Noname manuscript No.
(will be inserted by the editor)

Generalized Symmetric ADMM for Separable Convex Optimization

Jianchao Bai · Jicheng Li · Fengmin Xu · Hongchao Zhang

Received: date / Accepted: date

Abstract The Alternating Direction Method of Multipliers (ADMM) has been proved to be effective for solving separable convex optimization subject to linear constraints. In this paper, we propose a Generalized Symmetric ADMM (GS-ADMM), which updates the Lagrange multiplier twice with suitable stepsizes, to solve the multi-block separable convex programming. This GS-ADMM partitions the data into two group variables so that one group consists of p block variables while the other has q block variables, where $p \geq 1$ and $q \geq 1$ are two integers. The two grouped variables are updated in a *Gauss-Seidel* scheme, while the variables within each group are updated in a *Jacobi* scheme, which would make it very attractive for a big data setting. By adding proper proximal terms to the subproblems, we specify the domain of the stepsizes to guarantee that GS-ADMM is globally convergent with a worst-case $\mathcal{O}(1/t)$ ergodic convergence rate. It turns out that our convergence domain of the stepsizes is significantly larger than other convergence domains in the literature. Hence, the GS-ADMM is more flexible and attractive on choosing and using larger stepsizes of the dual variable. Besides, two special cases of GS-ADMM, which allows using zero penalty terms, are also discussed and analyzed. Compared with several state-of-the-art methods, preliminary numerical experiments on solving a sparse matrix minimization problem in the statistical learning show that our proposed method is effective and promising.

The work was supported by the National Science Foundation of China under grants 11671318 and 11571178, the Science Foundation of Fujian province of China under grant 2016J01028, and the National Science Foundation of U.S.A. under grant 1522654.

Jianchao Bai
School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, P.R. China
E-mail: bjc1987@163.com

Jicheng Li
School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, P.R. China
E-mail: jcli@mail.xjtu.edu.cn

Fengmin Xu
School of Economics and Finance, Xi'an Jiaotong University, Xi'an 710049, P.R. China
E-mail: fengminxu@mail.xjtu.edu.cn

Hongchao Zhang
Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803-4918, U.S.A.
E-mail: hozhang@math.lsu.edu

Keywords Separable convex programming · Multiple blocks · Parameter convergence domain · Alternating direction method of multipliers · Global convergence · Complexity · Statistical learning

Mathematics Subject Classification (2000) 65C60 · 65E05 · 68W40 · 90C06

1 Introduction

We consider the following grouped multi-block separable convex programming problem

$$\begin{array}{ll} \min & \sum_{i=1}^{p} f_i(x_i) + \sum_{j=1}^{q} g_j(y_j) \\ \text{s.t.} & \sum_{i=1}^{p} A_i x_i + \sum_{j=1}^{q} B_j y_j = c, \\ & x_i \in \mathcal{X}_i,\ i=1,\cdots,p, \\ & y_j \in \mathcal{Y}_j,\ j=1,\cdots,q, \end{array} \tag{1}$$

where $f_i(x_i):\mathcal{R}^{m_i}\to\mathcal{R},\ g_j(y_j):\mathcal{R}^{d_j}\to\mathcal{R}$ are closed and proper convex functions (possibly nonsmooth); $A_i\in\mathcal{R}^{n\times m_i}, B_j\in\mathcal{R}^{n\times d_j}$ and $c\in\mathcal{R}^n$ are given matrices and vectors, respectively; $\mathcal{X}_i\subset\mathcal{R}^{m_i}$ and $\mathcal{Y}_j\subset\mathcal{R}^{d_j}$ are closed convex sets; $p\geq 1$ and $q\geq 1$ are two integers. Throughout this paper, we assume that the solution set of the problem (1) is nonempty and all the matrices A_i, $i=1,\cdots,p$, and B_j, $j=1,\cdots,q$, have full column rank. And in the following, we denote $\mathcal{A}=(A_1,\cdots,A_p), \mathcal{B}=(B_1,\cdots,B_q)$, $\mathbf{x}=(x_1^\mathsf{T},\cdots,x_p^\mathsf{T})^\mathsf{T}$, $\mathbf{y}=(y_1^\mathsf{T},\cdots,y_q^\mathsf{T})^\mathsf{T}$, $\mathcal{X}=\mathcal{X}_1\times\mathcal{X}_2\times\cdots\mathcal{X}_p$, $\mathcal{Y}=\mathcal{Y}_1\times\mathcal{Y}_2\times\cdots\mathcal{Y}_q$ and $\mathcal{M}=\mathcal{X}\times\mathcal{Y}\times\mathcal{R}^n$.

In the last few years, the problem (1) has been extensively investigated due to its wide applications in different fields, such as the sparse inverse covariance estimation problem [21] in finance and statistics, the model updating problem [4] in the design of vibration structural dynamic system and bridges, the low rank and sparse representations [19] in image processing and so forth. One standard way to solve the problem (1) is the classical Augmented Lagrangian Method (ALM) [10], which minimizes the following augmented Lagrangian function

$$\mathcal{L}_\beta(\mathbf{x},\mathbf{y},\lambda) = L(\mathbf{x},\mathbf{y},\lambda) + \frac{\beta}{2}\|\mathcal{A}\mathbf{x}+\mathcal{B}\mathbf{y}-c\|^2,$$

where $\beta>0$ is a penalty parameter for the equality constraint and

$$L(\mathbf{x},\mathbf{y},\lambda) = \sum_{i=1}^{p} f_i(x_i) + \sum_{j=1}^{q} g_j(y_j) - \langle\lambda, \mathcal{A}\mathbf{x}+\mathcal{B}\mathbf{y}-c\rangle \tag{2}$$

is the Lagrangian function of the problem (1) with the Lagrange multiplier $\lambda\in\mathcal{R}^n$. Then, the ALM procedure for solving (1) can be described as follows:

$$\begin{cases} \left(\mathbf{x}^{k+1},\mathbf{y}^{k+1}\right) = \arg\min\left\{\mathcal{L}_\beta\left(\mathbf{x},\mathbf{y},\lambda^k\right) \mid \mathbf{x}\in\mathcal{X},\mathbf{y}\in\mathcal{Y}\right\}, \\ \lambda^{k+1} = \lambda^k - \beta(\mathcal{A}\mathbf{x}^{k+1}+\mathcal{B}\mathbf{y}^{k+1}-c). \end{cases}$$

However, ALM does not make full use of the separable structure of the objective function of (1) and hence, could not take advantage of the special properties of the component objective functions f_i and g_j in (1). As a result, in many recent real applications involving big data, solving the subproblems of ALM becomes very expensive.

One effective approach to overcome such difficulty is the Alternating Direction Method of Multipliers (ADMM), which was originally proposed in [8] and could be regarded as a splitting version of ALM. At each iteration, ADMM first sequentially optimize over one block variable while fixing all the other block variables, and then follows by updating the Lagrange multiplier. A natural extension of ADMM for solving the multi-block case problem (1) takes the following iterations:

$$\begin{cases} \text{For } i = 1, 2, \cdots, p, \\ \qquad x_i^{k+1} = \arg\min_{x_i \in \mathcal{X}_i} \mathcal{L}_\beta \left(x_1^{k+1}, \cdots, x_{i-1}^{k+1}, x_i, x_{i+1}^k, \cdots, x_p^k, \mathbf{y}^k, \lambda^k \right), \\ \text{For } j = 1, 2, \cdots, q, \\ \qquad y_j^{k+1} = \arg\min_{y_j \in \mathcal{Y}_j} \mathcal{L}_\beta \left(\mathbf{x}^{k+1}, y_1^{k+1}, \cdots, y_{j-1}^{k+1}, y_j, y_{j+1}^k, \cdots, y_q^k, \lambda^k \right), \\ \lambda^{k+1} = \lambda^k - \beta \left(\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c \right). \end{cases} \tag{3}$$

Obviously, the scheme (3) is a serial algorithm which uses the newest information of the variables at each iteration. Although the above scheme was proved to be convergent for the two-block, i.e., $p = q = 1$, separable convex minimization (see [11]), as shown in [3], the direct extension of ADMM (3) for the multi-block case, i.e., $p + q \geq 3$, without proper modifications is not necessarily convergent. Another natural extension of ADMM is to use the Jacobian fashion, where the variables are updated simultaneously after each iteration, that is,

$$\begin{cases} \text{For } i = 1, 2, \cdots, p, \\ \qquad x_i^{k+1} = \arg\min_{x_i \in \mathcal{X}_i} \mathcal{L}_\beta \left(x_1^{k}, \cdots, x_{i-1}^{k}, x_i, x_{i+1}^k, \cdots, x_p^k, \mathbf{y}^k, \lambda^k \right), \\ \text{For } j = 1, 2, \cdots, q, \\ \qquad y_j^{k+1} = \arg\min_{y_j \in \mathcal{Y}_j} \mathcal{L}_\beta \left(\mathbf{x}^{k}, y_1^{k}, \cdots, y_{j-1}^{k}, y_j, y_{j+1}^k, \cdots, y_q^k, \lambda^k \right), \\ \lambda^{k+1} = \lambda^k - \beta \left(\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c \right). \end{cases} \tag{4}$$

As shown in [13], however, the Jacobian scheme (4) is not necessarily convergent either. To ensure the convergence, He et al. [14] proposed a novel ADMM-type splitting method that by adding certain proximal terms, allowed some of the subproblems to be solved in parallel, i.e., in a Jacobian fashion. And in [14], some sparse low-rank models and image painting problems were tested to verify the efficiency of their method.

More recently, a Symmetric ADMM (S-ADMM) was proposed by He et al. [16] for solving the two-block separable convex minimization, where the algorithm performs the following updating scheme:

$$\begin{cases} \mathbf{x}^{k+1} = \arg\min \left\{ \mathcal{L}_\beta \left(\mathbf{x}, \mathbf{y}^k, \lambda^k \right) \mid \mathbf{x} \in \mathcal{X} \right\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \tau\beta \left(\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^k - c \right), \\ \mathbf{y}^{k+1} = \arg\min \left\{ \mathcal{L}_\beta (\mathbf{x}^{k+1}, \mathbf{y}, \lambda^{k+\frac{1}{2}}) \mid \mathbf{y} \in \mathcal{Y} \right\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta \left(\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c \right), \end{cases} \tag{5}$$

and the stepsizes (τ, s) were restricted into the domain

$$\mathcal{H} = \left\{ (\tau, s) \mid s \in (0, (\sqrt{5}+1)/2),\ \tau + s > 0,\ \tau \in (-1, 1),\ |\tau| < 1 + s - s^2 \right\} \tag{6}$$

in order to ensure its global convergence. The main improvement of [16] is that the scheme (5) largely extends the domain of the stepsizes (τ, s) of other ADMM-type methods [12]. What's more, the numerical performance of S-ADMM on solving the widely used basis pursuit model and the total-variational image debarring model significantly outperforms the original ADMM in both the CPU time and the number of iterations. Besides, Gu, et al.[9] also studied a semi-proximal-based strictly contractive Peaceman-Rachford splitting method, that is (5) with two additional proximal penalty terms for the x and y update. But their method has a nonsymmetric convergence domain of the stepsize and still focuses on the two-block case problem, which limits its applications for solving large-scale problems with multiple block variables.

Mainly motivated by the work of [14,16,9], we would like to generalize S-ADMM with more wider convergence domain of the stepsizes to tackle the multi-block separable convex programming model (1), which more frequently appears in recent applications involving big data [2,20]. Our algorithm framework can be described as follows:

$$(\textbf{GS-ADMM}) \quad \begin{cases} \text{For } i = 1, 2, \cdots, p, \\ \quad x_i^{k+1} = \arg\min\limits_{x_i \in \mathcal{X}_i} \mathcal{L}_\beta(x_1^k, \cdots, x_i, \cdots, x_p^k, \mathbf{y}^k, \lambda^k) + P_i^k(x_i), \\ \quad \text{where } P_i^k(x_i) = \frac{\sigma_1 \beta}{2} \left\| A_i(x_i - x_i^k) \right\|^2, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \tau\beta(\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^k - c), \\ \\ \text{For } j = 1, 2, \cdots, q, \\ \quad y_j^{k+1} = \arg\min\limits_{y_j \in \mathcal{Y}_j} \mathcal{L}_\beta(\mathbf{x}^{k+1}, y_1^k, \cdots, y_j, \cdots, y_q^k, \lambda^{k+\frac{1}{2}}) + Q_j^k(y_j), \\ \quad \text{where } Q_j^k(y_j) = \frac{\sigma_2 \beta}{2} \left\| B_j(y_j - y_j^k) \right\|^2, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c). \end{cases} \tag{7}$$

In the above Generalized Symmetric ADMM (**GS-ADMM**), τ and s are two stepsize parameters satisfying

$$(\tau, s) \in \mathcal{K} = \left\{ (\tau, s) \mid \tau + s > 0,\ \tau \leq 1,\ -\tau^2 - s^2 - \tau s + \tau + s + 1 > 0 \right\}, \tag{8}$$

and $\sigma_1 \in (p-1, +\infty), \sigma_2 \in (q-1, +\infty)$ are two proximal parameters[1] for the regularization terms $P_i^k(\cdot)$ and $Q_j^k(\cdot)$. He and Yuan[15] also investigated the above GS-ADMM (7) but restricted the stepsize $\tau = s \in (0, 1)$, which does not exploit the advantages of using flexible stepsizes given in (8) to improve its convergence.

Major contributions of this paper can be summarized as the following four aspects:

- Firstly, the new GS-ADMM could deal with the multi-block separable convex programming problem (1), while the original S-ADMM in [16] only works for the two block case and may not be convenient for solving large-scale problems. In addition, the convergence domain $\mathcal{K}$ for the stepsizes (τ, s) in (8), shown in Fig. 1, is significantly larger than the domain $\mathcal{H}$ given in (6) and the convergence domain in [9,15]. For example, the stepsize s can be arbitrarily close to $5/3$ when the stepsize τ is close to $-1/3$. Moreover, the above domain in (8) is later enlarged to a symmetric domain $\mathcal{G}$ defined in (73), shown in Fig. 2. Numerical experiments in Sec. 5.2.1 also validate that using more flexible and relatively larger stepsizes (τ, s) can often improve the convergence speed of GS-ADMM. On the other hand, we can see that when $\tau = 0$, the stepsize s can be chosen in the interval $(0, (\sqrt{5}+1)/2)$, which was firstly suggested by Fortin and Glowinski in [5,7].
- Secondly, the global convergence of GS-ADMM as well as its worst-case $\mathcal{O}(1/t)$ ergodic convergence rate are established. What's more, the total $p + q$ block variables are partitioned into two grouped variables. While a *Gauss-Seidel* fashion is taken between the two grouped variables, the block variables within each group are updated in a *Jacobi* scheme. Hence, parallel computing can be implemented for updating the variables within each group, which could be critical in some scenarios for problems involving big data.

[1] Note that these two parameters are strictly positive in (7). In Section 4, however, we analyze two special cases of GS-ADMM allowing either σ_1 or σ_2 to be zero.

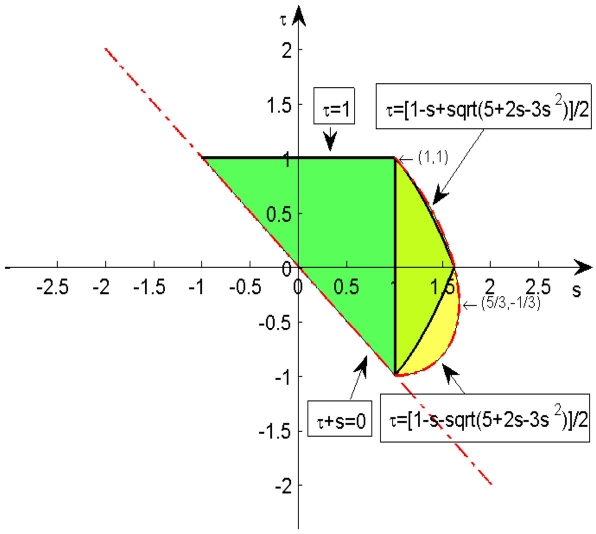

Fig 1. Stepsize region $\mathcal{K}$ of GS-ADMM

– Thirdly, we discuss two special cases of GS-ADMM, which is (7) with $p \geq 1, q = 1$ and $\sigma_2 = 0$ or with $p = 1, q \geq 1$ and $\sigma_1 = 0$. These two special cases of GS-ADMM were not discussed in [15] and in fact, to the best of our knowledge, they have not been studied in the literature. We show the convergence domain of the stepsizes (τ, s) for these two cases is still $\mathcal{K}$ defined in (8) that is larger than $\mathcal{H}$.
– Finally, numerical experiments are performed on solving a sparse matrix optimization problem arising from the statistical learning. We have investigated the effects of the stepsizes (τ, s) and the penal parameter β on the performance of GS-ADMM. And our numerical experiments demonstrate that by properly choosing the parameters, GS-ADMM could perform significantly better than other recently quite popular methods developed in [1, 14, 17, 23].

The paper is organized as follows. In Section 2, some preliminaries are given to reformulate the problem (1) into a variational inequality and to interpret the GS-ADMM (7) as a prediction-correction procedure. Section 3 investigates some properties of $\|\mathbf{w}^k - \mathbf{w}^*\|_H^2$ and provides a lower bound of $\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\|_G^2$, where H and G are some particular symmetric matrices. Then, we establish the global convergence of GS-ADMM and show its convergence rate in an ergodic sense. In Section 4, we discuss two special cases of GS-ADMM, in which either the penalty parameters σ_1 or σ_2 is allowed to be zero. Some preliminary numerical experiments are done in Section 5. We finally make some conclusions in Section 6.

