ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 3, March 2025

The convergence of the multi-block symmetric accelerated random ADMM with a larger step size

Meng jiang

P.G. Student, Department of Mathematics, Wuhan University of Science and Technology, Wuhan, Hubei, China

ABSTRACT: The symmetric ADMM algorithm with large step sizes has been widely applied to distributed and large-scale optimization problems. Building on this foundation, accelerated stochastic versions of the algorithm have also demonstrated excellent performance. This paper explores the separable convex minimization problem, where the objective function is expressed as the sum of multiple independent functions, each depending on different sets of variables. When the standard SAS-ADMM is directly applied to such problems, convergence is not always guaranteed. To address this challenge, we propose a multi-block symmetric accelerated stochastic ADMM (LAR-ADMM) algorithm with large step sizes. The objective function is formed by combining a potentially non-smooth convex function with the mean of several smooth convex functions. In each iteration, the proposed method updates the Lagrange multipliers twice, incorporating the principles of ADMM and techniques from accelerated stochastic gradient methods. It may also employ variance reduction techniques to solve the smooth subproblems more effectively. We demonstrate the global convergence of the proposed method under certain mild conditions and provide an analysis of its convergence rate.

KEY WORDS: Convex programming, Multi-block, Separable structure, Stochastic ADMM

I.INTRODUCTION

We investigate the structured separable convex optimization problems subject to linear equality constraints as follows:

$$\min\{f(\mathbf{x}) + \sum\nolimits_{i=1}^{q} g_i(\mathbf{y}_i) \mid x \in X,\ y_i \in Y_i,\ Ax + By_i = b\}, \tag{1}$$

where $\mathcal{X} \subset \mathbb{R}^{n_1}$, $\mathcal{Y}_i \subset \mathbb{R}^{n_i}$ are closed convex subsets, $A \in \mathbb{R}^{n\times n_1}$, $B_i \in \mathbb{R}^{n\times n_i}$, $\boldsymbol{b} \in \mathbb{R}^n$ are given,
g_i: $\mathcal{Y}_i \to \mathbb{R} \cup \{+\infty\}$ is a convex function that may be non-smooth, and is the average of convex functions mapping real values：

$$f(\mathbf{x}) = \frac{1}{N}\sum_{j=1}^{N} f_j(\mathbf{x}).$$

We assume that each function f_j, defined on an open set containing X, is differentiable on X and has Lipschitz continuous derivatives. Here, N represents the sample size, and f_j refers to the empirical loss. A significant challenge in dealing with problems of the form (1) is that N can be exceedingly large, making the evaluation of f or its gradient costly at each iteration. The augmented Lagrangian of equation (1) is given by

$$L_\beta(x, y_i, \lambda) = L(x, y_i, \lambda) + \frac{\beta}{2}\left\|Ax + B_i y_i - b\right\|^2, \tag{2}$$

here, $\beta > 0$ denotes the penalty parameter, λ represents the Lagrange multiplier, and the Lagrangian of (1) is given by

$$\mathcal{L}(x, y_i, \lambda) = f(x) + g_i(y_i) - \lambda^T(Ax + B_i y_i - b). \tag{3}$$

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 3, March 2025

First, let us consider the case of two blocks, specifically the scenario where $q=2$ in Model (1). We begin by examining a classical Alternating Direction Method of Multipliers (ADMM), which leverages separable structures in its algorithm design. This method [1] has received increasing attention [2] across various fields in recent years. The algorithm follows an iterative procedure described as follows:

$$\begin{cases} \mathbf{x}^{k+1} = \arg\min_{\mathbf{x}\in\mathcal{X}} \mathcal{L}_\beta(\mathbf{x},\mathbf{y}^k,\boldsymbol{\lambda}^k), \\ \mathbf{y}^{k+1} = \arg\min_{\mathbf{y}\in\mathcal{Y}} \mathcal{L}_\beta(\mathbf{x}^{k+1},\mathbf{y},\boldsymbol{\lambda}^k), \\ \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \beta(A\mathbf{x}^{k+1} + B\mathbf{y}^{k+1} - \mathbf{b}). \end{cases}$$

Here, $\lambda \in \mathbb{R}^l$ denotes the Lagrange multiplier associated with the linear constraint in (1), and $\beta > 0$ represents the penalty parameter [5].

Applying the Peaceman-Rachford Splitting Method (PRSM) to the dual formulation of (1) [6], we derive a variant of ADMM, this can be written in the following iterative form:

$$\begin{cases} \mathbf{x}^{k+1} & \in & \arg\min_{\mathbf{x}\in X} \mathrm{L}_\beta(\mathbf{x},\mathbf{y}^k,\boldsymbol{\lambda}^k), \\ \boldsymbol{\lambda}^{k+\frac{1}{2}} & = & \boldsymbol{\lambda}^k - \beta\left(A\mathbf{x}^{k+1} + B\mathbf{y}^k - \mathbf{b}\right), \\ \mathbf{y}^{k+1} & \in & \arg\min_{\mathbf{y}\in Y} \mathrm{L}_\beta(\mathbf{x}^{k+1},\mathbf{y},\boldsymbol{\lambda}^{k+\frac{1}{2}}), \\ \boldsymbol{\lambda}^{k+1} & = & \boldsymbol{\lambda}^{k+\frac{1}{2}} - \beta\left(A\mathbf{x}^{k+1} + B\mathbf{y}^{k+1} - \mathbf{b}\right). \end{cases}$$

PRSM [8] is also known as Symmetric ADMM (S-ADMM) because, in each iteration, the Lagrange multipliers are symmetrically updated twice. It is crucial to highlight that both updates of the dual variable in PRSM employ a constant step size of 1. Inspired by the concept of increasing the dual step size as discussed in [16], Gu et al. In [13], a symmetric proximal ADMM was proposed, where the dual variable is updated twice using different step sizes. Additionally, He et al. [17] proposed the following extension of the S-ADMM:

$$\begin{cases} \mathbf{x}^{k+1} & \in & \arg\min_{\mathbf{x}\in X} \mathrm{L}_\beta(\mathbf{x},\mathbf{y}^k,\boldsymbol{\lambda}^k), \\ \boldsymbol{\lambda}^{k+\frac{1}{2}} & = & \boldsymbol{\lambda}^k - \tau\beta\left(A\mathbf{x}^{k+1} + B\mathbf{y}^k - \mathbf{b}\right), \\ \mathbf{y}^{k+1} & \in & \arg\min_{\mathbf{y}\in Y} \mathrm{L}_\beta(\mathbf{x}^{k+1},\mathbf{y},\boldsymbol{\lambda}^{k+\frac{1}{2}}), \\ \boldsymbol{\lambda}^{k+1} & = & \boldsymbol{\lambda}^{k+\frac{1}{2}} - s\beta\left(A\mathbf{x}^{k+1} + B\mathbf{y}^{k+1} - \mathbf{b}\right). \end{cases}$$

(4)

For convergence, the step size pair (τ, s) [3] must fall within the following region:

$$\Delta = \{(\tau,s) \mid \tau + s > 0,\ \ \tau \le 1,\ \ -\tau^2 - s^2 - \tau s + \tau + s + 1 \ge 0\}. \qquad (5)$$

Δ represents the maximal convergence region for the dual step size [7] in the current symmetric ADMM-type algorithms.

