

A Schur Complement Based Semi-Proximal ADMM for Convex Quadratic Conic Programming and Extensions

Xudong Li*, Defeng Sun[†] and Kim-Chuan Toh[‡]

September 9, 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, Toh and Yang [arXiv preprint arXiv:1404.5378, (2014)]. 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.

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

*Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore (lixudong@nus.edu.sg).

[†]Department of Mathematics and Risk Management Institute, National University of Singapore, 10 Lower Kent Ridge Road, Singapore (matsundf@nus.edu.sg).

[‡]Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore (mattohk@nus.edu.sg).

semidefinite programming (QSDP)

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

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 , $\mathcal{A}_E : \mathcal{S}^n \rightarrow \mathfrak{R}^{m_E}$ and $\mathcal{A}_I : \mathcal{S}^n \rightarrow \mathfrak{R}^{m_I}$ are two linear maps, $C \in \mathcal{S}^n$, $b_E \in \mathfrak{R}^{m_E}$ and $b_I \in \mathfrak{R}^{m_I}$ 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

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

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

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

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

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

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

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

where p and q are given nonnegative integers, $f : \mathcal{U} \rightarrow (-\infty, +\infty]$, $g : \mathcal{V} \rightarrow (-\infty, +\infty]$, $\theta_i : \mathcal{Y}_i \rightarrow (-\infty, +\infty]$, $i = 1, \dots, p$, and $\varphi_j : \mathcal{Z}_j \rightarrow (-\infty, +\infty]$, $j = 1, \dots, q$ are closed proper convex functions, $\mathcal{F} : \mathcal{X} \rightarrow \mathcal{U}$, $\mathcal{G} : \mathcal{X} \rightarrow \mathcal{V}$, $\mathcal{A}_i : \mathcal{X} \rightarrow \mathcal{Y}_i$, $i = 1, \dots, p$ and $\mathcal{B}_j : \mathcal{X} \rightarrow \mathcal{Z}_j$, $j = 1, \dots, q$ are linear maps, $\mathcal{U}, \mathcal{V}, \mathcal{Y}_1, \dots, \mathcal{Y}_p, \mathcal{Z}_1, \dots, \mathcal{Z}_q$ 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 1.1 For $i = 1, \dots, p$ and $j = 1, \dots, q$, each $\theta_i(\cdot)$ and $\varphi_j(\cdot)$ are convex quadratic functions.

Note that, in general, problem (3) does not satisfy Assumption 1.1 unless y_I is vacuous from the model or $\mathcal{K} \equiv \mathcal{S}^n$. However, one can always reformulate problem (3) equivalently as

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

where $\mathcal{D} : \mathfrak{R}^{m_I} \rightarrow \mathfrak{R}^{m_I}$ is any given nonsingular linear operator and $\delta_{\mathfrak{R}_+^{m_I}}(\cdot)$ is the indicator function over $\mathfrak{R}_+^{m_I}$. Now, one can see that problem (6) satisfies Assumption 1.1.

There are many other important cases that take the form of model (5) satisfying Assumption 1.1. One prominent example comes from the matrix completion with fixed basis coefficients [15, 14, 20]. Indeed the nuclear semi-norm penalized least squares model in [14] can be written as

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

where $\|X\|_*$ is the nuclear norm of X defined as the sum of all its singular values, $\|\cdot\|_\infty$ is the elementwise l_∞ norm defined by $\|X\|_\infty := \max_{i=1, \dots, m} \{\max_{j=1, \dots, n} |X_{ij}|\}$, $\mathcal{A}_F : \mathfrak{R}^{m \times n} \rightarrow \mathfrak{R}^{n_F}$ and $\mathcal{A}_E : \mathfrak{R}^{m \times n} \rightarrow \mathfrak{R}^{n_E}$ are two linear maps, ρ and α are two given positive parameters, $d \in \mathfrak{R}^{n_F}$, $C \in \mathfrak{R}^{m \times n}$ and $b_E \in \mathfrak{R}^{n_E}$ are given data, $\Omega \subseteq \{1, \dots, m\} \times \{1, \dots, n\}$ is the set of the indices relative to which the basis coefficients are not fixed, $\mathcal{R}_\Omega : \mathfrak{R}^{m \times n} \rightarrow \mathfrak{R}^{|\Omega|}$ 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., $\Omega = \{1, \dots, m\} \times \{1, \dots, n\}$ and \mathcal{A}_E are vacuous), the above problem reduces to the model considered by Negahban and Wainwright in [16] and Klopp in [12]. By introducing slack variables η , R and W , we can reformulate problem (7) as

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

The dual of problem (8) takes the form of

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

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 [19]:

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

where $W \in \mathfrak{R}^{m \times n}$ 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:

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

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

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

Again, problem (11) satisfies Assumption 1.1. In [18], 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 $\mathcal{Y} := \mathcal{Y}_1 \times \mathcal{Y}_2 \times \dots \times \mathcal{Y}_p$, $\mathcal{Z} := \mathcal{Z}_1 \times \mathcal{Z}_2 \times \dots \times \mathcal{Z}_q$. 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 $\mathcal{A} : \mathcal{X} \rightarrow \mathcal{Y}$ 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 $\mathcal{B} : \mathcal{X} \rightarrow \mathcal{Z}$ 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:

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

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 $\| \cdot \|$, $\phi_i : \mathcal{W}_i \rightarrow (-\infty, +\infty]$ is a closed proper convex function, $\mathcal{H}_i : \mathcal{X} \rightarrow \mathcal{W}_i$ 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

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

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

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

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

where $\tau \in (0, 2)$ guarantees the convergence. Due to the non-separability of the quadratic penalty term in \mathcal{L}_σ , 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):

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

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

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

The convergence of 2-block ADMM has already been extensively studied in [8, 6, 7, 4, 5, 2] 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 [1] 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, Toh and Yang [17] 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 [3]. 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 [22, 21] 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 section 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

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

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

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

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

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

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

$$\mathcal{L}_\sigma(u, v; x) = f(u) + g(v) + \langle x, \mathcal{F}^*u + \mathcal{G}^*v - c \rangle + \frac{\sigma}{2} \|\mathcal{F}^*u + \mathcal{G}^*v - c\|^2. \tag{21}$$

The semi-proximal ADMM proposed in [3], when applied to (19), has the following template. Since the proximal terms added here are allowed to be positive semidefinite, the corresponding method is referred to as semi-proximal ADMM instead of proximal ADMM as in [3].

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 $(u^0, v^0, x^0) \in \operatorname{dom}(f) \times \operatorname{dom}(g) \times \mathcal{X}$. For $k = 0, 1, 2, \dots$, perform the k th iteration as follows:

Step 1. Compute

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

Step 2. Compute

$$v^{k+1} = \operatorname{argmin}_v \mathcal{L}_\sigma(u^{k+1}, v; x^k) + \frac{\sigma}{2} \|v - v^k\|_{\mathcal{T}_g}^2. \tag{23}$$

Step 3. Compute

$$x^{k+1} = x^k + \tau\sigma(\mathcal{F}^*u^{k+1} + \mathcal{G}^*v^{k+1} - c). \tag{24}$$

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 (22) and (23). In addition, they play important roles in ensuring the boundedness of the two generated sequences $\{y^{k+1}\}$ and $\{z^{k+1}\}$. 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 y^{k+1} and z^{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 $\tilde{\xi} \in \partial f(\tilde{u})$,

$$\langle \xi - \tilde{\xi}, u - \tilde{u} \rangle \geq \|u - \tilde{u}\|_{\Sigma_f}^2 \quad (25)$$

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

$$\langle \zeta - \tilde{\zeta}, v - \tilde{v} \rangle \geq \|v - \tilde{v}\|_{\Sigma_g}^2. \quad (26)$$

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

Assumption 2.1 *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 2.1 *Let Σ_f and Σ_g be the self-adjoint and positive semidefinite operators defined by (25) and (26), respectively. Suppose that the solution set of problem (19) is nonempty and that Assumption 2.1 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 2.1 *The conclusions of Theorem 2.1 follow essentially from the results given in [3, Theorem B.1]. See [17] for more detailed discussions.*

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

$$g(v) = \frac{1}{2} \langle v, \Sigma_g v \rangle - \langle b, v \rangle, \quad v \in \mathcal{V}, \quad (27)$$

where Σ_g 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

$$\begin{aligned} \min \quad & f(u) + \frac{1}{2} \langle v, \Sigma_g v \rangle - \langle b, v \rangle \\ \text{s.t.} \quad & \mathcal{F}^*u + \mathcal{G}^*v = c. \end{aligned} \quad (28)$$