1.1 Notation

Denoted by $\mathcal{R}, \mathcal{R}^n, \mathcal{R}^{m\times n}$ be the set of real numbers, the set of n dimensional real column vectors and the set of $m \times n$ real matrices, respectively. For any $x, y \in \mathcal{R}^n$, $\langle x, y\rangle = x^\mathsf{T} y$ denotes their inner product and $\|x\| = \sqrt{\langle x, x\rangle}$ denotes the Euclidean norm of x, where the superscript $^{\mathsf{T}}$ is the transpose. Given a symmetric matrix G, we define $\|x\|_G^2 = x^{\mathsf{T}}Gx$. Note that with this convention, $\|x\|_G^2$ is not necessarily nonnegative unless G is a positive definite matrix ($\succeq \mathbf{0}$). For convenience, we use I and $\mathbf{0}$ to stand respectively for the identity matrix and the zero matrix with proper dimension throughout the context.

2 Preliminaries

In this section, we first use a variational inequality to characterize the solution set of the problem (1). Then, we analyze that GS-ADMM (7) can be treated as a prediction-correction procedure involving a prediction step and a correction step.

2.1 Variational reformulation of (1)

We begin with the following standard lemma whose proof can be found in [16,18].

Lemma 1 *Let $f : \mathcal{R}^m \longrightarrow \mathcal{R}$ and $h : \mathcal{R}^m \longrightarrow \mathcal{R}$ be two convex functions defined on a closed convex set $\Omega \subset \mathcal{R}^m$ and h is differentiable. Suppose that the solution set $\Omega^* = \arg\min_{x\in\Omega}\{f(x)+h(x)\}$ is nonempty. Then, we have*

$$x^* \in \Omega^* \quad \textit{if and only if} \quad x^* \in \Omega,\ f(x) - f(x^*) + \langle x - x^*, \nabla h(x^*)\rangle \geq 0, \forall x \in \Omega.$$

It is well-known in optimization that a triple point $(\mathbf{x}^*, \mathbf{y}^*, \lambda^*) \in \mathcal{M}$ is called the saddle-point of the Lagrangian function (2) if it satisfies

$$L(\mathbf{x}^*, \mathbf{y}^*, \lambda) \leq L(\mathbf{x}^*, \mathbf{y}^*, \lambda^*) \leq L(\mathbf{x}, \mathbf{y}, \lambda^*), \quad \forall (\mathbf{x}, \mathbf{y}, \lambda) \in \mathcal{M}$$

which can be also characterized as

$$\begin{cases} \text{For } i = 1, 2, \cdots, p, \\ \qquad x_i^* = \arg\min\limits_{x_i\in\mathcal{X}_i} \mathcal{L}_\beta\left(x_1^*, \cdots, x_{i-1}^*, x_i, x_{i+1}^*, \cdots, x_p^*, \mathbf{y}^*, \lambda^*\right), \\ \text{For } j = 1, 2, \cdots, q, \\ \qquad y_j^* = \arg\min\limits_{y_j\in\mathcal{Y}_j} \mathcal{L}_\beta\left(\mathbf{x}^*, y_1^*, \cdots, y_{j-1}^*, y_j, y_{j+1}^*, \cdots, y_q^*, \lambda^*\right), \\ \lambda^* = \arg\max\limits_{\lambda\in\mathcal{R}^n} \mathcal{L}_\beta(\mathbf{x}^*, \mathbf{y}^*, \lambda). \end{cases}$$

Then, by Lemma 1, the above saddle-point equations can be equivalently expressed as

$$\begin{cases} \text{For } i = 1, 2, \cdots, p, \\ \qquad x_i^* \in \mathcal{X}_i, \quad f_i(x_i) - f_i(x_i^*) + \langle x_i - x_i^*, -A_i^{\mathsf{T}}\lambda^*\rangle \geq 0, \forall x_i \in \mathcal{X}_i, \\ \text{For } j = 1, 2, \cdots, q, \\ \qquad y_j^* \in \mathcal{Y}_j, \quad g_j(y_j) - g_j(y_j^*) + \langle y_j - y_j^*, -B_j^{\mathsf{T}}\lambda^*\rangle \geq 0, \forall y_j \in \mathcal{Y}_j, \\ \lambda^* \in \mathcal{R}^n, \quad \langle \lambda - \lambda^*, \mathcal{A}\mathbf{x}^* + \mathcal{B}\mathbf{y}^* - c\rangle \geq 0, \forall \lambda \in \mathcal{R}^n. \end{cases} \tag{9}$$

Rewriting (9) in a more compact variational inequality (VI) form, we have

$$h(\mathbf{u}) - h(\mathbf{u}^*) + \langle \mathbf{w} - \mathbf{w}^*, \mathcal{J}(\mathbf{w}^*)\rangle \geq 0, \quad \forall \mathbf{w} \in \mathcal{M}, \tag{10}$$

where

$$h(\mathbf{u}) = \sum_{i=1}^{p} f_i(x_i) + \sum_{j=1}^{p} g_j(y_j)$$

and

$$\mathbf{u} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \\ \lambda \end{pmatrix}, \quad \mathcal{J}(\mathbf{w}) = \begin{pmatrix} -\mathcal{A}^{\mathsf{T}}\lambda \\ -\mathcal{B}^{\mathsf{T}}\lambda \\ \mathcal{A}\mathbf{x} + \mathcal{B}\mathbf{y} - c \end{pmatrix}.$$

Noticing that the affine mapping $\mathcal{J}$ is skew-symmetric, we immediately get

$$\langle \mathbf{w} - \widehat{\mathbf{w}}, \mathcal{J}(\mathbf{w}) - \mathcal{J}(\widehat{\mathbf{w}}) \rangle = 0, \quad \forall\, \mathbf{w}, \widehat{\mathbf{w}} \in \mathcal{M}. \tag{11}$$

Hence, (10) can be also rewritten as

$$\mathrm{VI}(\mathbf{h}, \mathcal{J}, \mathcal{M}): \quad h(\mathbf{u}) - h(\mathbf{u}^*) + \langle \mathbf{w} - \mathbf{w}^*, \mathcal{J}(\mathbf{w}) \rangle \geq 0, \quad \forall \mathbf{w} \in \mathcal{M}. \tag{12}$$

Because of the nonempty assumption on the solution set of (1), the solution set $\mathcal{M}^*$ of the variational inequality VI$(h, \mathcal{J}, \mathcal{M})$ is also nonempty and convex, see e.g. Theorem 2.3.5 [6] for more details. The following theorem given by Theorem 2.1 [11] provides a concrete way to characterize the set $\mathcal{M}^*$.

Theorem 1 *The solution set of the variational inequality VI$(h, \mathcal{J}, \mathcal{M})$ is convex and can be expressed as*

$$\mathcal{M}^* = \bigcap_{\mathbf{w} \in \mathcal{M}} \{ \widehat{\mathbf{w}} \in \mathcal{M} |\ h(\mathbf{u}) - h(\widehat{\mathbf{u}}) + \langle \mathbf{w} - \widehat{\mathbf{w}}, \mathcal{J}(\mathbf{w}) \rangle \geq 0 \}.$$

2.2 A prediction-correction interpretation of GS-ADMM

Following a similar approach in [16], we next interpret GS-ADMM as a prediction-correction procedure. First, let

$$\widetilde{\mathbf{x}}^k = \begin{pmatrix} \widetilde{x}_1^k \\ \widetilde{x}_2^k \\ \vdots \\ \widetilde{x}_p^k \end{pmatrix} = \begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \\ \vdots \\ x_p^{k+1} \end{pmatrix}, \quad \widetilde{\mathbf{y}}^k = \begin{pmatrix} \widetilde{y}_1^k \\ \widetilde{y}_2^k \\ \vdots \\ \widetilde{y}_q^k \end{pmatrix} = \begin{pmatrix} y_1^{k+1} \\ y_2^{k+1} \\ \vdots \\ y_q^{k+1} \end{pmatrix}, \tag{13}$$

$$\widetilde{\lambda}^k = \lambda^k - \beta \left(\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^k - c \right), \tag{14}$$

and

$$\widetilde{\mathbf{u}} = \begin{pmatrix} \widetilde{\mathbf{x}} \\ \widetilde{\mathbf{y}} \end{pmatrix}, \quad \widetilde{\mathbf{w}}^k = \begin{pmatrix} \widetilde{\mathbf{x}}^k \\ \widetilde{\mathbf{y}}^k \\ \widetilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} \mathbf{x}^{k+1} \\ \mathbf{y}^{k+1} \\ \widetilde{\lambda}^k \end{pmatrix}. \tag{15}$$

Then, by using the above notations, we derive the following basic lemma.

Lemma 2 *For the iterates $\widetilde{\mathbf{u}}^k, \widetilde{\mathbf{w}}^k$ defined in (15), we have $\widetilde{\mathbf{w}}^k \in \mathcal{M}$ and*

$$h(\mathbf{u}) - h(\widetilde{\mathbf{u}}^k) + \left\langle \mathbf{w} - \widetilde{\mathbf{w}}^k, \mathcal{J}(\widetilde{\mathbf{w}}^k) + Q(\widetilde{\mathbf{w}}^k - \mathbf{w}^k) \right\rangle \geq 0, \quad \forall \mathbf{w} \in \mathcal{M}, \tag{16}$$

where

$$Q = \begin{bmatrix} H_{\mathbf{x}} & \mathbf{0} \\ \mathbf{0} & \widetilde{Q} \end{bmatrix} \tag{17}$$

with

$$H_{\mathbf{x}} = \beta \begin{bmatrix} \sigma_1 A_1^{\mathsf{T}} A_1 & -A_1^{\mathsf{T}} A_2 & \cdots & -A_1^{\mathsf{T}} A_p \\ -A_2^{\mathsf{T}} A_1 & \sigma_1 A_2^{\mathsf{T}} A_2 & \cdots & -A_2^{\mathsf{T}} A_p \\ \vdots & \vdots & \ddots & \vdots \\ -A_p^{\mathsf{T}} A_1 & -A_p^{\mathsf{T}} A_2 & \cdots & \sigma_1 A_p^{\mathsf{T}} A_p \end{bmatrix}, \tag{18}$$

$$\widetilde{Q} = \left[\begin{array}{cccc|c} (\sigma_2+1)\beta B_1^{\mathsf{T}} B_1 & \mathbf{0} & \cdots & \mathbf{0} & -\tau B_1^{\mathsf{T}} \\ \mathbf{0} & (\sigma_2+1)\beta B_2^{\mathsf{T}} B_2 & \cdots & \mathbf{0} & -\tau B_2^{\mathsf{T}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & (\sigma_2+1)\beta B_q^{\mathsf{T}} B_q & -\tau B_q^{\mathsf{T}} \\ \hline -B_1 & -B_2 & \cdots & -B_q & \frac{1}{\beta} I \end{array}\right]. \tag{19}$$

Proof Omitting some constants, it is easy to verify that the x_i-subproblem $(i = 1, 2, \cdots, p)$ of GS-ADMM can be written as

$$x_i^{k+1} = \arg\min_{x_i \in \mathcal{X}_i} \left\{ f_i(x_i) - \langle \lambda^k, A_i x_i \rangle + \frac{\beta}{2} \|A_i x_i - c_{x,i}\|^2 + \frac{\sigma_1 \beta}{2} \left\| A_i (x_i - x_i^k) \right\|^2 \right\},$$

where $c_{x,i} = c - \sum\limits_{l=1, l\neq i}^{p} A_l x_l^k - \mathcal{B}\mathbf{y}^k$. Hence, by Lemma 1, we have $x_i^{k+1} \in \mathcal{X}_i$ and

$$f_i(x_i) - f_i(x_i^{k+1}) + \left\langle A_i(x_i - x_i^{k+1}), -\lambda^k + \beta(A_i x_i^{k+1} - c_{x,i}) + \sigma_1 \beta A_i (x_i^{k+1} - x_i^k) \right\rangle \geq 0$$

for any $x_i \in \mathcal{X}_i$. So, by the definition of (13) and (14), we get

$$f_i(x_i) - f_i(\widetilde{x}_i^k) + \left\langle A_i(x_i - \widetilde{x}_i^k), -\widetilde{\lambda}^k - \beta \sum_{l=1, l\neq i}^{p} A_l (\widetilde{x}_l^k - x_l^k) + \sigma_1 \beta \sum_{l=1}^{p} A_l (\widetilde{x}_l^k - x_l^k) \right\rangle \geq 0 \tag{20}$$

for any $x_i \in \mathcal{X}_i$. By the way of generating $\lambda^{k+\frac{1}{2}}$ in (7) and the definition of (14), the following relation holds

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \tau(\lambda^k - \widetilde{\lambda}^k). \tag{21}$$

Similarly, the y_j-subproblem $(j = 1, \cdots, q)$ of GS-ADMM can be written as

$$y_j^{k+1} = \arg\min_{y_j \in \mathcal{Y}_j} \left\{ g_j(y_j) - \left\langle \lambda^{k+\frac{1}{2}}, B_j y_j \right\rangle + \frac{\beta}{2} \|B_j y_j - c_{y,j}\|^2 + \frac{\sigma_2 \beta}{2} \left\| B_j (y_j - y_j^k) \right\|^2 \right\},$$

where $c_{y,j} = c - \mathcal{A}\mathbf{x}^{k+1} - \sum\limits_{l=1, l\neq j}^{q} B_l y_l^k$. Hence, by Lemma 1, we have $y_j^{k+1} \in \mathcal{Y}_j$ and

$$g_j(y_j) - g_j(y_j^{k+1}) + \left\langle B_j(y_j - y_j^{k+1}), -\lambda^{k+\frac{1}{2}} + \beta(B_j y_j^{k+1} - c_{y,j}) + \sigma_2 \beta B_j (y_j^{k+1} - y_j^k) \right\rangle \geq 0$$

for any $y_j \in \mathcal{Y}_j$. This inequality, by using (21) and the definition of (13) and (14), can be rewritten as

$$g_j(y_j) - g_j(\widetilde{y}_j^k) + \left\langle B_j(y_j - \widetilde{y}_j^k), -\widetilde{\lambda}^k + (\sigma_2 + 1)\beta B_j(\widetilde{y}_j^k - y_j^k) - \tau(\widetilde{\lambda}^k - \lambda^k) \right\rangle \geq 0 \tag{22}$$

for any $y_j \in \mathcal{Y}_j$. Besides, the equality (14) can be rewritten as

$$\left(\mathcal{A}\widetilde{\mathbf{x}}^k + \mathcal{B}\widetilde{\mathbf{y}}^k - c\right) - \mathcal{B}\left(\widetilde{\mathbf{y}}^k - \mathbf{y}^k\right) + \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k) = 0,$$

which is equivalent to

$$\left\langle \lambda - \widetilde{\lambda}^k, (\mathcal{A}\widetilde{\mathbf{x}}^k + \mathcal{B}\widetilde{\mathbf{y}}^k - c) - \mathcal{B}(\widetilde{\mathbf{y}}^k - \mathbf{y}^k) + \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k) \right\rangle \geq 0, \ \forall \lambda \in \mathcal{R}^n. \tag{23}$$

Then, (16) follows from (20), (22) and (23). $\diamondsuit$

Lemma 3 *For the sequences* $\{\mathbf{w}^k\}$ *and* $\{\widetilde{\mathbf{w}}^k\}$ *generated by GS-ADMM, the following equality holds*

$$\mathbf{w}^{k+1} = \mathbf{w}^k - M(\mathbf{w}^k - \widetilde{\mathbf{w}}^k), \tag{24}$$

where

$$M = \left[\begin{array}{c|cccc} I & & & & \\ \hline & I & & & \\ & & \ddots & & \\ & & & I & \\ & -s\beta B_1 & \cdots & -s\beta B_q & (\tau+s)I \end{array}\right]. \tag{25}$$

Proof It follows from the way of generating λ^{k+1} in the algorithm and (21) that

$$\begin{aligned} \lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - s\beta(\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c) \\ &= \lambda^{k+\frac{1}{2}} - s\beta(\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k} - c) + s\beta\mathcal{B}(\mathbf{y}^k - \mathbf{y}^{k+1}) \\ &= \lambda^k - \tau(\lambda^k - \widetilde{\lambda}^k) - s(\lambda^k - \widetilde{\lambda}^k) + s\beta\mathcal{B}(\mathbf{y}^k - \widetilde{\mathbf{y}}^k) \\ &= \lambda^k - \left[-s\beta\mathcal{B}(\mathbf{y}^k - \widetilde{\mathbf{y}}^k) + (\tau+s)(\lambda^k - \widetilde{\lambda}^k)\right]. \end{aligned}$$

The above equality together with $x_i^{k+1} = \widetilde{x}_l^k$, for $i = 1, \cdots, p$, and $y_j^{k+1} = \widetilde{y}_j^k$, for $j = 1, \cdots, q$, imply

$$\begin{pmatrix} x_1^{k+1} \\ \vdots \\ x_p^{k+1} \\ y_1^{k+1} \\ \vdots \\ y_q^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} x_1^{k} \\ \vdots \\ x_p^{k} \\ y_1^{k} \\ \vdots \\ y_q^{k} \\ \lambda^{k} \end{pmatrix} - \left[\begin{array}{ccc|cccc} I & & & & & & \\ & \ddots & & & & & \\ & & I & & & & \\ \hline & & & I & & & \\ & & & & \ddots & & \\ & & & & & I & \\ & & & -s\beta B_1 & \cdots & -s\beta B_q & (\tau+s)I \end{array}\right] \begin{pmatrix} x_1^k - \widetilde{x}_1^k \\ \vdots \\ x_p^k - \widetilde{x}_p^k \\ y_1^k - \widetilde{y}_1^k \\ \vdots \\ y_q^k - \widetilde{y}_q^k \\ \lambda^k - \widetilde{\lambda}^k \end{pmatrix},$$

which immediately gives (24). $\diamondsuit$

Lemma 2 and Lemma 3 show that our GS-ADMM can be interpreted as a prediction-correction framework, where $\mathbf{w}^{k+1}$ and $\widetilde{\mathbf{w}}^k$ are normally called the predictive variable and the correcting variable, respectively.

