Math. Program., Ser. A (2016) 155:333–373
DOI 10.1007/s10107-014-0850-5

FULL LENGTH PAPER

A Schur complement based semi-proximal ADMM for convex quadratic conic programming and extensions

Xudong Li · Defeng Sun · Kim-Chuan Toh

Received: 9 September 2014 / Accepted: 2 December 2014 / Published online: 14 December 2014

Abstract This paper is devoted to the design of an efficient and convergent semi-proximal alternating direction method of multipliers (ADMM) for finding a solution of low to medium accuracy to convex quadratic conic programming and related problems. For this class of problems, the convergent two block semi-proximal ADMM can be employed to solve their primal form in a straightforward way. However, it is known that it is more efficient to apply the directly extended multi-block semi-proximal ADMM, though its convergence is not guaranteed, to the dual form of these problems. Naturally, one may ask the following question: can one construct a convergent multi-block semi-proximal ADMM that is more efficient than the directly extended semi-proximal ADMM? Indeed, for linear conic programming with 4-block constraints this has been shown to be achievable in a recent paper by Sun et al.

Defeng Sun: Research supported in part by the Ministry of Education, Singapore, Academic Research Fund under Grant R-146-000-194-112.
Kim-Chuan Toh: Research supported in part by the Ministry of Education, Singapore, Academic Research Fund under Grant R-146-000-194-112.

Electronic supplementary material The online version of this article (doi:10.1007/s10107-014-0850-5) contains supplementary material, which is available to authorized users.

X. Li · K.-C. Toh
Department of Mathematics, National University of Singapore,
10 Lower Kent Ridge Road, Singapore 119076, Singapore
e-mail: lixudong@nus.edu.sg

K.-C. Toh
e-mail: mattohkc@nus.edu.sg

D. Sun (✉)
Department of Mathematics and Risk Management Institute, National University of Singapore,
10 Lower Kent Ridge Road, Singapore 119076, Singapore
e-mail: matsundf@nus.edu.sg

(2014, arXiv:1404.5378). Inspired by the aforementioned work and with the convex quadratic conic programming in mind, we propose a Schur complement based convergent semi-proximal ADMM for solving convex programming problems, with a coupling linear equality constraint, whose objective function is the sum of two proper closed convex functions plus an arbitrary number of convex quadratic or linear functions. Our convergent semi-proximal ADMM is particularly suitable for solving convex quadratic semidefinite programming (QSDP) with constraints consisting of linear equalities, a positive semidefinite cone and a simple convex polyhedral set. The efficiency of our proposed algorithm is demonstrated by numerical experiments on various examples including QSDP.

Keywords Convex quadratic conic programming · Multiple-block ADMM · Semi-proximal ADMM · Convergence · QSDP

Mathematics Subject Classification 90C06 · 90C20 · 90C22 · 90C25 · 65F10

1 Introduction

In this paper, we aim to design an efficient yet simple first order convergent method for solving convex quadratic conic programming. An important special case is the following convex quadratic semidefinite programming (QSDP)

$$\begin{aligned} &\min \ \frac{1}{2}\langle X,\ \mathcal{Q}X\rangle + \langle C,\ X\rangle \\ &\text{s.t. } \mathcal{A}_E X = b_E, \quad \mathcal{A}_I X \geq b_I, \quad X \in \mathcal{S}^n_+ \cap \mathcal{K}, \end{aligned} \tag{1}$$

where $\mathcal{S}^n_+$ is the cone of $n \times n$ symmetric and positive semi-definite matrices in the space of $n \times n$ symmetric matrices $\mathcal{S}^n$ endowed with the standard trace inner product $\langle \cdot,\ \cdot\rangle$ and the Frobenius norm $\|\cdot\|$, $\mathcal{Q}$ is a self-adjoint positive semidefinite linear operator from $\mathcal{S}^n$ to $\mathcal{S}^n$, $\mathcal{A}_E : \mathcal{S}^n \to \Re^{m_E}$ and $\mathcal{A}_I : \mathcal{S}^n \to \Re^{m_I}$ are two linear maps, $C \in \mathcal{S}^n$, $b_E \in \Re^{m_E}$ and $b_I \in \Re^{m_I}$ are given data, $\mathcal{K}$ is a nonempty simple closed convex set, e.g., $\mathcal{K} = \{W \in \mathcal{S}^n : \ L \leq W \leq U\}$ with $L, U \in \mathcal{S}^n$ being given matrices. By introducing a slack variable $W \in \mathcal{S}^n$, we can equivalently recast (1) as

$$\begin{aligned} &\min \ \frac{1}{2}\langle X,\ \mathcal{Q}X\rangle + \langle C,\ X\rangle + \delta_{\mathcal{K}}(W) \\ &\text{s.t. } \mathcal{A}_E X = b_E, \quad \mathcal{A}_I X \geq b_I, \quad X = W, \quad X \in \mathcal{S}^n_+, \end{aligned} \tag{2}$$

where $\delta_{\mathcal{K}}(\cdot)$ is the indicator function of $\mathcal{K}$, i.e., $\delta_{\mathcal{K}}(X) = 0$ if $X \in \mathcal{K}$ and $\delta_{\mathcal{K}}(X) = \infty$ if $X \notin \mathcal{K}$. The dual of problem (2) is given by

$$\begin{aligned} &\max \ -\delta^*_{\mathcal{K}}(-Z) + \langle b_I,\ y_I\rangle - \frac{1}{2}\langle X,\ \mathcal{Q}X\rangle + \langle b_E,\ y_E\rangle \\ &\text{s.t. } Z + \mathcal{A}^*_I y_I - \mathcal{Q}X + S + \mathcal{A}^*_E y_E = C, \quad y_I \geq 0, \quad S \in \mathcal{S}^n_+, \end{aligned} \tag{3}$$

where for any $Z \in \mathcal{S}^n$, $\delta^*_{\mathcal{K}}(-Z)$ is given by

$$\delta^*_{\mathcal{K}}(-Z) = -\inf_{W \in \mathcal{K}} \langle Z,\ W \rangle = \sup_{W \in \mathcal{K}} \langle -Z,\ W \rangle. \tag{4}$$

Due to its wide applications and mathematical elegance, QSDP has been extensively studied in the literature, see, for examples [1–6], and references therein.

It is evident that the dual problem (3) is in the form of the following convex optimization model:

$$\begin{aligned} &\min\ f(u) + \sum_{i=1}^{p} \theta_i(y_i) + g(v) + \sum_{j=1}^{q} \varphi_j(z_j) \\ &\text{s.t.}\ \mathcal{F}^* u + \sum_{i=1}^{p} \mathcal{A}_i^* y_i + \mathcal{G}^* v + \sum_{j=1}^{q} \mathcal{B}_j^* z_j = c, \end{aligned} \tag{5}$$

where p and q are given nonnegative integers, $f : \mathcal{U} \to (-\infty, +\infty]$, $g : \mathcal{V} \to (-\infty, +\infty]$, $\theta_i : \mathcal{Y}_i \to (-\infty, +\infty]$, $i = 1, \ldots, p$, and $\varphi_j : \mathcal{Z}_j \to (-\infty, +\infty]$, $j = 1, \ldots, q$ are closed proper convex functions, $\mathcal{F} : \mathcal{X} \to \mathcal{U}$, $\mathcal{G} : \mathcal{X} \to \mathcal{V}$, $\mathcal{A}_i : \mathcal{X} \to \mathcal{Y}_i$, $i = 1, \ldots, p$ and $\mathcal{B}_j : \mathcal{X} \to \mathcal{Z}_j$, $j = 1, \ldots, q$ are linear maps, $\mathcal{U}, \mathcal{V}, \mathcal{Y}_1, \ldots, \mathcal{Y}_p$, $\mathcal{Z}_1, \ldots, \mathcal{Z}_q$ and $\mathcal{X}$ are all real finite dimensional Euclidean spaces each equipped with an inner product $\langle \cdot,\ \cdot \rangle$ and its induced norm $\| \cdot \|$.

In this paper, we make the following blanket assumption.

Assumption 1 For $i = 1, \ldots, p$ and $j = 1, \ldots, q$, each $\theta_i(\cdot)$ and $\varphi_j(\cdot)$ are convex quadratic functions.

Note that, in general, problem (3) does not satisfy Assumption 1 unless y_I is vacuous from the model or $\mathcal{K} \equiv \mathcal{S}^n$. However, one can always reformulate problem (3) equivalently as

$$\begin{aligned} &\min\ (\delta^*_{\mathcal{K}}(-Z) + \delta_{\Re^{m_I}_+}(u)) - \langle b_I,\ y_I \rangle + \frac{1}{2}\langle X,\ \mathcal{Q}X \rangle + \delta_{\mathcal{S}^n_+}(S) - \langle b_E,\ y_E \rangle \\ &\text{s.t.}\ Z + \mathcal{A}_I^* y_I - \mathcal{Q}X + S + \mathcal{A}_E^* y_E = C, \\ &\qquad \mathcal{D}^* u - \mathcal{D}^* y_I = 0, \end{aligned} \tag{6}$$

where $\mathcal{D} : \Re^{m_I} \to \Re^{m_I}$ is any given nonsingular linear operator and $\delta_{\Re^{m_I}_+}(\cdot)$ is the indicator function over $\Re^{m_I}_+$. Now, one can see that problem (7) satisfies Assumption 1.

There are many other important cases that take the form of model (5) satisfying Assumption 1. One prominent example comes from the matrix completion with fixed basis coefficients [7–9]. Indeed the nuclear semi-norm penalized least squares model in [8] can be written as

$$\begin{aligned} &\min_{X \in \Re^{m \times n}}\ \frac{1}{2}\|\mathcal{A}_F X - d\|^2 + \rho(\|X\|_* - \langle C,\ X \rangle) \\ &\text{s.t.}\quad \mathcal{A}_E X = b_E, \quad X \in \mathcal{K} := \{X \mid \|\mathcal{R}_\Omega X\|_\infty \le \alpha\}, \end{aligned} \tag{7}$$

where $\|X\|_*$ is the nuclear norm of X defined as the sum of all its singular values, $\|\cdot\|_\infty$ is the element-wise l_∞ norm defined by $\|X\|_\infty := \max_{i=1,\ldots,m} \max_{j=1,\ldots,n} |X_{ij}|$, $\mathcal{A}_F : \Re^{m\times n} \to \Re^{n_F}$ and $\mathcal{A}_E : \Re^{m\times n} \to \Re^{n_E}$ are two linear maps, ρ and α are two given positive parameters, $d \in \Re^{n_F}$, $C \in \Re^{m\times n}$ and $b_E \in \Re^{n_E}$ are given data, $\Omega \subseteq \{1,\ldots,m\}\times\{1,\ldots,n\}$ is the set of the indices relative to which the basis coefficients are not fixed, $\mathcal{R}_\Omega : \Re^{m\times n} \to \Re^{|\Omega|}$ is the linear map such that $\mathcal{R}_\Omega X := (X_{ij})_{ij\in\Omega}$. Note that when there are no fixed basis coefficients (i.e., $\Omega = \{1,\ldots,m\}\times\{1,\ldots,n\}$ and $\mathcal{A}_E$ are vacuous), the above problem reduces to the model considered by Negahban and Wainwright in [10] and Klopp in [11]. By introducing slack variables η, R and W, we can reformulate problem (7) as

$$\begin{aligned} &\min \ \frac{1}{2}\|\eta\|^2 + \rho\big(\|R\|_* - \langle C,\ X\rangle\big) + \delta_{\mathcal{K}}(W) \\ &\text{s.t.}\ \mathcal{A}_F X - d = \eta, \quad \mathcal{A}_E x = b_E, \quad X = R, \quad X = W. \end{aligned} \tag{8}$$

The dual of problem (8) takes the form of

$$\begin{aligned} &\max \ -\delta^*_{\mathcal{K}}(-Z) - \frac{1}{2}\|\xi\|^2 + \langle d,\ \xi\rangle + \langle b_E,\ y_E\rangle \\ &\text{s.t.} \quad Z + \mathcal{A}^*_F\xi + S + \mathcal{A}^*_E y_E = -\rho C, \quad \|S\|_2 \le \rho, \end{aligned} \tag{9}$$

where $\|S\|_2$ is the operator norm of S, which is defined to be its largest singular value.

Another compelling example is the so called robust PCA (principle component analysis) considered in [12]:

$$\begin{aligned} &\min \ \|A\|_* + \lambda_1\|E\|_1 + \frac{\lambda_2}{2}\|Z\|_F^2 \\ &\text{s.t.}\ A + E + Z = W, \quad A, E, Z \in \Re^{m\times n}, \end{aligned} \tag{10}$$

where $W \in \Re^{m\times n}$ is the observed data matrix, $\|\cdot\|_1$ is the elementwise l_1 norm given by $\|E\|_1 := \sum_{i=1}^m \sum_{j=1}^n |E_{ij}|$, $\|\cdot\|_F$ is the Frobenius norm, λ_1 and λ_2 are two positive parameters. There are many different variants to the robust PCA model. For example, one may consider the following model where the observed data matrix W is incomplete:

$$\begin{aligned} &\min \ \|A\|_* + \lambda_1\|E\|_1 + \frac{\lambda_2}{2}\|\mathcal{P}_\Omega(Z)\|_F^2 \\ &\text{s.t.}\ \mathcal{P}_\Omega(A + E + Z) = \mathcal{P}_\Omega(W), \quad A, E, Z \in \Re^{m\times n}, \end{aligned} \tag{11}$$

i.e. one assumes that only a subset $\Omega \subseteq \{1,\ldots,m\}\times\{1,\ldots,n\}$ of the entries of W can be observed. Here $\mathcal{P}_\Omega : \Re^{m\times n} \to \Re^{m\times n}$ is the orthogonal projection operator defined by

$$\mathcal{P}_\Omega(X) = \begin{cases} X_{ij} & \text{if } (i,j) \in \Omega, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Again, problem (11) satisfies Assumption 1. In [13], Tao and Yuan tested one of the equivalent forms of problem (11). In the numerical section, we will see other interesting examples.

For notational convenience, let $\mathcal{Y} := \mathcal{Y}_1 \times \mathcal{Y}_2 \times, \dots, \mathcal{Y}_p$, $\mathcal{Z} := \mathcal{Z}_1 \times \mathcal{Z}_2 \times, \dots, \mathcal{Z}_q$. We write $y \equiv (y_1, y_2, \dots, y_p) \in \mathcal{Y}$ and $z \equiv (z_1, z_2, \dots, z_q) \in \mathcal{Z}$. Define the linear map $\mathcal{A} : \mathcal{X} \to \mathcal{Y}$ such that its adjoint is given by

$$\mathcal{A}^* y = \sum_{i=1}^{p} \mathcal{A}_i^* y_i \quad \forall y \in \mathcal{Y}.$$

Similarly, we define the linear map $\mathcal{B} : \mathcal{X} \to \mathcal{Z}$ such that its adjoint is given by

$$\mathcal{B}^* z = \sum_{j=1}^{q} \mathcal{B}_j^* z_j \quad \forall z \in \mathcal{Z}.$$

Additionally, let $\theta(y) := \sum_{i=1}^{p} \theta_i(y_i)$, $y \in \mathcal{Y}$ and $\varphi(z) := \sum_{j=1}^{q} \varphi_j(z_j)$, $z \in \mathcal{Z}$. Now we can rewrite (5) in the following compact form:

$$\begin{aligned} &\min \ f(u) + \theta(y) + g(v) + \varphi(z) \\ &\text{s.t. } \mathcal{F}^* u + \mathcal{A}^* y + \mathcal{G}^* v + \mathcal{B}^* z = c. \end{aligned} \tag{13}$$

Problem (5) can be view as a special case of the following block-separable convex optimization problem:

$$\min \left\{ \sum_{i=1}^{n} \phi_i(w_i) \mid \sum_{i=1}^{n} \mathcal{H}_i^* w_i = c \right\}, \tag{14}$$

where for each $i \in \{1, \dots, n\}$, $\mathcal{W}_i$ is a finite dimensional real Euclidean space equipped with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$, $\phi_i : \mathcal{W}_i \to (-\infty, +\infty]$ is a closed proper convex function, $\mathcal{H}_i : \mathcal{X} \to \mathcal{W}_i$ is a linear map and $c \in \mathcal{X}$ is given. Note that when we rewrite problem (5) in terms of (14), the quadratic structure in (5) is hidden in the sense that each ϕ_i will be treated equally. However, this special quadratic structure will be thoroughly exploited in our search for an efficient yet simple ADMM-type method with guaranteed convergence.

Let $\sigma > 0$ be a given parameter. The augmented Lagrangian function for (14) is defined by

$$\mathcal{L}_\sigma(w_1, \dots, w_n; x) := \sum_{i=1}^{n} \phi_i(w_i) + \left\langle x \sum_{i=1}^{n} \mathcal{H}_i^* w_i - c \right\rangle + \frac{\sigma}{2} \left\| \sum_{i=1}^{n} \mathcal{H}_i^* w_i - c \right\|^2$$

for $w_i \in \mathcal{W}_i$, $i = 1, \dots, n$ and $x \in \mathcal{X}$. Choose any initial points $w_i^0 \in \operatorname{dom}(\phi_i)$, $i = 1, \dots, q$ and $x^0 \in \mathcal{X}$. The classical augmented Lagrangian method consists of the following iterations:

$$(w_1^{k+1}, \ldots, w_n^{k+1}) = \operatorname{argmin} \mathcal{L}_\sigma(w_1, \ldots, w_n; x^k), \tag{15}$$

$$x^{k+1} = x^k + \tau\sigma\left(\sum_{i=1}^{n} \mathcal{H}_i^* w_i^{k+1} - c\right), \tag{16}$$

where $\tau \in (0, 2)$ guarantees the convergence. Due to the non-separability of the quadratic penalty term in $\mathcal{L}_\sigma$, it is generally a challenging task to solve the joint minimization problem (15) exactly or approximately with high accuracy. To overcome this difficulty, one may consider the following n-block alternating direction methods of multipliers (ADMM):

$$\begin{aligned}
w_1^{k+1} &= \operatorname{argmin} \mathcal{L}_\sigma(w_1, w_2^k \ldots, w_n^k; x^k),\\
&\vdots\\
w_i^{k+1} &= \operatorname{argmin} \mathcal{L}_\sigma(w_1^{k+1}, \ldots, w_{i-1}^{k+1}, w_i, w_{i+1}^k, \ldots, w_n^k; x^k),\\
&\vdots\\
w_n^{k+1} &= \operatorname{argmin} \mathcal{L}_\sigma(w_1^{k+1}, \ldots, w_{n-1}^{k+1}, w_n; x^k),\\
x^{k+1} &= x^k + \tau\sigma\left(\sum_{i=1}^{n} \mathcal{H}_i^* w_i^{k+1} - c\right).
\end{aligned} \tag{17}$$

The above n-block ADMM is an direct extension of the ADMM for solving the following 2-block convex optimization problem

$$\min \left\{\phi_1(w_1) + \phi_2(w_2) \mid \mathcal{H}_1^* w_1 + \mathcal{H}_2^* w_2 = c\right\}. \tag{18}$$

The convergence of 2-block ADMM has already been extensively studied in [14–19] and references therein. However, the convergence of the n-block ADMM has been ambiguous for a long time. Fortunately this ambiguity has been addressed very recently in [20] where Chen, He, Ye, and Yuan showed that the direct extension of the ADMM to the case of a 3-block convex optimization problem is not necessarily convergent. On the other hand, the n-block ADMM with $\tau \geq 1$ often works very well in practice and this fact poses a big challenge if one attempts to develop new ADMM-type algorithms which have convergence guarantee but with competitive numerical efficiency and iteration simplicity as the n-block ADMM.

Recently, there is exciting progress in this active research area. Sun et al. [21] proposed a convergent semi-proximal ADMM (PADMM3c) for convex programming problems of three separable blocks in the objective function with the third part being linear. One distinctive feature of algorithm PADMM3c is that it requires only an inexpensive extra step, compared to the 3-block ADMM, but yields a convergent and faster algorithm. Extensive numerical tests on the doubly non-negative SDP problems with equality and/or inequality constraints demonstrate that PADMM3c can have superior numerical efficiency over the directly extended ADMM. This opens up the possibility of designing an efficient and convergent ADMM type method for solving multi-block convex optimization problems. Inspired by the aforementioned work, in this paper we shall propose a Schur complement based semi-proximal ADMM (SCB-SPADMM) to efficiently solve the convex quadratic conic programming problems to medium accuracy. The development of our algorithm is based on the simple yet elegant idea of the Schur complement and the convenient convergence results of the semi-proximal ADMM given in the appendix of [22]. Our primary motivation for designing the proposed SCB-SPADMM is to generate a good initial point quickly to warm-start locally fast convergent method such as the semismooth Newton-CG method used in [23,24] for solving linear SDP though the method proposed here is definitely of its own interest.

