CONVERGENCE ON A SYMMETRIC ACCELERATED STOCHASTIC ADMM WITH LARGER STEPSIZES *

JIANCHAO BAI1, DEREN HAN2, HAO SUN3, AND HONGCHAO ZHANG4

Abstract. In this paper, we develop a symmetric accelerated stochastic Alternating Direction Method of Multipliers (SAS-ADMM) for solving separable convex optimization problems with linear constraints. The objective function is the sum of a possibly nonsmooth convex function and an average function of many smooth convex functions. Our proposed algorithm combines both ideas of ADMM and the techniques of accelerated stochastic gradient methods using variance reduction to solve the smooth subproblem. One main feature of SAS-ADMM is that its dual variable is symmetrically updated after each update of the separated primal variable, which would allow a more flexible and larger convergence region of the dual variable compared with that of standard deterministic or stochastic ADMM. This new stochastic optimization algorithm is shown to converge in expectation with $O(1/T)$ convergence rate, where $T$ is the number of outer iterations. In addition, 3-block extensions of the algorithm and its variance of an accelerated stochastic augmented Lagrangian method are also discussed. Our preliminary numerical experiments indicate the proposed algorithm is very effective for solving separable optimization problems from big-data applications.

Key words. convex optimization, separable structure, stochastic ADMM, symmetric ADMM, inexact ADMM iteration, larger stepsize, proximal mapping, complexity, big data

AMS subject classifications. 65K10, 65Y20, 68W40, 90C25

1. Introduction. We consider the following structured composite convex optimization problem with linear equality constraints:

$$\min\{f(x) + g(y) \mid x \in \mathcal{X}, y \in \mathcal{Y}, Ax + By = b\},$$

where $\mathcal{X} \subseteq \mathbb{R}^{n_1}$, $\mathcal{Y} \subseteq \mathbb{R}^{n_2}$ are closed convex subsets, $A \in \mathbb{R}^{n \times n_1}$, $B \in \mathbb{R}^{n \times n_2}$, $b \in \mathbb{R}^{n}$ are given, $g : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ is a convex but possibly nonsmooth function, and $f$ is an average of $N$ real-valued convex functions:

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

We assume that each $f_j$ defined on an open set containing $\mathcal{X}$ is Lipschitz continuously differentiable on $\mathcal{X}$. Problem (1.1) is also referred as regularized empirical risk minimization in big-data applications [19, 26], including classification and regression models in machine learning, where $N$ denotes the sample size and $f_j$ corresponds to...
the empirical loss. A major difficulty for solving (1.1) is that the sample size $N$ can be very large or even huge such that it is often computationally prohibitive to evaluate either the full function value or the gradient of $f$ at each iteration of an algorithm. Hence, it is essential for an effective algorithm, e.g., a stochastic gradient method, to explore the summation structure of $f$ in the objective function.

The augmented Lagrangian function of (1.1) is generally given by

(1.2) $\mathcal{L}_\beta(x, y, \lambda) = \mathcal{L}(x, y, \lambda) + \frac{\beta}{2} \|Ax + By - b\|^2,$

where $\beta > 0$ is a penalty parameter, $\lambda$ is the Lagrange multiplier and the Lagrangian of (1.1) is defined as

$\mathcal{L}(x, y, \lambda) = f(x) + g(y) - \lambda^T(Ax + By - b).$

Although the Augmented Lagrangian Method (ALM) can be applied to solve (1.1), it does not take advantage of the separable structure of (1.1). Hence, it is commonly known that ALM could be quite inefficient to minimize both primal variables $x$ and $y$ simultaneously. As a splitting version of ALM, the standard Alternating Direction Method of Multipliers (ADMM, [9, 10]) exploits the separable structure of the objective function, and performs the following iterations:

\[
\begin{aligned}
x^{k+1} & \in \arg \min_{x \in X} \mathcal{L}_\beta(x, y^k, \lambda^k), \\
y^{k+1} & \in \arg \min_{y \in Y} \mathcal{L}_\beta(x^{k+1}, y, \lambda^k), \\
\lambda^{k+1} & = \lambda^k - s\beta \left(Ax^{k+1} + By^{k+1} - b\right),
\end{aligned}
\]

where $s \in (0, \frac{1+\sqrt{5}}{2})$ is considered as stepsize of the dual variable $\lambda$.

The classical ADMM is shown convergent for solving the 2-block separable optimization problem [10], but its direct extension for solving optimization problem with multi-block separable variables is not necessarily convergent [7] although its efficiency had been observed in many practical applications. Motivated from enlarging stepsize region of dual variable in [15], Gu, et al. [11] proposed a symmetric proximal ADMM whose dual variable is updated twice with different stepsizes. In contrast to the work [11], He, et al. [16] proposed the following symmetric ADMM (S-ADMM) without using additional proximal terms:

\[
\begin{aligned}
x^{k+1} & \in \arg \min_{x \in X} \mathcal{L}_\beta(x, y^k, \lambda^k), \\
y^{k+1} & \in \arg \min_{y \in Y} \mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+1/2}), \\
\lambda^{k+1} & = \lambda^{k+1/2} - s\beta \left(Ax^{k+1} + By^{k+1} - b\right),
\end{aligned}
\]

where $(\tau, s)$ denotes the stepsize parameter satisfying

$\Delta_0 = \{ (\tau, s) \mid s \in (0, (1 + \sqrt{5})/2), \tau + s > 0, \tau \in (-1, 1), \ | \tau | < 1 + s - s^2 \}.$

Recently, Bai, et al. [1] further designed a generalized S-ADMM (GS-ADMM) for solving a multi-block separable convex optimization and enlarged the above convergence region $\Delta_0$ to

$\Delta_1 = \{ (\tau, s) \mid \tau + s > 0, \tau \leq 1, -\tau^2 - s^2 - \tau s + \tau s + s + 1 > 0 \}.$
Moreover, numerical experiments show that symmetrically updating the dual variable in a more flexible way often increases the algorithm performance [11, 16]. The sublinear convergence rate of GS-ADMM in the nonergodic sense and its linear convergence rate can be found in [2]. To the best of our knowledge, $\Delta_1$ seems so far the largest convergence region of the dual stepsize for symmetric ADMM-type algorithms and has been used in the logarithmic-quadratic proximal based ADMM for solving the 2-block problems [21] and the grouped multi-block problems [4].

For convergence rate of ADMM, it is well-known that most of deterministic ADMM algorithms [1, 6, 12, 14, 20, 22, 23] enjoy global $O(1/T)$ ergodic convergence rate for convex separable convex optimization, where $T$ is the iteration number. Under the assumption that the subdifferential of each component objective function is piecewise linear, Yang-Han [25] established linear convergence rate of ADMM for two-block separable convex optimization. Assuming that an error bound condition holds, the dual stepsize is sufficiently small and the coefficient matrices in the equality constraint have full column rank, Hong-Luo [17] showed a linear convergence rate of their multi-block ADMM. Recently, Zhang et al. [27] developed a majorized ADMM with indefinite proximal terms (iPADMM) for a class of composite convex optimization problems, and analyzed the convergence of this iPADMM with a linear convergence rate under a local error bound condition. Moreover, Chang et al. [6] proposed a linearized symmetric ADMM but with indefinite proximal regularization and optimal proximal parameter for solving the multi-block separable convex optimization. More recently, Yuan-Zeng-Zhang [24] show that the local linear convergence of ADMM can be guaranteed without the strong convexity assumptions of the objective functions, or the full rank assumption of the coefficient matrices in the constraints, or the full polyhedral assumption on the subdifferential of the objective function. For more details about linear convergence rate under strongly convexity assumption, we refer the readers to [3, 5, 13, 18] and the references therein.

Motivated from the stochastic ADMM developed in [3] and the generalized symmetric ADMM proposed in [1], we propose a Symmetric Accelerated Stochastic ADMM (SAS-ADMM, i.e., Algorithm 1.1) with the following region of the dual stepsize:

\[
\Delta = \{ (\tau, s) \mid \tau + s > 0, \tau \leq 1, -\tau^2 - s^2 - \tau s + \tau + s + 1 \geq 0 \}.
\]

The major features of this SAS-ADMM can be summarized as the follows:

(i) This SAS-ADMM has many similar features to AS-ADMM developed in [3]. For example, SAS-ADMM has low memory requirement since there is no need to save previous stochastic gradients and iterates; the subroutine $\mathbf{x}_{\text{sub}}$ is a variant of deterministic accelerated gradient method where the full gradient is replaced by a stochastic gradient; and variance reduction techniques can be applied. We assume that the possibly nonsmooth convex function $g(y)$ is proper closed and the proximal $y$-subproblem is solvable. In addition, we also assume the projection on the constraint set $\mathcal{X}$ when $\mathcal{X} \neq \mathbb{R}^n_1$ can be efficiently computed.

(ii) Unlike the classic ADMM and AS-ADMM [3], the dual variable of SAS-ADMM is symmetrically updated twice with stepsize domain (1.3). Compared with the standard stepsize region for the dual variable, this symmetric updates of dual variable is more balanced, flexible and could lead better numerical performance.