3 Convergence analysis of GS-ADMM

Compared with (12) and (16), the key to proving the convergence of GS-ADMM is to verify that the extra term in (16) converges to zero, that is,

$$\lim_{k\to\infty} \left\langle \mathbf{w}-\widetilde{\mathbf{w}}^k, Q(\mathbf{w}^k-\widetilde{\mathbf{w}}^k)\right\rangle = 0, \quad \forall \mathbf{w}\in\mathcal{M}.$$

In this section, we first investigate some properties of the sequence $\{\|\mathbf{w}^k-\mathbf{w}^*\|_H^2\}$. Then, we provide a lower bound of $\|\mathbf{w}^k-\widetilde{\mathbf{w}}^k\|_G^2$. Based on these properties, the global convergence and worst-case $\mathcal{O}(1/t)$ convergence rate of GS-ADMM are established in the end.

3.1 Properties of $\{\left\|\mathbf{w}^k-\mathbf{w}^*\right\|_H^2\}$

It follows from (11) and (16) that $\widetilde{\mathbf{w}}^k\in\mathcal{M}$ and

$$h(\mathbf{u})-h(\widetilde{\mathbf{u}}^k)+\left\langle \mathbf{w}-\widetilde{\mathbf{w}}^k, \mathcal{J}(\mathbf{w})\right\rangle \geq \left\langle \mathbf{w}-\widetilde{\mathbf{w}}^k, Q(\mathbf{w}^k-\widetilde{\mathbf{w}}^k)\right\rangle, \quad \forall \mathbf{w}\in\mathcal{M}. \tag{26}$$

Suppose $\tau+s>0$. Then, the matrix M defined in (25) is nonsingular. So, by (24) and a direct calculation, the right-hand term of (26) is rewritten as

$$\left\langle \mathbf{w}-\widetilde{\mathbf{w}}^k, Q(\mathbf{w}^k-\widetilde{\mathbf{w}}^k)\right\rangle = \left\langle \mathbf{w}-\widetilde{\mathbf{w}}^k, H(\mathbf{w}^k-\mathbf{w}^{k+1})\right\rangle \tag{27}$$

where

$$H = QM^{-1} = \left[\begin{array}{c|c} H_{\mathbf{x}} & \mathbf{0} \\ \hline \mathbf{0} & \widetilde{H}\end{array}\right] \tag{28}$$

with H_x defined in (18) and

$$\widetilde{H} = \left[\begin{array}{ccc|c} (\sigma_2+1-\frac{\tau s}{\tau+s})\beta B_1^\mathsf{T} B_1 & \cdots & -\frac{\tau s}{\tau+s}\beta B_1^\mathsf{T} B_q & -\frac{\tau}{\tau+s}B_1^\mathsf{T} \\ \vdots & \ddots & \vdots & \vdots \\ -\frac{\tau s}{\tau+s}\beta B_q^\mathsf{T} B_1 & \cdots & (\sigma_2+1-\frac{\tau s}{\tau+s})\beta B_q^\mathsf{T} B_q & -\frac{\tau}{\tau+s}B_q^\mathsf{T} \\ \hline -\frac{\tau}{\tau+s}B_1 & \cdots & -\frac{\tau}{\tau+s}B_q & \frac{1}{\beta(\tau+s)}I \end{array}\right]. \tag{29}$$

The following lemma shows that H is a positive definite matrix for proper choice of the parameters $(\sigma_1,\sigma_2,\tau,s)$.

Lemma 4 *The matrix H defined in (28) is symmetric positive definite if*

$$\sigma_1\in(p-1,+\infty), \quad \sigma_2\in(q-1,+\infty), \quad \tau\leq 1 \quad \textit{and} \quad \tau+s>0. \tag{30}$$

Proof By the block structure of H, we only need to show that the blocks $H_{\mathbf{x}}$ and $\widetilde{H}$ in (28) are positive definite if the parameters $(\sigma_1,\sigma_2,\tau,s)$ satisfy (30). Note that

$$H_{\mathbf{x}} = \beta \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_p\end{bmatrix}^\mathsf{T} H_{\mathbf{x},0} \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_p\end{bmatrix}, \tag{31}$$

where

$$H_{\mathbf{x},0} = \begin{bmatrix} \sigma_1 I & -I & \cdots & -I \\ -I & \sigma_1 I & \cdots & -I \\ \vdots & \vdots & \ddots & \vdots \\ -I & -I & \cdots & \sigma_1 I \end{bmatrix}_{p\times p}. \tag{32}$$

If $\sigma_1 \in (p-1, +\infty)$, $H_{\mathbf{x},0}$ is positive definite. Then, it follows from (31) that $H_{\mathbf{x}}$ is positive definite if $\sigma_1 \in (p-1,+\infty)$ and all A_i, $i = 1, \cdots, p$, have full column rank.

Now, note that the matrix $\widetilde{H}$ can be decomposed as

$$\widetilde{H} = \widetilde{D}^{\mathsf{T}} \widetilde{H}_0 \widetilde{D}, \tag{33}$$

where

$$\widetilde{D} = \left[\begin{array}{cccc|c} \beta^{\frac{1}{2}} B_1 & & & & \\ & \beta^{\frac{1}{2}} B_2 & & & \\ & & \ddots & & \\ & & & \beta^{\frac{1}{2}} B_q & \\ \hline & & & & \beta^{-\frac{1}{2}} I \end{array}\right] \tag{34}$$

and

$$\widetilde{H}_0 = \left[\begin{array}{cccc|c} \left(\sigma_2 + 1 - \frac{\tau s}{\tau+s}\right) I & -\frac{\tau s}{\tau+s} I & \cdots & -\frac{\tau s}{\tau+s} I & -\frac{\tau}{\tau+s} I \\ -\frac{\tau s}{\tau+s} I & \left(\sigma_2 + 1 - \frac{\tau s}{\tau+s}\right) I & \cdots & -\frac{\tau s}{\tau+s} I & -\frac{\tau}{\tau+s} I \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{\tau s}{\tau+s} I & -\frac{\tau s}{\tau+s} I & \cdots & \left(\sigma_2 + 1 - \frac{\tau s}{\tau+s}\right) I & -\frac{\tau}{\tau+s} I \\ \hline -\frac{\tau}{\tau+s} I & -\frac{\tau}{\tau+s} I & \cdots & -\frac{\tau}{\tau+s} I & \frac{1}{\tau+s} I \end{array}\right].$$

According to the fact that

$$\begin{aligned} &\begin{bmatrix} I & & \tau I \\ & \ddots & \tau I \\ & & I \end{bmatrix} \widetilde{H}_0 \begin{bmatrix} I & & \tau I \\ & \ddots & \tau I \\ & & I \end{bmatrix}^{\mathsf{T}} \\ &= \left[\begin{array}{cccc|c} (\sigma_2 + 1 - \tau) I & -\tau I & \cdots & -\tau I & \mathbf{0} \\ -\tau I & (\sigma_2 + 1 - \tau) I & \cdots & -\tau I & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\tau I & -\tau I & \cdots & (\sigma_2 + 1 - \tau) I & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \frac{1}{\tau+s} I \end{array}\right] \\ &= \begin{bmatrix} H_{\mathbf{y},0} + (1-\tau) E E^{\mathsf{T}} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\tau+s} I \end{bmatrix}, \end{aligned}$$

where

$$E = \begin{bmatrix} I \\ I \\ \vdots \\ I \end{bmatrix} \quad \text{and} \quad H_{\mathbf{y},0} = \begin{bmatrix} \sigma_2 I & -I & \cdots & -I \\ -I & \sigma_2 I & \cdots & -I \\ \vdots & \vdots & \ddots & \vdots \\ -I & -I & \cdots & \sigma_2 I \end{bmatrix}_{q\times q}, \tag{35}$$

we have $\widetilde{H}_0$ is positive definite if and only if $H_{\mathbf{y},0} + (1-\tau)EE^{\mathsf{T}}$ is positive definite and $\tau + s > 0$. Note that $H_{y,0}$ is positive definite if $\sigma_2 \in (q-1, +\infty)$, and $(1-\tau)EE^{\mathsf{T}}$ is positive semidefinite if $\tau \le 1$. So, $\widetilde{H}_0$ is positive definite if $\sigma_2 \in (q-1, +\infty)$, $\tau \le 1$ and $\tau + s > 0$. Then, it follows from (33) that $\widetilde{H}$ is positive definite, if $\sigma_2 \in (q-1, +\infty)$, $\tau \le 1$, $\tau + s > 0$ and all the matrices B_j, $j = 1, \cdots, q$, have full column rank.

Summarizing the above discussions, the matrix H is positive definite if the parameters $(\sigma_1, \sigma_2, \tau, s)$ satisfy (30). $\diamondsuit$

Theorem 2 *The sequences* $\{\mathbf{w}^k\}$ *and* $\{\widetilde{\mathbf{w}}^k\}$ *generated by GS-ADMM satisfy*

$$\begin{aligned} h(\mathbf{u}) - h(\widetilde{\mathbf{u}}^k) + \left\langle \mathbf{w} - \widetilde{\mathbf{w}}^k, \mathcal{J}(\mathbf{w}) \right\rangle \ge{}& \frac{1}{2}\left\{ \left\|\mathbf{w} - \mathbf{w}^{k+1}\right\|_H^2 - \left\|\mathbf{w} - \mathbf{w}^k\right\|_H^2 \right\} \\ &+ \frac{1}{2}\left\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right\|_G^2, \ \forall \mathbf{w} \in \mathcal{M}, \end{aligned} \tag{36}$$

where

$$G = Q + Q^{\mathsf{T}} - M^{\mathsf{T}} H M. \tag{37}$$

Proof By substituting

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

into the following identity

$$2\langle a - b, H(c-d)\rangle = \|a-d\|_H^2 - \|a-c\|_H^2 + \|c-b\|_H^2 - \|d-b\|_H^2,$$

we have

$$\begin{aligned} & 2\left\langle \mathbf{w} - \widetilde{\mathbf{w}}^k, H(\mathbf{w}^k - \mathbf{w}^{k+1}) \right\rangle \\ ={}& \left\|\mathbf{w} - \mathbf{w}^{k+1}\right\|_H^2 - \left\|\mathbf{w} - \mathbf{w}^k\right\|_H^2 + \left\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right\|_H^2 - \left\|\mathbf{w}^{k+1} - \widetilde{\mathbf{w}}^k\right\|_H^2. \end{aligned} \tag{38}$$

Now, notice that

$$\begin{aligned} & \left\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right\|_H^2 - \left\|\mathbf{w}^{k+1} - \widetilde{\mathbf{w}}^k\right\|_H^2 \\ ={}& \left\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right\|_H^2 - \left\|\widetilde{\mathbf{w}}^k - \mathbf{w}^k + \mathbf{w}^k - \mathbf{w}^{k+1}\right\|_H^2 \\ ={}& \left\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right\|_H^2 - \left\|\widetilde{\mathbf{w}}^k - \mathbf{w}^k + M(\mathbf{w}^k - \widetilde{\mathbf{w}}^k)\right\|_H^2 \\ ={}& \left\langle \mathbf{w}^k - \widetilde{\mathbf{w}}^k, \left(HM + (HM)^{\mathsf{T}} - M^{\mathsf{T}}HM\right)(\mathbf{w}^k - \widetilde{\mathbf{w}}^k) \right\rangle \\ ={}& \left\langle \mathbf{w}^k - \widetilde{\mathbf{w}}^k, \left(Q + Q^{\mathsf{T}} - M^{\mathsf{T}}HM\right)(\mathbf{w}^k - \widetilde{\mathbf{w}}^k) \right\rangle, \end{aligned} \tag{39}$$

where the second equality holds by (24) and the fourth equality follows from (28). Then, (36) follows from (26)-(27), (38)-(39) and the definition of G in (37). $\diamondsuit$

Theorem 3 *The sequences* $\{\mathbf{w}^k\}$ *and* $\{\widetilde{\mathbf{w}}^k\}$ *generated by GS-ADMM satisfy*

$$\left\|\mathbf{w}^{k+1} - \mathbf{w}^*\right\|_H^2 \le \left\|\mathbf{w}^k - \mathbf{w}^*\right\|_H^2 - \left\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right\|_G^2, \ \forall \mathbf{w}^* \in \mathcal{M}^*. \tag{40}$$

Proof Setting $\mathbf{w} = \mathbf{w}^* \in \mathcal{M}^*$ in (36) we have

$$\frac{1}{2}\left\|\mathbf{w}^{k+1} - \mathbf{w}^*\right\|_H^2 \leq \frac{1}{2}\left\|\mathbf{w}^k - \mathbf{w}^*\right\|_H^2 - \frac{1}{2}\left\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right\|_G^2 + h(\mathbf{u}^*) - h(\widetilde{\mathbf{u}}^k) + \left\langle \mathbf{w}^* - \widetilde{\mathbf{w}}^k, \mathcal{J}(\mathbf{w}^*)\right\rangle .$$

The above inequality together with (10) reduces to the inequality (40). $\diamondsuit$

It can be observed that if $\left\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right\|_G^2$ is positive, then the inequality (40) implies the contractiveness of the sequence $\{\mathbf{w}^k - \mathbf{w}^*\}$ under the H-weighted norm. However, the matrix G defined in (37) is not necessarily positive definite when $\sigma_1 \in (p-1, +\infty)$, $\sigma_2 \in (q-1, +\infty)$ and $(\tau, s) \in \mathcal{K}$. Therefore, it is necessary to estimate the lower bound of $\left\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right\|_G^2$ for the sake of the convergence analysis.

3.2 Lower bound of $\left\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right\|_G^2$

This subsection provides a lower bound of $\left\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right\|_G^2$, for $\sigma_1 \in (p-1, +\infty)$, $\sigma_2 \in (q-1, +\infty)$ and $(\tau, s) \in \mathcal{K}$, where $\mathcal{K}$ is defined in (8).

By simple calculations, the G given in (37) can be explicitly written as

$$G = \left[\begin{array}{c|c} H_{\mathbf{x}} & \mathbf{0} \\ \hline \mathbf{0} & \widetilde{G} \end{array}\right],$$

where $H_{\mathbf{x}}$ is defined in (18) and

$$\widetilde{G} = \left[\begin{array}{ccc|c} (\sigma_2 + 1 - s)\beta B_1^{\mathsf{T}} B_1 & \cdots & -s\beta B_1^{\mathsf{T}} B_q & (s-1)B_1^{\mathsf{T}} \\ \vdots & \ddots & \vdots & \vdots \\ -s\beta B_q^{\mathsf{T}} B_1 & \cdots & (\sigma_2 + 1 - s)\beta B_q^{\mathsf{T}} B_q & s-1)B_q^{\mathsf{T}} \\ \hline (s-1)B_1 & \cdots & (s-1)B_q & \frac{2-\tau-s}{\beta} I \end{array}\right]. \tag{41}$$

In addition, we have

$$\widetilde{G} = \widetilde{D}^{\mathsf{T}} \widetilde{G}_0 \widetilde{D},$$

where $\widetilde{D}$ is defined in (34) and

$$\widetilde{G}_0 = \left[\begin{array}{cccc|c} (\sigma_2 + 1 - s)I & -sI & \cdots & -sI & (s-1)I \\ -sI & (\sigma_2 + 1 - s)I & \cdots & -sI & (s-1)I \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -sI & -sI & \cdots & (\sigma_2 + 1 - s)I & (s-1)I \\ \hline (s-1)I & (s-1)I & \cdots & (s-1)I & (2-\tau-s)I \end{array}\right]. \tag{42}$$

Now, we present the following lemma which provides a lower bound of $\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\|_G$.