The remaining parts of this paper are organized as follows. In the next section, we present a Schur complement based semi-proximal augmented Lagrangian method (SCB-SPALM) to solve a 2-block convex optimization problem where the second function g is quadratic and then show the relation between our SCB-SPALM and the generic 2-block semi-proximal ADMM (SPADMM). In Sect. 3, we propose our main algorithm SCB-SPADMM for solving the general convex model (5). Our main convergence results are presented in this section. Section 4 is devoted to the implementation and numerical experiments of using our SCB-SPADMM to solve convex quadratic conic programming problems and the various extensions. We conclude our paper in the final section.

Notation Define the spectral (or operator) norm of a given linear operator $\mathcal{T}$ by $\|\mathcal{T}\| := \sup_{\|w\|=1} \|\mathcal{T}w\|$. For any $w \in \mathcal{U}$, we let

$$\mathrm{Prox}_f(w) := \mathrm{argmin}_u \ f(u) + \frac{1}{2}\|u - w\|^2.$$

2 A Schur complement based semi-proximal augmented Lagrangian method

Before we introduce our approach for the multi-block case, we need to consider the convex optimization problem with the following 2-block separable structure

$$\begin{array}{ll} \min & f(u) + g(v) \\ \text{s.t.} & \mathcal{F}^*u + \mathcal{G}^*v = c, \end{array} \tag{19}$$

where $f : \mathcal{U} \to (-\infty, +\infty]$ and $g : \mathcal{V} \to (-\infty, +\infty]$ are closed proper convex functions, $\mathcal{F} : \mathcal{X} \to \mathcal{U}$ and $\mathcal{G} : \mathcal{X} \to \mathcal{V}$ are given linear maps. The dual of problem (19) is given by

$$\min \left\{ \langle c, \ x \rangle + f^*(s) + g^*(t) \mid \mathcal{F}x + s = 0, \ \mathcal{G}x + t = 0 \right\}. \tag{20}$$

Let $\sigma > 0$ be given. The augmented Lagrangian function associated with (19) is given as follows:

$$\mathcal{L}_\sigma(u, v; x) = f(u) + g(v) + \langle x, \ \mathcal{F}^*u + \mathcal{G}^*v - c\rangle + \frac{\sigma}{2}\|\mathcal{F}^*u + \mathcal{G}^*v - c\|^2.$$

The semi-proximal ADMM proposed in [22], when applied to (19), has the following template. Since the proximal terms added here are allowed to be positive semi-definite, the corresponding method is referred to as semi-proximal ADMM instead of proximal ADMM as in [22].

Algorithm SPADMM: A generic 2-block semi-proximal ADMM for solving (19). Let $\sigma > 0$ and $\tau \in (0, \infty)$ be given parameters. Let $\mathcal{T}_f$ and $\mathcal{T}_g$ be given self-adjoint positive semidefinite, not necessarily positive definite, linear operators defined on $\mathcal{U}$ and $\mathcal{V}$, respectively. Choose $(u^0, v^0, x^0) \in \text{dom}(f) \times \text{dom}(g) \times \mathcal{X}$. For $k = 0, 1, 2, \ldots$, perform the kth iteration as follows:

Step 1. Compute

$$u^{k+1} = \text{argmin}_u \ \mathcal{L}_\sigma(u, v^k; x^k) + \frac{\sigma}{2}\|u - u^k\|^2_{\mathcal{T}_f}. \tag{21}$$

Step 2. Compute

$$v^{k+1} = \text{argmin}_v \ \mathcal{L}_\sigma(u^{k+1}, v; x^k) + \frac{\sigma}{2}\|v - v^k\|^2_{\mathcal{T}_g}. \tag{22}$$

Step 3. Compute

$$x^{k+1} = x^k + \tau\sigma(\mathcal{F}^* u^{k+1} + \mathcal{G}^* v^{k+1} - c). \tag{23}$$

In the above 2-block semi-proximal ADMM for solving (19), the presence of $\mathcal{T}_f$ and $\mathcal{T}_g$ can help to guarantee the existence of solutions for the subproblems (21) and (22). In addition, they play important roles in ensuring the boundedness of the two generated sequences $\{u^{k+1}\}$ and $\{v^{k+1}\}$. Hence, these two proximal terms are preferred. The choices of $\mathcal{T}_f$ and $\mathcal{T}_g$ are very much problem dependent. The general principle is that both $\mathcal{T}_f$ and $\mathcal{T}_g$ should be as small as possible while u^{k+1} and v^{k+1} are still relatively easy to compute.

Let ∂f and ∂g be the subdifferential mappings of f and g, respectively. Since both ∂f and ∂g are maximally monotone, there exist two self-adjoint and positive semidefinite operators Σ_f and Σ_g such that for all $u, \tilde{u} \in \text{dom}(f)$, $\xi \in \partial f(u)$, and $\tilde{\xi} \in \partial f(\tilde{u})$,

$$\langle \xi - \tilde{\xi}, u - \tilde{u}\rangle \geq \|u - \tilde{u}\|^2_{\Sigma_f} \tag{24}$$

and for all $v, \tilde{v} \in \text{dom}(g)$, $\zeta \in \partial g(v)$, and $\tilde{\zeta} \in \partial g(\tilde{v})$,

$$\langle \zeta - \tilde{\zeta}, v - \tilde{v}\rangle \geq \|v - \tilde{v}\|^2_{\Sigma_g}. \tag{25}$$

For the convergence of the 2-block semi-proximal ADMM, we need the following assumption.

Assumption 2 There exists $(\hat{u}, \hat{v}) \in \text{ri}(\text{dom}\, f \times \text{dom}\, g)$ such that $\mathcal{F}^*\hat{u} + \mathcal{G}^*\hat{v} = c$.

Theorem 1 *Let Σ_f and Σ_g be the self-adjoint and positive semidefinite operators defined by (24) and (25), respectively. Suppose that the solution set of problem (19) is nonempty and that Assumption 2 holds. Assume that $\mathcal{T}_f$ and $\mathcal{T}_g$ are chosen such that the sequence $\{(u^k, v^k, x^k)\}$ generated by Algorithm SPADMM is well defined. Then, under the condition either (a) $\tau \in (0, (1+\sqrt{5})/2)$ or (b) $\tau \geq (1+\sqrt{5})/2$ but $\sum_{k=0}^{\infty}(\|\mathcal{G}^*(v^{k+1}-v^k)\|^2 + \tau^{-1}\|\mathcal{F}^*u^{k+1} + \mathcal{G}^*v^{k+1} - c\|^2) < \infty$, the following results hold:*

- (i) *If $(u^\infty, v^\infty, x^\infty)$ is an accumulation point of $\{(u^k, v^k, x^k)\}$, then (u^∞, v^∞) solves problem (19) and x^∞ solves (20), respectively.*
- (ii) *If both $\sigma^{-1}\Sigma_f + \mathcal{T}_f + \mathcal{F}\mathcal{F}^*$ and $\sigma^{-1}\Sigma_g + \mathcal{T}_g + \mathcal{G}\mathcal{G}^*$ are positive definite, then the sequence $\{(u^k, v^k, x^k)\}$, which is automatically well defined, converges to a unique limit, say, $(u^\infty, v^\infty, x^\infty)$ with (u^∞, v^∞) solving problem (19) and x^∞ solving (20), respectively.*
- (iii) *When the u-part disappears, the corresponding results in parts (i)–(ii) hold under the condition either $\tau \in (0, 2)$ or $\tau \geq 2$ but $\sum_{k=0}^{\infty}\|\mathcal{G}^*v^{k+1} - c\|^2 < \infty$.*

Remark 1 The conclusions of Theorem 1 follow essentially from the results given in [22, Theorem B.1]. See [21] for more detailed discussions.

Next, we shall pay particular attention to the case when g is a quadratic function:

$$g(v) = \frac{1}{2}\langle v,\ \Sigma_g v\rangle - \langle b,\ v\rangle, \quad v \in \mathcal{V}, \tag{26}$$

where Σ_g is a self-adjoint positive semidefinite linear operator defined on $\mathcal{V}$ and $b \in \mathcal{V}$ is a given vector. Problem (19) now takes the form of

$$\begin{aligned} &\min\ f(u) + \frac{1}{2}\langle v,\ \Sigma_g v\rangle - \langle b,\ v\rangle \\ &\text{s.t. } \mathcal{F}^*u + \mathcal{G}^*v = c. \end{aligned} \tag{27}$$

The dual of problem (27) is given by

$$\min\left\{\langle c,\ x\rangle + f^*(s) + g^*(t) \mid \mathcal{F}x + s = 0,\ \mathcal{G}x + t = 0\right\}. \tag{28}$$

In order to solve subproblem (22) in Algorithm SPADMM, we need to solve a linear system with the linear operator given by $\sigma^{-1}\Sigma_g + \mathcal{G}\mathcal{G}^*$. Hence, an appropriate proximal term should be chosen such that (22) can be solved efficiently. Here, we choose $\mathcal{T}_g$ as follows. Let $\mathcal{E}_g : \mathcal{V} \to \mathcal{V}$ be a self-adjoint positive definite linear operator such that it is a majorization of $\sigma^{-1}\Sigma_g + \mathcal{G}\mathcal{G}^*$, i.e.,

$$\mathcal{E}_g \succeq \sigma^{-1}\Sigma_g + \mathcal{G}\mathcal{G}^*.$$

We choose $\mathcal{E}_g$ such that its inverse can be computed at a moderate cost. Define

$$\mathcal{T}_g := \mathcal{E}_g - \sigma^{-1}\Sigma_g - \mathcal{G}\mathcal{G}^* \succeq 0. \tag{29}$$

Note that for numerical efficiency, we need the self-adjoint positive semidefinite linear operator $\mathcal{T}_g$ to be as small as possible. In order to fully exploit the structure of the quadratic function g, we add, instead of a naive proximal term, a proximal term based on the Schur complement as follows. For a given $\mathcal{T}_f \succeq 0$, we define the self-adjoint positive semidefinite linear operator

$$\widehat{\mathcal{T}}_f := \mathcal{T}_f + \mathcal{F}\mathcal{G}^*\mathcal{E}_g^{-1}\mathcal{G}\mathcal{F}^*. \tag{30}$$

For later developments, here we state a proposition which uses the Schur complement condition for establishing the positive definiteness of a linear operator.

Proposition 1 *It holds that*

$$\mathcal{W} := \begin{pmatrix} \mathcal{F} \\ \mathcal{G} \end{pmatrix} \begin{pmatrix} \mathcal{F} \\ \mathcal{G} \end{pmatrix}^* + \sigma^{-1} \begin{pmatrix} \Sigma_f & \\ & \Sigma_g \end{pmatrix} + \begin{pmatrix} \widehat{\mathcal{T}}_f & \\ & \mathcal{T}_g \end{pmatrix} \succ 0$$

if and only if $\mathcal{F}\mathcal{F}^* + \sigma^{-1}\Sigma_f + \mathcal{T}_f \succ 0$.

Proof We have that

$$\mathcal{W} = \begin{pmatrix} \mathcal{F}\mathcal{F}^* + \sigma^{-1}\Sigma_f + \widehat{\mathcal{T}}_f & \mathcal{F}\mathcal{G}^* \\ \mathcal{G}\mathcal{F}^* & \mathcal{G}\mathcal{G}^* + \sigma^{-1}\Sigma_g + \mathcal{T}_g \end{pmatrix}.$$

Since $\mathcal{E}_g = \mathcal{G}\mathcal{G}^* + \sigma^{-1}\Sigma_g + \mathcal{T}_g \succ 0$, by the Schur complement condition for ensuring the positive definiteness of linear operators, we have $\mathcal{W} \succ 0$ if and only if

$$\mathcal{F}\mathcal{F}^* + \sigma^{-1}\Sigma_f + \widehat{\mathcal{T}}_f - \mathcal{F}\mathcal{G}^*\mathcal{E}_g^{-1}\mathcal{G}\mathcal{F}^* \succ 0.$$

By (30), we know that the conclusion of this proposition holds. □

Now, we can propose our Schur complement based semi-proximal augmented Lagrangian method (SCB-SPALM) to solve (27) with a specially chosen proximal term involving $\widehat{\mathcal{T}}_f$ and $\mathcal{T}_g$.

Algorithm SCB-SPALM: A Schur complement based semi-proximal augmented Lagrangian method for solving (27).

Let $\sigma > 0$ and $\tau \in (0, \infty)$ be given parameters. Choose $(u^0, v^0, x^0) \in \text{dom}(f) \times \mathcal{V} \times \mathcal{X}$. For $k = 0, 1, 2, \ldots$, perform the kth iteration as follows:

Step 1. Compute

$$(u^{k+1}, v^{k+1}) = \text{argmin}_{u,v}\ \mathcal{L}_\sigma(u, v; x^k) + \frac{\sigma}{2}\|u - u^k\|^2_{\widehat{\mathcal{T}}_f} + \frac{\sigma}{2}\|v - v^k\|^2_{\mathcal{T}_g}. \tag{31}$$

Step 2. Compute

$$x^{k+1} = x^k + \tau\sigma(\mathcal{F}^*u^{k+1} + \mathcal{G}^*v^{k+1} - c). \tag{32}$$

Note that problem (31) in Step 1 is well defined if the the linear operator $\mathcal{W}$ defined in Proposition 1 is positive definite, or equivalently, if $\mathcal{F}\mathcal{F}^* + \sigma^{-1}\Sigma_f + \mathcal{T}_f \succ 0$. Also, note that in the context of the convex optimization problem (27), Assumption 2 is reduced to the following:

Assumption 3 There exists $(\hat{u}, \hat{v}) \in \text{ri}(\text{dom}\, f) \times \mathcal{V}$ such that $\mathcal{F}^*\hat{u} + \mathcal{G}^*\hat{v} = c$.

Now, we are ready to establish our convergence results for Algorithm SCB-SPALM for solving (27).

Theorem 2 *Let Σ_f, Σ_g and $\mathcal{T}_g$ be three self-adjoint and positive semidefinite operators defined by* (24)*,* (26) *and* (29)*, respectively. Suppose that the solution set of problem* (27) *is nonempty and that Assumption* 3 *holds. Assume that $\mathcal{T}_f$ is chosen such that the sequence $\{(u^k, v^k, x^k)\}$ generated by Algorithm SCB-SPALM is well defined. Then, under the condition either (a) $\tau \in (0, 2)$ or (b) $\tau \geq 2$ but $\sum_{k=0}^{\infty} \|\mathcal{F}^*u^{k+1} + \mathcal{G}^*v^{k+1} - c\|^2 < \infty$, the following results hold:*

(i) *If $(u^\infty, v^\infty, x^\infty)$ is an accumulation point of $\{(u^k, v^k, x^k)\}$, then (u^∞, v^∞) solves problem* (27) *and x^∞ solves* (28)*, respectively.*
(ii) *If $\sigma^{-1}\Sigma_f + \mathcal{T}_f + \mathcal{F}\mathcal{F}^*$ is positive definite, then the sequence $\{(u^k, v^k, x^k)\}$ is well defined and it converges to a unique limit, say, $(u^\infty, v^\infty, x^\infty)$ with (u^∞, v^∞) solving problem* (27) *and x^∞ solving* (28)*, respectively.*

Proof By combining Theorem 1 and Proposition 1, one can prove the results of this theorem directly. □

The relationship between Algorithm SCB-SPALM and Algorithm SPADMM for solving (27) will be revealed in the next proposition.

Let $\delta_g : \mathcal{U} \times \mathcal{V} \times \mathcal{X} \to \mathcal{U}$ be an auxiliary linear function associated with (27) defined by

$$\delta_g(u, v, x) := \mathcal{F}\mathcal{G}^*\mathcal{E}_g^{-1}(b - \mathcal{G}x - \Sigma_g v + \sigma\mathcal{G}(c - \mathcal{F}^*u - \mathcal{G}^*v)). \tag{33}$$

Let $\bar{u} \in \mathcal{U}$, $\bar{v} \in \mathcal{V}$, $\bar{x} \in \mathcal{X}$ and $c \in \mathcal{X}$ be given. Denote

$$\bar{c} := c - \mathcal{F}^*\bar{u} - \mathcal{G}^*\bar{v} \quad \text{and} \quad \bar{\delta}_g := \delta_g(\bar{u}, \bar{v}, \bar{x}) = \mathcal{F}\mathcal{G}^*\mathcal{E}_g^{-1}(b - \mathcal{G}\bar{x} - \Sigma_g\bar{v} + \sigma\mathcal{G}\bar{c}).$$

Let $(u^+, v^+) \in \mathcal{U} \times \mathcal{V}$ be defined by

$$(u^+, v^+) = \text{argmin}_{u,v}\ \mathcal{L}_\sigma(u, v; \bar{x}) + \frac{\sigma}{2}\|u - \bar{u}\|^2_{\mathcal{T}_f} + \frac{\sigma}{2}\|v - \bar{v}\|^2_{\mathcal{T}_g}. \tag{34}$$

Proposition 2 *Let $\bar{\alpha} := \sigma^{-1}b + \mathcal{T}_g\bar{v} + \mathcal{G}(c - \sigma^{-1}\bar{x})$. Define $v' \in \mathcal{V}$ by*

$$v' = \text{argmin}_v\ \mathcal{L}_\sigma(\bar{u}, v; \bar{x}) + \frac{\sigma}{2}\|v - \bar{v}\|^2_{\mathcal{T}_g} = \mathcal{E}_g^{-1}(\bar{\alpha} - \mathcal{G}\mathcal{F}^*\bar{u}). \tag{35}$$

The optimal solution (u^+, v^+) *to problem* (34) *is generated exactly by the following procedure*

$$
\begin{cases} u^+ = \operatorname{argmin}_u \mathcal{L}_\sigma(u, \bar{v}; \bar{x}) + \langle \bar{\delta}_g, u\rangle + \frac{\sigma}{2}\|u - \bar{u}\|^2_{\mathcal{T}_f}, \\ v^+ = \operatorname{argmin}_v \mathcal{L}_\sigma(u^+, v; \bar{x}) + \frac{\sigma}{2}\|v - \bar{v}\|^2_{\mathcal{T}_g} = \mathcal{E}_g^{-1}(\bar{\alpha} - \mathcal{G}\mathcal{F}^* u^+). \end{cases} \tag{36}
$$

Furthermore, (u^+, v^+) *can also be obtained by the following equivalent procedure*

$$
\begin{cases} u^+ = \operatorname{argmin}_u \mathcal{L}_\sigma(u, v'; \bar{x}) + \frac{\sigma}{2}\|u - \bar{u}\|^2_{\mathcal{T}_f}, \\ v^+ = \operatorname{argmin}_v \mathcal{L}_\sigma(u^+, v; \bar{x}) + \frac{\sigma}{2}\|v - \bar{v}\|^2_{\mathcal{T}_g} = \mathcal{E}_g^{-1}(\bar{\alpha} - \mathcal{G}\mathcal{F}^* u^+). \end{cases} \tag{37}
$$

Proof First we show that the equivalence between (34) and (36). Define

$$
\widetilde{\mathcal{L}}_\sigma(u, v; \bar{x}) := \mathcal{L}_\sigma(u, v; \bar{x}) + \frac{\sigma}{2}\|u - \bar{u}\|^2_{\widehat{\mathcal{T}}_f} + \frac{\sigma}{2}\|v - \bar{v}\|^2_{\mathcal{T}_g}, \quad (u, v) \in \mathcal{U} \times \mathcal{V}.
$$

By simple algebraic manipulations, we have that

$$
\widetilde{\mathcal{L}}_\sigma(u, v; \bar{x}) = f(u) + \frac{\sigma}{2}\|u - \bar{u}\|^2_{\widehat{\mathcal{T}}_f} + \phi(u, v) - \frac{1}{2\sigma}\|\bar{x}\|^2, \tag{38}
$$

where

$$
\begin{aligned}
\phi(u, v) &= g(v) + \frac{\sigma}{2}\|\mathcal{F}^* u + \mathcal{G}^* v + \sigma^{-1}\bar{x} - c\|^2 + \frac{\sigma}{2}\|v - \bar{v}\|^2_{\mathcal{T}_g} \\
&= \frac{\sigma}{2}\Big(\langle v, \mathcal{E}_g v\rangle + 2\langle v, \mathcal{G}\mathcal{F}^* u - \bar{\alpha}\rangle + \|\mathcal{F}^* u + \sigma^{-1}\bar{x} - c\|^2 + \|\bar{v}\|^2_{\mathcal{T}_g}\Big)
\end{aligned}
$$

with $\bar{\alpha}$ as defined in the proposition. For any given $u \in \mathcal{U}$, let

$$
v(u) := \operatorname{argmin}_{v\in\mathcal{V}} \phi(u, v) = \mathcal{E}_g^{-1}(\bar{\alpha} - \mathcal{G}\mathcal{F}^* u).
$$