The dual of problem (28) is given by

$$\min \{ \langle c, x \rangle + f^*(s) + g^*(t) \mid \mathcal{F}x + s = 0, \mathcal{G}x + t = 0 \}. \quad (29)$$

In order to solve subproblem (23) 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 (23) can be solved efficiently. Here, we choose \mathcal{T}_g as follows. Let $\mathcal{E}_g : \mathcal{V} \rightarrow \mathcal{V}$ 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

$$\mathcal{T}_g := \mathcal{E}_g - \sigma^{-1}\Sigma_g - \mathcal{G}\mathcal{G}^* \succeq 0. \quad (30)$$

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

$$\widehat{\mathcal{T}}_f := \mathcal{T}_f + \mathcal{F}\mathcal{G}^*\mathcal{E}_g^{-1}\mathcal{G}\mathcal{F}^*. \quad (31)$$

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

Proposition 2.1 *It holds that*

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

Proof. We have that

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

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 (31), we know that the conclusion of this proposition holds. \square

Now, we can propose our Schur complement based semi-proximal augmented Lagrangian method (SCB-SPALM) to solve (28) 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 (28).

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 $k = 0, 1, 2, \dots$, perform the k th iteration as follows:

Step 1. Compute

$$(u^{k+1}, v^{k+1}) = \operatorname{argmin}_{u,v} \mathcal{L}_\sigma(u, v; x^k) + \frac{\sigma}{2} \|u - u^k\|_{\mathcal{T}_f}^2 + \frac{\sigma}{2} \|v - v^k\|_{\mathcal{T}_g}^2. \quad (32)$$

Step 2. Compute

$$x^{k+1} = x^k + \tau\sigma(\mathcal{F}^*u^{k+1} + \mathcal{G}^*v^{k+1} - c). \quad (33)$$

Note that problem (32) in Step 1 is well defined if the the linear operator \mathcal{W} defined in Proposition 2.1 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 (28), Assumption 2.1 is reduced to the following:

Assumption 2.2 *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 (28).

Theorem 2.2 *Let Σ_f , Σ_g and \mathcal{T}_g be three self-adjoint and positive semidefinite operators defined by (25), (27) and (30), respectively. Suppose that the solution set of problem (28) is nonempty and that Assumption 2.2 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 (28) and x^∞ solves (29), respectively.*
- (ii) *If $\sigma^{-1}\Sigma_f + \mathcal{T}_f + \mathcal{F}\mathcal{F}^*$ is positive definite, then the sequence $\{(u^k, v^k, x^k)\}$, which is automatically well defined, converges to a unique limit, say, $(u^\infty, v^\infty, x^\infty)$ with (u^∞, v^∞) solving problem (28) and x^∞ solving (29), respectively.*

Proof. By combining Theorem 2.1 and Proposition 2.1, one can prove the results of this theorem directly. \square

The relationship between Algorithm SCB-SPALM and Algorithm SPADMM for solving (28) will be revealed in the next proposition.

Let $\delta_g : \mathcal{U} \times \mathcal{V} \times \mathcal{X} \rightarrow \mathcal{U}$ be an auxiliary linear function associated with (28) defined by

$$\delta_g(u, v, x) := \mathcal{F}\mathcal{G}^*\mathcal{E}_g^{-1}(b - \mathcal{G}x - \Sigma_g v + \sigma\mathcal{G}(c - \mathcal{F}^*u - \mathcal{G}^*v)). \quad (34)$$

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

$$(u^+, v^+) = \operatorname{argmin}_{u,v} \mathcal{L}_\sigma(u, v; \bar{x}) + \frac{\sigma}{2} \|u - \bar{u}\|_{\mathcal{T}_f}^2 + \frac{\sigma}{2} \|v - \bar{v}\|_{\mathcal{T}_g}^2. \quad (35)$$

Proposition 2.2 Let $\bar{\alpha} := \sigma^{-1}b + \mathcal{T}_g\bar{v} + \mathcal{G}(c - \sigma^{-1}\bar{x})$. Define $v' \in \mathcal{V}$ by

$$v' = \operatorname{argmin}_v \mathcal{L}_\sigma(\bar{u}, v; \bar{x}) + \frac{\sigma}{2} \|v - \bar{v}\|_{\mathcal{T}_g}^2 = \mathcal{E}_g^{-1}(\bar{\alpha} - \mathcal{G}\mathcal{F}^*\bar{u}). \quad (36)$$

The optimal solution (u^+, v^+) to problem (35) is generated exactly by the following procedure

$$\begin{cases} u^+ &= \operatorname{argmin}_u \mathcal{L}_\sigma(u, \bar{v}; \bar{x}) + \langle \bar{\delta}_g, u \rangle + \frac{\sigma}{2} \|u - \bar{u}\|_{\mathcal{T}_f}^2, \\ v^+ &= \operatorname{argmin}_v \mathcal{L}_\sigma(u^+, v; \bar{x}) + \frac{\sigma}{2} \|v - \bar{v}\|_{\mathcal{T}_g}^2 = \mathcal{E}_g^{-1}(\bar{\alpha} - \mathcal{G}\mathcal{F}^*u^+). \end{cases} \quad (37)$$

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

$$\begin{cases} u^+ &= \operatorname{argmin}_u \mathcal{L}_\sigma(u, v'; \bar{x}) + \frac{\sigma}{2} \|u - \bar{u}\|_{\mathcal{T}_f}^2, \\ v^+ &= \operatorname{argmin}_v \mathcal{L}_\sigma(u^+, v; \bar{x}) + \frac{\sigma}{2} \|v - \bar{v}\|_{\mathcal{T}_g}^2 = \mathcal{E}_g^{-1}(\bar{\alpha} - \mathcal{G}\mathcal{F}^*u^+). \end{cases} \quad (38)$$

Proof. First we show that the equivalence between (35) and (37). Define

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

By simple algebraic manipulations, we have that

$$\tilde{\mathcal{L}}_\sigma(u, v; \bar{x}) = f(u) + \frac{\sigma}{2} \|u - \bar{u}\|_{\mathcal{T}_f}^2 + \phi(u, v) - \frac{1}{2\sigma} \|\bar{x}\|^2, \quad (39)$$

where

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

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

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

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

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

From (39), we have that for any given $u \in \mathcal{U}$,

$$\begin{aligned} \tilde{\mathcal{L}}_\sigma(u, v(u); \bar{x}) &= f(u) + \frac{\sigma}{2} \|u - \bar{u}\|_{\mathcal{T}_f}^2 + \frac{\sigma}{2} \|\mathcal{G}\mathcal{F}^*(u - \bar{u})\|_{\mathcal{E}_g^{-1}}^2 + \phi(u, v(u)) - \frac{1}{2\sigma} \|\bar{x}\|^2 \\ &= f(u) + \frac{\sigma}{2} \|u - \bar{u}\|_{\mathcal{T}_f}^2 + \sigma \langle u, \mathcal{F}(\mathcal{G}^*\mathcal{E}_g^{-1}\bar{\alpha} + \sigma^{-1}\bar{x} - c) - \mathcal{F}\mathcal{G}^*\mathcal{E}_g^{-1}\mathcal{G}\mathcal{F}^*\bar{u} \rangle + \frac{\sigma}{2} \langle u, \mathcal{F}\mathcal{F}^*u \rangle + \kappa_1 \\ &= f(u) + \frac{\sigma}{2} \|u - \bar{u}\|_{\mathcal{T}_f}^2 + \langle u, \bar{\delta}_g \rangle + \langle u, \mathcal{F}(\bar{x} + \sigma(\mathcal{G}^*\bar{v} - c)) \rangle + \frac{\sigma}{2} \langle u, \mathcal{F}\mathcal{F}^*u \rangle + \kappa_1 \\ &= \mathcal{L}_\sigma(u, \bar{v}; \bar{x}) + \langle u, \bar{\delta}_g \rangle + \frac{\sigma}{2} \|u - \bar{u}\|_{\mathcal{T}_f}^2 + \kappa_2, \end{aligned} \quad (40)$$

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

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

Now, we have that

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

where $\tilde{\mathcal{L}}_\sigma(u, v(u); \bar{x})$ satisfies (40). From here, the equivalence between (35) and (37) follows.