(iii) When stepsize $\tau = 0$ and $L = 0$, SAS-ADMM will reduce to AS-ADMM with stepsize $s \in (0, \frac{\sqrt{5} - 1}{2}]$. If $m_k = 1, N = 1$, then SAS-ADMM will degrade
Parameters: $\beta > 0$, $\mathcal{H} > 0$, $L \succeq 0$ and $(\tau, s) \in \Delta$ given by (1.3).
Initialization: $(x^0, y^0, \lambda^0) \in \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^n$, $\hat{x}^0 = x^0$.

For $k = 0, 1, \ldots$

Choose $m_k > 0$, $\eta_k > 0$ and $\mathcal{M}_k$ such that $\mathcal{M}_k - \beta A^T A \succeq 0$.

\[ h^k := -A^T \left[ \lambda^k - \beta(Ax^k + By^k - b) \right]. \]

\( (x^{k+1}, \hat{x}^{k+1}) = \text{xsub} \ (x^k, \hat{x}^k, h^k). \)

\[ \lambda^{k+\frac{1}{2}} = \lambda^k - \tau \beta \left( Ax^{k+1} + By^{k+1} - b \right). \]

\( y^{k+1} \in \arg \min_{y \in \mathcal{Y}} \left( x^{k+1}, y, \lambda^{k+\frac{1}{2}} \right) + \frac{1}{2} \| y - y^k \|_L^2. \)

\[ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s \beta \left( Ax^{k+1} + By^{k+1} - b \right). \]

end

\((x^+, \hat{x}^+) = \text{xsub} \ (x_1, \hat{x}_1, h)\).

For $t = 1, 2, \ldots, m_k$

Randomly select $\xi_t \in \{1, 2, \ldots, N\}$ with uniform probability. $\beta_t = \frac{2}{(t + 1)}$, $\gamma_t = \frac{2}{tb_k}$, $\hat{x}_t = \beta_t \hat{x}_{t-1} + (1 - \beta_t) x_t$.

\[ d_t = g_t + e_t, \text{ where } g_t = \nabla f_{\xi_t}(\hat{x}_t) \text{ and } e_t \text{ is a random vector satisfying } \mathbb{E}[e_t] = 0. \]

\[ \hat{x}_{t+1} = \arg \min_{x \in \mathcal{X}} \left\{ \langle d_t + h, x \rangle + \frac{\tau}{2} \| x - \hat{x}_t \|_H^2 + \frac{1}{2} \| x - x^k \|_{\mathcal{M}_k}^2 : x \in \mathcal{X} \right\}. \]

\[ x_{t+1} = \beta_t \hat{x}_{t+1} + (1 - \beta_t) x_t. \]

end

Return \((x^+, \hat{x}^+) = (x_{m_k+1}, \hat{x}_{m_k+1})\).

Alg. 1.1. Symmetric accelerated stochastic ADMM (SAS-ADMM) with larger stepsizes

to a linearized symmetric ADMM. When $m_k > 1, N = 1$, SAS-ADMM is a multi-step deterministic inexact symmetric ADMM. Hence, the convergence properties developed in this paper also apply to these deterministic algorithms as special cases. Moreover, by taking $L = \gamma I - \beta B^T B$ for some $\gamma > 0$, the $y$-subproblem would become the following proximal mapping problem:

\begin{equation}
\text{prox}_g^\gamma(y^k) := \arg \min_{y \in \mathcal{Y}} g(y) + \frac{\gamma}{2} \| y - y^k \|_2^2,
\end{equation}

where $y^k = y^k - \beta B^T (Ax^{k+1} + By^k - b - \lambda^{k+\frac{1}{2}} / \beta) / \gamma$. Hence, closed-form solution of $y$-subproblem may exist when function $g$ has special structure.

(iv) We further discuss the 3-block extensions of SAS-ADMM and its variance of an accelerated stochastic augmented Lagrangian method (AS-ALM). With the aid of variational analysis, we show that SAS-ADMM converges in expectation with a $O(1/T)$ rate, where $T$ is the outer iteration number. Our preliminary numerical experiments show that SAS-ADMM performs competitively well and even slightly better than AS-ADMM [3] for solving some separable optimization problems from big-data applications.

2. Preliminaries

2.1. Notations and assumptions. The following notations are used in this paper. Let $\mathbb{R}$, $\mathbb{R}^n$, and $\mathbb{R}^{n \times l}$ be the sets of real numbers, $n$ dimensional real column vectors, and $n \times l$ dimensional real matrices, respectively. The $I$ and $0$ denote the
identity matrix and the zero matrix/vector, respectively. For any symmetric matrices $A$ and $B$ of the same dimension, $A \succ B$ ($A \succeq B$) means $A - B$ is a positive definite (semidefinite) matrix. For any symmetric matrix $G$, define $\|x\|_G^2 := x^T G x$ and $\|x\|_G := \sqrt{x^T G x}$ if $G \succeq 0$. We use $\| \cdot \|$ to denote the standard Euclidean norm equipped with inner product $\langle \cdot, \cdot \rangle$, $\nabla f(x)$ to represent the gradient of $f$ at $x$, and $\mathbb{E}[\cdot]$ to denote mathematical expectation. We also define

\begin{equation}
(2.1) \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad J(w) = \begin{pmatrix} -A^T \lambda \\ -B^T \lambda \\ Ax + By - b \end{pmatrix},
\end{equation}

and

\begin{equation}
(2.2) \quad w^k = \begin{pmatrix} x^k \\ y^k \\ \lambda^k \end{pmatrix}, \quad J(w^k) = \begin{pmatrix} -A^T \lambda^k \\ -B^T \lambda^k \\ Ax^k + By^k - b \end{pmatrix}.
\end{equation}

For convenience of analysis, we define $F(w) = f(x) + g(y)$.

Throughout the context, we make the following assumptions:

**Assumption 2.1.** The primal-dual solution set $\Omega^*$ of problem (1.1) is nonempty, and the problem $\min_{y \in Y} \{ g(y) + \frac{\nu}{2} y^T B^T B y + z^T y \}$ has a minimizer for any $z \in \mathbb{R}^n$.

**Assumption 2.2.** For any $H \succ 0$, there exists a constant $\nu > 0$ such that the gradients $\nabla f_j$ satisfy the Lipschitz condition

\begin{equation}
(2.3) \quad \| \nabla f_j(x_1) - \nabla f_j(x_2) \|_H \leq \nu \| x_1 - x_2 \|_H
\end{equation}

for every $x_1, x_2 \in \mathcal{X}$ and $j = 1, 2, \ldots, N$.

The first assumption is a basic assumption to illustrate the solvability of the problem. Under Assumption 2.2, it holds that for every $x, y \in \mathcal{X}$, we have

$$f(x_1) \leq f(x_2) + \langle \nabla f(x_2), x_1 - x_2 \rangle + \frac{\nu}{2} \| x_1 - x_2 \|_H^2.$$

**2.2. Variational characterization.** Denote $\Omega = \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^n$. It is well-known that a point $w^* := (x^*; y^*; \lambda^*) \in \Omega$ is called the saddle-point of $\mathcal{L}(x, y, \lambda)$ if it satisfies

\begin{equation}
(2.4) \quad \mathcal{L}(x^*, y^*, \lambda) \leq \mathcal{L}(x, y^*, \lambda^*) \leq \mathcal{L}(x, y, \lambda^*) , \quad \forall w \in \Omega,
\end{equation}

which is equivalent to

\begin{equation}
\begin{cases}
    f(x) - f(x^*) + (x - x^*)^T (-A^T \lambda^*) \geq 0, \\
    g(y) - g(y^*) + (y - y^*)^T (-B^T \lambda^*) \geq 0, \\
    Ax^* + By^* - b = 0.
\end{cases}
\end{equation}

Rewriting these inequalities as a more compact form, it gives

\begin{equation}
(2.5) \quad F(w) - F(w^*) + (w - w^*)^T J(w^*) \geq 0, \quad \forall w \in \Omega.
\end{equation}

Notice that the affine mapping $J(\cdot)$ is skew-symmetric. So, we have

\begin{equation}
(2.6) \quad (w - w^*)^T [J(w) - J(w^*)] \equiv 0, \quad \forall w, w^* \in \Omega.
\end{equation}
Hence, (2.5) is also equivalent to

\[(2.7) \quad F(w) - F(w^*) + (w - w^*)^T J(w) \geq 0, \quad \forall w \in \Omega.\]

For later analysis, we need the following lemma on the iterates generated by the \textbf{x}sub routine in Algorithm 1.1. The lemma was given in [3]. Hence, we omit its proof.

**Lemma 2.1.** [3, Lemma 3.2] Let $\delta_i = \nabla f(x_i) - d_i$ and $D_k = M_k - \beta A^T A$ respectively. Suppose $\eta_k \in (0, 1/\nu)$. Then, the iterates generated by Algorithm 1.1 satisfy

\[(2.8) \quad f(x) - f(x^{k+1}) + \langle x - x^{k+1}, -A^T \tilde{\lambda}^k \rangle \geq \langle x^{k+1} - x, D_k(x^{k+1} - x) \rangle + \zeta^k \]

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

\[(2.9) \quad \tilde{\lambda}^k = \lambda^k - \beta (Ax^{k+1} + By^k - b) \quad \text{and} \]

\[(2.10) \quad \zeta^k = \frac{2}{m_k(m_k+1)} \left[ \frac{1}{\eta_k} \left( \|x - x^{k+1}\|_H^2 - \|x - \tilde{x}^k\|_H^2 \right) - \sum_{i=1}^{m_k} t_i \langle \delta_i, \tilde{x}_i - x \rangle - \frac{\eta_k}{4(1-\eta_k \nu)} \sum_{i=1}^{m_k} t_i^2 \|\delta_i\|_{H^{-1}}^2 \right].\]

Based on the previous lemma, we have the following theorem.