Then by using the fact that $\min_v \frac{1}{2}\langle v, \mathcal{E}_g v\rangle + \langle q, v\rangle = -\frac{1}{2}\langle q, \mathcal{E}_g^{-1} q\rangle$ for any $q \in \mathcal{V}$, we have that

$$
\begin{aligned}
\phi(u, v(u)) &= \frac{\sigma}{2}\Big(- \langle \mathcal{G}\mathcal{F}^* u - \bar{\alpha}, \mathcal{E}_g^{-1}(\mathcal{G}\mathcal{F}^* u - \bar{\alpha})\rangle + \|\mathcal{F}^* u + \sigma^{-1}\bar{x} - c\|^2 + \|\bar{v}\|^2_{\mathcal{T}_g}\Big) \\
&= \frac{\sigma}{2}\Big(\langle u, (\mathcal{F}\mathcal{F}^* - \mathcal{F}\mathcal{G}^*\mathcal{E}_g^{-1}\mathcal{G}\mathcal{F}^*)u\rangle + 2\langle u, \mathcal{F}(\mathcal{G}^*\mathcal{E}_g^{-1}\bar{\alpha} + \sigma^{-1}\bar{x} - c)\rangle\Big) + \kappa_0,
\end{aligned}
$$

where $\kappa_0 = \frac{\sigma}{2}(\|\sigma^{-1}\bar{x} - c\|^2 + \|\bar{v}\|^2_{\mathcal{T}_g} - \|\bar{\alpha}\|^2_{\mathcal{E}_g^{-1}})$. Let

$$
\begin{aligned}
\kappa_1 &:= \kappa_0 + \frac{\sigma}{2}\|\mathcal{G}\mathcal{F}^* \bar{u}\|^2_{\mathcal{E}_g^{-1}} - \frac{1}{2\sigma}\|\bar{x}\|^2 \\
&= -\langle c, \bar{x}\rangle + \frac{\sigma}{2}(\|c\|^2 + \|\mathcal{G}\mathcal{F}^*\bar{u}\|^2_{\mathcal{E}_g^{-1}} + \|\bar{v}\|^2_{\mathcal{T}_g} - \|\bar{\alpha}\|^2_{\mathcal{E}_g^{-1}}).
\end{aligned}
$$

From (38), we have that for any given $u \in \mathcal{U}$,

$$\begin{aligned}
\widetilde{\mathcal{L}}_\sigma(u, v(u); \bar{x}) &= f(u) + \frac{\sigma}{2}\|u - \bar{u}\|^2_{\mathcal{T}_f} + \frac{\sigma}{2}\|\mathcal{G}\mathcal{F}^*(u - \bar{u})\|^2_{\mathcal{E}_g^{-1}} \\
&\quad + \phi(u, v(u)) - \frac{1}{2\sigma}\|\bar{x}\|^2 \\
&= f(u) + \frac{\sigma}{2}\|u - \bar{u}\|^2_{\mathcal{T}_f} + \sigma\langle u,\ \mathcal{F}(\mathcal{G}^*\mathcal{E}_g^{-1}\bar{\alpha} + \sigma^{-1}\bar{x} - c) - \mathcal{F}\mathcal{G}^*\mathcal{E}_g^{-1}\mathcal{G}\mathcal{F}^*\bar{u}\rangle \\
&\quad + \frac{\sigma}{2}\langle u,\ \mathcal{F}\mathcal{F}^*u\rangle + \kappa_1 \\
&= f(u) + \frac{\sigma}{2}\|u - \bar{u}\|^2_{\mathcal{T}_f} + \langle u,\ \bar{\delta}_g\rangle + \langle u,\ \mathcal{F}(\bar{x} + \sigma(\mathcal{G}^*\bar{v} - c))\rangle + \frac{\sigma}{2}\langle u,\ \mathcal{F}\mathcal{F}^*u\rangle + \kappa_1 \\
&= \mathcal{L}_\sigma(u, \bar{v}; \bar{x}) + \langle u,\ \bar{\delta}_g\rangle + \frac{\sigma}{2}\|u - \bar{u}\|^2_{\mathcal{T}_f} + \kappa_2, \qquad (39)
\end{aligned}$$

where $\kappa_2 = \kappa_1 - g(\bar{v}) - \frac{\sigma}{2}\|\mathcal{G}^*\bar{v} - c\|^2 - \langle \bar{x},\ \mathcal{G}^*\bar{v} - c\rangle$. Note that with some manipulations, we can show that the constant term

$$\kappa_2 = \frac{\sigma}{2}\|\mathcal{G}\mathcal{F}^*\bar{u}\|^2_{\mathcal{E}_g^{-1}} - \frac{\sigma}{2}\|\mathcal{E}_g\bar{v} - \bar{\alpha}\|^2_{\mathcal{E}_g^{-1}}.$$

Now, we have that

$$\min_{u\in\mathcal{U}, v\in\mathcal{V}} \widetilde{\mathcal{L}}_\sigma(u, v; \bar{x}) = \min_{u\in\mathcal{U}}\left(\min_{v\in\mathcal{V}} \widetilde{\mathcal{L}}_\sigma(u, v; \bar{x})\right) = \min_{u\in\mathcal{U}} \widetilde{\mathcal{L}}_\sigma(u, v(u); \bar{x}),$$

where $\widetilde{\mathcal{L}}_\sigma(u, v(u); \bar{x})$ satisfies (39). From here, the equivalence between (34) and (36) follows.

Next, we prove the equivalence between (36) and (37). Note that, the first minimization problem in (37) can be equivalently recast as

$$0 \in \partial f(u^+) + \mathcal{F}\bar{x} + \sigma\mathcal{F}(\mathcal{F}^*u^+ + \mathcal{G}^*v' - c) + \sigma\mathcal{T}_f(u^+ - \bar{u}),$$

which, together with the definition of v' given in (35), is equivalent to

$$0 \in \partial f(u^+) + \mathcal{F}\bar{x} + \sigma\mathcal{F}(\mathcal{F}^*u^+ - c + \mathcal{G}^*\mathcal{E}_g^{-1}(\bar{\alpha} - \mathcal{G}\mathcal{F}^*\bar{u})) + \sigma\mathcal{T}_f(u^+ - \bar{u}). \qquad (40)$$

The condition (40) can be reformulated as

$$\begin{aligned}
0 \in\ & \partial f(u^+) + \mathcal{F}\bar{x} + \sigma\mathcal{F}(\mathcal{F}^*u^+ + \mathcal{G}^*\bar{v} - c) \\
& + \sigma\mathcal{F}\mathcal{G}^*\mathcal{E}_g^{-1}(\bar{\alpha} - \mathcal{G}\mathcal{F}^*\bar{u} - \mathcal{E}_g\bar{v}) + \sigma\mathcal{T}_f(u^+ - \bar{u}).
\end{aligned}$$

Thus, we have

$$0 \in \partial f(u^+) + \mathcal{F}\bar{x} + \sigma\mathcal{F}(\mathcal{F}^*u^+ + \mathcal{G}^*\bar{v} - c) + \bar{\delta}_g + \sigma\mathcal{T}_f(u^+ - \bar{u}), \qquad (41)$$

which can equivalently be rewritten as

$$u^+ = \text{argmin}_u\ \mathcal{L}_\sigma(u, \bar{v}; \bar{x}) + \langle \bar{\delta}_g,\ u\rangle + \frac{\sigma}{2}\|u - \bar{u}\|^2_{\mathcal{T}_f}.$$

The equivalence between (36) and (37) then follows. This completes the proof of this proposition. □

Proposition 3 *Let $\delta_g^k := \delta_g(u^k, v^k, x^k)$ for $k = 0, 1, 2, \ldots$. We have that u^{k+1} and v^{k+1} obtained by Algorithm SCB-SPALM for solving (27) can be generated exactly according to the following procedure:*

$$\begin{cases} u^{k+1} = \text{argmin}_u\ \mathcal{L}_\sigma(u, v^k; x^k) + \langle \delta_g^k,\ u\rangle + \frac{\sigma}{2}\|u - u^k\|^2_{\mathcal{T}_f}, \\ v^{k+1} = \text{argmin}_v\ \mathcal{L}_\sigma(u^{k+1}, v; x^k) + \frac{\sigma}{2}\|v - v^k\|^2_{\mathcal{T}_g}, \\ x^{k+1} = x^k + \tau\sigma(\mathcal{F}^*u^{k+1} + \mathcal{G}^*v^{k+1} - c). \end{cases} \tag{42}$$

Equivalently, (u^{k+1}, v^{k+1}) can also be obtained exactly via:

$$\begin{cases} \bar{v}^k \quad = \text{argmin}_v\ \mathcal{L}_\sigma(u^k, v; x^k) + \frac{\sigma}{2}\|v - v^k\|^2_{\mathcal{T}_g}, \\ u^{k+1} = \text{argmin}_u\ \mathcal{L}_\sigma(u, \bar{v}^k; x^k) + \frac{\sigma}{2}\|u - u^k\|^2_{\mathcal{T}_f}, \\ v^{k+1} = \text{argmin}_v\ \mathcal{L}_\sigma(u^{k+1}, v; x^k) + \frac{\sigma}{2}\|v - v^k\|^2_{\mathcal{T}_g}, \\ x^{k+1} = x^k + \tau\sigma(\mathcal{F}^*u^{k+1} + \mathcal{G}^*v^{k+1} - c). \end{cases}$$

Proof The results follow directly from (36) and (37) in Proposition 2. □

Remark 2 (i) Note that comparing to (21) in Algorithm SPADMM, the first subproblem of (42) has an extra linear term $\langle \delta_g^k,\ \cdot\rangle$. It is this linear term that allows us to design a convergent SPADMM for solving multi-block convex optimization problems.

(ii) The linear term $\langle \delta_g^k,\ \cdot\rangle$ will vanish if $\Sigma_g = 0$, $\mathcal{E}_g = \mathcal{G}\mathcal{G}^* \succ 0$ and a proper starting point (u^0, v^0, x^0) is chosen. Specifically, if we choose $x^0 \in \mathcal{X}$ such that $\mathcal{G}x^0 = b$ and $(u^0, v^0) \in \text{dom}(f) \times \mathcal{V}$ such that $v^0 = \mathcal{E}_g^{-1}\mathcal{G}(c - \mathcal{F}^*u^0)$, then it holds that $\mathcal{G}x^k = b$ and $v^k = \mathcal{E}_g^{-1}\mathcal{G}(c - \mathcal{F}^*u^k)$, which imply that $\delta_g^k = 0$.

(iii) Observe that when $\mathcal{T}_f$ and $\mathcal{T}_g$ are chosen to be 0 in (42), apart from the range of τ, our Algorithm SCB-SPALM differs from the classical 2-block ADMM for solving problem (27) only in the linear term $\langle \delta_g^k,\ \cdot\rangle$. This shows that the classical 2-block ADMM for solving problem (27) has an unremovable deviation from the augmented Lagrangian method. This may explain why even when ADMM type methods suffer from slow local convergence, the latter can still enjoy fast local convergence.

In the following, we compare our Schur complement based proximal term $\frac{\sigma}{2}\|u - u^k\|^2_{\mathcal{T}_f} + \frac{\sigma}{2}\|v - v^k\|^2_{\mathcal{T}_g}$ used to derive the scheme (42) for solving (27) with the following proximal term which allows one to update u and v simultaneously:

$$\frac{\sigma}{2}(\|(u,v)-(u^k,v^k)\|^2_{\mathcal{M}}+\|u-u^k\|^2_{\mathcal{T}_f}+\|v-v^k\|^2_{\mathcal{T}_g}) \quad \text{with} \tag{43}$$
$$\mathcal{M}=\begin{pmatrix} \mathcal{D}_1 & -\mathcal{F}\mathcal{G}^* \\ -\mathcal{G}\mathcal{F}^* & \mathcal{D}_2 \end{pmatrix} \succeq 0,$$

where $\mathcal{D}_1 : \mathcal{U} \to \mathcal{U}$ and $\mathcal{D}_2 : \mathcal{V} \to \mathcal{V}$ are two self-adjoint positive semidefinite linear operators satisfying

$$\mathcal{D}_1 \succeq \sqrt{(\mathcal{F}\mathcal{G}^*)(\mathcal{F}\mathcal{G}^*)^*} \quad \text{and} \quad \mathcal{D}_2 \succeq \sqrt{(\mathcal{G}\mathcal{F}^*)(\mathcal{G}\mathcal{F}^*)^*}.$$

A common naive choice will be $\mathcal{D}_1 = \lambda_{\max}\mathcal{I}_1$ and $\mathcal{D}_2 = \lambda_{\max}\mathcal{I}_2$ where $\lambda_{\max} = \|\mathcal{F}\mathcal{G}^*\|_2$, $\mathcal{I}_1 : \mathcal{U} \to \mathcal{U}$ and $\mathcal{I}_2 : \mathcal{V} \to \mathcal{V}$ are identity maps. Simple calculations show that the resulting semi-proximal augmented Lagrangian method generates $(u^{k+1}, v^{k+1}, x^{k+1})$ as follows:

$$\begin{cases} u^{k+1} = \operatorname{argmin}_u \mathcal{L}_\sigma(u, v^k; x^k) + \frac{\sigma}{2}\|u-u^k\|^2_{\mathcal{D}_1+\mathcal{T}_f}, \\ v^{k+1} = \operatorname{argmin}_v \mathcal{L}_\sigma(u^k, v; x^k) + \frac{\sigma}{2}\|v-v^k\|^2_{\mathcal{D}_2+\mathcal{T}_g}, \\ x^{k+1} = x^k + \tau\sigma(\mathcal{F}^*u^{k+1} + \mathcal{G}^*v^{k+1} - c). \end{cases} \tag{44}$$

To ensure that the subproblems in (44) are well defined, we may require the following sufficient conditions to hold:

$$\sigma^{-1}\Sigma_f + \mathcal{T}_f + \mathcal{F}\mathcal{F}^* + \mathcal{D}_1 \succ 0 \quad \text{and} \quad \sigma^{-1}\Sigma_g + \mathcal{T}_g + \mathcal{G}\mathcal{G}^* + \mathcal{D}_2 \succ 0.$$

Comparing the proximal terms used in (31) and (43), we can easily see that the difference is:

$$\|u-u^k\|^2_{\mathcal{F}\mathcal{G}^*\mathcal{E}_g^{-1}\mathcal{G}\mathcal{F}^*} \quad \text{vs.} \quad \|(u,v)-(u^k,v^k)\|^2_{\mathcal{M}}.$$

To simplify the comparison, we assume that

$$\mathcal{D}_1 = \sqrt{(\mathcal{F}\mathcal{G}^*)(\mathcal{F}\mathcal{G}^*)^*} \quad \text{and} \quad \mathcal{D}_2 = \sqrt{(\mathcal{G}\mathcal{F}^*)(\mathcal{G}\mathcal{F}^*)^*}.$$

By rescaling the equality constraint in (27) if necessary, we may also assume that $\|\mathcal{F}\| = 1$. Now, we have that

$$\mathcal{F}\mathcal{G}^*\mathcal{E}_g^{-1}\mathcal{G}\mathcal{F}^* \preceq \mathcal{F}\mathcal{F}^*$$

and

$$\|u-u^k\|^2_{\mathcal{F}\mathcal{G}^*\mathcal{E}_g^{-1}\mathcal{G}\mathcal{F}^*} \leq \|u-u^k\|^2_{\mathcal{F}\mathcal{F}^*} \leq \|u-u^k\|^2.$$

In contrast, we have

$$\begin{aligned}\|(u, v)-(u^k, v^k)\|_{\mathcal{M}}^2 &\le 2\left(\|u-u^k\|_{\mathcal{D}_1}^2+\|v-v^k\|_{\mathcal{D}_2}^2\right)\\ &\le 2\|\mathcal{F}\mathcal{G}^*\|\left(\|u-u^k\|^2+\|v-v^k\|^2\right)\\ &\le 2\|\mathcal{G}\|\left(\|u-u^k\|^2+\|v-v^k\|^2\right),\end{aligned}$$

which is larger than the former upper bound $\|u-u^k\|^2$ if $\|\mathcal{G}\| \ge 1/2$. Thus we can conclude safely that the proximal term $\|u-u^k\|^2_{\mathcal{F}\mathcal{G}^*\mathcal{E}_g^{-1}\mathcal{G}\mathcal{F}^*}$ can be potentially much smaller than $\|(u, v)-(u^k, v^k)\|^2_{\mathcal{M}}$ unless $\|\mathcal{G}\|$ is very small. In fact, as a simple extension to (43), for the general multi-block case, one can always design a similar proximal term $\mathcal{M}$ to obtain an algorithm with a Jacobian type decomposition.

The above mentioned upper bounds difference is of course due to the fact that the SCB semi-proximal augmented Lagrangian method takes advantage of the fact that g is assumed to be a convex quadratic function. However, the key difference lies in the fact that (44) is a splitting version of the semi-proximal augmented Lagrangian method with a Jacobi type decomposition, whereas Algorithm SCB-SPALM is a splitting version of semi-proximal augmented Lagrangian method with a Gauss-Seidel type decomposition. It is this fact that provides us with the key idea to design Schur complement based proximal terms for multi-block convex optimization problems in the next section.

3 A Schur complement based semi-proximal ADMM

In this section, we focus on the problem

$$\begin{aligned}&\min\ f(u)+\sum_{i=1}^{p}\theta_i(y_i)+g(v)+\sum_{j=1}^{q}\varphi_j(z_j)\\ &\text{s.t. } \mathcal{F}^*u+\sum_{i=1}^{p}\mathcal{A}_i^*y_i+\mathcal{G}^*v+\sum_{j=1}^{q}\mathcal{B}_j^*z_j=c\end{aligned} \tag{45}$$

with all θ_i and φ_j being assumed to be convex quadratic functions:

$$\theta_i(y_i)=\frac{1}{2}\langle y_i,\ \mathcal{P}_i y_i\rangle-\langle b_i,\ y_i\rangle, \qquad \varphi_j(z_j)=\frac{1}{2}\langle z_j,\ \mathcal{Q}_j z_j\rangle-\langle d_j,\ z_j\rangle,$$

for $i=1,\dots,p$ and $j=1,\dots,q$, where $\mathcal{P}_i$ and $\mathcal{Q}_j$ are given self-adjoint positive semidefinite linear operators. The dual of (45) is given by

$$\max\left\{-\langle c,\ x\rangle-f^*(-\mathcal{F}x)-\sum_{i=1}^{p}\theta_i^*(-\mathcal{A}_i x)-g^*(-\mathcal{G}x)-\sum_{j=1}^{q}\varphi_j^*(-\mathcal{B}_j x)\right\}, \tag{46}$$

which can equivalently be written as

$$\begin{aligned} &\min \langle c,\ x\rangle + f^*(s) + \sum_{i=1}^{p} \theta_i^*(r_i) + g^*(t) + \sum_{j=1}^{q} \varphi_j^*(w_j) \\ &\text{s.t. } \mathcal{F}x + s = 0, \quad \mathcal{A}_i x + r_i = 0, \quad i = 1, \ldots, p, \\ &\quad \mathcal{G}x + t = 0, \quad \mathcal{B}_j x + w_j = 0, \quad j = 1, \ldots, q. \end{aligned} \tag{47}$$

For $i = 1, \ldots, p$, let $\mathcal{E}_{\theta_i}$ be a self-adjoint positive definite linear operator on $\mathcal{Y}_i$ such that it is a majorization of $\sigma^{-1}\mathcal{P}_i + \mathcal{A}_i\mathcal{A}_i^*$, i.e.,

$$\mathcal{E}_{\theta_i} \succeq \sigma^{-1}\mathcal{P}_i + \mathcal{A}_i\mathcal{A}_i^*.$$

We choose $\mathcal{E}_{\theta_i}$ in a way that its inverse can be computed at a moderate cost. Define

$$\mathcal{T}_{\theta_i} := \mathcal{E}_{\theta_i} - \sigma^{-1}\mathcal{P}_i - \mathcal{A}_i\mathcal{A}_i^* \succeq 0, \quad i = 1, \ldots, p. \tag{48}$$

Note that for numerical efficiency, we need the self-adjoint positive semidefinite linear operator $\mathcal{T}_{\theta_i}$ to be as small as possible for each i. Similarly, for $j = 1, \ldots, q$, let $\mathcal{E}_{\varphi_j}$ be a self-adjoint positive definite linear operator on $\mathcal{Z}_j$ that majorizes $\sigma^{-1}\mathcal{Q}_j + \mathcal{B}_j\mathcal{B}_j^*$ in a way that $\mathcal{E}_{\varphi_j}^{-1}$ can be computed relatively easily. Denote

$$\mathcal{T}_{\varphi_j} := \mathcal{E}_{\varphi_j} - \sigma^{-1}\mathcal{Q}_j - \mathcal{B}_j\mathcal{B}_j^* \succeq 0, \quad j = 1, \ldots, q. \tag{49}$$

Again, we need the self-adjoint positive semidefinite linear operator $\mathcal{T}_{\varphi_j}$ to be as small as possible for each j.

