

# On the Equivalence of Inexact Proximal ALM and ADMM for a Class of Convex Composite Programming

Liang Chen · Xudong Li · Defeng Sun · Kim-Chuan Toh

Received: date / Accepted: date

**Abstract** In this paper, we show that for a class of linearly constrained convex composite optimization problems, an (inexact) symmetric Gauss-Seidel based majorized multi-block proximal alternating direction method of multipliers (ADMM) is equivalent to an *inexact* proximal augmented Lagrangian method (ALM). This equivalence not only provides new perspectives for understanding some ADMM-type algorithms but also supplies meaningful guidelines on implementing them to achieve better computational efficiency. Even for the two-block case, a by-product of this equivalence is the convergence of the whole sequence generated by the classic ADMM with a step-length that exceeds the conventional upper bound of  $(1 + \sqrt{5})/2$ , if one part of the objective is linear. This is exactly the problem setting in which the very first convergence analysis of ADMM was conducted by Gabay and Mercier in 1976, but, even under notably stronger assumptions, only the convergence of the primal sequence was known. A collection of illustrative examples are provided to demonstrate the breadth of applications for which our results can be used. Numerical experiments on solving a large number of linear and convex quadratic semidefinite programming problems are conducted to illustrate how the theoretical results established here can lead to improvements on the corresponding practical implementations.

**Keywords** Alternating direction method of multipliers · Augmented Lagrangian method · Symmetric Gauss-Seidel decomposition · Proximal term

**Mathematics Subject Classification (2010)** 90C25 · 65K05 · 90C06 · 49M27 · 90C20

## 1 Introduction

Let  $\mathbb{X}$ ,  $\mathbb{Y}$  and  $\mathbb{Z}$  be three finite-dimensional real Hilbert spaces each endowed with an inner product denoted by  $\langle \cdot, \cdot \rangle$  and its induced norm denoted by  $\| \cdot \|$ , where  $\mathbb{Y} := \mathbb{Y}_1 \times \cdots \times \mathbb{Y}_s$  is the Cartesian product of  $s$  finite-dimensional real Hilbert spaces  $\mathbb{Y}_i, i = 1, \dots, s$ , each endowed with the inner product, as well as the induced

---

The research of the first author was supported by the National Natural Science Foundation of China (11801158, 11871205) and the Fundamental Research Funds for the Central Universities in China. The research of the second author was supported by the Fundamental Research Funds for the Central Universities in China. The research of the third author was supported in part by a start-up research grant from the Hong Kong Polytechnic University. The research of the fourth author was supported in part by the Ministry of Education, Singapore, Academic Research Fund (R-146-000-257-112).

---

Liang Chen  
College of Mathematics and Econometrics, Hunan University, Changsha 410082, China; and  
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong  
E-mail: chl@hnu.edu.cn

Xudong Li  
School of Data Science, Fudan University, Shanghai 200433, China; and  
Shanghai Center for Mathematical Sciences, Fudan University, Shanghai 200433, China  
E-mail: lixudong@fudan.edu.cn

Defeng Sun  
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong  
E-mail: defeng.sun@polyu.edu.hk

Kim-Chuan Toh  
Department of Mathematics, and Institute of Operations Research and Analytics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076, Singapore  
E-mail: mattohc@nus.edu.sg

norm, inherited from  $\mathbb{Y}$ . For any given  $y \in \mathbb{Y}$ , we can write  $y = (y_1; \dots; y_s)$  with  $y_i \in \mathbb{Y}_i, \forall i = 1, \dots, s$ . Here, and throughout this paper, we use the notation  $(y_1; \dots; y_s)$  to mean that the vectors  $y_1, \dots, y_s$  are written symbolically in a column format.

In this paper, we shall focus on the following multi-block convex composite optimization problem

$$\min_{y \in \mathbb{Y}, z \in \mathbb{Z}} \{p(y_1) + f(y) - \langle b, z \rangle \mid \mathcal{F}^*y + \mathcal{G}^*z = c\}, \quad (1.1)$$

where  $p : \mathbb{Y}_1 \rightarrow (-\infty, +\infty]$  is a (possibly nonsmooth) closed proper convex function,  $f : \mathbb{Y} \rightarrow (-\infty, +\infty)$  is a continuously differentiable convex function whose gradient is Lipschitz continuous,  $b \in \mathbb{Z}$  and  $c \in \mathbb{X}$  are the given data, and  $\mathcal{F}^*$  and  $\mathcal{G}^*$  are the adjoints of the given linear mappings  $\mathcal{F} : \mathbb{X} \rightarrow \mathbb{Y}$  and  $\mathcal{G} : \mathbb{X} \rightarrow \mathbb{Z}$ , respectively. Despite the simple appearance of problem (1.1), we shall see in the next section that this model actually encompasses various important classes of convex optimization problems in both classical core convex programming as well as recently emerged models from a broad range of real-world applications. A quintessential example of problem (1.1) is the dual of the following convex composite quadratic programming

$$\min_x \left\{ \psi(x) + \frac{1}{2} \langle x, Qx \rangle - \langle c, x \rangle \mid \mathcal{G}x = b \right\}, \quad (1.2)$$

where  $\psi : \mathbb{X} \rightarrow (-\infty, +\infty]$  is a closed proper convex function,  $Q : \mathbb{X} \rightarrow \mathbb{X}$  is a self-adjoint positive semidefinite linear operator,  $\mathcal{G} : \mathbb{X} \rightarrow \mathbb{Z}$  is a linear mapping, and  $c \in \mathbb{X}$ ,  $b \in \text{Range}(\mathcal{G})$  (i.e.,  $b$  is in the range space of the linear operator  $\mathcal{G}$ ) are the given data. The dual of problem (1.2) in the minimization form can be written as follows:

$$\min_{y_1, y_2, z} \left\{ \psi^*(y_1) + \frac{1}{2} \langle y_2, Qy_2 \rangle - \langle b, z \rangle \mid y_1 + Qy_2 - \mathcal{G}^*z = c \right\}, \quad (1.3)$$

where  $\psi^*$  is the Fenchel conjugate of  $\psi$ ,  $y_1 \in \mathbb{X}$ ,  $y_2 \in \mathbb{X}$  and  $z \in \mathbb{Z}$ , so that problem (1.3) constitutes an instance of problem (1.1).

To solve problem (1.1), one of the most preferred approaches is the augmented Lagrangian method (ALM) initiated by Hestenes [23] and Powell [42], and elegantly studied for general (without taking into account of the multi-block structure) convex optimization problems in the seminal work of Rockafellar [45]. Given a penalty parameter  $\sigma > 0$ , the augmented Lagrangian function corresponding to problem (1.1) is defined by

$$L_\sigma(y, z; x) := p(y_1) + f(y) - \langle b, z \rangle + \langle x, \mathcal{F}^*y + \mathcal{G}^*z - c \rangle + \frac{\sigma}{2} \|\mathcal{F}^*y + \mathcal{G}^*z - c\|^2, \\ \forall (x, y, z) \in \mathbb{X} \times \mathbb{Y} \times \mathbb{Z}.$$

Starting from a given initial multiplier  $x^0 \in \mathbb{X}$ , the ALM performs the following steps at the  $k$ -th iteration:

- (1) compute  $(y^{k+1}, z^{k+1})$  to (approximately) minimize the function  $L_\sigma(y, z; x^k)$ , and
- (2) update the multipliers  $x^{k+1} := x^k + \tau\sigma(\mathcal{F}^*y^{k+1} + \mathcal{G}^*z^{k+1} - c)$ , where  $\tau \in (0, 2)$  is the step-length.

While one would really want to solve  $\min_{y, z} L_\sigma(y, z; x^k)$  as it is without modifying the augmented Lagrangian function, it can be expensive to minimize  $L_\sigma(y, z; x^k)$  with respect to both  $y$  and  $z$  simultaneously, due to the coupled quadratic term in  $y$  and  $z$ . Thus, in practice, unless the ALM is converging rapidly, one would generally want to replace the augmented Lagrangian subproblem with an easier-to-solve surrogate by modifying the augmented Lagrangian function to decouple the minimization with respect to  $y$  and  $z$ . Such a modification is especially desirable during the initial phase of the ALM when its local superlinear convergence has yet to kick in. The most obvious approach to decouple the subproblem for obtaining  $(y^{k+1}, z^{k+1})$  is to add to  $L_\sigma(y, z; x^k)$  the proximal term  $\frac{\sigma}{2} \|(y; z) - (y^k; z^k)\|_\Lambda^2$ , where  $\Lambda = \lambda^2 \mathcal{I} - (\mathcal{F}; \mathcal{G})(\mathcal{F}; \mathcal{G})^*$  with  $\lambda$  being the largest singular value of  $(\mathcal{F}; \mathcal{G})$  and  $\mathcal{I}$  being the identity operator in  $\mathbb{Y} \times \mathbb{Z}$ . However, such a modification to the augmented Lagrangian function is generally too drastic and has the undesirable effect of significantly slowing down the convergence of the ALM [6, Section 7]. This naturally leads us to the important question on what is an appropriate proximal term to add to  $L_\sigma(y, z; x^k)$  such that the ALM subproblem is easier to solve while at the same time it is less drastic than the obvious choice we have just mentioned in the previous sentences.

We shall show in this paper that by adding an appropriately designed proximal term to  $L_\sigma(y, z; x^k)$ , we can reduce the computation of the modified ALM subproblem to sequentially updating  $y$  and  $z$  via computing

$$y^{k+1} \approx \min_y \{L_\sigma(y, z^k; x^k)\} \quad \text{and} \quad z^{k+1} \approx \min_z \{L_\sigma(y^{k+1}, z; x^k)\}.$$

The reader would have observed that the resulting proximal ALM updating scheme is the same as the classic two-block ADMM (pioneered by Glowinski and Marroco [21] and Gabay and Mercier [18]) that is applied to

problem (1.1). However, there is a *crucial difference* in that our convergence result holds true for the step-length  $\tau$  in the range  $(0, 2)$ , whereas the classic two-block ADMM only allows the step-length to be in the interval  $(0, (1 + \sqrt{5})/2)$  if the convergence of the full sequence generated by the algorithm is required. It is important to note that even with the sequential minimization of  $y$  and  $z$  in the modified ALM subproblem, the minimization subproblem with respect to  $y$  can still be very difficult to solve due to the coupling of the blocks  $y_1, \dots, y_s$  in (1.1). One of the main contributions we made in this paper is to show that by majorizing the function  $f(y)$  at  $y^k$  with a quadratic function and by adding an extra proximal term that is derived based on the block symmetric Gauss-Seidel (sGS) decomposition theorem [32] for the quadratic term associated with  $y$ , we are able to update the sub-blocks in  $y$  individually in a symmetric Gauss-Seidel fashion. A crucial implication of this result is that the (inexact) block sGS decomposition based multi-block majorized ADMM is equivalent to an inexact majorized proximal ALM. Consequently, we are able to prove the convergence of the whole sequence generated by the former even when the step-length is in the range  $(0, 2)$ .

In this paper, we shall not delve into the vast literature on both ALM and ADMM, as well as their variants, and their relationships to the proximal point method and operator splitting methods. They are simply too abundant for us to list even a few of them here. Thus, we shall only refer to those that are most relevant for our work in this paper. Here we should mention that many attempts have been made in recent years on designing convergent multi-block ADMM-type algorithms that can outperform the directly extended multi-block (proximal ADMM) numerically. While the latter is not guaranteed to converge even under the strong assumption that  $f \equiv 0$ , paradoxically its practical numerical performance is often better than many convergent variants that have been developed in the past; see for example [48]. Against this backdrop, we should mention that the ADMM-type algorithms that have been progressively designed in [48, 30, 6] not only come with convergence guarantee but they have also been demonstrated to have superior numerical performance than the directly extended ADMM, at least for a large number of convex conic programming problems. More recently, those algorithms have found applications in various areas [1, 2, 9, 17, 27, 31, 53, 55, 56]. Among those algorithms, the most general and versatile one is the recently developed inexact majorized multi-block proximal ADMM in Chen et al. [6], which we shall briefly describe in the next paragraph.

Under the assumption that the gradient of  $f$  is Lipschitz continuous, we know that one can specify a fixed self-adjoint positive semidefinite linear operator  $\widehat{\Sigma}^f : \mathbb{Y} \rightarrow \mathbb{Y}$  and define at each  $y' \in \mathbb{Y}$  the following convex quadratic function

$$\widehat{f}(y, y') := f(y') + \langle \nabla f(y'), y - y' \rangle + \frac{1}{2} \|y - y'\|_{\widehat{\Sigma}^f}^2, \quad \forall y \in \mathbb{Y}, \quad (1.4)$$

such that

$$f(y) \leq \widehat{f}(y, y'), \quad \forall y, y' \in \mathbb{Y} \quad \text{and} \quad f(y') = \widehat{f}(y', y'), \quad \forall y' \in \mathbb{Y}.$$

Thus, we say that at each  $y' \in \mathbb{Y}$ , the function  $\widehat{f}(\cdot, y')$  constitutes a majorization of the function  $f$ . Let  $\sigma > 0$  be the penalty parameter. Based on the notion of majorization described above, the *majorized augmented Lagrangian* function of problem (1.1) is defined by

$$\begin{aligned} \widehat{\mathcal{L}}_\sigma(y, z; (x, y')) &:= p(y_1) + \widehat{f}(y, y') - \langle b, z \rangle + \langle \mathcal{F}^* y + \mathcal{G}^* z - c, x \rangle \\ &\quad + \frac{\sigma}{2} \|\mathcal{F}^* y + \mathcal{G}^* z - c\|^2, \quad \forall (y, z, x, y') \in \mathbb{Y} \times \mathbb{Z} \times \mathbb{X} \times \mathbb{Y}. \end{aligned} \quad (1.5)$$

Let  $(x^0, y^0, z^0) \in \mathbb{X} \times \mathbb{Y} \times \mathbb{Z}$  be a given initial point with  $y_1^0 \in \text{dom } p$ , and  $\mathcal{D} : \mathbb{Y} \rightarrow \mathbb{Y}$  be a given self-adjoint linear operator for the purpose of facilitating the computations of the subproblems. For convenience, we define

$$\mathcal{L}_\sigma(y, z; (x, y')) := \widehat{\mathcal{L}}_\sigma(y, z; (x, y')) + \frac{1}{2} \|y - y'\|_{\mathcal{D}}^2, \quad \forall (y, z, x, y') \in \mathbb{Y} \times \mathbb{Z} \times \mathbb{X} \times \mathbb{Y}, \quad (1.6)$$

where  $\|y\|_{\mathcal{D}}^2 := \langle y, \mathcal{D}y \rangle$ ,  $\forall y \in \mathbb{Y}$  even if  $\mathcal{D}$  is not positive semidefinite, and denote for any  $y = (y_1; \dots; y_s) \in \mathbb{Y}_1 \times \dots \times \mathbb{Y}_s$ ,

$$y_{<i} := (y_1; \dots; y_{i-1}) \quad \text{and} \quad y_{>i} := (y_{i+1}; \dots; y_s), \quad \forall i = 1, \dots, s.$$

Then, the  $k$ -th step of the (inexact) block sGS decomposition based majorized multi-block proximal ADMM, which was adapted for solving problem (1.1) from [6], takes the following form

$$\begin{cases} y_i^{k+\frac{1}{2}} \approx \arg \min_{y_i \in \mathbb{Y}_i} \mathcal{L}_\sigma(y_{<i}^k; y_i; y_{>i}^{k+\frac{1}{2}}), z^k; (x^k, y^k), & i = s, \dots, 2; \\ y_i^{k+1} \approx \arg \min_{y_i \in \mathbb{Y}_i} \mathcal{L}_\sigma(y_{<i}^{k+1}; y_i; y_{>i}^{k+\frac{1}{2}}), z^k; (x^k, y^k), & i = 1, \dots, s; \\ z^{k+1} \approx \arg \min_{z \in \mathbb{Z}} \mathcal{L}_\sigma(y^{k+1}, z; (x^k, y^k)); \\ x^{k+1} = x^k + \tau \sigma (\mathcal{F}^* y^{k+1} + \mathcal{G}^* z^{k+1} - c), \end{cases} \quad (1.7)$$

where  $\tau \in (0, (1 + \sqrt{5})/2)$  was allowed in [6]. As one can observe from (1.5) and (1.7), the quadratic majorization technique in Li et al. [29] was used to replace the original augmented Lagrangian function by the majorized augmented Lagrangian function. This in turn enables us to employ the inexact block sGS decomposition technique in Li et al. [32] to sequentially update the sub-blocks of  $y$  individually. More importantly, the algorithm is highly flexible in that all the subproblems are allowed to be solved approximately to overcome possible numerical obstacles such as, for example, when iterative solvers must be employed to solve large-scale linear systems to overcome extreme memory requirement and prohibitive computing cost. It has already been demonstrated in [6] that the inexact block sGS decomposition based multi-block ADMM (1.7) is far superior to the directly extended ADMM in solving high-dimensional linear and convex quadratic semidefinite programming with the step-length in (1.7) being restricted to be less than  $(1 + \sqrt{5})/2$ .

Our focus in this paper is to investigate whether the framework in (1.7) can be proven to be convergent for problem (1.1) when the step-length  $\tau$  is in the range  $(0, 2)$ . In particular, we will show that the inexact block sGS decomposition based multi-block ADMM (1.7) is equivalent to an *inexact* majorized proximal ALM in the sense that computations of  $y^{k+1}$ ,  $z^{k+1}$  and  $x^{k+1}$  in (1.7) can equivalently be written as

$$\begin{cases} (y^{k+1}, z^{k+1}) \approx \arg \min_{(y, z) \in \mathbb{Y} \times \mathbb{Z}} \left\{ \mathcal{L}_\sigma(y, z; (x^k, y^k)) + \frac{1}{2} \|(y; z) - (y^k; z^k)\|_{\mathcal{T}}^2 \right\}; \\ x^{k+1} = x^k + \tau \sigma(\mathcal{F}^* y^{k+1} + \mathcal{G}^* z^{k+1} - c), \end{cases}$$

where  $\mathcal{T} : \mathbb{Y} \times \mathbb{Z} \rightarrow \mathbb{Y} \times \mathbb{Z}$  is a self-adjoint (not necessarily positive definite) linear operator whose precise definition will be given later, and  $\|(y; z)\|_{\mathcal{T}}^2 := \langle (y; z), \mathcal{T}(y; z) \rangle$ ,  $\forall (y, z) \in \mathbb{Y} \times \mathbb{Z}$ . This connection not only provides new theoretical perspectives for analyzing multi-block ADMM-type algorithms, but also has the potential of allowing them to achieve even better computational efficiency since a larger step-length beyond  $(1 + \sqrt{5})/2$  can now be taken in (1.7), without adding any extra conditions or any additional verification steps such as those extensively used in [48, 30, 6, 5].

The main contributions of this paper are as follows.

- We derive the equivalence of an (inexact) block sGS decomposition based multi-block majorized proximal ADMM to an inexact majorized proximal ALM, and establish the global and local convergence properties of the latter with the step-length  $\tau \in (0, 2)$ . As a result, the global and local convergence properties of the former even with  $\tau \in (0, 2)$  are also established.
- Even for the most conventional two-block case, we are able for the first time to rigorously characterize the connection between ADMM and proximal ALM. Note that given the form of the updating rules of the classic ADMM and ALM, although it is natural to view ADMM as an approximate version of the ALM, this is not completely true as can be seen from our analysis in this paper. Indeed, to alleviate the difficulty of solving the subproblems in the ALM, the classic ADMM uses a single cycle of the Gauss-Seidel block minimization to replace the full minimization of the augmented Lagrangian function in the ALM. This viewpoint in fact motivated the study of the classic ADMM in the very first paper [21]. However, as was mentioned in [13, 14], there were no known results in quantifying this interpretation.
- As a by-product of the second contribution, this paper gives an affirmative answer to the open question on whether the dual sequence generated by the classic ADMM with  $\tau \in (0, 2)$  is convergent if one of the two functions in the objective is linear<sup>1</sup>. This is the problem setting of the very first proof for the ADMM in Gabay and Mercier [18, Theorem 3.1] in which the dual sequence is only guaranteed to be bounded, even under very strong assumptions. The later proof of Glowinski [20, Chapter 5, Theorem 5.1] established stronger results than [18] but it requires  $\tau \in (0, (1 + \sqrt{5})/2)$ . Thereafter, only the latter interval, and especially the unit step-length, has been considered. In fact, in a rigorous proof presented recently in [5] for the classic two-block ADMM with  $\tau \in (0, (1 + \sqrt{5})/2)$ , it was shown that the convergence of the dual sequence can be guaranteed under pretty weak conditions but the convergence of the primal sequence requires more. Hence, it is of much theoretical interest to clarify whether the dual sequence is convergent if the objective contains a linear part while  $\tau \geq (1 + \sqrt{5})/2$ .
- We provide a fairly general criterion for choosing the possibly indefinite<sup>2</sup> linear operator  $\mathcal{D}$  in the proximal term, which unifies those used in Chen et al. [6] and those used in Zhang et al. [57] to guarantee the viability of the block sGS decomposition techniques and the convergence of the whole sequence generated by the algorithm in (1.7). Recall that the proximal terms in [6] should be positive semidefinite while in [57] the functions being majorized should be separable with respect to each block of variables and the corresponding

<sup>1</sup> This question was first resolved in [48] when the initial multiplier  $x^0$  satisfies  $\mathcal{G}x^0 - b = 0$  and all the subproblems are solved exactly.

<sup>2</sup> One may refer to [29] for the details that motivating the use of indefinite proximal terms in the 2-block majorized proximal ADMM, especially [29, Section 6] on their computational merits, as well as [57] for the similar results in multi-block cases.

**$\mathcal{D}$  should also be block diagonal.** Here, we do not require  $f$  to be separable and indefinite proximal terms are allowed.

- We use a unified criterion, which is weaker than those used in [6], for choosing the proximal terms in the algorithmic framework (1.7) and analyzing its convergence. Note that in [6], compared with the condition [6, (3.2)] imposed on choosing the proximal terms, a stronger condition ([6, (5.26)] of Theorem 5.1) was used to guarantee the convergence of the algorithm. Here, we are able to get rid of such a gap while using a weaker condition.
- We conduct extensive numerical experiments on solving the linear and convex quadratic semidefinite programming (SDP) problems to demonstrate how the theoretical results obtained here can be exploited to improve the numerical efficiency of the implementation on ADMM. Based on the numerical results, together with the theoretical analysis in this paper, we are able to give a plausible explanation as to why ADMM often performs well when the dual step-length is chosen to be the golden ratio of 1.618. Meanwhile, a guiding principle on choosing the step-length during the practical implementation of the algorithmic framework in (1.7) is derived.