Next, we prove the equivalence between (37) and (38). Note that, the first minimization problem in (38) 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 (36), is equivalent to

$$0 \in \partial f(u^+) + \mathcal{F} \bar{x} + \sigma \mathcal{F}(\mathcal{F}^* u^+ - c + \mathcal{G}^* \mathcal{E}_g^{-1}(\bar{\alpha} - \mathcal{G} \mathcal{F}^* \bar{u})) + \sigma \mathcal{T}_f(u^+ - \bar{u}). \quad (41)$$

The condition (41) can be reformulated as

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

Thus, we have

$$0 \in \partial f(u^+) + \mathcal{F} \bar{x} + \sigma \mathcal{F}(\mathcal{F}^* u^+ + \mathcal{G}^* \bar{v} - c) + \bar{\delta}_g + \sigma \mathcal{T}_f(u^+ - \bar{u}), \quad (42)$$

which can equivalently be rewritten as

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

The equivalence between (37) and (38) then follows. This completes the proof of this proposition. \square

Proposition 2.3 *Let $\delta_g^k := \delta_g(u^k, v^k, x^k)$ for $k = 0, 1, 2, \dots$. We have that u^{k+1} and v^{k+1} obtained by Algorithm SCB-SPALM for solving (28) can be generated exactly according to the following procedure:*

$$\begin{cases} u^{k+1} &= \operatorname{argmin}_u \mathcal{L}_\sigma(u, v^k; x^k) + \langle \delta_g^k, u \rangle + \frac{\sigma}{2} \|u - u^k\|_{\mathcal{T}_f}^2, \\ v^{k+1} &= \operatorname{argmin}_v \mathcal{L}_\sigma(u^{k+1}, v; x^k) + \frac{\sigma}{2} \|v - v^k\|_{\mathcal{T}_g}^2, \\ x^{k+1} &= x^k + \tau \sigma (\mathcal{F}^* u^{k+1} + \mathcal{G}^* v^{k+1} - c). \end{cases} \quad (43)$$

Proof. The conclusion follows directly from (37) in Proposition 2.2. \square

Remark 2.2 (i) Note that comparing to (22) in Algorithm SPADMM, the first subproblem of (43) 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 (43), apart from the range of τ , our Algorithm SCB-SPALM differs from the classical 2-block ADMM for solving problem (28) only in the linear term $\langle \delta_g^k, \cdot \rangle$. This shows that the classical 2-block ADMM for solving problem (28) 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 $\frac{\sigma}{2}\|u - u^k\|_{\mathcal{T}_f}^2 + \frac{\sigma}{2}\|v - v^k\|_{\mathcal{T}_g}^2$ used to derive the scheme (43) for solving (28) with the following proximal term which allows one to update u and v simultaneously:

$$\frac{\sigma}{2}(\|(u, v) - (u^k, v^k)\|_{\mathcal{M}}^2 + \|u - u^k\|_{\mathcal{T}_f}^2 + \|v - v^k\|_{\mathcal{T}_g}^2) \quad \text{with} \quad \mathcal{M} = \begin{pmatrix} \mathcal{D}_1 & -\mathcal{F}\mathcal{G}^* \\ -\mathcal{G}\mathcal{F}^* & \mathcal{D}_2 \end{pmatrix} \succeq 0, \quad (44)$$

where $\mathcal{D}_1 : \mathcal{U} \rightarrow \mathcal{U}$ and $\mathcal{D}_2 : \mathcal{V} \rightarrow \mathcal{V}$ 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 $\lambda_{\max} = \|\mathcal{F}\mathcal{G}^*\|_2$, $\mathcal{I}_1 : \mathcal{U} \rightarrow \mathcal{U}$ and $\mathcal{I}_2 : \mathcal{V} \rightarrow \mathcal{V}$ 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:

$$\begin{cases} u^{k+1} &= \operatorname{argmin}_u \mathcal{L}_\sigma(u, v^k; x^k) + \frac{\sigma}{2}\|u - u^k\|_{\mathcal{D}_1 + \mathcal{T}_f}^2, \\ v^{k+1} &= \operatorname{argmin}_v \mathcal{L}_\sigma(u^k, v; x^k) + \frac{\sigma}{2}\|v - v^k\|_{\mathcal{D}_2 + \mathcal{T}_g}^2, \\ x^{k+1} &= x^k + \tau\sigma(\mathcal{F}^*u^{k+1} + \mathcal{G}^*v^{k+1} - c). \end{cases} \quad (45)$$

To ensure that the subproblems in (45) 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 (32) and (44), we can easily see that the difference is:

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

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 (28) 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

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

In contrast, we have

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

which is larger than the former upper bound $\|u - u^k\|^2$ if $\|\mathcal{G}\| \geq 1/2$. Thus we can conclude safely that the proximal term $\|u - u^k\|_{\mathcal{F}\mathcal{G}^*\mathcal{E}_g^{-1}\mathcal{G}\mathcal{F}^*}^2$ can be potentially much smaller than $\|(u, v) - (u^k, v^k)\|_{\mathcal{M}}^2$ unless $\|\mathcal{G}\|$ is very small.

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 (45) 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

$$\begin{aligned} \min \quad & f(u) + \sum_{i=1}^p \theta_i(y_i) + g(v) + \sum_{j=1}^q \varphi_j(z_j) \\ \text{s.t.} \quad & \mathcal{F}^*u + \sum_{i=1}^p \mathcal{A}_i^*y_i + \mathcal{G}^*v + \sum_{j=1}^q \mathcal{B}_j^*z_j = c \end{aligned} \quad (46)$$

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

$$\theta_i(y_i) = \frac{1}{2}\langle y_i, \mathcal{P}_i y_i \rangle - \langle b_i, y_i \rangle, \quad i = 1, \dots, p, \quad \varphi_j(z_j) = \frac{1}{2}\langle z_j, \mathcal{Q}_j z_j \rangle - \langle d_j, z_j \rangle, \quad j = 1, \dots, q,$$

where \mathcal{P}_i and \mathcal{Q}_j are given self-adjoint positive semidefinite linear operators. The dual of (46) is given by

$$\max \left\{ -\langle c, x \rangle - f^*(-\mathcal{F}x) - \sum_{i=1}^p \theta_i^*(-\mathcal{A}_i x) - g^*(-\mathcal{G}x) - \sum_{j=1}^q \varphi_j^*(-\mathcal{B}_j x) \right\}, \quad (47)$$

which can equivalently be written as

$$\begin{aligned}
\min \quad & \langle c, x \rangle + f^*(s) + \sum_{i=1}^p \theta_i^*(r_i) + g^*(t) + \sum_{j=1}^q \varphi_j^*(w_j) \\
\text{s.t.} \quad & \mathcal{F}x + s = 0, \quad \mathcal{A}_i x + r_i = 0, \quad i = 1, \dots, p, \\
& \mathcal{G}x + t = 0, \quad \mathcal{B}_j x + w_j = 0, \quad j = 1, \dots, q.
\end{aligned} \tag{48}$$

For $i = 1, \dots, p$, let \mathcal{E}_{θ_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}_{θ_i} in a way that its inverse can be computed at a moderate cost. Define

$$\mathcal{T}_{\theta_i} := \mathcal{E}_{\theta_i} - \sigma^{-1}\mathcal{P}_i - \mathcal{A}_i\mathcal{A}_i^* \succeq 0, \quad i = 1, \dots, p. \tag{49}$$

Note that for numerical efficiency, we need the self-adjoint positive semidefinite linear operator \mathcal{T}_{θ_i} to be as small as possible for each i . Similarly, for $j = 1, \dots, q$, let \mathcal{E}_{φ_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

$$\mathcal{T}_{\varphi_j} := \mathcal{E}_{\varphi_j} - \sigma^{-1}\mathcal{Q}_j - \mathcal{B}_j\mathcal{B}_j^* \succeq 0, \quad j = 1, \dots, q. \tag{50}$$

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

For notational convenience, we define

$$y_{\leq i} := (y_1, y_2, \dots, y_i), \quad y_{\geq i} := (y_i, y_{i+1}, \dots, y_p), \quad i = 0, \dots, p+1$$

with the convention that $y_0 = y_{p+1} = y_{\leq 0} = y_{\geq p+1} = \emptyset$. For $i = 1, \dots, p$, define the linear operator $\mathcal{A}_{\leq i} : \mathcal{X} \rightarrow \mathcal{Y}$ by

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

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

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

$$\Gamma(u, y, v, z) := \mathcal{F}^*u + \mathcal{A}^*y + \mathcal{G}^*v + \mathcal{B}^*z - c \quad \forall (u, y, v, z) \in \mathcal{U} \times \mathcal{Y} \times \mathcal{V} \times \mathcal{Z}. \tag{51}$$

Let $\sigma > 0$ be given. The augmented Lagrangian function associated with (46) is given as follows:

$$\mathcal{L}_\sigma(u, y, v, z; x) = f(u) + \theta(y) + g(v) + \varphi(z) + \langle x, \Gamma(u, y, v, z) \rangle + \frac{\sigma}{2} \|\Gamma(u, y, v, z)\|^2 \tag{52}$$

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 (46).