For notational convenience, we define

$$y_{\le i} := (y_1, y_2, \ldots, y_i), \quad y_{\ge i} := (y_i, y_{i+1}, \ldots, y_p), \quad i = 0, \ldots, p+1$$

with the convention that $y_0 = y_{p+1} = y_{\le 0} = y_{\ge p+1} = \emptyset$. For $i = 1, \ldots, p$, define the linear operator $\mathcal{A}_{\le i} : \mathcal{X} \to \mathcal{Y}$ by

$$\begin{pmatrix} \mathcal{A}_1 x \\ \mathcal{A}_2 x \\ \vdots \\ \mathcal{A}_i x \end{pmatrix} \equiv \mathcal{A}_{\le i} x \ := \ \mathcal{A}_1 x \times \mathcal{A}_2 x \ldots \times \mathcal{A}_i x \quad \forall x \in \mathcal{X}.$$

In a similar manner, we can define $z_{\le j}, z_{\ge j}$ for $j = 0, \ldots, q+1$ and define the linear operator $\mathcal{B}_{\le j}$ for $j = 1, \ldots, q$. Note that by definition, we have $y = y_{\le p}$, $z = z_{\le q}$, $\mathcal{A} = \mathcal{A}_{\le p}$ and $\mathcal{B} = \mathcal{B}_{\le q}$.

Define the affine function $\Gamma : \mathcal{U} \times \mathcal{Y} \times \mathcal{V} \times \mathcal{Z} \to \mathcal{X}$ by

$$\Gamma(u, y, v, z) := \mathcal{F}^*u + \mathcal{A}^*y + \mathcal{G}^*v + \mathcal{B}^*z - c \quad \forall (u, y, v, z) \in \mathcal{U} \times \mathcal{Y} \times \mathcal{V} \times \mathcal{Z}.$$

Let $\sigma > 0$ be given. The augmented Lagrangian function associated with (45) is given as follows:

$$\mathcal{L}_\sigma(u, y, v, z; x) = f(u) + \theta(y) + g(v) + \varphi(z) + \langle x,\ \Gamma(u, y, v, z)\rangle + \frac{\sigma}{2}\|\Gamma(u, y, v, z)\|^2$$

where $\theta(y) = \sum_{i=1}^{p} \theta_i(y_i)$ and $\varphi(z) = \sum_{j=1}^{q} \varphi_j(z_j)$.

Now we are ready to present our SCB-SPADMM (Schur complement based semi-proximal alternating direction method of multipliers) algorithm for solving (45).

Algorithm SCB-SPADMM: A Schur complement based SPADMM for solving (45).

Let $\sigma > 0$ and $\tau \in (0, \infty)$ be given parameters. Let $\mathcal{T}_f$ and $\mathcal{T}_g$ be given self-adjoint positive semidefinite operators defined on $\mathcal{U}$ and $\mathcal{V}$ respectively. Choose $(u^0, y^0, v^0, z^0, x^0) \in \mathrm{dom}(f) \times \mathcal{Y} \times \mathrm{dom}(g) \times \mathcal{Z} \times \mathcal{X}$. For $k = 0, 1, 2, \ldots$, generate $(u^{k+1}, y^{k+1}, v^{k+1}, z^{k+1})$ and x^{k+1} according to the following iteration.

Step 1. Compute for $i = p, \ldots, 1$,

$$\overline{y}_i^k \;=\; \mathrm{argmin}_{y_i}\ \mathcal{L}_\sigma(u^k, (y^k_{\le i-1}, y_i, \overline{y}^k_{\ge i+1}), v^k, z^k; x^k) + \frac{\sigma}{2}\|y_i - y_i^k\|^2_{\mathcal{T}_{\theta_i}},$$

where $\mathcal{T}_{\theta_i}$ is defined as in (48). Then compute

$$u^{k+1} = \mathrm{argmin}_u\ \mathcal{L}_\sigma(u, \overline{y}^k, v^k, z^k; x^k) + \frac{\sigma}{2}\|u - u^k\|^2_{\mathcal{T}_f}.$$

Step 2. Compute for $i = 1, \ldots, p$,

$$y_i^{k+1} \;=\; \mathrm{argmin}_{y_i}\ \mathcal{L}_\sigma(u^{k+1}, (y^{k+1}_{\le i-1}, y_i, \overline{y}^k_{\ge i+1}), v^k, z^k; x^k) + \frac{\sigma}{2}\|y_i - y_i^k\|^2_{\mathcal{T}_{\theta_i}}.$$

Step 3. Compute for $j = q, \ldots, 1$,

$$\overline{z}_j^k \;=\; \mathrm{argmin}_{z_j}\ \mathcal{L}_\sigma(u^{k+1}, y^{k+1}, v^k, (z^k_{\le j-1}, z_j, \overline{z}^k_{\ge j+1}); x^k) + \frac{\sigma}{2}\|z_j - z_j^k\|^2_{\mathcal{T}_{\varphi_j}},$$

where $\mathcal{T}_{\varphi_j}$ is defined as in (49). Then compute

$$v^{k+1} = \mathrm{argmin}_v\ \mathcal{L}_\sigma(u^{k+1}, y^{k+1}, v, \overline{z}^k; x^k) + \frac{\sigma}{2}\|v - v^k\|^2_{\mathcal{T}_g}.$$

Step 4. Compute for $j = 1, \ldots, q$,

$$z_j^{k+1} \;=\; \mathrm{argmin}_{z_j} \left\{ \begin{array}{l} \mathcal{L}_\sigma(u^{k+1}, y^{k+1}, v^{k+1}, (z^{k+1}_{\le j-1}, z_j, \overline{z}^k_{\ge j+1}); x^k) \\ +\frac{\sigma}{2}\|z_j - z_j^k\|^2_{\mathcal{T}_{\varphi_j}} \end{array} \right\}.$$

Step 5. Compute

$$x^{k+1} = x^k + \tau\sigma(\mathcal{F}^* u^{k+1} + \mathcal{A}^* y^{k+1} + \mathcal{G}^* v^{k+1} + \mathcal{B}^* z^{k+1} - c).$$

In order to prove the convergence of Algorithm SCB-SPADMM for solving (45), we need first to study the relationship between SCB-SPADMM and the generic 2-block semi-proximal ADMM for solving a two-block convex optimization problem discussed in the previous section.

Define for $l = 1, \ldots, p$,

$$f_1(u) := f(u), \quad f_{l+1}(u, y_{\le l}) := f(u) + \sum_{i=1}^{l} \theta_i(y_i) \quad \forall\, (u, y_{\le l}) \in \mathcal{U} \times \mathcal{Y}_{\le l},$$

where $\mathcal{Y}_{\le l} = \mathcal{Y}_1 \times \mathcal{Y}_2 \times \ldots \times \mathcal{Y}_l$. Similarly, for $l = 1, \ldots, q$, define $\mathcal{Z}_{\le l} = \mathcal{Z}_1 \times \mathcal{Z}_2 \times \ldots \times \mathcal{Z}_l$, and

$$g_1(v) := g(v), \quad g_{l+1}(v, z_{\le l}) := g(v) + \sum_{j=1}^{l} \varphi_j(z_j) \quad \forall\, (v, z_{\le l}) \in \mathcal{V} \times \mathcal{Z}_{\le l}.$$

Denote $\mathcal{A}_0^* \equiv \mathcal{F}_1^* \equiv \mathcal{F}^*$ and $\mathcal{B}_0^* \equiv \mathcal{G}_1^* \equiv \mathcal{G}^*$. Let

$$\mathcal{F}_{i+1}^* = \left(\mathcal{F}^*, \mathcal{A}_1^*, \ldots, \mathcal{A}_i^*\right), \qquad \mathcal{G}_{j+1}^* = \left(\mathcal{G}^*, \mathcal{B}_1^*, \ldots, \mathcal{B}_j^*\right),$$

for $i = 1, \ldots, p$ and $j = 1, \ldots, q$. Define the following self-adjoint linear operators: $\widehat{\mathcal{T}}_{f_1} := \mathcal{T}_f + \mathcal{F}_1 \mathcal{A}_1^* \mathcal{E}_{\theta_1}^{-1} \mathcal{A}_1 \mathcal{F}_1^*$,

$$\widehat{\mathcal{T}}_{f_i} := \begin{pmatrix} \widehat{\mathcal{T}}_{f_{i-1}} & \\ & \mathcal{T}_{\theta_{i-1}} \end{pmatrix} + \mathcal{F}_i \mathcal{A}_i^* \mathcal{E}_{\theta_i}^{-1} \mathcal{A}_i \mathcal{F}_i^*, \qquad i = 2, \ldots, p \tag{50}$$

and $\widehat{\mathcal{T}}_{g_1} := \mathcal{T}_g + \mathcal{G}_1 \mathcal{B}_1^* \mathcal{E}_{\varphi_1}^{-1} \mathcal{B}_1 \mathcal{G}_1^*$,

$$\widehat{\mathcal{T}}_{g_j} := \begin{pmatrix} \widehat{\mathcal{T}}_{g_{j-1}} & \\ & \mathcal{T}_{\varphi_{j-1}} \end{pmatrix} + \mathcal{G}_j \mathcal{B}_j^* \mathcal{E}_{\varphi_j}^{-1} \mathcal{B}_j \mathcal{G}_j^*, \qquad j = 2, \ldots, q. \tag{51}$$

Let $(\bar{v}, \bar{z}, \bar{x}, c) \in \mathcal{V} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{X}$ be given. Denote

$$\bar{c} := c - \mathcal{G}^* \bar{v} - \mathcal{B}^* \bar{z} \quad \text{and} \quad \bar{\gamma} := -\Gamma(\bar{u}, \bar{y}, \bar{v}, \bar{z}).$$

Define

$$\beta_{p,j} := \mathcal{A}_{j-1} \mathcal{A}_p^* \mathcal{E}_{\theta_p}^{-1} (b_p - \mathcal{A}_p \bar{x} - \mathcal{P}_p \bar{y}_p + \sigma \mathcal{A}_p \bar{\gamma}), \quad j = 1, \ldots, p \tag{52}$$

and for $i = p-1, \ldots, 1$,

$$\beta_{i,j} := \mathcal{A}_{j-1} \mathcal{A}_i^* \mathcal{E}_{\theta_i}^{-1} \left(b_i - \sum_{k=i+1}^{p} \beta_{k,i+1} - \mathcal{A}_i \bar{x} - \mathcal{P}_i \bar{y}_i + \sigma \mathcal{A}_i \bar{\gamma} \right), \quad j = 1, \ldots, i.$$

Let

$$\bar{\delta}_\theta := \sum_{i=1}^{p} \beta_{i,1}. \tag{53}$$

We will show later in Proposition 4 that $\bar{\delta}_\theta$ is the auxiliary linear term associated with problem (45). Recall that

$$\mathcal{L}_\sigma(u, y, \bar{v}, \bar{z}; \bar{x}) = f(u) + \theta(y) + g(\bar{v}) + \varphi(\bar{z}) + \langle \bar{x},\ \Gamma(u, y, \bar{v}, \bar{z}) \rangle + \frac{\sigma}{2} \|\Gamma(u, y, \bar{v}, \bar{z})\|^2.$$

For $i = p, \dots, 1$, let $y_i' \in \mathcal{Y}_i$ be defined by

$$\begin{aligned} y_i' &:= \operatorname{argmin}_{y_i}\ \mathcal{L}_\sigma(\bar{u}, (\bar{y}_{\le i-1}, y_i, y'_{\ge i+1}), \bar{v}, \bar{z}; \bar{x}) + \frac{\sigma}{2}\|y_i - \bar{y}_i\|^2_{\mathcal{T}_{\theta_i}} \\ &= \mathcal{E}_{\theta_i}^{-1}\big(\sigma^{-1} b_i - \sigma^{-1}\mathcal{A}_i \bar{x} + \mathcal{T}_{\theta_i}\bar{y}_i + \mathcal{A}_i\mathcal{A}_i^*\bar{y}_i - \mathcal{A}_i \Gamma(\bar{u}, (\bar{y}_{\le i-1}, \bar{y}_i, y'_{\ge i+1}), \bar{v}, \bar{z})\big) \end{aligned} \tag{54}$$

with the convention $y'_{p+1} = \emptyset$. Define $(u^+, y^+) \in \mathcal{U} \times \mathcal{Y}$ by

$$(u^+, y^+) := \operatorname{argmin}_{u,y} \left\{ \begin{array}{l} \mathcal{L}_\sigma(u, y, \bar{v}, \bar{z}; \bar{x}) + \frac{\sigma}{2}\|(u, y_{\le p-1}) - (\bar{u}, \bar{y}_{\le p-1})\|^2_{\mathcal{T}_{f_p}} \\ + \frac{\sigma}{2}\|y_p - \bar{y}_p\|^2_{\mathcal{T}_{\theta_p}} \end{array} \right\}. \tag{55}$$

The following proposition about two other equivalent procedures for computing (u^+, y^+) is the key ingredient for our algorithmic developments. The idea of proving this proposition is very simple: use Proposition 2 repeatedly though the proof itself is rather lengthy due to the multi-layered nature of the problems involved. For (55), we first express y_p as a function of $(u, y_{\le p-1})$ to obtain a problem involving only $(u, y_{\le p-1})$, and from the resulting problem, express y_{p-1} as a function of $(u, y_{\le p-2})$ to get another problem involving only $(u, y_{\le p-2})$. We continue this way until we get a problem involving only (u, y_1).

Proposition 4 *The optimal solution (u^+, y^+) defined by (55) can be obtained exactly by*

$$\begin{cases} u^+ = \operatorname{argmin}_u\ \mathcal{L}_\sigma(u, \bar{y}, \bar{v}, \bar{z}; \bar{x}) + \langle \bar{\delta}_\theta,\ u\rangle + \frac{\sigma}{2}\|u - \bar{u}\|^2_{\mathcal{T}_f}, \\ y_i^+ = \operatorname{argmin}_{y_i}\ \mathcal{L}_\sigma(u^+, (y^+_{\le i-1}, y_i, y'_{\ge i+1}), \bar{v}, \bar{z}; \bar{x}) + \frac{\sigma}{2}\|y_i - \bar{y}_i\|^2_{\mathcal{T}_{\theta_i}},\ i = 1, \dots, p \end{cases} \tag{56}$$

where the auxiliary linear term $\bar{\delta}_\theta$ is defined by (53). Furthermore, (u^+, y^+) can also be generated by the following equivalent procedure

$$\begin{cases} u^+ = \operatorname{argmin}_u\ \mathcal{L}_\sigma(u, y', \bar{v}, \bar{z}; \bar{x}) + \frac{\sigma}{2}\|u - \bar{u}\|^2_{\mathcal{T}_f}, \\ y_i^+ = \operatorname{argmin}_{y_i}\ \mathcal{L}_\sigma(u^+, (y^+_{\le i-1}, y_i, y'_{\ge i+1}), \bar{v}, \bar{z}; \bar{x}) + \frac{\sigma}{2}\|y_i - \bar{y}_i\|^2_{\mathcal{T}_{\theta_i}},\ i=1,\dots,p. \end{cases} \tag{57}$$

Proof We will separate our proof into two parts and for each part we prove our conclusions by induction.

Part one. In this part we show that (u^+, y^+) defined by (55) can be obtained exactly by (56). For the case $p = 1$, this follows directly from Proposition 2.

Assume that the equivalence between (55) and (56) holds for all $p \le l$. We need to show that for $p = l + 1$, this equivalence also holds. For this purpose, we consider the following optimization problem with respect to $(u, y_{\le l})$ and y_{l+1}:

$$\begin{aligned} &\min\ f_{l+1}(u, y_{\le l}) + \theta_{l+1}(y_{l+1}) + g(\bar{v}) + \varphi(\bar{z}) \\ &\text{s.t. } \mathcal{F}_{l+1}^*(u, y_{\le l}) + \mathcal{A}_{l+1}^* y_{l+1} = \bar{c}. \end{aligned} \tag{58}$$

The augmented Lagrangian function associated with problem (58) is given by

$$\begin{aligned} \mathcal{L}_\sigma^{l+1}((u, y_{\le l}), y_{l+1}; \bar{v}, \bar{z}, x) = &\ f_{l+1}(u, y_{\le l}) + \theta_{l+1}(y_{l+1}) + g(\bar{v}) + \varphi(\bar{z}) \\ &+ \langle x,\ \Gamma(u, y, \bar{v}, \bar{z})\rangle + \frac{\sigma}{2}\|\Gamma(u, y, \bar{v}, \bar{z})\|^2. \end{aligned} \tag{59}$$

We denote the vector $\delta_{\theta_{l+1}}$ as the auxiliary linear term associated with problem (58) by

$$\delta_{\theta_{l+1}} := \mathcal{F}_{l+1}\mathcal{A}_{l+1}^*\mathcal{E}_{\theta_{l+1}}^{-1}(b_{l+1} - \mathcal{A}_{l+1}\bar{x} - \mathcal{P}_{l+1}\bar{y}_{l+1} + \sigma\mathcal{A}_{l+1}\bar{\gamma}). \tag{60}$$

Note that by the definition of $\mathcal{F}_{l+1}$ and $p = l + 1$, we have

$$\langle \delta_{\theta_p},\ (u, y_{\le l})\rangle = \langle \beta_{p,1},\ u\rangle + \sum_{j=1}^{l}\langle \beta_{p,j+1},\ y_j\rangle$$

with $\beta_{p,j}$, $j = 1, \ldots, l + 1$, defined as in (52).

