* **111**, 207–237 (2008)
- [58] Han, D., Yuan, X.: Convergence analysis of the Peaceman-Rachford splitting method for nonsmooth convex optimization (2012). Manuscript
- [59] Han, D., Yuan, X.: A note on the alternating direction method of multipliers. *J. Optim. Theory Appl.* **155**, 227–238 (2012)
- [60] Han, D., Yuan, X.: Local linear convergence of the alternating direction method of multipliers for quadratic programs. *SIAM J. Numer. Anal.* **51(6)**, 3446–3457 (2013)
- [61] Han, D., Yuan, X., Zhang, W.: An augmented Lagrangian based parallel splitting method for separable convex minimization with applications to image processing. *Math. Comput.* **83**, 2263–2291 (2014)

- [62] Han, D., Yuan, X., Zhang, W., Cai, X.: An ADM-based splitting method for separable convex programming. *Comput. Optim. Appl.* **54(2)**, 343–369 (2013)
- [63] He, B.: A class of projection and contraction methods for monotone variational inequalities. *Appl. Math. Optim.* **35(1)**, 79–114 (1997)
- [64] He, B.: Parallel splitting augmented lagrangian methods for monotone structured variational inequalities. *Comput. Optim. Appl.* **42(2)**, 195–212 (2009)
- [65] He, B.: PPA-like contraction methods for convex optimization: a framework using variational inequality approach. *Journal of the Operations Research Society of China* **3(4)**, 391–420 (2015)
- [66] He, B., Fu, X., Jiang, Z.: Proximal-point algorithm using a linear proximal term. *J. Optim. Theory Appl.* **141**, 299–319 (2009)
- [67] He, B., Hou, L., Yuan, X.: On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming. *SIAM J. Optim.* **25(4)**, 2274–2312 (2015)
- [68] He, B., Liu, H., Lu, J., Yuan, X.: Application of the strictly contractive peaceman-rachford splitting method to multi-block separable convex programming. In: *Splitting Methods in Communication, Imaging, Science, and Engineering*, pp. 195–235. Springer International Publishing (2016)
- [69] He, B., Liu, H., Wang, Z., Yuan, X.: A strictly contractive Peaceman–Rachford splitting method for convex programming. *SIAM J. Optim.* **24(3)**, 1011–1040 (2014)
- [70] He, B., Peng, Z., Wang, X.: Proximal alternating direction-based contraction methods for separable linearly constrained convex optimization. *Front. Math. China* **6(1)**, 79–114 (2011)
- [71] He, B., Tao, M., Xu, M., Yuan, X.: Alternating directions based contraction method for generally separable linearly constrained convex programming problems. *Optimization* **62**, 573–596 (2013)

- [72] He, B., Tao, M., Yuan, X.: Alternating direction method with Gaussian-back substitution for separable convex programming. *SIAM J. Optim.* **22**, 313–340 (2012)
- [73] He, B., Tao, M., Yuan, X.: A splitting method for separable convex programming. *IMA J. Numer. Anal.* **35**(1), 394–426 (2015)
- [74] He, B., Tao, M., Yuan, X.: Convergence rate analysis for the alternating direction method of multipliers with a substitution procedure for separable convex programming. *Math. Oper. Res.* (2017). DOI <http://dx.doi.org/10.1287/moor.2016.0822>
- [75] He, B., Xu, H., Yuan, X.: On the proximal Jacobian decomposition of ALM for multiple-block separable convex minimization problems and its relationship to ADMM. *J. Sci. Comput.* **66**(3), 1204–1217 (2016)
- [76] He, B., Xu, M., Yuan, X.: Block-wise ADMM with a relaxation factor for multi-block convex programming. Tech. rep., Nanjing University (2014). Preprint
- [77] He, B., Yuan, X.: Convergence analysis of primal-dual algorithms for a saddle-point problem: form contraction perspective. *SIAM J. Imaging Sci.* **5**, 119–149 (2012)
- [78] He, B., Yuan, X.: On the $O(1/n)$ convergence rate of Douglas-Rachford alternating direction method. *SIAM J. Numer. Anal.* **50**, 700–709 (2012)
- [79] He, B., Yuan, X.: The unified framework of some proximal-based decomposition methods for monotone variational inequalities with separable structure. *Pac. J. Optim.* **8**(4), 817–844 (2012)
- [80] He, B., Yuan, X.: Block-wise alternating direction method of multipliers for multiple-block convex programming and beyond (2014). Preprint
- [81] He, B., Yuan, X.: On the direct extension of ADMM for multi-block separable convex programming and beyond: from variational inequality perspective. (2014). Preprint
- [82] He, B., Yuan, X.: On nonergodic convergence rate of Douglas-Rachford alternating direction method of multipliers. *Numer. Math.* **130**, 567–577 (2015)