This paper examines the use of S-ADMM [4] to solve separable convex minimization problems (1) [9], where the objective function [18] is the sum of two or more functions, each independent of coupling variables. In this scenario, when extending the shrinkage of symmetric ADMM [19] (4) to directly address problem (1), the following iterative scheme is derived [12]:

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 3, March 2025

$$\begin{cases} x^{k+1} \in \operatorname{argmin} f(x) + \frac{\beta}{2}|Ax + \sum_{i=1}^{q} B_i y_i^k - b - \frac{\lambda^k}{\beta}|^2, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \tau\beta(Ax^{k+1} + B_i y_i^k - b), \\ \text{For}\, i = 1, 2, \cdots, q\ , \\ y_i^{k+1} \in \operatorname{argmin}_{y_i \in Y_i} g_i(y_i) + \frac{\beta}{2}|Ax^{k+1} + B_i y_i + \sum_{l \neq i, l=1}^{q} B_l y_l^k - b - \frac{\lambda^{k+\frac{1}{2}}}{\beta}|^2 + \frac{1}{2}|y_i - y_i^k|_{L_i}^2, \\ end \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(Ax^{k+1} + B_i y_i^{k+1} - b). \end{cases} \tag{6}$$

Please note that in the standard deterministic S-ADMM for (1), the gradient method [15] is commonly used to address the subproblem associated with f . As a result, evaluating the gradient of f at each iteration requires calculating the gradients of the individual component functions f_i. In large-scale data applications [1,14], computing the full gradient at each iteration can become too computationally expensive when N is large. Nevertheless, the stochastic gradient [19] can be utilized to develop a stochastic gradient-based approach, allowing for the fast yet inexact solution of $\boldsymbol{xsub}$ [9,10,11].

Building on the previous analysis and observations, we propose a multi-block accelerated stochastic S-ADMM splitting method [20] to tackle the structural empirical risk minimization problem (1).

II. PRELIMINARIES

2.1 Notations and assumptions

Let $\mathcal{R}$ denote the set of real numbers, $\mathcal{R}^n$ the space of n-dimensional real column vectors, and $\mathcal{R}^{n\times m}$ the set of real $n\times m$ matrices. The symbols $\boldsymbol{I}$ and $\mathbf{0}$ represent the identity matrix and the zero matrix/vector, respectively. For symmetric matrices A and B_i of the same dimension, the notation $A \succ B_i$ (or $A \succcurlyeq B_i$) indicates that the matrix $A - B_i$ is positive definite (or semidefinite). The symbol $\|\cdot\|$ represents the Euclidean norm associated with the inner product $\langle\cdot,\cdot\rangle$, while $\nabla f(\mathbf{x})$ denotes the gradient of the function f at the point $\mathbf{x}$. Additionally, $\mathrm{E}[\cdot]$ signifies the expected value of a random variable. let us denote

$$w = \begin{pmatrix} x \\ y_1 \\ \vdots \\ y_q \\ \lambda \end{pmatrix}, \quad J(w) = \begin{pmatrix} -A^T\lambda \\ -{B_1}^T\lambda \\ \vdots \\ -{B_q}^T\lambda \\ Ax + B_q y_q - b \end{pmatrix}, \tag{7}$$

And

$$w^k = \begin{pmatrix} x^k \\ y_1^k \\ \vdots \\ y_q^k \\ \lambda^k \end{pmatrix}, \quad J(w^k) = \begin{pmatrix} -A^T\lambda^k \\ -{B_1}^T\lambda^k \\ \vdots \\ -{B_q}^T\lambda^k \\ Ax^k + B_q {y_q}^k - b \end{pmatrix}. \tag{8}$$

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 3, March 2025

We also define

$$F(w) = f(x) + g_i(y_i).$$

We posit the following dual hypotheses:
Assumption 1. The solution set Ω^* of the primal-dual problem (1) is nonempty.
Assumption 2. For any positive definite matrix $\mathcal{H} \succ 0$, there exists a constant $\nu > 0$ such that the gradients ∇f_j satisfy the general Lipschitz condition.

$$\| \nabla f_j(\mathbf{x}_1) - \nabla f_j(\mathbf{x}_2) \|_{\mathrm{H}^{-1}}, \nu \| \mathbf{x}_1 - \mathbf{x}_2 \|_{\mathrm{H}}. \tag{9}$$

for every $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$ and $j = 1, 2, \cdots, N$.

2.2Variational characterization of (1)
Let $\Omega = \mathcal{X} \times \mathcal{Y}_i \times \mathbb{R}^n$. It is widely recognized that a point

$$\mathbf{w}^* := \begin{pmatrix} \mathbf{u}^* \\ \lambda^* \end{pmatrix} = \begin{pmatrix} \mathbf{x}^* \\ \mathbf{y}_1^* \\ \vdots \\ \mathbf{y}_q^* \\ \lambda^* \end{pmatrix} \in \Omega := X \times Y_i \times {}^n.$$

which is referred to as the saddle-point of $L(\mathbf{x}, \mathbf{y}_i, \lambda)$, if it satisfies the following inequalities

$$\begin{aligned} & \mathcal{L}(\mathbf{x}^*, \quad \mathbf{y}_1^*, \quad \cdots \quad \mathbf{y}_i^*, \quad \boldsymbol{\lambda}) \\ \leqslant & \mathcal{L}(\mathbf{x}^*, \quad \mathbf{y}_1^*, \quad \cdots \quad \mathbf{y}_i^*, \quad \boldsymbol{\lambda}^*) \\ \leqslant & \mathcal{L}(\mathbf{x}, \quad \mathbf{y}_1, \quad \cdots \quad \mathbf{y}_i, \quad \boldsymbol{\lambda}) \end{aligned} \tag{10}$$

which can be expressed as

$$\begin{cases} f(x) - f(x^*) + (x - x^*)^{\mathrm{T}}(-A^{\mathrm{T}}\lambda^*) & \geq 0, \\ g_1(y_1) - g_1(y_1^*) + (y_1 - y_q^*)^{\mathrm{T}}(-B_1^{\mathrm{T}}\lambda^*) & \geq 0, \\ \vdots & \\ g_q(y_q) - g_q(y_q^*) + (y_q - y_q^*)^{\mathrm{T}}(-B_q^{\mathrm{T}}\lambda^*) & \geq 0, \\ Ax^* + B_1 y_1^* + \cdots + B_q y_q^* - b & = 0. \end{cases} \tag{11}$$

Using the previously defined notations, these inequalities can be reformulated as follows

$$F(\boldsymbol{w}) - F(\boldsymbol{w}^*) + (\boldsymbol{w} - \boldsymbol{w}^*)^{\mathrm{T}} \mathcal{J}(\boldsymbol{w}^*) \geq 0, \quad \forall \boldsymbol{w} \in \Omega. \tag{12}$$

As the affine mapping $\mathcal{J}(\cdot)$ is skew-symmetric,

$$\left\langle \mathbf{w} - \overline{\mathbf{w}}, \mathcal{J}(\mathbf{w}) - \mathcal{J}\left(\overline{\mathbf{w}}\right) \right\rangle \equiv 0, \quad \forall \mathbf{w}, \overline{\mathbf{w}} \in \Omega. \tag{13}$$

Therefore,

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 3, March 2025

$$F(\boldsymbol{w}) - F(\boldsymbol{w}^*) + (\boldsymbol{w} - \boldsymbol{w}^*)^{\mathrm{T}} \mathcal{J}(\boldsymbol{w}) \geq 0, \quad \forall \boldsymbol{w} \in \Omega. \tag{14}$$

The previous analysis shows that the saddle point $\boldsymbol{w}^*$ can also be described by the variational inequality presented in equation (14).

III. SECURITY GAME MODELING UNDER COGNITIVE UNCERTAINTY

Building on the stochastic AS-ADMM introduced in [3], we present a multi-block Symmetric Accelerated Stochastic ADMM (LAR-ADMM), which features a dual step-size region Δ (5) similar to that of GS-ADMM.