Algorithm SCB-SPADMM: A Schur complement based SPADMM for solving (46).

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 $(u^0, y^0, v^0, z^0, x^0) \in \text{dom}(f) \times \mathcal{Y} \times \text{dom}(g) \times \mathcal{Z} \times \mathcal{X}$. For $k = 0, 1, 2, \dots$, 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 $i = p, \dots, 1$,

$$\bar{y}_i^k = \operatorname{argmin}_{y_i} \mathcal{L}_\sigma(u^k, (y_{\leq i-1}^k, y_i, \bar{y}_{\geq i+1}^k), v^k, z^k; x^k) + \frac{\sigma}{2} \|y_i - y_i^k\|_{\mathcal{T}_{\theta_i}}^2, \quad (53)$$

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

$$u^{k+1} = \operatorname{argmin}_u \mathcal{L}_\sigma(u, \bar{y}^k, v^k, z^k; x^k) + \frac{\sigma}{2} \|u - u^k\|_{\mathcal{T}_f}^2. \quad (54)$$

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

$$y_i^{k+1} = \operatorname{argmin}_{y_i} \mathcal{L}_\sigma(u^{k+1}, (y_{\leq i-1}^{k+1}, y_i, \bar{y}_{\geq i+1}^k), v^k, z^k; x^k) + \frac{\sigma}{2} \|y_i - y_i^k\|_{\mathcal{T}_{\theta_i}}^2. \quad (55)$$

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

$$\bar{z}_j^k = \operatorname{argmin}_{z_j} \mathcal{L}_\sigma(u^{k+1}, y^{k+1}, v^k, (z_{\leq j-1}^k, z_j, \bar{z}_{\geq j+1}^k); x^k) + \frac{\sigma}{2} \|z_j - z_j^k\|_{\mathcal{T}_{\varphi_j}}^2, \quad (56)$$

where \mathcal{T}_{φ_j} is defined as in (50). Then compute

$$v^{k+1} = \operatorname{argmin}_v \mathcal{L}_\sigma(u^{k+1}, y^{k+1}, v, \bar{z}^k; x^k) + \frac{\sigma}{2} \|v - v^k\|_{\mathcal{T}_g}^2. \quad (57)$$

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

$$z_j^{k+1} = \operatorname{argmin}_{z_j} \mathcal{L}_\sigma(u^{k+1}, y^{k+1}, v^{k+1}, (z_{\leq j-1}^{k+1}, z_j, \bar{z}_{\geq j+1}^k); x^k) + \frac{\sigma}{2} \|z_j - z_j^k\|_{\mathcal{T}_{\varphi_j}}^2. \quad (58)$$

Step 5. Compute

$$x^{k+1} = x^k + \tau \sigma (\mathcal{F}^* u^{k+1} + \mathcal{A}^* y^{k+1} + \mathcal{G}^* v^{k+1} + \mathcal{B}^* z^{k+1} - c). \quad (59)$$

In order to prove the convergence of Algorithm SCB-SPADMM for solving (46), 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 $l = 1, \dots, p$,

$$f_l(u) := f(u), \quad f_{l+1}(u, y_{\leq l}) := f(u) + \sum_{i=1}^l \theta_i(y_i) \quad \forall (u, y_{\leq l}) \in \mathcal{U} \times \mathcal{Y}_{\leq l}, \quad (60)$$

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

$$g_1(v) := g(v), \quad g_{l+1}(v, z_{\leq l}) := g(v) + \sum_{j=1}^l \varphi_j(z_j) \quad \forall (v, z_{\leq l}) \in \mathcal{V} \times \mathcal{Z}_{\leq l}. \quad (61)$$

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

$$\mathcal{F}_{i+1}^* = \left(\mathcal{F}^*, \mathcal{A}_1^*, \dots, \mathcal{A}_i^* \right), \quad i = 1, \dots, p, \quad \mathcal{G}_{j+1}^* = \left(\mathcal{G}^*, \mathcal{B}_1^*, \dots, \mathcal{B}_j^* \right), \quad j = 1, \dots, q.$$

Define the following self-adjoint linear operators: $\widehat{\mathcal{T}}_{f_1} := \mathcal{T}_f + \mathcal{F}_1 \mathcal{A}_1^* \mathcal{E}_{\theta_1}^{-1} \mathcal{A}_1 \mathcal{F}_1^*$,

$$\widehat{\mathcal{T}}_{f_i} := \begin{pmatrix} \widehat{\mathcal{T}}_{f_{i-1}} & \\ & \mathcal{T}_{\theta_{i-1}} \end{pmatrix} + \mathcal{F}_i \mathcal{A}_i^* \mathcal{E}_{\theta_i}^{-1} \mathcal{A}_i \mathcal{F}_i^*, \quad i = 2, \dots, p \quad (62)$$

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

$$\widehat{\mathcal{T}}_{g_j} := \begin{pmatrix} \widehat{\mathcal{T}}_{g_{j-1}} & \\ & \mathcal{T}_{\varphi_{j-1}} \end{pmatrix} + \mathcal{G}_j \mathcal{B}_j^* \mathcal{E}_{\varphi_j}^{-1} \mathcal{B}_j \mathcal{G}_j^*, \quad j = 2, \dots, q. \quad (63)$$

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

$$\beta_{p,j} := \mathcal{A}_{j-1} \mathcal{A}_p^* \mathcal{E}_{\theta_p}^{-1} (b_p - \mathcal{A}_p \bar{x} - \mathcal{P}_p \bar{y}_p + \sigma \mathcal{A}_p \bar{\gamma}), \quad j = 1, \dots, p \quad (64)$$

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

$$\beta_{i,j} := \mathcal{A}_{j-1} \mathcal{A}_i^* \mathcal{E}_{\theta_i}^{-1} \left(b_i - \sum_{k=i+1}^p \beta_{k,i+1} - \mathcal{A}_i \bar{x} - \mathcal{P}_i \bar{y}_i + \sigma \mathcal{A}_i \bar{\gamma} \right), \quad j = 1, \dots, i. \quad (65)$$

Let

$$\bar{\delta}_\theta := \sum_{i=1}^p \beta_{i,1}. \quad (66)$$

We will show later in Proposition 3.1 that $\bar{\delta}_\theta$ is the auxiliary linear term associated with problem (46). 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

$$\begin{aligned} y'_i &:= \operatorname{argmin}_{y_i} \mathcal{L}_\sigma(\bar{u}, (\bar{y}_{\leq i-1}, y_i, y'_{\geq i+1}), \bar{v}, \bar{z}; \bar{x}) + \frac{\sigma}{2} \|y_i - \bar{y}_i\|_{\mathcal{T}_{\theta_i}}^2 \\ &= \mathcal{E}_{\theta_i}^{-1} (\sigma^{-1} b_i - \sigma^{-1} \mathcal{A}_i \bar{x} + \mathcal{T}_{\theta_i} \bar{y}_i + \mathcal{A}_i \mathcal{A}_i^* \bar{y}_i - \mathcal{A}_i \Gamma(\bar{u}, (\bar{y}_{\leq i-1}, \bar{y}_i, y'_{\geq i+1}), \bar{v}, \bar{z})) \end{aligned} \quad (67)$$

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

$$(u^+, y^+) := \operatorname{argmin}_{u, y} \mathcal{L}_\sigma(u, y, \bar{v}, \bar{z}; \bar{x}) + \frac{\sigma}{2} \|(u, y_{\leq p-1}) - (\bar{u}, \bar{y}_{\leq p-1})\|_{\mathcal{F}_p}^2 + \frac{\sigma}{2} \|y_p - \bar{y}_p\|_{\mathcal{T}_{\theta_p}}^2. \quad (68)$$

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 2.2 repeatedly though the proof itself is rather lengthy due to the multi-layered nature of the problems involved. For (68), we first express y_p as a function of $(u, y_{\leq p-1})$ to obtain a problem involving only $(u, y_{\leq p-1})$, and from the resulting problem, express y_{p-1} as a function of $(u, y_{\leq p-2})$ to get another problem involving only $(u, y_{\leq p-2})$. We continue this way until we get a problem involving only (u, y_1) .