- [83] He, B., Yuan, X., Zhang, W.: A customized proximal point algorithm for convex minimization with linear constraints. *Comput. Optim. A* **56**, 559–572 (2013)
- [84] He, B.S., You, Y., Yuan, X.M.: On the convergence of primal-dual hybrid gradient algorithm. *SIAM J. Imaging Sci.* **7**, 2526–2537 (2015)
- [85] He, H., Han, D.: A distributed Douglas-Rachford splitting method for multi-block convex minimization problems. *Adv. Comput. Math.* **42(1)**, 27–53 (2016)
- [86] He, H., Ling, C., Xu, H.: An implementable splitting algorithm for the ℓ_1 -norm regularized split feasibility problem. *J. Sci. Comput.* **67(1)**, 281–298 (2016)
- [87] Hestenes, M.: Multiplier and gradient methods. *J. Optim. Theory Appl.* **4**, 303–320 (1969)
- [88] Hong, M., Luo, Z.Q.: On the linear convergence of the alternating direction method of multipliers. *Math. Program.* **162(1)**, 165–199 (2017)
- [89] Hou, L., He, H., Yang, J.: A partially parallel splitting method for multiple-block separable convex programming with applications to robust PCA. *Comput. Optim. Appl.* **63(1)**, 273–303 (2016)
- [90] Kiwiel, K.C., Rosa C. H. Ruszczynski, A.: Proximal decomposition via alternating linearization. *SIAM J. Optim.* **9(3)**, 668–689 (1999)
- [91] Komodakis, N., Pesquet, J.C.: Playing with duality an overview of recent primal dual approaches for solving large scale optimization problems. *IEEE Signal Process Mag.* **32(6)**, 31–54 (2015)
- [92] Konnov, I.: Partial proximal point method for nonmonotone equilibrium problems. *Optim. Method Softw.* **21(3)**, 373–384 (2006)
- [93] Li, M., Sun, D., Toh, K.C.: A convergent 3-block semi-proximal ADMM for convex minimization problems with one strongly convex block. *Asia-Pac. J. Oper. Res.* **32(4)**, 1–19 (2015). 150024

REFERENCES

- [94] Li, M., Sun, D., Toh, K.C.: A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization. *SIAM J. Optim.* **26(2)**, 922–950 (2016)
- [95] Li, X., Sun, D., Toh, K.C.: A Schur complement based semi-proximal ADMM for convex quadratic conic programming and extensions. *Math. Program.* **155(1-2)**, 333–373 (2016)
- [96] Lin, T., Ma, S., Zhang, S.: On the global linear convergence of the ADMM with multiblock variables. *SIAM J. Optim.* **25(3)**, 1478–1497 (2015)
- [97] Lin, T., Ma, S., Zhang, S.: On the sublinear convergence rate of multi-block ADMM. *Journal of the Operations Research Society of China* **3(3)**, 251–274 (2015)
- [98] Lions, P., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* **16**, 964–979 (1979)
- [99] Lu, Z., Pong, T., Zhang, Y.: An alternating direction method for finding dantzig selectors. *Comput. Stat. Data Anal.* **56(12)**, 4037–4046 (2012)
- [100] Marcotte, P., Dussault, J.: A note on a globally convergent Newton method for solving variational inequalities. *Operations Research Letters* **6**, 35–42 (1987)
- [101] Martinet, B.: Régularization d’inéquations variationnelles par approximations successives. *Rev. Fr. Inform. Rech. Opér.* **4**, 154–159 (1970)
- [102] McLachlan, G.: Discriminant analysis and statistical pattern recognition, vol. 544. John Wiley & Sons (2004)
- [103] Medhi, D., Ha, C.D.: Generalized proximal point algorithm for convex optimization. *J. Optim. Theory Appl.* **88(2)**, 475–488 (1996)
- [104] Nedić, A., Ozdaglar, A.: Subgradient methods for saddle point problems. *J. Optim. Theory Appl.* **142**, 205–228 (2009)
- [105] Nemirovski, A.: Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM J. Optim.* **15(1)**, 229–251 (2004)

REFERENCES

- [106] Nesterov, Y.: A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$. Dokl. Akad.Nauk SSSR **269(3)**, 543–547 (1983)
- [107] Nesterov, Y.: Gradient methods for minimizing composite functions. Math. Program. Ser. B. **140(1)**, 125–161 (2013)
- [108] Ng, M., Weiss, P., Yuan, X.: Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods. SIAM J. Sci. Comput. **32(5)**, 2710–2736 (2010)
- [109] Ng, M., Yuan, X., Zhang, W.: Coupled variational image decomposition and restoration model for blurred cartoon-plus-texture images with missing pixels. IEEE Trans. Image Process. **22(6)**, 2233–2246 (2013)
- [110] Oreifej, O., Li, X., Shah, M.: Simultaneous video stabilization and moving object detection in turbulence. IEEE Trans. Pattern Anal. Mach. Intell. **35**, 450–462 (2013)
- [111] Parikh, N., Boyd, S.: Proximal algorithms. Found. Trends Optim. **1**, 123–231 (2013)
- [112] Peaceman, D., Rachford, H.: The numerical solution of parabolic elliptic differential equations. Journal of the Society of Industry and Applied Mathematics **3**, 28–41 (1955)
- [113] Powell, M.: A method for nonlinear constraints in minimization problems. In: R. Fletcher (ed.) Optimization, pp. 283–298. Academic Press, New York (1969)
- [114] Qi, L., Sun, D., Zhou, G.: A new look at smoothing newton methods for nonlinear complementarity problems and box constrained variational inequalities. Math. Program. **87(1)**, 1–35 (2000)
- [115] Robinson, D.P., Tappenden, R.: A flexible ADMM algorithm for big data applications. J. Sci. Comput. **71(1)**, 435–467 (2017)
- [116] Rockafellar, R.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper. Res. **1**, 97–116 (1976)