**Theorem 2.2.** Suppose $\eta_k \in (0, 1/\nu)$. Then, the iterates generated by Algorithm 1.1 satisfy

\[(2.11) \quad F(w) - F(\tilde{w}^k) + \langle w - \tilde{w}^k, J(w) \rangle \geq (w - \tilde{w}^k)^T Q_k (w^k - \tilde{w}^k) + \zeta^k \]

for all $w \in \Omega$, where $\zeta^k$ and $\tilde{\lambda}$ are defined in (2.9) and (2.10),

\[(2.12) \quad \tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k \end{pmatrix} \quad \text{and} \quad Q_k = \begin{bmatrix} D_k & -B B^T & -\tau B^T \\ L + \beta B^T B & -B & \frac{1}{\nu} I \end{bmatrix}.\]

**Proof.** By the first-order optimality condition of y-subproblem, we have

\[(2.13) \quad g(y) - g(y^{k+1}) + \langle y - y^{k+1}, p_k \rangle \geq 0, \quad \forall y \in \mathcal{Y},\]

where $p_k$ is the gradient of the smooth terms in the objective function of the y-subproblem:

\[
p_k = -B^T \lambda^{k+\frac{1}{2}} + \beta B^T (Ax^{k+1} + By^{k+1} - b) + L(y^{k+1} - y^k)
\]

\[
= -B^T \lambda^{k+\frac{1}{2}} + \beta B^T (Ax^{k+1} + By^k - b) + [L + \beta B^T B] (y^{k+1} - y^k)
\]

\[
= -B^T \lambda^{k+\frac{1}{2}} + B^T (\lambda^k - \tilde{\lambda}^k) + [L + \beta B^T B] (y^{k+1} - y^k)
\]

\[
= -B^T \lambda^k + \tau B^T (\lambda^k - \tilde{\lambda}^k) + [L + \beta B^T B] (y^{k+1} - y^k),
\]

where we use the following relation

\[(2.14) \quad \lambda^{k+\frac{1}{2}} = \lambda^k - \tau (\lambda^k - \tilde{\lambda}^k).\]
By the definition of $\tilde{\lambda}^k$, we have

\[(2.15) \quad (A\tilde{x}^k + B\tilde{y}^k - b) - B(y^k - \tilde{y}^k) + \frac{1}{\beta} (\tilde{\lambda}^k - \lambda^k) = 0.\]

Taking inner product of the above equality with $\lambda^k - \tilde{\lambda}^k$, we get

\[(2.16) \quad \langle \lambda^k - \tilde{\lambda}^k, A\tilde{x}^k + B\tilde{y}^k - b \rangle = \langle \lambda^k - \tilde{\lambda}^k, -B(y^k - \tilde{y}^k) + \frac{1}{\beta} (\lambda^k - \tilde{\lambda}^k) \rangle.\]

Then, the inequality (2.11) is achieved by combining (2.8), (2.13), (2.16) together with the property (2.6).

3. Convergence analysis.

3.1. Basic lemmas and theorems. We first show the following corollary before establishing the final convergence of Algorithm 1.1.

**Corollary 3.1.** Suppose $\eta_k \in (0, 1/\nu)$. Then, the iterates generated by Algorithm 1.1 satisfy

\[(3.1) \quad F(w) - F(\tilde{w}^k) + (w - \tilde{w}^k)^T J(w) \geq \frac{1}{2} \left\{ \|w - w^{k+1}\|^2_{\tilde{Q}_k} - \|w - w^k\|^2_{\tilde{Q}_k} + \|w^k - \tilde{w}^k\|^2_{\tilde{G}_k} \right\} + \zeta^k \]

for all $w \in \Omega$, where $\zeta^k$ is defined in (2.9) and

\[(3.2) \quad \tilde{Q}_k = D_k \left[ L + \left( \frac{1 - \tau_s}{\tau + s} \right) \beta B^T B - \frac{\tau_s}{\tau + s} I \right], \quad \tilde{G}_k = D_k \left[ L + \left( \frac{1 - \tau_s}{\tau + s} \right) \beta B^T B - \frac{\tau_s}{\tau + s} I \right].\]

**Proof.** By (2.14) and the way of generating $\lambda^{k+1}$, we have

\[(3.3) \quad -s\beta \langle y^k - \tilde{y}^k \rangle + (\tau + s) \left( \lambda^k - \tilde{\lambda}^k \right) = \lambda^k - \lambda^{k+1},\]

which, by the definition of $\tilde{w}^k$ in (2.12), further shows

\[(3.4) \quad w^k - w^{k+1} = P (w^k - \tilde{w}^k) \quad \text{with} \quad P = \left[ \begin{array}{cc} I & 1 \\ -s \beta B & (\tau + s) I \end{array} \right].\]

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

\[Q_k P^{-1} = \left[ D_k \left[ L + \left( \frac{1 - \tau_s}{\tau + s} \right) \beta B^T B - \frac{\tau_s}{\tau + s} I \right] \right] = \tilde{Q}_k.\]

So, for any $w \in \Omega$, it follows from (2.11) and the above relation that

\[(3.5) \quad F(w) - F(\tilde{w}^k) + (w - \tilde{w}^k)^T J(w) \geq \zeta^k + (w - \tilde{w}^k)^T \tilde{Q}_k (w^k - w^{k+1}) = \zeta^k + \frac{1}{2} \left\{ \|w - w^{k+1}\|^2_{\tilde{Q}_k} - \|w^k - \tilde{w}^k\|^2_{\tilde{Q}_k} + \|w^k - \tilde{w}^k\|^2_{\tilde{G}_k} - \|w^{k+1} - \tilde{w}^k\|^2_{\tilde{G}_k} \right\},\]
where the equality uses the identity
\[ 2(a - b)^T \tilde{Q}_k(c - d) = \|a - d\|^2_{\tilde{Q}_k} - \|a - c\|^2_{\tilde{Q}_k} + \|c - b\|^2_{\tilde{Q}_k} - \|b - d\|^2_{\tilde{Q}_k} \]
with specifications \(a := w, b := \tilde{w}^k, c := w^k, d := w^{k+1}\).

Now, by (3.4) again, we deduce
\[
\|w^k - \tilde{w}^k\|^2_{\tilde{Q}_k} - \|w^{k+1} - \tilde{w}^k\|^2_{\tilde{Q}_k} + \|w^k - \tilde{w}^k - P(w^k - \tilde{w}^k)\|^2_{\tilde{Q}_k} - \|w^k - \tilde{w}^k\|^2_{\tilde{Q}_k} = \|w^k - \tilde{w}^k\|^2_{\tilde{Q}_k},
\]
where we use the relation \(\tilde{Q}_k = P^T \tilde{Q}_k + \tilde{Q}_k P - P^T \tilde{Q}_k P\) for the last equality. Then, (3.1) follows from (3.5).

The above Corollary 3.1 and its proof abuse the notation \(\|w^k\|^2_{\tilde{Q}_k} := (w^k)^T \tilde{Q}_k w^k\) since \(\tilde{Q}_k\) is symmetric but not always positive semidefinite for any parameter \(\tau\). Next, we provide a sufficient condition to ensure the positive semidefiniteness of \(\tilde{Q}_k\).

**Lemma 3.2.** Let \(L \succeq (\tau - 1)B^TB\). Then, the matrix \(\tilde{Q}_k\) given by (3.2) is symmetric positive semidefinite for any \((\tau, s) \in \Delta\).

**Proof.** Clearly, we just need to check the lower-upper 2-by-2 block of \(\tilde{Q}_k\), i.e.,
\[
\tilde{Q}_k^T = \begin{bmatrix} L + \left(1 - \frac{\tau}{\tau + s}\right)B^TB & -\frac{\tau}{\tau + s}B^T \\ -\frac{\tau}{\tau + s}B & \frac{1}{\tau + s}I \end{bmatrix} \begin{bmatrix} \beta^2 B & \beta^{-\frac{s}{2}} \beta^{-\frac{s}{2}} \beta^{-\frac{s}{2}} \end{bmatrix}
\]
\[
= \begin{bmatrix} 1 & \tau I & I \end{bmatrix} \begin{bmatrix} \frac{\tau}{\tau + s} I & -\frac{\tau}{\tau + s} I & -\frac{\tau}{\tau + s} I \\ -\frac{\tau}{\tau + s} I & \frac{1}{\tau + s} I & -\frac{\tau}{\tau + s} I \\ -\frac{\tau}{\tau + s} I & -\frac{\tau}{\tau + s} I & \frac{1}{\tau + s} I \end{bmatrix} \begin{bmatrix} \beta^2 B & \beta^{-\frac{s}{2}} \beta^{-\frac{s}{2}} \beta^{-\frac{s}{2}} \end{bmatrix}
\]
is positive semidefinite. Notice that
\[
\begin{bmatrix} 1 & \tau I & I \end{bmatrix} \begin{bmatrix} \frac{\tau}{\tau + s} I & -\frac{\tau}{\tau + s} I & -\frac{\tau}{\tau + s} I \\ -\frac{\tau}{\tau + s} I & \frac{1}{\tau + s} I & -\frac{\tau}{\tau + s} I \\ -\frac{\tau}{\tau + s} I & -\frac{\tau}{\tau + s} I & \frac{1}{\tau + s} I \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}.
\]
So, \(\tilde{Q}_k^T\) is positive semidefinite since \(\tau + s > 0\) for any \((\tau, s) \in \Delta\). \(\square\)

To show the global convergence of Algorithm 1.1, we need to further establish a useful lower bound on the term \(\|w^k - \tilde{w}^k\|^2_{\tilde{Q}_k}\), since \(\tilde{G}_k\) is not necessarily positive definite for any \((\tau, s) \in \Delta\). In the following theorem, we assume \(L \succeq 0\), which implies \(L \succeq (\tau - 1)B^TB\) since \(\tau \leq 1\). Hence, with \(L \succeq 0\), the Lemma 3.2 holds as well.