Here we emphasize again that for solving large-scale instances of the multi-block problem (1.1), a successful multi-block ADMM-type algorithm must not only possess convergence guarantee but should also numerically perform at least as fast as the directly extended ADMM. Based on our work in this paper, we can conclude that the inexact block sGS decomposition based majorized proximal ADMM studied in [6] indeed does possess those desirable properties. Moreover, this algorithm is a versatile framework and one can apply it to problem (1.1) in different routines other than (1.7). The reason that we are more interested in the iteration scheme (1.7) is not only for the theoretical improvement one can achieve, but also for the practical merit it features for solving large scale problems, especially when the dominating computational cost is in performing the evaluations associated with the linear mappings  $\mathcal{G}$  and  $\mathcal{G}^*$ . A particular case in point is the following problem:

$$\min_{x \in \mathbb{X}} \left\{ \psi(x) + \frac{1}{2} \langle x, \mathcal{Q}x \rangle - \langle c, x \rangle \mid \mathcal{G}_E x = b_E, \mathcal{G}_I x \geq \mathbf{b}_I \right\}, \quad (1.8)$$

where  $\mathcal{Q}$ ,  $\psi$ , and  $c$  have the same meaning as in (1.3),  $\mathcal{G}_E : \mathbb{X} \rightarrow \mathbb{Z}_E$  and  $\mathcal{G}_I : \mathbb{X} \rightarrow \mathbb{Z}_I$  are the given linear mappings, and  $b = (\mathbf{b}_E; \mathbf{b}_I) \in \mathbb{Z} := \mathbb{Z}_E \times \mathbb{Z}_I$  is a given vector. By introducing a slack variable  $x' \in \mathbb{Z}_I$ , the above problem can be equivalently reformulated as

$$\min_{x \in \mathbb{X}, x' \in \mathbb{Z}_I} \left\{ \psi(x) + \frac{1}{2} \langle x, \mathcal{Q}x \rangle - \langle c, x \rangle \mid \begin{pmatrix} \mathcal{G}_E & 0 \\ \mathcal{G}_I & \mathcal{I}_{\mathbb{Z}_I} \end{pmatrix} \begin{pmatrix} x \\ x' \end{pmatrix} = b, x' \leq 0 \right\},$$

where  $\mathcal{I}_{\mathbb{Z}_I}$  is the identity operator in  $\mathbb{Z}_I$ . The corresponding dual problem in the minimization form is then given by

$$\min_{y_1, y_2, z} \left\{ p(y_1) + \frac{1}{2} \langle y_2, \mathcal{Q}y_2 \rangle - \langle b, z \rangle \mid \begin{pmatrix} y_{11} \\ y_{12} \end{pmatrix} + \begin{pmatrix} \mathcal{Q} \\ 0 \end{pmatrix} y_2 - \begin{pmatrix} \mathcal{G}_E^* & \mathcal{G}_I^* \\ 0 & \mathcal{I}_{\mathbb{Z}_2} \end{pmatrix} z = \begin{pmatrix} c \\ 0 \end{pmatrix} \right\},$$

where  $y_1 := (y_{11}; y_{12}) \in \mathbb{X} \times \mathbb{Z}_I$ ,  $p(y_1) := \psi_1^*(y_{11}) + \delta_+(y_{12})$  with  $\delta_+$  being the indicator function of the nonnegative orthant in  $\mathbb{Z}_I$ ,  $y_2 \in \mathbb{X}$  and  $z \in \mathbb{Z}$ . It is clear that when problem (1.8) has a large number of inequality constraints, the dimension of  $\mathbb{Z}$  can be much larger than that of  $\mathbb{X}$ . For such a scenario, the iteration scheme (1.7) is more preferable since the more difficult subproblem involving  $z$  is solved only once in each iteration.

### Organization

This paper is organized as follows. In Section 2, we present a few important classes of problems that can be handled by (1.1) to illustrate the wide applicability of this model. In Section 3, we design an *inexact* majorized proximal ALM framework and establish its global and local convergence properties. In Section 4, we show the key result that the sequence generated by the inexact block sGS decomposition based majorized proximal ADMM (1.7), together with a simple error tolerance criterion, is equivalent to the sequence generated by the inexact ALM framework introduced in Section 3. Accordingly, the convergence of the two-block ADMM with the step-length in the interval of  $(0, 2)$  is also established for problem (1.1) with  $s = 1$ . In Section 5, we conduct extensive numerical experiments on the 2-block dual linear SDP problems and the *multi*-block dual convex quadratic SDP problems to illustrate the numerical efficiency of the proposed algorithm, as well as the impact of the step-length on its numerical performance. A few important practical observations from the numerical results are also presented. Finally, we conclude this paper in the last section.

### Notation

- Let  $\mathbb{H}$  and  $\mathbb{H}'$  be two finite-dimensional real Hilbert spaces each endowed with an inner product  $\langle \cdot, \cdot \rangle$  and its induced norm  $\| \cdot \|$ . We also use  $\| \cdot \|$  to denote the norm induced on the product space  $\mathbb{H} \times \mathbb{H}'$  by the inner product  $\langle (v_1, v'_1), (v_2, v'_2) \rangle := \langle v_1, v_2 \rangle + \langle v'_1, v'_2 \rangle, \forall v_1, v_2 \in \mathbb{H}, \forall v'_1, v'_2 \in \mathbb{H}'$ .
- For any linear map  $\mathcal{O} : \mathbb{H} \rightarrow \mathbb{H}'$ , we use  $\mathcal{O}^*$  to denote its adjoint,  $\mathcal{O}^{-1}$  to denote its inverse (if invertible),  $\mathcal{O}^\dagger$  to denote its Moore–Penrose pseudoinverse,  $\text{Range}(\mathcal{O})$  to denote its range space, and  $\|\mathcal{O}\|$  to denote its spectral norm.
- If  $\mathbb{H}' = \mathbb{H}$  and  $\mathcal{O}$  is self-adjoint, we denote  $\|v\|_{\mathcal{O}}^2 := \langle v, \mathcal{O}v \rangle$  even if  $\mathcal{O}$  is not positive semidefinite. If, additionally,  $\mathcal{O}$  is positive semidefinite, there must be a unique self-adjoint positive semidefinite operator, denoted by  $\mathcal{O}^{1/2}$ , such that  $\mathcal{O}^{1/2}\mathcal{O}^{1/2} = \mathcal{O}$ . In this case, for any  $v, v' \in \mathbb{H}$  we define  $\langle v, v' \rangle_{\mathcal{O}} := \langle \mathcal{O}v, v' \rangle$  and  $\|v\|_{\mathcal{O}} := \sqrt{\langle v, \mathcal{O}v \rangle} = \|\mathcal{O}^{1/2}v\|$ . If  $\mathcal{O}$  is also invertible,  $\mathcal{O}^{1/2}$  is invertible and we use the notation that  $\mathcal{O}^{-1/2} := (\mathcal{O}^{1/2})^{-1}$ .
- Let  $\mathcal{O}_1, \dots, \mathcal{O}_k$  be  $k$  self-adjoint linear operators, we used  $\text{Diag}(\mathcal{O}_1, \dots, \mathcal{O}_k)$  to denote the block-diagonal linear operator whose block-diagonal elements are in the order of  $\mathcal{O}_1, \dots, \mathcal{O}_k$ .
- For any convex set  $H \subseteq \mathbb{H}$ , we denote the relative interior of  $H$  by  $\text{ri}(H)$ . When the self-adjoint linear operator  $\mathcal{O} : \mathbb{H} \rightarrow \mathbb{H}$  is positive definite, we define, for any  $v \in \mathbb{H}$ ,

$$\text{dist}_{\mathcal{O}}(v, H) := \inf_{v' \in H} \|v - v'\|_{\mathcal{O}} \quad \text{and} \quad \Pi_H^{\mathcal{O}}(v) = \arg \min_{v' \in H} \|v - v'\|_{\mathcal{O}}.$$

If  $\mathcal{O}$  is the identity operator we just omit it from the notation so that  $\text{dist}(\cdot, H)$  and  $\Pi_H(\cdot)$  are the standard distance function and the metric projection operator, respectively.

- Let  $\theta : \mathbb{H} \rightarrow (-\infty, +\infty]$  be an arbitrary closed proper convex function. We use  $\text{dom } \theta$  to denote its effective domain,  $\partial\theta$  to denote its subdifferential mapping, and  $\theta^*$  to denote its conjugate function. Moreover, for a given self-adjoint and positive definite linear operator  $\mathcal{O} : \mathbb{H} \rightarrow \mathbb{H}$ , we use  $\text{Prox}_{\theta}^{\mathcal{O}}$  to denote the Moreau–Yosida proximal mapping of  $\theta$ , which is defined by

$$\text{Prox}_{\theta}^{\mathcal{O}}(v) := \arg \min_{v' \in \mathbb{H}} \left\{ \theta(v') + \frac{1}{2} \|v - v'\|_{\mathcal{O}}^2 \right\}, \quad \forall v \in \mathbb{H}.$$

Note that the mapping  $\text{Prox}_{\theta}^{\mathcal{O}}$  is globally Lipschitz continuous. If  $\mathcal{O}$  is the identity operator, we will drop  $\mathcal{O}$  from  $\text{Prox}_{\theta}^{\mathcal{O}}(\cdot)$ .

## 2 Illustrative Examples

In this section, we present a few important classes of concrete problems, including those in the classic core convex programming as well as those which are popularly used in various real-world applications. As will be shown, these problems and/or their dual problems have the form given by (1.1), so that the algorithm designed in this paper can be utilized to solve them.

### 2.1 Convex Composite Quadratic Programming

It is well known that many problems are subsumed under the convex composite quadratic programming model (1.2) or the more concrete form (1.8). For example, it includes the important classes of convex quadratic programming (QP), the convex quadratic semidefinite programming (QSDP), and the convex quadratic programming and weighted centering [41] (QPWC). As an illustration, consider a convex QSDP problem in the following form

$$\min_{X \in \mathbb{S}^n} \left\{ \frac{1}{2} \langle X, \mathcal{Q}X \rangle - \langle C, X \rangle \mid \mathcal{A}_E X = \mathbf{b}_E, \mathcal{A}_I X \geq \mathbf{b}_I, X \in \mathbb{S}_+^n \right\}, \quad (2.1)$$

where  $\mathbb{S}^n$  is the space of  $n \times n$  real symmetric matrices and  $\mathbb{S}_+^n$  is the closed convex cone of positive semidefinite matrices in  $\mathbb{S}^n$ ,  $\mathcal{Q} : \mathbb{S}^n \rightarrow \mathbb{S}^n$  is a positive semidefinite linear operator,  $C \in \mathbb{S}^n$  is a given matrix, and  $\mathcal{A}_E$  and  $\mathcal{A}_I$  are the linear maps from  $\mathbb{S}^n$  to the two finite-dimensional Euclidean spaces  $\mathbb{R}^{m_E}$  and  $\mathbb{R}^{m_I}$  that containing  $\mathbf{b}_E$  and  $\mathbf{b}_I$ , respectively. To solve this problem, one may consult the recently developed software QSDPNAL in Li et al. [31] and the references therein. The algorithm implemented in QSDPNAL is a two-phase augmented Lagrangian method in which the first phase is an inexact sGS decomposition based multi-block proximal ADMM whose convergence was established in [6, Theorem 5.1]. The solution generated in the first phase is used as the initial point to warm-start the second phase algorithm, which is an ALM with the inner subproblem in each iteration



being solved via an inexact semismooth Newton algorithm. In Section 5, we will use the QSDP problem (2.1) to test the algorithm studied in this paper.

Besides the core optimization problems just mentioned above, there are many problems from real-world applications that can be cast in the form of (1.2) and the following are only a few such examples.

### *Penalized and Constrained Regression Models*

In various statistical applications, the penalized and constrained (PAC) regression [25] often arises in high-dimensional generalized linear models with linear equality and inequality constraints. A concrete example of the PAC regression is the following constrained lasso problem

$$\min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\Phi x - \eta\|^2 + \lambda \|x\|_1 \mid A_E x = b_E, A_I x \geq b_I \right\}, \quad (2.2)$$

where  $\Phi \in \mathbb{R}^{m \times n}$ ,  $A_E \in \mathbb{R}^{m_E \times n}$ ,  $A_I \in \mathbb{R}^{m_I \times n}$ ,  $\eta \in \mathbb{R}^m$ ,  $b_E \in \mathbb{R}^{m_E}$  and  $b_I \in \mathbb{R}^{m_I}$  are the given data, and  $\lambda > 0$  is a given regularization parameter. The statistical properties of problem (2.2) have been studied in [25]. For more details on the applications of the model (2.2), one may refer to [25, 19] and the references therein. In Gaines et al. [19], the authors considered solving (2.2) by first reformulating it as a conventional QP via letting  $x = x_+ - x_-$  and adding the extra constraints  $x_+ \geq 0$ ,  $x_- \geq 0$ , and then applying the primal ADMM to solve the conventional QP, in which all the subproblems should be solved exactly (or to very high accuracy) by iterative methods. Such a combination may perform well for low dimensional problems with moderate sample sizes. But for the more challenging and interesting high-dimensional cases where  $n$  is extremely large and  $m \ll n$ , the approach in [19] is likely to face severe numerical difficulties because of the presence of a huge number of constraints. Fortunately, the algorithm we designed in this paper can precisely handle those difficult cases because the large linear systems associated with the huge number of constraints are not required to solve to very high accuracy by an iterative solver.

### *Noisy Matrix Completion and Rank-Correction Step*

In Miao et al. [36], the authors introduced a rank-correction step for matrix completion with fixed basis coefficients to overcome the shortcomings of the nuclear norm penalization model for such problems. Let  $\bar{X} \in \mathbb{V}^{n_1 \times n_2}$  (where  $\mathbb{V}^{n_1 \times n_2}$  may represent the space of  $n_1 \times n_2$  real or complex matrices or the space of  $n \times n$  real symmetric or Hermitian matrices) be the unknown true low-rank matrix and  $\bar{X}_m$  is an initial estimator of  $\bar{X}$  from the nuclear norm penalized least squares model. The rank-correction step is to solve the following convex optimization problem

$$\begin{aligned} \min_X \quad & \frac{1}{2m} \|y - \mathcal{P}_o(X)\|^2 + \rho_m (\|X\|_* - \langle F(\bar{X}_m), X \rangle) \\ \text{s.t.} \quad & \mathcal{P}_A(X) = \mathcal{P}_A(\bar{X}), \quad \|\mathcal{P}_B(X)\|_\infty \leq b, \end{aligned} \quad (2.3)$$

where  $y = \mathcal{P}_o(\bar{X}) + \epsilon \in \mathbb{R}^m$  is the observed data for the matrix  $\bar{X}$ ,  $\mathcal{P}_o$  is the linear map corresponding to the observed entries,  $\epsilon \in \mathbb{R}^m$  is the unknown error,  $\rho_m > 0$  is a given penalty parameter, and  $F : \mathbb{V}^{n_1 \times n_2} \rightarrow \mathbb{V}^{n_1 \times n_2}$  is a spectral operator [10] whose precise definition can be found in [36, Section 5]. Here the constraints  $\mathcal{P}_A(X) = \mathcal{P}_A(\bar{X})$  and  $\|\mathcal{P}_B(X)\|_\infty \leq b$  represent the fixed elements and bounded elements of  $X$ , respectively. If  $F$  and the equality constraints are vacuous, problem (2.3) is exactly the noise matrix completion model considered in [37], and a similar matrix completion model can be found in [26]. One may view (2.3) as an instance of problem (1.2), and whose corresponding linear operator  $\mathcal{Q}$  admits a very simple form.

## 2.2 Two-Block Problems

Next we present a few important classes of two-block problems whose objective functions contain a linear part.

### *Semidefinite Programming*

One of the most prominent examples of problem (1.1) with 2 blocks of variables (i.e.,  $s = 1$ ) is the dual linear semidefinite programming (SDP) problem given by

$$\min_{Y, z} \{ \delta_{\mathbb{S}_+^n}(Y) - \langle b, z \rangle \mid Y + \mathcal{A}^* z = C \}, \quad (2.4)$$

where  $\mathcal{A} : \mathbb{S}^n \rightarrow \mathbb{R}^m$  is a given linear map, and  $\mathbf{b} \in \mathbb{R}^m$  and  $C \in \mathbb{S}^n$  are given data. The notation  $\delta_{\mathbb{S}_+^n}$  denotes the indicator function of  $\mathbb{S}_+^n$ . For problem (2.4), various ADMM algorithms have been employed to solve the problem. As far as we are aware of, the classic two-block ADMM with unit step-length was first employed in Povh et al. [43] under the name of boundary point method for solving the SDP problem (2.4). It was later extended in Malick et al. [34] with a convergence proof. The ADMM approach was later used in the software SDPNAL developed by Zhao et al. [58] to warm-start a semismooth Newton method based ALM for solving problem (2.4).

In section 5, we will conduct extensive numerical experiments on solving a few classes of linear SDP problems via the two-block ADMM algorithm but with the dual step-length being chosen in the interval  $(0, 2)$ , as it is guaranteed by this paper.

### Equality Constrained Problems

Consider the equality constrained problem

$$\min_{x \in \mathbb{X}} \{\theta(x) \mid \mathcal{G}x = b\}, \quad (2.5)$$

where  $\mathcal{G} : \mathbb{X} \rightarrow \mathbb{R}^m$  is a linear map,  $b \in \mathbb{R}^m$  is a given vector, and  $\theta : \mathbb{X} \rightarrow (-\infty, +\infty]$  is a simple closed proper convex function such that its proximal mapping can be computed efficiently. The dual problem of (2.5) can be written in the minimization form as

$$\min_{y, z} \{\theta^*(y) - \langle b, z \rangle \mid y - \mathcal{G}^*z = 0\}. \quad (2.6)$$

A concrete example of problem (2.5), with  $\mathbb{X} := \mathbb{R}^n$  and  $\theta(x) := \|x\|_1$ , is the basis pursuit (BP) problem [7], which has been widely used in sparse signal recovery and image restoration. Another example of (2.5) is the nuclear norm based matrix completion problem for which  $\mathbb{X} := \mathbb{R}^{n_1 \times n_2}$  and  $\theta(x) = \|x\|_*$ . Moreover, the so called tensor completion problem [33] also falls into this category.

We note that for the application problems just mentioned above, the dimension of  $\mathbb{X}$  is generally much larger than  $m$ , i.e., the dimension of the linear constraints. Therefore from the computational viewpoint, it is generally more economical to apply the two-block ADMM to the dual problem (2.6) instead of the primal problem (2.5) (by introducing an extra variable  $x'$  and adding the condition  $x - x' = 0$ ) because the former will solve smaller  $m \times m$  linear systems in each iteration whereas the latter will correspondingly need to solve much larger linear systems.

### Composite Problems

A composite problem can take the following form

$$\min_{z \in \mathbb{Z}} f(c - \mathcal{G}^*z), \quad (2.7)$$

where  $f : \mathbb{Z} \rightarrow (-\infty, +\infty]$  is a (possibly nonsmooth) closed proper convex function whose proximal mapping can be computed efficiently,  $\mathcal{G} : \mathbb{Z} \rightarrow \mathbb{X}$  is a given linear operator and  $c \in \mathbb{X}$  is given data. By introducing a slack variable, problem (2.7) can be recast as

$$\min_{y, z} \{f(y) \mid y + \mathcal{G}^*z = c\}.$$

Problem (2.7) contains many real-world applications such as the well-known least absolute deviation (LAD) problem (also known as least absolute error (LAE), least absolute value (LAV), least absolute residual (LAR), sum of absolute deviations, or the  $\ell_1$ -norm condition). The model (2.7) also includes the Huber fitting problem [24]. We shall not continue with more examples as there are too many applications to be listed here to serve as a literature review.



### Consensus Optimization

Consider the following problem

$$\min_{z \in \mathbb{Z}} \left\{ \sum_{i=1}^n f_i(\mathcal{G}_i^* z) \right\}, \quad (2.8)$$

where each  $f_i$  is a closed proper convex function and each  $\mathcal{G}_i : \mathbb{Y}_i \rightarrow \mathbb{Z}$  is a linear operator. The model (2.8) includes the global variable consensus optimization and general variable optimization, as well as their regularized versions (see [4, Section 7]), which have been well applied in many areas such as machine learning, signal processing and wireless communication [4, 3, 49, 46, 59]. In the consensus optimization setting, it is usually preferable to solve subproblems each involving a subset of the component functions  $f_1, \dots, f_n$  instead of all of them. Therefore, one can equivalently recast problem (2.8) as

$$\min_{y, z} \left\{ \sum_{i=1}^n f_i(y_i) \mid y_i - \mathcal{G}_i^* z = 0, \ 1 \leq i \leq n \right\}. \quad (2.9)$$

Obviously, when applying the two-block ADMM to solve (2.9), the subproblem with respect to  $y$  is separated into  $n$  independent problems that can be solved in parallel. In [4], the variable  $z$  in (2.9) is called the central collector. Besides, the network based decentralized and distributed computation of the consensus optimization, such as the distributed lasso in [35], also falls in the problems setting in this paper.

### 3 An Inexact Majorized ALM with Indefinite Proximal Terms

In this section, we present an inexact majorized indefinite-proximal ALM. This algorithm, as well as its global and local convergence properties, not only constitutes a generalization of the original (proximal) ALM, but also paves the way for us to establish its equivalence relationship with the inexact block sGS decomposition based indefinite-proximal multi-block ADMM in the next section.

Let  $\mathbb{X}$  and  $\mathbb{W}$  be two finite-dimensional real Hilbert spaces each endowed with an inner product  $\langle \cdot, \cdot \rangle$  and its induced norm  $\| \cdot \|$ . We consider the following fairly general linearly constrained convex optimization problem

$$\min_{w \in \mathbb{W}} \{ \varphi(w) + h(w) \mid \mathcal{A}^* w = c \}, \quad (3.1)$$

where  $\varphi : \mathbb{W} \rightarrow (-\infty, +\infty]$  is a closed proper convex function,  $h : \mathbb{W} \rightarrow (-\infty, +\infty)$  is a continuously differentiable convex function whose gradient is Lipschitz continuous,  $\mathcal{A} : \mathbb{X} \rightarrow \mathbb{W}$  is a linear mapping and  $c \in \mathbb{X}$  is the given data. The Karush-Kuhn-Tucker (KKT) system of problem (3.1) is given by

$$0 \in \partial\varphi(w) + \nabla h(w) + \mathcal{A}x, \quad \mathcal{A}^* w - c = 0. \quad (3.2)$$

For any  $(w, x) \in \mathbb{W} \times \mathbb{X}$  that solve the KKT system (3.2),  $w$  is a solution to problem (3.1) while  $x$  is a dual solution of (3.1).

The fact that the gradient of  $h$  is Lipschitz continuous implies that there exists a self-adjoint positive semidefinite linear operator  $\widehat{\Sigma}^h : \mathbb{W} \rightarrow \mathbb{W}$ , such that for any  $w' \in \mathbb{W}$ ,  $h(w) \leq \widehat{h}(w, w')$ , where

$$\widehat{h}(w, w') := h(w') + \langle \nabla h(w'), w - w' \rangle + \frac{1}{2} \|w - w'\|_{\widehat{\Sigma}^h}^2, \quad \forall w \in \mathbb{W}. \quad (3.3)$$

We call the function  $\widehat{h}(\cdot, w') : \mathbb{W} \rightarrow (-\infty, +\infty)$  a majorization of  $h$  at  $w'$ . The following result, whose proof can be found in [57, Lemma 3.2], will be used later.

**Lemma 3.1** Suppose that (3.3) holds for any given  $w' \in \mathbb{W}$ . Then, it holds that

$$\langle \nabla h(w) - \nabla h(w'), w'' - w' \rangle \geq -\frac{1}{4} \|w - w''\|_{\widehat{\Sigma}^h}^2, \quad \forall w, w', w'' \in \mathbb{W}.$$

Let  $\sigma > 0$  be a given penalty parameter. The majorized augmented Lagrangian function associated with problem (3.1) is defined by

$$\begin{aligned} \widehat{\mathcal{L}}_\sigma(w; (x, w')) &:= \varphi(w) + \widehat{h}(w, w') + \langle \mathcal{A}^* w - c, x \rangle + \frac{\sigma}{2} \|\mathcal{A}^* w - c\|^2, \\ &\quad \forall (w, x, w') \in \mathbb{W} \times \mathbb{X} \times \mathbb{W}. \end{aligned} \quad (3.4)$$

In the following, we propose an inexact majorized indefinite-proximal ALM in Algorithm **iPALM** for solving problem (3.1). This algorithm is an extension of the proximal method of multipliers developed by Rockafellar [45], with new ingredients added based on the recent progress on using proximal terms which are not necessarily positive definite [16, 29, 57] and the implementable inexact minimization criteria studied in [6]. For the convenience of later convergence analysis, we make the following blanket assumption.

**Assumption 3.1** *The solution set to the KKT system (3.2) is nonempty and  $\mathcal{S} : \mathbb{W} \rightarrow \mathbb{W}$  is a given self-adjoint (not necessarily positive semidefinite) linear operator such that*

$$\mathcal{S} \succeq -\frac{1}{2}\widehat{\Sigma}^h \quad \text{and} \quad \frac{1}{2}\widehat{\Sigma}^h + \sigma\mathcal{A}\mathcal{A}^* + \mathcal{S} \succ 0. \quad (3.5)$$

We are now ready to present Algorithm **iPALM** that will be studied in this section.