Algorithm 3 Multi-block Large-Step Symmetric Accelerated Stochastic ADMM

Parameters: $\beta > 0$, $L_i \succeq (q-1)\beta B_i^\top B_i$ for all $i = 1, \dots, q$ and $(\tau, s) \in \Delta$ given by (5).
Initialization: $(x^0, y^0, \lambda^0) \in \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^n$, $\tilde{x}^0 = x^0$.
for $k = 0, 1, \dots$
  Choose $m_k > 0$, $\eta_k > 0$ and $\mathcal{M}_k$ such that $\mathcal{M}_k - \beta A^\top A \succeq 0$.
  $h^k := -A^\top \left[\lambda^k - \beta(Ax^k + \sum_{i=1}^{q} B_i y_i^k - b)\right]$.
  $(x^{k+1}, \tilde{x}^{k+1}) = \mathrm{xsub}(x^k, \tilde{x}^k, h^k)$.
  $\lambda^{k+\frac{1}{2}} = \lambda^k - \tau\beta(Ax^{k+1} + \sum_{i=1}^{q} B_i y_i^k - b)$.
  $y_i^{k+1} \in \arg\min_{y_i \in \mathcal{Y}} \mathcal{L}_\beta(x^{k+1}, y_i, \lambda^{k+\frac{1}{2}}) + \frac{1}{2}\left\|y_i - y_i^k\right\|_{\mathcal{L}_i}^2$
  $\lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(Ax^{k+1} + \sum_{i=1}^{q} B_i y_i^k - b)$.
end

$(x^+, \breve{x}^+) = \mathrm{xsub}(x_1, \breve{x}_1, h)$.
for $t = 1, 2, \dots, m_k$
  Randomly select $\xi_t \in \{1, 2, \dots, N\}$ with uniform probability.
  $\beta_t = \frac{2}{t+1}$, $\gamma_t = \frac{2}{t\eta_k}$, $\hat{x}_t = \beta_t \tilde{x}_t + (1 - \beta_t) x_t$.
  $d_t = \hat{g}_t + e_t$, where $\hat{g}_t = \nabla f_{\xi_t}(\hat{x}_t)$ and e_t is a random vector satisfying $\mathbb{E}[e_t] = 0$.
  $\tilde{x}_{t+1} = \arg\min_{x \in \mathcal{X}} \left\{\langle d_t + h, x\rangle + \frac{\gamma_t}{2}\|x - \tilde{x}_t\|_{\mathcal{H}}^2 + \frac{1}{2}\|x - x^k\|_{\mathcal{M}_k}^2\right\}$.
  $x_{t+1} = \beta_t \tilde{x}_{t+1} + (1 - \beta_t) x_t$.
end
Return: $(x^+, \tilde{x}^+) = (x_{m_k+1}, \tilde{x}_{m_k+1})$.

IV. CONVERGENCE ANALYSIS

To establish the convergence of Algorithm 3, we first present the following lemma, which addresses the iterates generated by the $\boldsymbol{xsub}$ routine in Algorithm 3. This lemma, originally introduced in [3], is provided here without a proof.

Lemma 4.1. [3, Lemma 3.2] Let $\boldsymbol{\delta}_t = \nabla f\left(\widehat{\boldsymbol{x}}_t\right) - \boldsymbol{d}_t$. Assume that $\eta_k \in (0, 1/\nu)$ and Assumption 2 holds. Under these conditions, the iterates produced by Algorithm 3 satisfy the following

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 3, March 2025

$$f(\mathbf{x})-f(\mathbf{x}^{k+1})+\left\langle \mathbf{x}-\mathbf{x}^{k+1},-A^{\mathsf{T}}\tilde{\boldsymbol{\lambda}}^{k}\right\rangle \geq \left\langle \mathbf{x}^{k+1}-\mathbf{x},\mathrm{D}_{k}(\mathbf{x}^{k+1}-\mathbf{x}^{k})\right\rangle +\zeta^{k}, \tag{15}$$

for all $x\in\mathcal{X}$, where

$$\tilde{\boldsymbol{\lambda}}^{k}=\boldsymbol{\lambda}^{k}-\beta\left(A\mathbf{x}^{k+1}+B\mathbf{y}^{k}-\mathbf{b}\right),\ \mathrm{D}_{k}=\mathrm{M}_{k}-\beta A^{\mathsf{T}}A, \tag{16}$$

and

$$\begin{aligned}\zeta^{k}=&\frac{2}{m_k(m_k+1)}\Big[\frac{1}{\eta_k}\left(\left\|x-\breve{x}^{k+1}\right\|_{\mathrm{H}}^{2}-\left\|x-\breve{x}^{k}\right\|_{\mathrm{H}}^{2}\right)\\&-\sum_{t=1}^{m_k}t\langle\boldsymbol{\delta}_t,\breve{x}_t-\mathbf{x}\rangle-\frac{\eta_k}{4(1-\eta_k\nu)}\sum_{t=1}^{m_k}t^{2}\left\|\boldsymbol{\delta}_t\right\|_{\mathrm{H}^{-1}}^{2}\Big].\end{aligned} \tag{17}$$

Lemma 4.2. Assume that $\eta_k\in(0,1/\nu)$. Under this condition, the iterates produced by Algorithm 3 satisfy the following properties:

$$F(\mathbf{w})-F(\tilde{\mathbf{w}}^{k})+\left\langle \mathbf{w}-\tilde{\mathbf{w}}^{k},\mathrm{J}\ \ (\mathbf{w})\right\rangle \geq(\mathbf{w}-\tilde{\mathbf{w}}^{k})^{\mathsf{T}}Q_k(\mathbf{w}^{k}-\tilde{\mathbf{w}}^{k})+\zeta^{k}, \tag{18}$$

For every $\boldsymbol{w}\in\Omega$, where ζ^{k} and $\tilde{\lambda}^{k}$ are given by (17) and (16), respectively:

$$\tilde{w}^{k}=\begin{pmatrix}\tilde{x}^{k}\\ \tilde{y}_1^{k}\\ \vdots\\ \tilde{y}_q^{k}\\ \tilde{\lambda}^{k}\end{pmatrix}=\begin{pmatrix}x^{k+1}\\ y_1^{k+1}\\ \vdots\\ y_q^{k+1}\\ \tilde{\lambda}^{k}\end{pmatrix} \tag{19}$$

and

$$Q_k=\begin{bmatrix}D_k&&&&&\\&L_1+\beta B_1^{T}B_1&&\cdots&&-\tau B_1^{T}\\&&L_2+\beta B_2^{T}B_2&&&-\tau B_2^{T}\\&\vdots&\vdots&\ddots&&\vdots\\&&&&L_q+\beta B_q^{T}B_q&-\tau B_q^{T}\\&-B_1&-B_2&\cdots&-B_q&\frac{1}{\beta}I\end{bmatrix}. \tag{20}$$

Proof: The y_i-subproblem in Algorithm 3 can be reformulated as

$$g(y_i)-g(y_i^{k+1})+\left\langle y_i-y_i^{k+1},p_{y_i}^{k}\right\rangle \geq 0,\quad \forall y_i\in Y, \tag{21}$$

where

$$\begin{aligned}p_{y_i}^{k}&=-B_i^{T}\lambda^{k+\frac{1}{2}}+\beta B_i^{T}\left(Ax^{k+1}+B_iy_i^{k+1}+\sum\nolimits_{l\neq i,l=1}^{q}B_ly_l^{k}-b\right)+L_i(y_i^{k+1}-y_i^{k})\\&=-B_i^{T}\lambda^{k+\frac{1}{2}}+\beta B_i^{T}\left(Ax^{k+1}+B_iy_i^{k}+\sum\nolimits_{l\neq i,l=1}^{q}B_ly_l^{k}-b\right)+[L_i+\beta B_i^{T}B_i](y_i^{k+1}-y_i^{k})\\&=-B_i^{T}\tilde{\lambda}^{k}+\tau B_i^{T}\left(\lambda^{k}-\tilde{\lambda}^{k}\right)+[L_i+\beta B_i^{T}B_i](y_i^{k+1}-y_i^{k}).\end{aligned}$$

(22)

The preceding equality employs the following relation

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 3, March 2025

$$\boldsymbol{\lambda}^{k+\frac{1}{2}} = \boldsymbol{\lambda}^k - \tau(\boldsymbol{\lambda}^k - \tilde{\boldsymbol{\lambda}}^k). \tag{23}$$

Based on the definition of $\tilde{\lambda}^k$, it follows that

$$\left(A\tilde{x}^k + \sum_{i=1}^{q} B_i\tilde{y}_i^k - b\right) - \sum_{i=1}^{q} B_i(\tilde{y}_i^k - y_i^k) + \frac{1}{\beta}\left(\tilde{\lambda}^k - \lambda^k\right) = \mathbf{0}. \tag{24}$$