**Theorem 3.3.** Let \(L \succeq 0\). Then, for any \((\tau, s) \in \Delta\) defined in (1.3), we have
\[
\begin{align*}
\|w^k - \tilde{w}^k\|^2_{\tilde{Q}_k} &\geq \|x^k - x^{k+1}\|^2_{\tilde{Q}_k} + \omega_0 \|Ax^{k+1} + By^{k+1} - b\|^2 \\
&+ \omega_1 \left(\|Ax^{k+1} + By^{k+1} - b\|^2 - \|Ax^k + By^k - b\|^2\right) \\
&+ \omega_2 \left(\|y^k - y^{k+1}\|^2_L - \|y^{k-1} - y^k\|^2_L\right),
\end{align*}
\]
where $\omega_i \geq 0$, $i = 0, 1, 2$, are given as
\begin{equation}
\omega_0 = \left(2 - \tau - s - \frac{(1 - s)^2}{1 + \tau}\right) \beta, \quad \omega_1 = \frac{(1 - s)^2}{1 + \tau} \beta, \quad \omega_2 = \frac{1 - \tau}{1 + \tau}.
\end{equation}

Proof. First of all, for any $(\tau, s) \in \Omega$, we have $1 + \tau > 0$. By the structure of $\tilde{G}_k$ in (3.2), $L \succeq 0$ and (2.15), we have
\begin{align}
\|w^k - \tilde{w}^k\|_G^2 &= \|x^k - x^{k+1}\|_{D_k}^2 + \|x^k - y^{k+1}\|_{L(1-s)\beta B^T B}^2 \\
&\quad + 2(s - 1) \left(\lambda^k - \tilde{\lambda}^k\right)^T B (y^k - y^{k+1}) + \frac{2 - \tau - s}{\beta} \|\lambda^k - \tilde{\lambda}^k\|^2 \\
&\geq \|x^k - x^{k+1}\|_{D_k}^2 + (2 - \tau - s) \beta \|Ax^{k+1} + By^{k+1} - b\|^2 \\
&\quad + (1 - \tau) \beta \|B (y^k - y^{k+1})\|^2 \\
&\quad + 2(1 - \tau) \beta (Ax^{k+1} + By^{k+1} - b)^T B (y^k - y^{k+1})).
\end{align}

Next, we estimate the last crossing term in the inequality of (3.8). Taking $y = y^k$ in the first-order optimality condition (2.13) yields
\begin{align}
g(y^k) - g(y^{k+1}) + \left(y^k - y^{k+1}, -B^T \lambda^{k+\frac{1}{2}} + \beta B^T (Ax^{k+1} + By^{k+1} - b) + L(y^{k+1} - y^k)\right) \geq 0.
\end{align}

Similarly, letting $y = y^{k+1}$ in the first-order optimality condition of $y$-subproblem at the $(k+1)$-th iteration gives
\begin{align}
g(y^{k+1}) - g(y^k) + \left(y^{k+1} - y^k, -B^T \lambda^{k+\frac{1}{2}} + \beta B^T (Ax^k + By^k - b) + L(y^k - y^{k-1})\right) \geq 0.
\end{align}

Summing up the above two inequalities together with the relation
\begin{align}
\lambda^{k+\frac{1}{2}} - \lambda^{k+\frac{1}{2}} = \tau \beta (Ax^{k+1} + By^{k+1} - b) + s \beta (Ax^k + By^k - b) + \tau \beta B(y^k - y^{k+1}),
\end{align}

and noticing that $1 + \tau > 0$, we obtain
\begin{align}
(Ax^{k+1} + By^{k+1} - b)^T B (y^k - y^{k+1}) \\
&\geq \frac{1 - s}{1 + \tau} \left((Ax^k + By^k - b)^T B (y^k - y^{k+1}) - \frac{\tau}{1 + \tau} \|B(y^k - y^{k+1})\|^2 \right) \\
&\quad + \frac{1}{\beta(1 + \tau)} (y^k - y^{k+1})^T L (y^k - y^{k+1} - (y^{k-1} - y^k)) \\
&\geq \frac{1 - s}{1 + \tau} \left((Ax^k + By^k - b)^T B (y^k - y^{k+1}) - \frac{\tau}{1 + \tau} \|B(y^k - y^{k+1})\|^2 \right) \\
&\quad + \frac{1}{2\beta(1 + \tau)} \left(\|y^k - y^{k+1}\|_L^2 - \|y^{k-1} - y^k\|_L^2\right).
\end{align}

Then, combining (3.8) and (3.9) we have
\begin{align}
\|w^k - \tilde{w}^k\|_{\tilde{G}_k}^2 \\
&\geq \|x^k - x^{k+1}\|_{D_k}^2 + (2 - \tau - s) \beta \|Ax^{k+1} + By^{k+1} - b\|^2 \\
&\quad + \frac{2\beta}{1 + \tau} (1 - \tau)(1 - s) \left((Ax^k + By^k - b)^T B (y^k - y^{k+1}) \right)
\end{align}
\[+(1-\tau)\beta \|B(y^k-y^{k+1})\|^2 - \frac{2\tau(1-\tau)}{1+\tau} \beta \|B(y^k-y^{k+1})\|^2\]
\[+\frac{1-\tau}{1+\tau} \left(\|y^k-y^{k+1}\|_L^2 - \|y^{k-1}-y^k\|_L^2\right)\]
\[\geq \|x^k-x^{k+1}\|_D^2 + \left(2 - \tau - s - \frac{(1-s)^2}{1+\tau}\right) \beta \|Ax^{k+1} + By^{k+1} - b\|^2\]
\[+\frac{(1-s)^2}{1+\tau} \beta \left(\|Ax^{k+1} + By^{k+1} - b\|^2 - \|Ax^k + By^k - b\|^2\right)\]
\[+\frac{1-\tau}{1+\tau} \left(\|y^k-y^{k+1}\|_L^2 - \|y^{k-1}-y^k\|_L^2\right)\]
\[= \|x^k-x^{k+1}\|_D^2 + \left(2 - \tau - s - \frac{(1-s)^2}{1+\tau}\right) \beta \|Ax^{k+1} + By^{k+1} - b\|^2\]
\[+\frac{(1-s)^2}{1+\tau} \beta \left(\|Ax^{k+1} + By^{k+1} - b\|^2 - \|Ax^k + By^k - b\|^2\right)\]
\[+\frac{1-\tau}{1+\tau} \left(\|y^k-y^{k+1}\|_L^2 - \|y^{k-1}-y^k\|_L^2\right),\]
where the second inequality follows from the Cauchy-Schwartz inequality
\[2(1-s)(1-\tau) (Ax^k + By^k - b)\top B(y^k-y^{k+1})\]
\[\geq -(1-s)^2 \|Ax^k + By^k - b\|^2 - (1-\tau)^2 \|B(y^k-y^{k+1})\|^2.\]

So, (3.6) holds with \(\omega_i, i = 0, 1, 2\), defined as in (3.7). Moreover, for any \((\tau, s) \in \Delta\), we can derive \(\omega_i \geq 0\) for \(i = 0, 1, 2\). This completes the whole proof. \(\square\)

3.2. Iteration complexity in expectation. In this subsection, we analyze the convergence and the iteration complexity of Algorithm 1.1. We first have the following lemma.

**Lemma 3.4.** Suppose \(L \succeq 0\) and \((\tau, s) \in \Delta\) defined in (1.3). If for some integers \(\kappa, T > 0\), the following conditions hold for all \(k \in [\kappa, \kappa + T]\): (I) \(\eta_k \in (0, 1/(2\nu)]\) and the sequence \(\{\eta_k m_k(m_k + 1)\}\) is nondecreasing; (II) \(D_k \succeq D_{k+1} \succeq 0\) and \(E(\|\Delta_t\|_{H^{-1}}^2) \leq \sigma^2\) for some \(\sigma > 0\), where \(\Delta_t\) and \(D_k\) are defined in Lemma 2.1. Then, for any \(w \in \Omega\), we have

\[E[F(w_T) - F(w) + (w_T - w)\top J(w)]\]
\[\leq \frac{1}{2T} \left\{\sigma^2 \sum_{k=\kappa}^{\kappa+T} \eta_k m_k + \frac{4}{m_k(m_k + 1)} \|x - \bar{x}_k\|_H^2 + \|w - w^\kappa\|_Q^2\right\}
\[+ \omega_1 \|Ax^k + By^k - b\|^2 + \omega_2 \|y^{k-1} - y^k\|_L^2\},\]
where \(w_T = \frac{1}{T} \sum_{k=\kappa}^{\kappa+T} \bar{w}_k\), \(\omega_1 \geq 0\) and \(\omega_2 \geq 0\) are defined in (3.7).