REFERENCES

- [117] Rockafellar, R.: Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **14**, 877–898 (1976)
- [118] Setzer, S., Steidl, G., Teuber, T.: Deblurring Poissonian images by split Bregman techniques. *J. Vis. Commun. Image R.* **21**, 193–199 (2010)
- [119] Sun, D., Toh, K.C., Yang, L.: A convergent 3-block semiproximal alternating direction method of multipliers for conic programming with 4-type constraints. *SIAM J. Optim.* **25(2)**, 882–915 (2015)
- [120] Tao, M.: Some parallel splitting methods for separable convex programming with the $O(1/t)$ convergence rate. *Pac. J. Optim.* **10(2)**, 359–384 (2014)
- [121] Tao, M., Yuan, X.: Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM J. Optim.* **21**, 57–81 (2011)
- [122] Tao, M., Yuan, X.: An inexact parallel splitting augmented Lagrangian method for monotone variational inequalities with separable structures. *Comput. Optim. Appl.* **52(2)**, 439–461 (2012)
- [123] Tao, M., Yuan, X.: Convergence analysis of the direct extension of ADMM for multiple-block separable convex minimization (2016). URL <http://arxiv.org/abs/1609.07221>. Preprint.
- [124] Tian, W., Yuan, X.: Linearized primal-dual methods for linear inverse problems with total variation regularization and finite element discretization. *Inverse Problems* **32(11)**, 115,011 (2016). Manuscript
- [125] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., Knight, K.: Sparsity and smoothness via the fused LASSO. *J. R. Stat. Soc. Ser. B.* **67**, 91–108 (2005)
- [126] Tseng, P.: Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM J. Control Optim.* **29(1)**, 119–138 (1991)
- [127] Tseng, P.: On accelerated proximal gradient methods for convex-concave optimization. Tech. rep., Department of Mathematics, University of Washington (2008)

REFERENCES

- [128] Wang, K., Han, D.R., Xu, L.L.: A parallel splitting method for separable convex programs. *J. Optim. Theory Appl.* **159**(1), 138–158 (2013)
- [129] Wang, X., Hong, M., Ma, S., Luo, Z.: Solving multiple-block separable convex minimization problems using two-block alternating direction method of multipliers. *Pac. J. Optim.* **11**(4), 645–667 (2015)
- [130] Wang, X., Yuan, X.: The linearized alternating direction method of multipliers for dantzig selector. *SIAM J. Sci. Comput.* **34**(5), 2792–2811 (2012)
- [131] Wang, Y., Yang, J., Yin, W., Zhang, Y.: A new alternating minimization algorithm for total variation image reconstruction. *SIAM J. Imaging Sci.* **1**(3), 248–272 (2008)
- [132] Wu, Z., Cai, X., Han, D.: Linearized block-wise alternating direction method of multipliers for multi-block convex programming (2016). Manuscript.
- [133] Yang, J., Yuan, X.: An inexact alternating direction method for trace norm regularized least squares problem. (2010). Manuscript.
- [134] Yang, J., Yuan, X.: Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization. *Math. Comput.* **82**(2), 301–329 (2013)
- [135] Yang, J., Zhang, Y.: Alternating direction algorithms for ℓ_1 -problems in compressive sensing. *SIAM J. Sci. Comput.* **33**(2), 250–278 (2011)
- [136] Yang, W.H., Han, D.: Linear convergence of the alternating direction method of multipliers for a class of convex optimization problems. *SIAM J. Numer. Anal.* **54**(2), 625–640 (2016)
- [137] Ye, C., Yuan, X.: A descent method for structured monotone variational inequalities. *Optim. Method Softw.* **22**(2), 329–338 (2007)
- [138] Yuan, X., Yang, J.: Sparse and low rank sparse matrix decomposition via alternating directions methods. *Pac. J. Optim.* **9**(1), 167–180 (2012)
- [139] Zhang, W., Han, D., Yuan, X.: An efficient simultaneous method for the constrained multiple-sets split feasibility problem. *Comput. Optim. Appl.* **52**(3), 825–843 (2012)

REFERENCES

- [140] Zhang, X., Burger, M., Osher, S.: A unified primal-dual algorithm framework based on Bregman iteration. *J.Sci. Comput.* **6**, 20–46 (2010)
- [141] Zhu, M., Chan, T.: An efficient primal-dual hybrid gradient algorithm for total variation image restoration. *CAM Reports 08-34*, UCLA (2008)