By taking the inner product of both sides of the above equation with $\lambda - \tilde{\lambda}^k$, we derive the following result

$$\left\langle \lambda - \tilde{\lambda}^k, A\tilde{x}^k + \sum_{i=1}^{q} B_i\tilde{y}_i^k - b \right\rangle = \left\langle \lambda - \tilde{\lambda}^k, -B\left(y_i^k - \tilde{y}_i^k\right) + \frac{1}{\beta}\left(\lambda^k - \tilde{\lambda}^k\right) \right\rangle. \tag{25}$$

Thus, inequality (18) is derived by integrating (21) and (25) along with property (13), this certificate completes. In addition, combining equation (21) with the definition of $p_{y_i}^k$, we obtain

$$\begin{aligned} &g_i(y_i) - g_i(y_i^{k+1}) + \Big\langle y_i - y_i^{k+1}, -B_i^T\lambda^{k+\frac{1}{2}} \\ &+\beta B_i^T\left(Ax^{k+1} + B_i y_i^{k+1} + \sum_{l\neq i,l=1}^{q} B_l y_l^{k+1} - b\right) \\ &-\sum_{l\neq i,l=1}^{q} \beta B_i^T B_l(y_l^{k+1} - y_l^k) + L_i(y_i^{k+1} - y_i^k)\Big\rangle \geqslant 0. \end{aligned} \tag{26}$$

By summing the inequality above from $i=1$ to q, we can deduce that $\boldsymbol{y}^{k+1}$ satisfies the first-order optimality condition, thereby acting as a solution to the following problem:

$$y^{k+1} \in \arg\min_{y\in Y} g(y) + \frac{1}{2}\left\|y - y^k\right\|_{\bar{L}}^2 + \frac{\beta}{2}\left\|Ax^{k+1} + By - b - \frac{\lambda^{k+\frac{1}{2}}}{\beta}\right\|^2,$$

where

$$\bar{L} = \begin{bmatrix} L_1 & -\beta B_1^T B_2 & \cdots & -\beta B_1^T B_q \\ -\beta B_2^T B_1 & L_2 & \cdots & -\beta B_2^T B_q \\ \vdots & \vdots & \ddots & \vdots \\ -\beta B_q^T B_1 & -\beta B_q^T B_2 & \cdots & L_q \end{bmatrix}. \tag{27}$$

The following corollaries are presented to support the establishment of the main convergence theorem for Algorithm 3.

Corollary4.1. Assume that $\eta_k \in (0, 1/\nu)$. Then, the iterates generated by Algorithm 3 satisfy the following conditions:

$$\begin{aligned} &F(\mathbf{w}) - F(\tilde{\mathbf{w}}^k) + (\mathbf{w} - \tilde{\mathbf{w}}^k)^{\mathsf{T}} \mathrm{J}\ (\mathbf{w}) \\ &\geq \frac{1}{2}\left\{\left\|\mathbf{w} - \mathbf{w}^{k+1}\right\|_{\tilde{Q}_k}^2 - \left\|\mathbf{w} - \mathbf{w}^k\right\|_{\tilde{Q}_k}^2 + \left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_{\tilde{G}_k}^2\right\} + \zeta^k, \end{aligned} \tag{28}$$

where ζ^k is defined in (18), and

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 3, March 2025

$$\tilde{Q}_k = \begin{bmatrix} D_k & & & & \\ & L_1 + \left(1 - \frac{\tau s}{\tau + s}\right)\beta B_1^\top B_1 & -\frac{\tau s}{\tau + s}\beta B_1^\top B_2 & \cdots & -\frac{\tau s}{\tau + s}\beta B_1^\top B_q & -\frac{\tau}{\tau + s}B_1^\top \\ & -\frac{\tau s}{\tau + s}\beta B_2^\top B_1 & L_2 + \left(1 - \frac{\tau s}{\tau + s}\right)\beta B_2^\top B_2 & \cdots & -\frac{\tau s}{\tau + s}\beta B_2^\top B_q & -\frac{\tau}{\tau + s}B_2^\top \\ & \vdots & \vdots & \ddots & \vdots & \vdots \\ & -\frac{\tau s}{\tau + s}\beta B_q^\top B_1 & -\frac{\tau s}{\tau + s}\beta B_q^\top B_2 & \cdots & L_q + \left(1 - \frac{\tau s}{\tau + s}\right)\beta B_q^\top B_q & -\frac{\tau}{\tau + s}B_q^\top \\ & -\frac{\tau}{\tau + s}B_1 & -\frac{\tau}{\tau + s}B_2 & \cdots & -\frac{\tau}{\tau + s}B_q & \frac{1}{(\tau + s)\beta}I \end{bmatrix},$$

$$\tilde{G}_k = \begin{bmatrix} D_k & & & & \\ & L_1 + (1 - s)\beta B_1^\top B_1 & -s\beta B_1^\top B_2 & \cdots & -s\beta B_1^\top B_q & (s-1)B_1^\top \\ & -s\beta B_2^\top B_1 & L_2 + (1 - s)\beta B_2^\top B_2 & \cdots & -s\beta B_2^\top B_q & (s-1)B_2^\top \\ & \vdots & \vdots & \ddots & \vdots & \vdots \\ & -s\beta B_q^\top B_1 & -s\beta B_q^\top B_2 & \cdots & L_q + (1 - s)\beta B_q^\top B_q & (s-1)B_q^\top \\ & (s-1)B_1 & (s-1)B_2 & \cdots & (s-1)B_q & \frac{2 - \tau - s}{\beta}I \end{bmatrix}. \quad (30)$$

Proof: According to the method of generating λ^{k+1}, we obtain

$$-s\beta \sum_{i=1}^{q} B_i \left(y_i^k - \tilde{y}_i^k\right) + (\tau + s)\left(\lambda^k - \tilde{\lambda}^k\right) = \lambda^k - \lambda^{k+1}, \quad (31)$$

and by the definition of $\boldsymbol{w}^k$, we have

$$\mathbf{w}^k - \mathbf{w}^{k+1} = P(\mathbf{w}^k - \tilde{\mathbf{w}}^k) \ \ with \ \ P = \begin{bmatrix} I & & & & \\ & I & & & \\ & & \ddots & & \\ & & & I & \\ & -s\beta B_1 & \cdots & -s\beta B_q & (\tau + s)I \end{bmatrix}. \quad (32)$$

Therefore, the relation $Q_k(\mathbf{w}^k - \tilde{\mathbf{w}}^k) = Q_k P^{-1}(\mathbf{w}^k - \mathbf{w}^{k+1})$ holds and

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 3, March 2025

$$Q_k P^{-1} = \begin{bmatrix} D_k & & & & \\ L_1 + \left(1 - \frac{\tau s}{\tau + s}\right)\beta B_1^T B_1 & -\frac{\tau s}{\tau + s}\beta B_1^{B_2} & \cdots & -\frac{\tau s}{\tau + s}\beta B_1^{B_q} & -\frac{\tau}{\tau + s}B_1^T \\ -\frac{\tau s}{\tau + s}\beta B_2^T B_1 & L_2 + \left(1 - \frac{\tau s}{\tau + s}\right)\beta B_2^T B_2 & \cdots & -\frac{\tau s}{\tau + s}\beta B_2^{B_q} & -\frac{\tau}{\tau + s}B_2^T \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{\tau s}{\tau + s}\beta B_q^T B_1 & -\frac{\tau s}{\tau + s}\beta B_q^T B_2 & \cdots & L_q + \left(1 - \frac{\tau s}{\tau + s}\right)\beta B_q^T B_q & -\frac{\tau}{\tau + s}B_q^T \\ -\frac{\tau}{\tau + s}B_1 & -\frac{\tau}{\tau + s}B_2 & \cdots & -\frac{\tau}{\tau + s}B_q & \frac{1}{(\tau + s)\beta}I \end{bmatrix} = \tilde{Q}_k. \tag{33}$$