**Proof.** By the assumption, \(D_k \succeq D_{k+1} \succeq 0\) which implies that the matrix \(\tilde{Q}_k\) in (3.2) satisfies \(\tilde{Q}_k \succeq \tilde{Q}_{k+1} \succeq 0\). Substituting (3.6) into (3.1) and utilizing the relation \(\tilde{Q}_k \succeq \tilde{Q}_{k+1}\), it follows from Theorem 3.3 that

\[F(\tilde{w}^k) - F(w) + (\tilde{w}^k - w)\top J(w)\leq\]
\[-\zeta^k + \frac{1}{2} \left\{ \|w - w^k\|^2_{\tilde{Q}_k} - \|w - w^{k+1}\|^2_{\tilde{Q}_{k+1}} \right\} + \frac{\omega_1}{2} \left( \|Ax^k + By^k - b\|^2 - \|Ax^{k+1} + By^{k+1} - b\|^2 \right) + \frac{\omega_2}{2} \left( \|y^{k-1} - y^k\|^2_L - \|y^k - y^{k+1}\|^2_L \right), \]

where \(\omega_1, \omega_2\) are defined in (3.7). Summing the above inequality over \(k\) between \(\kappa\) and \(\kappa + T\), we deduce by Lemma 3.2 that

\[
(3.11) \quad \sum_{k=\kappa}^{\kappa+T} F(\tilde{w}^k) - T \left\{ F(w) + (w_T - w)^T J(w) \right\} \leq -\sum_{k=\kappa}^{\kappa+T} \zeta^k + \frac{1}{2} \left\{ \|w - w^k\|^2_{\tilde{Q}_k} + \omega_1 \|Ax^k + By^k - b\|^2 + \omega_2 \|y^{k-1} - y^k\|^2_L \right\}.
\]

Then, it follows from convexity of \(F\) and the definition of \(w_T\) that

\[
(3.12) \quad F(w_T) \leq \frac{1}{T} \sum_{k=\kappa}^{\kappa+T} F(\tilde{w}^k).
\]

Dividing (3.11) by \(T\) and using (3.12), we obtain

\[
(3.13) \quad F(w_T) - F(w) + (w_T - w)^T J(w) \leq \frac{1}{T} \left[ -\sum_{k=\kappa}^{\kappa+T} \zeta^k + \frac{1}{2} \left\{ \|w - w^k\|^2_{\tilde{Q}_k} + \omega_1 \|Ax^k + By^k - b\|^2 + \omega_2 \|y^{k-1} - y^k\|^2_L \right\} \right].
\]

Let us now focus on the terms involving \(\zeta^k\). By assumption, the sequence \(\{m_k(m_k + 1)\eta_k\}\) is nondecreasing for \(k \in [\kappa, \kappa + T]\) and \(H > 0\), thus we have

\[
(3.14) \quad \sum_{k=\kappa}^{\kappa+T} \frac{2}{m_k(m_k + 1)\eta_k} \left( \|x - \tilde{x}^k\|^2_H - \|x - \tilde{x}^{k+1}\|^2_H \right) \leq \frac{2}{\eta_k} \frac{\|x - \tilde{x}^\kappa\|^2_H}{m_k(m_k + 1)}.
\]

Note that

\[
\delta_t = \nabla f(\tilde{x}_t) - d_t = \nabla f(\tilde{x}_t) - \nabla f_{\xi_t}(\tilde{x}_t) - e_t
\]

only depends on the index \(\xi_t\). So we have \(E[\delta_t] = 0\) since the random variable \(\xi_t \in \{1, 2, \ldots, N\}\) is chosen with uniform probability and \(E[e_t] = 0\). Also, since \(\tilde{x}_t\) depends on \(\xi_{t-1}, \xi_{t-2}, \ldots\), we have \(E[\delta_t, \xi_t, x] = 0\). By the assumption that \(E(\|\delta_t\|^2_H) \leq \sigma^2\), we have

\[
E \left[ \sum_{t=1}^{\kappa+T} t^2 \|\delta_t\|^2_H \right] \leq \frac{\sigma^2 m_k(m_k + 1)(2m_k + 1)}{6} \leq \frac{\sigma^2}{2} m_k^2(m_k + 1)
\]

since \(m_k \geq 1\). Combining these bounds for the terms in \(\zeta^k\) with the condition \(\eta_k \leq 1/(2\nu)\) gives

\[
-E \left[ \sum_{k=\kappa}^{\kappa+T} \zeta^k \right] \leq \frac{2}{\eta_k} \frac{\|x - \tilde{x}^\kappa\|^2_H}{m_k(m_k + 1)} + \frac{\sigma^2}{2} \sum_{k=\kappa}^{\kappa+T} \eta_k m_k.
\]
Finally, applying the expectation operator to (3.13) and substituting this bound into the $c^{k}$ term complete the proof. □

**Theorem 3.5.** Suppose the conditions in Theorem 3.4 hold. Let

$$
\eta_{k} = \min \left\{ \frac{c_{1}}{m_{k}(m_{k}+1)}, c_{2} \right\} \quad \text{and} \quad m_{k} = \max \{\lceil c_{3}k^{d} \rceil, m\},
$$

where $c_{1}, c_{2}, c_{3} > 0$, $\varrho \geq 1$ are constants and $m > 0$ is a given integer. Then, for every $w^{*} \in \Omega^{*}$, we have

$$
E_{u} \left[ F(w_{T}) - F(w^{*}) \right] = E_{\varrho}(T) = E_{u} \left[ \|Ax_{T} + By_{T} - b\| \right],
$$

where $E_{\varrho}(T) = O(1/T)$ for $\varrho > 1$ and $E_{\varrho}(T) = O(T^{-1} \log T)$ for $\varrho = 1$.

**Proof.** The proof is the same as Theorem 4.2 [3] and thus is omitted here. □

In practice, the matrix $M_{k}$ in Algorithm 1.1 could be adaptively adjusted as $M_{k} = \rho_{k} I$, where $\rho_{k} = \max \{\rho_{\text{min}}, \beta \delta_{2}^{k}/\delta_{1}^{k} \}$ with $\rho_{\text{min}} > 0$,

$$
\delta_{1}^{k} = \|x^{k} - x^{k-1}\|^{2} \quad \text{and} \quad \delta_{2}^{k} = \|A(x^{k} - x^{k-1})\|^{2}.
$$

Since $\delta_{2}^{k}/\delta_{1}^{k}$ is an underestimate of the largest eigenvalue of $A^{T}A$, to ensure convergence, the safeguard lower bound $\rho_{\text{min}}$ should be increased during the optimization if necessary. One may see [3, Remark 4.2] for more details.

**4. Further discussions.** In this part, we consider 3-block extensions of Algorithm 1.1 and its variance of a stochastic augmented Lagrangian method.

**4.1. A Stochastic ALM.** We first consider a stochastic augmented Lagrangian method, a variant of SAS-ADMM, to solve

$$
\min \{f(x) \mid Ax = b, \ x \in \mathcal{X}\},
$$

where $\mathcal{X} \subset \mathbb{R}^{n}$ is a closed convex subset, and $f$ is an average of $N$ smooth convex functions as defined in (1.1). Now, the augmented Lagrangian of (4.1) is

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

where $\mathcal{L}(x, \lambda) = f(x) - \lambda^{T}(Ax - b)$. Then, based on Algorithm 1.1, we can propose the following Accelerated Stochastic Augmented Lagrange Method (AS-ALM), Algorithm 4.1. Similar to SAS-ADMM, we can easily establish the following lemmas and algorithms on AS-ALM. However, in this case, the convergence region for the dual stepsize can be enlarged from $(0, (\sqrt{5} + 1)/2]$ of AS-ADMM [3] to $(0, 2]$, which is an extended interval from the convergence region $(0, 2)$ of standard deterministic ALM.

**Lemma 4.1.** Let $\{x^{k}\}$ be generated by Algorithm 4.1 and $\eta_{k} \in (0, 1/\nu)$. Then, the inequality (2.8) holds with

$$
\hat{\lambda}^{k} = \lambda^{k} - \beta (Ax^{k+1} - b).
$$

For the iterates generated by Algorithm 4.1, in this subsection we denote $w^{k} = (x^{k}, \hat{\lambda}^{k})$ and $\tilde{w}^{k} = (x^{k+1}, \hat{\lambda}^{k})$, where $\hat{\lambda}$ is defined in (4.2).
Parameters: $\beta > 0, s \in (0, 2]$ and $\mathcal{H} > 0$.

Initialization: $(x^0, \xi^0) \in \mathcal{X} \times \mathbb{R}^n := \Omega$ and $\xi^0 = x^0$.