By noting that $\mathcal{L}_\sigma^{l+1}((u, y_{\le l}), y_{l+1}; \bar{v}, \bar{z}, \bar{x}) = \mathcal{L}_\sigma(u, y_{\le l}, y_{l+1}, \bar{v}, \bar{z}; \bar{x})$, we can rewrite problem (55) for $p = l + 1$ equivalently as

$$\begin{aligned} &((u^+, y_{\le l}^+), y_{l+1}^+) \\ &= \operatorname{argmin}\left\{ \begin{array}{l} \mathcal{L}_\sigma^{l+1}((u, y_{\le l}), y_{l+1}; \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2}\|(u, y_{\le l}) - (\bar{u}, \bar{y}_{\le l})\|_{\widehat{\mathcal{T}}_{f_{l+1}}}^2 \\ + \frac{\sigma}{2}\|y_{l+1} - \bar{y}_{l+1}\|_{\mathcal{T}_{\theta_{l+1}}}^2 \end{array} \right\}. \end{aligned} \tag{61}$$

Then, from Proposition 2, we know that problem (61) is equivalent to

$$(u^+, y_{\le l}^+) = \operatorname{argmin}_{(u, y_{\le l})} \left\{ \begin{array}{l} \mathcal{L}_\sigma^{l+1}((u, y_{\le l}), \bar{y}_{l+1}; \bar{v}, \bar{z}, \bar{x}) + \langle \delta_{\theta_{l+1}},\ (u, y_{\le l})\rangle \\ + \frac{\sigma}{2}\|(u, y_{\le l-1}) - (\bar{u}, \bar{y}_{\le l-1})\|_{\widehat{\mathcal{T}}_{f_l}}^2 \\ + \frac{\sigma}{2}\|y_l - \bar{y}_l\|_{\mathcal{T}_{\theta_l}}^2 \end{array} \right\}, \tag{62}$$

$$y_{l+1}^+ = \operatorname{argmin}_{y_{l+1}}\ \mathcal{L}_\sigma^{l+1}((u^+, y_{\le l}^+), y_{l+1}; \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2}\|y_{l+1} - \bar{y}_{l+1}\|_{\mathcal{T}_{\theta_{l+1}}}^2. \tag{63}$$

By observing that $\mathcal{L}_\sigma^{l+1}((u^+, y_{\le l}^+), y_{l+1}; \bar{v}, \bar{z}, \bar{x}) = \mathcal{L}_\sigma(u^+, y_{\le l}^+, y_{l+1}, \bar{v}, \bar{z}; \bar{x})$, we know that problem (63) can equivalently be rewritten as

$$y_{l+1}^+ = \operatorname{argmin}_{y_{l+1}} \mathcal{L}_\sigma(u^+, y_{\le l}^+, y_{l+1}, \bar{v}, \bar{z}; \bar{x}) + \frac{\sigma}{2}\|y_{l+1} - \bar{y}_{l+1}\|^2_{\mathcal{T}_{\theta_{l+1}}}. \tag{64}$$

In order to apply our induction assumption to problem (62), we need to construct a corresponding optimization problem. Define for $i = 1, \ldots, l$, $\widetilde{b}_i := b_i - \beta_{p,i+1}$,

$$\widetilde{\theta}_i(y_i) := \theta_i(y_i) + \langle \beta_{p,i+1},\ y_i\rangle = \tfrac{1}{2}\langle y_i,\ \mathcal{P}_i y_i\rangle - \langle \widetilde{b}_i,\ y_i\rangle \quad \forall\, y_i \in \mathcal{Y}_i,$$

$\widetilde{f}_1(u) := f(u) + \langle \beta_{p,1},\ u\rangle$, and

$$\widetilde{f}_{i+1}(u, y_{\le i}) := \widetilde{f}_1(u) + \sum_{j=1}^{i} \widetilde{\theta}_j(y_j) \quad \forall\, (u, y_{\le i}) \in \mathcal{U} \times \mathcal{Y}_{\le i}.$$

We shall now consider the following optimization problem with respect to $(u, y_{\le l})$:

$$\begin{aligned} &\min\ \widetilde{f}_1(u) + \sum_{i=1}^{l} \widetilde{\theta}_i(y_i) + \theta_{l+1}(\bar{y}_{l+1}) + g(\bar{v}) + \varphi(\bar{z}) \\ &\text{s.t. } \mathcal{F}^*u + \mathcal{A}_{\le l}^* y_{\le l} \ = \ \bar{c} - \mathcal{A}_{l+1}^* \bar{y}_{l+1}. \end{aligned} \tag{65}$$

The augmented Lagrangian function for the problem (65) is defined by

$$\begin{aligned} \widetilde{\mathcal{L}}_\sigma(u, y_{\le l}; \bar{y}_{l+1}, \bar{v}, \bar{z}, x) = {} & \widetilde{f}_1(u) + \sum_{i=1}^{l} \widetilde{\theta}_i(y_i) + \theta_{l+1}(\bar{y}_{l+1}) + g(\bar{v}) + \varphi(\bar{z}) \\ & + \langle x,\ \Gamma(u, (y_{\le l}, \bar{y}_{l+1}), \bar{v}, \bar{z})\rangle + \frac{\sigma}{2}\|\Gamma(u, (y_{\le l}, \bar{y}_{l+1}), \bar{v}, \bar{z})\|^2. \end{aligned}$$

Define

$$\mathcal{T}_{\widetilde{\theta}_i} \equiv \mathcal{T}_{\theta_i} \quad \text{and} \quad \mathcal{T}_{\widetilde{f}_i} \equiv \mathcal{T}_{f_i}, \quad i = 1, \ldots, l.$$

By using the definitions of $\widetilde{\theta}_i$ and $\widetilde{f}_i$, $i = 1, \ldots, l$, we have

$$\mathcal{E}_{\widetilde{\theta}_i} \equiv \mathcal{E}_{\theta_i} \quad \text{and} \quad \widehat{\mathcal{T}}_{\widetilde{f}_i} \equiv \widehat{\mathcal{T}}_{f_i}, \quad i = 1, \ldots, l. \tag{66}$$

Therefore, problem (62) can equivalently be rewritten as

$$(u^+, y_{\le l}^+) = \operatorname{argmin}_{(u, y_{\le l})} \left\{ \begin{array}{l} \widetilde{\mathcal{L}}_\sigma(u, y_{\le l}; \bar{y}_{l+1}, \bar{v}, \bar{z}, \bar{x}) \\ + \frac{\sigma}{2}\|(u, y_{\le l-1}) - (\bar{u}, \bar{y}_{\le l-1})\|^2_{\mathcal{T}_{\widetilde{f}_l}} + \frac{\sigma}{2}\|y_l - \bar{y}_l\|^2_{\mathcal{T}_{\widetilde{\theta}_l}} \end{array} \right\}. \tag{67}$$

Define

$$\widetilde{\beta}_{l,j} := \mathcal{A}_{j-1}\mathcal{A}_l^*\mathcal{E}_{\widetilde{\theta}_l}^{-1}(\widetilde{b}_l - \mathcal{A}_l\bar{x} - \mathcal{P}_l\bar{y}_l + \sigma\mathcal{A}_l\bar{\gamma}), \quad j = 1, \ldots, l$$

and for $i = l-1, \ldots, 1$, and $j = 1, \ldots, i$,

$$\widetilde{\beta}_{i,j} := \mathcal{A}_{j-1}\mathcal{A}_i^*\mathcal{E}_{\widetilde{\theta}_i}^{-1}\left(\widetilde{b}_i - \sum_{k=i+1}^{l}\widetilde{\beta}_{k,i+1} - \mathcal{A}_i\bar{x} - \mathcal{P}_i\bar{y}_i + \sigma\mathcal{A}_i\bar{\gamma}\right).$$

The auxiliary linear term $\delta_{\widetilde{\theta}}$ associated with problem (67) is given by

$$\delta_{\widetilde{\theta}} := \sum_{i=1}^{l}\widetilde{\beta}_{i,1}. \tag{68}$$

We will show that for $i = l, \ldots, 1$,

$$\widetilde{\beta}_{i,j} = \beta_{i,j} \quad \forall\, j = 1, \ldots, i. \tag{69}$$

First, by using (66), we have for $j = 1, \ldots, l$ that

$$\begin{aligned}\widetilde{\beta}_{l,j} &= \mathcal{A}_{j-1}\mathcal{A}_l^*\mathcal{E}_{\widetilde{\theta}_l}^{-1}(\widetilde{b}_l - \mathcal{A}_l\bar{x} - \mathcal{P}_l\bar{y}_l + \sigma\mathcal{A}_l\bar{\gamma})\\ &= \mathcal{A}_{j-1}\mathcal{A}_l^*\mathcal{E}_{\theta_l}^{-1}(b_l - \beta_{l+1,l+1} - \mathcal{A}_l\bar{x} - \mathcal{P}_l\bar{y}_l + \sigma\mathcal{A}_l\bar{\gamma}) \;=\; \beta_{l,j}.\end{aligned}$$

That is, (69) holds for $i = l$ and $j = 1, \ldots, l$. Now assume that we have proven $\widetilde{\beta}_{i,j} = \beta_{i,j}$ for all $i \geq k+1$ with $k+1 \leq l$ and $j = 1, \ldots, i$. We shall next prove that (69) holds for $i = k$ and $j = 1, \ldots, k$. Again, by using (66), we have for $j = 1, \ldots, k$ that

$$\begin{aligned}\widetilde{\beta}_{k,j} &= \mathcal{A}_{j-1}\mathcal{A}_k^*\mathcal{E}_{\widetilde{\theta}_k}^{-1}\left(\widetilde{b}_k - \sum_{s=k+1}^{l}\widetilde{\beta}_{s,k+1} - \mathcal{A}_k\bar{x} - \mathcal{P}_k\bar{y}_k + \sigma\mathcal{A}_k\bar{\gamma}\right)\\ &= \mathcal{A}_{j-1}\mathcal{A}_k^*\mathcal{E}_{\theta_k}^{-1}\left(b_k - \beta_{p,k+1} - \sum_{s=k+1}^{l}\beta_{s,k+1} - \mathcal{A}_k\bar{x} - \mathcal{P}_k\bar{y}_k + \sigma\mathcal{A}_k\bar{\gamma}\right)\\ &= \mathcal{A}_{j-1}\mathcal{A}_k^*\mathcal{E}_{\theta_k}^{-1}\left(b_k - \sum_{s=k+1}^{l+1}\beta_{s,k+1} - \mathcal{A}_k\bar{x} - \mathcal{P}_k\bar{y}_k + \sigma\mathcal{A}_k\bar{\gamma}\right) \;=\; \beta_{k,j},\end{aligned}$$

which, shows that (69) holds for $i = k$ and $j = 1, \ldots, k$. Thus, (69) is proven.

For $i = l, \ldots, 1$, define $\widetilde{y}'_i \in \mathcal{Y}_i$ by

$$\begin{aligned}\widetilde{y}'_i &:= \operatorname{argmin}_{y_i}\; \widetilde{\mathcal{L}}_\sigma(\bar{u}, (\bar{y}_{\leq i-1}, y_i, \widetilde{y}'_{\geq i+1}); \bar{y}_{l+1}, \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2}\|y_i - \bar{y}_i\|^2_{\mathcal{T}_{\widetilde{\theta}_i}},\\ &= \mathcal{E}_{\widetilde{\theta}_i}^{-1}\big(\sigma^{-1}\widetilde{b}_i - \sigma^{-1}\mathcal{A}_i\bar{x} + \mathcal{T}_{\widetilde{\theta}_i}\bar{y}_i + \mathcal{A}_i\mathcal{A}_i^*\bar{y}_i\\ &\quad - \mathcal{A}_i\varGamma(\bar{u}, (\bar{y}_{\leq i-1}, \bar{y}_i, \widetilde{y}'_{\geq i+1}, \bar{y}_{l+1}), \bar{v}, \bar{z})\big),\end{aligned}$$

where we use the convention $\widetilde{y}'_{l+1} = \emptyset$. We will prove that

$$\widetilde{y}'_i = y'_i \quad \forall\, i = l, \dots, 1. \tag{70}$$

We first calculate

$$\begin{aligned} y'_{l+1} - \bar{y}_{l+1} &= \mathcal{E}^{-1}_{\theta_{l+1}}(\sigma^{-1} b_{l+1} - \sigma^{-1}\mathcal{A}_{l+1}\bar{x} + \mathcal{T}_{\theta_{l+1}}\bar{y}_{l+1} + \mathcal{A}_{l+1}\mathcal{A}^*_{l+1}\bar{y}_{l+1} \\ &\quad + \mathcal{A}_{l+1}\bar{\gamma} - \mathcal{E}_{\theta_{l+1}}\bar{y}_{l+1}) \\ &= \mathcal{E}^{-1}_{\theta_{l+1}}(\sigma^{-1} b_{l+1} - \sigma^{-1}\mathcal{A}_{l+1}\bar{x} - \sigma^{-1}\mathcal{P}_{l+1}\bar{y}_{l+1} + \mathcal{A}_{l+1}\bar{\gamma}), \end{aligned}$$

which, together with the definitions of $\beta_{p,i}$ in (52), implies

$$\mathcal{A}_i\mathcal{A}^*_{l+1}(y'_{l+1} - \bar{y}_{l+1}) = \sigma^{-1}\beta_{p,i+1} \quad \forall\, i = 0, \dots, l. \tag{71}$$

Now, by using (66), (71) and the definitions of $\widetilde{y}'_l$ and y'_l, we have

$$\begin{aligned} y'_l - \widetilde{y}'_l &= \mathcal{E}^{-1}_{\theta_l}\big(\sigma^{-1}\beta_{p,l+1} + \mathcal{A}_l\mathcal{A}^*_{l+1}(\bar{y}_{l+1} - y'_{l+1})\big) \\ &= \mathcal{E}^{-1}_{\theta_l}(\sigma^{-1}\beta_{p,l+1} - \sigma^{-1}\beta_{p,l+1}) = 0. \end{aligned}$$

That is, (70) holds for $i = l$. Now assume that we have proven $\widetilde{y}'_i = y'_i$ for all $i \geq k+1$ with $k+1 \leq l$. We shall next prove that (70) holds for $i = k$. Again, by using the definitions of $\widetilde{y}'_k$ and y'_k and noting

$$\Gamma(\bar{u}, (\bar{y}_{\leq k}, \widetilde{y}'_{\geq k+1}, \bar{y}_{l+1}), \bar{v}, \bar{z}) - \Gamma(\bar{u}, (\bar{y}_{\leq k}, y'_{\geq k+1}), \bar{v}, \bar{z}) = \mathcal{A}^*_{l+1}(\bar{y}_{l+1} - y'_{l+1}),$$

we obtain that

$$\begin{aligned} y'_k - \widetilde{y}'_k &= \mathcal{E}^{-1}_{\theta_k}\big(\sigma^{-1}(b_k - \widetilde{b}_k) + \mathcal{A}_k\mathcal{A}^*_{l+1}(\bar{y}_{l+1} - y'_{l+1})\big) \\ &= \mathcal{E}^{-1}_{\theta_k}\big(\sigma^{-1}\beta_{p,k+1} + \mathcal{A}_k\mathcal{A}^*_{l+1}(\bar{y}_{l+1} - y'_{l+1})\big) \\ &= \mathcal{E}^{-1}_{\theta_l}(\sigma^{-1}\beta_{p,k+1} - \sigma^{-1}\beta_{p,k+1}) = 0, \end{aligned}$$

which, shows that (70) holds for $i = k$. Thus, (70) holds.

By applying our induction assumption to problem (67), we obtain equivalently that

$$u^+ = \operatorname{argmin}_u\ \widetilde{\mathcal{L}}_\sigma(u, \bar{y}_{\leq l}; \bar{y}_{l+1}, \bar{v}, \bar{z}, \bar{x}) + \langle \delta_{\widetilde{\theta}}, u\rangle + \frac{\sigma}{2}\|u - \bar{u}\|^2_{\mathcal{T}_f}, \tag{72}$$

$$y_i^+ = \operatorname{argmin}_{y_i}\ \widetilde{\mathcal{L}}_\sigma(u^+, (y^+_{\leq i-1}, y_i, \widetilde{y}'_{\geq i+1}); \bar{y}_{l+1}, \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2}\|y_i - \bar{y}_i\|^2_{\mathcal{T}_{\theta_i}} \tag{73}$$

for $i = 1, \dots, l$, where we use the facts that $\mathcal{T}_{\widetilde{f_1}} = \mathcal{T}_f$ and $\mathcal{T}_{\widetilde{\theta}_i} = \mathcal{T}_{\theta_i}$ for $i = 1, \dots, l$. By combining (69) and the definitions of $\bar{\delta}_\theta$ and $\delta_{\widetilde{\theta}}$ defined in (53) and (68), respectively, we derive that

$$\bar{\delta}_\theta = \sum_{i=1}^{l} \beta_{i,1} + \beta_{l+1,1} = \sum_{i=1}^{l} \widetilde{\beta}_{i,1} + \beta_{l+1,1} = \delta_{\widetilde{\theta}} + \beta_{l+1,1}. \tag{74}$$

By direct calculations,

$$\widetilde{\mathcal{L}}_\sigma(u, \bar{y}_{\le l}; \bar{y}_{l+1}, \bar{v}, \bar{z}, \bar{x}) = \mathcal{L}_\sigma(u, \bar{y}, \bar{v}, \bar{z}; \bar{x}) + \langle \beta_{l+1,1},\ u\rangle + \sum_{i=1}^{l} \langle \beta_{l+1,i+1},\ \bar{y}_i\rangle. \tag{75}$$

Using (70), (71) and the definition of $\widetilde{\mathcal{L}}_\sigma$, we have for $i = 1, \ldots, l$ that

$$\begin{aligned}
&\widetilde{\mathcal{L}}_\sigma(u^+, (y^+_{\le i-1}, y_i, \widetilde{y}'_{\ge i+1}); \bar{y}_{l+1}, \bar{v}, \bar{z}, \bar{x}) - \mathcal{L}_\sigma(u^+, (y^+_{\le i-1}, y_i, y'_{\ge i+1}), \bar{v}, \bar{z}; \bar{x}) \\
&\quad = \widetilde{\mathcal{L}}_\sigma(u^+, (y^+_{\le i-1}, y_i, y'_{i+1}, \ldots, y'_l); \bar{y}_{l+1}, \bar{v}, \bar{z}, \bar{x}) \\
&\qquad - \mathcal{L}_\sigma(u^+, (y^+_{\le i-1}, y_i, y'_{\ge i+1}), \bar{v}, \bar{z}; \bar{x}) \\
&\quad = \langle \beta_{p,i+1},\ y_i\rangle + \langle \sigma \mathcal{A}_i \mathcal{A}^*_{l+1}(\bar{y}_{l+1} - y'_{l+1}),\ y_i\rangle + c_i \\
&\quad = c_i,
\end{aligned} \tag{76}$$

where c_i is a constant term given by

$$\begin{aligned}
c_i &= \langle \beta_{l+1,1},\ u^+\rangle + \sum_{j=1}^{i-1} \langle \beta_{l+1,j+1},\ y^+_j\rangle + \sum_{j=i+1}^{l} \langle \beta_{l+1,j+1},\ y'_j\rangle \\
&\quad + \theta_{l+1}(\bar{y}_{l+1}) - \theta_{l+1}(y'_{l+1}) + \langle \bar{x},\ \mathcal{A}^*_{l+1}(\bar{y}_{l+1} - y'_{l+1})\rangle \\
&\quad + \frac{\sigma}{2}\langle \mathcal{A}^*_{l+1}(\bar{y}_{l+1} - y'_{l+1}), 2(\mathcal{F}^* u^+ + \mathcal{A}^*_{\le i-1} y^+_{\le i-1} \\
&\quad + \sum_{j=i+1}^{l} \mathcal{A}^*_j y'_j - \bar{c}) + \mathcal{A}^*_{l+1}(\bar{y}_{l+1} + y'_{l+1})\rangle.
\end{aligned}$$

Thus, by using (74), (75) and (76) we know that (72) and (73) can be rewritten as

$$\begin{cases}
u^+ = \operatorname{argmin}_u\ \mathcal{L}_\sigma(u, \bar{y}, \bar{v}, \bar{z}; \bar{x}) + \langle \bar{\delta}_\theta,\ u\rangle + \frac{\sigma}{2}\|u - \bar{u}\|^2_{\mathcal{T}_f}, \\
y^+_i = \operatorname{argmin}_{y_i}\ \mathcal{L}_\sigma(u^+, (y^+_{\le i-1}, y_i, y'_{\ge i+1}), \bar{v}, \bar{z}; \bar{x}) + \frac{\sigma}{2}\|y_i - \bar{y}_i\|^2_{\mathcal{T}_{\theta_i}}, \quad i = 1, \ldots, l,
\end{cases}$$

which, together with (64), shows that the equivalence between (55) and (56) holds for $p = l + 1$. The proof of this part is completed.

Part two. In this part, we prove the equivalence between (56) and (57). Again, for the case $p = 1$, it follows directly from Proposition 2.