Lemma 5 *Suppose $\sigma_1 \in (p-1,+\infty)$ and $\sigma_2 \in (q-1,+\infty)$. For the sequences $\{\mathbf{w}^k\}$ and $\{\widetilde{\mathbf{w}}^k\}$ generated by GS-ADMM, there exists $\xi_1 > 0$ such that*

$$\begin{aligned}\left\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right\|_G^2 \geq \xi_1 &\left(\sum_{i=1}^{p}\left\|A_i\left(x_i^k - x_i^{k+1}\right)\right\|^2 + \sum_{j=1}^{q}\left\|B_j\left(y_j^k - y_j^{k+1}\right)\right\|^2\right)\\ &+(1-\tau)\beta\left\|\mathcal{B}\left(\mathbf{y}^k - \mathbf{y}^{k+1}\right)\right\|^2 + (2-\tau-s)\beta\left\|\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c\right\|^2\\ &+2(1-\tau)\beta\left(\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c\right)^{\mathsf{T}}\mathcal{B}\left(\mathbf{y}^k - \mathbf{y}^{k+1}\right). \qquad (43)\end{aligned}$$

Proof First, it is easy to derive that

$$\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\|_G^2 = \left\|\begin{pmatrix} A_1\left(x_1^k - x_1^{k+1}\right)\\ \vdots \\ A_p\left(x_p^k - x_p^{k+1}\right)\\ B_1\left(y_1^k - y_1^{k+1}\right)\\ \vdots \\ B_q\left(y_q^k - y_q^{k+1}\right)\\ \mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c\end{pmatrix}\right\|_{\overline{G}}^2,$$

where

$$\overline{G} = \beta\left[\begin{array}{c|ccc} I & & & \\ \hline & I & & I\\ & & \ddots & \vdots\\ & & & I\end{array}\right]\begin{bmatrix} H_{\mathbf{x},0} & \\ & \widetilde{G}_0\end{bmatrix}\left[\begin{array}{c|ccc} I & & & \\ \hline & I & & I\\ & & \ddots & \vdots\\ & & & I\end{array}\right]^{\mathsf{T}} = \beta\begin{bmatrix} H_{\mathbf{x},0} & \\ & \overline{G}_0\end{bmatrix},$$

with $H_{\mathbf{x},0}$ and $\widetilde{G}_0$ defined in (32) and (42), respectively, and

$$\overline{G}_0 = \left[\begin{array}{cccc|c} (\sigma_2+1-\tau)I & -\tau I & \cdots & -\tau I & (1-\tau)I\\ -\tau I & (\sigma_2+1-\tau)I & \cdots & -\tau I & (1-\tau)I\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ -\tau I & -\tau I & \cdots & (\sigma_2+1-\tau)I & (1-\tau)I\\ \hline (1-\tau)I & (1-\tau)I & \cdots & (1-\tau)I & (2-\tau-s)I\end{array}\right]$$

$$= \left[\begin{array}{ccc|c} & & & (1-\tau)I\\ & H_{\mathbf{y},0} + (1-\tau)EE^{\mathsf{T}} & & \vdots\\ & & & (1-\tau)I\\ \hline (1-\tau)I & \cdots & (1-\tau)I & (2-\tau-s)I\end{array}\right].$$

In the above matrix $\overline{G}_0$, E and $H_{\mathbf{y},0}$ are defined in (35). Hence, we have

$$\frac{1}{\beta}\left\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right\|_G^2 = \left\|\begin{pmatrix} A_1\left(x_1^k - x_1^{k+1}\right)\\ A_2\left(x_2^k - x_2^{k+1}\right)\\ \vdots\\ A_p\left(x_p^k - x_p^{k+1}\right)\end{pmatrix}\right\|_{H_{\mathbf{x},0}}^2 + \left\|\begin{pmatrix} B_1\left(y_1^k - y_1^{k+1}\right)\\ B_2\left(y_2^k - y_2^{k+1}\right)\\ \vdots\\ B_q\left(y_q^k - y_q^{k+1}\right)\end{pmatrix}\right\|_{H_{\mathbf{y},0}}^2$$

$$+\left\|\begin{pmatrix} B_1\left(y_1^k-y_1^{k+1}\right)\\ B_2\left(y_2^k-y_2^{k+1}\right)\\ \vdots\\ B_q\left(y_q^k-y_q^{k+1}\right)\end{pmatrix}\right\|^2_{(1-\tau)EE^\mathsf{T}} + (2-\tau-s)\left\|\mathcal{A}\mathbf{x}^{k+1}+\mathcal{B}\mathbf{y}^{k+1}-c\right\|^2$$

$$+2(1-\tau)\left(\mathcal{A}\mathbf{x}^{k+1}+\mathcal{B}\mathbf{y}^{k+1}-c\right)^\mathsf{T}\mathcal{B}\left(\mathbf{y}^k-\mathbf{y}^{k+1}\right). \tag{44}$$

Since $\sigma_1 \in (p-1,+\infty)$ and $\sigma_2 \in (q-1,+\infty)$, $H_{\mathbf{x},0}$ and $H_{\mathbf{y},0}$ are positive definite. So, there exists a $\xi_1 > 0$ such that

$$\left\|\begin{pmatrix} A_1\left(x_1^k-x_1^{k+1}\right)\\ A_2\left(x_2^k-x_2^{k+1}\right)\\ \vdots\\ A_p\left(x_p^k-x_p^{k+1}\right)\end{pmatrix}\right\|^2_{H_{\mathbf{x},0}} + \left\|\begin{pmatrix} B_1\left(y_1^k-y_1^{k+1}\right)\\ B_2\left(y_2^k-y_2^{k+1}\right)\\ \vdots\\ B_q\left(y_q^k-y_q^{k+1}\right)\end{pmatrix}\right\|^2_{H_{\mathbf{y},0}} \tag{45}$$

$$\geq \xi_1\left(\sum_{i=1}^p\|A_i\left(x_i^k-x_i^{k+1}\right)\|^2+\sum_{j=1}^q\|B_j\left(y_j^k-y_j^{k+1}\right)\|^2\right).$$

In view of the definition of E in (35), we have

$$\left\|\begin{pmatrix} B_1\left(y_1^k-y_1^{k+1}\right)\\ B_2\left(y_2^k-y_2^{k+1}\right)\\ \vdots\\ B_q\left(y_q^k-y_q^{k+1}\right)\end{pmatrix}\right\|^2_{(1-\tau)EE^\mathsf{T}} = (1-\tau)\left\|\mathcal{B}\left(\mathbf{y}^k-\mathbf{y}^{k+1}\right)\right\|^2.$$

Then, the inequality (43) follows from (44) and (45). ◇

Lemma 6 *Suppose $\tau > -1$. Then, the sequence $\{\mathbf{w}^k\}$ generated by GS-ADMM satisfies*

$$\begin{aligned}&\left(\mathcal{A}\mathbf{x}^{k+1}+\mathcal{B}\mathbf{y}^{k+1}-c\right)^\mathsf{T}\mathcal{B}\left(\mathbf{y}^k-\mathbf{y}^{k+1}\right)\\ \geq\ & \frac{1-s}{1+\tau}\left(\mathcal{A}\mathbf{x}^k+\mathcal{B}\mathbf{y}^k-c\right)^\mathsf{T}\mathcal{B}\left(\mathbf{y}^k-\mathbf{y}^{k+1}\right)-\frac{\tau}{1+\tau}\left\|\mathcal{B}\left(\mathbf{y}^k-\mathbf{y}^{k+1}\right)\right\|^2\\ &+\frac{1}{2(1+\tau)\beta}\left(\left\|\mathbf{y}^{k+1}-\mathbf{y}^k\right\|^2_{H_\mathbf{y}}-\left\|\mathbf{y}^k-\mathbf{y}^{k-1}\right\|^2_{H_\mathbf{y}}\right).\end{aligned} \tag{46}$$

Proof It follows from the optimality condition of y_l^{k+1}-subproblem that $y_l^{k+1} \in \mathcal{Y}_l$ and for any $y_l \in \mathcal{Y}_l$, we have

$$g_l(y_l)-g_l(y_l^{k+1})+\left\langle B_l(y_l-y_l^{k+1}), -\lambda^{k+\frac{1}{2}}+\sigma_2\beta B_l\left(y_l^{k+1}-y_l^k\right)+\beta(B_l y_l^{k+1}-c_{y,l})\right\rangle \geq 0$$

with $c_{y,l} = c-\mathcal{A}\mathbf{x}^{k+1}-\sum\limits_{j=1,j\neq l}^q B_j y_j^k$, which implies

$$\begin{aligned}&g_l(y_l)-g_l(y_l^{k+1})\\ &+\left\langle B_l(y_l-y_l^{k+1}), -\lambda^{k+\frac{1}{2}}+\sigma_2\beta B_l\left(y_l^{k+1}-y_l^k\right)+\beta(\mathcal{A}\mathbf{x}^{k+1}+\mathcal{B}\mathbf{y}^{k+1}-c)\right\rangle\\ &-\beta\left\langle B_l(y_l-y_l^{k+1}), \sum_{j=1,j\neq l}^q B_j(y_j^{k+1}-y_j^k)\right\rangle \geq 0.\end{aligned}$$

For $l=1,2,\cdots,q$, letting $y_l=y_l^k$ in the above inequality and summing them together, we can deduce that

$$\begin{aligned}&\sum_{l=1}^{q}\left(g_l(y_l^k)-g_l(y_l^{k+1})\right)+\left\langle \mathcal{B}(\mathbf{y}^k-\mathbf{y}^{k+1}),-\lambda^{k+\frac{1}{2}}+\beta\left(\mathcal{A}\mathbf{x}^{k+1}+\mathcal{B}\mathbf{y}^{k+1}-c\right)\right\rangle \\ &\geq \left\|\mathbf{y}^{k+1}-\mathbf{y}^k\right\|_{H_{\mathbf{y}}}^2,\end{aligned} \tag{47}$$

where

$$\begin{aligned}H_{\mathbf{y}}&=\beta\begin{bmatrix}\sigma_2 B_1^{\mathsf{T}}B_1 & -B_1^{\mathsf{T}}B_2 & \cdots & -B_1^{\mathsf{T}}B_q\\ -B_2^{\mathsf{T}}B_1 & \sigma_2 B_2^{\mathsf{T}}B_2 & \cdots & -B_2^{\mathsf{T}}B_q\\ \vdots & \vdots & \ddots & \vdots\\ -B_q^{\mathsf{T}}B_1 & -B_q^{\mathsf{T}}B_2 & \cdots & \sigma_2 B_q^{\mathsf{T}}B_q\end{bmatrix}\\ &=\beta\begin{bmatrix}B_1 & & \\ & \ddots & \\ & & B_q\end{bmatrix}^{\mathsf{T}} H_{\mathbf{y},0}\begin{bmatrix}B_1 & & \\ & \ddots & \\ & & B_q\end{bmatrix}\end{aligned} \tag{48}$$

and $H_{\mathbf{y},0}$ is defined in (35). Similarly, it follows from the optimality condition of y_l^k-subproblem that

$$\begin{aligned}&g_l(y_l)-g_l(y_l^k)+\left\langle B_l(y_l-y_l^k),-\lambda^{k-\frac{1}{2}}+\sigma_2\beta B_l\left(y_l^k-y_l^{k-1}\right)+\beta(\mathcal{A}\mathbf{x}^k+\mathcal{B}\mathbf{y}^k-c)\right\rangle\\ &-\beta\left\langle B_l(y_l-y_l^k),\sum_{j=1,j\neq l}^{q}B_j(y_j^k-y_j^{k-1})\right\rangle\geq 0.\end{aligned}$$

For $l=1,2,\cdots,q$, letting $y_l=y_l^{k+1}$ in the above inequality and summing them together, we obtain

$$\begin{aligned}&\sum_{l=1}^{q}\left(g_l(y_l^{k+1})-g_p(y_l^k)\right)+\left\langle \mathcal{B}(\mathbf{y}^{k+1}-\mathbf{y}^k),-\lambda^{k-\frac{1}{2}}+\beta\left(\mathcal{A}\mathbf{x}^k+\mathcal{B}\mathbf{y}^k-c\right)\right\rangle\\ &\geq(\mathbf{y}^k-\mathbf{y}^{k+1})^{\mathsf{T}}H_{\mathbf{y}}(\mathbf{y}^k-\mathbf{y}^{k-1}).\end{aligned} \tag{49}$$

Since $\sigma_2\in(q-1,\infty)$ and all B_j, $j=1,\cdots,q$, have full column rank, we have from (48) that $H_{\mathbf{y}}$ is positive definite. Meanwhile, by the Cauchy-Schwartz inequality, we get

$$\left\|\mathbf{y}^{k+1}-\mathbf{y}^k\right\|_{H_{\mathbf{y}}}^2+\left(\mathbf{y}^k-\mathbf{y}^{k+1}\right)^{\mathsf{T}}H_{\mathbf{y}}\left(\mathbf{y}^k-\mathbf{y}^{k-1}\right)\geq\frac{1}{2}\left(\left\|\mathbf{y}^{k+1}-\mathbf{y}^k\right\|_{H_{\mathbf{y}}}^2-\left\|\mathbf{y}^k-\mathbf{y}^{k-1}\right\|_{H_{\mathbf{y}}}^2\right) \tag{50}$$

By adding (47) to (49) and using (50), we achieve

$$\begin{aligned}&\left\langle \mathcal{B}(\mathbf{y}^k-\mathbf{y}^{k+1}),\lambda^{k-\frac{1}{2}}-\lambda^{k+\frac{1}{2}}+\beta(\mathcal{A}\mathbf{x}^{k+1}+\mathcal{B}\mathbf{y}^{k+1}-c)\right\rangle\\ &\geq\langle\mathcal{B}(\mathbf{y}^k-\mathbf{y}^{k+1}),\beta(\mathcal{A}\mathbf{x}^k+\mathcal{B}\mathbf{y}^k-c)\rangle+\frac{1}{2}\left(\left\|\mathbf{y}^{k+1}-\mathbf{y}^k\right\|_{H_{\mathbf{y}}}^2-\left\|\mathbf{y}^k-\mathbf{y}^{k-1}\right\|_{H_{\mathbf{y}}}^2\right).\end{aligned} \tag{51}$$

From the update of $\lambda^{k+\frac{1}{2}}$, i.e., $\lambda^{k+\frac{1}{2}}=\lambda^k-\tau\beta\left(\mathcal{A}\mathbf{x}^{k+1}+\mathcal{B}\mathbf{y}^k-c\right)$ and the update of λ^k, i.e., $\lambda^k=\lambda^{k-\frac{1}{2}}-s\beta\left(\mathcal{A}\mathbf{x}^k+\mathcal{B}\mathbf{y}^k-c\right)$, we have

$$\lambda^{k-\frac{1}{2}}-\lambda^{k+\frac{1}{2}}=\tau\beta(\mathcal{A}\mathbf{x}^{k+1}+\mathcal{B}\mathbf{y}^{k+1}-c)+s\beta(\mathcal{A}\mathbf{x}^k+\mathcal{B}\mathbf{y}^k-c)+\tau\beta\mathcal{B}(\mathbf{y}^k-\mathbf{y}^{k+1}).$$

Substituting the above inequality into the left-term of (51), the proof is completed. $\diamondsuit$

Theorem 4 *Suppose $\sigma_1 \in (p-1,+\infty)$, $\sigma_2 \in (q-1,+\infty)$ and $\tau > -1$. For the sequences $\{\mathbf{w}^k\}$ and $\{\widetilde{\mathbf{w}}^k\}$ generated by GS-ADMM , there exists $\xi_1 > 0$ such that*

$$\begin{aligned}
&\left\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right\|_G^2 \\
&\geq \xi_1 \left(\sum_{i=1}^p \left\|A_i\left(x_i^k - x_i^{k+1}\right)\right\|^2 + \sum_{j=1}^q \left\|B_j\left(y_j^k - y_j^{k+1}\right)\right\|^2 \right) \\
&+ (2-\tau-s)\beta\left\|\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c\right\|^2 + \frac{1-\tau}{1+\tau}\left(\left\|\mathbf{y}^{k+1} - \mathbf{y}^k\right\|_{H_\mathbf{y}}^2 - \left\|\mathbf{y}^k - \mathbf{y}^{k-1}\right\|_{H_\mathbf{y}}^2\right) \\
&+ \frac{(1-\tau)^2}{1+\tau}\beta\left\|\mathcal{B}\left(\mathbf{y}^k - \mathbf{y}^{k+1}\right)\right\|^2 + \frac{2(1-\tau)(1-s)}{1+\tau}\beta\left(\mathcal{A}\mathbf{x}^k + \mathcal{B}\mathbf{y}^k - c\right)^\mathsf{T}\mathcal{B}\left(\mathbf{y}^k - \mathbf{y}^{k+1}\right).
\end{aligned} \tag{52}$$

Proof The inequality (52) is directly obtained from (43) and (46). ◇

The following theorem gives another variant of the lower bound of $\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\|_G^2$, which plays a key role in showing the convergence of GS-ADMM.