---

**Algorithm iPALM** An inexact majorized indefinite-proximal ALM

---

Let  $\{\varepsilon_k\}$  be a summable sequence of nonnegative numbers. Choose an initial point  $(x^0, w^0) \in \mathbb{X} \times \mathbb{W}$ . For  $k = 0, 1, \dots$ , perform the following steps in each iteration.

**Step 1.** Compute

$$w^{k+1} \approx \bar{w}^{k+1} := \arg \min_{w \in \mathbb{W}} \left\{ \widehat{\mathcal{L}}_\sigma(w; (x^k, w^k)) + \frac{1}{2} \|w - w^k\|_{\mathcal{S}}^2 \right\} \quad (3.6)$$

such that there exists a vector  $d^k \in \mathbb{W}$  satisfying  $\|d^k\| \leq \varepsilon_k$  and

$$d^k \in \partial_w \widehat{\mathcal{L}}_\sigma(w^{k+1}; (x^k, w^k)) + \mathcal{S}(w^{k+1} - w^k). \quad (3.7)$$

**Step 2.** Compute  $x^{k+1} := x^k + \tau\sigma(\mathcal{A}^*w^{k+1} - c)$  with  $\tau \in (0, 2)$  being the step-length.

---

We shall next proceed to analyze the global convergence, the rate of local convergence and the iteration complexity of Algorithm **iPALM**. For notational convenience, we collect the total quadratic information in the objective function of (3.6) as the following linear operator

$$\mathcal{M} := \widehat{\Sigma}^h + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^*. \quad (3.8)$$

The following result presents two important inequalities for the subsequent analysis. The first one characterizes the distance (with  $\mathcal{M}$  being involved in the metric) from the computed solution to the true solution of the subproblem in (3.6), while the second one presents a non-monotone descent property about the sequence generated by Algorithm **iPALM**.

**Proposition 3.1** *Suppose that Assumption 3.1 holds. Then,*

(a) *the sequence  $\{(x^k, w^k)\}$  generated by Algorithm **iPALM** and the auxiliary sequence  $\{\bar{w}^k\}$  defined in (3.6) are well-defined, and it holds that*

$$\|w^{k+1} - \bar{w}^{k+1}\|_{\mathcal{M}}^2 \leq \langle d^k, w^{k+1} - \bar{w}^{k+1} \rangle; \quad (3.9)$$

(b) *for any given  $(x^*, w^*) \in \mathbb{X} \times \mathbb{W}$  that solves the KKT system (3.2) and  $k \geq 1$ , it holds that*

$$\begin{aligned} & \left( \frac{1}{2\tau\sigma} \|x_e^{k+1}\|^2 + \frac{1}{2} \|w_e^{k+1}\|_{\widehat{\Sigma}^h + \mathcal{S}}^2 \right) - \left( \frac{1}{2\tau\sigma} \|x_e^k\|^2 + \frac{1}{2} \|w_e^k\|_{\widehat{\Sigma}^h + \mathcal{S}}^2 \right) \\ & \leq - \left( \frac{(2-\tau)\sigma}{2} \|\mathcal{A}^*w_e^{k+1}\|^2 + \frac{1}{2} \|w^{k+1} - w^k\|_{\frac{1}{2}\widehat{\Sigma}^h + \mathcal{S}}^2 - \langle d^k, w_e^{k+1} \rangle \right), \end{aligned} \quad (3.10)$$

where  $x_e := x - x^*$ ,  $\forall x \in \mathbb{X}$  and  $w_e := w - w^*$ ,  $\forall w \in \mathbb{W}$ .

*Proof (a)* From (3.5) and (3.8) we know that  $\mathcal{M} \succ 0$ . Hence, each of the subproblems in Algorithm **iPALM** is strongly convex so that each  $\bar{w}^{k+1}$  is uniquely determined by  $(x^k, w^k)$ . Note that, for the given  $\varepsilon_k \geq 0$ , one can always find a certain  $w^{k+1}$  such that  $\|d^k\| \leq \varepsilon_k$  with  $d^k$  being given in (3.7), see [12, Lemma 4.5]. Hence, Algorithm **iPALM** is well-defined. According to (3.3) and (3.4), the objective function in (3.6) is given by

$$\varphi(w) + \langle \nabla h(w^k) + \mathcal{A}x^k, w \rangle + \frac{\sigma}{2} \|\mathcal{A}^*w - c\|^2 + \frac{1}{2} \|w - w^k\|_{\widehat{\Sigma}^h + \mathcal{S}}^2,$$

so that (3.7) implies that

$$d^k \in \partial\varphi(w^{k+1}) + \nabla h(w^k) + \mathcal{A}x^k + \sigma\mathcal{A}(\mathcal{A}^*w^{k+1} - c) + (\widehat{\Sigma}^h + \mathcal{S})(w^{k+1} - w^k). \quad (3.11)$$

Therefore, from the definitions of the Moreau-Yosida proximal mapping and  $\mathcal{M}$  in (3.8), one has that

$$w^{k+1} = \text{Prox}_{\varphi}^{\mathcal{M}} \left( \mathcal{M}^{-1} [d^k - (\nabla h(w^k) + \mathcal{A}x^k - \sigma\mathcal{A}c - (\widehat{\Sigma}^h + \mathcal{S})w^k)] \right).$$

Consequently, by the Lipschitz continuity of  $\text{Prox}_{\varphi}^{\mathcal{M}}$  [28, Proposition 2.3] and the fact that  $d^k$  can be set as zero if  $w^{k+1} = \overline{w}^{k+1}$ , one can readily get (3.9).

(b) Let  $(x^*, w^*) \in \mathbb{X} \times \mathbb{W}$  be an arbitrary solution to the KKT system (3.2). Obviously, one has that  $-\nabla h(w^*) - \mathcal{A}x^* \in \partial\varphi(w^*)$  and  $\mathcal{A}^*w^* = c$ . This, together with (3.11) and the maximal monotonicity of  $\partial\varphi$ , implies that

$$\langle d^k - \nabla h(w^k) + \nabla h(w^*) - \mathcal{A}x_e^k - \sigma\mathcal{A}(\mathcal{A}^*w^{k+1} - c) - (\widehat{\Sigma}^h + \mathcal{S})(w^{k+1} - w^k), w_e^{k+1} \rangle \geq 0.$$

Therefore, by using the fact that

$$\mathcal{A}^*w^{k+1} - c = \mathcal{A}^*w_e^{k+1} = \frac{1}{\tau\sigma}(x^{k+1} - x^k), \quad (3.12)$$

one can obtain from the above inequality and Lemma 3.1 that

$$\begin{aligned} & \langle d^k, w_e^{k+1} \rangle - \frac{1}{\tau\sigma} \langle x_e^k, x^{k+1} - x^k \rangle - \sigma \|\mathcal{A}^*w_e^{k+1}\|^2 - \langle (\widehat{\Sigma}^h + \mathcal{S})(w^{k+1} - w^k), w_e^{k+1} \rangle \\ & \geq \langle \nabla h(w^k) - \nabla h(w^*), w_e^{k+1} \rangle \geq -\frac{1}{4} \|w^{k+1} - w^k\|_{\widehat{\Sigma}^h}^2 = -\frac{1}{2} \|w^{k+1} - w^k\|_{\frac{1}{2}\widehat{\Sigma}^h}^2. \end{aligned} \quad (3.13)$$

Note that  $\langle x_e^k, x^{k+1} - x^k \rangle = \frac{1}{2} \|x_e^k\|^2 - \frac{1}{2} \|x_e^k\|^2 - \frac{1}{2} \|x^{k+1} - x^k\|^2$  and

$$\langle (\widehat{\Sigma}^h + \mathcal{S})(w^{k+1} - w^k), w_e^{k+1} \rangle = \frac{1}{2} \|w^{k+1} - w^k\|_{\widehat{\Sigma}^h + \mathcal{S}}^2 + \frac{1}{2} \|w_e^{k+1}\|_{\widehat{\Sigma}^h + \mathcal{S}}^2 - \frac{1}{2} \|w_e^k\|_{\widehat{\Sigma}^h + \mathcal{S}}^2.$$

Then, (3.10) follows from (3.13) and this completes the proof of the proposition.  $\square$

### 3.1 Global Convergence

For the convenience of our analysis, we define the following two linear operators

$$\Xi := \tau\sigma \left( \frac{1}{2}\widehat{\Sigma}^h + \mathcal{S} + \frac{(2-\tau)\sigma}{6}\mathcal{A}\mathcal{A}^* \right) \quad \text{and} \quad \Theta := \tau\sigma \left( \widehat{\Sigma}^h + \mathcal{S} + \frac{(2-\tau)\sigma}{3}\mathcal{A}\mathcal{A}^* \right), \quad (3.14)$$

which will be used in defining metrics in  $\mathbb{W}$ . Note that  $\tau \in (0, 2)$ . If (3.5) in Assumption 3.1 holds, one has that

$$\begin{cases} \frac{1}{\tau\sigma}\Xi = \frac{2-\tau}{6} \left( \frac{1}{2}\widehat{\Sigma}^h + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^* \right) + \left( 1 - \frac{2-\tau}{6} \right) \left( \frac{1}{2}\widehat{\Sigma}^h + \mathcal{S} \right) \\ \quad \succeq \frac{2-\tau}{6} \left( \frac{1}{2}\widehat{\Sigma}^h + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^* \right) \succ 0, \\ \frac{1}{\tau\sigma}\Theta = \frac{1}{\tau\sigma}\Xi + \frac{1}{2}\widehat{\Sigma}^h + \frac{(2-\tau)\sigma}{6}\mathcal{A}\mathcal{A}^* \succeq \frac{1}{\tau\sigma}\Xi \succ 0. \end{cases} \quad (3.15)$$

Moreover, we define the block-diagonal linear operator  $\Omega : \mathbb{U} \rightarrow \mathbb{U}$  by

$$\Omega(x; w) := (x; \Theta^{\frac{1}{2}}w), \quad \forall (x, w) \in \mathbb{X} \times \mathbb{W}, \quad (3.16)$$

where  $\Theta$  is given by (3.14). Now we establish the convergence theorem of Algorithm iPALM. The corresponding proof mainly follows from the proof of [6, Theorem 5.1] for the convergence of an inexact majorized semi-proximal ADMM and the following result on quasi-Fejér monotone sequence will be used.

**Lemma 3.2** *Let  $\{a_k\}_{k \geq 0}$  be a sequence of nonnegative real numbers satisfying  $a_{k+1} \leq a_k + \varepsilon_k$  for all  $k \geq 0$ , where  $\{\varepsilon_k\}_{k \geq 0}$  is a nonnegative and summable sequence of real numbers. Then the  $\{a_k\}$  converges to a unique limit point.*

**Theorem 3.1** *Suppose that Assumption 3.1 holds and the sequence  $\{(x^k, w^k)\}$  is generated by Algorithm iPALM. Then,*

(a) for any solution  $(x^*, w^*) \in \mathbb{X} \times \mathbb{W}$  of the KKT system (3.2) and  $k \geq 1$ , we have that

$$\begin{aligned} & \| (x_e^{k+1}; w_e^{k+1}) \|_{\Omega}^2 - \| (x_e^k; w_e^k) \|_{\Omega}^2 \\ & \leq - \left( \frac{(2-\tau)}{3\tau} \|x^{k+1} - x^k\|^2 + \|w^{k+1} - w^k\|_{\Xi}^2 - 2\tau\sigma \langle d^k, w_e^{k+1} \rangle \right), \end{aligned} \quad (3.17)$$

where  $x_e := x - x^*$ ,  $\forall x \in \mathbb{X}$  and  $w_e := w - w^*$ ,  $\forall w \in \mathbb{W}$ ;

(b) the sequence  $\{(x^k, w^k)\}$  is bounded;

(c) any accumulation point of the sequence  $\{(x^k, w^k)\}$  solves the KKT system (3.2);

(d) the whole sequence  $\{(x^k, w^k)\}$  converges to a solution to the KKT system (3.2).

*Proof (a)* By using (3.14), together with the definitions of  $\Xi$  and  $\Theta$  in (3.10), and the fact that  $\mathcal{A}^* w_e^{k+1} = \frac{1}{\tau\sigma}(x^{k+1} - x^k)$ , one can get

$$\begin{aligned} & \| (x_e^{k+1}; w_e^{k+1}) \|_{\Omega}^2 - \| (x_e^k; w_e^k) \|_{\Omega}^2 = (\|x_e^{k+1}\|^2 + \|w_e^{k+1}\|_{\Theta}^2) - (\|x_e^k\|^2 + \|w_e^k\|_{\Theta}^2) \\ & \leq -\tau\sigma \left( \frac{2(2-\tau)\sigma}{3} \|\mathcal{A}^* w_e^{k+1}\|^2 + \frac{(2-\tau)\sigma}{3} \|\mathcal{A}^* w_e^k\|^2 + \|w^{k+1} - w^k\|_{\frac{1}{2}\widehat{\Sigma}^h + \mathcal{S}}^2 - 2\langle d^k, w_e^{k+1} \rangle \right) \\ & \leq - \left( \frac{(2-\tau)}{3\tau} \|x^{k+1} - x^k\|^2 + \tau\sigma \frac{(2-\tau)}{3} (\sigma \|\mathcal{A}^* w_e^{k+1}\|^2 + \sigma \|\mathcal{A}^* w_e^k\|^2) \right. \\ & \quad \left. + \tau\sigma \|w^{k+1} - w^k\|_{\frac{1}{2}\widehat{\Sigma}^h + \mathcal{S}}^2 - 2\tau\sigma \langle d^k, w_e^{k+1} \rangle \right) \\ & \leq - \left( \frac{(2-\tau)}{3\tau} \|x^{k+1} - x^k\|^2 + \sigma\tau \|w^k - w^{k+1}\|_{\frac{(2-\tau)\sigma}{6}\mathcal{A}\mathcal{A}^* + \frac{1}{2}\widehat{\Sigma}^h + \mathcal{S}}^2 - 2\tau\sigma \langle d^k, w_e^{k+1} \rangle \right), \end{aligned}$$

which, together with the definition of the linear operator  $\Xi$  in (3.14), implies (3.17).

(b) Define  $\bar{x}^{k+1} := x^k + \tau\sigma(\mathcal{A}^* \bar{w}^{k+1} - c)$ ,  $\forall k \geq 0$ . From (3.6), (3.7) and (3.17) one can get that for any  $k \geq 0$ ,

$$\| (\bar{x}_e^{k+1}; \bar{w}_e^{k+1}) \|_{\Omega}^2 \leq \| (x_e^k; w_e^k) \|_{\Omega}^2 - \left( \frac{(2-\tau)}{3\tau} \|\bar{x}^{k+1} - x^k\|^2 + \|\bar{w}^{k+1} - w^k\|_{\Xi}^2 \right).$$

Meanwhile, one can get that  $\|(\bar{x}_e^{k+1}; \bar{w}_e^{k+1})\|_{\Omega} \leq \| (x_e^k; w_e^k) \|_{\Omega}$ ,  $\forall k \geq 0$ . Therefore, it holds that

$$\begin{aligned} & \| (x_e^{k+1}; w_e^{k+1}) \|_{\Omega} \leq \| (\bar{x}_e^{k+1}; \bar{w}_e^{k+1}) \|_{\Omega} + \| (x^{k+1} - \bar{x}^{k+1}; w^{k+1} - \bar{w}^{k+1}) \|_{\Omega} \\ & \leq \| (x_e^k; w_e^k) \|_{\Omega} + \| (\tau\sigma\mathcal{A}^*(w^{k+1} - \bar{w}^{k+1}); \Theta^{1/2}(w^{k+1} - \bar{w}^{k+1})) \| \\ & = \| (x_e^k; w_e^k) \|_{\Omega} + \| w^{k+1} - \bar{w}^{k+1} \|_{\tau^2\sigma^2\mathcal{A}\mathcal{A}^* + \Theta}, \quad \forall k \geq 0. \end{aligned} \quad (3.18)$$

From (3.9) we know that  $\|w^{k+1} - \bar{w}^{k+1}\|_{\mathcal{M}}^2 \leq \langle \mathcal{M}^{-1/2}d^k, \mathcal{M}^{1/2}(w^{k+1} - \bar{w}^{k+1}) \rangle$ , so that

$$\| \mathcal{M}^{1/2}(w^{k+1} - \bar{w}^{k+1}) \| \leq \| \mathcal{M}^{-1/2}d^k \| \leq \| \mathcal{M}^{-1/2} \| \|d^k\|.$$

Therefore, it holds that

$$\begin{aligned} & \| w^{k+1} - \bar{w}^{k+1} \| \leq \| \mathcal{M}^{-1/2} \| \| \mathcal{M}^{1/2}(w^{k+1} - \bar{w}^{k+1}) \| \\ & \leq \| \mathcal{M}^{-1/2} \| \| \mathcal{M}^{-1/2} \| \|d^k\| \leq \| \mathcal{M}^{-1} \| \varepsilon_k, \quad \forall k \geq 0. \end{aligned} \quad (3.19)$$

Therefore, by combining (3.18) and (3.19) together we can get

$$\| (x_e^{k+1}; w_e^{k+1}) \|_{\Omega} \leq \| (x_e^k; w_e^k) \|_{\Omega} + \sqrt{\tau^2\sigma^2\mathcal{A}\mathcal{A}^* + \Theta} \| \mathcal{M}^{-1} \| \varepsilon_k, \quad \forall k \geq 0.$$

Hence, the sequence  $\{\| (x_e^{k+1}; w_e^{k+1}) \|_{\Omega}\}$  is quasi-Fejér monotone, which converges to a unique limit point by Lemma 3.2. Since  $\Omega$  defined in (3.16) is positive definite, we further know that the sequence  $\{(x^k, w^k)\}$  is bounded.

(c) From (3.17) we know that for any  $k \geq 0$ ,

$$\begin{aligned} & \sum_{i=0}^k \left\{ \| (x_e^i; w_e^i) \|_{\Omega}^2 - \| (x_e^{i+1}; w_e^{i+1}) \|_{\Omega}^2 + 2\tau\sigma\varepsilon_i \| w_e^{i+1} \| \right\} \\ & \geq \sum_{i=0}^k \left\{ \frac{(2-\tau)}{3\tau} \|x^{i+1} - x^i\|^2 + \|w^{i+1} - w^i\|_{\Xi}^2 \right\}. \end{aligned} \quad (3.20)$$

Since  $\{(x^k, w^k)\}$  is bounded and  $\{\varepsilon_k\}$  is summable, it holds that

$$\sum_{i=0}^{\infty} \left\{ \|(x_e^i; w_e^i)\|_{\Omega}^2 - \|(x_e^{i+1}; w_e^{i+1})\|_{\Omega}^2 + 2\tau\sigma\varepsilon_i \|w_e^{i+1}\| \right\} < \infty,$$

which, together with (3.20) and the fact that  $\Xi \succ 0$ , implies

$$\lim_{k \rightarrow \infty} (x^{k+1} - x^k) = 0 \quad \text{and} \quad \lim_{k \rightarrow \infty} (w^{k+1} - w^k) = 0. \quad (3.21)$$

Suppose that the subsequence  $\{(x^{k_j}, w^{k_j})\}$  of  $\{(x^k, w^k)\}$  converges to some limit point  $(x^\infty, w^\infty)$ . By taking limits on both sides of (3.11) and (3.12) along with  $k^j$  and using (3.21) and [44, Theorem 24.6], one can get

$$0 \in \partial\varphi(w^\infty) + \nabla h(w^\infty) + \mathcal{A}x^\infty \quad \text{and} \quad \mathcal{A}^*w^\infty - c = 0,$$

which implies that  $(x^\infty, w^\infty)$  is a solution to the KKT system (3.2).

(d) Note that (3.17) holds for any  $(x^*, w^*)$  satisfying the KKT system (3.2). Therefore, we can choose  $x^* = x^\infty$  and  $w^* = w^\infty$  in (3.17):

$$\|(x^{k+1} - x^\infty; w^{k+1} - w^\infty)\|_{\Omega}^2 \leq \|(x^k - x^\infty; w^k - w^\infty)\|_{\Omega}^2 + 2\tau\sigma \|w^{k+1} - w^\infty\| \varepsilon_k.$$

Note that  $\{w^k\}$  is bounded. Then, the above inequality, together with Lemma 3.2, implies that the quasi-Fejér monotone sequence  $\{\|(x^k - x^\infty; w^k - w^\infty)\|_{\Omega}^2\}$  converges. Since  $(x^\infty, w^\infty)$  is a limit point of  $\{(x^k, w^k)\}$ , one has that

$$\lim_{k \rightarrow \infty} \|(x^k - x^\infty; w^k - w^\infty)\|_{\Omega}^2 = 0,$$

which, together with the fact that  $\Omega \succ 0$ , implies that the whole sequence  $\{(x^k, w^k)\}$  converges to  $(x^\infty, w^\infty)$ . This completes the proof of the theorem.  $\square$

### 3.2 Local Convergence Rate

In this section, we present the local convergence rate analysis of Algorithm **iPALM**. For this purpose, we denote  $\mathbb{U} := \mathbb{X} \times \mathbb{W}$  and consider the KKT residual mapping of problem (3.1) defined by

$$\mathcal{R}(u) = \mathcal{R}(x, w) := \begin{pmatrix} c - \mathcal{A}^*w \\ w - \text{Prox}_{\varphi}(w - \nabla h(w) - \mathcal{A}x) \end{pmatrix}, \quad \forall u = (x, w) \in \mathbb{X} \times \mathbb{W}. \quad (3.22)$$

Note that  $\mathcal{R}(u) = 0$  if and only if  $u = (x, w)$  is a solution to the KKT system (3.2), whose solution set can therefore be characterized by  $\mathbf{K} := \{u \mid \mathcal{R}(u) = 0\}$ . Moreover, the residual mapping  $\mathcal{R}$  has the following property.

**Lemma 3.3** *Suppose that Assumption 3.1 holds and the sequence  $\{u^k := (x^k, w^k)\}$  is generated by Algorithm **iPALM**. Then, for any  $k \geq 0$ ,*

$$\begin{aligned} \|\mathcal{R}(u^{k+1})\|^2 &\leq \frac{1}{\tau^2\sigma^2} \|x^k - x^{k+1}\|^2 + \frac{2\|\widehat{\Sigma}^h + \mathcal{S}\|}{\tau\sigma} \|w^{k+1} - w^k\|_{\Theta}^2 \\ &\quad + 2\|(1 - \tau^{-1})\mathcal{A}(x^{k+1} - x^k) + d^k\|^2. \end{aligned} \quad (3.23)$$

*Proof* Note that (3.11) holds. Then, one can see that

$$w^{k+1} = \text{Prox}_{\varphi} \left( w^{k+1} + d^k - \nabla h(w^k) - \mathcal{A} \left( x^k + \frac{x^{k+1} - x^k}{\tau} \right) - (\widehat{\Sigma}^h + \mathcal{S})(w^{k+1} - w^k) \right).$$

By taking the above equality and  $c - \mathcal{A}^*w^{k+1} = \frac{1}{\tau\sigma}(x^k - x^{k+1})$  into the definition of  $\mathcal{R}(u^{k+1})$  in (3.22) and using the Lipschitz continuity of  $\text{Prox}_{\varphi}$ , one can get

$$\begin{aligned} \|\mathcal{R}(u^{k+1})\|^2 &\leq \frac{1}{\tau^2\sigma^2} \|x^k - x^{k+1}\|^2 + 2\|(1 - \tau^{-1})\mathcal{A}(x^{k+1} - x^k) + d^k\|^2 \\ &\quad + 2\|\nabla h(w^{k+1}) - \nabla h(w^k) - (\widehat{\Sigma}^h + \mathcal{S})(w^{k+1} - w^k)\|^2. \end{aligned} \quad (3.24)$$

By using Clarke's mean value theorem [8, Proposition 2.6.5] we know that for any  $k \geq 0$  there exists a linear operator  $\Sigma_k : \mathbb{W} \rightarrow \mathbb{W}$  such that  $\nabla h(w^{k+1}) - \nabla h(w^k) = \Sigma_k(w^{k+1} - w^k)$  with  $0 \preceq \Sigma_k \preceq \widehat{\Sigma}^h$ , so that

$$\begin{aligned}
& \|\nabla h(w^{k+1}) - \nabla h(w^k) - (\widehat{\Sigma}^h + \mathcal{S})(w^{k+1} - w^k)\|^2 \\
&= \|(\widehat{\Sigma}^h + \mathcal{S} - \Sigma_k)(w^{k+1} - w^k)\|^2 \\
&\leq \|\widehat{\Sigma}^h + \mathcal{S} - \Sigma_k\| \langle w^{k+1} - w^k, (\widehat{\Sigma}^h + \mathcal{S} - \Sigma_k)(w^{k+1} - w^k) \rangle \\
&\leq \|\widehat{\Sigma}^h + \mathcal{S}\| \langle w^{k+1} - w^k, (\widehat{\Sigma}^h + \mathcal{S})(w^{k+1} - w^k) \rangle \\
&\leq \frac{\|\widehat{\Sigma}^h + \mathcal{S}\|}{\tau\sigma} \left\langle w^{k+1} - w^k, \tau\sigma \left( \widehat{\Sigma}^h + \mathcal{S} + \frac{(2-\tau)\sigma}{3} \mathcal{A}\mathcal{A}^* \right) (w^{k+1} - w^k) \right\rangle, \quad \forall k \geq 0,
\end{aligned} \tag{3.25}$$

where the last inequality comes from the fact that  $0 < \tau < 2$ . Then, by using the definition of  $\Theta$  in (3.14), one can readily see from (3.24) and (3.25) that (3.23) holds. This completes the proof.  $\square$

To analyze the linear convergence rate of Algorithm **iPALM**, we shall introduce the following error bound condition.