For $k = 0, 1, \ldots$

Choose $m_k > 0$, $\eta_k > 0$ and $\mathcal{M}_k$ such that $\mathcal{M}_k - \beta A^T A \succeq 0$.

$h^k := -A^T \left[ \xi^k - \beta(Ax^k - b) \right]$.

$(x^{k+1}, \xi^{k+1}) = x_{\text{sub}} (x^k, \xi^k, h^k)$ with $x_{\text{sub}}$ given in Alg. 4.1.

$\lambda^{k+1} = \lambda^k - s \xi(Ax^{k+1} - b)$.

4.2. Three-block extensions.

Consider the following 3-block extension of problem (1.1)

$$
\min \quad F(w) := f(x) + g(y) + l(z)
$$

$$
\text{s.t.} \quad Kw := Ax + By + Cz = b, \quad x \in \mathcal{X}, \ y \in \mathcal{Y}, \ z \in \mathcal{Z}.
$$
where \( l \) is a closed convex function, \( C \in \mathbb{R}^{n \times n} \) is a given matrix, \( Z \subset \mathbb{R}^n \) is a simple closed convex subset, and the other functions and variables remain the same definitions as those in problem (1.1). In many applications, the additional function \( l \) in the objective function is often used to further promote some data structure different from the structure promoted by \( g \). For convenience, in this subsection, let us define
\[
Kw := Ax + By + Cz,
\]
and set \( w = (z; x; y; \lambda) \),
\[
(4.4) \quad w^k = \begin{pmatrix} x^k \\ y^k \\ z^k \\ \lambda^k \end{pmatrix}, \quad \tilde{w}^k := \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{z}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ z^{k+1} \\ \lambda^{k+1} \end{pmatrix} \text{ and } J(w) = \begin{pmatrix} -AT\lambda \\ -BT\lambda \\ -CT\lambda \\ Kw - b \end{pmatrix},
\]
where \( \tilde{\lambda}^k \) will be specified differently in the following two discussion cases.

### 4.2.1. Extension in Gauss-Seidel update

For this case, we need an assumption that \( C^TA = 0 \). Then SAS-ADMM can be directly extended to Algorithm 4.2 for solving the 3-block problem (4.3), where the variable updating order is \( z^{k+1} \rightarrow x^{k+1} \rightarrow y^{k+1} \rightarrow \lambda^{k+1} \) in a Gauss-Seidel scheme. For this case, we let
\[
(4.5) \quad \tilde{\lambda}^k = \lambda^k - \beta (Cz^{k+1} + Ax^{k+1} + By^k - b),
\]
\[
(4.6) \quad \tilde{v}^{k+1} = (x^{k+1}; y^{k+1}; \lambda^{k+1}) \quad \text{and} \quad \tilde{v}^k = (\tilde{x}^k; \tilde{y}^k; \tilde{\lambda}^k).
\]
Then, we have the following main theorem for the convergence of Algorithm 4.2.

**Theorem 4.4.** Assume \( C^TA = 0 \) and \( \eta_k \in (0, 1/\nu) \). Then, the iterates generated by Algorithm 4.2 satisfy \( \tilde{w}^k \in \Omega \) and
\[
F(w) - F(\tilde{w}^k) + (w - \tilde{w}^k)^T J(w) \geq \frac{1}{2} \left\{ \|v - v^{k+1}\|^2_{Q_k} + \|v - v^k\|^2_{Q_k} + \|v^k - \tilde{v}^k\|^2_{\tilde{G}_k} \right\} + \zeta_k
\]
for any \( w \in \Omega \), where \( \tilde{Q}_k, \tilde{G}_k \) and \( \zeta_k \) are given in Corollary 3.1 and (2.10), respectively. Moreover, we have
\[
\|v^k - \tilde{v}^k\|^2_{\tilde{G}_k} \geq \|x^k - x^{k+1}\|^2_{D_k} + \omega_0 \|Ax^{k+1} + By^{k+1} - b\|^2
\]
where $\omega_0, \omega_1, \omega_2 \geq 0$ are given in (3.7).

Proof. By the updates of $h^k$ and $\tilde{x}^k$ in Algorithm 4.2, it is easy to derive (2.8) as before. Then, according to the first-order optimality condition of $z$-subproblem and the assumption that $C^TA = 0$, we have

$$z^{k+1} \in Z, \, l(z) - l(z^{k+1}) + \langle z - z^{k+1}, p_x^k \rangle \geq 0, \, \forall z \in Z,$$

where

$$p_x^k = -C^T \lambda^k + \beta C^T (Cz^{k+1} + Ax^k + By^k - b)
= -C^T \lambda^k - \beta C^T A(x^{k+1} - x^k)
= -C^T \lambda^k.$$

Similarly, we have by the $y$-update that

$$y^{k+1} \in Y, \, g(y) - g(y^{k+1}) + \langle y - y^{k+1}, p_y^k \rangle \geq 0, \, \forall y \in Y,$$

where

$$p_y^k = -B^T \lambda^{k+\frac{1}{2}} + \beta B^T (Kw^{k+1} - b) + L(y^{k+1} - y^k)
= -B^T \lambda^{k+\frac{1}{2}} + B^T (\lambda^k - \tilde{\lambda}^k) + [L + \beta B^T B] (y^{k+1} - y^k)
= -B^T \lambda^k + \tau B^T (\lambda^k - \tilde{\lambda}^k) + [L + \beta B^T B] (y^{k+1} - y^k).$$

Besides, it follows from the updates of $\tilde{\lambda}^k$ that

$$\langle \lambda - \tilde{\lambda}^k, K\tilde{w}^k - b + \frac{1}{\beta} (\tilde{\lambda}^k - \lambda^k) - B (\tilde{y}^k - y^k) \rangle = 0, \, \forall \lambda \in \mathbb{R}^n.$$

Combining the above inequalities (4.7), (4.8), (4.9) with (2.8), we can get

$$F(w) - F(\tilde{w}^k) + (w - \tilde{w}^k)^T J(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q_k (v^k - \tilde{v}^k) + \zeta^k,$$

where $\zeta^k, Q_k$ are given by (2.10) and (2.12), respectively. Then, the rest proof will be similar to that of Corollary 3.1 and Theorem 3.3. \(\square\)

Based on the above Theorem 4.4, the convergence of Algorithm 4.2 with a sublinear convergence rate can be similarly established under the conditions of Theorem 3.5. Here, we omit the detailed proof for the sake of conciseness. To eliminate the condition $C^TA = 0$, we can consider a partially Jacobi update for the primal variables as discussed in the next subsection.

### 4.2.2. Extension in partially Jacobi update

In this subsection, let us consider Algorithm 4.3, where the block variables $y$ and $z$ are updated in a Jacobi fashion.

To establish the global convergence of Algorithm 4.3, we first have the following observations. By denoting

$$\tilde{\lambda}^k = \lambda^k - \beta \left(Ax^{k+1} + By^k + Cz^k - b\right)$$

$$+\omega_1 \left(\|Ax^{k+1} + By^{k+1} - b\|^2 - \|Ax^k + By^k - b\|^2\right)
+\omega_2 \left(\|y^k - y^{k+1}\|_L^2 - \|y^{k-1} - y^k\|_L^2\right),$$

where $\omega_0, \omega_1, \omega_2 \geq 0$ are given in (3.7).
Adding the above two inequalities (4.14) and (4.15), we can see (x_{k+1}, z_{k+1}) with x_{k+1} given in ALG.1.1.

\[ \lambda_{k+\frac{1}{2}} = \lambda_k - \tau \beta (A x_{k+1} + B y_k + C z_k - b). \]

and by using the first-order optimality condition of the y-subproblem, we have

\[ y_{k+1} \in Y, \quad g(y) - g(y_{k+1}) + \langle y - y_{k+1}, p_f^k \rangle \geq 0, \quad \forall y \in Y, \]

where

\[ p_f^k = -B^T \lambda_{k+\frac{1}{2}} + \beta B^T (A x_{k+1} + B y_{k+1} + C z_k - b) + L_1 (y_{k+1} - y_k) \]

\[ = -B^T \lambda_{k+\frac{1}{2}} + \beta B^T (A x_{k+1} + B y_k + C z_k - b) + (L_1 + \beta B^T B) (y_{k+1} - y_k) \]

and we use the relationship

\[ \lambda_{k+\frac{1}{2}} = \lambda_k - \tau (\lambda_k - \bar{\lambda}_k). \]

Looking at (4.12) directly and the first equation about p_f^k, we have

\[ g(y) - g(y_{k+1}) + \langle y - y_{k+1}, -B^T \lambda_{k+\frac{1}{2}} + \beta B^T (K w_{k+1} - b) \rangle - \beta B^T C (z_{k+1} - z_k) + L_1 (y_{k+1} - y_k) \geq 0. \]

Similarly, by the first-order optimality condition of the z-subproblem, we have

\[ l(z) - l(z_{k+1}) + \langle z - z_{k+1}, -B^T \lambda_{k+\frac{1}{2}} + \beta C^T (K w_{k+1} - b) - \beta C B (y_{k+1} - y_k) + L_2 (z_{k+1} - z_k) \geq 0. \]

Adding the above two inequalities (4.14) and (4.15), we can see (y_{k+1}, z_{k+1}) satisfies the first-order optimality condition, hence is a solution, of the following problem

\[ (y_{k+1}, z_{k+1}) \in \arg \min_{y \in Y, z \in Z} g(y) + l(z) + \frac{\beta}{2} \left( A x_{k+1} + B y + C z - b - \frac{\lambda_{k+\frac{1}{2}}}{\beta} \right)^2 + \frac{1}{2} \left\| (y - y_{k}, z - z_{k+1}) \right\|^2_T. \]
where

\[
(4.17) \quad \mathcal{L} = \begin{bmatrix} L_1 & -\beta B^T C \\ -\beta^T C B & L_2 \end{bmatrix}.
\]

Hence, by considering \((y, z)\) as one block variable, Algorithm 4.3 is essentially a particular version of Algorithm 1.1 for solving a 2-block problem with \(L\) and \(B\) being replaced by \(\mathcal{L}\) and \((B, C)\), respectively.