For all $\boldsymbol{w} \in \Omega$, the consequences of equation (18) and the preceding relationship imply that

$$\begin{aligned} & F(\mathbf{w}) - F(\tilde{\mathbf{w}}^k) + (\mathbf{w} - \tilde{\mathbf{w}}^k)^{\mathsf{T}} \mathrm{J}\ (\mathbf{w}) \\ & \geq \zeta^k + (\mathbf{w} - \tilde{\mathbf{w}}^k)^{\mathsf{T}} \tilde{Q}_k (\mathbf{w}^k - \mathbf{w}^{k+1}) = \zeta^k + \frac{1}{2}\left\{\left\|\mathbf{w} - \mathbf{w}^{k+1}\right\|_{\tilde{Q}_k}^2\right. \\ & \left. - \left\|\mathbf{w} - \mathbf{w}^k\right\|_{\tilde{Q}_k}^2 + \left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_{\tilde{Q}_k}^2 - \left\|\mathbf{w}^{k+1} - \tilde{\mathbf{w}}^k\right\|_{\tilde{Q}_k}^2\right\}, \end{aligned} \tag{34}$$

where the equation uses the constant

$$2(a-b)^{\mathsf{T}} \tilde{Q}_k (c-d) = \| a-d\|_{\tilde{Q}_k}^2 - \| a-c\|_{\tilde{Q}_k}^2 + \| c-b\|_{\tilde{Q}_k}^2 - \| b-d\|_{\tilde{Q}_k}^2, \tag{35}$$

and

$$a := \mathbf{w},\ b := \tilde{\mathbf{w}}^k,\ c := \mathbf{w}^k,\ d := \mathbf{w}^{k+1}.$$

Now, we again derive from (32) that

$$\begin{aligned} & \left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_{\tilde{Q}_k}^2 - \left\|\mathbf{w}^{k+1} - \tilde{\mathbf{w}}^k\right\|_{\tilde{Q}_k}^2 \\ & = \left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_{\tilde{Q}_k}^2 - \left\|\mathbf{w}^{k+1} - \mathbf{w}^k + \mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_{\tilde{Q}_k}^2 \\ & = \left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_{\tilde{Q}_k}^2 - \left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k - P(\mathbf{w}^k - \tilde{\mathbf{w}}^k)\right\|_{\tilde{Q}_k}^2 \\ & = \left\|\mathbf{w}^k - \tilde{\mathbf{w}}^k\right\|_{\tilde{G}_k}^2. \end{aligned} \tag{36}$$

This certificate completes.

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 3, March 2025

In Corollary 4.1 and its proof above, since $\tilde{Q}_k$ may not be positive semidefinite for any given parameter τ, we define the notation $|\boldsymbol{w}^k|^2_{\tilde{Q}_k} := (\boldsymbol{w}^k)^{\mathrm{T}} \widetilde{Q}_k \boldsymbol{w}^k$. We will now proceed to demonstrate the positive semi-definiteness of the matrix $\tilde{Q}_k$.

Lemma 4.3. Given that $L_i \succeq (\tau-1)\beta B_i^T B_i$ for $i=1,\ldots,q$, the matrix $\tilde{Q}_k$ defined in (5) is symmetric and positive semidefinite for all $(\tau,s)\in\Delta$.

$$\tilde{Q}_k^L = \begin{bmatrix} L_1+(1-\frac{\tau s}{\tau+s})\beta B_1^T B_1 & -\frac{\tau s}{\tau+s}\beta B_1^{B_2} & \cdots & -\frac{\tau s}{\tau+s}\beta B_1^{B_q} & -\frac{\tau}{\tau+s}B_1^T \\ -\frac{\tau s}{\tau+s}\beta B_2^T B_1 & L_2+(1-\frac{\tau s}{\tau+s})\beta B_2^T B_2 & \cdots & -\frac{\tau s}{\tau+s}\beta B_2^{B_q} & -\frac{\tau}{\tau+s}B_2^T \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{\tau s}{\tau+s}\beta B_q^T B_1 & -\frac{\tau s}{\tau+s}\beta B_q^T B_2 & \cdots & L_q+(1-\frac{\tau s}{\tau+s})\beta B_q^T B_q & -\frac{\tau}{\tau+s}B_q^T \\ -\frac{\tau}{\tau+s}B_1 & -\frac{\tau}{\tau+s}B_2 & \cdots & -\frac{\tau}{\tau+s}B_q & \frac{1}{(\tau+s)\beta}I \end{bmatrix}$$

$$\pm \begin{bmatrix} \left(\tau-\frac{\tau s}{\tau+s}\right)\beta B_1^T B_1 & -\frac{\tau s}{\tau+s}\beta B_1^{B_2} & \cdots & -\frac{\tau s}{\tau+s}\beta B_1^{B_q} & -\frac{\tau}{\tau+s}B_1^T \\ -\frac{\tau s}{\tau+s}\beta B_2^T B_1 & \left(\tau-\frac{\tau s}{\tau+s}\right)\beta B_2^T B_2 & \cdots & -\frac{\tau s}{\tau+s}\beta B_2^{B_q} & -\frac{\tau}{\tau+s}B_2^T \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{\tau s}{\tau+s}\beta B_q^T B_1 & -\frac{\tau s}{\tau+s}\beta B_q^T B_2 & \cdots & \left(\tau-\frac{\tau s}{\tau+s}\right)\beta B_q^T B_q & -\frac{\tau}{\tau+s}B_q^T \\ -\frac{\tau}{\tau+s}B_1 & -\frac{\tau}{\tau+s}B_2 & \cdots & -\frac{\tau}{\tau+s}B_q & \frac{1}{\beta(\tau+s)}I \end{bmatrix}$$

$$= \begin{bmatrix} \beta^{\frac{1}{2}}[B_i] & \\ & \beta^{-\frac{1}{2}}I \end{bmatrix}^T \begin{bmatrix} \frac{\tau^2}{\tau+s}I & \cdots & -\frac{\tau}{\tau+s}I \\ \vdots & \ddots & \vdots \\ -\frac{\tau}{\tau+s}I & \cdots & \frac{1}{\tau+s}I \end{bmatrix} \begin{bmatrix} \beta^{\frac{1}{2}}[B_i] & \\ & \beta^{-\frac{1}{2}}I \end{bmatrix}, \tag{37}$$

is semipositive,

where $B_{ii} := \begin{bmatrix} B_1 & & \\ & \ddots & \\ & & B_q \end{bmatrix}_{i\times i}$.

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 3, March 2025

Next prove the global convergence of Algorithm 3.

Lemma 4.4. Assume that $L_i \succeq 0$. Then, for any pair (τ, s) in Δ as defined in (5), the following holds

$$\begin{aligned} \left\|w^k-\tilde{w}^k\right\|_{\tilde{G}_k}^2 \geq & \left\|x^k-x^{k+1}\right\|_{D_k}^2+r_1\left\|Ax^{k+1}+\sum_{i=1}^q B_i y_i^{k+1}-b\right\|^2 \\ & +r_2\left(\left\|Ax^{k+1}+\sum_{i=1}^q B_i y_i^{k+1}-b\right\|^2-\left\|Ax^k+\sum_{i=1}^q B_i y_i^k-b\right\|^2\right) \\ & +r_3\left(\left\|y_i^k-y_i^{k+1}\right\|_{\bar{L}}^2-\left\|y_i^{k-1}-y_i^k\right\|_{\bar{L}}^2\right), \end{aligned} \tag{38}$$

where $r_i \geqslant 1, i=1,2,3$, $\bar{L}$ given by equation (27), and

$$r_1=\left(2-\tau-s-\frac{(1-s)^2}{1+\tau}\right)\beta, \quad r_2=\frac{(1-s)^2}{1+\tau}\beta \quad and \quad r_3=\frac{1-\tau}{1+\tau}. \tag{39}$$

Proof: Based on Lemma 4.5 from reference [4] and

$$\bar{L}=\begin{bmatrix} L_1 & -\beta B_1^T B_2 & \cdots & -\beta B_1^T B_q \\ -\beta B_2^T B_1 & L_2 & \cdots & -\beta B_2^T B_q \\ \vdots & \vdots & \ddots & \vdots \\ -\beta B_q^T B_1 & -\beta B_q^T B_2 & \cdots & L_q \end{bmatrix} \beta \begin{bmatrix} w_1 B_1^T B_1 & -\beta B_1^T B_2 & \cdots & -\beta B_1^T B_q \\ -\beta B_2^T B_1 & w_2 B_2^T B_2 & \cdots & -\beta B_2^T B_q \\ \vdots & \vdots & \ddots & \cdots \\ -\beta B_q^T B_1 & -\beta B_q^T B_2 & \cdots & w_q B_q^T B_q \end{bmatrix}, \tag{40}$$

given that $L_i \succeq w_l \beta B_i^T B_i$, it follows from $w_l > 0$, and $w_i \sum_{l \neq i, l=1}^{q} w_l > 0$ that inequality (38) holds.