**Definition 3.1** The KKT residual mapping  $\mathcal{R}$  defined in (3.22) is said to be metric subregular<sup>3</sup> [11, 3.8 [3H]] (with the modulus  $\kappa > 0$ ) at  $\bar{u} \in \mathbf{K}$  for  $0 \in \mathbb{U}$  if there exists a constant  $r > 0$  such that

$$\text{dist}(u, \mathbf{K}) \leq \kappa \|\mathcal{R}(u)\|, \quad \forall u \in \{u \in \mathbb{U} \mid \|u - \bar{u}\| \leq r\}. \tag{3.26}$$

Suppose that Assumption 3.1 holds. We know from (3.14) and (3.15) that  $\Xi \succ 0$ . Hence, one can let  $\zeta > 0$  be the smallest real number such that  $\zeta \Xi \succeq \Theta$ . For notational convenience, we define the following positive constants:

$$\rho := \max \left\{ \frac{6\sigma^2(\tau-1)^2 \|\mathcal{A}^* \mathcal{A}\| + 3}{\tau\sigma^2(2-\tau)}, \frac{2\zeta \|\widehat{\Sigma}^h + \mathcal{S}\|}{\tau\sigma} \right\} \max \{\|\Theta\|, 1\}, \tag{3.27}$$

$$\beta := \max \left\{ \sqrt{\zeta}, \sqrt{3\tau/(2-\tau)} \right\}, \tag{3.28}$$

$$\mu := \sqrt{\tau\sigma \|\widehat{\Sigma}^h + \mathcal{S} + \frac{2}{3}(1+\tau)\sigma \mathcal{A}\mathcal{A}^*\| \|\mathcal{M}^{-1}\|}. \tag{3.29}$$

To ensure the local linear rate convergence of Algorithm **iPALM**, we need extra conditions to control the error variable  $d^k$  in each iteration. Hence, we make the following assumption.

**Assumption 3.2** There exists an integer  $k_0 > 0$  and a sequence of nonnegative real numbers  $\{\eta_k\}$  such that

$$\sup_{k \geq k_0} \{\eta_k\} < 1/\mu \quad \text{and} \quad \|d^k\| \leq \eta_k \|u^k - u^{k+1}\|, \quad \forall k \geq k_0. \tag{3.30}$$

Now we are ready to present the local convergence rate of Algorithm **iPALM**.

**Theorem 3.2** Suppose that Assumptions 3.1 and 3.2 hold. Let  $\{u^k = (x^k, w^k)\}$  be the sequence generated by Algorithm **iPALM** that converges to  $u^* := (x^*, w^*) \in \mathbf{K}$ . Suppose that the KKT residual mapping  $\mathcal{R}$  defined in (3.22) is metric subregular at  $u^*$  for  $0 \in \mathbb{U}$  with the modulus  $\kappa > 0$ , in the sense that there exists a constant  $r > 0$  such that (3.26) holds with  $\bar{u} = u^*$ . Then, there exists a threshold  $\bar{k} > 0$  such that for all  $k \geq \bar{k}$ ,

$$\text{dist}_\Omega(u^{k+1}, \mathbf{K}) \leq \vartheta^k \text{dist}_\Omega(u^k, \mathbf{K}) \quad \text{with} \quad \vartheta^k := \frac{1}{1 - \mu\eta_k} \left( \sqrt{\frac{\kappa^2 \rho}{1 + \kappa^2 \rho}} + \mu\eta_k(1 + \beta) \right). \tag{3.31}$$

Moreover, if it holds that

$$\sup_{k \geq \bar{k}} \{\eta_k\} < \frac{1}{\mu(2 + \beta)} \left( 1 - \sqrt{\frac{\kappa^2 \rho}{1 + \kappa^2 \rho}} \right), \tag{3.32}$$

then one has  $\sup_{k \geq \bar{k}} \{\vartheta^k\} < 1$ , and the convergence rate of  $\text{dist}_\Omega(u^k, \mathbf{K})$  is  $Q$ -linear when  $k \geq \bar{k}$ .

<sup>3</sup> This is equivalent to say that  $\mathcal{R}^{-1}$  is calm at  $0 \in \mathbb{U}$  for  $\bar{u} \in \mathbf{K}$  with the same modulus  $\kappa > 0$ , see [11, Theorem 3H.3].



*Proof* Denote  $u_e := u - u^*$  for all  $u \in \mathbb{U}$  and define  $\bar{x}^{k+1} := x^k + \tau\sigma(\mathcal{A}^* \bar{w}^{k+1} - c)$  and  $\bar{u}^{k+1} := (\bar{x}^{k+1}, \bar{w}^{k+1})$ ,  $\forall k \geq 0$ . Since  $\{u^k\}$  converges to  $u^*$  and  $\{d^k\}$  converges to 0 as  $k \rightarrow \infty$ , one has from (3.9) that  $\{\bar{u}^k\}$  also converges to  $u^*$  as  $k \rightarrow \infty$ . Therefore, there exists a threshold  $\bar{k} > 0$  such that

$$\|\bar{u}_e^{k+1}\| \leq r \quad \text{and} \quad \|u_e^{k+1}\| \leq r, \quad \forall k \geq \bar{k}. \quad (3.33)$$

According to Lemma 3.3, one can let  $u^{k+1} = \bar{u}^{k+1}$  and  $d^k = 0$  in (3.23) and use the fact that  $\zeta\Xi \succeq \Theta$  to obtain that

$$\begin{aligned} & \|\mathcal{R}(\bar{u}^{k+1})\|^2 \\ & \leq \left( \frac{2(\tau-1)^2 \|\mathcal{A}^* \mathcal{A}\|}{\tau^2} + \frac{1}{(\tau\sigma)^2} \right) \|\bar{x}^{k+1} - x^k\|^2 + \frac{2\|\widehat{\Sigma}^h + \mathcal{S}\|}{\tau\sigma} \|\bar{w}^{k+1} - w^k\|_{\Theta}^2 \\ & \leq \max \left\{ \frac{6\sigma^2(\tau-1)^2 \|\mathcal{A}^* \mathcal{A}\| + 3}{\tau\sigma^2(2-\tau)}, \frac{2\zeta\|\widehat{\Sigma}^h + \mathcal{S}\|}{\tau\sigma} \right\} \left( \frac{(2-\tau)}{3\tau} \|\bar{x}^{k+1} - x^k\|^2 + \|\bar{w}^{k+1} - w^k\|_{\Xi}^2 \right). \end{aligned} \quad (3.34)$$

Moreover, according to the definition of  $\Omega$  in (3.16), one has that

$$\text{dist}_{\Omega}^2(u, \mathbf{K}) \leq \max\{\|\Theta\|, 1\} \text{dist}^2(u, \mathbf{K}), \quad \forall u \in \mathbb{U}.$$

Then, by using the above inequality, together with (3.26), (3.33) and (3.34), we can obtain with the constant  $\rho > 0$  being defined in (3.27) that

$$\begin{aligned} \text{dist}_{\Omega}^2(\bar{u}^{k+1}, \mathbf{K}) & \leq \kappa^2 \max\{\|\Theta\|, 1\} \|\mathcal{R}(\bar{u}^{k+1})\|^2 \\ & \leq \kappa^2 \rho \left( \frac{(2-\tau)}{3\tau} \|\bar{x}^{k+1} - x^k\|^2 + \|\bar{w}^{k+1} - w^k\|_{\Xi}^2 \right), \quad \forall k \geq \bar{k}. \end{aligned} \quad (3.35)$$

It is easy to see from (3.17) that, for any  $k \geq 0$ ,

$$\text{dist}_{\Omega}^2(\bar{u}^{k+1}, \mathbf{K}) \leq \text{dist}_{\Omega}^2(u^k, \mathbf{K}) - \left( \frac{(2-\tau)}{3\tau} \|\bar{x}^{k+1} - x^k\|^2 + \|\bar{w}^{k+1} - w^k\|_{\Xi}^2 \right). \quad (3.36)$$

Therefore, by combining (3.35) and (3.36) together we can get

$$\text{dist}_{\Omega}^2(\bar{u}^{k+1}, \mathbf{K}) \leq \frac{\kappa^2 \rho}{1 + \kappa^2 \rho} \text{dist}_{\Omega}^2(u^k, \mathbf{K}), \quad \forall k \geq \bar{k}. \quad (3.37)$$

From (3.36) and the fact that  $\zeta\Xi \succeq \Theta$  we know that

$$\text{dist}_{\Omega}^2(u^k, \mathbf{K}) \geq \min \left\{ \frac{(2-\tau)}{3\tau}, \frac{1}{\zeta} \right\} \|u^k - \bar{u}^{k+1}\|_{\Omega}^2, \quad \forall k \geq 0.$$

Therefore, it holds that

$$\|u^k - \bar{u}^{k+1}\|_{\Omega} \leq \beta \text{dist}_{\Omega}(u^k, \mathbf{K}), \quad \forall k \geq 0, \quad (3.38)$$

where the constant  $\beta > 0$  is given in (3.28). By using the triangle inequality, we have that

$$\|u^k - \Pi_{\mathbf{K}}^{\Omega}(\bar{u}^{k+1})\|_{\Omega} \leq \text{dist}_{\Omega}(u^k, \mathbf{K}) + \|\Pi_{\mathbf{K}}^{\Omega}(u^k) - \Pi_{\mathbf{K}}^{\Omega}(\bar{u}^{k+1})\|_{\Omega}, \quad \forall k \geq 0. \quad (3.39)$$

Moreover, from [28, Proposition 2.3] we know that

$$\|\Pi_{\mathbf{K}}^{\Omega}(u^k) - \Pi_{\mathbf{K}}^{\Omega}(\bar{u}^{k+1})\|_{\Omega}^2 \leq \langle \Pi_{\mathbf{K}}^{\Omega}(u^k) - \Pi_{\mathbf{K}}^{\Omega}(\bar{u}^{k+1}), \Omega(u^k - \bar{u}^{k+1}) \rangle, \quad \forall k \geq 0.$$

Thus, one has

$$\|\Pi_{\mathbf{K}}^{\Omega}(u^k) - \Pi_{\mathbf{K}}^{\Omega}(\bar{u}^{k+1})\|_{\Omega} \leq \|u^k - \bar{u}^{k+1}\|_{\Omega}, \quad \forall k \geq 0,$$

which together with (3.38) and (3.39), implies that

$$\|u^k - \Pi_{\mathbf{K}}^{\Omega}(\bar{u}^{k+1})\|_{\Omega} \leq (1 + \beta) \text{dist}_{\Omega}(u^k, \mathbf{K}), \quad \forall k \geq 0. \quad (3.40)$$

From the definitions of  $\Theta$  in (3.14) and  $\Omega$  in (3.16) we know that

$$\begin{aligned} \|u^{k+1} - \bar{u}^{k+1}\|_{\Omega}^2 & = (\tau\sigma)^2 \|\mathcal{A}^*(w^{k+1} - \bar{w}^{k+1})\|^2 + \|w^{k+1} - \bar{w}^{k+1}\|_{\Theta}^2 \\ & = \left\langle w^{k+1} - \bar{w}^{k+1}, \tau\sigma(\widehat{\Sigma}^h + \mathcal{S} + \frac{2}{3}(1+\tau)\sigma\mathcal{A}\mathcal{A}^*)(w^{k+1} - \bar{w}^{k+1}) \right\rangle. \end{aligned}$$

Based on the above equality, one can see from (3.19) and (3.29) that

$$\|u^{k+1} - \bar{u}^{k+1}\|_{\Omega} \leq \mu \|d^k\|. \quad (3.41)$$

Since Assumption 3.2 holds, by using (3.30), (3.41) and the triangle inequality one can get that

$$\begin{aligned} \|u^{k+1} - \Pi_{\mathbf{K}}^{\Omega}(\bar{u}^{k+1})\|_{\Omega} &\leq \|u^{k+1} - \bar{u}^{k+1}\|_{\Omega} + \text{dist}_{\Omega}(\bar{u}^{k+1}, \mathbf{K}) \\ &\leq \mu \|d^k\| + \text{dist}_{\Omega}(\bar{u}^{k+1}, \mathbf{K}) \leq \mu \eta_k \|u^k - u^{k+1}\| + \text{dist}_{\Omega}(\bar{u}^{k+1}, \mathbf{K}) \\ &\leq \mu \eta_k \|u^{k+1} - \Pi_{\mathbf{K}}^{\Omega}(\bar{u}^{k+1})\|_{\Omega} + \mu \eta_k \|u^k - \Pi_{\mathbf{K}}^{\Omega}(\bar{u}^{k+1})\|_{\Omega} + \text{dist}_{\Omega}(\bar{u}^{k+1}, \mathbf{K}), \quad \forall k \geq \bar{k}. \end{aligned}$$

Then, by using the fact that  $\|u^{k+1} - \Pi_{\mathbf{K}}^{\Omega}(\bar{u}^{k+1})\|_{\Omega} \geq \text{dist}_{\Omega}(u^{k+1}, \mathbf{K})$  and (3.40), we can obtain that when  $k \geq 0$ ,

$$(1 - \mu \eta_k) \text{dist}_{\Omega}(u^{k+1}, \mathbf{K}) \leq \mu \eta_k (1 + \beta) \text{dist}_{\Omega}(u^k, \mathbf{K}) + \text{dist}_{\Omega}(\bar{u}^{k+1}, \mathbf{K}),$$

which, together with (3.37), implies (3.31). Finally, it is easy to see that  $\sup_{k \geq \bar{k}} \{\vartheta^k\} < 1$  from (3.30) and (3.32). This completes the proof.  $\square$

*Remark 3.1* Note that if  $\{\eta_k\} \rightarrow 0$  as  $k \rightarrow \infty$ , condition (3.32) holds eventually for  $\bar{k}$  sufficiently large.

### 3.3 Non-Ergodic Iteration Complexity

With the inequalities established in the previous subsections, one can easily get the following non-ergodic iteration complexity results for Algorithm **iPALM**.

**Theorem 3.3** Suppose that Assumption 3.1 holds. Let  $\{u^k = (x^k, w^k)\}$  be the sequence generated by Algorithm **iPALM** that converges to  $u^* := (x^*, w^*) \in \mathbf{K}$ . Then, the KKT residual mapping  $\mathcal{R}$  defined in (3.22) satisfies

$$\min_{0 \leq j \leq k} \|\mathcal{R}(u^j)\|^2 \leq \varrho/k \quad \text{and} \quad \lim_{k \rightarrow \infty} (k \cdot \min_{0 \leq j \leq k} \|\mathcal{R}(u^j)\|^2) = 0, \quad (3.42)$$

where the constant  $\varrho$  is defined by

$$\varrho := \max \left\{ \frac{12\sigma^2(\tau-1)^2 \|\mathcal{A}^* \mathcal{A}\| + 3}{\tau\sigma^2(2-\tau)}, \frac{2\zeta \|\widehat{\Sigma}^h + \mathcal{S}\|}{\tau\sigma} \right\} \mathbf{e} \quad (3.43)$$

with  $\mathbf{e} := \|u_e^0\|_{\Omega}^2 + 2\tau\sigma \|\Theta^{-1/2}\|(\sum_{j=0}^{\infty} \varepsilon_j)(\|u_e^0\|_{\Omega} + \mu \sum_{j=0}^{\infty} \varepsilon_j) + 4 \sum_{j=1}^{\infty} \varepsilon_j^2$ .

*Proof* From (3.17) in Theorem 3.1(a) we know that  $\|\bar{u}_e^{j+1}\|_{\Omega} \leq \|u_e^j\|_{\Omega}$ ,  $\forall j \geq 0$ . Moreover, (3.41) still holds with  $\mu$  being given in (3.29), so that

$$\|u^{j+1} - \bar{u}^{j+1}\|_{\Omega} \leq \mu \|d^j\|, \quad \forall j \geq 0.$$

Therefore,

$$\|w_e^{j+1}\|_{\Theta} \leq \|u_e^{j+1}\|_{\Omega} \leq \|u_e^j\|_{\Omega} + \mu \|d^j\| \leq \|u_e^0\|_{\Omega} + \mu \sum_{j=0}^{\infty} \varepsilon_j, \quad \forall j \geq 0.$$

Consequently, for any  $k \geq 0$

$$\begin{aligned} \sum_{j=0}^k \langle d^j, w_e^{j+1} \rangle &\leq \|\Theta^{-1/2}\| \left( \sum_{j=0}^k \|d^j\| \right) \|w_e^{j+1}\|_{\Theta} \\ &\leq \|\Theta^{-1/2}\| \left( \sum_{j=0}^{\infty} \varepsilon_j \right) \left( \|u_e^0\|_{\Omega} + \mu \sum_{j=0}^{\infty} \varepsilon_j \right). \end{aligned} \quad (3.44)$$

Also, from (3.17) of Theorem 3.1(a) we know that for any  $k \geq 0$ ,

$$\begin{aligned} \|u_e^0\|_{\Omega}^2 &\geq \|u_e^0\|_{\Omega}^2 - \|u_e^{k+1}\|_{\Omega}^2 = \sum_{j=0}^k (\|u_e^j\|_{\Omega}^2 - \|u_e^{j+1}\|_{\Omega}^2) \\ &\geq \sum_{j=0}^k \left( \|w^{j+1} - w^j\|_{\Xi}^2 + \frac{2-\tau}{3\tau} \|x^{j+1} - x^j\|^2 \right) - 2\tau\sigma \sum_{j=0}^k \langle d^j, w_e^{j+1} \rangle. \end{aligned} \quad (3.45)$$

Moreover, from (3.23) we know that

$$\|\mathcal{R}(u^{k+1})\|^2 \leq \frac{4\sigma^2(\tau-1)^2 \|\mathcal{A}^* \mathcal{A}\| + 1}{\tau^2\sigma^2} \|x^{k+1} - x^k\|^2 + \frac{2\zeta \|\widehat{\Sigma}^h + \mathcal{S}\|}{\tau\sigma} \|w^{k+1} - w^k\|_{\Xi}^2 + 4\|d^k\|^2.$$

Therefore, we can get from (3.44) and (3.45) that  $\sum_{j=0}^{\infty} \|\mathcal{R}(u^{j+1})\|^2 \leq \varrho$ . From here, we can easily get required results in (3.42).  $\square$

#### 4 The Equivalence Property

In this section, we establish the equivalence of an inexact block sGS decomposition based multi-block indefinite-proximal ADMM for solving problem (1.1) to the inexact indefinite-proximal ALM presented in the previous section. The iteration scheme of the former has already been briefly sketched in (1.7) in the introduction. Here we shall formally present it as Algorithm sGS-iPADMM.

Recall that the majorized augmented Lagrangian function of problem (1.1) was given in (1.5) and the function  $\mathcal{L}_\sigma$  that will be used in Algorithm sGS-iPADMM was defined (1.6).

---

**Algorithm sGS-iPADMM** An inexact block sGS decomposition based indefinite-Proximal ADMM

---

Let  $\{\tilde{\varepsilon}_k\}$  be a summable sequence of nonnegative real numbers,  $\tau \in (0, 2)$  be the (dual) step-length, and  $(x^0, y^0, z^0) \in \mathbb{X} \times \text{dom } p \times \mathbb{Y}_2 \times \cdots \times \mathbb{Y}_s \times \mathbb{Z}$  be the given initial point. Choose the self-adjoint linear operators  $\mathcal{D}_i : \mathbb{Y}_i \rightarrow \mathbb{Y}_i$ ,  $i = 1, \dots, s$ . For  $k = 0, 1, \dots$ , perform the following steps in each iteration.

**Step 1.** For  $i = s, \dots, 2$ , compute

$$y_i^{k+\frac{1}{2}} \approx \arg \min_{y_i \in \mathbb{Y}_i} \mathcal{L}_\sigma \left( (y_{<i}^k; y_i; y_{>i}^{k+\frac{1}{2}}), z^k; (x^k, y^k) \right),$$

such that there exists  $\tilde{\delta}_i^k$  satisfying  $\|\tilde{\delta}_i^k\| \leq \tilde{\varepsilon}_k$  and  $\tilde{\delta}_i^k \in \partial_{y_i} \mathcal{L}_\sigma \left( (y_{<i}^k; y_i^{k+\frac{1}{2}}; y_{>i}^{k+\frac{1}{2}}), z^k; (x^k, y^k) \right)$ .

**Step 2.** For  $i = 1, \dots, s$ , compute

$$y_i^{k+1} \approx \arg \min_{y_i \in \mathbb{Y}_i} \mathcal{L}_\sigma \left( (y_{<i}^{k+1}; y_i; y_{>i}^{k+\frac{1}{2}}), z^k; (x^k, y^k) \right),$$

such that there exists  $\delta_i^k$  satisfying  $\|\delta_i^k\| \leq \tilde{\varepsilon}_k$  and  $\delta_i^k \in \partial_{y_i} \mathcal{L}_\sigma \left( (y_{<i}^{k+1}; y_i^{k+1}; y_{>i}^{k+\frac{1}{2}}), z^k; (x^k, y^k) \right)$ .

**Step 3.** Compute  $z^{k+1} \approx \arg \min_{z \in \mathbb{Z}} \{ \mathcal{L}_\sigma (y^{k+1}, z; (x^k, y^k)) \}$ , such that  $\|\gamma^k\| \leq \tilde{\varepsilon}_k$  with

$$\gamma^k := \nabla_z \mathcal{L}_\sigma (y^{k+1}, z^{k+1}; (x^k, y^k)) = \mathcal{G}x^k - b + \sigma \mathcal{G}(\mathcal{F}^* y^{k+1} + \mathcal{G}^* z^{k+1} - c). \quad (4.1)$$

**Step 4.** Compute  $x^{k+1} := x^k + \tau \sigma (\mathcal{F}^* y^{k+1} + \mathcal{G}^* z^{k+1} - c)$ .

---

Recall that the KKT system of problem (1.1) is defined by

$$0 \in \begin{pmatrix} \partial p(y_1) \\ 0 \end{pmatrix} + \nabla f(y) + \mathcal{F}x, \quad \mathcal{G}x - b = 0, \quad \mathcal{F}^* y + \mathcal{G}^* z = c. \quad (4.2)$$

We make the following assumption on problem (1.1) throughout this section.