Proposition 3.1 *The optimal solution (u^+, y^+) defined by (68) can be obtained exactly by*

$$\begin{cases} u^+ &= \operatorname{argmin}_u \mathcal{L}_\sigma(u, \bar{y}, \bar{v}, \bar{z}; \bar{x}) + \langle \bar{\delta}_\theta, u \rangle + \frac{\sigma}{2} \|u - \bar{u}\|_{\mathcal{F}_p}^2, \\ y_i^+ &= \operatorname{argmin}_{y_i} \mathcal{L}_\sigma(u^+, (y_{\leq i-1}^+, y_i, y'_{\geq i+1}), \bar{v}, \bar{z}; \bar{x}) + \frac{\sigma}{2} \|y_i - \bar{y}_i\|_{\mathcal{T}_{\theta_i}}^2, \quad i = 1, \dots, p, \end{cases} \quad (69)$$

where the auxiliary linear term $\bar{\delta}_\theta$ is defined by (66). Furthermore, (u^+, y^+) can also be generated by the following equivalent procedure

$$\begin{cases} u^+ &= \operatorname{argmin}_u \mathcal{L}_\sigma(u, y', \bar{v}, \bar{z}; \bar{x}) + \frac{\sigma}{2} \|u - \bar{u}\|_{\mathcal{F}}^2, \\ y_i^+ &= \operatorname{argmin}_{y_i} \mathcal{L}_\sigma(u^+, (y_{\leq i-1}^+, y_i, y'_{\geq i+1}), \bar{v}, \bar{z}; \bar{x}) + \frac{\sigma}{2} \|y_i - \bar{y}_i\|_{\mathcal{T}_{\theta_i}}^2, \quad i = 1, \dots, p. \end{cases} \quad (70)$$

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 (68) can be obtained exactly by (69). For the case $p = 1$, this follows directly from Proposition 2.2.

Assume that the equivalence between (68) and (69) holds for all $p \leq l$. 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 $(u, y_{\leq l})$ and y_{l+1} :

$$\begin{aligned} \min \quad & f_{l+1}(u, y_{\leq l}) + \theta_{l+1}(y_{l+1}) + g(\bar{v}) + \varphi(\bar{z}) \\ \text{s.t.} \quad & \mathcal{F}_{l+1}^*(u, y_{\leq l}) + \mathcal{A}_{l+1}^* y_{l+1} = \bar{c}. \end{aligned} \quad (71)$$

The augmented Lagrangian function associated with problem (71) is given by

$$\begin{aligned} \mathcal{L}_\sigma^{l+1}((u, y_{\leq l}), y_{l+1}; \bar{v}, \bar{z}, x) &= f_{l+1}(u, y_{\leq l}) + \theta_{l+1}(y_{l+1}) + g(\bar{v}) + \varphi(\bar{z}) \\ &\quad + \langle x, \Gamma(u, y, \bar{v}, \bar{z}) \rangle + \frac{\sigma}{2} \|\Gamma(u, y, \bar{v}, \bar{z})\|^2. \end{aligned} \quad (72)$$

We denote the vector $\delta_{\theta_{l+1}}$ as the auxiliary linear term associated with problem (71) by

$$\delta_{\theta_{l+1}} := \mathcal{F}_{l+1} \mathcal{A}_{l+1}^* \mathcal{E}_{\theta_{l+1}}^{-1} (b_{l+1} - \mathcal{A}_{l+1} \bar{x} - \mathcal{P}_{l+1} \bar{y}_{l+1} + \sigma \mathcal{A}_{l+1} \bar{\gamma}). \quad (73)$$

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

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

with $\beta_{p,j}$, $j = 1, \dots, l+1$, defined as in (64).

By noting that $\mathcal{L}_\sigma^{l+1}((u, y_{\leq l}), y_{l+1}; \bar{v}, \bar{z}, \bar{x}) = \mathcal{L}_\sigma(u, y_{\leq l}, y_{l+1}, \bar{v}, \bar{z}, \bar{x})$, we can rewrite problem (68) for $p = l+1$ equivalently as

$$((u^+, y_{\leq l}^+), y_{l+1}^+) = \operatorname{argmin} \left\{ \begin{array}{l} \mathcal{L}_\sigma^{l+1}((u, y_{\leq l}), y_{l+1}; \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2} \|(u, y_{\leq l}) - (\bar{u}, \bar{y}_{\leq l})\|_{\widehat{\mathcal{T}}_{f_{l+1}}}^2 \\ + \frac{\sigma}{2} \|y_{l+1} - \bar{y}_{l+1}\|_{\mathcal{T}_{\theta_{l+1}}}^2 \end{array} \right\}. \quad (74)$$

Then, from Proposition 2.2, we know that problem (74) is equivalent to

$$(u^+, y_{\leq l}^+) = \operatorname{argmin}_{(u, y_{\leq l})} \left\{ \begin{array}{l} \mathcal{L}_\sigma^{l+1}((u, y_{\leq l}), \bar{y}_{l+1}; \bar{v}, \bar{z}, \bar{x}) + \langle \delta_{\theta_{l+1}}, (u, y_{\leq l}) \rangle \\ + \frac{\sigma}{2} \|(u, y_{\leq l-1}) - (\bar{u}, \bar{y}_{\leq l-1})\|_{\widehat{\mathcal{T}}_{f_l}}^2 + \frac{\sigma}{2} \|y_l - \bar{y}_l\|_{\mathcal{T}_{\theta_l}}^2 \end{array} \right\}, \quad (75)$$

$$y_{l+1}^+ = \operatorname{argmin}_{y_{l+1}} \mathcal{L}_\sigma^{l+1}((u^+, y_{\leq l}^+), y_{l+1}; \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2} \|y_{l+1} - \bar{y}_{l+1}\|_{\mathcal{T}_{\theta_{l+1}}}^2. \quad (76)$$

By observing that $\mathcal{L}_\sigma^{l+1}((u^+, y_{\leq l}^+), y_{l+1}; \bar{v}, \bar{z}, \bar{x}) = \mathcal{L}_\sigma(u^+, y_{\leq l}^+, y_{l+1}, \bar{v}, \bar{z}, \bar{x})$, we know that problem (76) can equivalently be rewritten as

$$y_{l+1}^+ = \operatorname{argmin}_{y_{l+1}} \mathcal{L}_\sigma(u^+, y_{\leq l}^+, y_{l+1}, \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2} \|y_{l+1} - \bar{y}_{l+1}\|_{\mathcal{T}_{\theta_{l+1}}}^2. \quad (77)$$

In order to apply our induction assumption to problem (75), we need to construct a corresponding optimization problem. Define for $i = 1, \dots, l$,

$$\begin{aligned} \tilde{b}_i &:= b_i - \beta_{p,i+1} \quad \text{and} \quad \tilde{\theta}_i(y_i) := \theta_i(y_i) + \langle \beta_{p,i+1}, y_i \rangle = \frac{1}{2} \langle y_i, \mathcal{P}_i y_i \rangle - \langle \tilde{b}_i, y_i \rangle \quad \forall y_i \in \mathcal{Y}_i, \\ \tilde{f}_1(u) &:= f(u) + \langle \beta_{p,1}, u \rangle, \quad \tilde{f}_{i+1}(u, y_{\leq i}) := \tilde{f}_1(u) + \sum_{j=1}^i \tilde{\theta}_j(y_j) \quad \forall (u, y_{\leq i}) \in \mathcal{U} \times \mathcal{Y}_{\leq i}. \end{aligned}$$

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

$$\begin{aligned} \min \quad & \tilde{f}_1(u) + \sum_{i=1}^l \tilde{\theta}_i(y_i) + \theta_{l+1}(\bar{y}_{l+1}) + g(\bar{v}) + \varphi(\bar{z}) \\ \text{s.t.} \quad & \mathcal{F}^* u + \mathcal{A}_{\leq l}^* y_{\leq l} = \bar{c} - \mathcal{A}_{l+1}^* \bar{y}_{l+1}. \end{aligned} \quad (78)$$