The subsequent theorem provides a foundational convergence result for Algorithm 3.

Let us assume that for some integers $\kappa \geq 0$ and $T > 0$, the following conditions are satisfied for all $k \in [\kappa, \kappa+T]$:

(1) $\eta_k \in (0, \frac{1}{2\nu}]$ and the sequence $\{\eta_k M_k (M_k+1)\}$ is nondecreasing;

(2) we have that $\mathcal{D}_k \succeq \mathcal{D}_{k+1} \succeq \mathbf{0}$, and for some $u > 0$, it holds that $\mathbb{E}\left(|\boldsymbol{\delta}_t|_{\mathcal{H}^{-1}}^2\right) \leq u^2$.

Then for every $\mathbf{w} \in \Omega$, we have

$$\begin{aligned} & \left[F(w_T)-F(w)+(w_T-w)^T J(w)\right] \\ & \leq \frac{1}{2T}\left\{\sigma^2 \sum_{k=\kappa}^{\kappa+T} \eta_k M_k+\frac{4}{M_\kappa(M_\kappa+1)\eta_\kappa}\left\|x-\breve{x}^\kappa\right\|_H^2+\left\|w-w^\kappa\right\|_{\tilde{Q}_\kappa}^2\right. \\ & \left.+r_2\left\|Ax^\kappa+\sum_{i=1}^q B_i y_i^\kappa-b\right\|^2+r_3\left\|y_i^{\kappa-1}-y_i^\kappa\right\|_{\bar{L}}^2\right\}, \end{aligned} \tag{41}$$

where

$$\boldsymbol{w}_T=\frac{1}{T}\sum_{k=\kappa}^{\kappa+T} \boldsymbol{w}^k, \ \omega_2 \geq 0 \ and \ \omega_3 \geq 0.$$

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 3, March 2025

Proof: Assume $\mathcal{D}_k \succeq \mathcal{D}_{k+1} \succeq \mathbf{0}$, it follows that $\widetilde{Q}_k \succeq \widetilde{Q}_{k+1} \succeq \mathbf{0}$. By substituting equation (38) into (28) and applying the relation $\widetilde{Q}_k \succeq \widetilde{Q}_{k+1}$, Lemma 4.5 leads to the conclusion that

$$
\begin{aligned}
& F(\tilde{w}^k) - F(w) + (\tilde{w}^k - w)^T J(w) \leq -\zeta^k \\
& + \frac{1}{2}\left\{ \left\| w - w^k \right\|_{\tilde{Q}_k}^2 - \left\| w - w^{k+1} \right\|_{\tilde{Q}_{k+1}}^2 \right\} \\
& + \frac{r_2}{2}\left(\left\| Ax^k + \sum_{i=1}^{q} B_i y_i^k - b \right\|^2 - \left\| Ax^{k+1} + \sum_{i=1}^{q} B_i y_i^{k+1} - b \right\|^2 \right) \\
& + \frac{r_3}{2}\left(\left\| y_i^{k-1} - y_i^k \right\|_{\bar{L}}^2 - \left\| y_i^k - y_i^{k+1} \right\|_{\bar{L}}^2 \right),
\end{aligned} \tag{42}
$$

summing the above inequality over k in the interval $[\kappa, \kappa + T]$:

$$
\begin{aligned}
& \sum_{k=\kappa}^{\kappa+T} F(\tilde{w}^k) - T\left\{ F(w) + (w_T - w)^T J(w) \right\} \leq -\sum_{k=\kappa}^{\kappa+T} \zeta^k \\
& + \frac{1}{2}\left\{ \left\| w - w^k \right\|_{\tilde{Q}_k}^2 + r_2 \left\| Ax^\kappa + \sum_{i=1}^{q} B_i y_i^\kappa - b \right\|^2 + r_3 \left\| y_i^{\kappa-1} - y_i^\kappa \right\|_{\bar{L}}^2 \right\}.
\end{aligned} \tag{43}
$$

This conclusion is derived from the convexity of F and the definition of $\mathbf{W}_T$.

$$
F(\mathbf{w}_T) \leq \frac{1}{T} \sum_{k=\kappa}^{\kappa+T} F(\tilde{\mathbf{w}}^k). \tag{44}
$$

By dividing (43) by T and applying (44), we derive

$$
\begin{aligned}
& F(w_T) - F(w) + (w_T - w)^T J(w) \leq \frac{1}{T}\left[-\sum_{k=\kappa}^{\kappa+T} \zeta^k \right. \\
& \left. + \frac{1}{2}\left\{ \left\| w - w^k \right\|_{\tilde{Q}_k}^2 + r_2 \left\| Ax^\kappa + \sum_{i=1}^{q} B_i y_i^\kappa - b \right\|^2 + r_3 \left\| y_i^{\kappa-1} - y_i^\kappa \right\|_{\bar{L}}^2 \right\} \right].
\end{aligned} \tag{45}
$$

We will now focus on the terms involving $-\sum_{k=\kappa}^{\kappa+T} \zeta^k$ and proceed to evaluate its expected value. The sequence $\{M_k(M_k+1)\eta_k\}$ is non-decreasing for $k \in [\kappa, \kappa+T]$, and since $\mathcal{H} \succ \mathbf{0}$, we obtain the following result:

$$
\sum_{k=\kappa}^{\kappa+T} \frac{2}{M_k(M_k+1)\eta_k}\left(\left\| \mathbf{x} - \breve{\mathbf{x}}^k \right\|_{\mathrm{H}}^2 - \left\| \mathbf{x} - \breve{\mathbf{x}}^{k+1} \right\|_{\mathrm{H}}^2 \right),, \frac{2\left\| \mathbf{x} - \breve{\mathbf{x}}^\kappa \right\|_{\mathrm{H}}^2}{M_\kappa(M_\kappa+1)\eta_\kappa},
$$

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 3, March 2025

where $\delta_t = \nabla f(\mathbf{x}_t) - \mathbf{d}_t = \nabla f(\mathbf{x}_t) - \nabla f_{\xi_t}(\mathbf{x}_t) - \mathbf{e}_t$,

the quantity depends solely on the index ξ_t. As a result, since the random variable $\xi_t \in \{1, 2, ..., N\}$ is chosen with equal probability, it follows that $\mathbb{E}[\boldsymbol{\delta}_t] = \boldsymbol{0}$, which also implies $\mathbb{E}[\mathbf{e}_t] = \mathbf{0}$. Moreover, the iterate $\breve{\mathbf{x}}_t$, which depends on $\xi_{t-1}, \xi_{t-2}, ...,$, results in $\mathbb{E}[\langle \boldsymbol{\delta}_t, \breve{\mathbf{x}}_t - \mathbf{x} \rangle] = \mathbf{0}$. Coupled with the assumption $\mathbb{E}(|\boldsymbol{\delta}_t|^2_{\mathcal{H}^{-1}}) \leqslant u^2$, it follows that

$$\mathrm{E}\left[\sum_{t=1}^{M_k} t^2 \left\|\delta_t\right\|^2_{\mathrm{H}^{-1}}\right] ,, \frac{u^2 M_k (M_k+1)(2M_k+1)}{6} ,, \frac{u^2}{2} M_k^2 (M_k+1).$$

Given that $M_k \geqslant 1$ and based on the definition of ζ^k in equation (14), along with the previously mentioned inequality and the assumption $\eta_k \leqslant 1/(2\nu)$, we derive the following result

$$-\mathrm{E}\left[\sum_{k=\kappa}^{\kappa+T} \zeta^k\right] ,, \frac{2}{M_\kappa (M_\kappa+1)\eta_\kappa} \left\|\mathbf{x} - \breve{\mathbf{x}}^\kappa\right\|^2_{\mathrm{H}} + \frac{u^2}{2} \sum_{k=\kappa}^{\kappa+T} \eta_k M_k. \tag{46}$$

The proof is completed by taking the expectation of both sides of equation (45) and utilizing the result from equation (46).