**Assumption 4.1** The solution set to the KKT system (4.2) is nonempty.

*Remark 4.1* Note that if Slater's constraint qualification (SCQ) holds for problem (1.1), i.e.,

$$\{(y, z) \mid y_1 \in \text{ri}(\text{dom } p), \mathcal{F}^* y + \mathcal{G}^* z = c\} \neq \emptyset,$$

then we know from [44, Corollaries 28.2.2 & 28.3.1] that a vector  $(y, z) \in \mathbb{Y} \times \mathbb{Z}$  is a solution to problem (1.1) if and only if there exists a Lagrangian multiplier  $x \in \mathbb{X}$  such that  $(x, y, z)$  is a solution to the KKT system (4.2). Therefore, Assumption 4.1 holds if the SCQ holds and (1.1) has at least one optimal solution. Moreover, for any  $(x, y, z) \in \mathbb{X} \times \mathbb{Y} \times \mathbb{Z}$  satisfying (4.2), we know from [44, Corollary 30.5.1] that  $(y, z)$  is an optimal solution to problem (1.1) and  $x$  is an optimal solution to its dual problem.

We are now ready for proving the equivalence of Algorithm 1 and Algorithm sGS-iPADMM for solving problem (1.1). Note that one can always write  $\mathcal{F}x = (\mathcal{F}_1 x; \dots; \mathcal{F}_s x)$ ,  $\forall x \in \mathbb{X}$  with each  $\mathcal{F}_i : \mathbb{X} \rightarrow \mathbb{Y}_i$  being a

given linear mapping. For the convenience of the later discussions, we will symbolically decompose an arbitrary self-adjoint linear operator  $\mathcal{O} : \mathbb{Y} \rightarrow \mathbb{Y}$ , if necessary, to the following from

$$\mathcal{O} = \begin{pmatrix} \mathcal{O}_{11} & \cdots & \mathcal{O}_{1s} \\ \vdots & \ddots & \vdots \\ \mathcal{O}_{s1} & \cdots & \mathcal{O}_{ss} \end{pmatrix} \quad \text{with} \quad \mathcal{O}_{ij} : \mathbb{Y}_j \rightarrow \mathbb{Y}_i, \quad \forall 1 \leq i, j \leq s, \quad (4.3)$$

so that  $\mathcal{O}_{ij} = \mathcal{O}_{ji}^*$ ,  $\forall 1 \leq i, j \leq s$ . Moreover, we use  $\mathcal{O}_d : \mathbb{Y} \rightarrow \mathbb{Y}$  to denote the block diagonal part of  $\mathcal{O}$ , i.e.,  $\mathcal{O}_d = \text{Diag}(\mathcal{O}_{11}, \dots, \mathcal{O}_{ss})$  and  $\mathcal{O}_u$  to denote the strictly block upper triangular part of  $\mathcal{O}$ .

Based on the above decomposition and notation, we begin by applying the inexact block sGS decomposition theorem [32, Theorem 1] to express the procedure for computing  $y^{k+1}$  in Steps 1 and 2 of Algorithm **sGS-iPADMM** in a more compact fashion. For this purpose we define the following linear operator

$$\mathcal{N} := \widehat{\Sigma}^f + \sigma \mathcal{F} \mathcal{F}^* + \mathcal{D}. \quad (4.4)$$

Moreover, as can be seen from (1.6), for any given  $(x, y', z) \in \mathbb{X} \times \mathbb{Y} \times \mathbb{Z}$ , the linear operator  $\mathcal{N}$  contains all the quadratic information of the function  $\mathcal{L}_\sigma(y, z; (x, y'))$  with respect to  $y$ . Based on (4.3), the linear operator  $\mathcal{N}$  can be decomposed as  $\mathcal{N} = \mathcal{N}_d + \mathcal{N}_u + \mathcal{N}_u^*$  with  $\mathcal{N}_d$  and  $\mathcal{N}_u$  being the block-diagonal and the strict block-upper triangular parts of  $\mathcal{N}$ , respectively, i.e.,

$$\mathcal{N}_d := \text{Diag}(\mathcal{N}_{11}, \dots, \mathcal{N}_{ss}) \quad \text{with} \quad \mathcal{N}_{ii} := \widehat{\Sigma}_{ii}^f + \sigma \mathcal{F}_i \mathcal{F}_i^* + \mathcal{D}_{ii}, \quad i = 1, \dots, s$$

and

$$\mathcal{N}_u := \begin{pmatrix} 0 & \mathcal{N}_{12} & \cdots & \mathcal{N}_{1s} \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \mathcal{N}_{(s-1)s} \\ 0 & 0 & \cdots & 0 \end{pmatrix} \quad \text{with} \quad \mathcal{N}_{ij} = \widehat{\Sigma}_{ij}^f + \sigma \mathcal{F}_i \mathcal{F}_j^* + \mathcal{D}_{ij}, \quad \forall 1 \leq i < j \leq s. \quad (4.5)$$

For convenience, we denote in Algorithm **sGS-iPADMM** for each  $k \geq 0$ ,  $\delta_1^k := \delta_1^k$ ,  $\tilde{\delta}^k := (\tilde{\delta}_1^k, \tilde{\delta}_2^k, \dots, \tilde{\delta}_s^k)$  and  $\delta^k := (\delta_1^k, \dots, \delta_s^k)$ . When  $\mathcal{N}_d \succ 0$ , one can define the sequence  $\{\delta_{\text{sGS}}^k\}$  in  $\mathbb{Y}$  by

$$\delta_{\text{sGS}}^k := \delta^k + \mathcal{N}_u \mathcal{N}_d^{-1} (\delta^k - \tilde{\delta}^k). \quad (4.6)$$

Meanwhile, one can define the linear operator

$$\mathcal{N}_{\text{sGS}} := \mathcal{N}_u \mathcal{N}_d^{-1} \mathcal{N}_u^*. \quad (4.7)$$

The following assumption on the proximal terms  $\mathcal{N}$  and  $\mathcal{D}$  is also needed for using the sGS decomposition theorem.

**Assumption 4.2** *The self-adjoint linear operator  $\mathcal{D} : \mathbb{Y} \rightarrow \mathbb{Y}$  in (1.6) is chosen such that*

$$\mathcal{N} = \widehat{\Sigma}^f + \sigma \mathcal{F} \mathcal{F}^* + \mathcal{D} \succeq 0, \quad \mathcal{N}_d \succ 0. \quad (4.8)$$

We note that Assumption 4.2 is rather weak and is merely used to guarantee the convexity of function  $\mathcal{L}_\sigma(y, z; (x, y'))$  in terms of  $y$  and the well-definedness of the steps 1 and 2 in Algorithm **sGS-iPADMM**. The following result is a direct consequence of [32, Theorem 1].

**Lemma 4.1** *Suppose that Assumption 4.2 holds. Then, it holds that  $\mathcal{N} + \mathcal{N}_{\text{sGS}} = (\mathcal{N}_d + \mathcal{N}_u) \mathcal{N}_d^{-1} (\mathcal{N}_d + \mathcal{N}_u^*) \succ 0$ . Moreover, the iterate  $y^{k+1}$  in Step 2 of Algorithm **sGS-iPADMM** is the unique solution to the following perturbed proximal minimization problem:*

$$\begin{aligned} y^{k+1} &= \arg \min_{y \in \mathbb{Y}} \left\{ \mathcal{L}_\sigma(y, z^k; (x^k, y^k)) + \frac{1}{2} \|y - y^k\|_{\mathcal{N}_{\text{sGS}}}^2 - \langle \delta_{\text{sGS}}^k, y \rangle \right\} \\ &= \arg \min_{y \in \mathbb{Y}} \left\{ \widehat{\mathcal{L}}_\sigma(y, z^k; (x^k, y^k)) + \frac{1}{2} \|y - y^k\|_{\mathcal{D} + \mathcal{N}_{\text{sGS}}}^2 - \langle \delta_{\text{sGS}}^k, y \rangle \right\}. \end{aligned} \quad (4.9)$$

**Remark 4.2** From (4.9) one can get the interpretation of the linear operator  $\mathcal{N}_{\text{sGS}}$  defined in (4.7). That is, by adding the proximal term  $\frac{1}{2}\|y - y^k\|_{\mathcal{D}}^2$  to the majorized augmented Lagrangian function and conduct one cycle of the block sGS-type block coordinate minimization via Steps 1 and 2 in Algorithm sGS-iPADMM, the resulted  $y^{k+1}$  is then an inexact solution to the following problem

$$\min_{y \in \mathbb{Y}} \left\{ \widehat{\mathcal{L}}_{\sigma}(y, z^k; (x^k, y^k)) + \frac{1}{2}\|y - y^k\|_{\mathcal{D}}^2 + \frac{1}{2}\|y - y^k\|_{\mathcal{N}_{\text{sGS}}}^2 \right\},$$

where the proximal term  $\frac{1}{2}\|y - y^k\|_{\mathcal{N}_{\text{sGS}}}^2$  is generated due to the sGS-type iteration with the linear operator  $\mathcal{N}_{\text{sGS}}$  being defined by (4.7) and (4.4), while  $\delta_{\text{sGS}}^k$  defined in (4.6) represents the error accumulation resulted from the inexact computations in Steps 1 and 2 in Algorithm sGS-iPADMM.

The following elementary result<sup>4</sup> will be frequently used later.

**Lemma 4.2** *The self-adjoint linear operator  $\mathcal{G}\mathcal{G}^*$  is nonsingular (positive definite) on the subspace  $\text{Range}(\mathcal{G})$  of  $\mathbb{Z}$ .*

Now, we start to establish the equivalence between Algorithm sGS-iPADMM and Algorithm iPALM. The first step is to show that the procedure of obtaining  $(y^{k+1}, z^{k+1})$  in Algorithm sGS-iPADMM can be viewed as the procedure of getting  $w^{k+1}$  in Algorithm iPALM. For this purpose, we define the block diagonal linear operator  $\mathcal{T} : \mathbb{Y} \times \mathbb{Z} \rightarrow \mathbb{Y} \times \mathbb{Z}$  by

$$\mathcal{T}(y; z) := \begin{pmatrix} (\mathcal{D} + \mathcal{N}_{\text{sGS}} + \sigma \mathcal{F}\mathcal{G}^*[\mathcal{G}\mathcal{G}^*]^{\dagger} \mathcal{G}\mathcal{F}^*)y \\ 0 \end{pmatrix}, \quad \forall (y, z) \in \mathbb{Y} \times \mathbb{Z}. \quad (4.10)$$

Moreover, we define the sequence  $\{\Delta^k\}$  in  $\mathbb{Y}$  by

$$\Delta^k := \delta_{\text{sGS}}^k - \mathcal{F}\mathcal{G}^*[\mathcal{G}\mathcal{G}^*]^{\dagger} (\gamma^{k-1} - \gamma^k - \mathcal{G}(x^{k-1} - x^k)), \quad k \geq 0 \quad (4.11)$$

with the convention that

$$\begin{cases} x^{-1} := x^0 - \tau\sigma(\mathcal{F}^*y^0 + \mathcal{G}^*z^0 - c), \\ \gamma^{-1} := -b + \mathcal{G}x^{-1} + \sigma\mathcal{G}(\mathcal{F}^*y^0 + \mathcal{G}^*z^0 - c). \end{cases} \quad (4.12)$$

Based on the above definitions and Lemma 4.1, we have the following result.

**Proposition 4.1** *Suppose that Assumption 4.2 holds. Then,*

- (a) *Algorithm sGS-iPADMM is well-defined;*
- (b) *the sequence  $\{(x^k, y^k, z^k)\}$  generated by Algorithm sGS-iPADMM satisfies*

$$(\Delta^k; \gamma^k) \in \partial_{(y,z)} \widehat{\mathcal{L}}_{\sigma}((y^{k+1}, z^{k+1}); (x^k, y^k)) + \mathcal{T}(y^{k+1} - y^k; z^{k+1} - z^k), \quad \forall k \geq 0. \quad (4.13)$$

*Proof (a)* Since Assumption 4.2 holds, it is easy to see from Lemma 4.1 that Steps 1 and 2 in algorithm sGS-iPADMM are well-defined for any  $k \geq 0$ . Moreover, from (4.1) we know that Step 3 of Algorithm sGS-iPADMM is well-defined if, for any  $k \geq 0$ , the following linear system with respect to  $z$

$$\mathcal{G}x^k - b + \sigma\mathcal{G}(\mathcal{F}^*y^{k+1} + \mathcal{G}^*z - c) = 0 \quad (4.14)$$

has a solution. Since  $b \in \text{Range}(\mathcal{G})$ , we know that  $(b - \mathcal{G}x^k)/\sigma - \mathcal{G}(\mathcal{F}^*y^{k+1} - c) \in \text{Range}(\mathcal{G})$ . Therefore, Lemma 4.2 implies that the linear system

$$\mathcal{G}\mathcal{G}^*z = (b - \mathcal{G}x^k)/\sigma - \mathcal{G}(\mathcal{F}^*y^{k+1} - c)$$

or equivalently the linear system (4.14), has a solution. Consequently, Algorithm sGS-iPADMM is well-defined.

(b) From (4.1) and (4.12) we know that for any  $k \geq 0$ ,

$$\gamma^{k-1} = -b + \mathcal{G}x^{k-1} + \sigma\mathcal{G}(\mathcal{F}^*y^k + \mathcal{G}^*z^k - c) \quad (4.15)$$

so that  $\gamma^{k-1} \in \text{Range}(\mathcal{G})$  and  $\mathcal{G}\mathcal{G}^*z^k = (\gamma^{k-1} + b - \mathcal{G}x^{k-1})/\sigma - \mathcal{G}\mathcal{F}^*y^k + \mathcal{G}c$ . Hence,

$$\mathcal{G}\mathcal{G}^*(z^k - z^{k+1}) = \frac{1}{\sigma}(\gamma^{k-1} - \gamma^k - \mathcal{G}x^{k-1} + \mathcal{G}x^k) - \mathcal{G}\mathcal{F}^*(y^k - y^{k+1}), \quad \forall k \geq 0.$$

<sup>4</sup> This lemma can be directly verified via the singular value decomposition of the linear operator  $\mathcal{G}$  and some basic calculations from linear functional analysis.

Therefore, one can get<sup>5</sup> that for any  $k \geq 0$ ,

$$\begin{aligned} & \sigma \mathcal{F} \mathcal{G}^*(z^k - z^{k+1}) \\ &= \mathcal{F} \mathcal{G}^* [\mathcal{G} \mathcal{G}^*]^\dagger (\gamma^{k-1} - \gamma^k - \mathcal{G}(x^{k-1} - x^k)) + \sigma \mathcal{F} \mathcal{G}^* [\mathcal{G} \mathcal{G}^*]^\dagger \mathcal{G} \mathcal{F}^*(y^{k+1} - y^k). \end{aligned} \quad (4.16)$$

From (4.9) in Lemma 4.1 we know that, for any  $k \geq 0$ ,

$$\begin{aligned} \delta_{\text{sGS}}^k &\in \partial_y \widehat{\mathcal{L}}_\sigma(y^{k+1}, z^k; (x^k, y^k)) + (\mathcal{D} + \mathcal{N}_{\text{sGS}})(y^{k+1} - y^k) \\ &= \partial_y \widehat{\mathcal{L}}_\sigma(y^{k+1}, z^{k+1}; (x^k, y^k)) + (\mathcal{D} + \mathcal{N}_{\text{sGS}})(y^{k+1} - y^k) + \sigma \mathcal{F} \mathcal{G}^*(z^k - z^{k+1}). \end{aligned} \quad (4.17)$$

Then, by substituting (4.16) into (4.17) and using the definition of  $\Delta^k$  in (4.11), one has that

$$\Delta^k \in \partial_y \widehat{\mathcal{L}}_\sigma(y^{k+1}, z^{k+1}; (x^k, y^k)) + (\mathcal{D} + \mathcal{N}_{\text{sGS}} + \sigma \mathcal{F} \mathcal{G}^* [\mathcal{G} \mathcal{G}^*]^\dagger \mathcal{G} \mathcal{F}^*)(y^{k+1} - y^k),$$

which, together with (4.1), implies that (4.13) holds. This completes the proof.  $\square$

The following important result will be used later.

**Proposition 4.2** *Suppose that Assumption 4.2 holds. Let  $\{(x^k, y^k, z^k)\}$  be the sequence generated by Algorithm sGS-iPADMM. Define  $\xi_0 := \|b - \mathcal{G}x^0\|$  and*

$$\xi_k := |1 - \tau|^k \xi_0 + \tau \sum_{i=1}^k |1 - \tau|^{k-i} \tilde{\varepsilon}_{i-1}, \quad \forall k \geq 1.$$

Then, it holds that for all  $k \geq 0$ ,  $\|b - \mathcal{G}x^k\| \leq \xi_k$  and

$$\sum_{k=0}^{\infty} \|b - \mathcal{G}x^k\| \leq \sum_{k=0}^{\infty} \xi_k < +\infty.$$

*Proof* We know from Step 4 of Algorithm sGS-iPADMM and (4.12) that

$$x^k = x^{k-1} + \tau \sigma (\mathcal{F}^* y^k + \mathcal{G}^* z^k - c), \quad \forall k \geq 0.$$

Hence, one has that

$$b - \mathcal{G}x^k = b - \mathcal{G}x^{k-1} - \tau \sigma \mathcal{G}(\mathcal{F}^* y^k + \mathcal{G}^* z^k - c), \quad \forall k \geq 0.$$

Moreover, from (4.12) and (4.15) we know that

$$\tau(\gamma^{k-1} + b - \mathcal{G}x^{k-1}) = \tau \sigma \mathcal{G}(\mathcal{F}^* y^k + \mathcal{G}^* z^k - c), \quad \forall k \geq 0.$$

Thus, by combining the above two equalities together, one can get

$$b - \mathcal{G}x^k = b - \mathcal{G}x^{k-1} - \tau(\gamma^{k-1} + b - \mathcal{G}x^{k-1}) = (1 - \tau)(b - \mathcal{G}x^{k-1}) - \tau\gamma^{k-1}, \quad \forall k \geq 0.$$

Consequently, it holds that

$$\|b - \mathcal{G}x^k\| \leq |1 - \tau| \|b - \mathcal{G}x^{k-1}\| + \tau \|\gamma^{k-1}\|, \quad \forall k \geq 0,$$

and hence

$$\|b - \mathcal{G}x^k\| \leq |1 - \tau|^k \|b - \mathcal{G}x^0\| + \tau \sum_{i=1}^k |1 - \tau|^{k-i} \|\gamma^{i-1}\| \leq \xi_k, \quad \forall k \geq 0. \quad (4.18)$$

Note that  $\tau \in (0, 2)$ . It is easy to see that

$$\begin{aligned} \sum_{k=0}^{\infty} \|b - \mathcal{G}x^k\| &\leq \sum_{k=0}^{\infty} \xi_k \leq \left( \sum_{k=0}^{\infty} |1 - \tau|^k \right) \xi_0 + \tau \sum_{k=1}^{\infty} \sum_{i=1}^k |1 - \tau|^{k-i} \tilde{\varepsilon}_{i-1} \\ &\leq \left( \sum_{k=0}^{\infty} |1 - \tau|^k \right) \xi_0 + \tau \left( \sum_{k=0}^{\infty} |1 - \tau|^k \right) \left( \sum_{i=0}^{\infty} \tilde{\varepsilon}_i \right) < +\infty. \end{aligned}$$

$\square$

<sup>5</sup> This can be routinely derived by using the singular value decomposition of  $\mathcal{G}$  and the definition of the Moore–Penrose pseudoinverse.



Now, we start to show that the sequence  $\{(x^k, y^k, z^k)\}$  generated by Algorithm **sGS-iPADMM** can be viewed as a sequence generated by Algorithm **iPALM** from the same initial point. For this purpose, we define the space  $\mathbb{V} := \mathbb{Y} \times \text{Range}(\mathcal{G})$ , and we define the linear operators  $\mathcal{B} : \mathbb{X} \rightarrow \mathbb{V}$  and  $\mathcal{P} : \mathbb{V} \rightarrow \mathbb{V}$  by

$$\mathcal{B}x := (\mathcal{F}x; \mathcal{G}x), \quad \forall x \in \mathbb{X} \quad \text{and} \quad \mathcal{P}(y, z) := (\widehat{\Sigma}^f y; 0), \quad \forall (y, z) \in \mathbb{V}. \quad (4.19)$$

Moreover, we define the closed proper convex function  $\phi : \mathbb{V} \rightarrow (-\infty, +\infty]$  by

$$\phi(v) = \phi(y, z) := p(y_1) + f(y) - \langle b, z \rangle, \quad \forall v = (y, z) \in \mathbb{V}$$

and define

$$\widehat{\mathcal{L}}_\sigma(v; (x, v')) := \widehat{\mathcal{L}}_\sigma(y, z; (x, y')), \quad \forall v = (y, z) \in \mathbb{V}, \quad v' = (y', z') \in \mathbb{V}. \quad (4.20)$$

Based on the above definitions, problem (1.1) can be viewed as an instance of problem (3.1). Then, we show in the next theorem that Algorithm **sGS-iPADMM** can be viewed an instance of Algorithm **iPALM**.

**Theorem 4.1** *Suppose that Assumption 4.2 holds. Let  $\{(x^k, y^k, z^k)\}$  be the sequence generated by Algorithm **sGS-iPADMM**. Define*

$$v^k := (y^k; \Pi_{\text{Range}(\mathcal{G})}(z^k)), \quad \forall k \geq 0. \quad (4.21)$$

*Then, for any  $k \geq 0$ , it holds that*

$$v^{k+1} \approx \arg \min_{v \in \mathbb{V}} \left\{ \widehat{\mathcal{L}}_\sigma(v; (x^k, v^k)) + \frac{1}{2} \|v - v^k\|_{\mathcal{T}}^2 \right\}$$

*in the sense that there exists a nonnegative summable sequence  $\{\widehat{\varepsilon}_k\}$  such that*

$$(\Delta^k; \gamma^k) \in \partial_v \widehat{\mathcal{L}}_\sigma(v^{k+1}; (x^k, v^k)) + \mathcal{T}(v^{k+1} - v^k) \quad \text{and} \quad \|(\Delta^k; \gamma^k)\| \leq \widehat{\varepsilon}_k,$$

*where  $\gamma^k$  and  $\Delta^k$  are defined in (4.1) and (4.11), respectively.*

*Proof* From definitions (1.4) and (1.5), it is easy to see that

$$\partial_y \widehat{\mathcal{L}}_\sigma(y, z; (x, y')) = \begin{pmatrix} \partial_{y_1} p(y_1) \\ 0 \end{pmatrix} + \nabla f(y') + \widehat{\Sigma}^f(y - y') + \mathcal{F}x + \sigma \mathcal{F}(\mathcal{F}^* y + \mathcal{G}^* z - c)$$

and

$$\nabla_z \widehat{\mathcal{L}}_\sigma(y, z; (x, y')) = -b + \mathcal{G}x + \sigma \mathcal{G}(\mathcal{F}^* y + \mathcal{G}^* z - c).$$

Since  $\mathcal{G}^* z^{k+1} = \mathcal{G}^* \Pi_{\text{Range}(\mathcal{G})}(z^{k+1})$ ,  $\forall k \geq 0$ , we know that

$$\partial_{(y,z)} \widehat{\mathcal{L}}_\sigma((y^{k+1}, z^{k+1}); (x^k, y^k)) = \partial_{(y,z)} \widehat{\mathcal{L}}_\sigma((y^{k+1}, \Pi_{\text{Range}(\mathcal{G})}(z^{k+1})); (x^k, y^k)), \quad \forall k \geq 0.$$

According to (4.10), (4.13) and (4.20), we obtain that

$$(\Delta^k; \gamma^k) \in \partial_v \widehat{\mathcal{L}}_\sigma(v^{k+1}; (x^k, v^k)) + \mathcal{T}(v^{k+1} - v^k), \quad \forall k \geq 0.$$

Next, we focus on bounding the error term  $\|(\Delta^k; \gamma^k)\|$ . From the definition of  $\{\Delta^k\}$  in (4.11) one can see that for all  $k \geq 0$ ,

$$\|\Delta^k\| \leq \|\delta_{\text{sGS}}^k\| + \|\mathcal{F}\mathcal{G}^*[\mathcal{G}\mathcal{G}^*]^\dagger\| \|\gamma^{k-1} - \gamma^k - \mathcal{G}(x^{k-1} - x^k)\|.$$

Since, in the  $k$ -th iteration of Algorithm **sGS-iPADMM**,  $\max_{i=1, \dots, s} \{\|\tilde{\delta}_i^k\|, \|\delta_i^k\|\} \leq \tilde{\varepsilon}_k$ , it is not difficult to bound  $\|\delta_{\text{sGS}}^k\|$  as follows:

$$\|\delta_{\text{sGS}}^k\| \leq \|\delta^k\| + \|\mathcal{N}_u \mathcal{N}_d^{-1}\| (\|\delta^k\| + \|\tilde{\delta}^k\|) \leq s\tilde{\varepsilon}_k + 2s\|\mathcal{N}_u \mathcal{N}_d^{-1}\| \tilde{\varepsilon}_k.$$

Meanwhile, by  $\|\gamma_k\| \leq \tilde{\varepsilon}_k$  and (4.18), we know that for all  $k \geq 1$

$$\begin{aligned} \|\gamma^{k-1} - \gamma^k - \mathcal{G}(x^{k-1} - x^k)\| &\leq \|\gamma^{k-1}\| + \|\gamma^k\| + \|\mathcal{G}x^{k-1} - b\| + \|\mathcal{G}x^k - b\| \\ &\leq \tilde{\varepsilon}_{k-1} + \xi_{k-1} + \tilde{\varepsilon}_k + \xi_k. \end{aligned}$$

Therefore, it holds for all  $k \geq 1$  that

$$\begin{aligned} \|(\Delta^k; \gamma^k)\| &\leq \|\gamma^k\| + \|\Delta^k\| \\ &\leq \widehat{\varepsilon}_k := (s+1)\tilde{\varepsilon}_k + 2s\|\mathcal{N}_u \mathcal{N}_d^{-1}\| \tilde{\varepsilon}_k + \|\mathcal{F}\mathcal{G}^*[\mathcal{G}\mathcal{G}^*]^\dagger\| (\tilde{\varepsilon}_{k-1} + \xi_{k-1} + \tilde{\varepsilon}_k + \xi_k). \end{aligned}$$

Moreover, we define  $\widehat{\varepsilon}_0 := \|(\Delta^0; \gamma^0)\|$ . Since the sequences  $\{\tilde{\varepsilon}_k\}$  and  $\{\xi_k\}$  are summable, we know that  $\sum_{k=0}^{\infty} \widehat{\varepsilon}_k < +\infty$ . We thus complete the proof.  $\square$

**Remark 4.3** One can see that in Algorithm **sGS-iPADMM**, the sequence  $\{(x^k, y^k, z^k)\}$  was generated, while the sequence  $\{\Pi_{\text{Range}(\mathcal{G})}(z^k)\}$  has never been explicitly calculated. Note that once  $z^k$  is computed, only the vector  $\mathcal{G}^* z^k$  is needed during the next iteration, instead of  $z^k$  itself. Since  $\mathcal{G}^* z^k = \mathcal{G}^* \Pi_{\text{Range}(\mathcal{G})}(z^k)$ ,  $\forall k \geq 0$ , one may view the sequence  $\{\Pi_{\text{Range}(\mathcal{G})}(z^k)\} \in \text{Range}(\mathcal{G})$  as a shadow sequence of  $\{z^k\}$ . It has never been explicitly computed, but still plays an important role on establishing the convergence of the algorithm. In fact, similar observations have been made and extensively used in [30, 31].