Theorem 5 *Let the sequences $\{\mathbf{w}^k\}$ and $\{\widetilde{\mathbf{w}}^k\}$ be generated by GS-ADMM. Then, for any*

$$\sigma_1 \in (p-1,+\infty), \quad \sigma_2 \in (q-1,+\infty) \quad \text{and} \quad (\tau,s) \in \mathcal{K}, \tag{53}$$

where $\mathcal{K}$ is defined in (8), there exist constants $\xi_i (i=1,2) > 0$ and $\xi_j (j=3,4) \geq 0$, such that

$$\begin{aligned}
\left\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right\|_G^2 \geq\ & \xi_1 \left(\sum_{i=1}^p \left\|A_i\left(x_i^k - x_i^{k+1}\right)\right\|^2 + \sum_{j=1}^q \left\|B_j\left(y_j^k - y_j^{k+1}\right)\right\|^2 \right) \\
&+\xi_2\left\|\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c\right\|^2 \\
&+\xi_3\left(\left\|\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c\right\|^2 - \left\|\mathcal{A}\mathbf{x}^k + \mathcal{B}\mathbf{y}^k - c\right\|^2\right) \\
&+\xi_4\left(\left\|\mathbf{y}^{k+1} - \mathbf{y}^k\right\|_{H_\mathbf{y}}^2 - \left\|\mathbf{y}^k - \mathbf{y}^{k-1}\right\|_{H_\mathbf{y}}^2\right).
\end{aligned} \tag{54}$$

Proof By the Cauchy-Schwartz inequality, we have

$$\begin{aligned}
&2(1-\tau)(1-s)\left(\mathcal{A}\mathbf{x}^k + \mathcal{B}\mathbf{y}^k - c\right)^\mathsf{T}\mathcal{B}\left(\mathbf{y}^k - \mathbf{y}^{k+1}\right) \\
&\geq -(1-s)^2\left\|\mathcal{A}\mathbf{x}^k + \mathcal{B}\mathbf{y}^k - c\right\|^2 - (1-\tau)^2\left\|\mathcal{B}\left(\mathbf{y}^k - \mathbf{y}^{k+1}\right)\right\|^2.
\end{aligned} \tag{55}$$

Since

$$-\tau^2 - s^2 - \tau s + \tau + s + 1 = -\tau^2 + (1-s)(\tau+s) + 1,$$

we have $\tau > -1$ when $(\tau,s) \in \mathcal{K}$. Then, combining (55) with Theorem 4, we deduce

$$\begin{aligned}
\left\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right\|_G^2 \geq\ & \xi_1 \left(\sum_{i=1}^p \left\|A_i\left(x_i^k - x_i^{k+1}\right)\right\|^2 + \sum_{j=1}^q \left\|B_j\left(y_j^k - y_j^{k+1}\right)\right\|^2 \right) \\
&+\left(2-\tau-s-\frac{(1-s)^2}{1+\tau}\right)\beta\left\|\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c\right\|^2 \\
&+\frac{(1-s)^2}{1+\tau}\beta\left(\left\|\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c\right\|^2 - \left\|\mathcal{A}\mathbf{x}^k + \mathcal{B}\mathbf{y}^k - c\right\|^2\right) \\
&+\frac{1-\tau}{1+\tau}\left(\left\|\mathbf{y}^{k+1} - \mathbf{y}^k\right\|_{H_\mathbf{y}}^2 - \left\|\mathbf{y}^k - \mathbf{y}^{k-1}\right\|_{H_\mathbf{y}}^2\right),
\end{aligned} \tag{56}$$

where $\xi_1 > 0$ is the constant in Theorem 4. Since $-1 < \tau \leq 1$ and $\beta > 0$, we have

$$\xi_3 := \frac{(1-s)^2}{1+\tau}\beta \geq 0 \quad \text{and} \quad \xi_4 := \frac{1-\tau}{1+\tau} \geq 0. \tag{57}$$

In addition, when $(\tau, s) \in \mathcal{K}$, there is

$$-\tau^2 - s^2 - \tau s + \tau + s + 1 > 0,$$

which, by $\tau > -1$ and $\beta > 0$, implies

$$\xi_2 := \left(2 - \tau - s - \frac{(1-s)^2}{1+\tau}\right)\beta > 0. \tag{58}$$

Hence, the proof is completed. $\diamondsuit$

3.3 Global convergence

In this subsection, we show the global convergence and the worst-case $O(1/t)$ convergence rate of GS-ADMM. The following corollary is obtained directly from Theorems 2-3 and Theorem 5.

Corollary 1 *Let the sequences $\{\mathbf{w}^k\}$ and $\{\widetilde{\mathbf{w}}^k\}$ be generated by GS-ADMM. For any $(\sigma_1, \sigma_2, \tau, s)$ satisfying (53), there exist constants $\xi_i (i = 1, 2) > 0$ and $\xi_j (j = 3, 4) \geq 0$ such that*

$$\begin{aligned} &\left\|\mathbf{w}^{k+1} - \mathbf{w}^*\right\|_H^2 + \xi_3 \left\|\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c\right\|^2 + \xi_4 \left\|\mathbf{y}^{k+1} - \mathbf{y}^k\right\|_{H_\mathbf{y}}^2 \\ \leq &\left\|\mathbf{w}^k - \mathbf{w}^*\right\|_H^2 + \xi_3 \left\|\mathcal{A}\mathbf{x}^k + \mathcal{B}\mathbf{y}^k - c\right\|^2 + \xi_4 \left\|\mathbf{y}^k - \mathbf{y}^{k-1}\right\|_{H_\mathbf{y}}^2 \\ &-\xi_1 \left(\sum_{i=1}^p \|A_i(x_i^k - x_i^{k+1})\|^2 + \sum_{j=1}^q \left\|B_j(y_j^k - y_j^{k+1})\right\|^2\right) \\ &-\xi_2 \left\|\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c\right\|^2, \quad \forall \mathbf{w}^* \in \mathcal{M}^*, \end{aligned} \tag{59}$$

and

$$\begin{aligned} &h(\mathbf{u}) - h(\widetilde{\mathbf{u}}^k) + \langle w - \widetilde{\mathbf{w}}^k, \mathcal{J}(\mathbf{w})\rangle \\ \geq &\frac{1}{2}\left(\|\mathbf{w} - \mathbf{w}^{k+1}\|_H^2 + \xi_3 \left\|\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c\right\|^2 + \xi_4 \left\|\mathbf{y}^{k+1} - \mathbf{y}^k\right\|_{H_\mathbf{y}}^2\right) \\ &-\frac{1}{2}\left(\|\mathbf{w} - \mathbf{w}^k\|_H^2 + \xi_3 \left\|\mathcal{A}\mathbf{x}^k + \mathcal{B}\mathbf{y}^k - c\right\|^2 + \xi_4 \left\|\mathbf{y}^k - \mathbf{y}^{k-1}\right\|_{H_\mathbf{y}}^2\right), \quad \forall \mathbf{w} \in \mathcal{M}. \end{aligned} \tag{60}$$

Theorem 6 *Let the sequences $\{\mathbf{w}^k\}$ and $\{\widetilde{\mathbf{w}}^k\}$ be generated by GS-ADMM. Then, for any $(\sigma_1, \sigma_2, \tau, s)$ satisfying (53), we have*

$$\lim_{k\to\infty} \left(\sum_{i=1}^p \left\|A_i(x_i^k - x_i^{k+1})\right\|^2 + \sum_{j=1}^q \left\|B_j(y_j^k - y_j^{k+1})\right\|^2\right) = 0, \tag{61}$$

$$\lim_{k\to\infty} \left\|\mathcal{A}\mathbf{x}^k + \mathcal{B}\mathbf{y}^k - c\right\| = 0, \tag{62}$$

and there exists a $\mathbf{w}^\infty \in \mathcal{M}^*$ such that

$$\lim_{k\to\infty} \widetilde{\mathbf{w}}^k = \mathbf{w}^\infty. \tag{63}$$

Proof Summing the inequality (59) over $k=1,2,\cdots,\infty$, we have

$$\xi_1\sum_{k=1}^{\infty}\left(\sum_{i=1}^{p}\left\|A_i\left(x_i^k-x_i^{k+1}\right)\right\|^2+\sum_{j=1}^{q}\left\|B_j\left(y_j^k-y_j^{k+1}\right)\right\|^2\right)+\xi_2\sum_{k=1}^{\infty}\left\|\mathcal{A}\mathbf{x}^{k+1}+\mathcal{B}\mathbf{y}^{k+1}-c\right\|^2$$
$$\leq \|\mathbf{w}^1-\mathbf{w}^*\|_H^2+\xi_1\left\|\mathcal{A}\mathbf{x}^1+\mathcal{B}\mathbf{y}^1-c\right\|^2+\xi_2\left\|\mathbf{y}^1-\mathbf{y}^0\right\|_{H_{\mathbf{y}}}^2,$$

which implies that (61) and (62) hold since $\xi_1>0$ and $\xi_2>0$.

Because $(\sigma_1,\sigma_2,\tau,s)$ satisfy (53), we have by Lemma 4 that H is positive definite. So, it follows from (59) that the sequence $\{\mathbf{w}^k\}$ is uniformly bounded. Therefore, there exits a subsequence $\{\mathbf{w}^{k_j}\}$ converging to a point $\mathbf{w}^\infty=(\mathbf{x}^\infty,\mathbf{y}^\infty,\lambda^\infty)\in\mathcal{M}$. In addition, by the definitions of $\widetilde{x}_k$, $\widetilde{y}_k$ and $\widetilde{\lambda}_k$ in (13) and (14), it follows from (61), (62) and the full column rank assumption of all the matrices A_i and B_j that

$$\lim_{k\to\infty}x_i^k-\widetilde{x}_i^k=\mathbf{0},\quad \lim_{k\to\infty}y_j^k-\widetilde{y}_j^k=\mathbf{0}\quad\text{and}\quad\lim_{k\to\infty}\lambda^k-\widetilde{\lambda}^k=\mathbf{0}, \tag{64}$$

for all $i=1,\cdots,p$ and $j=1,\cdots,q$. So, we have $\lim\limits_{k\to\infty}\mathbf{w}^k-\widetilde{\mathbf{w}}^k=\mathbf{0}$. Thus, for any fixed $\mathbf{w}\in\mathcal{M}$, taking $\widetilde{\mathbf{w}}^{k_j}$ in (16) and letting j go to ∞, we obtain

$$h(\mathbf{u})-h(\mathbf{u}^\infty)+\langle w-\mathbf{w}^\infty,\mathcal{J}(\mathbf{w}^\infty)\rangle\geq 0. \tag{65}$$

Hence, $\mathbf{w}^\infty\in\mathcal{M}^*$ is a solution point of $\mathrm{VI}(h,\mathcal{J},\mathcal{M})$ defined in (12).

Since (59) holds for any $\mathbf{w}^*\in\mathcal{M}^*$, by (59) and $\mathbf{w}^\infty\in\mathcal{M}^*$, for all $l\geq k_j$, we have

$$\left\|\mathbf{w}^l-\mathbf{w}^\infty\right\|_H^2+\xi_3\left\|\mathcal{A}\mathbf{x}^l+\mathcal{B}\mathbf{y}^l-c\right\|^2+\xi_4\left\|\mathbf{y}^l-\mathbf{y}^{l-1}\right\|_{H_{\mathbf{y}}}^2$$
$$\leq\left\|\mathbf{w}^{k_j}-\mathbf{w}^\infty\right\|_H^2+\xi_3\left\|\mathcal{A}\mathbf{x}^{k_j}+\mathcal{B}\mathbf{y}^{k_j}-c\right\|^2+\xi_4\left\|\mathbf{y}^{k_j}-\mathbf{y}^{k_j-1}\right\|_{H_{\mathbf{y}}}^2.$$

This together with (62), (64), $\lim\limits_{j\to\infty}\mathbf{w}^{k_j}=\mathbf{w}^\infty$ and the positive definiteness of H illustrate $\lim\limits_{l\to\infty}\mathbf{w}^l=\mathbf{w}^\infty$. Therefore, the whole sequence $\{\mathbf{w}^k\}$ converges to the solution $\mathbf{w}^\infty\in\mathcal{M}^*$. This completes the whole proof. $\diamondsuit$

The above Theorem 6 shows the global convergence of our GS-ADMM. Next, we show the $\mathcal{O}(1/t)$ convergence rate for the ergodic iterates

$$\mathbf{w}_t:=\frac{1}{t}\sum_{k=1}^{t}\widetilde{\mathbf{w}}^k\quad\text{and}\quad\mathbf{u}_t:=\frac{1}{t}\sum_{k=1}^{t}\widetilde{\mathbf{u}}^k. \tag{66}$$

Theorem 7 *Let the sequences $\{\mathbf{w}^k\}$ and $\{\widetilde{\mathbf{w}}^k\}$ be generated by GS-ADMM. Then, for any $(\sigma_1,\sigma_2,\tau,s)$ satisfying (53), there exist $\xi_j(j=3,4)\geq 0$ such that*

$$h(\mathbf{u}_t)-h(\mathbf{u})+\langle\mathbf{w}_t-\mathbf{w},\mathcal{J}(\mathbf{w})\rangle \tag{67}$$
$$\leq\frac{1}{2t}\left(\left\|\mathbf{w}-\mathbf{w}^1\right\|_H^2+\xi_3\left\|\mathcal{A}\mathbf{x}^1+\mathcal{B}\mathbf{y}^1-c\right\|^2+\xi_4\left\|\mathbf{y}^1-\mathbf{y}^0\right\|_{H_{\mathbf{y}}}^2\right),\quad\forall\mathbf{w}\in\mathcal{M}.$$

Proof For $k = 1, \cdots, t$, summing the inequality (60), we have

$$
\begin{aligned}
& th(\mathbf{u}) - \sum_{k=1}^{t} h(\widetilde{\mathbf{u}}^k) + \left\langle t\mathbf{w} - \sum_{k=1}^{t} \widetilde{\mathbf{w}}^k, \mathcal{J}(\mathbf{w}) \right\rangle \qquad (68)\\
\geq & \frac{1}{2}\left(\left\|\mathbf{w} - \mathbf{w}^{t+1}\right\|_H^2 + \xi_3 \left\|\mathcal{A}\mathbf{x}^{t+1} + \mathcal{B}\mathbf{y}^{t+1} - c\right\|^2 + \xi_4 \left\|\mathbf{y}^{t+1} - \mathbf{y}^t\right\|_{H_\mathbf{y}}^2 \right)\\
& -\frac{1}{2}\left(\|\mathbf{w} - \mathbf{w}^1\|_H^2 + \xi_3 \left\|\mathcal{A}\mathbf{x}^1 + \mathcal{B}\mathbf{y}^1 - c\right\|^2 + \xi_4 \left\|\mathbf{y}^1 - \mathbf{y}^0\right\|_{H_\mathbf{y}}^2 \right), \quad \forall \mathbf{w} \in \mathcal{M}.
\end{aligned}
$$

Since $(\sigma_1, \sigma_2, \tau, s)$ satisfy (53), $H_\mathbf{y}$ is positive definite. And by Lemma 4, H is also positive definite. So, it follows from (68) that

$$
\begin{aligned}
& \frac{1}{t}\sum_{k=1}^{t} h(\widetilde{\mathbf{u}}^k) - h(\mathbf{u}) + \left\langle \frac{1}{t}\sum_{k=1}^{t} \widetilde{\mathbf{w}}^k - \mathbf{w}, \mathcal{J}(\mathbf{w}) \right\rangle \qquad (69)\\
\leq & \frac{1}{2t}(\left\|\mathbf{w} - \mathbf{w}^1\right\|_H^2 + \xi_3 \left\|\mathcal{A}\mathbf{x}^1 + \mathcal{B}\mathbf{y}^1 - c\right\|^2 + \xi_4 \left\|\mathbf{y}^1 - \mathbf{y}^0\right\|_{H_\mathbf{y}}^2), \quad \forall \mathbf{w} \in \mathcal{M}.
\end{aligned}
$$

By the convexity of h and (66), we have

$$
h(\mathbf{u}_t) \leq \frac{1}{t}\sum_{k=1}^{t} h(\widetilde{\mathbf{u}}^k).
$$

Then, (67) follows from (69). $\diamondsuit$

Remark 1 In the above Theorem 6 and Theorem 7, we assume the parameters $(\sigma_1, \sigma_2, \tau, s)$ satisfy (53). However, because of the symmetric role played by the x and y iterates in the GS-ADMM, substituting the index $k+1$ by k for the x and λ iterates, the GS-ADMM algorithm (7) can be clearly written as

$$
\left\{
\begin{array}{l}
\text{For } j = 1, 2, \cdots, q,\\
\quad y_j^{k+1} = \arg\min_{y_j \in \mathcal{Y}_j} \mathcal{L}_\beta(\mathbf{x}^k, y_1^k, \cdots, y_j, \cdots, y_q^k, \lambda^{k-\frac{1}{2}}) + Q_j^k(y_j),\\
\quad \text{where } Q_j^k(y_j) = \frac{\sigma_2\beta}{2}\left\|B_j(y_j - y_j^k)\right\|^2,\\
\lambda^k = \lambda^{k-\frac{1}{2}} - s\beta(\mathcal{A}\mathbf{x}^k + \mathcal{B}\mathbf{y}^{k+1} - c)\\
\\
\text{For } i = 1, 2, \cdots, p,\\
\quad x_i^{k+1} = \arg\min_{x_i \in \mathcal{X}_i} \mathcal{L}_\beta(x_1^k, \cdots, x_i, \cdots, x_p^k, \mathbf{y}^{k+1}, \lambda^k) + P_i^k(x_i),\\
\quad \text{where } P_i^k(x_i) = \frac{\sigma_1\beta}{2}\left\|A_i(x_i - x_i^k)\right\|^2,\\
\lambda^{k+\frac{1}{2}} = \lambda^k - \tau\beta(\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c).
\end{array}
\right. \qquad (70)
$$

So, by applying Theorem 6 and Theorem 7 on the algorithm (70), it also converges and has the $\mathcal{O}(1/t)$ convergence rate when $(\sigma_1, \sigma_2, \tau, s)$ satisfy

$$
\sigma_1 \in (p-1, +\infty), \quad \sigma_2 \in (q-1, +\infty) \quad \text{and} \quad (\tau, s) \in \overline{\mathcal{K}}, \qquad (71)
$$

where

$$
\overline{\mathcal{K}} = \left\{(\tau, s) \mid \tau + s > 0,\ s \leq 1,\ -\tau^2 - s^2 - \tau s + \tau + s + 1 > 0\right\}. \qquad (72)
$$

Hence, the convergence domain $\mathcal{K}$ in Theorem 6 and Theorem 7 can be easily enlarged to the symmetric domain, shown in Fig. 2,

$$
\mathcal{G} = \mathcal{K} \cup \overline{\mathcal{K}} = \left\{(\tau, s) \mid \tau + s > 0,\ -\tau^2 - s^2 - \tau s + \tau + s + 1 > 0\right\}. \qquad (73)
$$

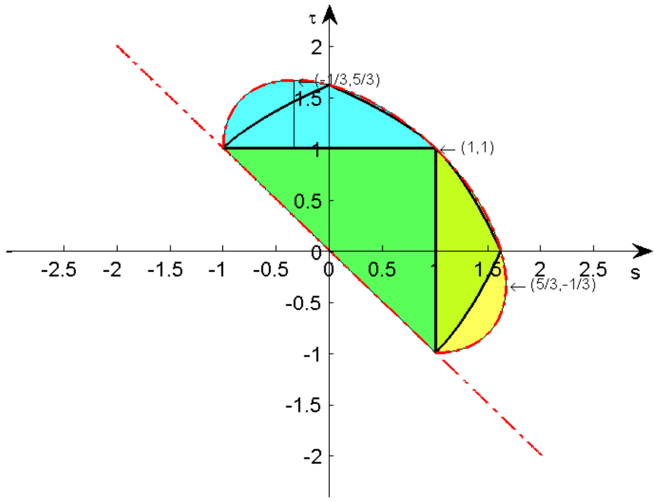

Fig 2. Stepsize region $\mathcal{G}$ of GS-ADMM

Remark 2 Theorem 6 implies that we can apply the following easily usable stopping criterion for GS-ADMM:

$$\max\left\{\left\|x_i^k - x_i^{k+1}\right\|_\infty, \left\|y_j^k - y_j^{k+1}\right\|_\infty, \left\|\mathcal{A}\mathbf{x}^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c\right\|_\infty\right\} \leq tol,$$

where $tol > 0$ is a given tolerance error. One the other hand, Theorem 7 tells us that for any given compact set $\mathcal{K} \subset \mathcal{M}$, if

$$\eta = \sup_{\mathbf{w}\in\mathcal{K}}\left\{\|\mathbf{w} - \mathbf{w}^1\|_H^2 + \xi_3\left\|\mathcal{A}\mathbf{x}^1 + \mathcal{B}\mathbf{y}^1 - c\right\|^2 + \xi_4\left\|\mathbf{y}^1 - \mathbf{y}^0\right\|_{H_\mathbf{y}}^2\right\} < \infty,$$

we have

$$h(\mathbf{u}_t) - h(\mathbf{u}) + \langle \mathbf{w}_t - \mathbf{w}, \mathcal{J}(\mathbf{w})\rangle \leq \frac{\eta}{2t},$$

which shows that GS-ADMM has a worst-case $\mathcal{O}(1/t)$ convergence rate in an ergodic sense.