From the above observations, we can establish the following properties of Algorithm 4.3 straightforwardly.

**Theorem 4.5.** The iterates generated by Algorithm 4.3 satisfy

\[
F(w) - F(\bar{w}^k) + \langle w - \bar{w}^k, J(w) \rangle \geq (w - \bar{w}^k)^T Q_k (w^k - \bar{w}^k) + \zeta^k
\]

for any \(w \in \Omega\), where \(\zeta^k\) is given by (2.10),

\[
(4.18) \quad Q_k = \begin{bmatrix} \mathcal{D}_k & L_1 + \beta B^T B \\ -B & L_2 + \beta C^T C \end{bmatrix} - \tau B^T B - B \tau C^T C - \frac{\tau}{\beta} I
\]

**Proof.** Notice that

\[
L + \beta (B, C)^T (B, C) = \begin{bmatrix} L_1 + \beta B^T B \\ L_2 + \beta C^T C \end{bmatrix} - \tau B^T B - B \tau C^T C - \frac{\tau}{\beta} I
\]

So, replacing \(L\) and \(B\) in Theorem 2.2 by \(L\) and \((B, C)\), respectively, this theorem directly follows from Theorem 2.2.

Similarly, identifying \((L, B)\) in (3.2) by \((L, B)\), respectively, it follows from Corollary 3.1 that

\[
\begin{align*}
F(w) - F(\bar{w}^k) + \langle w - \bar{w}^k, J(w) \rangle & \geq \frac{1}{2} \left\{ ||w - w^{k+1}||_{Q_k}^2 - ||w - w^k||_{Q_k}^2 + ||w^k - \bar{w}^k||_{G_k}^2 \right\} + \zeta^k,
\end{align*}
\]

where \(\zeta^k\) is given by (2.10) and

\[
(4.20) \quad \bar{Q}_k = \begin{bmatrix} \mathcal{D}_k & L_1 + \frac{(1 - s)\beta B^T B}{\tau + s} \tau C^T B \\ L_1 + \frac{(1 - s)\beta C^T C}{\tau + s} \tau B^T B - \frac{\tau}{\tau + s} B & \tau + s \end{bmatrix},
\]

\[
(4.21) \quad \bar{G}_k = \begin{bmatrix} \mathcal{D}_k & L_1 + (s)\beta B^T B \\ L_1 + (s)\beta C^T C - s \beta B^T B \\ (s - 1)B \\ \tau \tau C^T C \end{bmatrix} - \frac{2}{\tau + s} I
\]

Then, we have the following theorem on a lower bound of \(\|w^k - \bar{w}^k\|_{\tilde{G}_k}^2\).

**Theorem 4.6.** Suppose there exist \(\gamma_1 > 0\) and \(\gamma_2 > 0\) with \(\gamma_1 \gamma_2 \geq 1\) such that

\[
(4.22) \quad L_1 \geq \gamma_1 \beta B^T B \quad \text{and} \quad L_2 \geq \gamma_2 \beta C^T C.
\]
then, for any \((\tau, s) \in \Delta\) defined in (1.3), we have \(\tilde{Q}_k\) defined in (4.20) is positive semidefinite and

\[
\|w^k - \tilde{w}^k\|_{\tilde{G}_k}^2 \geq \|x^k - x^{k+1}\|_{D_k}^2 + \omega_0 \|Kw^{k+1} - b\|^2
\]

\[
+ \omega_1 \left( \|Kw^{k+1} - b\|^2 - \|Kw^k - b\|^2 \right)
\]

\[
+ \omega_2 \left( \left\| \left( \begin{array}{c} y^k - y^{k+1} \\ z^k - z^{k+1} \end{array} \right) \right\|_T^2 - \left\| \left( \begin{array}{c} y^k - y^{k-1} \\ z^k - z^{k-1} \end{array} \right) \right\|_T^2 \right),
\]

where \(\omega_0, \omega_1, \omega_2 \geq 0\) is defined in (3.7) and \(\bar{T}\) is defined in (4.17).

**Proof.** First, since \(L_1 \geq \gamma_1 \beta B^T B\) and \(L_2 \geq \gamma_2 \beta C^T C\), it follows from \(\gamma_1 > 0, \gamma_2 > 0\) and \(\gamma_1 \gamma_2 \geq 1\) that

\[
\bar{T} = \left[ \begin{array}{cc} L_1 & -\beta B^T C \\ -\beta C^T B & L_2 \end{array} \right] \succeq \beta \left[ \begin{array}{cc} \gamma_1 B^T B & -B^T C \\ -C^T B & \gamma_2 C^T C \end{array} \right] \succeq 0.
\]

By Lemma 3.2, we have \(\tilde{Q}_k\) defined in (4.20) is positive semidefinite if

\[
\bar{T} \succeq (\tau - 1) \beta (B, C)^T (B, C).
\]

Since \(\tau \leq 1\) for any \((\tau, s) \in \Delta\), we have \(0 \succeq (\tau - 1) \beta (B, C)^T (B, C)\). Therefore, we have from (4.24) that (4.25) holds and therefore, \(\tilde{Q}_k\) defined in (4.20) is positive semidefinite. Furthermore, it follows from Theorem 3.3 that (4.23) holds as long as \(\bar{T} \succeq 0\) which is verified by (4.24).

Now, defining \(w_T := \frac{1}{T} \sum_{k=1}^{k=\infty} \tilde{w}^k\) for some integers \(T > 0\) and \(\kappa > 0\), under the same conditions in Theorem 3.4, by Theorem 3.4 and Theorem 4.6, we will have

\[
\mathbb{E} \left[ F(w_T) - F(w) + (w_T - w)^T J(w) \right] 
\leq \frac{1}{2T} \left\{ \sigma^2 \sum_{k=\kappa}^{k=\infty} \eta_k m_k + \frac{4}{m_\kappa (m_\kappa + 1) \eta_\kappa} \|x - \tilde{x}\|_H^2 + \|w - w^\kappa\|_{\tilde{Q}_k}^2 \right.
\]

\[
+ \omega_1 \|Kw^\kappa - b\|^2 + \omega_2 \left( \left\| \left( \begin{array}{c} y^\kappa - y^{\kappa-1} \\ z^\kappa - z^{\kappa-1} \end{array} \right) \right\|_T^2 \right),
\]

where \(\omega_1 \geq 0\) and \(\omega_2 \geq 0\) given in (3.7). So, by the choice of the parameters \((\eta_k, m_k)\) chosen in Theorem 3.5, we can obtain

\[
\mathbb{E} \left[ F(w_T) - F(w^\star) \right] = E_\theta(T) = \mathbb{E} \left[ \|Ax_T + By_T + Cz_T - b\| \right],
\]

where \(E_\theta(T) = \mathcal{O}(1/T)\) for the parameter \(\theta > 1\) and \(E_\theta(T) = \mathcal{O}(T^{-1} \log T)\) for \(\theta = 1\).

**Remark 4.1.** Observing from the above analysis, Algorithm 4.3 could be in fact generalized to Algorithm 4.4 for solving the multi-block separable convex optimization:

\[
\min \ F(w) := f(x) + \sum_{i=1}^{q} g_i(y_i)
\]

s.t. \(Kw := Ax + \sum_{i=1}^{q} B_i y_i = b\),

\(x \in \mathcal{X}, \ y_i \in \mathcal{Y}_i, \ i = 1, 2, \ldots, q,\)

where \(f\) has the same definition as in (1.1), \(g_i : \mathcal{Y}_i \to \mathbb{R} \cup \{\infty\}\) is a convex but possibly nonsmooth function, \(B_i \in \mathbb{R}^{n \times n_i}\) and \(\mathcal{Y}_i \subset \mathbb{R}^{n_i}\) is a closed convex subset. The
Parameters: \( \beta > 0, (\tau, s) \in \Delta \) and \( L_j \geq (q - 1)\beta B_j^T B_j \) for all \( j = 1, \ldots, q \).

Initialization: \( (x^0, y^0, \lambda^0) \in \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^n \), \( \tilde{x}^0 = x^0 \).