The equivalence between Algorithm **sGS-iPADMM** and Algorithm **iPALM** derived in Theorem 4.1 allows us to study the convergence properties of Algorithm **sGS-iPADMM** via using Theorem 3.1. For this purpose, we need the following assumption on choosing the proximal terms in Algorithm **sGS-iPADMM**.

**Assumption 4.3** The self-adjoint linear operator  $\mathcal{D} : \mathbb{Y} \rightarrow \mathbb{Y}$  in (1.6) is chosen such that

$$\mathcal{D} \succeq -\frac{1}{2}\widehat{\Sigma}^f, \quad \mathcal{N}_d \succ 0 \quad \text{and} \quad \frac{1}{2}\widehat{\Sigma}^f + \sigma\mathcal{F}\mathcal{F}^* + \mathcal{D} + \mathcal{N}_{\text{sGS}} \succ 0. \quad (4.22)$$

Under this assumption, we obtain the following proposition on the properties of  $\mathcal{T}$  which will be used in the convergence analysis for Algorithm **sGS-iPADMM**.

**Proposition 4.3** Suppose that Assumption 4.3 holds. Then, the linear operators  $\mathcal{T}$ ,  $\mathcal{B}$  and  $\mathcal{P}$  defined in (4.10) and (4.19) satisfy

$$\mathcal{T} \succeq -\frac{1}{2}\mathcal{P} \quad \text{and} \quad \langle v, (\frac{1}{2}\mathcal{P} + \sigma\mathcal{B}\mathcal{B}^* + \mathcal{T})v \rangle > 0, \quad \forall v \in \mathbb{V} \setminus \{0\}. \quad (4.23)$$

*Proof* Since  $\mathcal{D} \succeq -\frac{1}{2}\widehat{\Sigma}^f$  and  $\mathcal{N}_d \succ 0$ , we know that  $\mathcal{N}_{\text{sGS}} \succeq 0$  and thus

$$\mathcal{D} + \mathcal{N}_{\text{sGS}} + \sigma\mathcal{F}\mathcal{G}^*[\mathcal{G}\mathcal{G}^*]^\dagger\mathcal{G}\mathcal{F}^* \succeq -\frac{1}{2}\widehat{\Sigma}^f.$$

Thus, one can readily see from (4.10) and (4.19) that  $\mathcal{T} \succeq -\frac{1}{2}\mathcal{P}$ . Meanwhile, one can symbolically do the decomposition that

$$\frac{1}{2}\mathcal{P} + \sigma\mathcal{B}\mathcal{B}^* + \mathcal{T} = \begin{pmatrix} \frac{1}{2}\widehat{\Sigma}^f + \sigma\mathcal{F}\mathcal{F}^* + \mathcal{D} + \mathcal{N}_{\text{sGS}} + \sigma\mathcal{F}\mathcal{G}^*[\mathcal{G}\mathcal{G}^*]^\dagger\mathcal{G}\mathcal{F}^* & \sigma\mathcal{F}\mathcal{G}^* \\ \sigma\mathcal{G}\mathcal{F}^* & \sigma\mathcal{G}\mathcal{G}^* \end{pmatrix}.$$

From Lemma 4.2, we know that  $\mathcal{G}\mathcal{G}^*$  is nonsingular on the  $\text{Range}(\mathcal{G})$ . Therefore, by using the definition of  $\mathbb{V}$  and the Schur complement condition for ensuring the positive definiteness of a linear operator, we only need that  $\frac{1}{2}\widehat{\Sigma}^f + \sigma\mathcal{F}\mathcal{F}^* + \mathcal{D} + \mathcal{N}_{\text{sGS}} \succ 0$  on  $\mathbb{Y}$ , and this is guaranteed by (4.22) in Assumption 4.3. This completes the proof of the proposition.  $\square$

**Remark 4.4** We shall discuss here the Assumption 4.3. Since  $\mathcal{D} \succeq -\frac{1}{2}\widehat{\Sigma}^f$  and  $\mathcal{N}_d \succ 0$ , we know that  $\mathcal{N} \succeq 0$  and  $\mathcal{N}_{\text{sGS}} \succeq 0$ . Obviously, one has that

$$\frac{1}{2}\widehat{\Sigma}^f + \sigma\mathcal{F}\mathcal{F}^* + \mathcal{D} + \mathcal{N}_{\text{sGS}} \succeq 0.$$

Meanwhile, by Lemma 4.1, it holds that  $\mathcal{N} + \mathcal{N}_{\text{sGS}} = \widehat{\Sigma}^f + \sigma\mathcal{F}\mathcal{F}^* + \mathcal{D} + \mathcal{N}_{\text{sGS}} \succ 0$ , which, together with  $\mathcal{D} \succeq -\frac{1}{2}\widehat{\Sigma}^f$ , further implies that for any  $\mu > 0$

$$(\frac{1}{2} + \mu)\widehat{\Sigma}^f + \sigma\mathcal{F}\mathcal{F}^* + \mathcal{D} + \mathcal{N}_{\text{sGS}} \succ 0.$$

Hence, under the condition  $\mathcal{D} \succeq -\frac{1}{2}\widehat{\Sigma}^f$ , Assumption 4.3 is only slightly stronger than Assumption 4.2. Next, we provide an simple illustration about the construction of  $\mathcal{D}$  such that Assumption 4.3 holds. We choose  $\mathcal{D}$  as a self-adjoint block diagonal linear operator (particularly for the case that  $\widehat{\Sigma}^f$  is block diagonal)  $\mathcal{D} = \text{Diag}(\mathcal{D}_1, \dots, \mathcal{D}_s)$  such that

$$\mathcal{D} \succeq -\frac{1}{2}\widehat{\Sigma}^f \quad \text{and} \quad \frac{1}{2}\widehat{\Sigma}_{ii}^f + \sigma\mathcal{F}_i\mathcal{F}_i^* + \mathcal{D}_i \succ 0. \quad (4.24)$$

Clearly, condition (4.24) implies that  $\mathcal{N}_d \succ 0$ . Next, we show  $\frac{1}{2}\widehat{\Sigma}^f + \sigma\mathcal{F}\mathcal{F}^* + \mathcal{D} + \mathcal{N}_{\text{sGS}} \succ 0$ . Indeed, if the linear operator  $\frac{1}{2}\widehat{\Sigma}^f + \sigma\mathcal{F}\mathcal{F}^* + \mathcal{D} + \mathcal{N}_{\text{sGS}}$  fails to be positive definite, there exists a nonzero vector  $y \in \mathbb{Y}$  such that

$$\langle y, (\frac{1}{2}\widehat{\Sigma}^f + \sigma\mathcal{F}\mathcal{F}^* + \mathcal{D} + \mathcal{N}_{\text{sGS}})y \rangle = \langle y, (\frac{1}{2}\widehat{\Sigma}^f + \mathcal{D} + \sigma\mathcal{F}\mathcal{F}^*)y \rangle + \langle y, \mathcal{N}_{\text{sGS}}y \rangle = 0.$$

Since  $\frac{1}{2}\widehat{\Sigma}^f + \mathcal{D} + \sigma\mathcal{F}\mathcal{F}^* \succeq 0$  and  $\mathcal{N}_{\text{sGS}} \succeq 0$ , it holds that

$$\langle y, \left(\frac{1}{2}\widehat{\Sigma}^f + \mathcal{D} + \sigma\mathcal{F}\mathcal{F}^*\right)y \rangle = 0 = \langle y, \mathcal{N}_{\text{sGS}}y \rangle.$$

Then, by using (4.7) we can get that  $\mathcal{N}_u^*y = 0$ . This, together with (4.5), implies that

$$\begin{aligned} 0 &= \langle y, \left(\frac{1}{2}\widehat{\Sigma}^f + \mathcal{D} + \sigma\mathcal{F}\mathcal{F}^*\right)y \rangle \\ &= \frac{1}{2} \langle y, \left(\widehat{\Sigma}^f + \sigma\mathcal{F}\mathcal{F}^*\right)y \rangle + \langle y, \left(\frac{1}{2}\sigma\mathcal{F}\mathcal{F}^* + \mathcal{D}\right)y \rangle \\ &= \frac{1}{2} \langle y, \left(\widehat{\Sigma}^f + \sigma\mathcal{F}\mathcal{F}^*\right)_d y \rangle + \langle y, \left(\frac{1}{2}\sigma\mathcal{F}\mathcal{F}^* + \mathcal{D}\right)y \rangle \\ &= \langle y, \left(\frac{1}{2}\widehat{\Sigma}^f\right)_d y \rangle + \frac{\sigma}{2} \langle y, (\mathcal{F}\mathcal{F}^*)_d y \rangle + \frac{\sigma}{2} \langle y, \mathcal{F}\mathcal{F}^*y \rangle. \end{aligned} \quad (4.25)$$

Since  $\mathcal{D} \succeq -\frac{1}{2}\widehat{\Sigma}^f$  implies  $\frac{1}{2}(\widehat{\Sigma}^f)_d + \mathcal{D} \succeq 0$ , we obtain from (4.25) that

$$\langle y, \left(\frac{1}{2}\widehat{\Sigma}^f\right)_d y \rangle = \frac{\sigma}{2} \langle y, (\mathcal{F}\mathcal{F}^*)_d y \rangle = \frac{\sigma}{2} \langle y, \mathcal{F}\mathcal{F}^*y \rangle = 0,$$

which contradicts the fact that  $\mathcal{N}_d > 0$ .

By combining the results of Theorems 3.1 and 4.1 and Proposition 4.3, one can readily get the following convergence theorem of Algorithm sGS-iPADMM.

**Theorem 4.2** Suppose that Assumptions 4.1 and 4.3 hold. Let  $\{(x^k, y^k, z^k)\}$  be the sequence generated by Algorithm sGS-iPADMM. Then,

- (a) the sequence  $\{(y^k, \Pi_{\text{Range}(\mathcal{G})}(z^k))\}$  converges to a solution to problem (1.1) and the sequence  $\{x^k\}$  converges to a solution to the dual of (1.1);
- (b) any accumulation point of the sequence  $\{(y^k, z^k)\}$  is a solution to problem (1.1);
- (c) the sequence  $\{p(y_1^k) + f(y^k) - \langle b, z^k \rangle\}$  of the objective values converges to the optimal value of problem (1.1), and

$$\lim_{k \rightarrow \infty} (\mathcal{F}^*y^k + \mathcal{G}^*z^k - c) = 0;$$

- (d) it holds with  $\mathbf{K}$  being the solution set to the KKT system (4.2) that

$$\lim_{k \rightarrow \infty} \text{dist}((x^k, y^k, z^k), \mathbf{K}) = 0;$$

- (e) if the linear operator  $\mathcal{G}$  is surjective, the whole sequence  $\{(x^k, y^k, z^k)\}$  converges to a solution to the KKT system (4.2) of problem (1.1).

*Proof* (a) Note that the sequence  $\{y^k = (y^k; \Pi_{\text{Range}(\mathcal{G})}(z^k))\}$  defined in (4.21) lies in  $\mathbb{Y} \times \text{Range}(\mathcal{G})$ . By using Theorem 4.1, one can treat the sequence  $\{(x^k, y^k)\}$  generated by Algorithm sGS-iPADMM as the one generated by Algorithm iPALM with the given initial point  $(x^0, y^0)$ . In addition, (4.23) in Proposition 4.3 guarantees that condition (3.5) in Assumption 3.1 holds. Thus, by Theorem 3.1, the sequence  $\{(x^k, y^k)\}$  converges to a solution to the KKT system (4.2), i.e., the sequences  $\{(y^k, \Pi_{\text{Range}(\mathcal{G})}(z^k))\}$  and  $\{x^k\}$  converge to a solution to problem (1.1) and its dual, respectively.

(b) From (a), we see that  $\lim_{k \rightarrow \infty} (x^k, y^k, \Pi_{\text{Range}(\mathcal{G})}(z^k)) = (x^*, y^*, z^*)$  which is a solution to the KKT system (4.2). Since  $\mathcal{G}^*z^k = \mathcal{G}^*\Pi_{\text{Range}(\mathcal{G})}(z^k)$ ,  $\forall k \geq 1$ , any accumulation point, say  $z^\infty$  of  $\{z^k\}$  satisfies  $\mathcal{G}^*z^\infty = \mathcal{G}^*z^*$ . Then, it is easy to verify that  $(x^*, y^*, z^\infty)$  also satisfy the KKT system (4.2). Therefore,  $(y^*, z^\infty)$  is a solution to problem (1.1).

(c) From (a) and the fact that the objective function of problem (1.1) is continuous on its domain, we know that  $\{p(y_1^k) + f(y^k) - \langle b, \Pi_{\text{Range}(\mathcal{G})}(z^k) \rangle\}$  converges to the optimal value of problem (1.1). Since  $b \in \text{Range}(\mathcal{G})$ , it holds that for any  $k \geq 1$ ,  $\langle b, z^k \rangle = \langle b, \Pi_{\text{Range}(\mathcal{G})}(z^k) \rangle$ . Thus,

$$p(y_1^k) + f(y^k) - \langle b, \Pi_{\text{Range}(\mathcal{G})}(z^k) \rangle = p(y_1^k) + f(y^k) - \langle b, z^k \rangle, \quad \forall k \geq 1.$$

Therefore, the sequence  $\{p(y_1^k) + f(y^k) - \langle b, z^k \rangle\}$  converges to the optimal value of problem (1.1). Meanwhile, since  $\mathcal{G}^*z^k = \mathcal{G}^*\Pi_{\text{Range}(\mathcal{G})}(z^k)$ , we further have that

$$\lim_{k \rightarrow \infty} (\mathcal{F}^*y^k + \mathcal{G}^*z^k - c) = \lim_{k \rightarrow \infty} (\mathcal{F}^*y^k + \mathcal{G}^*\Pi_{\text{Range}(\mathcal{G})}(z^k) - c) = 0.$$

(d) From (a), we have that  $(x^*, y^*, z^*)$ , the limit point of  $\{(x^k, y^k, \Pi_{\text{Range}(\mathcal{G})}(z^k))\}$ , is a solution to the KKT system (4.2), i.e.,  $(x^*, y^*, z^*) \in \mathbf{K}$ . Since  $\mathcal{G}^*(z^k - \Pi_{\text{Range}(\mathcal{G})}(z^k)) = 0$  for any  $k \geq 1$ , it is not difficult to see that

$$(x^*, y^*, z^* + (z^k - \Pi_{\text{Range}(\mathcal{G})}(z^k))) \in \mathbf{K}, \quad \forall k \geq 1.$$

Therefore, it holds for all  $k \geq 1$

$$\text{dist}((x^k, y^k, z^k), \mathbf{K}) \leq \|x^k - x^*\| + \|y^k - y^*\| + \|\Pi_{\text{Range}(\mathcal{G})}(z^k) - z^*\|$$

and  $\lim_{k \rightarrow \infty} \text{dist}((x^k, y^k, z^k), \mathbf{K}) = 0$ .

(e) In this case, it holds that  $\text{Range}(\mathcal{G}) = \mathbb{Z}$  and  $z^k = \Pi_{\text{Range}(\mathcal{G})}(z^k)$ ,  $\forall k \geq 0$ . The result follows from (a), which completes the proof of the theorem.  $\square$

We make the following remark on Theorem 4.2.

*Remark 4.5* Without any additional assumptions on  $\mathcal{G}$ , one can observe that the solution set of problem (1.1) is unbounded and the sequence  $\{z^k\}$  generated by Algorithm **sGS-iPADMM** may also be unbounded. Fortunately, we are still able to show in Theorem 4.2(a) and (c) that the sequence  $\{(x^k, y^k, \Pi_{\text{Range}(\mathcal{G})}(z^k))\}$  converges to a solution to the KKT system (4.2), and both the objective and the feasibility converge to the optimal value and zero, respectively. Meanwhile, we would like to emphasize that the surjectivity assumption on  $\mathcal{G}$  in Theorem 4.2(e) is not restrictive at all. Indeed, this assumption simply means that there are no redundant equations in the linear constraints  $\mathcal{G}x = b$  in the primal problem (1.2). If necessary, well established numerical linear algebra techniques can be used to remove redundant equations from  $\mathcal{G}x = b$ .

#### 4.1 The Two-Block Case

Consider the two-block case that  $\mathbb{Y} = \mathbb{Y}_1$  and  $f$  is vacuous, i.e., the following problem

$$\min_{y, z} \{p(y) - \langle b, z \rangle \mid \mathcal{F}^*y + \mathcal{G}^*z = c\}. \quad (4.26)$$

Assume that the KKT system of problem (4.26) admits a nonempty solution set  $\mathbf{K}$ . For such a two-block problem, Algorithm **sGS-iPADMM** without the proximal terms and the inexact computations reduces to the classic ADMM. Then, by Theorem 4.2, the sequence  $\{(x^k, y^k, \Pi_{\text{Range}(\mathcal{G})}(z^k))\}$  generated by the classic ADMM or its inexact variants with  $\tau \in (0, 2)$  (in the order that the  $y$ -subproblem is solved before the  $z$ -subproblem) converges to a point in  $\mathbf{K}$  if either  $\mathcal{F}$  is surjective or  $p$  is strongly convex. Moreover, if  $\mathcal{G}$  is also surjective, we have that the sequence  $\{(x^k, y^k, z^k)\}$  converges to a point in  $\mathbf{K}$ . Note that the assumptions we made for problem (4.26) are apparently weaker than those in [18], where  $\mathcal{F}$  is assumed to be the identity operator,  $\mathcal{G}$  is surjective, and  $p$  is assumed to be strongly convex. Moreover, in [18, Theorem 3.1], only the convergence of the primal sequence  $\{(y^k, z^k)\}$  and the boundedness of the dual sequence  $\{x^k\}$  were obtained.

The detailed comparison between the results in this paper and those in [18] is presented in Table 1. As can be observed from this table, the convergence result on the dual sequence  $\{x^k\}$  is easier to be derived than that of the primal sequence  $\{(y^k, z^k)\}$ , and this result is consistent with the results in [5] for the classic ADMM and the ALM in [45]. Hence, the results derived in this paper properly resolves the questions we have mentioned in the introduction.

At last, we should mention that, in Sun et al. [48, Theorem 3.3 (iv)], a similar result to ours has been derived with the requirements that the initial multiplier  $x^0$  satisfies  $\mathcal{G}x^0 - b = 0$  and all the subproblems are solved exactly. Here, we are able to relax these requirements to the most general case and extend our results to the more interesting and challenging multi-block problems.

#### 4.2 Linear Rate of Convergence and Iteration Complexity

Theorem 3.2 has provided a tool, which can be used together with Theorem 4.1 to analyze the linear convergence rate of the sequence generated by Algorithm **sGS-iPADMM**, i.e., one only need to verify whether (3.30) is valid for this sequence, provided that the metric subregular property (3.26) holds. However, such a verification is not as straightforward as it conceptually seems.

**Table 1** Comparison between [18] and this paper. In the table ‘**SOL**’ denotes the solution set to problem (4.26), ‘**X**’ denotes the set of multipliers (the solution set to the dual problem) to problem (4.26), and ‘**K**’ denotes the solution set to the KKT system or problem (4.26), i.e.,  $\mathbf{K} = \mathbf{X} \times \mathbf{SOL}$ . The symbol  $\rightarrow$  means that the sequence on its left-hand-side is convergent, and converges to a point in its right-hand-side.