We will now analyse the convergence rate of Algorithm 3:

$$\begin{aligned} \lambda_T &= \frac{1}{1+T} \sum_{k=\kappa}^{\kappa+T} \tilde{\lambda}^k, \quad \mathbf{x}_T = \frac{1}{1+T} \sum_{k=\kappa}^{\kappa+T} \tilde{\mathbf{x}}^k, \\ (\mathbf{y}_i)_T &= \frac{1}{1+T} \sum_{k=\kappa}^{\kappa+T} \tilde{\mathbf{y}}_i^k. \end{aligned} \tag{47}$$

By selecting appropriate values for η_k and m_k as outlined in (41), we can achieve the following results:

$$\eta_k = \min\left\{\frac{u_1}{M_k (M_k+1)}, c_2\right\} \quad \text{and} \quad M_k = \max\left\{\left\lceil u_3 k^{\text{ñ}} \right\rceil, M\right\}. \tag{48}$$

Considering that $u_1, u_2, u_3 > 0$, $\varrho \geqslant 1$ are constants, and $M > 0$ is a fixed integer, Theorem 3 directly implies that

$$\mathrm{E}\left[F(\mathbf{u}_T) - F(\mathbf{u}) + (\mathbf{w}_T - \mathbf{w})^{\boldsymbol{\cdot}} \,\mathrm{J}\; (\mathbf{w})\right] ,, \mathrm{O}\left(\frac{1}{T}\left(1 + \sum_{k=\kappa}^{\kappa+T} \frac{1}{k^{\text{ñ}}}\right)\right).$$

In practice, users can choose a sufficiently large value of k such that $M_k = \lceil u_3 k^{\varrho} \rceil$. As k increases, both M_k and η_k approach infinity and zero, respectively. To satisfy the condition $\eta_k = \dfrac{c_1}{M_k(M_k+1)} \leqslant \dfrac{1}{2\nu}$ in condition (1), it is feasible to choose a sufficiently large k if necessary. Given that $\eta_k M_k (M_k+1) = u_1$ is constant and $\eta_k \in (0, \dfrac{1}{2\nu}]$ for $k \geqslant \kappa$, condition (1) is satisfied for this choice of k. As a result, we obtain the following convergence theorem, analogous to the proof in [2, Theorem 4.2].

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 3, March 2025

Assuming the conditions in Theorem 1 are met, and with η_k and m_k selected as per equation (48), it follows that for all $\mathbf{w}^* = (\mathbf{u}^*, \boldsymbol{\lambda}^*) \in \Omega^*$, it follows that

$$\left|\mathrm{E}\left[F(\mathbf{u}_T) - F(\mathbf{u}^*)\right]\right| = E_{\tilde{n}}(T) = E[\Box A\mathbf{x}_T + B\mathbf{y}_T + C\mathbf{z}_T - \mathbf{b}\Box] \tag{49}$$

for $\varrho > 1$, $E_\varrho(T) = \mathcal{O}(1/T)$, while for $\varrho = 1$, $E_\varrho(T) = \mathcal{O}(T^{-1}\log T)$.

V. NUMERICAL EXPERIMENTS

In this section, we conduct numerical experiments to evaluate the performance of the proposed multi-block extension of the accelerated stochastic ADMM algorithm. Specifically, we consider the graph-guided fused lasso problem, a representative sparse optimization problem widely applied in machine learning for handling tasks with graph-structured sparsity patterns. Given feature-label pairs $(a_j, b_j) \in \mathbb{R}^l \times \{-1, 1\}$, where N denotes the number of samples ($N > l$), the graph-guided fused lasso problem can be formulated as:

$$\min_x \frac{1}{N}\sum_{j=1}^{N} f_j(x) + \mu|Ax|_1,$$

where $f_j(x)$ represents the logistic loss function:

$$f_j(x) = \log(1 + \exp(-b_j a_j^\top x)),$$

and $\mu > 0$ is the regularization parameter controlling sparsity. The matrix $A = [G; I]$ encodes the graph structure G, derived from sparse inverse covariance estimation, combined with an identity matrix I.

To exploit the efficiency of the ADMM framework for separable structures, we introduce an auxiliary variable y, reformulating the above problem as:

$$\min_{x,y} F(x,y) = \frac{1}{N}\sum_{j=1}^{N} f_j(x) + \mu\|y\|_1, \quad \text{s.t. } Ax - y = 0. \tag{50}$$

The reformulated problem allows for efficient computation by leveraging the separable structure. Specifically, the subproblem related to y benefits from a closed-form solution when the coefficient matrix associated with y in the constraints is $-I$. In this case, setting $L = 0$ in the proposed multi-block stochastic ADMM framework simplifies the y-update step. With $L = 0$, the subproblems in the multi-block framework have the following updates:

$$\begin{aligned} \tilde{x}_{t+1} &= \left[\gamma_t \mathrm{H} + M_k\right]^{-1}\left[\gamma_t \mathrm{H}\tilde{x}_t + M_k x^k - d_t - h^k\right], \\ y^{k+1} &= \mathrm{Shrink}\left(\frac{\mu}{\beta}, Ax^{k+1} - \frac{\lambda^{k+1/2}}{\beta}\right), \end{aligned} \tag{51}$$

where the operator $\mathrm{Shrink}(\cdot,\cdot)$ applies soft-thresholding, which can be efficiently computed using available numerical libraries.

In the numerical experiments, the penalty parameter β is set to 0.001, while the adaptive matrix updates M_k follow the dynamic adjustment strategy outlined in the literature. Specifically:

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 12, Issue 3, March 2025

Initial values are configured as $\rho_0 = 1$, $\rho_{\min} = 10^{-5}$. The matrix $\mathcal{H}$ is set to $2 \times 10^{-5} I$ to maintain stability and enhance convergence.

To address stochastic variations, the vector e_t introduced for variance reduction is configured as:

$$e_t = \begin{cases} \nabla f(x_{k-1}) - \nabla f_{\xi_t}(x_{k-1}), & \text{if } m_k > l, \\ 0, & \text{otherwise,} \end{cases} \tag{52}$$

where x_k represents the ergodic mean of the x_k-iterates. This design ensures effective gradient estimation while balancing computational cost.
Following Theorem 1, the performance is evaluated using two primary metrics:

$$\text{Obj.err} = \frac{|F(x,y) - F^*|}{\max\{F^*, 1\}}, \quad \text{Equ.err} = \| Ax - y \|. \tag{53}$$

The optimization error combines these two components to provide a comprehensive measure of the algorithm's convergence behaviour under varying conditions.
To evaluate the effectiveness of the proposed algorithm, we measure the maximum of the relative objective value error and the constraint violation error, defined as:

$$\text{Opt\_err} = \max(\text{Obj\_err}, \text{Equ\_err}), \tag{54}$$

where Obj_err represents the relative difference between the objective value $F(x,y)$ and its approximate optimal value F^* (computed using prolonged runs of Algorithm 3), and Equ_err measures the norm of constraint violations $\| Ax - y \|$. This unified metric provides a comprehensive assessment of both the objective accuracy and the feasibility of solutions.
The starting point for all experiments is initialized as $(x^0, y^0, \lambda^0) = (0, 0, 0)$.
The dataset used is the widely studied `mnist` dataset, which includes 11,791 samples and 784 features (i.e., $N = 11,791$, $l = 784$). It was obtained from the LIBSVM repository, and the regularization parameter μ in problem (50) is set to 10^{-5}. We conducted multiple consecutive runs for each algorithm under CPU time budgets of 20 s, 50 s, 100 s, and 200 s, respectively. The average comparative results are illustrated in the following figures, where the horizontal axis represents CPU time and the vertical axis denotes the optimization error $Opt_e rr$.