The augmented Lagrangian function associated with problem (78) is defined by

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

Define

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

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

$$\mathcal{E}_{\tilde{\theta}_i} \equiv \mathcal{E}_{\theta_i} \quad \text{and} \quad \widehat{\mathcal{T}}_{\tilde{f}_i} \equiv \widehat{\mathcal{T}}_{f_i}, \quad i = 1, \dots, l. \quad (79)$$

Therefore, problem (75) can equivalently be rewritten as

$$(u^+, y_{\leq l}^+) = \operatorname{argmin}_{(u, y_{\leq l})} \left\{ \begin{array}{l} \tilde{\mathcal{L}}_\sigma(u, y_{\leq l}; \bar{y}_{l+1}, \bar{v}, \bar{z}, \bar{x}) \\ + \frac{\sigma}{2} \|(u, y_{\leq l-1}) - (\bar{u}, \bar{y}_{\leq l-1})\|_{\widehat{\mathcal{T}}_{\tilde{f}_l}}^2 + \frac{\sigma}{2} \|y_l - \bar{y}_l\|_{\mathcal{T}_{\tilde{\theta}_l}}^2 \end{array} \right\}. \quad (80)$$

Define

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

and for $i = l-1, l-2, \dots, 1$,

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

The auxiliary linear term $\delta_{\tilde{\theta}}$ associated with problem (80) is given by

$$\delta_{\tilde{\theta}} := \sum_{i=1}^l \tilde{\beta}_{i,1}. \quad (81)$$

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

$$\tilde{\beta}_{i,j} = \beta_{i,j} \quad \forall j = 1, \dots, i. \quad (82)$$

First, by using (79), we have for $j = 1, \dots, l$ that

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

That is, (82) holds for $i = l$ and $j = 1, \dots, l$. Now assume that we have proven $\tilde{\beta}_{i,j} = \beta_{i,j}$ for all $i \geq k+1$ with $k+1 \leq l$ and $j = 1, \dots, i$. We shall next prove that (82) holds for $i = k$ and $j = 1, \dots, k$. Again, by using (79), we have for $j = 1, \dots, k$ that

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

which, shows that (82) holds for $i = k$ and $j = 1, \dots, k$. Thus, (82) is proven.

For $i = l, l-1, \dots, 1$, define $\tilde{y}'_i \in \mathcal{Y}_i$ by

$$\begin{aligned} \tilde{y}'_i &:= \operatorname{argmin}_{y_i} \tilde{\mathcal{L}}_{\sigma}(\bar{u}, (\bar{y}_{\leq i-1}, y_i, \tilde{y}'_{\geq i+1}); \bar{y}_{l+1}, \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2} \|y_i - \bar{y}_i\|_{\mathcal{T}_{\tilde{\theta}_i}}^2, \\ &= \mathcal{E}_{\tilde{\theta}_i}^{-1} (\sigma^{-1} \tilde{b}_i - \sigma^{-1} \mathcal{A}_i \bar{x} + \mathcal{T}_{\tilde{\theta}_i} \bar{y}_i + \mathcal{A}_i \mathcal{A}_i^* \bar{y}_i - \mathcal{A}_i \Gamma(\bar{u}, (\bar{y}_{\leq i-1}, \bar{y}_i, \tilde{y}'_{\geq i+1}, \bar{y}_{l+1}), \bar{v}, \bar{z})), \quad (83) \end{aligned}$$

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

$$\tilde{y}'_i = y'_i \quad \forall i = l, l-1, \dots, 1. \quad (84)$$

We first calculate

$$\begin{aligned} y'_{l+1} - \bar{y}_{l+1} &= \mathcal{E}_{\theta_{l+1}}^{-1} (\sigma^{-1} b_{l+1} - \sigma^{-1} \mathcal{A}_{l+1} \bar{x} + \mathcal{T}_{\theta_{l+1}} \bar{y}_{l+1} + \mathcal{A}_{l+1} \mathcal{A}_{l+1}^* \bar{y}_{l+1} + \mathcal{A}_{l+1} \bar{\gamma} - \mathcal{E}_{\theta_{l+1}} \bar{y}_{l+1}) \\ &= \mathcal{E}_{\theta_{l+1}}^{-1} (\sigma^{-1} b_{l+1} - \sigma^{-1} \mathcal{A}_{l+1} \bar{x} - \sigma^{-1} \mathcal{P}_{l+1} \bar{y}_{l+1} + \mathcal{A}_{l+1} \bar{\gamma}), \quad (85) \end{aligned}$$

which, together with the definitions of $\beta_{p,i}$ in (64), implies

$$\mathcal{A}_i \mathcal{A}_{l+1}^* (y'_{l+1} - \bar{y}_{l+1}) = \sigma^{-1} \beta_{p,i+1} \quad \forall i = 0, \dots, l. \quad (86)$$

Now, by using (79), (86) and the definitions of \tilde{y}'_l and y'_l , we have

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

That is, (84) holds for $i = l$. Now assume that we have proven $\tilde{y}'_i = y'_i$ for all $i \geq k+1$ with $k+1 \leq l$. We shall next prove that (84) holds for $i = k$. Again, by using the definitions of \tilde{y}'_k and y'_k and noting

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

we obtain that

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

which, shows that (84) holds for $i = k$. Thus, (84) holds.

By applying our induction assumption to problem (80), we obtain equivalently that

$$u^+ = \operatorname{argmin}_u \tilde{\mathcal{L}}_\sigma(u, \bar{y}_{\leq l}; \bar{y}_{l+1}, \bar{v}, \bar{z}, \bar{x}) + \langle \delta_{\tilde{\theta}}, u \rangle + \frac{\sigma}{2} \|u - \bar{u}\|_{\mathcal{T}_f}^2, \quad (87)$$

$$y_i^+ = \operatorname{argmin}_{y_i} \tilde{\mathcal{L}}_\sigma(u^+, (y_{\leq i-1}^+, y_i, \tilde{y}'_{\geq i+1}); \bar{y}_{l+1}, \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2} \|y_i - \bar{y}_i\|_{\mathcal{T}_{\theta_i}}^2, \quad i = 1, \dots, l, \quad (88)$$

where we use the facts that $\mathcal{T}_{\tilde{f}_1} = \mathcal{T}_f$ and $\mathcal{T}_{\tilde{\theta}_i} = \mathcal{T}_{\theta_i}$ for $i = 1, \dots, l$. By combining (82) and the definitions of $\bar{\delta}_\theta$ and $\delta_{\tilde{\theta}}$ defined in (66) and (81), respectively, we derive that

$$\bar{\delta}_\theta = \sum_{i=1}^l \beta_{i,1} + \beta_{l+1,1} = \sum_{i=1}^l \tilde{\beta}_{i,1} + \beta_{l+1,1} = \delta_{\tilde{\theta}} + \beta_{l+1,1}. \quad (89)$$

By direct calculations,

$$\tilde{\mathcal{L}}_\sigma(u, \bar{y}_{\leq l}; \bar{y}_{l+1}, \bar{v}, \bar{z}, \bar{x}) = \mathcal{L}_\sigma(u, \bar{y}, \bar{v}, \bar{z}; \bar{x}) + \langle \beta_{l+1,1}, u \rangle + \sum_{i=1}^l \langle \beta_{l+1,i+1}, \bar{y}_i \rangle. \quad (90)$$

Using (84), (86) and the definition of $\tilde{\mathcal{L}}_\sigma$, we have for $i = 1, \dots, l$ that

$$\begin{aligned} &\tilde{\mathcal{L}}_\sigma(u^+, (y_{\leq i-1}^+, y_i, \tilde{y}'_{\geq i+1}); \bar{y}_{l+1}, \bar{v}, \bar{z}, \bar{x}) - \mathcal{L}_\sigma(u^+, (y_{\leq i-1}^+, y_i, y'_{\geq i+1}), \bar{v}, \bar{z}; \bar{x}) \\ &= \tilde{\mathcal{L}}_\sigma(u^+, (y_{\leq i-1}^+, y_i, y'_{i+1}, \dots, y'_l); \bar{y}_{l+1}, \bar{v}, \bar{z}, \bar{x}) - \mathcal{L}_\sigma(u^+, (y_{\leq i-1}^+, y_i, y'_{\geq i+1}), \bar{v}, \bar{z}; \bar{x}) \\ &= \langle \beta_{p,i+1}, y_i \rangle + \langle \sigma \mathcal{A}_i \mathcal{A}_{l+1}^* (\bar{y}_{l+1} - y'_{l+1}), y_i \rangle + c_i \\ &= c_i, \end{aligned} \quad (91)$$

where c_i is a constant term given by

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

Thus, by using (89), (90) and (91) we know that (87) and (88) can be rewritten as

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

which, together with (77), shows that the equivalence between (68) and (69) holds for $p = l + 1$. The proof of this part is completed.