Item \ Ref	[18]	This paper	
Updating rules	$z \Rightarrow y \Rightarrow x \quad \& \quad \tau \in (0, 2)$	$y \Rightarrow z \Rightarrow x \quad \& \quad \tau \in (0, 2)$	
Assumptions- $y$	$p$ strongly convex and $\mathcal{F}$ the identity operator	$p$ strongly convex or $\mathcal{F}$ surjective	$p$ strongly convex or $\mathcal{F}$ surjective
Assumptions- $z$	$\mathcal{G}$ surjective	-	$\mathcal{G}$ surjective
Sequences	$\{(y^k, z^k)\} \rightarrow \mathbf{SOL}$ $\{x^k\}$ bounded	$\text{dist}((x^k, y^k, z^k), \mathbf{K}) \rightarrow 0$ $\{x^k\} \rightarrow \mathbf{X}$	$\{(y^k, z^k)\} \rightarrow \mathbf{SOL}$ $\{x^k\} \rightarrow \mathbf{X}$

Here, we establish a linear convergence result for the case that the linear system in step 3 of Algorithm **sGS-iPADMM** is solve exactly, but leave the general cases as a topic for further study. For this purpose, we view problem (1.1) as an instance of (3.1) with

$$\begin{cases} \varphi(w) := p(y_1), \\ h(w) := f(y) - \langle b, z \rangle, \quad \forall w = (y, z) \in \mathbb{W} := \mathbb{Y} \times \mathbb{Z}. \\ \mathcal{A}^*w := \mathcal{F}^*y + \mathcal{G}^*z, \end{cases} \quad (4.27)$$

Then, the corresponding KKT residual mapping of problem (1.1) can be given by (3.22). Moreover, the self-adjoint linear operator  $\Omega$  defined in (3.16) is given by  $\Omega(x; (y; z)) = (x; \Theta^{\frac{1}{2}}(y; z))$ , where  $\Theta = \tau\sigma(\mathcal{P} + \mathcal{T} + \frac{(2-\tau)\sigma}{3}\mathcal{A}\mathcal{A}^*)$  with  $\mathcal{T}$  and  $\mathcal{P}$  being defined in (4.10) and (4.19), respectively. In fact, we further have that

$$\Omega(x; (y; z)) = \Omega(x; (y; \Pi_{\text{Range}(\mathcal{G})}(z))), \quad \forall (x, y, z) \in \mathbb{X} \times \mathbb{Y} \times \mathbb{Z}. \quad (4.28)$$

**Theorem 4.3** Suppose that Assumptions 4.1 and 4.3 hold. Let  $\{u^k = (x^k, w^k)\}$  with  $w^k := (y^k; z^k)$  be the sequence generated by Algorithm **sGS-iPADMM** such that  $\{v^k := (x^k, y^k, \Pi_{\text{Range}(\mathcal{G})}(z^k))\}$  converges to  $v^* \in \mathbf{K}$ . It holds that

$$\text{dist}_{\Omega}(u^k, \mathbf{K}) = \text{dist}_{\Omega}(v^k, \mathbf{K}), \quad \forall k \geq 0. \quad (4.29)$$

Suppose that  $b - \mathcal{G}x^0 = 0$  and  $\gamma^k = 0$  for all  $k \geq 0$ . Suppose that the KKT residual mapping  $\mathcal{R}$  defined in (3.22) (with the notation in (4.27)) is metric subregular at  $v^*$  for  $0 \in \mathbb{U}$  with the modulus  $\kappa > 0$ , in the sense that there exists a constant  $r > 0$  such that (3.26) holds with  $\bar{u} = v^*$ . Let  $\{\tilde{\eta}_k\}$  be a given sequence of nonnegative numbers that converges to 0 in the limit. Suppose that in addition to satisfying  $\max\{\|\tilde{\delta}_i^k\|, \|\delta_i^k\| \mid i = 1, \dots, s\} \leq \tilde{\epsilon}_k$ , there exists an integer  $k_0 > 0$  such that for any  $k \geq k_0$ , it holds that

$$\max_{1 \leq i \leq s} \{\|\tilde{\delta}_i^k\|, \|\delta_i^k\|\} \leq \tilde{\eta}_k \|v^k - v^{k+1}\|. \quad (4.30)$$

Then, for all  $k$  sufficiently large, it holds that  $\text{dist}_{\Omega}(u^{k+1}, \mathbf{K}) \leq \vartheta^k \text{dist}_{\Omega}(u^k, \mathbf{K})$  with  $\sup_{k \geq k_0} \{\vartheta_k\} < 1$ , i.e., the convergence rate of  $\text{dist}_{\Omega}(u^k, \mathbf{K})$  is  $Q$ -linear when  $k$  is sufficiently large.

*Proof* By (4.28), we have that for all  $k \geq 0$ ,

$$\begin{aligned} \text{dist}_{\Omega}^2(u^k, \mathbf{K}) &= \inf_{u \in \mathbf{K}} \frac{1}{2} \langle u^k - u, \Omega(u^k - u) \rangle = \inf_{u \in \mathbf{K}} \frac{1}{2} \langle u^k - u, \Omega(v^k) - \Omega(u) \rangle \\ &= \inf_{u \in \mathbf{K}} \frac{1}{2} \langle v^k - u, \Omega(v^k - u) \rangle = \text{dist}_{\Omega}^2(v^k, \mathbf{K}), \end{aligned}$$

i.e., (4.29) holds. Since  $b - \mathcal{G}x^0 = 0$  and  $\gamma^k = 0$  for all  $k \geq 0$ , according to (4.18) one has that

$$\|b - \mathcal{G}x^k\| \leq |1 - \tau|^k \|b - \mathcal{G}x^0\| + \tau \sum_{i=1}^k |1 - \tau|^{k-i} \|\gamma^{i-1}\| = 0, \quad \forall k \geq 0.$$

Therefore, by (4.6) and (4.11) one knows that

$$\Delta^k := \delta_{\text{sGS}}^k - \mathcal{F}\mathcal{G}^*(\mathcal{G}\mathcal{G}^*)^{-1}\mathcal{G}(x^k - x^{k-1}) = \delta^k + \mathcal{N}_u\mathcal{N}_d^{-1}(\delta^k - \tilde{\delta}^k).$$

Thus, we can get that for all  $k \geq 0$ ,

$$\begin{aligned} \|d^k\| &= \|\Delta^k\| \leq (1 + 2\|\mathcal{N}_u\mathcal{N}_d^{-1}\|) \max\{\|\tilde{\delta}\|, \|\tilde{\delta}\|\} \\ &\leq \sqrt{s}(1 + 2\|\mathcal{N}_u\mathcal{N}_d^{-1}\|) \tilde{\eta}_k \|v^k - v^{k+1}\|, \end{aligned}$$

where  $d^k := (\widehat{\Delta}^k; \gamma^k) \in \mathbb{W}$ . Define  $\eta_k = \sqrt{s}(1 + 2\|\mathcal{N}_u\mathcal{N}_d^{-1}\|)\tilde{\eta}_k$ . Then, it holds that  $\eta_k \rightarrow 0$  and  $\|d^k\| \leq \eta_k \|v^k - v^{k+1}\|$ . Therefore, by Theorem 3.2, we know that for all  $k$  sufficiently large

$$\text{dist}_\Omega(v^k, \mathbf{K}) \leq \vartheta^k \text{dist}_\Omega(v^k, \mathbf{K})$$

with  $\sup_{k \geq k_0} \{\vartheta_k\} < 1$ , which, together with (4.29), implies

$$\text{dist}_\Omega(u^k, \mathbf{K}) \leq \vartheta^k \text{dist}_\Omega(u^k, \mathbf{K}) \quad \text{for all } k \text{ sufficiently large.}$$

This completes the proof.  $\square$

*Remark 4.6* Note that, different from the condition (3.30) in Assumption 3.2, the condition (4.30) here is generally not directly verifiable during the numerical implementation. However, Theorem 4.3 does provide us a very important theoretical guideline on implementing Algorithm **sGS-iPADMM**, i.e., in the  $k$ -th iteration, it is likely to be beneficial to solve the subproblems to an accuracy higher than the dual feasibility  $\|\mathcal{F}^*y^k + \mathcal{G}^*z^k - c\|$ . In fact, this phenomenon has already been observed during our numerical experiments. We should also mention that even for the 2-block case, the study on the linear convergence of inexact ADMMs with shorter step-length  $\tau \in (0, \frac{1+\sqrt{5}}{2})$  is still not as mature as the study for their exact counterparts, especially when compared with the recently developed results, e.g., in [22, 57]. Suitable criteria that generalize the condition (3.30) for terminating the subproblems are still lacking. We note that the results presented in Theorem 4.3 are still far from complete, and more effort should be spent on this part in the future.

Finally, different from the above discussions on the convergence rate, we can establish the following non-ergodic iteration complexity for the sequence generated by Algorithm **sGS-iPADMM** by a direct application of Theorem 4.1.

**Theorem 4.4** Suppose that Assumptions 4.3 and 4.1 hold. Let  $\{u^k = (x^k, w^k)\}$  with  $w^k := (y^k; z^k)$  be the sequence generated by Algorithm **sGS-iPADMM** such that  $\{v^k := (x^k, y^k, \Pi_{\text{Range}(\mathcal{G})}(z^k))\}$  converges to  $v^* \in \mathbf{K}$ . It holds that the KKT residual (3.22), with  $\mathcal{B}$  and  $\mathcal{P}$  given by (4.19), satisfies

$$\min_{0 \leq j \leq k} \|\mathcal{R}(u^j)\|^2 \leq \varrho/k, \quad \text{and} \quad \lim_{k \rightarrow \infty} \{k \cdot \min_{0 \leq j \leq k} \|\mathcal{R}(u^j)\|^2\} = 0,$$

where the constant  $\varrho$  is defined as in (3.43) but with

$$\mathbf{e} := \|u_e^0\|_\Omega^2 + 2\tau\sigma\|\Theta^{-1/2}\| \left( \sum_{j=0}^\infty \tilde{\varepsilon}_j \right) \left( \|u_e^0\|_\Omega + \mu \sum_{j=0}^\infty \tilde{\varepsilon}_j \right) + 4 \left( \sum_{j=1}^\infty \tilde{\varepsilon}_j^2 \right)$$

and  $u_e^0 = u^0 - v^*$ .

*Proof* From (4.28), we know that

$$\|u^0 - v^*\|_\Omega^2 = \langle \Omega(u^0 - v^*), u^0 - v^* \rangle = \langle \Omega(v^0) - \Omega(v^*), u^0 - v^* \rangle = \|v^0 - v^*\|_\Omega^2. \quad (4.31)$$

According to (3.22), (4.2) and (4.27), one has that

$$\mathcal{R}(u) = \begin{pmatrix} c - \mathcal{F}^*y + \mathcal{G}^*z \\ y - \text{Prox}_P(y - \nabla f(y) - \mathcal{F}x) \\ \mathcal{G}x - b \end{pmatrix}, \quad \forall u = (x, y, z) \in \mathbb{X} \times \mathbb{Y} \times \mathbb{Z}.$$

Since for all  $k \geq 0$ ,  $\mathcal{G}^*z^k = \mathcal{G}\Pi_{\text{Range}(\mathcal{G}^*)}(z^k)$ , one has that  $\mathcal{R}(u^k) = \mathcal{R}(v^k)$ . Therefore, by using (4.23) in Theorem 4.1(a), Theorem 3.3 and (4.31), one has the results of this theorem holds. This completes the proof.  $\square$



## 5 Numerical Experiments

In this section, we conduct numerical experiments on solving dual linear SDP and dual convex quadratic SDP problems via Algorithm **sGS-iPADMM** with the dual step-length  $\tau$  taking values beyond the standard restriction of  $(1 + \sqrt{5})/2$ . For linear SDP problems, the algorithm reduces to the two-block ADMM, and the aim is two-fold. On the one hand, as the ADMM is among the most important first-order algorithms for solving SDP problems, it is of importance to know to what extent can the numerical efficiency be improved if the observation on the dual step-length made in this paper is incorporated. On the other hand, as the upper bound of the step-length has been enlarged, it is also important to see whether a step-length that is very close to the upper bound will lead to better or worse numerical performance.

A standard linear SDP problem has the following form:

$$\min_X \{ \langle C, X \rangle \mid \mathcal{A}X = \mathbf{b}, X \in \mathbb{S}_+^n \} \quad (5.1)$$

and its corresponding dual is given as in (2.4). To avoid repetition, we refer the reader to (2.4) for the notation used. The (majorized) augmented Lagrangian function associated with problem (2.4) is given by

$$\begin{aligned} \mathcal{L}_\sigma(S, \mathbf{z}; X) &= \delta_{\mathbb{S}_+^n}(S) - \langle \mathbf{b}, \mathbf{z} \rangle + \langle X, S + \mathcal{A}^* \mathbf{z} - C \rangle + \frac{\sigma}{2} \|S + \mathcal{A}^* \mathbf{z} - C\|^2, \\ &\quad \forall (S, \mathbf{z}, X) \in \mathbb{S}^n \times \mathbb{R}^m \times \mathbb{S}^n, \end{aligned}$$

where  $\sigma > 0$  is the given penalty parameter. When applied to solving problem (2.4), at the  $k$ -th step of the two-block ADMM the following steps are performed:

$$\begin{cases} S^{k+1} = \Pi_{\mathbb{S}_+^n}(C - \mathcal{A}^* \mathbf{z}^k - X^k/\sigma), \\ \mathbf{z}^{k+1} = (\mathcal{A}\mathcal{A}^*)^{-1}(\mathcal{A}(C - S^{k+1}) - (\mathcal{A}X^k - \mathbf{b})/\sigma), \\ X^{k+1} = X^k + \tau\sigma(S^{k+1} + \mathcal{A}^* \mathbf{z}^{k+1} - C), \end{cases}$$

where the step-length  $\tau$  is allowed to be in the range  $(0, 2)$  based on Theorem 4.1 and the discussions in Section 4.1. We emphasize again that this is in contrast to the usual interval of  $(0, (1 + \sqrt{5})/2)$  allowed by the convergence analysis of Glowinski in [20, Theorem 5.1].

On the other hand, as was briefly introduced in Section 2.1, the convex QSDP problem was formally given in (2.1), whose dual problem, in minimization form, is a multi-block problem given by

$$\begin{aligned} \min_{S, W, \mathbf{z}_E, \mathbf{z}_I} \quad & \delta_{\mathbb{S}_+^n}(S) + \delta_{\mathbb{R}_+^{m_I}}(\mathbf{z}_I) + \frac{1}{2} \langle W, \mathcal{Q}W \rangle - \langle \mathbf{b}_E, \mathbf{z}_E \rangle - \langle \mathbf{b}_I, \mathbf{z}_I \rangle \\ \text{s.t.} \quad & S - \mathcal{Q}W + \mathcal{A}_E^* \mathbf{z}_E + \mathcal{A}_I^* \mathbf{z}_I + C = 0. \end{aligned} \quad (5.2)$$

Note that problem (2.1) was subsumed as an instance of the convex quadratic composition optimization problem (1.8). Therefore, to fit the framework of Algorithm **sGS-iPADMM**, we write the dual of problem (2.1) in the minimization form as

$$\begin{aligned} \min_{S, W, \mathbf{z}_E, \mathbf{z}_I} \quad & \delta_{\mathbb{S}_+^n}(S) + \delta_{\mathbb{R}_+^{m_I}}(\mathbf{z}_I) + \frac{1}{2} \langle W, \mathcal{Q}W \rangle - \langle \mathbf{b}_E, \mathbf{z}_E \rangle - \langle \mathbf{b}_I, \mathbf{z}_I \rangle \\ \text{s.t.} \quad & \begin{cases} S - \mathcal{Q}W + \mathcal{A}_E^* \mathbf{z}_E + \mathcal{A}_I^* \mathbf{z}_I + C = 0, \\ D(\mathbf{s} - \mathbf{z}_I) = 0, \end{cases} \end{aligned} \quad (5.3)$$

where  $D \in \mathbb{R}^{m_I \times m_I}$  is a given positive definite diagonal matrix which is incorporated here for the purpose of scaling the variables to ensure the numerical stability.

The convex QSDP problem (2.1) is solved via its dual (5.3), whose (majorized) augmented Lagrangian function is defined by

$$\begin{aligned} \mathcal{L}_\sigma(S, W, \mathbf{z}_E, \mathbf{z}_I, s; X, \mathbf{x}) &:= (\delta_{\mathbb{S}_+^n}(S) + \delta_{\mathbb{R}_+^{m_I}}(\mathbf{s})) + \frac{1}{2} \langle W, \mathcal{Q}W \rangle - \langle \mathbf{b}_E, \mathbf{z}_E \rangle - \langle \mathbf{b}_I, \mathbf{z}_I \rangle \\ &\quad + \langle X, S - \mathcal{Q}W + \mathcal{A}_E^* \mathbf{z}_E + \mathcal{A}_I^* \mathbf{z}_I + C \rangle + \langle D(\mathbf{s} - \mathbf{z}_I), \mathbf{x} \rangle \\ &\quad + \frac{\sigma}{2} \|S - \mathcal{Q}W + \mathcal{A}_E^* \mathbf{z}_E + \mathcal{A}_I^* \mathbf{z}_I + C\|^2 + \frac{\sigma}{2} \|D(\mathbf{s} - \mathbf{z}_I)\|^2, \\ &\quad \forall (S, W, \mathbf{z}_E, \mathbf{z}_I, s; X, \mathbf{x}) \in \mathbb{S}^n \times \mathbb{S}^n \times \mathbb{R}^{m_E} \times \mathbb{R}^{m_I} \times \mathbb{R}^{m_I} \times \mathbb{S}^n \times \mathbb{R}^{m_I}. \end{aligned}$$

where  $\sigma > 0$  is the given penalty parameter and we have used  $X \in \mathbb{S}^n$  and  $\mathbf{x} \in \mathbb{R}^m$  to denote the Lagrange multipliers which are introduced for the two groups of equality constraints in (5.3). During the  $k$ -th iteration of Algorithm **sGS-iPADMM** with given  $(S^k, W^k, \mathbf{z}_E^k, \mathbf{z}_I^k, \mathbf{s}^k)$  and  $(X^k, \mathbf{x}^k)$ , we update the variables in the order of

$$\underbrace{(\mathbf{z}_E^{k+1/2} \Rightarrow W^{k+1/2})}_{\text{backward GS}} \Rightarrow \underbrace{(S^{k+1}, \mathbf{s}^{k+1}) \Rightarrow W^{k+1} \Rightarrow \mathbf{z}_E^{k+1}}_{\text{forward GS}} \Rightarrow \mathbf{z}_I^{k+1} \Rightarrow \underbrace{(X^{k+1}, \mathbf{x}^{k+1})}_{\tau \in (0,2)}.$$

Note that the term  $\langle \mathbf{b}_I, \mathbf{z}_I \rangle$  is treated as the linear term in the framework of (1.1). We made this choice because for the test instances that we will consider later, the linear system that must be solved to update  $\mathbf{z}_I$  is much larger than that for updating  $\mathbf{z}_E$ , and in this way, the larger linear system will be solved only once in each iteration.

The numerical results in the subsequent two subsections are obtained by using MATLAB R2017b on a HP Elitedesk (64-bit Windows 10 system) with one Intel Core i7-4770S Processor (4 Cores, 3.1 – 3.9 GHz) and 16 GB RAM (with the virtual memory turned off).

### 5.1 Numerical Results on Linear SDP Problems

Based on the first-order optimality condition for problem (5.1), we terminate all the tested algorithms if

$$\eta_{\text{SDP}} := \max\{\eta_D, \eta_P, \eta_S\} \leq 10^{-6},$$

where

$$\begin{cases} \eta_D = \frac{\|\mathcal{A}^* \mathbf{z} + S - C\|}{1 + \|C\|}, \quad \eta_P = \frac{\|\mathcal{A}X - \mathbf{b}\|}{1 + \|\mathbf{b}\|}, \\ \eta_S = \max \left\{ \frac{\|X - \Pi_{\mathbb{S}_+^n}(X)\|}{1 + \|X\|}, \frac{|\langle X, S \rangle|}{1 + \|X\| + \|S\|} \right\} \end{cases}$$

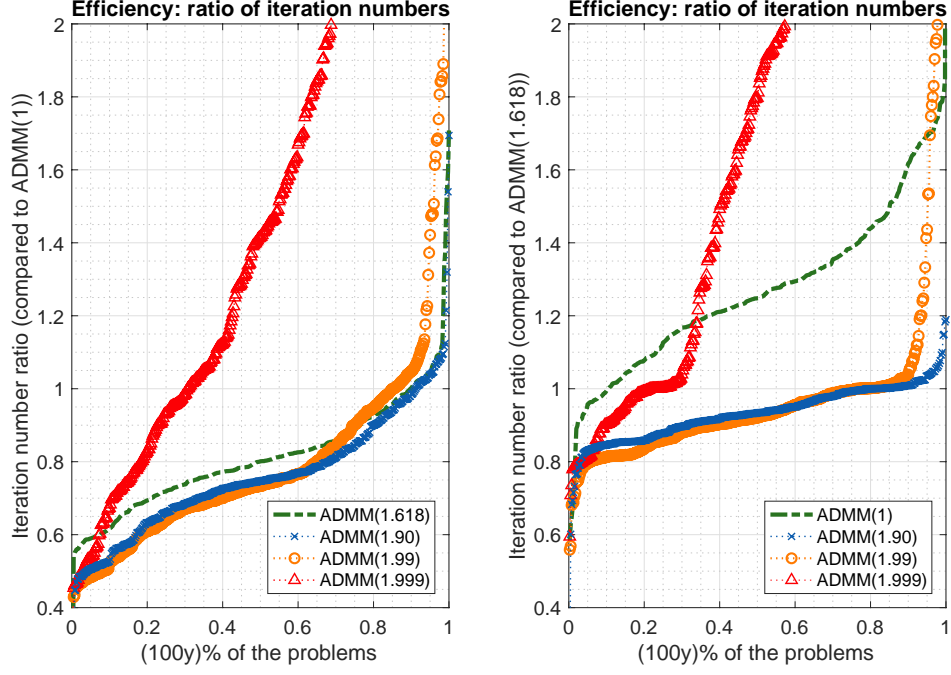
with the maximum number of iterations set at 100,000. In addition, we also measure the duality gap:

$$\eta_{\text{gap}} := \frac{\langle C, X \rangle - \langle \mathbf{b}, \mathbf{z} \rangle}{1 + |\langle C, X \rangle| + |\langle \mathbf{b}, \mathbf{z} \rangle|}.$$

During our preliminary tests, we found that using a step-length smaller than 1 is not as good as using the unit step-length. Therefore, we shall only consider the cases that  $\tau \geq 1$ . Note that the known theoretical upper bound of the step-length  $\tau$  in the classic ADMM for solving general convex programming is  $\frac{1+\sqrt{5}}{2} \approx 1.618034$ . Although it has been observed empirically that the ADMM with the step-length  $\tau = 1.618$  works quite well, this phenomenon still requires further understanding since the value 1.618 is quite close to the theoretical upper bound and such an aggressive choice may result in unstable numerical performance. Fortunately, the above concern is partially alleviated by the theoretical results obtained in this paper. Indeed, for a large class convex optimization problems, one can use  $\tau = 1.618$  confidently since it has a “safe” distance to the renewed theoretical upper bound of 2. For this class of problems, it is thus very interesting to see what would happen if the step-length  $\tau$  is very close to 2. Therefore, we tested five choices of the step-length, i.e.,  $\tau = 1, 1.618, 1.90, 1.99$  and 1.999. For convenience, we use  $\text{ADMM}(\tau)$  to denote the algorithm with the specific step-length  $\tau$ .

We tested 6 categories of linear SDP problems, including the random sparse SDP problems tested in [34], the semidefinite relaxation of frequency assignment problems (FAP) [15], the relaxation of maximum stable set problems [50, 52, 47], the SDP relaxation of binary integer quadratic (BIQ) problems from [54], the SDP relaxation of rank-1 tensor approximations (R1TA) [38, 39], and the SDP relaxations of clustering problems [40]. One may refer to [58, 56] for detailed descriptions and the data sources of these problems. All these algorithms are tested by running the MATLAB package SDPNAL+ (version 1.0, available at <http://www.math.nus.edu.sg/~mattokhc/SDPNALplus.html>). The records of the computational results are provided in the first table of the supplementary materials. Here, we should mention that even though all the problems we tested have been successfully solved by at least one of the tested algorithms, there are a few categories of SDP problems that are beyond the capability of the ADMM, see, e.g., [58].

Figure 1 presents the computational performance of the ADMM with all the five choices of step-lengths. The original scalable MATLAB-generated figure with more details is available in the supplementary materials. The left panel shows the comparison between  $\text{ADMM}(1)$  and all the other algorithms, while the right panel shows the comparison between  $\text{ADMM}(1.618)$  and all the others. As can be seen from Figure 1,  $\text{ADMM}(1.618)$  has an impressive improvement over  $\text{ADMM}(1)$  and  $\text{ADMM}(1.9)$  works even better than  $\text{ADMM}(1.618)$  for more than 80% of the tested instances. Furthermore,  $\text{ADMM}(1.99)$  can perform marginally better than  $\text{ADMM}(1.9)$



**Fig. 1** Comparison of the computational efficiency of the classic two-block ADMM with different step-lengths

for about 60% of the tested instances but for about 10% of them, its performance is apparently worse. However, ADMM(1.999) has a significantly worse performance than ADMM(1.99) even though its step-length is just slightly larger than 1.99. This can be partially explained by the fact that the step-length of 1.999 is too close to the theoretical upper bound of 2.