4 Two special cases of GS-ADMM

Note that in GS-ADMM (7), the two proximal parameters σ_1 and σ_2 are required to be strictly positive for the generalized $p + q$ block separable convex programming. However, when taking $\sigma_1 = \sigma_2 = 0$ for the two-block problem, i.e., $p = q = 1$, GS-ADMM would reduce to the scheme (5), which is globally convergent [16]. Such observations motivate us to further investigate the following two special cases:

(a) GS-ADMM (7) with $p \geq 1, q = 1, \sigma_1 \in (p - 1, \infty)$ and $\sigma_2 = 0$;
(b) GS-ADMM (7) with $p = 1, q \geq 1, \sigma_1 = 0$ and $\sigma_2 \in (q - 1, \infty)$.

4.1 Convergence for the case (a)

The corresponding GS-ADMM for the first case (a) can be simplified as follows:

$$\begin{cases} \text{For } i = 1, 2, \cdots, p, \\ \quad x_i^{k+1} = \arg\min_{x_i \in \mathcal{X}_i} \mathcal{L}_\beta\left(x_1^k, \cdots, x_i, \cdots, x_p^k, y^k, \lambda^k\right) + P_i^k(x_i), \\ \quad \text{where } P_i^k(x_i) = \frac{\sigma_1 \beta}{2}\left\|A_i\left(x_i - x_i^k\right)\right\|^2, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \tau\beta\left(\mathcal{A}\mathbf{x}^{k+1} + By^k - c\right), \\ y^{k+1} = \arg\min_{y \in \mathcal{Y}} \mathcal{L}_\beta\left(\mathbf{x}^{k+1}, y, \lambda^{k+\frac{1}{2}}\right), \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta\left(\mathcal{A}\mathbf{x}^{k+1} + By^{k+1} - c\right). \end{cases} \tag{74}$$

And, the corresponding matrices Q, M, H and G in this case are the following:

$$Q = \left[\begin{array}{c|c} H_{\mathbf{x}} & \mathbf{0} \\ \hline \mathbf{0} & \widetilde{Q} \end{array}\right], \tag{75}$$

where $H_{\mathbf{x}}$ is defined in (18) and

$$\widetilde{Q} = \begin{bmatrix} \beta B^\mathsf{T} B & -\tau B^\mathsf{T} \\ -B & \frac{1}{\beta} I \end{bmatrix}, \tag{76}$$

$$M = \left[\begin{array}{c|cc} I & & \\ \hline & I & \\ & -s\beta B & (\tau+s)I \end{array}\right], \tag{77}$$

$$H = QM^{-1} = \left[\begin{array}{c|cc} H_{\mathbf{x}} & & \\ \hline & \left(1 - \frac{\tau s}{\tau+s}\right)\beta B^\mathsf{T} B & -\frac{\tau}{\tau+s} B^\mathsf{T} \\ & -\frac{\tau}{\tau+s} B & \frac{1}{(\tau+s)\beta} I \end{array}\right], \tag{78}$$

$$G = Q + Q^\mathsf{T} - M^\mathsf{T} H M = \left[\begin{array}{c|cc} H_{\mathbf{x}} & & \\ \hline & (1-s)\beta B^\mathsf{T} B & (s-1)B^\mathsf{T} \\ & (s-1)B & \frac{2-\tau-s}{\beta} I \end{array}\right]. \tag{79}$$

It can be verified that H in (78) is positive definite as long as

$$\sigma_1 \in (p-1, +\infty), \qquad (\tau, s) \in \{(\tau, s) \mid \tau + s > 0, \tau < 1\}.$$

Analogous to (44), we have

$$\begin{aligned} & \frac{1}{\beta}\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\|_G^2 \qquad (80) \\ = & \left\| \begin{pmatrix} A_1\left(x_1^k - x_1^{k+1}\right) \\ A_2\left(x_2^k - x_2^{k+1}\right) \\ \vdots \\ A_p\left(x_p^k - x_p^{k+1}\right) \end{pmatrix} \right\|_{H_{\mathbf{x},0}}^2 + (2-\tau-s)\left\|\mathcal{A}\mathbf{x}^{k+1} + By^{k+1} - c\right\|^2 \\ & + (1-\tau)\left\|B\left(y^k - y^{k+1}\right)\right\|^2 + 2(1-\tau)\left(\mathcal{A}\mathbf{x}^{k+1} + By^{k+1} - c\right)^\mathsf{T} B\left(y^k - y^{k+1}\right) \\ \geq & \ \xi_1 \sum_{i=1}^p \left\|A_i\left(x_i^k - x_i^{k+1}\right)\right\|^2 + (2-\tau-s)\left\|\mathcal{A}\mathbf{x}^{k+1} + By^{k+1} - c\right\|^2 \\ & + (1-\tau)\left\|B\left(y^k - y^{k+1}\right)\right\|^2 + 2(1-\tau)\left(\mathcal{A}\mathbf{x}^{k+1} + By^{k+1} - c\right)^\mathsf{T} B\left(y^k - y^{k+1}\right), \end{aligned}$$

for some $\xi_1 > 0$, since $H_{\mathbf{x},0}$ defined in (32) is positive definite. When $\sigma_2 = 0$, by a slight modification for the proof of Lemma 6, we have the following lemma.

Lemma 7 *Suppose $\tau > -1$. The sequence $\{\mathbf{w}^k\}$ generated by the algorithm (74) satisfies*

$$\begin{aligned}&\left(\mathcal{A}\mathbf{x}^{k+1} + By^{k+1} - c\right)^\top B\left(y^k - y^{k+1}\right)\\ \geq\ & \frac{1-s}{1+\tau}\left(\mathcal{A}\mathbf{x}^k + By^k - c\right)^\top B\left(y^k - y^{k+1}\right) - \frac{\tau}{1+\tau}\left\|B\left(y^k - y^{k+1}\right)\right\|^2.\end{aligned}$$

Then, the following two theorems are similar to Theorem 5 and Theorem 6.

Theorem 8 *Let the sequences $\{\mathbf{w}^k\}$ and $\{\widetilde{\mathbf{w}}^k\}$ be generated by the algorithm (74). For any*

$$\sigma_1 \in (p-1, +\infty) \quad \textit{and} \quad (\tau, s) \in \mathcal{K},$$

where $\mathcal{K}$ is given in (8), there exist constants $\xi_i (i = 1, 2) > 0$ and $\xi_3 \geq 0$, such that

$$\begin{aligned}\left\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right\|_G^2 \geq\ & \xi_1 \sum_{i=1}^p \|A_i\left(x_i^k - x_i^{k+1}\right)\|^2 + \xi_2\left\|\mathcal{A}\mathbf{x}^{k+1} + By^{k+1} - c\right\|^2 \\ & + \xi_3\left(\left\|\mathcal{A}\mathbf{x}^{k+1} + By^{k+1} - c\right\|^2 - \left\|\mathcal{A}\mathbf{x}^k + By^k - c\right\|^2\right).\end{aligned} \tag{81}$$

Proof For any $(\tau, s) \in \mathcal{K}$, we have $\tau > -1$. Then, by Lemma 7, the inequality (80) and the Cauchy-Schwartz inequality (55), we can deduce that (81) holds with ξ_1 given in (80), ξ_2 and ξ_3 given in (58) and (57), respectively. ♢

Theorem 9 *Let the sequences $\{\mathbf{w}^k\}$ and $\{\widetilde{\mathbf{w}}^k\}$ be generated by the algorithm (74). For any*

$$\sigma_1 \in (p-1, +\infty) \quad \textit{and} \quad (\tau, s) \in \mathcal{K}_1, \tag{82}$$

where

$$\mathcal{K}_1 = \left\{(\tau, s) \mid \tau + s > 0,\ \tau < 1,\ -\tau^2 - s^2 - \tau s + \tau + s + 1 > 0\right\},$$

we have

$$\lim_{k\to\infty} \sum_{i=1}^p \left\|A_i\left(x_i^k - x_i^{k+1}\right)\right\|^2 = 0 \quad \textit{and} \quad \lim_{k\to\infty}\left\|\mathcal{A}\mathbf{x}^k + By^k - c\right\| = 0, \tag{83}$$

and there exists a $\mathbf{w}^\infty \in \mathcal{M}^*$ such that $\lim_{k\to\infty} \widetilde{\mathbf{w}}^k = \mathbf{w}^\infty$.

Proof First, it is clear that Theorem 3 still holds, which, combining with Theorem 8, gives

$$\begin{aligned}&\|\mathbf{w}^{k+1} - \mathbf{w}^*\|_H^2 + \xi_3\left\|\mathcal{A}\mathbf{x}^{k+1} + By^{k+1} - c\right\|^2 \\ \leq\ & \|\mathbf{w}^k - \mathbf{w}^*\|_H^2 + \xi_3\left\|\mathcal{A}\mathbf{x}^k + By^k - c\right\|^2 \\ & -\xi_1 \sum_{i=1}^p \|A_i\left(x_i^k - x_i^{k+1}\right)\|^2 - \xi_2\left\|\mathcal{A}\mathbf{x}^{k+1} + By^{k+1} - c\right\|^2, \quad \forall \mathbf{w}^* \in \mathcal{M}^*.\end{aligned} \tag{84}$$

Hence, (83) holds. For (σ_1, τ, s) satisfying (82), H in (78) is positive definite. So, by (84), $\{\mathbf{w}^k\}$ is uniformly bounded and therefore, there exits a subsequence $\{\mathbf{w}^{k_j}\}$ converging to a point $\mathbf{w}^\infty = (\mathbf{x}^\infty, y^\infty, \lambda^\infty) \in \mathcal{M}$. So, it follows from (83) and the full column rank assumption of all the matrices A_i that

$$\lim_{k\to\infty} x_i^k - \widetilde{x}_i^k = \lim_{k\to\infty} x_i^k - x_i^{k+1} = \mathbf{0} \quad \text{and} \quad \lim_{k\to\infty} \lambda^k - \widetilde{\lambda}^k = \mathbf{0}, \tag{85}$$

for all $i = 1, \cdots, p$. Hence, by $\lim_{j\to\infty} \mathbf{w}^{k_j} = \mathbf{w}^\infty$ and (83), we have

$$\lim_{j\to\infty} \mathbf{x}^{k_j+1} = \mathbf{x}^\infty \quad \text{and} \quad \mathcal{A}\mathbf{x}^\infty + By^\infty - c = \mathbf{0},$$

and therefore, by the full column rank assumption of B and (83),

$$\lim_{j\to\infty} y^{k_j+1} = \lim_{j\to\infty} \widetilde{y}^{k_j} = y^\infty.$$

Hence, by (85), we have $\lim_{j\to\infty} \mathbf{w}^{k_j} - \widetilde{\mathbf{w}}^{k_j} = \mathbf{0}$. Thus, by taking $\widetilde{\mathbf{w}}^{k_j}$ in (16) and letting j go to ∞, the inequality (65) still holds. Then, the rest proof of this theorem follows from the same proof of Theorem 6. $\diamondsuit$

Based on Theorem 8 and by a similar proof to those of Theorem 7, we can also show that the algorithm (74) has the worst-case $\mathcal{O}(1/t)$ convergence rate, which is omitted here for conciseness.

4.2 Convergence for the case (b)

The corresponding GS-ADMM for the case (b) can be simplified as follows

$$\begin{cases} x^{k+1} = \arg\min_{x\in\mathcal{X}} \mathcal{L}_\beta(x, \mathbf{y}^k, \lambda^k) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \tau\beta(Ax^{k+1} + \mathcal{B}\mathbf{y}^k - c), \\ \text{For } j = 1, 2, \cdots, q, \\ \quad y_j^{k+1} = \arg\min_{y_j\in\mathcal{Y}_j} \mathcal{L}_\beta(x^{k+1}, y_1^k, \cdots, y_j, \cdots, y_q^k, \lambda^{k+\frac{1}{2}}) + Q_j^k(y_j), \\ \quad \text{where } Q_j^k(y_j) = \frac{\sigma_2\beta}{2}\left\|B_j(y_j - y_j^k)\right\|^2, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(Ax^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c). \end{cases} \tag{86}$$

In this case, the corresponding matrices Q, M, H and G become $\widetilde{Q}, \widetilde{M}, \widetilde{H}$ and $\widetilde{G}$, which are defined in (19), the lower-right block of M in (25), (29) and (41), respectively.

In what follows, let us define

$$\mathbf{v}^k = \begin{pmatrix} \mathbf{y}^k \\ \lambda^k \end{pmatrix} \quad \text{and} \quad \widetilde{\mathbf{v}}^k = \begin{pmatrix} \widetilde{\mathbf{y}}^k \\ \widetilde{\lambda}^k \end{pmatrix}.$$

Then, by the proof of Theorem 5, we can deduce the following theorem.

Theorem 10 *Let the sequences $\{\mathbf{v}^k\}$ and $\{\widetilde{\mathbf{v}}^k\}$ be generated by the algorithm (86). For any*

$$\sigma_2 \in (q-1, +\infty) \quad \textit{and} \quad (\tau, s) \in \mathcal{K},$$

where $\mathcal{K}$ is defined in (8), there exist constants $\xi_i(i = 1, 2) > 0$ and $\xi_j(j = 3, 4) \geq 0$ such that

$$\begin{aligned} \left\|\mathbf{v}^k - \widetilde{\mathbf{v}}^k\right\|_{\widetilde{G}}^2 \geq \xi_1 \sum_{j=1}^{q} \left\|B_j\left(y_j^k - y_j^{k+1}\right)\right\|^2 + \xi_2\left\|Ax^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c\right\|^2 \\ + \xi_3\left(\left\|Ax^{k+1} + \mathcal{B}\mathbf{y}^{k+1} - c\right\|^2 - \left\|Ax^k + \mathcal{B}\mathbf{y}^k - c\right\|^2\right) \\ + \xi_4\left(\left\|\mathbf{y}^{k+1} - \mathbf{y}^k\right\|_{H_\mathbf{y}}^2 - \left\|\mathbf{y}^k - \mathbf{y}^{k-1}\right\|_{H_\mathbf{y}}^2\right). \end{aligned}$$

By slight modifications of the proof of Theorem 6 and Theorem 9, we have the following global convergence theorem.

Theorem 11 *Let the sequences* $\{\mathbf{w}^k\}$ *and* $\{\widetilde{\mathbf{w}}^k\}$ *be generated by the algorithm (74). Then, for any*

$$\sigma_2 \in (q-1, +\infty) \quad \text{and} \quad (\tau, s) \in \mathcal{K},$$

where $\mathcal{K}$ *is defined in (8), we have*

$$\lim_{k\to\infty} \sum_{j=1}^{q} \left\| B_j \left(y_j^k - y_j^{k+1} \right) \right\|^2 = 0 \quad \text{and} \quad \lim_{k\to\infty} \left\| Ax^k + \mathcal{B}\mathbf{y}^k - c \right\| = 0,$$

and there exists a $\mathbf{w}^\infty \in \mathcal{M}^*$ *such that* $\lim_{k\to\infty} \widetilde{\mathbf{w}}^k = \mathbf{w}^\infty$.

By a similar proof to that of Theorem 7, the algorithm (86) also has the worst-case $\mathcal{O}(1/t)$ convergence rate.