Assume that the equivalence between (56) and (57) holds for all $p \le l$. We shall prove that this equivalence also holds for $p = l + 1$. Write $f_0(\cdot) \equiv f(\cdot) +$

$\sum_{i=1}^{l}\langle \beta_{i,1},\ \cdot\rangle$. Since f_0 differs from f only with an extra linear term, we define $\mathcal{T}_{f_0} \equiv \mathcal{T}_f$. In order to use Proposition 2, we consider the following optimization problem with respect to u and y_{l+1}:

$$\begin{aligned} &\min\ f_0(u) + \theta_{l+1}(y_{l+1}) + \sum_{i=1}^{l} \theta_i(\bar{y}_i) + g(\bar{v}) + \varphi(\bar{z}) \\ &\text{s.t. } \mathcal{F}^* u + \mathcal{A}_{l+1}^* y_{l+1} = \bar{c} - \mathcal{A}_{\leq l}^* \bar{y}_{\leq l}. \end{aligned} \tag{77}$$

The augmented Lagrangian function associated with problem (77) is given as follows:

$$\begin{aligned} \mathcal{L}_\sigma^0(u, y_{l+1}; \bar{y}_{\leq l}, \bar{v}, \bar{z}, x) = {} & f_0(u) + \theta_{l+1}(y_{l+1}) + \sum_{i=1}^{l} \theta_i(\bar{y}_i) + g(\bar{v}) + \varphi(\bar{z}) \\ & + \langle x,\ \Gamma(u, (\bar{y}_{\leq l}, y_{l+1}), \bar{v}, \bar{z})\rangle + \frac{\sigma}{2}\|\Gamma(u, (\bar{y}_{\leq l}, y_{l+1}), \bar{v}, \bar{z})\|^2. \end{aligned}$$

By observing that

$$\mathcal{L}_\sigma^0(u, \bar{y}_{l+1}; \bar{y}_{\leq l}, \bar{v}, \bar{z}, \bar{x}) = \mathcal{L}_\sigma(u, \bar{y}, \bar{v}, \bar{z}; \bar{x}) + \sum_{i=1}^{l}\langle \beta_{i,1},\ u\rangle \quad \text{and} \quad \mathcal{T}_{f_0} \equiv \mathcal{T}_f,$$

we can rewrite the first subproblem in (56) as

$$u^+ = \operatorname{argmin}_u\ \mathcal{L}_\sigma^0(u, \bar{y}_{l+1}; \bar{y}_{\leq l}, \bar{v}, \bar{z}, \bar{x}) + \langle \beta_{l+1,1},\ u\rangle + \frac{\sigma}{2}\|u - \bar{u}\|_{\mathcal{T}_{f_0}}^2. \tag{78}$$

By using the definition of y'_{l+1} given in (54), we have

$$y'_{l+1} = \mathcal{E}_{\theta_{l+1}}^{-1}\big(\sigma^{-1}(b_{l+1} - \mathcal{A}_{l+1}\bar{x}) + \mathcal{T}_{\theta_{l+1}}\bar{y}_{l+1} + \mathcal{A}_{l+1}\mathcal{A}_{l+1}^*\bar{y}_{l+1} + \mathcal{A}_{l+1}\bar{\gamma}\big). \tag{79}$$

Since

$$\mathcal{L}_\sigma^0(\bar{u}, y_{l+1}; \bar{y}_{\leq l}, \bar{v}, \bar{z}, \bar{x}) = \mathcal{L}_\sigma(\bar{u}, (\bar{y}_{\leq l}, y_{l+1}), \bar{v}, \bar{z}; \bar{x}) + \sum_{i=1}^{l}\langle \beta_{i,1},\ \bar{u}\rangle,$$

the point y'_{l+1} can be rewritten equivalently as

$$y'_{l+1} = \operatorname{argmin}_{y_{l+1}}\ \mathcal{L}_\sigma^0(\bar{u}, y_{l+1}; \bar{y}_{\leq l}, \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2}\|y_{l+1} - \bar{y}_{l+1}\|_{\mathcal{T}_{\theta_{l+1}}}^2. \tag{80}$$

Then, by applying Proposition 2 to problem (77) with respect to u and y_{l+1}, we know that problem (78) is equivalent to

$$u^+ = \operatorname{argmin}_u\ \mathcal{L}_\sigma^0(u, y'_{l+1}; \bar{y}_{\leq l}, \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2}\|u - \bar{u}\|_{\mathcal{T}_{f_0}}^2. \tag{81}$$

In order to apply our induction assumption to problem (81), we need to consider the following optimization problem with respect to $(u, y_{\le l})$:

$$\begin{aligned} &\min\ f(u) + \sum_{i=1}^{l} \theta_i(y_i) + \theta_{l+1}(y'_{l+1}) + g(\bar{v}) + \varphi(\bar{z}) \\ &\text{s.t. } \mathcal{F}^*(u) + \mathcal{A}^*_{\le l} y_{\le l} = \bar{c} - \mathcal{A}^*_{l+1} y'_{l+1}. \end{aligned} \tag{82}$$

The augmented Lagrangian function associated with problem (82) is given by

$$\begin{aligned} \widehat{\mathcal{L}}_\sigma(u, y_{\le l}; y'_{l+1}, \bar{v}, \bar{z}, x) = &\ f(u) + \sum_{i=1}^{l} \theta_i(y_i) + \theta_{l+1}(y'_{l+1}) + g(\bar{v}) + \varphi(\bar{z}) \\ &+ \langle x,\ \Gamma(u, (y_{\le l}, y'_{l+1}), \bar{v}, \bar{z})\rangle + \frac{\sigma}{2}\|\Gamma(u, (y_{\le l}, y'_{l+1}), \bar{v}, \bar{z})\|^2. \end{aligned}$$

Define

$$\widehat{\gamma} := -\Gamma(\bar{u}, (\bar{y}_{\le l}, y'_{l+1}), \bar{v}, \bar{z}) \quad \text{and} \quad h_i := b_i - \mathcal{A}_i \bar{x} - \mathcal{P}_i \bar{y}_i, \quad i = 1, \ldots, l.$$

For problem (82), we define the following associated terms

$$\widehat{\beta}_{l,j} := \mathcal{A}_{j-1}\mathcal{A}_l^* \mathcal{E}_{\theta_l}^{-1}(h_l + \sigma\mathcal{A}_l\widehat{\gamma}), \quad j = 1, \ldots, l$$

and for $i = l-1, \ldots, 1$,

$$\widehat{\beta}_{i,j} := \mathcal{A}_{j-1}\mathcal{A}_i^* \mathcal{E}_{\theta_i}^{-1}\left(h_i - \sum_{k=i+1}^{l} \widehat{\beta}_{k,i+1} + \sigma\mathcal{A}_i\widehat{\gamma}\right), \quad j = 1, \ldots, i.$$

The auxiliary linear term $\widehat{\delta}$ associated with problem (82) is given by

$$\widehat{\delta} = \sum_{i=1}^{l} \widehat{\beta}_{i,1}. \tag{83}$$

We will show that, for $i = l, \ldots, 1$,

$$\widehat{\beta}_{i,j} = \beta_{i,j} \quad \forall\, j = 1, \ldots, i. \tag{84}$$

Similar to what we have done in part one, we shall first prove that $\widehat{\beta}_{l,j} = \beta_{l,j}$ for $j = 1, \ldots, l$. In fact, for $j = 1, \ldots, l$, we have

$$\begin{aligned} \beta_{l,j} &= \mathcal{A}_{j-1}\mathcal{A}_l^* \mathcal{E}_{\theta_l}^{-1}(h_l - \beta_{l+1,l+1} + \sigma\mathcal{A}_l\bar{\gamma}) \\ &= \mathcal{A}_{j-1}\mathcal{A}_l^* \mathcal{E}_{\theta_l}^{-1}(h_l - \mathcal{A}_l\mathcal{A}^*_{l+1}\mathcal{E}^{-1}_{\theta_{l+1}}(h_{l+1} + \sigma\mathcal{A}_{l+1}\bar{\gamma}) + \sigma\mathcal{A}_l\bar{\gamma}) \\ &= \mathcal{A}_{j-1}\mathcal{A}_l^* \mathcal{E}_{\theta_l}^{-1}(h_l - \sigma\mathcal{A}_l\Gamma(\bar{u}, (\bar{y}_{\le l}, y'_{l+1}), \bar{v}, \bar{z})) \\ &= \mathcal{A}_{j-1}\mathcal{A}_l^* \mathcal{E}_{\theta_l}^{-1}(h_l + \sigma\mathcal{A}_l\widehat{\gamma}) = \widehat{\beta}_{l,j}, \end{aligned}$$

where the third equation follows from (79) and simple calculations. This shows that (84) holds for $i = l$ and $j = 1, \ldots, l$. Now we assume that $\widehat{\beta}_{i,j} = \beta_{i,j}$ for all $i \geq k+1$ with $k+1 \leq l$ and $j = 1, \ldots, i$. Next, we shall prove that (84) holds for $i = k$ and $j = 1, \ldots, k$. By direct calculations, we know for $j = 1, \ldots, k$ that

$$
\begin{aligned}
\beta_{k,j} &= \mathcal{A}_{j-1}\mathcal{A}_k^*\mathcal{E}_{\theta_k}^{-1}\left(h_k - \sum_{s=k+1}^{l+1} \beta_{s,k} + \sigma\mathcal{A}_k\bar{\gamma}\right) \\
&= \mathcal{A}_{j-1}\mathcal{A}_k^*\mathcal{E}_{\theta_k}^{-1}\left(h_k - \sum_{s=k+1}^{l} \widehat{\beta}_{s,k} - \beta_{l+1,k} + \sigma\mathcal{A}_k\bar{\gamma}\right) \\
&= \mathcal{A}_{j-1}\mathcal{A}_k^*\mathcal{E}_{\theta_k}^{-1}\left(h_k - \sum_{s=k+1}^{l} \widehat{\beta}_{s,k} - \mathcal{A}_k\mathcal{A}_{l+1}^*\mathcal{E}_{\theta_{l+1}}^{-1}(h_{l+1} + \sigma\mathcal{A}_{l+1}\bar{\gamma}) + \sigma\mathcal{A}_k\bar{\gamma}\right) \\
&= \mathcal{A}_{j-1}\mathcal{A}_k^*\mathcal{E}_{\theta_k}^{-1}\left(h_k - \sum_{s=k+1}^{l} \widehat{\delta}_{\theta_s,k} - \sigma\mathcal{A}_k\Gamma(\bar{u}, (\bar{y}_{\leq l}, y'_{l+1}), \bar{v}, \bar{z})\right) \\
&= \mathcal{A}_{j-1}\mathcal{A}_k^*\mathcal{E}_{\theta_k}^{-1}\left(h_k - \sum_{s=k+1}^{l} \widehat{\delta}_{\theta_s,k} + \sigma\mathcal{A}_k\widehat{\gamma}\right) = \widehat{\beta}_{k,j},
\end{aligned}
$$

which, shows that (84) holds for $i = k$ and $j = 1, \ldots, k$. Therefore, we have shown that (84) holds.

For $i = l, \ldots, 1$, define $\widehat{y}'_i \in \mathcal{Y}_i$ as

$$
\begin{aligned}
\widehat{y}'_i &= \operatorname{argmin}_{y_i}\ \widehat{\mathcal{L}}_\sigma(\bar{u}, (\bar{y}_{\leq i-1}, y_i, \widehat{y}'_{\geq i+1}); y'_{l+1}, \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2}\|y_i - \bar{y}_i\|^2_{\mathcal{T}_{\theta_i}} \\
&= \mathcal{E}_{\theta_i}^{-1}\big(\sigma^{-1}b_i - \sigma^{-1}\mathcal{A}_i\bar{x} + \mathcal{T}_{\theta_i}\bar{y}_i + \mathcal{A}_i\mathcal{A}_i^*\bar{y}_i \\
&\quad -\mathcal{A}_i\Gamma(\bar{u}, (\bar{y}_{\leq i-1}, \bar{y}_i, \widehat{y}'_{\geq i+1}, y'_{l+1}), \bar{v}, \bar{z})\big),
\end{aligned}
$$

where we use the convention $\widehat{y}'_{l+1} = \emptyset$. We will prove that

$$
\widehat{y}'_i = y'_i \quad \forall\, i = 1, \ldots, l. \tag{85}
$$

From (85), we know that

$$
\widehat{y}'_l = \mathcal{E}_{\theta_l}^{-1}\big(\sigma^{-1}b_l - \sigma^{-1}\mathcal{A}_l\bar{x} + \mathcal{T}_{\theta_l}\bar{y}_l + \mathcal{A}_l\mathcal{A}_l^*\bar{y}_l - \mathcal{A}_l\Gamma(\bar{u}, (\bar{y}_{\leq i-1}, \bar{y}_l, y'_{l+1}), \bar{v}, \bar{z})\big),
$$

which is exactly the same as y'_l defined in (54). This shows that (85) holds for $i = l$. Now we assume that $\widehat{y}'_i = y'_i$ for all $i \geq k+1$ with $k+1 \leq l$. Next, we shall prove that (85) holds for $i = k$. Again, by using the definition of $\widehat{y}'_k$ in (85) and the definition of y'_k in (54), we see that

$$
\begin{aligned}
\widehat{y}'_k &= \mathcal{E}_{\theta_k}^{-1}\begin{pmatrix} \sigma^{-1}b_k - \sigma^{-1}\mathcal{A}_k\bar{x} + \mathcal{T}_{\theta_k}\bar{y}_k + \mathcal{A}_k\mathcal{A}_k^*\bar{y}_k \\ -\mathcal{A}_k\Gamma(\bar{u}, (\bar{y}_{\le k-1}, \bar{y}_k, \widehat{y}'_{\ge k+1}, y'_{l+1}), \bar{v}, \bar{z}) \end{pmatrix} \\
&= \mathcal{E}_{\theta_k}^{-1}\big(\sigma^{-1}b_k - \sigma^{-1}\mathcal{A}_k\bar{x} + \mathcal{T}_{\theta_k}\bar{y}_k + \mathcal{A}_k\mathcal{A}_k^*\bar{y}_k - \mathcal{A}_k\Gamma(\bar{u}, (\bar{y}_{\le k-1}, \bar{y}_k, y'_{\ge k+1}), \bar{v}, \bar{z})\big) \\
&= y'_k.
\end{aligned}
$$

Thus, (85) is proven to be true.

By direct calculations, we obtain from (83) and (84) that

$$
\mathcal{L}_\sigma^0(u, y'_{l+1}; \bar{y}_{\le l}, \bar{v}, \bar{z}, \bar{x}) - \widehat{\mathcal{L}}_\sigma(u, \bar{y}_{\le l}; y'_{l+1}, \bar{v}, \bar{z}, \bar{x}) = \sum_{i=1}^{l}\langle \beta_{i,1},\ u\rangle \ = \ \langle \widehat{\delta},\ u\rangle. \tag{86}
$$

By using (86) and $\mathcal{T}_{f_0} \equiv \mathcal{T}_f$, we can reformulate problem (81) equivalently as

$$
u^+ = \operatorname{argmin}_u\ \widehat{\mathcal{L}}_\sigma(u, \bar{y}_{\le l}; y'_{l+1}, \bar{v}, \bar{z}, \bar{x}) + \langle \widehat{\delta},\ u\rangle + \frac{\sigma}{2}\|u - \bar{u}\|^2_{\mathcal{T}_f}. \tag{87}
$$

Then, from our induction assumption we know that problem (87) can be equivalently recast as,

$$
\begin{cases} \widehat{y}'_i &= \operatorname{argmin}_{y_i}\ \widehat{\mathcal{L}}_\sigma(\bar{u}, (\bar{y}_{\le i-1}, y_i, \widehat{y}'_{\ge i+1}); y'_{l+1}, \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2}\|y_i - \bar{y}_i\|^2_{\mathcal{T}_{\theta_i}},\ i = l, \ldots, 1 \\ u^+ &= \operatorname{argmin}_u\ \widehat{\mathcal{L}}_\sigma(u, \widehat{y}'_{\le l}; y'_{l+1}, \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2}\|u - \bar{u}\|^2_{\mathcal{T}_f}. \end{cases} \tag{88}
$$

By using (85) and observing

$$
\widehat{\mathcal{L}}_\sigma(u, y_{\le l}; y'_{l+1}, \bar{v}, \bar{z}, \bar{x}) = \mathcal{L}_\sigma(u, y_{\le l}, y'_{l+1}, \bar{v}, \bar{z}; \bar{x}),
$$

we know that (88) is equivalent to

$$
\begin{cases} y'_i &= \operatorname{argmin}_{y_i}\ \mathcal{L}_\sigma(\bar{u}, (\bar{y}_{\le i-1}, y_i, y'_{\ge i+1}), \bar{v}, \bar{z}; \bar{x}) + \frac{\sigma}{2}\|y_i - \bar{y}_i\|^2_{\mathcal{T}_{\theta_i}}, \quad i = l, \ldots, 1, \\ u^+ &= \operatorname{argmin}_u\ \mathcal{L}_\sigma(u, (y'_{\le l}, y'_{l+1}), \bar{v}, \bar{z}; \bar{x}) + \frac{\sigma}{2}\|u - \bar{u}\|^2_{\mathcal{T}_f}, \end{cases}
$$

which, together with (80), shows that the equivalence between (56) and (57) holds for $p = l + 1$. This completes the proof to the second part of this proposition. □

Proposition 5 *For any $k \ge 0$, the point $(x^{k+1}, y^{k+1}, v^{k+1}, z^{k+1})$ obtained by Algorithm SCB-SPADMM for solving problem (45) can be generated exactly according to the following iteration:*

$$
\begin{aligned}
(u^{k+1}, y^{k+1}) &= \operatorname{argmin}_{u,y} \left\{ \begin{array}{l} \mathcal{L}_\sigma(u, y, v^k, z^k; x^k) \\ +\frac{\sigma}{2}\|(u, y_{\le p-1}) - (u^k, y^k_{\le p-1})\|^2_{\widehat{\mathcal{T}}_{f_p}} + \frac{\sigma}{2}\|y_p - y_p^k\|^2_{\mathcal{T}_{\theta_p}} \end{array} \right\}, \\
(v^{k+1}, z^{k+1}) &= \operatorname{argmin}_{v,z} \left\{ \begin{array}{l} \mathcal{L}_\sigma(u^{k+1}, y^{k+1}, v, z; x^k) \\ +\frac{\sigma}{2}\|(v, z_{\le q-1}) - (v^k, z^k_{\le q-1})\|^2_{\widehat{\mathcal{T}}_{g_q}} + \frac{\sigma}{2}\|z_q - z_q^k\|^2_{\mathcal{T}_{\varphi_q}} \end{array} \right\}, \\
x^{k+1} &= x^k + \tau\sigma(\mathcal{F}^* u^{k+1} + \mathcal{A}^* y^{k+1} + \mathcal{G}^* v^{k+1} + \mathcal{B}^* z^{k+1} - c).
\end{aligned}
$$

Proof The (u^{k+1}, y^{k+1}) part directly follows from Proposition 4. The conclusion for the (v^{k+1}, z^{k+1}) part can be obtained in similar arguments to the part about (u^{k+1}, y^{k+1}). Hence, the required result follows. □

Write $\Sigma_{f_1} \equiv \Sigma_f$ and $\Sigma_{g_1} \equiv \Sigma_g$. Define

$$
\Sigma_{f_i} := \begin{pmatrix} \Sigma_{f_{i-1}} & \\ & \mathcal{P}_{i-1} \end{pmatrix}, \quad i = 2, \ldots, p+1
$$

and

$$
\Sigma_{g_j} := \begin{pmatrix} \Sigma_{g_{j-1}} & \\ & \mathcal{Q}_{j-1} \end{pmatrix}, \quad j = 2, \ldots, q+1.
$$

In order to prove the convergence of our algorithm SCB-SPADMM for solving problem (45), we need the following proposition.

Proposition 6 *It holds that*

$$
\mathcal{F}_{p+1}\mathcal{F}^*_{p+1} + \sigma^{-1}\Sigma_{f_{p+1}} + \begin{pmatrix} \widehat{\mathcal{T}}_{f_p} & \\ & \mathcal{T}_{\theta_p} \end{pmatrix} \succ 0 \Leftrightarrow \mathcal{F}\mathcal{F}^* + \sigma^{-1}\Sigma_f + \mathcal{T}_f \succ 0, \tag{89}
$$

$$
\mathcal{G}_{q+1}\mathcal{G}^*_{q+1} + \sigma^{-1}\Sigma_{g_{q+1}} + \begin{pmatrix} \widehat{\mathcal{T}}_{g_q} & \\ & \mathcal{T}_{\varphi_q} \end{pmatrix} \succ 0 \Leftrightarrow \mathcal{G}\mathcal{G}^* + \sigma^{-1}\Sigma_g + \mathcal{T}_g \succ 0. \tag{90}
$$

Proof We only need to prove (89) as (90) can be obtained in the similar manner. For $i = 3, \ldots, p+1$, we have

$$
\begin{aligned}
&\mathcal{F}_i\mathcal{F}_i^* + \sigma^{-1}\Sigma_{f_i} + \begin{pmatrix} \widehat{\mathcal{T}}_{f_{i-1}} & \\ & \mathcal{T}_{\theta_{i-1}} \end{pmatrix} \\
&= \begin{pmatrix} \mathcal{F}_{i-1}\mathcal{F}^*_{i-1} + \sigma^{-1}\Sigma_{f_{i-1}} + \widehat{\mathcal{T}}_{f_{i-1}} & \mathcal{F}_{i-1}\mathcal{A}^*_{i-1} \\ \mathcal{A}_{i-1}\mathcal{F}^*_{i-1} & \mathcal{A}_{i-1}\mathcal{A}^*_{i-1} + \sigma^{-1}\mathcal{P}_{i-1} + \mathcal{T}_{\theta_{i-1}} \end{pmatrix}.
\end{aligned}
$$

Since $\mathcal{E}_{\theta_{i-1}} = \mathcal{A}_{i-1}\mathcal{A}^*_{i-1} + \sigma^{-1}\mathcal{P}_{i-1} + \mathcal{T}_{\theta_{i-1}} \succ 0$ for all $i \ge 3$, by the Schur complement condition for ensuring the positive definiteness of linear operators, we have

$$
\begin{pmatrix} \mathcal{F}_{i-1}\mathcal{F}_{i-1}^* + \sigma^{-1}\Sigma_{f_{i-1}} + \widehat{\mathcal{T}}_{f_{i-1}} & \mathcal{F}_{i-1}\mathcal{A}_{i-1}^* \\ \mathcal{A}_{i-1}\mathcal{F}_{i-1}^* & \mathcal{E}_{\theta_{i-1}} \end{pmatrix} \succ 0
$$
$$\Updownarrow$$
$$
\mathcal{F}_{i-1}\mathcal{F}_{i-1}^* + \sigma^{-1}\Sigma_{f_{i-1}} + \widehat{\mathcal{T}}_{f_{i-1}} - \mathcal{F}_{i-1}\mathcal{A}_{i-1}^*\mathcal{E}_{\theta_{i-1}}^{-1}\mathcal{A}_{i-1}\mathcal{F}_{i-1}^* \succ 0
$$
$$\Updownarrow$$
$$
\mathcal{F}_{i-1}\mathcal{F}_{i-1}^* + \sigma^{-1}\Sigma_{f_{i-1}} + \begin{pmatrix} \widehat{\mathcal{T}}_{f_{i-2}} & \\ & \mathcal{T}_{\theta_{i-2}} \end{pmatrix} \succ 0.
$$