From both the theoretical analysis and numerical experiments in this paper, one can see that in general it is a good idea to use a step-length that is larger than 1, e.g.,  $\tau = 1.618$ , when solving linear SDP problems. Meanwhile, we can even set the step-length to be larger than 1.618, say  $\tau = 1.9$ , to get even better numerical performance.

## 5.2 Numerical Results on Convex QSDP Problems

The KKT system of problem (5.2) is given by

$$\begin{cases} S - \mathcal{Q}W + \mathcal{A}_E^* z_E + \mathcal{A}_I^* z_I + C = 0, \mathcal{A}_E X - \mathbf{b}_E = 0, \\ \mathcal{Q}X - \mathcal{Q}W = 0, X \in \mathbb{S}_+^n, S \in \mathbb{S}_+^n, \langle X, S \rangle = 0, \\ \mathcal{A}_I X - \mathbf{b}_I \geq 0, z_I \geq 0, \langle \mathcal{A}_I X - \mathbf{b}_I, z_I \rangle = 0. \end{cases} \quad (5.4)$$

Based on the optimality conditions given in (5.4), we measure the accuracy of an approximate (computed) solution  $(X, Z, W, S, y_E, y_I)$  for the convex QSDP (2.1) and its dual (5.2) via

$$\eta_{\text{qsdP}} = \max \{ \eta_D, \eta_P, \eta_W, \eta_S, \eta_I \}, \quad (5.5)$$

where

$$\begin{cases} \eta_D = \frac{\|S - \mathcal{Q}W + \mathcal{A}_E^* z_E + \mathcal{A}_I^* z_I + C\|}{1 + \|C\|}, \quad \eta_P = \frac{\|\mathcal{A}_E X - \mathbf{b}_E\|}{1 + \|\mathbf{b}_E\|}, \\ \eta_W = \frac{\|\mathcal{Q}X - \mathcal{Q}W\|}{1 + \|\mathcal{Q}\|}, \quad \eta_S = \max \left\{ \frac{\|X - \Pi_{\mathbb{S}_+^n}(X)\|}{1 + \|X\|}, \frac{|\langle X, S \rangle|}{1 + \|X\| + \|S\|} \right\}, \\ \eta_I = \max \left\{ \frac{\|\min(0, z_I)\|}{1 + \|z_I\|}, \frac{\|\min(0, \mathcal{A}_I X - \mathbf{b}_I)\|}{1 + \|\mathbf{b}_I\|}, \frac{|\langle \mathcal{A}_I X - \mathbf{b}_I, z_I \rangle|}{1 + \|\mathcal{A}_I X - \mathbf{b}_I\| + \|z_I\|} \right\}. \end{cases}$$

Additionally, we measure the objective values and the duality gap:

$$\eta_{\text{gap}} := \frac{\text{Obj}_{\text{primal}} - \text{Obj}_{\text{dual}}}{1 + |\text{Obj}_{\text{primal}}| + |\text{Obj}_{\text{dual}}|},$$

**Table 2** 7 types of algorithms tested for the convex QSDP problems. In the table, ‘Dir’ denotes the corresponding subproblems are solved via direct solvers, ‘Proj.’ means that the corresponding subproblems are the calculation of projections, ‘Prox.’ means that the corresponding subproblems are solved via adding appropriate proximal terms to ensure closed-form solutions, ‘Inex.’ means that the subproblems are solved approximately via an iterative scheme, ‘(rep.)’ means that the corresponding subproblems in the forward sGS sweeps are also solved, ‘(check)’ means that the corresponding subproblems in the forward sGS sweeps are not directly solved but the most recently updated variables are checked to see if they are admissible approximate solution to the subproblems.

Algorithm \ Variable	$\tau$	$\mathbf{z}_E$	$W$	$S$	$s$	$\mathbf{z}_I$
Directly Extended	1	Dir.	Prox.	Proj.	Proj.	Prox.
sGS-PADMM	1	Dir. (rep.)	Prox. (rep.)	Proj.	Proj.	Prox.
	1.618	Dir. (rep.)	Prox. (rep.)	Proj.	Proj.	Prox.
	1.9	Dir. (rep.)	Prox. (rep.)	Proj.	Proj.	Prox.
sGS-iPADMM	1	Dir. (check)	Inex. (check)	Proj.	Proj.	Inex.
	1.618	Dir. (check)	Inex. (check)	Proj.	Proj.	Inex.
	1.9	Dir. (check)	Inex. (check)	Proj.	Proj.	Inex.

where

$$\text{Obj}_{\text{primal}} := \frac{1}{2} \langle X, \mathcal{Q}X \rangle - \langle C, X \rangle, \quad \text{and} \quad \text{Obj}_{\text{dual}} := -\frac{1}{2} \langle W, \mathcal{Q}W \rangle + \langle \mathbf{b}_E, \mathbf{z}_E \rangle + \langle \mathbf{b}_I, \mathbf{z}_I \rangle.$$

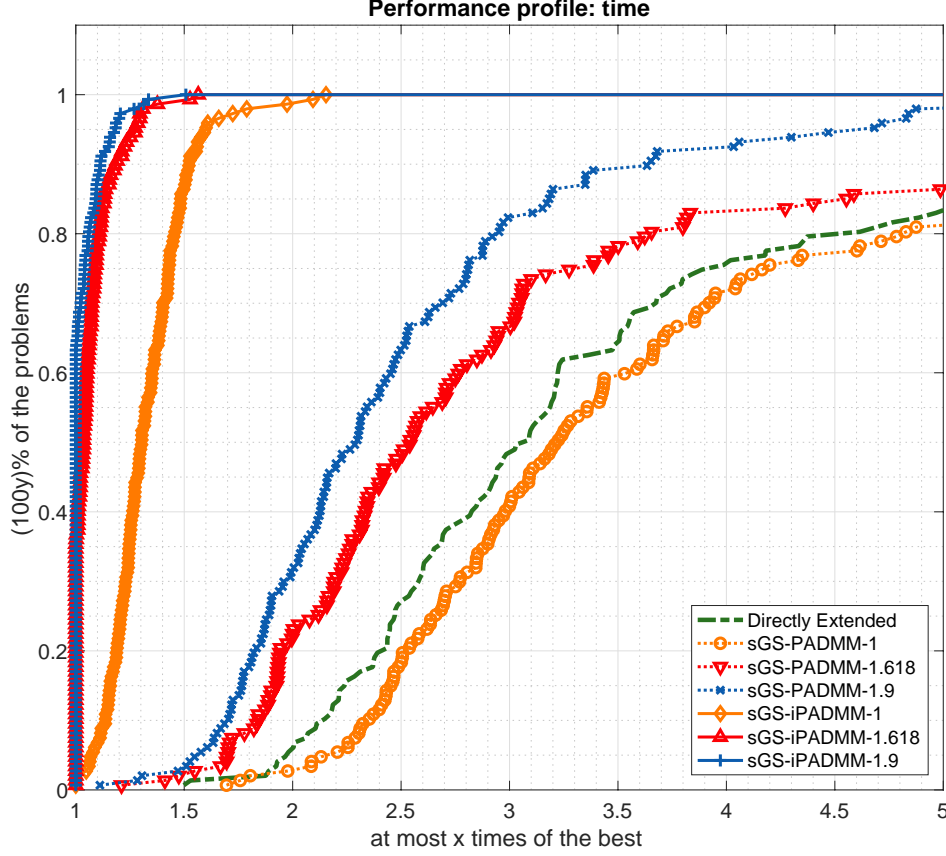
In our numerical experiments, similar to [6], we used QSDP test instances based on the SDP problems arising from the relaxation of the binary integer quadratic (BIQ) programming with a large number of inequality constraints that was introduced by Sun et al. [48] for getting tighter bounds. The problems that we actually solve have the following form:

$$\begin{aligned} \min \quad & \frac{1}{2} \langle X, \mathcal{Q}X \rangle + \frac{1}{2} \langle Q, \bar{X} \rangle + \langle c, x \rangle \\ \text{s.t.} \quad & \begin{cases} \text{diag}(\bar{X}) - x = 0, X = \begin{pmatrix} \bar{X} & x \\ x^T & 1 \end{pmatrix} \in \mathbb{S}_+^n, \\ x_i - \bar{X}_{ij} \geq 0, x_j - \bar{X}_{ij} \geq 0, \bar{X}_{ij} - x_i - x_j \geq -1, \forall 1 \leq i < j \leq n-1. \end{cases} \end{aligned}$$

The data for  $Q$  and  $c$  are taken from the Biq Mac Library maintained by Wiegale [54].

We solve the QSDP (2.1) via its dual (5.3) with the matrix  $D = (\sqrt{\|\mathcal{A}_I\|}/2)I_{\mathbb{R}^{m_I}}$ . We use the directly extended multi-block ADMM with step-length  $\tau = 1$  as the benchmark (which we named as ‘Directly Extended’), and compare it with Algorithm **sGS-iPADMM**, which was implemented in 6 different ways, i.e., 2 groups of algorithms with each using 3 types of different step-lengths, i.e.,  $\tau = 1, 1.618$  and  $1.9$ , which are chosen according to the numerical results in Section 5.1. For convenience, we use the name ‘sGS-PADMM’ to mean that Algorithm **sGS-iPADMM** is implemented such that all the subproblems are solved exactly via direct solvers or adding appropriate proximal terms, and use ‘sGS-iPADMM’ to mean that Algorithm **sGS-iPADMM** is implemented such that the subproblems are allowed to be solved inexactly via iterative solvers. The details of all the seven tested algorithms are presented in Table 2. For all the algorithms applied to problem (5.3), the subproblems corresponding to the block variable  $(S, s)$  can be solved analytically by computing the projections onto  $\mathbb{S}_+^n \times \mathbb{R}_+^{m_I}$ . For the subproblems corresponding to  $\mathbf{z}_E$ , linear systems of equations must be solved with the same coefficient matrix  $\mathcal{A}_E \mathcal{A}_E^*$ . As the linear systems are not too large, we solve them via the Cholesky factorization (computed only once) of  $\mathcal{A}_E \mathcal{A}_E^*$ . For the subproblems corresponding to  $\mathbf{z}_I$  and  $W$ , we need to solve very large scale linear systems of equations, so that they are either solved via a preconditioned conjugate gradient method with preconditioners that are described in [6, Section 7.1] (for sGS-iPADMM), or solved directly by adding an appropriate proximal term to the subproblems to get closed-form solutions (for sGS-PADMM). Moreover, in the implementation of the sGS-PADMM, all the subproblems in the forward Gauss-Seidel sweep are directly solved, while in the implementation of sGS-iPADMM we used the strategy described in [6, Remark 4.1(b)] to decide whether the computation of the subproblems in the forward GS sweep can be skipped (see [6, Section 7.2] for more details on using this technique). We used a similar strategy as described in [27, Section 4.4] to adaptively adjust the penalty parameter  $\sigma$  and used the same technique as in [6] to control the error tolerance for solving the subproblems, i.e.,  $\{\tilde{\epsilon}_k\}_{k \geq 0}$  is chosen such that  $\alpha \tilde{\epsilon}_k \leq 1/k^{1.2}$ , where  $\alpha > 0$  is a positive constant based on the problem data.

We have tested 147 instances of convex QSDP problems with  $n$  ranging from 51 to 501. The linear operator  $\mathcal{Q}$  was chosen as the symmetrized Kronecker operator  $\mathcal{Q}(X) = \frac{1}{2}(AXB + BXA)$  with  $A$  and  $B$  being two randomly



**Fig. 2** Comparison of the computational efficiency of the 7 algorithms

generated symmetric positive semidefinite matrices such that  $\text{rank}(A) = 10$  and  $\text{rank}(B) \approx n/5$ , respectively, as was used in [51, 6]. The maximum iteration number is set at 500,000. The detail computational results are provided in the second table of the supplementary materials.

Figure 2 shows the numerical performance of the 7 tested algorithms described in Table 2 on solving the convex QSDP problems to the accuracy of  $10^{-6}$  in terms of  $\eta_{\text{qsdP}}$  in (5.5). The original scalable MATLAB-generated figure with more details is available in the supplementary materials. One can readily see from the figure that sGS-iPADMM overwhelmingly outperforms sGS-PADMM, no matter which step-length  $\tau$  was used. This evidently shows the considerable advantage of catering for approximate solutions in the subproblems of Algorithm sGS-iPADMM. Moreover, for both sGS-PADMM and sGS-iPADMM, the step-length  $\tau = 1.618$  is able to bring a noticeable improvement on the numerical efficiency, compared with using the unit step-length. Meanwhile, the choice of  $\tau = 1.9$  can perform even better in general. Even this is more apparent for sGS-PADMM, in which all the subproblems are solved exactly. We can see that sGS-iPADMM with  $\tau = 1.9$  performs the best among all the tested algorithms for almost 65% of all the tested problems. Hence, the numerical results clearly demonstrate the merit of using a larger step-length and the flexibility of inexactly solving the subproblems.

## 6 Conclusions

In this paper, we have shown that, for a class of convex composite programming problems, the sequence generated by an inexact sGS decomposition based multi-block majorized (proximal) ADMM is equivalent to the sequence generated by an inexact proximal ALM starting with the same initial point. The convergence of the inexact majorized proximal ALM was first established, and the convergence of the multi-block ADMM-type algorithm follows readily because of the newly discovered equivalence. As a consequence of this equivalence, we are able to provide a very general answer to the open question on whether the whole sequence generated by the classic ADMM with  $\tau \in (0, 2)$  for a conventional two-block problem with one part of its objective function being linear, is convergent. Numerical experiments on solving a large number of linear and convex quadratic SDP problems are conducted. The numerical results show that one can achieve even better numerical performance of the ADMM if the step-length is chosen to be larger than the conventional upper bound of  $(1 + \sqrt{5})/2$ , and one

can get a considerable improvement by allowing inexact subproblems together with the large step-lengths on the multi-block ADMM for convex quadratic SDP problems. We hope that our theoretical analysis and numerical results can inspire more insightful studies on the ADMM-type algorithms.

## Acknowledgments

We would like to thank the two anonymous referees for their careful reading of this paper, and their insightful comments and suggestions which have helped to improve the quality of this paper.

## References

1. Bai, M., Zhang, X., Ni, G. and Cui, C.: An adaptive correction approach for tensor completion. *SIAM J. Imaging Sci.*, **9**, 1298–1323 (2016)
2. Bai, S. and Qi, H.-D.: Tackling the flip ambiguity in wireless sensor network localization and beyond. *Digit. Signal Process.* **55**, 85–97 (2016)
3. Bertsekas, D.P. and Tsitsiklis, J.N.: *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, Belmont, Massachusetts (1997)
4. Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**, 1–122 (2011)
5. Chen, L., Sun, D.F. and Toh, K.-C.: A note on the convergence of ADMM for linearly constrained convex optimization problems. *Comput. Optim. Appl.* **66**, 327–343 (2017)
6. Chen, L., Sun, D.F. and Toh, K.-C.: An efficient inexact symmetric Gauss-Seidel based majorized ADMM for high-dimensional convex composite conic programming. *Math. Program.* **161**(1-2), 237–270 (2017)
7. Chen, S.S., Donoho, D.L. and Saunders, M.A.: Atomic decomposition by basis pursuit. *SIAM Rev.* **43**, 129–159 (2001)
8. Clarke, F.H.: *Optimization and Nonsmooth Analysis*. Wiley, New York (1983)
9. Ding, C. and Qi, H.-D.: Convex optimization learning of faithful Euclidean distance representations in nonlinear dimensionality reduction. *Math. Program.* **164**(1-2), 341–381 (2017)
10. Ding, C., Sun, D.F., Sun, J. and Toh, K.-C.: Spectral operators of matrices. *Math. Program.* **168**, 509–531 (2018)
11. Dontchev, A.L. and Rockafellar, R.T.: *Implicit Functions and Solution Mappings*, Second Edition. Springer, New York (2014)
12. Du, M.Y.: *A Two-Phase Augmented Lagrangian Method for Convex Composite Quadratic Programming*. PhD thesis, Department of Mathematics, National University of Singapore (2015)
13. Eckstein, J.: *Augmented Lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results*. RUTCOR Research Reports (2012)
14. Eckstein, J. and Yao, W.: Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives. *Pac. J. Optim.* **11**, 619–644 (2015)
15. Eisenblätter, A., Grötschel, M. and Koster, A.: Frequency planning and ramification of coloring. *Discuss. Math. Graph Theory* **22**, 51–88 (2002)
16. Fazel, M., Pong, T.K., Sun D.F. and Tseng, P.: Hankel matrix rank minimization with applications to system identification and realization. *SIAM J. Matrix Anal.* **34**(3), 946–977 (2013)
17. Ferreira, J., Khoo, Y. and Singer, A.: Semidefinite programming approach for the quadratic assignment problem with a sparse graph. *Comput. Optim. Appl.* **69**(3), 677–712 (2018)
18. Gabay, D. and Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.* **2**(1), 17–40 (1976)
19. Gaines, B.R., Kim, J. and Zhou, H.: Algorithms for fitting the constrained lasso. *J. Comput. Graph. Stat.* **27**(4), 861–871 (2018)
20. Glowinski, R.: *Lectures on Numerical Methods for Non-Linear Variational Problems*. Published for the Tata Institute of Fundamental Research, Bombay. Springer-Verlag (1980)
21. Glowinski, R. and Marroco, A.: Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation-dualité d’une classe de problèmes de Dirichlet non linéaires. *Revue française d’automatique, Informatique Recherche Opérationnelle. Analyse Numérique* **9**(2), 41–76 (1975)
22. Han, D.R., Sun, D.R. and Zhang, L.W.: Linear rate convergence of the alternating direction method of multipliers for convex composite programming. *Math. Oper. Res.* **43**(2), 622–637 (2018)
23. Hestenes, M.: Multiplier and gradient methods. *J. Optim. Theory Appl.* **4**(5), 303–320 (1969)
24. Huber, P.J.: Robust estimation of a location parameter. *Ann. Math. Stat.* **35**, 73–101 (1964)
25. James, G.M., Paulson, C. and Rusmevichientong, P.: Penalized and constrained regression. Unpublished Manuscript, the latest version is available at <http://www-bcf.usc.edu/~gareth/research/Research.html> (2013)
26. Klopp, O.: Noisy low-rank matrix completion with general sampling distribution. *Bernoulli* **20**(1), 282–303 (2014)
27. Lam, X.Y., Marron, J.S., Sun, D.F. and Toh, K.-C.: Fast algorithms for large scale generalized distance weighted discrimination. *J. Comput. Graph. Stat.* **27**(2), 368–379 (2018)
28. Lemaréchal, C. and Sagastizábal, C.: Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries. *SIAM J. Optim.* **7**(2), 367–385 (1997)
29. Li, M., Sun, D.F. and Toh, K.-C.: A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization. *SIAM J. Optim.* **26**(2), 922–950 (2016)
30. Li, X.D., Sun, D.F. and Toh, K.-C.: A Schur complement based semi-proximal ADMM for convex quadratic conic programming and extensions. *Math. Program.* **155**, 333–373 (2016)
31. Li, X.D., Sun, D.F. and Toh, K.-C.: QSDPNAL: A two-phase augmented Lagrangian method for convex quadratic semidefinite programming. *Math. Program. Comput.* **10**(4), 703–743 (2018)
32. Li, X.D., Sun, D.F. and Toh, K.-C.: A block symmetric Gauss-Seidel decomposition theorem for convex composite quadratic programming and its applications. *Math. Program.* DOI:10.1007/s10107-018-1247-7 (2018)



33. Liu, J., Musialski, P., Wonka P. and Ye, J.: Tensor completion for estimating missing values in visual data. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 208–220 (2013)
34. Malick, J., Povh, J., Rendl, F. and Wiegeler, A.: Regularization methods for semidefinite programming. *SIAM J. Optim.* **20**, 336–356 (2009)
35. Mateos G., Bazerque, J.-A. and Giannakis G.B.: Distributed sparse linear regression. *IEEE Trans. Signal Proces.* **58**, 5262–5276 (2010)
36. Miao, W.M., Pan, S.H. and Sun, D.F.: A rank-corrected procedure for matrix completion with fixed basis coefficients. *Math. Program.* **159**, 289–338 (2016)
37. Negahban, S. and Wainwright, M.J.: Restricted strong convexity and weighted matrix completion: optimal bounds with noise. *J. Mach. Learn. Res.* **13**, 1665–1697 (2012)
38. Nie, J. and Wang, L.: Regularization methods for SDP relaxations in large-scale polynomial optimization. *SIAM J. Optim.* **22**, 408–428 (2012)
39. Nie, J. and Wang, L.: Semidefinite relaxations for best rank-1 tensor approximations. *SIAM J. Matrix Anal. Appl.* **35**, 1155–1179 (2014)
40. Peng, J. and Wei, Y.: Approximating k-means-type clustering via semidefinite programming. *SIAM J. Optim.* **18**, 186–205 (2007)
41. Potra F.A.: Weighted complementarity problems—a new paradigm for computing equilibria. *SIAM J. Optim.* **22**, 1634–1654 (2012)
42. Powell, M.: A method for nonlinear constraints in minimization problems. In: Fletcher, R. (ed.) *Optimization*, pp. 283–298. Academic, New York (1969)
43. Povh, J., Rendl, F. and Wiegeler, A.: A boundary point method to solve semidefinite programs. *Computing* **78**, 277–286 (2006)
44. Rockafellar, R.T.: *Convex Analysis*. Princeton University Press (1970)
45. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.* **1**, 97–116 (1976)
46. Schizas, I.D., Ribeiro A. and Giannakis G.B.: Consensus in ad hoc WSNs with noisy links - Part I: distributed estimation of deterministic signals. *IEEE Trans. Signal Process.* **56**, 350–364 (2008)
47. Sloane, N.: Challenge Problems: Independent Sets in Graphs, <http://www.research.att.com/~njas/doc/graphs.html>
48. Sun, D.F., Toh, K.-C. and Yang, L.Q.: A convergent 3-block semi-proximal alternating direction method of multipliers for conic programming with 4-type constraints. *SIAM J. Optim.* **25**(2), 882–915 (2015)
49. Teo, C.H., Vishwanathan, S.V.N., Smola A. and V.Le, Q.: Bundle methods for regularized risk minimization. *J. Mach. Learn. Res.* **11**, 313–365 (2010)
50. Toh, K.-C.: Solving large scale semidefinite programs via an iterative solver on the augmented systems. *SIAM J. Optim.* **14**, 670–698 (2004)
51. Toh, K.-C.: An inexact primal-dual path-following algorithm for convex quadratic SDP. *Math. Program.* **112**(1), 221–254 (2008)
52. Trick, M., Chvatal, V., Cook, W., Johnson, D., McGeoch, C. and Tarjan, R.: The Second DIMACS Implementation Challenge: NP Hard Problems: Maximum Clique, Graph Coloring, and Satisfiability. Rutgers University, <http://dimacs.rutgers.edu/Challenges/> (1992)
53. Wang, B. and Zou, H.: Another look at distance-weighted discrimination. *J. R. Stat. Soc. B* **80**, 177–198 (2018)
54. Wiegeler, A.: Biq Mac library—a collection of Max-Cut and quadratic 0 – 1 programming instances of medium size. Technical report (2007) <http://biqmac.uni-klu.ac.at/biqmaclib.pdf>
55. Yan, Z., Gao, S.Y. and Teo C.P.: On the design of sparse but efficient structures in operations. *Manage. Sci.* **64**, 2973–3468 (2018)
56. Yang, L.Q., Sun, D.F. and Toh, K.-C.: SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints. *Math. Program. Comput.* **7**, 331–366 (2015)
57. Zhang, N., Wu, J. and Zhang, L.W.: A linearly convergent majorized ADMM with indefinite proximal terms for convex composite programming and its applications. arXiv: 1706.01698v2 (2018)
58. Zhao, X.Y., Sun, D.F. and Toh K.-C.: A Newton-CG augmented Lagrangian method for semidefinite programming. *SIAM J. Optim.* **20**, 1737–1765 (2010)
59. Zhu, H., Cano, A. and Giannakis, G.B.: Distributed consensus-based demodulation: algorithms and error analysis. *IEEE Trans. Wirel. Commun.* **9**, 2044–2054 (2010)