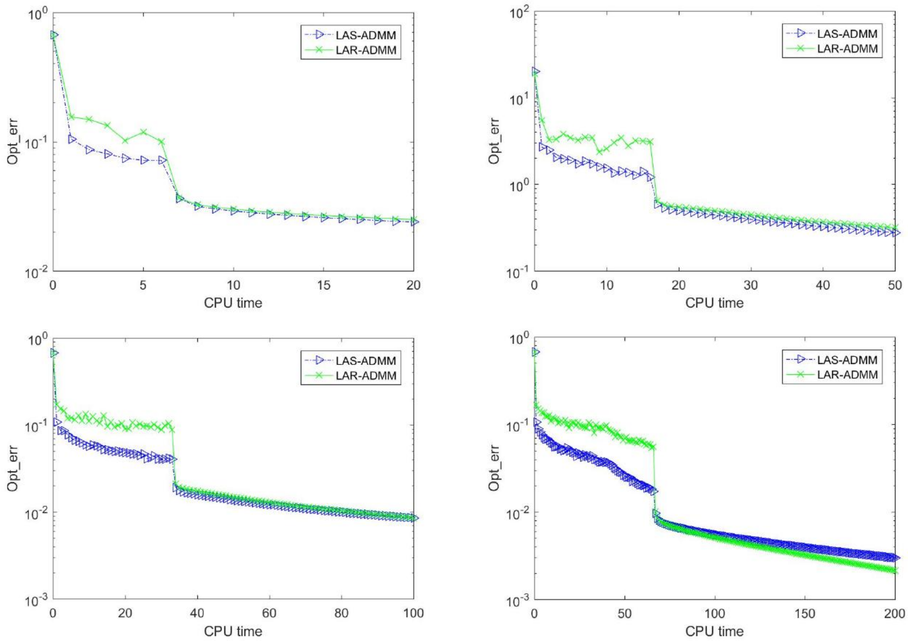

Fig. 1: Comparison of optimization error versus CPU time for solving problem (50) on the MNIST dataset

From the numerical experimental results presented in the figure, LAS-ADMM and LAR-ADMM exhibit different convergence behaviours under varying CPU time budgets.

1. **Convergence Speed and Stability of Optimization Error:** In the initial stages, LAS-ADMM shows a faster decrease in optimization error, especially within the short time frame, demonstrating superior early-stage performance. However, LAR-ADMM exhibits better stability over extended time periods, with its error curve becoming smoother as the optimization progresses. This suggests that although LAR-ADMM may have a slower initial convergence rate, its dynamic adjustment mechanism provides greater robustness in long-term optimization, ensuring stable progress even in the face of complex tasks.

2. **Advantages of the Dynamic Adjustment Mechanism:** LAR-ADMM'S dynamic adjustment mechanism is particularly beneficial for more complex optimization tasks. For long-duration optimization processes or intricate problems, LAR-ADMM effectively addresses fluctuations in the error through step size adjustments, ensuring stable results in large-scale problems. Notably, under extended time budgets (such as 100 s or 200 s), LAR-ADMM's optimization error stabilizes, demonstrating strong global convergence and robustness.

3. **Global Convergence:** While LAR-ADMM may exhibit slower convergence in the early and middle phases, it ultimately reaches a comparable optimization error level to that of LAS-ADMM. This indicates that LAR-ADMM excels in global convergence. Especially for complex or large-scale problems, LAR-ADMM ensures stability in convergence to near-optimal solutions, highlighting its advantage in handling challenging optimization tasks.

4. **Applicable Scenarios:** LAR-ADMM'S primary advantage lies in its suitability for larger-scale optimization problems, particularly those requiring longer optimization durations. Its dynamic adjustment mechanism enables it to effectively handle more complex tasks, ensuring stability over time. In contrast, LAS-ADMM is more suited for problems requiring rapid convergence within shorter time frames. In practical applications, LAR-ADMM is the preferred choice for optimization tasks that demand higher global convergence and stability.

5. **Final Performance:** Even with longer CPU time budgets (such as 100 s or 200 s), LAR-ADMM maintains a stable

optimization error at a low level. Although its error decreases more slowly in the initial stages, its stability and global convergence in the later stages ensure that it ultimately reaches comparable results to LAS-ADMM. This demonstrates that LAR-ADMM is better suited for large-scale optimization tasks or those requiring long-term optimization, providing reliable results while maintaining robustness and stability.

VI. CONCLUSION

LAR-ADMM exhibits notable advantages in terms of stability and global convergence over extended time horizons. While its convergence rate in the initial and intermediate phases is slower compared to LAS-ADMM, it ultimately achieves a comparable level of optimization error, making it well-suited for large-scale or long-duration optimization tasks. The dynamic adjustment mechanism of LAR-ADMM ensures robust performance in complex scenarios where stability is a priority.

REFERENCES

[1] Chen Y, Hager W W, Yashtini M, et al. Bregman operator splitting with variable stepsize for total variation image reconstruction. Computational optimization and applications, 2013, 54(2): 317-342.
[2]Deng W, Lai M J, Peng Z, et al. Parallel multi-block ADMM with o(1/k) convergence. Journal of Scientific Computing, 2017, 71: 712-736.
[3]Hou L, He H, Yang J. A partially parallel splitting method for multiple-block separable convex programming with applications to robust PCA. Computational Optimization and Applications, 2016, 63: 273-303.
[4]Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine learning, 2011, 3(1): 1-122.
[5] Eckstein J, Bertsekas D P. On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators. Mathematical programming, 1992, 55: 293-318.
[6]Shen Y, Zhang X, Zhang X. A partial PPA block-wise ADMM for multi-block linearly constrained separable convex optimization. Optimization, 2021, 70(3): 631-657.
[7]Fortin M, Glowinski R. Augmented Lagrangian methods, volume 15 of Studies in Mathematics and its Applications. 1983.
[8]Kim S, Xing E P. Statistical estimation of correlated genome associations to a quantitative trait network. PLoS genetics, 2009, 5(8): e1000587.
[9]Zhu Y N, Zhang X. Stochastic primal dual fixed point method for composite optimization. Journal of Scientific Computing, 2020, 84(1): 16.
[10]Esser E. Applications of Lagrangian-based alternating direction methods and connections to split Bregman. CAM report, 2009, 9: 31.
[11]Tao M, Yuan X. Recovering low-rank and sparse components of matrices from incomplete and noisy observations. SIAM Journal on Optimization, 2011, 21(1): 57-81.
[12] Xu M H, Wu T. A class of linearized proximal alternating direction methods. Journal of Optimization Theory and Applications, 2011, 151: 321-337.
[13]Hager W W, Zhang H. Inexact alternating direction methods of multipliers for separable convex optimization. Computational Optimization and Applications, 2019, 73: 201-235.
[14]Chen C, He B, Ye Y, et al. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. Mathematical Programming, 2016, 155(1): 57-79.
[15]Corman E, Yuan X. A generalized proximal point algorithm and its convergence rate. SIAM Journal on Optimization, 2014, 24(4): 1614-1638.
[16]He B, Ma F, Yuan X. Convergence study on the symmetric version of ADMM with larger step sizes. SIAM journal on imaging sciences, 2016, 9(3): 1467-1501.
[17]He B, Ma F, Yuan X. Optimally linearizing the alternating direction method of multipliers for convex programming. Computational Optimization and Applications, 2020, 75(2): 361-388.
[18]Eckstein J, Yao W. Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives. Pac. J. Optim., 2015, 11(4): 619-644.
[19]Bai J, Li J, Xu F, et al. Generalized symmetric ADMM for separable convex optimization. Computational optimization and applications, 2018, 70(1): 129-170.
[20]Gu G, He B, Yang J. Inexact alternating-direction-based contraction methods for separable linearly constrained convex optimization. Journal of Optimization Theory and Applications, 2014, 163: 105-129.