Part two. In this part, we prove the equivalence between (69) and (70). Again, for the case $p = 1$, it follows directly from Proposition 2.2.

Assume that the equivalence between (69) and (70) holds for all $p \leq l$. 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 2.2, we consider the following optimization problem with respect to u and y_{l+1} :

$$\begin{aligned}
\min \quad & f_0(u) + \theta_{l+1}(y_{l+1}) + \sum_{i=1}^l \theta_i(\bar{y}_i) + g(\bar{v}) + \varphi(\bar{z}) \\
\text{s.t.} \quad & \mathcal{F}^*u + \mathcal{A}_{l+1}^* y_{l+1} = \bar{c} - \mathcal{A}_{\leq l}^* \bar{y}_{\leq l}.
\end{aligned} \tag{92}$$

The augmented Lagrangian function associated with problem (92) is given as follows:

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

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 (69) as

$$u^+ = \operatorname{argmin}_u \mathcal{L}_\sigma^0(u, \bar{y}_{l+1}; \bar{y}_{\leq l}, \bar{v}, \bar{z}, \bar{x}) + \langle \beta_{l+1,1}, u \rangle + \frac{\sigma}{2} \|u - \bar{u}\|_{\mathcal{T}_{f_0}}^2. \tag{93}$$

By using the definition of y_{l+1}' given in (67), we have

$$y_{l+1}' = \mathcal{E}_{\theta_{l+1}}^{-1} (\sigma^{-1} (b_{l+1} - \mathcal{A}_{l+1} \bar{x}) + \mathcal{T}_{\theta_{l+1}} \bar{y}_{l+1} + \mathcal{A}_{l+1} \mathcal{A}_{l+1}^* \bar{y}_{l+1} + \mathcal{A}_{l+1} \bar{\gamma}). \tag{94}$$

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

$$y'_{l+1} = \operatorname{argmin}_{y_{l+1}} \mathcal{L}_\sigma^0(\bar{u}, y_{l+1}; \bar{y}_{\leq l}, \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2} \|y_{l+1} - \bar{y}_{l+1}\|_{\mathcal{T}_{\theta_{l+1}}}^2. \quad (95)$$

Then, by applying Proposition 2.2 to problem (92) with respect to u and y_{l+1} , we know that problem (93) is equivalent to

$$u^+ = \operatorname{argmin}_u \mathcal{L}_\sigma^0(u, y'_{l+1}; \bar{y}_{\leq l}, \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2} \|u - \bar{u}\|_{\mathcal{T}_{f_0}}^2. \quad (96)$$

In order to apply our induction assumption to problem (96), we need to consider the following optimization problem with respect to $(u, y_{\leq l})$:

$$\begin{aligned} \min \quad & f(u) + \sum_{i=1}^l \theta_i(y_i) + \theta_{l+1}(y'_{l+1}) + g(\bar{v}) + \varphi(\bar{z}) \\ \text{s.t.} \quad & \mathcal{F}^*(u) + \mathcal{A}_{\leq l}^* y_{\leq l} = \bar{c} - \mathcal{A}_{l+1}^* y'_{l+1}. \end{aligned} \quad (97)$$

The augmented Lagrangian function associated with problem (97) is given by

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

Define

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

For problem (97), we define the following associated terms

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

and for $i = l-1, l-2, \dots, 1$,

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

The auxiliary linear term $\widehat{\delta}$ associated with problem (97) is given by

$$\widehat{\delta} = \sum_{i=1}^l \widehat{\beta}_{i,1}. \quad (98)$$

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

$$\widehat{\beta}_{i,j} = \beta_{i,j} \quad \forall j = 1, \dots, i. \quad (99)$$

Similar to what we have done in part one, we shall first prove that $\widehat{\beta}_{l,j} = \beta_{l,j}$ for $j = 1, 2, \dots, l$. In fact, for $j = 1, \dots, l$, we have

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

where the third equation follows from (94) and simple calculations. This shows that (99) holds for $i = l$ and $j = 1, \dots, l$. Now we assume that $\widehat{\beta}_{i,j} = \beta_{i,j}$ for all $i \geq k+1$ with $k+1 \leq l$ and $j = 1, \dots, i$. Next, we shall prove that (99) holds for $i = k$ and $j = 1, \dots, k$. By direct calculations, we know for $j = 1, \dots, k$ that

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

which, shows that (99) holds for $i = k$ and $j = 1, \dots, k$. Therefore, we have shown that (99) holds. For $i = l, l-1, \dots, 1$, define $\widehat{y}'_i \in \mathcal{Y}_i$ as

$$\begin{aligned}
\widehat{y}'_i &= \operatorname{argmin}_{y_i} \widehat{\mathcal{L}}_{\sigma}(\bar{u}, (\bar{y}_{\leq i-1}, y_i, \widehat{y}'_{\geq i+1}); y'_{l+1}, \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2} \|y_i - \bar{y}_i\|_{\mathcal{T}_{\theta_i}}^2 \\
&= \mathcal{E}_{\theta_i}^{-1}(\sigma^{-1} b_i - \sigma^{-1} \mathcal{A}_i \bar{x} + \mathcal{T}_{\theta_i} \bar{y}_i + \mathcal{A}_i \mathcal{A}_i^* \bar{y}_i - \mathcal{A}_i \Gamma(\bar{u}, (\bar{y}_{\leq i-1}, \bar{y}_i, \widehat{y}'_{\geq i+1}, y'_{l+1}), \bar{v}, \bar{z})), \quad (100)
\end{aligned}$$

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

$$\widehat{y}'_i = y'_i \quad \forall i = 1, \dots, l. \quad (101)$$

From (100), we know that

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

which is exactly the same as y'_l defined in (67). This shows that (101) 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 (101) holds for $i = k$. Again, by using the definition of \widehat{y}'_k in (100) and the definition of y'_k in (67), we see that

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

Thus, (101) is proven to be true.

By direct calculations, we obtain from (98) and (99) that

$$\mathcal{L}_{\sigma}^0(u, y'_{l+1}; \bar{y}_{\leq l}, \bar{v}, \bar{z}, \bar{x}) - \widehat{\mathcal{L}}_{\sigma}(u, \bar{y}_{\leq l}; y'_{l+1}, \bar{v}, \bar{z}, \bar{x}) = \sum_{i=1}^l \langle \beta_{i,1}, u \rangle = \langle \widehat{\delta}, u \rangle. \quad (102)$$

By using (102) and $\mathcal{T}_{f_0} \equiv \mathcal{T}_f$, we can reformulate problem (96) equivalently as

$$u^+ = \operatorname{argmin}_u \widehat{\mathcal{L}}_{\sigma}(u, \bar{y}_{\leq l}; y'_{l+1}, \bar{v}, \bar{z}, \bar{x}) + \langle \widehat{\delta}, u \rangle + \frac{\sigma}{2} \|u - \bar{u}\|_{\mathcal{T}_f}^2. \quad (103)$$

Then, from our induction assumption we know that problem (103) can be equivalently recast as

$$\begin{cases} \hat{y}'_i &= \operatorname{argmin}_{y_i} \hat{\mathcal{L}}_\sigma(\bar{u}, (\bar{y}_{\leq i-1}, y_i, \hat{y}'_{\geq i+1}); y'_{l+1}, \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2} \|y_i - \bar{y}_i\|_{\mathcal{T}_{\theta_i}}^2, & i = l, l-1, \dots, 1, \\ u^+ &= \operatorname{argmin}_u \hat{\mathcal{L}}_\sigma(u, \hat{y}'_{\leq l}; y'_{l+1}, \bar{v}, \bar{z}, \bar{x}) + \frac{\sigma}{2} \|u - \bar{u}\|_{\mathcal{T}_f}^2. \end{cases} \quad (104)$$

By using (101) and observing

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

we know that (104) is equivalent to

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

which, together with (95), shows that the equivalence between (69) and (70) holds for $p = l + 1$. This completes the proof to the second part of this proposition. \square

Proposition 3.2 *For any $k \geq 0$, the point $(x^{k+1}, y^{k+1}, v^{k+1}, z^{k+1})$ obtained by Algorithm SCB-SPADMM for solving problem (46) can be generated exactly according to the following iteration:*

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

Proof. The (u^{k+1}, y^{k+1}) part directly follows from Proposition 3.1. 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. \square

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

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

and

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

In order to prove the convergence of our algorithm SCB-SPADMM for solving problem (46), we need the following proposition.