For \( k = 0, 1, \ldots \),

Choose \( m_k > 0 \), \( \eta_k > 0 \) and \( M_k \) such that \( M_k - \beta A^T A \succeq 0 \).

\[
h^k := -A^T \left[ \lambda^k - \beta (Ax^k + By^k - b) \right].
\]

\((x^{k+1}, \tilde{x}^{k+1}) = x_{sub} \) with \( x_{sub} \) given in ALG.1.1.

\[
\lambda^{k+1} = \lambda^k - \tau \beta (Ax^{k+1} + By^{k+1} - b).
\]

For \( i = 1, 2, \ldots, q \),

\[
y_{i+1}^{k+1} = \arg \min_{y_i \in \mathcal{Y}_i} g_i(y_i) + \frac{\beta}{2} \left\| Ax^{k+1} + \sum_{i=1}^{q} B_i y_i - b - \lambda^{k+1}_{i} \right\|_2^2 + \frac{\tau}{2} \left\| y_i - y_i^{k} \right\|_{L_i}^2.
\]

end

end

ALG. 4.4. Muts-block extension of SAS-ADMM in partially Jacobi update

The convergence of Algorithm 4.4 can be analogously established with proper modifications on the convergence proof of Algorithm 4.3. Here, we give a simple analysis on the convergence of Algorithm 4.4. We first denote \( g(y) = \sum_{i=1}^{q} g_i(y_i), B = (B_1, \ldots, B_q), y = (y_1; \ldots; y_q), y^k = (y^k_1; \ldots; y^k_q) \) and \( \mathcal{Y} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_q \). Then, by the first-order optimality condition of \( y_i \)-subproblem, we have \( y_{i+1}^{k+1} \in \mathcal{Y}_i \) and

\[
g_i(y_i) - g_i(y_i^{k+1}) + \left( y_i - y_i^{k+1}, -B_i^T \lambda^{k+1}_i + \beta B_i^T (Kw^{k+1} - b) - \beta \sum_{j \neq i} B_j^T B_j (y_j^{k+1} - y_j) + L_i (y_i^{k+1} - y_i^k) \right) \geq 0, \quad \forall y_i \in \mathcal{Y}_i.
\]

After adding the above inequality from \( i = 1 \) to \( q \), we can see \( y^{k+1} \) satisfies the first-order optimality condition, hence is a solution, of the following problem:

\[
y^{k+1} = \arg \min_{y \in \mathcal{Y}} \left\{ g(y) + \frac{\beta}{2} \left\| y - y_k \right\|_{\tilde{L}}^2 \right\}
\]

where

\[
\tilde{L} = \begin{bmatrix}
L_1 & -\beta B_1^T B_2 & \cdots & -\beta B_1^T B_q \\
-\beta B_2^T B_1 & L_2 & \cdots & -\beta B_2^T B_q \\
\vdots & \vdots & \ddots & \vdots \\
-\beta B_q^T B_1 & -\beta B_q^T B_2 & \cdots & L_q
\end{bmatrix}.
\]

So, by a similar analysis to Algorithm 4.3, the inequality (4.19) holds with
and

\[
G_k = \begin{bmatrix}
D_k & L_1 + (1-s)\beta B_1^T B_1 & -s\beta B_1^T B_2 & \cdots & -s\beta B_1^T B_q & (s-1)B_1^T \\
-s\beta B_2^T B_1 & L_2 + (1-s)\beta B_2^T B_2 & \cdots & -s\beta B_2^T B_q & (s-1)B_2^T \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
-s\beta B_q^T B_1 & -s\beta B_q^T B_2 & \cdots & L_q + (1-s)\beta B_q^T B_q & (s-1)B_q^T \\
(s-1)B_1 & (s-1)B_2 & \cdots & (s-1)B_q & (s-1)I
\end{bmatrix}
\]

If \( L_i \geq (q - 1)\beta B_i^T B_i \) for \( i = 1, \ldots, q \), then for any \((r,s) \in \Delta \) given by (1.3), the above matrix \( \hat{Q}_k \) is positive semidefinite and

\[
\|w^k - \hat{w}^k\|_{\hat{Q}_k}^2 \geq \|x^k - x^{k+1}\|_{D_k}^2 + \omega_1 \|Kw^{k+1} - b\|^2 + \omega_2 \left( \|y^{k+1} - y^k\|^2_L - \|y^k - y^{k-1}\|^2_L \right),
\]

where \( \omega_0, \omega_1, \omega_2 \geq 0 \) is defined in (3.7) and \( \bar{L} \) is defined in (4.27). The above discussions imply that Algorithm 4.4 converges with the same convergence results as Algorithm 4.3. This remark further shows that Algorithm 4.4 is a stochastic extension of our previous deterministic GS-ADMM [1] for solving the grouped multi-block separable convex optimization problem.

5. Numerical experiments. In this section, we apply the proposed algorithm to solve the following graph-guided fused lasso problem in machine learning:

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

where \( f_j(x) = \log(1 + \exp(-b_j a_j^T x)) \) denotes the logistic loss function on the feature-label pair \((a_j, b_j) \in \mathbb{R}^\ell \times \{-1, 1\}\), \( N \) is the data size, \( \mu > 0 \) is a given regularization parameter, and \( A = [G; I] \) is a matrix encoding the feature sparsity pattern. Here, \( G \) is the sparsity pattern of the graph that is obtained by sparse inverse covariance estimation [8]. Introducing an auxiliary variable \( y \), the above problem is equivalent to the problem

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

which has the format of our model (1.1). In addition, it can be easily verified that the Assumptions 2.1-2.2 hold. Since the coefficient matrix of the \( y \)-variable in the constraints of (5.1) is \(-I\), the \( y \)-subproblem will have a closed-form solution by simply setting \( L = 0 \) in the Algorithm 1.1. Otherwise, the linearization techniques discussed in (1.4) on choosing \( L \) can be applied to obtain a closed-form solution of the \( y \)-subproblem. With \( L = 0 \), the subproblems in Algorithm 1.1 would have the following closed-form solution:

\[
\begin{align*}
\hat{x}_{t+1} &= \left[ \gamma_t H + M_k \right]^{-1} \left[ \gamma_t H \hat{x}_t + M_k x^k - d_t - h^k \right], \\
y_{t+1} &= \text{Shrink} \left( \frac{\mu}{\gamma_t}, Ax^{k+1} - \frac{X^{k+\frac{1}{2}}}{\beta} \right).
\end{align*}
\]
Here, $\text{Shrink}(\cdot, \cdot)$ denotes the soft shrinkage operator and can be evaluated using the MATLAB built-in function "\texttt{wthresh}".

In the numerical experiments, the penalty parameter in SAS-ADMM is taken as $\beta = 0.001$. The other parameters as well as the vector $e_i$ in SAS-ADMM (i.e. Algorithm 1.1) are chosen the same way as that used in [3, Section 7.1]. Same as in [3], we use

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

to denote the relative objective error and the constraint error. Here, $F^*$ is the approximate optimal objective function value obtained by running Algorithm 1.1 for more than 10 minutes. To measure the performance of an algorithm, we plot the maximum of the relative objective error and the constraint error, that is

$$\text{Opt.err} = \max(\text{Obj.err}, \text{Equ.err}),$$

against the CPU time used. All experiments are implemented in MATLAB R2018a (64-bit) with the same starting point $(x^0, y^0, \lambda^0) = (0, 0, 0)$ and performed on a PC with Windows 10 operating system, with an Intel i7-8700K CPU and 16GB RAM.

We compare the numerical performance of the proposed algorithm SAS-ADMM using stepsizes $(\tau, s) = (0.9, 1.09)$ as suggested in [1, Section 5.2.2] and AS-ADMM [3] for solving problem (5.1) on the dataset \textit{mnist} (including 11,791 samples and 784 features) downloaded from LIBSVM website. The regularization parameter $\mu$ in (5.1) is set as $10^{-5}$. For both SAS-ADMM and AS-ADMM, we plot the error associated with the iterates over the first 1/3 of the total CPU time budget, followed by the error associated with the ergodic iterates over the last 2/3 of the budget. We make 20 successive runs of each algorithm under the CPU time budgets 120s and 200s, and the average comparison results are shown in Figure 5.1. Here, we only compare SAS-ADMM with AS-ADMM since in [3] AS-ADMM was shown better than other state-of-the-art methods. Note that Opt.err has a big drop at around 1/3 of the CPU time budget, the point where the ergodic iterates are started to use for reporting the objective value. From Figure 5.1, we can see that SAS-ADMM initially performs
worse than AS-ADMM at the beginning iterations. But after the first $1/3$ of the total CPU time budget, the SAS-ADMM eventually seems to perform slightly better than AS-ADMM.

6. Conclusion. We have studied the convergence and the ergodic iteration complexity of a symmetric accelerated stochastic alternating direction method of multipliers, called SAS-ADMM, whose dual variables are symmetrically updated. We give the specific region where the dual stepsizes should be restricted to for ensuring global convergence. Under proper choice of the algorithm parameters, SAS-ADMM is proved convergent in expectation with the worst-case $O(1/T)$ convergence rate, where $T$ represents the number of outer iterations. We also discuss an accelerated stochastic augmented Lagrangian algorithm (AS-ALM) as a variance of SAS-ADMM. Furthermore, we generalize SAS-ADMM to solve the multi-block separable convex optimization problem in both a Gauss-Seidel or a partially Jacobi scheme. Our preliminary experiments show that by symmetrically updating the dual variables using a more flexible region, SAS-ADMM could outperform AS-ADMM, which only updates the dual variable once, for solving some structured optimization problems arising in machine learning.

REFERENCES