Therefore, by taking $i = 3$, we obtain that

$$
\mathcal{F}_{p+1}\mathcal{F}_{p+1}^* + \sigma^{-1}\Sigma_{f_{p+1}} + \begin{pmatrix} \widehat{\mathcal{T}}_{f_p} & \\ & \mathcal{T}_{\theta_p} \end{pmatrix} \succ 0 \Leftrightarrow \mathcal{F}_2\mathcal{F}_2^* + \sigma^{-1}\Sigma_{f_2} + \begin{pmatrix} \widehat{\mathcal{T}}_{f_1} & \\ & \mathcal{T}_{\theta_1} \end{pmatrix} \succ 0.
$$

Note that

$$
\mathcal{F}_2\mathcal{F}_2^* + \sigma^{-1}\Sigma_{f_2} + \begin{pmatrix} \widehat{\mathcal{T}}_{f_1} & \\ & \mathcal{T}_{\theta_1} \end{pmatrix} = \begin{pmatrix} \mathcal{F}_1\mathcal{F}_1^* + \sigma^{-1}\Sigma_{f_1} + \widehat{\mathcal{T}}_{f_1} & \mathcal{F}_1\mathcal{A}_1^* \\ \mathcal{A}_1\mathcal{F}_1^* & \mathcal{A}_1\mathcal{A}_1^* + \sigma^{-1}\mathcal{P}_1 + \mathcal{T}_{\theta_1} \end{pmatrix}.
$$

Since $\mathcal{E}_{\theta_1} = \mathcal{A}_1\mathcal{A}_1^* + \sigma^{-1}\mathcal{P}_1 + \mathcal{T}_{\theta_1} \succ 0$, again by the Schur complement condition for ensuring the positive definiteness of linear operators, we have

$$
\begin{pmatrix} \mathcal{F}_1\mathcal{F}_1^* + \sigma^{-1}\Sigma_{f_1} + \widehat{\mathcal{T}}_{f_1} & \mathcal{F}_1\mathcal{A}_1^* \\ \mathcal{A}_1\mathcal{F}_1^* & \mathcal{E}_{\theta_1} \end{pmatrix} \succ 0
$$
$$\Updownarrow$$
$$
\mathcal{F}_1\mathcal{F}_1^* + \sigma^{-1}\Sigma_{f_1} + \widehat{\mathcal{T}}_{f_1} - \mathcal{F}_1\mathcal{A}_1^*\mathcal{E}_{\theta_1}^{-1}\mathcal{A}_1\mathcal{F}_1^* \succ 0
$$
$$\Updownarrow$$
$$
\mathcal{F}\mathcal{F}^* + \sigma^{-1}\Sigma_f + \mathcal{T}_f \succ 0.
$$

Thus, we have

$$
\mathcal{F}_{p+1}\mathcal{F}_{p+1}^* + \sigma^{-1}\Sigma_{f_{p+1}} + \begin{pmatrix} \mathcal{T}_{f_p} & \\ & \mathcal{T}_{\theta_p} \end{pmatrix} \succ 0 \Leftrightarrow \mathcal{F}\mathcal{F}^* + \sigma^{-1}\Sigma_f + \mathcal{T}_f \succ 0.
$$

The proof of this proposition is completed. □

Note that in the context of the multi-block convex optimization problem (45), Assumption 2 takes the following form:

Assumption 4 There exists $(\hat{u}, \hat{y}, \hat{v}, \hat{z}) \in \operatorname{ri}(\operatorname{dom} f) \times \mathcal{Y} \times \operatorname{ri}(\operatorname{dom} g) \times \mathcal{Z}$ such that $\mathcal{F}^*\hat{u} + \mathcal{A}^*\hat{y} + \mathcal{G}^*\hat{v} + \mathcal{B}^*\hat{z} = c$.

After all these preparations, we can finally state our main convergence theorem.

Theorem 3 *Let Σ_f and Σ_g be the two self-adjoint and positive semidefinite operators defined by (24) and (25), respectively. Suppose that the solution set of problem (45) is nonempty and that Assumption 4 holds. Assume that $\mathcal{T}_f$ and $\mathcal{T}_g$ are chosen such that the sequence $\{(u^k, y^k, v^k, z^k, x^k)\}$ generated by Algorithm SCB-SPADMM is well* *defined. Recall that* $\mathcal{T}_{\theta_i}$ *is defined in (48) for* $1 \le i \le p$ *and* $\mathcal{T}_{\varphi_j}$ *is defined in (49) for* $1 \le j \le q$*. Then, under the condition either (a)* $\tau \in (0, (1+\sqrt{5})/2)$ *or (b)* $\tau \ge (1+\sqrt{5})/2$ *but* $\sum_{k=0}^{\infty}(\|\mathcal{G}^*(v^{k+1}-v^k)+\mathcal{B}^*(z^{k+1}-z^k)\|^2+\tau^{-1}\|\mathcal{F}^*u^{k+1}+\mathcal{A}^*y^{k+1}+\mathcal{G}^*v^{k+1}+\mathcal{B}^*z^{k+1}-c\|^2)<\infty$*, the following results hold:*

- (i) *If* $(u^\infty, y^\infty, v^\infty, z^\infty, x^\infty)$ *is an accumulation point of* $\{(u^k, y^k, v^k, z^k, x^k)\}$*, then* $(u^\infty, y^\infty, v^\infty, z^\infty)$ *solves problem (45) and* x^∞ *solves (48), respectively.*
- (ii) *If both* $\sigma^{-1}\Sigma_f+\mathcal{T}_f+\mathcal{F}\mathcal{F}^*$ *and* $\sigma^{-1}\Sigma_g+\mathcal{T}_g+\mathcal{G}\mathcal{G}^*$ *are positive definite, then the sequence* $\{(u^k, y^k, v^k, z^k, x^k)\}$*, which is automatically well defined, converges to a unique limit, say,* $(u^\infty, y^\infty, v^\infty, z^\infty, x^\infty)$ *with* $(u^\infty, y^\infty, v^\infty, z^\infty)$ *solving problem (45) and* x^∞ *solving (48), respectively.*
- (iii) *When the* u, y*-part disappears, the corresponding results in parts (i)–(ii) hold under the condition either* $\tau \in (0,2)$ *or* $\tau \ge 2$ *but* $\sum_{k=0}^{\infty}\|\mathcal{G}^*v^{k+1}+\mathcal{B}^*z^{k+1}-c\|^2<\infty$*.*

Proof By combining Theorem 1 with Propositions 5 and 6, we can readily obtain the conclusions of this theorem. □

Remark 3 Our SCB-SPADMM algorithm actually provides a potentially efficient approach to handle large-scale and dense linear constraints. When dealing with such difficult linear systems, instead of being trapped with the possible convergence issues brought about by inexact solvers such as conjugate gradient methods, one can always first decompose the large systems into several smaller pieces, and then apply our SCB-SPADMM algorithm to the decomposed problems. As a result, these smaller systems can always be handled by adding suitable proximal terms or by solving them exactly.

4 Numerical experiments

We first examine the optimality condition for the general problem (45) and its dual (46). Suppose that the solution set of problem (45) is nonempty and that Assumption 4 holds. Then in order that (u^*, y^*, v^*, z^*) be an optimal solution for (45) and x^* be an optimal solution for (46), it is necessary and sufficient that (u^*, y^*, v^*, z^*) and x^* satisfy

$$\begin{cases} \mathcal{F}^*u+\sum_{i=1}^{p}\mathcal{A}_i^*y_i+\mathcal{G}^*v+\sum_{j=1}^{q}\mathcal{B}_j^*z_j=c, \\ f(u)+f^*(-\mathcal{F}x)=\langle -\mathcal{F}x,\ u\rangle, \quad \theta_i(y_i)+\theta_i^*(-\mathcal{A}_ix)=\langle -\mathcal{A}_ix,\ y_i\rangle, \\ g(v)+g^*(-\mathcal{G}x)=\langle -\mathcal{G}x,\ v\rangle, \quad \varphi_i(z_i)+\varphi_i^*(-\mathcal{B}_ix)=\langle -\mathcal{B}_ix,\ z_i\rangle, \end{cases} \tag{91}$$

for $i=1,\ldots,p$, and $j=1,\ldots,q$. We will measure the accuracy of an approximate solution based on the above optimality condition. If the given problem is properly scaled, the following relative residual is a natural choice to be used in our stopping criterion:

$$\eta=\max\{\eta_P, \eta_f, \eta_g, \eta_\theta, \eta_\varphi\}, \tag{92}$$

where

$$
\begin{aligned}
\eta_P &= \frac{\|\mathcal{F}^*u + \mathcal{A}^*y + \mathcal{G}^*v + \mathcal{B}^*z - c\|}{1+\|c\|}, \quad \eta_f = \frac{\|u - \operatorname{Prox}_f(u - \mathcal{F}x)\|}{1+\|u\|+\|\mathcal{F}x\|},\\
\eta_g &= \frac{\|v - \operatorname{Prox}_g(v - \mathcal{G}x)\|}{1+\|v\|+\|\mathcal{G}x\|}, \quad \eta_\theta = \max_{i=1,\ldots,p} \frac{\|y_i - \operatorname{Prox}_{\theta_i}(y_i - \mathcal{A}_i x)\|}{1+\|y_i\|+\|\mathcal{A}_i x\|},\\
\eta_\varphi &= \max_{j=1,\ldots,q} \frac{\|z_j - \operatorname{Prox}_{\varphi_j}(z_j - \mathcal{B}_j x)\|}{1+\|z_j\|+\|\mathcal{B}_j x\|}.
\end{aligned}
$$

Additionally, we compute the relative gap by

$$
\eta_{\text{gap}} = \frac{\text{obj}_P - \text{obj}_D}{1+|\text{obj}_P|+|\text{obj}_D|},
$$

where $\text{obj}_P := f(u) + \sum_{i=1}^p \theta_i(y_i) + g(v) + \sum_{j=1}^q \varphi_j(z_j)$ and $\text{obj}_D := -\langle c, x\rangle - f^*(s) - \sum_{i=1}^p \theta_i^*(r_i) - g^*(t) - \sum_{j=1}^q \varphi_j^*(w_j)$. We test the following problem sets.

4.1 Numerical results for convex quadratic SDP

Consider the following QSDP problem

$$
\begin{aligned}
&\min \ \frac{1}{2}\langle X, \mathcal{Q}X\rangle + \langle C, X\rangle\\
&\text{s.t. } \mathcal{A}_E X = b_E, \quad \mathcal{A}_I X \geq b_I, \quad X \in \mathcal{S}_+^n \cap \mathcal{K}
\end{aligned} \tag{93}
$$

and its dual problem

$$
\begin{aligned}
&\max \ -\delta_{\mathcal{K}}^*(-Z) + \langle b_I, y_I\rangle - \frac{1}{2}\langle X', \mathcal{Q}X'\rangle + \langle b_E, y_E\rangle\\
&\text{s.t. } Z + \mathcal{A}_I^* y_I - \mathcal{Q}X' + S + \mathcal{A}_E^* y_E = C, \quad y_I \geq 0, \quad S \in \mathcal{S}_+^n.
\end{aligned} \tag{94}
$$

We use X' here to indicate the fact that X' can be different from the primal variable X. Despite this fact, we have that at the optimal point, $\mathcal{Q}X = \mathcal{Q}X'$. Since $\mathcal{Q}$ is only assumed to be a self-adjoint positive semidefinite linear operator, the augmented Lagrangian function associated with (94) may not be strongly convex with respect to X'. Without further adding a proximal term, we propose the following strategy to rectify this difficulty. Since $\mathcal{Q}$ is positive semidefinite, $\mathcal{Q}$ can be decomposed as $\mathcal{Q} = \mathcal{B}^*\mathcal{B}$ for some linear map $\mathcal{B}$. By introducing a new variable $\Xi = -\mathcal{B}X'$, the problem (94) can be rewritten as follows:

$$
\begin{aligned}
&\max \ -\delta_{\mathcal{K}}^*(-Z) + \langle b_I, y_I\rangle - \frac{1}{2}\|\Xi\|_F^2 + \langle b_E, y_E\rangle\\
&\text{s.t. } Z + \mathcal{A}_I^* y_I + \mathcal{B}^*\Xi + S + \mathcal{A}_E^* y_E = C, \quad y_I \geq 0, \quad S \in \mathcal{S}_+^n.
\end{aligned} \tag{95}
$$

Note that now the augmented Lagrangian function associated with (95) is strongly convex with respect to Ξ. Surprisingly, much to our delight, we can update the iterations in our SCB-SPADMM without explicitly computing $\mathcal{B}$ or $\mathcal{B}^*$. Given $\overline{Z}$, $\bar{y}_I$, $\overline{S}$, $\bar{y}_E$ and $\overline{X}$, denote

$$\begin{aligned}\Xi^+ &:= \operatorname{argmin}_{\Xi}\ \frac{1}{2}\|\Xi\|^2 + \frac{\sigma}{2}\|\overline{Z} + \mathcal{A}_I^*\bar{y}_I + \mathcal{B}^*\Xi + \overline{S} + \mathcal{A}_E^*\bar{y}_E - C + \sigma^{-1}\overline{X}\|^2 \\ &= -(\mathcal{I} + \sigma\mathcal{B}\mathcal{B}^*)^{-1}\mathcal{B}\overline{R},\end{aligned}$$

where $\overline{R} = \overline{X} + \sigma(\overline{Z} + \mathcal{A}_I^*\bar{y}_I + \overline{S} + \mathcal{A}_E^*\bar{y}_E - C)$. In updating the SCB-SPADMM iterations, we actually do not need Ξ^+ explicitly, but only need $\Upsilon^+ := -\mathcal{B}^*\Xi^+$. From the condition that $(\mathcal{I}+\sigma\mathcal{B}\mathcal{B}^*)(-\Xi^+) = \mathcal{B}\overline{R}$, we get $(\mathcal{I}+\sigma\mathcal{B}^*\mathcal{B})(-\mathcal{B}^*\Xi^+) = \mathcal{B}^*\mathcal{B}\overline{R}$, hence we can compute Υ^+ via $\mathcal{Q}$:

$$\Upsilon^+ = (\mathcal{I} + \sigma\mathcal{Q})^{-1}(\mathcal{Q}\overline{R}).$$

In fact, $\Upsilon := -\mathcal{B}^*\Xi$ can be viewed as the shadow of $\mathcal{Q}X'$. Meanwhile, for the function $\delta_{\mathcal{K}}^*(-Z)$, we have the following useful observation that for any $\lambda > 0$,

$$Z^+ = \operatorname{argmin}\ \delta_{\mathcal{K}}^*(-Z) + \frac{\lambda}{2}\|Z - \overline{Z}\|^2 = \overline{Z} + \frac{1}{\lambda}\Pi_{\mathcal{K}}(-\lambda\overline{Z}), \tag{96}$$

where (96) follows from the following Moreau decomposition:

$$x = \mathrm{Prox}_{\tau f^*}(x) + \tau\mathrm{Prox}_{f/\tau}(x/\tau), \quad \forall\,\tau > 0.$$

In our numerical experiments, we test QSDP problems without inequality constraints (i.e., $\mathcal{A}_I$ and b_I are vacuous). We consider first the linear operator $\mathcal{Q}$ given by $\mathcal{Q}(X) = \frac{1}{2}(BX + XB)$ for a given matrix $B \in \mathcal{S}_+^n$. Suppose that we have the eigenvalue decomposition $B = P\Lambda P^T$, where $\Lambda = \mathrm{diag}(\lambda)$ and $\lambda = (\lambda_1, \ldots, \lambda_n)^T$ is the vector of eigenvalues of B. Then

$$\begin{aligned}\langle X,\ \mathcal{Q}X\rangle &= \frac{1}{2}\langle\widehat{X},\ \Lambda\widehat{X} + \widehat{X}\Lambda\rangle = \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \widehat{X}_{ij}^2(\lambda_i + \lambda_j) \\ &= \sum_{i=1}^n\sum_{j=1}^n \widehat{X}_{ij}^2 H_{ij}^2 = \langle X,\ \mathcal{B}^*\mathcal{B}X\rangle,\end{aligned}$$

where $\widehat{X} = P^TXP$, $H_{ij} = \sqrt{\frac{\lambda_i+\lambda_j}{2}}$, $\mathcal{B}X = H\circ(P^TXP)$ and $\mathcal{B}^*\Xi = P(H\circ\Xi)P^T$. In our numerical experiments, the matrix B is a low rank random symmetric positive semidefinite matrix. Note that when $\mathrm{rank}(B) = 0$ and $\mathcal{K}$ is a polyhedral cone, problem (93) reduces to the SDP problem considered in [21]. In our experiments, we test both the cases where $\mathrm{rank}(B) = 5$ and $\mathrm{rank}(B) = 10$. All the linear constraints are extracted from the numerical test examples in [21] (Sect. 4.1). For instance, we construct QSDP-BIQ problem sets based on the formulation in [21] as follows:

$$\begin{aligned} &\min\ \frac{1}{2}\langle X,\ \mathcal{Q}X\rangle + \frac{1}{2}\langle Q,\ X_0\rangle + \langle c,\ x\rangle \\ &\text{s.t. } \operatorname{diag}(X_0) - x = 0, \quad \alpha = 1, \\ &\quad X = \begin{pmatrix} X_0 & x \\ x^T & \alpha \end{pmatrix} \in \mathcal{S}^n_+, \quad X \in \mathcal{K} := \{X \in \mathcal{S}^n :\ X \geq 0\}. \end{aligned}$$

In our numerical experiments, the test data for Q and c are taken from Biq Mac Library maintained by Wiegele, which is available at http://biqmac.uni-klu.ac.at/biqmaclib.html. In the same sprit, we construct test problems QSDP-BIQ, QSDP-θ_+, QSDP-QAP and QSDP-RCP.

Here we compare our algorithm SCB- SPADMM with the directly extended ADMM (with step length $\tau = 1$) and the convergent alternating direction method with a Gaussian back substitution proposed in [25] (we call the method ADMMGB here and use the parameter $\alpha = 0.99$ in the Gaussian back substitution step). We have implemented all the algorithms SCB- SPADMM, ADMM and ADMMGB in MATLAB version 7.13. The numerical results reported later are obtained from a PC with 24 GB memory and 2.80GHz dual-core CPU running on 64-bit Windows Operating System.

We measure the accuracy of an approximate optimal solution (X, Z, Ξ, S, y_E) for QSDP (93) and its dual (95) by using the following relative residual obtained from the general optimality condition (91):

$$\eta_{\text{qsdp}} = \max\{\eta_P, \eta_D, \eta_Z, \eta_{S_1}, \eta_{S_2}\}, \tag{97}$$

where

$$\begin{aligned} &\eta_P = \frac{\|\mathcal{A}_E X - b_E\|}{1 + \|b_E\|}, \quad \eta_D = \frac{\|Z + \mathcal{B}^*\Xi + S + \mathcal{A}_E^* y_E - C\|}{1 + \|C\|}, \\ &\eta_Z = \frac{\|X - \Pi_{\mathcal{K}}(X - Z)\|}{1 + \|X\| + \|Z\|}, \quad \eta_{S_1} = \frac{|\langle S,\ X\rangle|}{1 + \|S\| + \|X\|}, \quad \eta_{S_2} = \frac{\|X - \Pi_{\mathcal{S}^n_+}(X)\|}{1 + \|X\|}. \end{aligned}$$

We terminate the solvers SCB- SPADMM, ADMM and ADMMGB when $\eta_{\text{qsdp}} < 10^{-6}$ with the maximum number of iterations set at 25,000.

The table in the supplementary material (Online Resource 1) reports detailed numerical results for SCB- SPADMM, ADMM and ADMMGB in solving some large scale QSDP problems. Here, we only list the results for the case of $\operatorname{rank}(B) = 10$, since we obtain similar results for the case of $\operatorname{rank}(B) = 5$. Our numerical experience also indicates that the order of solving the subproblems generally has no influence on the performance of SCB- SPADMM . From the numerical results, one can observe that SCB- SPADMM is generally the fastest in terms of the computing time, especially when the problem size is large. In addition, we can see that SCB- SPADMM and ADMM solved all instances to the required accuracy, while ADMMGB failed in certain cases.