Remark 3 Again, substituting the index $k+1$ by k for the x and λ iterates, the algorithm (74) can be also written as

$$\begin{cases} y^{k+1} = \arg\min\limits_{y\in\mathcal{Y}} \mathcal{L}_\beta(\mathbf{x}^k, y, \lambda^{k-\frac{1}{2}}), \\ \lambda^k = \lambda^{k-\frac{1}{2}} - s\beta(\mathcal{A}\mathbf{x}^k + By^{k+1} - c) \\ \text{For } i = 1, 2, \cdots, p, \\ \quad x_i^{k+1} = \arg\min\limits_{x_i\in\mathcal{X}_i} \mathcal{L}_\beta(x_1^k, \cdots, , x_i, \cdots, x_p^k, y^{k+1}, \lambda^k) + P_i^k(x_i), \\ \quad \text{where } P_i^k(x_i) = \frac{\sigma_1\beta}{2} \left\| A_i(x_i - x_i^k) \right\|^2, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \tau\beta(\mathcal{A}\mathbf{x}^{k+1} + By^{k+1} - c). \end{cases}$$

So, by applying Theorem 11 on the above algorithm, we know the algorithm (74) also converges globally when (σ_1, τ, s) satisfy

$$\sigma_1 \in (p-1, +\infty), \quad \text{and} \quad (\tau, s) \in \overline{\mathcal{K}},$$

where $\overline{\mathcal{K}}$ is given in (72). Hence, the convergence domain $\mathcal{K}_1$ in Theorem 9 can be enlarged to the symmetric domain $\mathcal{G} = \mathcal{K}_1 \cup \overline{\mathcal{K}}$ given in (73). By a similar reason, the convergence domain $\mathcal{K}$ in Theorem 11 can be enlarged to $\mathcal{G}$ as well.

5 Numerical experiments

In this section, we investigate the performance of the proposed GS-ADMM for solving a class of sparse matrix minimization problems. All the algorithms are coded and simulated in MATLAB 7.10(R2010a) on a PC with Intel Core i5 processor(3.3GHz) with 4 GB memory.

5.1 Test problem

Consider the following Latent Variable Gaussian Graphical Model Selection (LVGGMS) problem arising in the statistical learning [2,20]:

$$\begin{array}{ll} \min\limits_{X,S,L\in\mathcal{R}^{n\times n}} & F(X,S,L) := \langle X, C\rangle - \log\det(X) + \nu\|S\|_1 + \mu tr(L) \\ \text{s.t.} & X - S + L = \mathbf{0}, \ L \succeq \mathbf{0}, \end{array} \tag{87}$$

where $C \in \mathcal{R}^{n\times n}$ is the covariance matrix obtained from observation, ν and μ are two given positive weight parameters, $tr(L)$ stands for the trace of the matrix L and $\|S\|_1 = \sum_{ij}|S_{ij}|$. Clearly, by two different ways of partitioning the variables of (87), the GS-ADMM (7) can lead to the following two algorithms:

$$\left\{\begin{array}{l} X^{k+1} = \arg\min\limits_{X}\left\{\langle X, C\rangle - \log\det(X) + \frac{\beta}{2}\left\|X - S^k + L^k - \frac{\Lambda^k}{\beta}\right\|_F^2 + \frac{\sigma_1\beta}{2}\left\|X - X^k\right\|_F^2\right\},\\ S^{k+1} = \arg\min\limits_{S}\left\{\nu\|S\|_1 + \frac{\beta}{2}\left\|X^k - S + L^k - \frac{\Lambda^k}{\beta}\right\|_F^2 + \frac{\sigma_1\beta}{2}\left\|S - S^k\right\|_F^2\right\},\\ \Lambda^{k+\frac{1}{2}} = \Lambda^k - \tau\beta(X^{k+1} - S^{k+1} + L^k),\\ L^{k+1} = \arg\min\limits_{L\succeq\mathbf{0}}\left\{\mu tr(L) + \frac{\beta}{2}\left\|X^{k+1} - S^{k+1} + L - \frac{\Lambda^{k+\frac{1}{2}}}{\beta}\right\|_F^2 + \frac{\sigma_2\beta}{2}\left\|L - L^k\right\|_F^2\right\},\\ \Lambda^{k+1} = \Lambda^{k+\frac{1}{2}} - s\beta(X^{k+1} - S^{k+1} + L^{k+1}); \end{array}\right. \tag{88}$$

$$\left\{\begin{array}{l} X^{k+1} = \arg\min\limits_{X}\left\{\langle X, C\rangle - \log\det(X) + \frac{\beta}{2}\left\|X - S^k + L^k - \frac{\Lambda^k}{\beta}\right\|_F^2 + \frac{\sigma_1\beta}{2}\left\|X - X^k\right\|_F^2\right\},\\ \Lambda^{k+\frac{1}{2}} = \Lambda^k - \tau\beta(X^{k+1} - S^k + L^k),\\ S^{k+1} = \arg\min\limits_{S}\left\{\nu\|S\|_1 + \frac{\beta}{2}\left\|X^{k+1} - S + L^k - \frac{\Lambda^{k+\frac{1}{2}}}{\beta}\right\|_F^2 + \frac{\sigma_2\beta}{2}\left\|S - S^k\right\|_F^2\right\},\\ L^{k+1} = \arg\min\limits_{L\succeq\mathbf{0}}\left\{\mu tr(L) + \frac{\beta}{2}\left\|X^{k+1} - S^k + L - \frac{\Lambda^{k+\frac{1}{2}}}{\beta}\right\|_F^2 + \frac{\sigma_2\beta}{2}\left\|L - L^k\right\|_F^2\right\},\\ \Lambda^{k+1} = \Lambda^{k+\frac{1}{2}} - s\beta(X^{k+1} - S^{k+1} + L^{k+1}). \end{array}\right. \tag{89}$$

Note that all the subproblems in (88) and (89) have closed formula solutions. Next, we take the scheme (88) for an example to show how to get the explicit solutions of the subproblem. By the first-order optimality condition of the X-subproblem in (88), we derive

$$\mathbf{0} = C - X^{-1} + \beta\left(X - S^k + L^k - \Lambda^k/\beta\right) + \sigma_1\beta(X - X^k)$$

which is equivalent to

$$(\sigma_1 + 1)\beta X^2 + \left[C + \beta(L^k - S^k) - \Lambda^k - \sigma_1\beta X^k\right]X - \mathbf{I} = \mathbf{0}.$$

Then, from the eigenvalue decomposition

$$U\mathrm{Diag}(\rho)U^{\mathsf{T}} = C + \beta(L^k - S^k) - \Lambda^k - \sigma_1\beta X^k,$$

where $\mathrm{Diag}(\rho)$ is a diagonal matrix with $\rho_i, i = 1, \cdots, n$, on the diagonal, we obtain that the solution of the X-subproblem in (88) is

$$X^{k+1} = U\mathrm{Diag}(\gamma)U^{\mathsf{T}},$$

where $\mathrm{Diag}(\gamma)$ is the diagonal matrix with diagonal elements

$$\gamma_i = \frac{-\rho_i + \sqrt{\rho_i^2 + 4(\sigma_1 + 1)\beta}}{2(\sigma_1 + 1)\beta}, \quad i = 1, 2, \cdots, n.$$

On the other hand, the S-subproblem in (88) is equivalent to

$$\begin{aligned} S^{k+1} &= \arg\min_{S}\left\{\nu\|S\|_1 + \frac{(\sigma_1+1)\beta}{2}\left\|S - \frac{X^k + L^k + \sigma_1 S^k - \Lambda^k/\beta}{(\sigma_1+1)}\right\|_F^2\right\}\\ &= \mathrm{Shrink}\left(\frac{X^k + L^k + \sigma_1 S^k - \Lambda^k/\beta}{(\sigma_1+1)}, \frac{\nu}{(\sigma_1+1)\beta}\right), \end{aligned}$$

where Shrink$(\cdot,\cdot)$ is the soft shrinkage operator (see e.g.[22]). Meanwhile, it is easy to verify that the L-subproblem is equivalent to

$$\begin{aligned} L^{k+1} &= \arg\min_{L\succeq \mathbf{0}} \tfrac{(\sigma_2+1)\beta}{2} \left\| L - \widetilde{L} \right\|_F^2 \\ &= V\mathrm{Diag}(\max\{\rho, \mathbf{0}\})V^\mathsf{T}, \end{aligned}$$

where $\max\{\rho, \mathbf{0}\}$ is taken component-wise and $V\mathrm{Diag}(\rho)V^\mathsf{T}$ is the eigenvalue decomposition of the matrix

$$\widetilde{L} = \frac{\sigma_2 L^k + S^{k+1} + \Lambda^{k+\frac{1}{2}}/\beta - X^{k+1} - \mu\mathbf{I}/\beta}{(\sigma_2+1)}.$$

5.2 Numerical results

In the following, we investigate the performance of several algorithms for solving the LVGGMS problem, where all the corresponding subproblems can be solved in a similar way as shown in the above analysis. For all algorithms, the maximal number of iterations is set as 1000, the starting iterative values are set as $(X^0, S^0, L^0, \Lambda^0) = (\mathbf{I}, \mathbf{2I}, \mathbf{I}, \mathbf{0})$, and motivated by Remark 2, the following stopping conditions are used

$$\begin{aligned} \mathrm{IER(k)} &:= \max\left\{ \left\|X^k - X^{k-1}\right\|_\infty, \left\|S^k - S^{k-1}\right\|_\infty, \left\|L^k - L^{k-1}\right\|_\infty \right\} \le \mathrm{TOL}, \\ \mathrm{OER(k)} &:= \frac{|F(X^k, S^k, L^k) - F^*|}{|F^*|} \le \mathrm{Tol}, \end{aligned}$$

together with $\mathrm{CER(k)} := \left\|X^k - S^k + L^k\right\|_F \le 10^{-4}$, where F^* is the approximate optimal objective function value obtained by running GS-ADMM (88) after 1000 iterations. In (87), we set $(\nu, \mu) = (0.005, 0.05)$ and the given data C is randomly generated by the following MATLAB code with $m = 100$, which are downloaded from S. Boyd's homepage[2]:

```
randn('seed',0); rand('seed',0); n=m; N=10*n;
Sinv=diag(abs(ones(n,1))); idx=randsample(n^2,0.001*n^2);
Sinv(idx)=ones(numel(idx),1); Sinv=Sinv+Sinv';
if min(eig(Sinv))<0
   Sinv=Sinv+1.1*abs(min(eig(Sinv)))*eye(n);
end
S=inv(Sinv);
D=mvnrnd(zeros(1,n),S,N); C=cov(D);
```

5.2.1 Performance of different versions of GS-ADMM

In the following, we denote

GS-ADMM (88) by "**GS-ADMM-I**";
GS-ADMM (89) by "**GS-ADMM-II**";
GS-ADMM (88) with $\sigma_2 = 0$ by "**GS-ADMM-III**";
GS-ADMM (89) with $\sigma_1 = 0$ by "**GS-ADMM-IV**".

[2] http://web.stanford.edu/∼boyd/papers/admm/covsel/covsel_example.html

GS-ADMM-I/β	Iter(k)	CPU(s)	CER	IER	OER
0.5	1000	15.29	7.2116e-8	*5.0083e-6*	3.2384e-10
0.2	493	8.58	1.4886e-8	9.8980e-8	5.7847e-11
0.1	254	4.24	1.6105e-8	9.7867e-8	5.6284e-11
0.08	202	3.27	1.7112e-8	9.8657e-8	5.6063e-11
0.07	175	3.03	1.7548e-8	9.7091e-8	5.4426e-11
0.06	146	2.42	1.9200e-8	9.9841e-8	5.4499e-11
0.05	115	**1.84**	1.9174e-8	8.8302e-8	4.4919e-11
0.03	**112**	2.21	1.7788e-7	9.9591e-8	2.2472e-11
0.01	270	4.50	6.4349e-7	9.9990e-8	2.5969e-10
0.006	424	7.57	1.0801e-6	9.8883e-8	5.0542e-10
0.004	604	10.74	1.6490e-6	9.9185e-8	8.7172e-10
GS-ADMM-II/β	Iter(k)	CPU(s)	CER	IER	OER
0.5	1000	15.80	8.8857e-8	*3.2511e-6*	4.0156e-10
0.2	603	11.35	3.7706e-9	9.9070e-8	1.2204e-12
0.1	312	4.93	6.0798e-9	9.9239e-8	2.3994e-12
0.08	250	4.40	7.1384e-9	9.6234e-8	2.8127e-12
0.07	217	3.42	8.2861e-9	9.8471e-8	3.1878e-12
0.06	183	3.09	9.7087e-8	9.8298e-8	3.4898e-12
0.05	147	2.85	1.1335e-8	9.1450e-8	3.3405e-12
0.03	**114**	**1.85**	1.5606e-7	9.1283e-8	1.9479e-11
0.01	271	4.70	6.2003e-7	9.6960e-8	2.4594e-10
0.006	424	7.38	1.0774e-6	9.8852e-8	5.0224e-10
0.004	604	10.01	1.6461e-6	9.9114e-8	8.6812 e-10
GS-ADMM-III/β	Iter(k)	CPU(s)	CER	IER	OER
0.5	579	9.36	1.2740e-8	9.9818e-8	5.2821e-11
0.2	247	5.52	1.2043e-8	9.6354e-8	4.5217e-11
0.1	125	2.14	1.1737e-8	9.5170e-8	3.6207e-11
0.08	97	1.55	1.2078e-8	9.7603e-8	2.8773e-11
0.07	82	1.36	1.1854e-8	9.5322e-8	1.6215e-11
0.06	**69**	**1.27**	1.2680e-8	8.2352e-8	1.5087e-11
0.05	71	1.40	9.1560e-8	9.8745e-8	8.1869e-12
0.03	110	1.71	1.8118e-7	9.4257e-8	2.7549e-11
0.01	271	4.46	6.3390e-7	9.7803e-8	2.5210e-10
0.006	424	6.92	1.0856e-6	9.9123e-8	5.0717e-10
0.004	604	10.11	1.6527	9.9275	8.7303e-10
GS-ADMM-IV/β	Iter(k)	CPU(s)	CER	IER	OER
0.5	1000	15.76	7.1259e-8	*2.6323e-6*	6.9956e-12
0.2	587	9.08	3.8200e-9	9.9214e-8	1.3291e-12
0.1	304	4.80	6.0296e-9	9.6197e-8	2.4309e-12
0.08	243	4.91	7.2062e-9	9.4484e-8	2.8670e-12
0.07	211	3.25	8.1772e-9	9.4133e-8	3.1477e-12
0.06	177	2.81	9.9510e-9	9.6911e-8	3.5342e-12
0.05	140	3.07	1.3067e-8	9.9446e-8	3.6691e-12
0.03	**115**	**1.80**	1.6886e-7	9.5844e-8	2.1829e-11
0.01	271	4.67	6.2006e-7	9.7151e-8	2.4927e-10
0.006	424	6.94	1.0758e-6	9.8755e-8	5.0454e-10
0.004	604	10.21	1.6454e-6	9.9088e-8	8.6995e-10

Table 1: Numerical results of GS-ADMM-I, II, III and IV with different β.

First, we would like to investigate the performance of the above different versions of GS-ADMM for solving the LVGGMS problem with variance of the penalty parameter β. The results are reported in Table 1 with $\text{TOL} = \text{Tol} = 1.0 \times 10^{-7}$, and $(\tau, s) = (0.8, 1.17)$. For GS-ADMM-I and GS-ADMM-II, $(\sigma_1, \sigma_2) = (2, 3)$. Here, "Iter" and "CPU" denote the iteration number and the CPU time in seconds, and the bold letter indicates the best result of each algorithm. From Table 1, we can observe that:

– Both the iteration number and the CPU time of all the algorithms have a similar changing pattern, which decreases originally and then increases along with the decrease of the value of β.
– For the same value of β, the results of GS-ADMM-III are slightly better than other algorithms in terms of the iteration number, CPU time, and the feasibility errors CER, IER and OER.
– GS-ADMM-III with $\beta = 0.5$ can terminate after 579 iterations to achieve the tolerance 10^{-7}, while the other algorithms with $\beta = 0.5$ fail to achieve this tolerance within given number of iterations.

In general, the algorithm (88) with $\beta = 0.06$ performs better than other cases. Hence, in the following experiments for GS-ADMM, we adapt GS-ADMM-III with default $\beta = 0.06$. Also note that $\sigma_2 = 0$, which is not allowed by the algorithms discussed in [15].

(τ, s)	Iter(k)	CPU(s)	CER	IER	OER
(1, -0.8)	256	4.20	9.8084e-5	7.8786e-6	1.1298e-7
(1, -0.6)	144	2.39	5.7216e-5	9.9974e-6	3.8444e-8
(1, -0.4)	105	1.80	3.5144e-5	9.7960e-6	1.3946e-8
(1, -0.2)	84	1.45	2.3513e-5	9.3160e-6	6.4220e-9
(1, 0)	70	1.14	1.7899e-5	9.4261e-6	3.9922e-9
(1, 0.2)	61	0.98	1.3141e-5	8.9191e-6	1.7780e-9
(1, 0.4)	54	0.88	1.0549e-5	9.1564e-6	4.6063e-10
(1, 0.6)	49	0.82	9.0317e-5	9.4051e-6	2.7938e-9
(1, 0.8)	**49**	**0.80**	3.5351e-5	8.0885e-6	1.4738e-9
(-0.8, 1)	229	3.91	9.9324e-5	8.4462e-6	1.9906e-7
(-0.6, 1)	127	2.06	6.1118e-5	9.6995e-6	7.8849e-8
(-0.4, 1)	96	1.61	3.4111e-5	9.6829e-6	2.7549e-8
(-0.2, 1)	79	1.30	2.2004e-5	9.6567e-6	1.2015e-8
(0, 1)	67	1.16	1.6747e-5	9.9244e-6	6.2228e-9
(0.2, 1)	59	0.93	1.2719e-5	9.4862e-6	2.9997e-9
(0.4, 1)	53	0.88	1.0253e-5	9.3461e-6	3.4811e-10
(0.6, 1)	49	0.85	8.0343e-6	8.8412e-6	2.9837e-9
(0.8, 1)	**49**	**0.81**	3.3831e-6	8.1998e-6	2.1457e-9
(1.6, -0.3)	60	0.99	1.2111e-5	9.4583e-6	1.1705e-9
(1.6, -0.6)	74	1.22	1.8012e-5	9.6814e-6	2.7562e-9
(1.5, -0.8)	97	1.68	3.1310e-5	9.8972e-6	1.4911e-8
(1.3, 0.3)	50	0.83	8.5476e-6	8.9655e-6	3.4389e-10
(0.2, 0.5)	87	1.44	2.7160e-5	9.4503e-6	1.7906e-8
(0.4, 0.9)	56	0.98	1.1060e-5	9.1081e-6	1.7179e-9
(0.8, 1.17)	49	0.86	1.5419e-6	8.5023e-6	2.5529e-9
(0, 1.618)	50	0.90	5.5019e-6	8.6980e-6	1.4722e-9
(0.9, 1.09)	**49**	**0.78**	1.4874e-6	8.4766e-6	2.2194e-9
(0.1, 0.1)	229	4.42	9.8698e-5	8.3622e-6	2.3575e-7
(0.2, 0.2)	130	2.32	5.5559e-5	9.9888e-6	7.5859e-8
(0.3, 0.3)	97	1.75	3.4344e-5	9.9362e-6	2.8190e-8
(0.4, 0.4)	79	1.43	2.4256e-5	9.8539e-6	1.2790e-8
(0.5, 0.5)	68	1.15	1.6805e-5	9.2144e-6	5.5121e-9
(0.6, 0.6)	59	0.98	1.3862e-5	9.7793e-6	2.8580e-9
(0.7, 0.7)	53	0.91	1.1091e-5	9.6433e-6	3.9013e-12
(0.8, 0.8)	49	0.84	8.4235e-6	8.9432e-6	3.0519e-9
(0.9, 0.9)	**49**	**0.83**	3.4493e-6	8.1314e-6	1.8888e-9

Table 2: Numerical results of GS-ADMM-III with different stepsizes (τ, s).

Second, we investigate how the stepsizes $(\tau, s) \in \mathcal{G}$ with different values would affect the performance of GS-ADMM-III. Table 2 reports the comparison results with variance of (τ, s) for TOL = Tol = 1.0×10^{-5}. One obvious observation from Table 2 is that both the iteration number and the CPU time decrease along with the increase of s (or τ) for fixed value of τ (or s), which indicates that the stepsizes of $(\tau, s) \in \mathcal{G}$ could influence the performance of GS-ADMM significantly. In addition, the results in Table 2 also indicate that using more flexible but with both relatively larger stepsizes τ and s of the dual variables often gives the best convergence speed. Comparing all the reported results in Table 2, by setting $(\tau, s) = (0.9, 1.09)$, GS-ADMM-III gives the relative best performance for solving the problem (87).

5.2.2 Comparison of GS-ADMM with other state-of-the-art algorithms

In this subsection, we would like to carry out some numerical comparison of solving the problem (87) by using GS-ADMM-III and the other four methods:

	TOL	Tol	Iter(k)	CPU(s)	CER	IER	OER
	1e-3	1e-7	**33**	**0.46**	8.3280e-5	2.5770e-4	4.0973e-9
		1e-12	**83**	**1.16**	9.5004e-9	1.0413e-8	8.3240e-13
GS-ADMM-III	1e-6	1e-8	**58**	**0.84**	8.3812e-7	9.0995e-7	7.6372e-11
		1e-14	**108**	**1.55**	1.0936e-10	1.2072e-10	9.5398e-15
	1e-9	1e-7	**97**	**1.39**	7.7916e-10	8.5759e-10	6.8775e-14
		1e-15	**118**	**1.72**	1.8412e-11	2.0361e-11	6.6557e-16
	1e-3	1e-7	62	0.88	9.6422e-5	3.6934e-5	5.8126e-8
		1e-12	187	2.74	9.4636e-9	3.4868e-9	9.4447e-13
PJALM	1e-6	1e-8	111	1.67	2.4977e-6	9.4450e-7	4.1335e-10
		1e-14	249	3.63	1.0173e-10	3.7225e-11	8.6506e-15
	1e-9	1e-7	205	3.06	2.5369e-9	9.3210e-10	2.4510e-13
		1e-15	276	4.08	1.4143e-11	5.2002e-11	6.6543e-16
	1e-3	1e-7	62	0.85	4.8548e-5	1.7123e-5	9.3737e-8
		1e-12	176	2.64	2.7059e-9	9.7709e-10	9.1783e-13
HTY	1e-6	1e-8	92	1.35	2.7184e-6	9.6661e-7	4.0385e-9
		1e-14	223	3.15	1.1042e-10	4.0226e-11	9.3091e-15
	1e-9	1e-7	176	2.78	2.7059e-9	9.7709e-10	9.1783e-13
		1e-15	243	3.70	2.8377e-11	1.0533e-11	4.4329e-16
	1e-3	1e-7	***61***	***0.82***	7.4082e-5	3.3954e-5	1.5195e-9
		1e-12	***127***	***1.84***	5.8944e-8	1.6729e-7	1.3001e-13
GR-PPA	1e-6	1e-8	108	1.52	5.5130e-7	6.2676e-7	3.4315e-11
		1e-14	***172***	***2.56***	2.9521e-10	8.3742e-10	8.8742e-16
	1e-9	1e-7	***167***	***2.42***	5.3963e-10	7.3383e-10	3.7050e-14
		1e-15	***172***	***2.41***	2.9521e-10	8.3742e-10	8.8742e-16
	1e-3	1e-7	*40*	*0.55*	9.8495e-5	1.2096e-4	2.7440e-8
		1e-12	*112*	*1.53*	7.1036e-9	4.8224e-9	8.9763e-13
T-ADMM	1e-6	1e-8	*72*	*1.02*	1.3128e-6	8.9570e-7	2.3510e-10
		1e-14	*147*	*2.12*	7.4334e-11	5.0156e-11	9.7617e-15
	1e-9	1e-7	*125*	*1.70*	1.3053e-9	8.8746e-10	1.5974e-13
		1e-15	*160*	*2.01*	1.3669e-11	9.4374e-12	6.6557e-16

Table 3: Comparative results of different algorithms under different tolerances.

– The Proximal Jacobian Decomposition of ALM [17] (denoted by "PJALM");
– The splitting method in [14] (denoted by "HTY");
– The generalized parametrized proximal point algorithm [1] (denoted by "GR-PPA").
– The twisted version of the proximal ADMM [23] (denoted by "T-ADMM").

We set $(\tau, s) = (0.9, 1.09)$ for GS-ADMM-III and the parameter $\beta = 0.05$ for all the comparison algorithms. The default parameter $\mu = 2.01$ and $H = \beta \mathbf{I}$ are used for HTY [14]. As suggested by the theory and numerical experiments in [17], the proximal parameter is set as 2 for PJALM. As shown in [1], the relaxation factor of GR-PPA is set as 1.8 and other default parameters are chosen as

$$(\sigma_1, \sigma_2, \sigma_3, s, \tau, \varepsilon) = \left(0.178, 0.178, 0.178, 10, \frac{\sqrt{5}-1}{2}, \frac{\sqrt{5}-1}{2}\right).$$

For T-ADMM, the symmetric matrices therein are chosen as $M_2 = M_2 = v\mathbf{I}$ with $v = \beta$ and the correction factor is set as $a = 1.6$ [23]. The results obtained by the above algorithms under different tolerances are reported in Table 3. With fixed tolerance TOL $= 10^{-9}$ and Tol $= 10^{-15}$, the convergence behavior of the error measurements IER(k) and OER(k) by the five algorithms using different starting points are shown in Figs. 3-5. From Table 3 and Figures 3-5, we may have the following observation:

– Under all different tolerances, GS-ADMM-III performs significantly better than other four algorithms in both the number of iterations and CPU time.
– GR-PPA is slightly better than PJALM and HTY, and T-ADMM outperforms PJALM, HTY and GR-PPA.
– the convergence curves in Figs. 3-5 illustrate that using different starting points, GS-ADMM-III also converges fastest among the comparing methods.

All these numerical results demonstrate the effectiveness and robustness of GS-ADMM-III, which is perhaps due to the symmetric updating of the Lagrange multipliers and the proper choice of the stepsizes.

6 Conclusion

Since the direct extension of ADMM in a Gauss-Seidel fashion for solving the three-block separable convex optimization problem is not necessarily convergent analyzed by Chen et al. [3], there has been a constantly increasing interest in developing and improving the theory of the ADMM for solving the multi-block separable convex optimization. In this paper, we propose an algorithm, called GS-ADMM, which could solve the general model (1) by taking advantages of the multi-block structure. In our GS-ADMM, the Gauss-Seidel fashion is taken for updating the two grouped variables, while the block variables within each group are updated in a Jacobi scheme, which would make the algorithm more be attractive and effective for solving big size problems. We provide a new convergence domain for the stepsizes of the dual variables, which is significantly larger than the convergence domains given in the literature. Global convergence as well as the $\mathcal{O}(1/t)$ ergodic convergence rate of the GS-ADMM is established. In addition, two special cases of GS-ADMM, which allows one of the proximal parameters to be zero, are also discussed.

This paper simplifies the analysis in [16] and provides an easy way to analyze the convergence of the symmetric ADMM. Our preliminary numerical experiments show that with proper choice of parameters, the performance of the GS-ADMM could be very promising. Besides, from the presented convergence analysis, we can see that the theories in the paper can be naturally extended to use more general proximal terms, such as letting $P_i^k(x_k) := \frac{\beta}{2}\|x_i - x_i^k\|_{\mathbf{P}_i}$ and $Q_j^k(y_j) := \frac{\beta}{2}\|y_j - y_j^k\|_{\mathbf{Q}_j}$ in (7), where $\mathbf{P}_i$ and $\mathbf{Q}_i$ are matrices such that $\mathbf{P}_i \succ (p-1)A_i^\mathsf{T} A_i$ and $\mathbf{Q}_j \succ (q-1)B_j^\mathsf{T} B_j$ for all $i = 1, \cdots, p$ and $j = 1, \cdots, q$. Finally, the different ways of partitioning the variables of the problem also gives the flexibility of GS-ADMM.

Acknowledgements. The authors would like to thank the anonymous referees for providing very constructive comments. The first author also wish to thank Prof. Defeng Sun at National University of Singapore for his valuable discussions on ADMM and Prof. Pingfan Dai at Xi'an Jiaotong University for discussion on an early version of the paper.

References

1. Bai, J.C., Li, J.C., Li, J.F.: A novel parameterized proximal point algorithm with applications in statistical learning. Optimization Online, March, (2017) http://www.optimization-online.org/DB_HTML/2017/03/5901.html.
2. Chandrasekaran, V., Parrilo, P.A., Willsky, A.S.: Latent variable graphical model selection via convex optimization. Ann. Stat. 40, 1935-1967 (2012)
3. Chen, C.H., He, B.S., Ye, Y.Y., Yuan, X.M.: The direct extension of ADMM for multi-block minimization problems is not necessarily convergent. Math. Program. 155, 57-79 (2016)
4. Dong, B., Yu, Y., Tian, D.D.: Alternating projection method for sparse model updating problems. J. Eng. Math. 93, 159-173 (2015)
5. Fortin, M., Glowinski, R.: Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems. P.299-331. North-Holland, Amsterdam (1983)
6. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer Ser. Oper. Res. 1, Springer, New York (2003)
7. Fortin, M.: Numerical methods for Nonlinear Variational Problems. Springer-Verlag, New York (1984)
8. Glowinski, R. Marrocco, A.: Approximation parélements finis d'rdre un et résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. Rev. Fr. Autom. Inform. Rech. Opér. Anal. Numér. 2, 41-76 (1975)
9. Gu, Y., Jiang, B., Han, D.: A semi-proximal-based strictly contractive Peaceman-Rachford splitting method. arXiv:1506.02221, (2015)
10. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. 4, 303-320 (1969)
11. He, B.S., Yuan, X.M.: On the $\mathcal{O}(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. SIAM J. Num. Anal. 50, 700-709 (2012)
12. He, B.S., Liu, H., Wang, Z.R., Yuan, X.M.: A strictly contractive Peaceman-Rachford splitting method for convex programming. SIAM J. Optim. 24, 1011-1040 (2014)
13. He, B.S., Hou, L.S., Yuan, X.M.: On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming. SIAM J. Optim. 25, 2274-2312 (2015)
14. He, B.S., Tao, M., Yuan, X.M.: A splitting method for separable convex programming. IMA J. Numer. Anal. 31, 394-426 (2015)
15. He, B.S., Yuan, X.M.: Block-wise alternating direction method of multipliers for multiple-block convex programming and beyond. SMAI J. Comput. Math. 1, 145-174 (2015)
16. He, B.S., Ma, F., Yuan, X.M.: Convergence study on the symmetric version of ADMM with larger step sizes. SIAM J. Imaging Sci. 9, 1467-1501 (2016)
17. He, B.S., Xu, H.K., Yuan, X.M.: On the proximal Jacobian decomposition of ALM for multiple-block separable convex minimization problems and its relationship to ADMM. J. Sci. Comput. 66, 1204-1217 (2016)
18. He, B.S.: On the convergence properties of alternating direction method of multipliers. Numer. Math.: A Journal of Chinese Universities(Chinese Series) 39, 81-96 (2017).
19. Liu, Z.S., Li, J.C., Li, G., Bai, J.C., Liu, X.N.: A new model for sparse and low-rank matrix decomposition. J. Appl. Anal. Comput. 7, 600-616 (2017)
20. Ma, S.Q.: Alternating proximal gradient method for convex minimization. J. Sci. Comput. 68, 546-572 (2016)
21. Rothman, A.J., Bickel, P.J., Levina, E., Zhu, J.: Sparse permutation invariant covariance estimation. Electron. J. Stat. 2, 494-515 (2008)
22. Tao, M., Yuan, X.M.: Recovering low-rank and sparse components of matrices from incomplete and noisy observations. SIAM J. Optim. 21, 57-81 (2011)
23. Wang, J.J., Song, W.: An algorithm twisted from generalized ADMM for multi-block separable convex minimization models. J. Comput. Appl. Math. 309, 342-358 (2017)

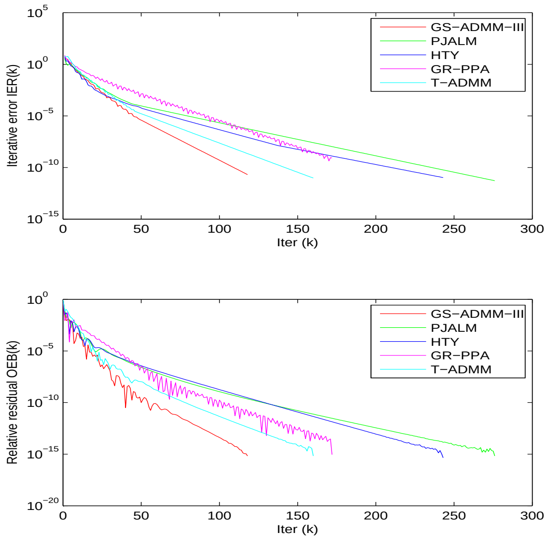

Fig. 1 Convergence curves of IER and OER with initial values $(X^0, S^0, L^0, \Lambda^0) = (\mathbf{I}, 2\mathbf{I}, \mathbf{I}, \mathbf{0})$.

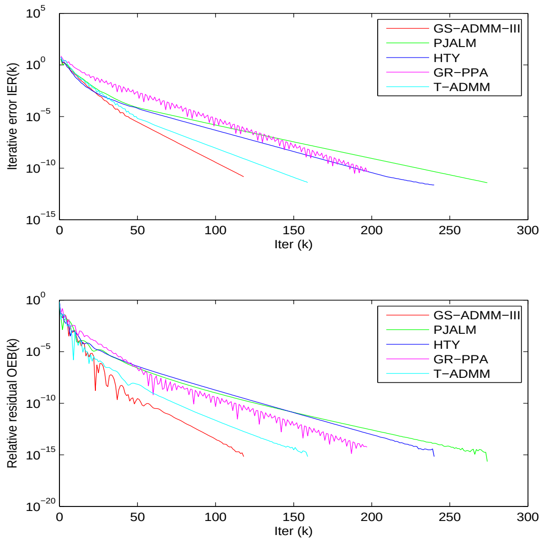

Fig. 2 Convergence curves of IER and OER with initial values $(X^0, S^0, L^0, \Lambda^0) = (\mathbf{I}, \mathbf{I}, \mathbf{0}, \mathbf{0})$.

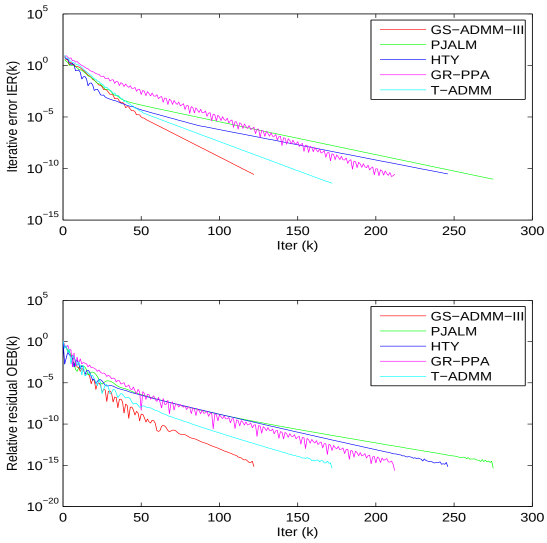

Fig. 3 Convergence curves of IER and OER with initial values $(X^0, S^0, L^0, \Lambda^0) = (\mathbf{I}, 4\mathbf{I}, 3\mathbf{I}, \mathbf{0})$.