Figure 1 shows the performance profiles in terms of the number of iterations and computing time for SCB- SPADMM, ADMM and ADMMGB, for all the tested large scale

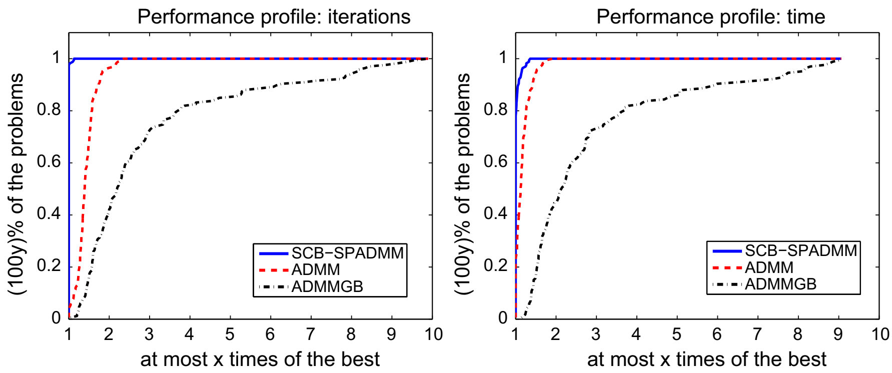

Fig. 1 Performance profiles of SCB- SPADMM, ADMM and ADMMGB for the tested large scale QSDP

QSDP problems. We recall that a point (x, y) is in the performance profiles curve of a method if and only if it can solve $(100y)\%$ of all the tested problems no slower than x times of any other methods. We may observe that for the majority of the tested problems, SCB- SPADMM takes the least number of iterations. Besides, in terms of computing time, it can be seen that both SCB- SPADMM and ADMM outperform ADMMGB by a significant margin, even though ADMM has no convergence guarantee.

4.2 Numerical results for nearest correlation matrix (NCM) approximations

In this subsection, we first consider the problem of finding the nearest correlation matrix (NCM) to a given matrix $G \in \mathcal{S}^n$:

$$\begin{aligned} &\min \ \frac{1}{2}\|H \circ (X - G)\|_F^2 + \langle C,\ X\rangle \\ &\text{s.t. } \mathcal{A}_E X \ = \ b_E, \quad X \in \mathcal{S}_+^n \cap \mathcal{K}, \end{aligned} \tag{98}$$

where $H \in \mathcal{S}^n$ is a nonnegative weight matrix, $\mathcal{A}_E : \mathcal{S}^n \to \Re^{m_E}$ is a linear map, $G \in \mathcal{S}^n$, $C \in \mathcal{S}^n$ and $b_E \in \Re^{m_E}$ are given data, $\mathcal{K}$ is a nonempty simple closed convex set, e.g., $\mathcal{K} = \{W \in \mathcal{S}^n : \ L \leq W \leq U\}$ with $L, U \in \mathcal{S}^n$ being given matrices. In fact, this is also an instance of the general model of problem (93) with no inequality constraints, $\mathcal{Q}X = H \circ H \circ X$ and $\mathcal{B}X = H \circ X$. We place this special example of QSDP here since an extension will be considered next.

Now, let's consider an interesting variant of the above NCM problem:

$$\begin{aligned} &\min \ \|H \circ (X - G)\|_2 + \langle C,\ X\rangle \\ &\text{s.t. } \mathcal{A}_E X \ = \ b_E, \quad X \in \mathcal{S}_+^n \cap \mathcal{K}. \end{aligned} \tag{99}$$

Note, in (99), instead of the Frobenius norm, we use the spectral norm. By introducing a slack variable Y, we can reformulate problem (99) as

$$\begin{aligned} &\min \ \|Y\|_2 + \langle C,\ X\rangle \\ &\text{s.t. } H \circ (X - G) = Y, \quad \mathcal{A}_E X = b_E, \quad X \in \mathcal{S}_+^n \cap \mathcal{K}. \end{aligned} \tag{100}$$

The dual of problem (100) is given by

$$\begin{aligned}
&\max\ -\delta^*_{\mathcal{K}}(-Z) + \langle H \circ G,\ \Xi \rangle + \langle b_E,\ y_E \rangle \\
&\text{s.t. } Z + H \circ \Xi + S + \mathcal{A}^*_E y_E = C, \quad \|\Xi\|_* \leq 1, \quad S \in \mathcal{S}^n_+,
\end{aligned} \tag{101}$$

which is obviously equivalent to the following problem

$$\begin{aligned}
&\max\ -\delta^*_{\mathcal{K}}(-Z) + \langle H \circ G,\ \Xi \rangle + \langle b_E,\ y_E \rangle \\
&\text{s.t. } Z + H \circ \Xi + S + \mathcal{A}^*_E y_E = C, \quad \|\Gamma\|_* \leq 1, \quad S \in \mathcal{S}^n_+, \\
&\quad \mathcal{D}^*\Gamma - \mathcal{D}^*\Xi = 0,
\end{aligned} \tag{102}$$

where $\mathcal{D} : \mathcal{S}^n \to \mathcal{S}^n$ is a nonsingular linear operator. Note that SCB- SPADMM can not be directly applied to solve the problem (101) while the equivalent reformulation (102) fits our model nicely.

In our numerical test, matrix $\widehat{G}$ is the gene correlation matrix from [26]. For testing purpose we perturb $\widehat{G}$ to

$$G := (1-\alpha)\widehat{G} + \alpha E,$$

where $\alpha \in (0, 1)$ and E is a randomly generated symmetric matrix with entries in $[-1, 1]$. We also set $G_{ii} = 1,\ i = 1, \ldots, n$. The weight matrix H is generated from a weight matrix H_0 used by a hedge fund company. The matrix H_0 is a 93×93 symmetric matrix with all positive entries. It has about 24 % of the entries equal to 10^{-5} and the rest are distributed in the interval $[2, 1.28 \times 10^3]$. It has 28 eigenvalues in the interval $[-520, -0.04]$, 11 eigenvalues in the interval $[-5 \times 10^{-13}, 2 \times 10^{-13}]$, and the rest of 54 eigenvalues in the interval $[10^{-4}, 2 \times 10^4]$. The MATLAB code for generating the matrix H is given by

```
tmp = kron(ones(25,25),H0); H = tmp(1:n,1:n); H = (H'+H)/2.
```

The reason for using such a weight matrix is because the resulting problems generated are more challenging to solve as opposed to a randomly generated weight matrix. Note that the matrices G and H are generated in the same way as in [27]. For simplicity, we further set $C = 0$ and $\mathcal{K} = \{X \in \mathcal{S}^n :\ X \geq -0.5\}$.

Generally speaking, there is no widely accepted stopping criterion for spectral norm H-weighted NCM problem (100). Here, with reference to the general relative residue (92), we measure the accuracy of an approximate optimal solution (X, Z, Ξ, S, y_E) for spectral norm H-weighted NCM problem problem (99) [equivalently (100)] and its dual (101) [equivalently (102)] by using the following relative residual derived from the general optimality condition (91):

$$\eta_{\text{sncm}} = \max\{\eta_P, \eta_D, \eta_Z, \eta_{S_1}, \eta_{S_2}, \eta_\Xi\}, \tag{103}$$

Table 1 The performance of SCB- SPADMM, ADMM, ADMMGB on Frobenius norm H-weighted NCM problems [dual of (98)] (accuracy $= 10^{-6}$)

Problem	n_s	α	Iteration scb|admm|gb	η_{qsdp} scb|admm|gb	η_{gap} scb|admm|gb	Time scb|admm|gb
Lymph	587	0.10	263|522|696	9.9-7|9.9-7|9.9-7	-4.4-7|-4.5-7|-4.0-7	30|53|1:23
	587	0.05	264|356|592	9.9-7|9.9-7|9.9-7	-3.9-7|-3.4-7|-3.0-7	29|35|1:08
ER	692	0.10	268|355|711	9.9-7|9.9-7|9.9-7	-5.1-7|-4.7-7|-4.2-7	43|51|1:58
	692	0.05	226|293|603	9.9-7|9.9-7|9.9-7	-4.2-7|-3.8-7|-3.3-7	37|43|1:54
Arabid.	834	0.10	510|528|725	9.9-7|9.9-7|9.9-7	-5.9-7|-5.3-7|-3.9-7	2:11|2:02|3:03
	834	0.05	444|470|650	9.9-7|9.9-7|9.9-7	-5.8-7|-5.2-7|-4.8-7	1:51|1:43|2:44
Leukemia	1,255	0.10	292|420|826	9.9-7|9.9-7|9.9-7	-5.4-7|-5.3-7|-4.4-7	3:13|4:11|9:13
	1,255	0.05	251|408|670	9.9-7|9.7-7|9.6-7	-5.4-7|-4.9-7|-4.0-7	2:48|4:03|7:35
Heredit.	1,869	0.10	555|634|871	9.9-7|9.9-7|9.9-7	-9.1-7|-9.1-7|-7.0-7	17:39|18:38|28:01
	1,869	0.05	530|626|839	9.9-7|9.9-7|9.9-7	-8.7-7|-8.7-7|-5.2-7	16:50|18:15|26:34

In the table, "scb" stands for SCB- SPADMM and "gb" stands for ADMMGB, respectively. The computation time is in the format of "hours:minutes:seconds"

where

$$
\begin{aligned}
\eta_P &= \frac{\|\mathcal{A}_E X - b_E\|}{1+\|b_E\|}, \quad \eta_D = \frac{\|Z + H\circ \Xi + S + \mathcal{A}_E^* y_E\|}{1+\|Z\|+\|S\|},\\
\eta_Z &= \frac{\|X - \Pi_{\mathcal{K}}(X-Z)\|}{1+\|X\|+\|Z\|}, \quad \eta_{S_1} = \frac{|\langle S,\ X\rangle|}{1+\|S\|+\|X\|}, \quad \eta_{S_2} = \frac{\|X - \Pi_{\mathcal{S}^n_+}(X)\|}{1+\|X\|},\\
\eta_\Xi &= \frac{\|\Xi - \Pi_{\{X\in\Re^{n\times n}\,:\|X\|_*\le 1\}}(\Xi - H\circ(X-G))\|}{1+\|\Xi\|+\|H\circ(X-G)\|}.
\end{aligned}
$$

Firstly, numerical results for solving F-norm H-weighted NCM problems (99) are reported. We compare all three algorithms, namely SCB- SPADMM, ADMM, ADMMGB using the relative residue (97). We terminate the solvers when $\eta_{\text{qsdp}} < 10^{-6}$ with the maximum number of iterations set at 25,000.

In Table 1, we report detailed numerical results for SCB- SPADMM, ADMM and ADMMGB in solving various instances of F-norm H-weighted NCM problem. As we can see from Table 1, our SCB- SPADMM is certainly more efficient than the other two algorithms on most of the problems tested.

The rest of this subsection is devoted to the numerical results of the spectral norm H-weighted NCM problem (99). As mentioned before, SCB- SPADMM is applied to solve the problem (102) rather than (101). We implemented all the algorithms for solving problem (102) using the relative residue (103). We terminate the solvers when $\eta_{\text{sncm}} < 10^{-5}$ with the maximum number of iterations set at 25,000. In Table 2, we report detailed numerical results for SCB- SPADMM, ADMM and ADMMGB in solving various instances of spectral norm H-weighted NCM problem. As we can see from Table 2, our SCB- SPADMM is much more efficient than the other two algorithms.

Table 2 The performance of SCB- SPADMM, ADMM, ADMMGB on spectral norm H-weighted NCM problem (102) (accuracy $= 10^{-5}$)

Problem	n_s	α	Iteration scb|admm|gb	η_{sncm} scb|admm|gb	η_{gap} scb|admm|gb	Time scb|admm|gb
Lymph	587	0.10	4110|6048|7131	9.9-6|9.9-6|1.0-5	-3.4-5|-2.8-5|-2.7-5	13:21|17:10|21:43
	587	0.05	5001|7401|8101	9.8-6|9.9-6|9.9-6	-2.0-5|-2.3-5|-8.1-6	19:41|21:25|25:13
ER	692	0.10	3251|4844|6478	9.9-6|9.9-6|1.0-5	-3.1-5|-2.6-5|-6.0-6	15:06|19:30|28:03
	692	0.05	4201|5851|7548	9.3-6|9.8-6|1.0-5	-3.5-5|-2.9-5|-3.4-5	18:44|23:46|32:57
Arabid.	834	0.10	3344|6251|7965	9.9-6|9.7-6|1.0-5	-3.8-5|-2.0-5|-3.7-5	23:20|40:12|54:31
	834	0.05	2496|3101|3231	9.9-6|9.9-6|1.0-5	-9.1-5|-4.3-5|-5.3-5	17:03|19:53|21:56
Leukemia	1,255	0.10	4351|6102|7301	9.9-6|9.9-6|1.0-5	-3.7-5|-3.3-5|-3.0-5	1:22:42|1:49:02|2:16:52
	1,255	0.05	3957|5851|10151	9.9-6|9.7-6|9.5-6	-7.2-5|-5.7-5|-1.1-5	1:18:19|1:44:47|3:26:08

In the table, "scb" stands for SCB- SPADMM and "gb" stands for ADMMGB, respectively. The computation time is in the format of "hours:minutes:seconds"

Table 3 The performance of LADMM, LADMMGB on spectral norm H-weighted NCM problem(101) (accuracy $= 10^{-5}$)

Problem	n_s	α	Iteration ladmm|lgb	η_{sncm} ladmm|lgb	η_{gap} ladmm|lgb	Time ladmm|lgb
Lymph	587	0.10	8,401 | 25,000	9.9-6 | 1.4-5	-1.6-5 | -2.1-5	23:59 | 1:22:58
Lymph	587	0.05	13,609 | 25,000	9.9-6 | 2.3-5	-1.6-5 | -4.2-5	39:29 | 1:18:50

In the table, "lgb" stands for LADMMGB. The computation time is in the format of "hours:minutes:seconds"

Observe that although there is no convergence guarantee, one may still apply the directly extended ADMM with 4 blocks to the original dual problem (101) by adding a proximal term for the Ξ part. We call this method LADMM. Moreover, by using the same proximal strategy for Ξ, a convergent linearized alternating direction method with a Gausssian back substitution proposed in [28] (we call the method LADMMGB here and use the parameter $\alpha = 0.99$ in the Gasussian back substitution step) can also be applied to the original problem (101). We have also implemented LADMM and LADMMGB in MATLAB. Our experiments show that solving the problem (101) directly is much slower than solving the equivalent problem (102). Thus, the reformulation of (101)–(102) is in fact advantageous for both ADMM and ADMMGB. In Table 3, for the purpose of illustration we list a couple of detailed numerical results on the performance of LADMM and LADMMGB.

5 Conclusions

In this paper, we have proposed a Schur complement based convergent yet efficient semi-proximal ADMM for solving convex programming problems, with a coupling linear equality constraint, whose objective function is the sum of two proper closed convex functions plus an arbitrary number of convex quadratic or linear functions. The ability of dealing with an arbitrary number of convex quadratic or linear functions in the objective function makes the proposed algorithm very flexible in solving various multi-block convex optimization problems. By conducting numerical experiments on QSDP and its extensions, we have presented convincing numerical results to demonstrate the superior performance of our proposed SCB-SPADMM. As mentioned in the introduction, our primary motivation of introducing this SCB-SPADMM is to quickly generate a good initial point so as to warm-start methods which have fast local convergence properties. For standard linear SDP and linear SDP with doubly nonnegative constraints, this has already been done by Zhao et al. [23] and Yang, Sun and Toh in [24], respectively. Naturally, our next target is to extend the approach of [23,24] to solve QSDP with an initial point generated by SCB-SPADMM. We will report our corresponding findings in subsequent works.

Acknowledgments The authors would like to thank Mr Liuqin Yang at National University of Singapore for suggestions on the numerical implementations of the algorithms described in the paper and the two referees for their careful reading of our paper.

References

1. Qi, H.: Local duality of nonlinear semidefinite programming. Math. Oper. Res. **34**(1), 124–141 (2009)
2. Sun, J., Zhang, S.: A modified alternating direction method for convex quadratically constrained quadratic semidefinite programs. Eur. J. Oper. Res. **207**, 1210–1220 (2010)
3. Toh, K.C.: An inexact primal-dual path following algorithm for convex quadratic sdp. Math. Program. **112**(1), 221–254 (2008)
4. Sun, D.F.: The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. Math. Oper. Res. **31**(4), 761–776 (2006)
5. Bi, S., Pan, S., Chen, J.S.: Nonsingularity conditions for the Fischer–Burmeister system of nonlinear sdps. SIAM J. Optim. **21**(4), 1392–1417 (2011)
6. Zhao, X.Y.: A semismooth newton-cg augmented lagrangian method for large scale linear and convex quadratic sdps. Ph.D. thesis, Department of Mathematics, National University of Singapore (2009)
7. Miao, W., Pan, S., Sun, D.F.: A rank-corrected procedure for matrix completion with fixed basis coefficients. Technical Report, National University of Singapore (2014)
8. Miao, W.: Matrix completion models with fixed basis coefficients and rank regularized prbolems with hard constraints. Ph.D. thesis, Department of Mathematics, Nationla University of Singapore (2013)
9. Wu, B.: High-dimensional analysis on matrix decomposition with application to correlation matrix estimation in factor models. Ph.D. thesis, Department of Mathematics, Nationla University of Singapore (2014)
10. Negahban, S., Wainwright, M.J.: Restricted strong convexity and weighted matrix completion: optimal bounds with noise. J. Mach. Learn. Res. **13**(1), 1665–1697 (2012)
11. Klopp, O.: Noisy low-rank matrix completion with general sampling distribution. Bernoulli **20**(1), 282–303 (2014)
12. Wright, J., Ganesh, A., Rao, S., Peng, Y., Ma, Y.: Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization. In: Proceedings of Neural Information Processing Systems, vol. 3 (2009)
13. Tao, M., Yuan, X.: Recovering low-rank and sparse components of matrices from incomplete and noisy observations. SIAM J. Optim. **21**(1), 57–81 (2011)
14. Glowinski, R., Marrocco, A.: Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de dirichlet non linéares. Revue Francaise d'Automatique, Informatique et Recherche Opérationelle **9**, 41–76 (1975)
15. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. **2**(1), 17–40 (1976)
16. Glowinski, R.: Lectures on numerical methods for nonlinear variational problems. In: Tata Institute of Fundamental Research Lectures on Mathematics and Physics, vol. 65. Tata Institute of Fundamental Research, Bombay. Notes by M. G. Vijayasundaram and M. Adimurthi (1980)

17. Fortin, M., Glowinski, R.: Augmented Lagrangian Methods, Studies in Mathematics and its Applications, vol. 15. North-Holland Publishing Co., Amsterdam. Applications to the numerical solution of boundary value problems. Translated from the French by B. Hunt and D. C. Spicer (1983)
18. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R. (eds.) Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, Studies in Mathematics and Its Applications, vol. 15, pp. 299–331. Elsevier (1983). doi:10.1016/S0168-2024(08)70034-1. http://www.sciencedirect.com/science/article/pii/S0168202408700341
19. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**(1–3), 293–318 (1992)
20. Chen, C., He, B., Ye, Y., Yuan, X.: The direct extension of admm for multi-block convex minimization problems is not necessarily convergent. Math. Program. 1–23 (2014). doi:10.1007/s10107-014-0826-5
21. Sun, D.F., Toh, K.C., Yang, L.: A convergent 3-block semi-proximal alternating direction method of multipliers for conic programming with 4-type of constraints (2014). arXiv:1404.5378
22. Fazel, M., Pong, T.K., Sun, D.F., Tseng, P.: Hankel matrix rank minimization with applications to system identification and realization. SIAM J. Matrix Anal. Appl. **34**(3), 946–977 (2013)
23. Zhao, X.Y., Sun, D.F., Toh, K.C.: A Newton-cg augmented lagrangian method for semidefinite programming. SIAM J. Optim. **20**(4), 1737–1765 (2010)
24. Yang, L., Sun, D.F., Toh, K.C.: Sdpnal +: a majorized semismooth Newton-cg augmented lagrangian method for semidefinite programming with nonnegative constraints (2014). arXiv:1406.0942
25. He, B., Tao, M., Yuan, X.: Alternating direction method with gaussian back substitution for separable convex programming. SIAM J. Optim. **22**(2), 313–340 (2012)
26. Li, L., Toh, K.C.: An inexact interior point method for l1-regularized sparse covariance selection. Math. Program. Comput. **2**(3–4), 291–315 (2010)
27. Jiang, K., Sun, D.F., Toh, K.C.: An inexact accelerated proximal gradient method for large scale linearly constrained convex sdp. SIAM J. Optim. **22**(3), 1042–1064 (2012)
28. He, B., Yuan, X.: Linearized alternating direction method of multipliers with Gaussian back substitution for separable convex programming. Numer. Algebra Control Optim. **3**(2), 247–260 (2013)