Proposition 3.3 *It holds that*

$$\mathcal{F}_{p+1}\mathcal{F}_{p+1}^* + \sigma^{-1}\Sigma_{f_{p+1}} + \begin{pmatrix} \hat{\mathcal{T}}_{f_p} & \\ & \mathcal{T}_{\theta_p} \end{pmatrix} \succ 0 \Leftrightarrow \mathcal{F}\mathcal{F}^* + \sigma^{-1}\Sigma_f + \mathcal{T}_f \succ 0, \quad (105)$$

$$\mathcal{G}_{q+1}\mathcal{G}_{q+1}^* + \sigma^{-1}\Sigma_{g_{q+1}} + \begin{pmatrix} \hat{\mathcal{T}}_{g_q} & \\ & \mathcal{T}_{\varphi_q} \end{pmatrix} \succ 0 \Leftrightarrow \mathcal{G}\mathcal{G}^* + \sigma^{-1}\Sigma_g + \mathcal{T}_g \succ 0. \quad (106)$$

Proof. We only need to prove (105) as (106) can be obtained in the similar manner. For $i = 3, \dots, p+1$, we have

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

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 $i \geq 3$, by the Schur complement condition for ensuring the positive definiteness of linear operators, we have

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

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

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

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. \square

Note that in the context of the multi-block convex optimization problem (46), Assumption 2.1 takes the following form:

Assumption 3.1 *There exists $(\hat{u}, \hat{y}, \hat{v}, \hat{z}) \in \text{ri}(\text{dom } f) \times \mathcal{Y} \times \text{ri}(\text{dom } g) \times \mathcal{Z}$ 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 3.1 *Let Σ_f and Σ_g be the two self-adjoint and positive semidefinite operators defined by (25) and (26), respectively. Suppose that the solution set of problem (46) is nonempty and that Assumption 3.1 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}_{θ_i} is defined in (49) for $1 \leq i \leq p$ and \mathcal{T}_{φ_j} is defined in (50) for $1 \leq j \leq q$. Then, under the condition either (a) $\tau \in (0, (1 + \sqrt{5})/2)$ or (b) $\tau \geq (1 + \sqrt{5})/2$ but $\sum_{k=0}^{\infty} (\|\mathcal{G}^*(v^{k+1} - v^k) + \mathcal{B}^*(z^{k+1} - z^k)\|^2 + \tau^{-1} \|\mathcal{F}^*u^{k+1} + \mathcal{A}^*y^{k+1} + \mathcal{G}^*v^{k+1} + \mathcal{B}^*z^{k+1} - c\|^2) < \infty$, 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 (46) 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 (46) 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 $\tau \geq 2$ but $\sum_{k=0}^{\infty} \|\mathcal{G}^*v^{k+1} + \mathcal{B}^*z^{k+1} - c\|^2 < \infty$.*

Proof. By combining Theorem 2.1 with Proposition 3.2 and Proposition 3.3, we can readily obtain the conclusions of this theorem. \square

Remark 3.1 *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 (46) and its dual (47). Suppose that the solution set of problem (46) is nonempty and that Assumption 3.1 holds. Then in order that (u^*, y^*, v^*, z^*) be an optimal solution for (46) and x^* be an optimal solution for (47), it is necessary and sufficient that (u^*, y^*, v^*, z^*) and x^* satisfy

$$\begin{cases} \mathcal{F}^*u + \sum_{i=1}^p \mathcal{A}_i^*y_i + \mathcal{G}^*v + \sum_{j=1}^q \mathcal{B}_j^*z_j = c, \\ f(u) + f^*(-\mathcal{F}x) = \langle -\mathcal{F}x, u \rangle, \quad \theta_i(y_i) + \theta_i^*(-\mathcal{A}_i x) = \langle -\mathcal{A}_i x, y_i \rangle, \quad i = 1, \dots, p, \\ g(v) + g^*(-\mathcal{G}x) = \langle -\mathcal{G}x, v \rangle, \quad \varphi_j(z_j) + \varphi_j^*(-\mathcal{B}_j x) = \langle -\mathcal{B}_j x, z_j \rangle, \quad j = 1, \dots, q. \end{cases} \quad (107)$$

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:

$$\eta = \max\{\eta_P, \eta_f, \eta_g, \eta_\theta, \eta_\varphi\}, \quad (108)$$

where

$$\eta_P = \frac{\|\mathcal{F}^*u + \mathcal{A}^*y + \mathcal{G}^*v + \mathcal{B}^*z - c\|}{1 + \|c\|}, \quad \eta_f = \frac{\|u - \text{Prox}_f(u - \mathcal{F}x)\|}{1 + \|u\| + \|\mathcal{F}x\|}, \quad \eta_g = \frac{\|v - \text{Prox}_g(v - \mathcal{G}x)\|}{1 + \|v\| + \|\mathcal{G}x\|},$$

$$\eta_\theta = \max_{i=1,\dots,p} \frac{\|y_i - \text{Prox}_{\theta_i}(y_i - \mathcal{A}_i x)\|}{1 + \|y_i\| + \|\mathcal{A}_i x\|}, \quad \eta_\varphi = \max_{j=1,\dots,q} \frac{\|z_j - \text{Prox}_{\varphi_j}(z_j - \mathcal{B}_j x)\|}{1 + \|z_j\| + \|\mathcal{B}_j x\|}.$$

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

$$\begin{aligned} \min \quad & \frac{1}{2} \langle X, \mathcal{Q}X \rangle + \langle C, X \rangle \\ \text{s.t.} \quad & \mathcal{A}_E X = b_E, \quad \mathcal{A}_I X \geq b_I, \quad X \in \mathcal{S}_+^n \cap \mathcal{K} \end{aligned} \quad (109)$$

and its dual problem

$$\begin{aligned} \max \quad & -\delta_{\mathcal{K}}^*(-Z) + \langle b_I, y_I \rangle - \frac{1}{2} \langle X', \mathcal{Q}X' \rangle + \langle b_E, y_E \rangle \\ \text{s.t.} \quad & Z + \mathcal{A}_I^* y_I - \mathcal{Q}X' + S + \mathcal{A}_E^* y_E = C, \quad y_I \geq 0, \quad S \in \mathcal{S}_+^n. \end{aligned} \quad (110)$$

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 (110) 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 (110) can be rewritten as follows:

$$\begin{aligned} \max \quad & -\delta_{\mathcal{K}}^*(-Z) + \langle b_I, y_I \rangle - \frac{1}{2} \|\Xi\|_F^2 + \langle b_E, y_E \rangle \\ \text{s.t.} \quad & Z + \mathcal{A}_I^* y_I + \mathcal{B}^* \Xi + S + \mathcal{A}_E^* y_E = C, \quad y_I \geq 0, \quad S \in \mathcal{S}_+^n. \end{aligned} \quad (111)$$

Note that now the augmented Lagrangian function associated with (111) 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 $\bar{Z}, \bar{y}_I, \bar{S}, \bar{y}_E$ and \bar{X} , denote

$$\Xi^+ := \underset{\Xi}{\text{argmin}} \frac{1}{2} \|\Xi\|^2 + \frac{\sigma}{2} \|\bar{Z} + \mathcal{A}_I^* \bar{y}_I + \mathcal{B}^* \Xi + \bar{S} + \mathcal{A}_E^* \bar{y}_E - C + \sigma^{-1} \bar{X}\|^2 = -(\mathcal{I} + \sigma \mathcal{B} \mathcal{B}^*)^{-1} \mathcal{B} \bar{R},$$

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

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

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 $\lambda > 0$,

$$Z^+ = \operatorname{argmin} \delta_{\mathcal{K}}^*(-Z) + \frac{\lambda}{2}\|Z - \bar{Z}\|^2 = \bar{Z} + \frac{1}{\lambda}\Pi_{\mathcal{K}}(-\lambda\bar{Z}), \quad (112)$$

where (112) follows from the following Moreau decomposition:

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

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 $\Lambda = \operatorname{diag}(\lambda)$ and $\lambda = (\lambda_1, \dots, \lambda_n)^T$ is the vector of eigenvalues of B . Then

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

where $\hat{X} = P^T X P$, $H_{ij} = \sqrt{\frac{\lambda_i + \lambda_j}{2}}$, $\mathcal{B}X = H \circ (P^T X P)$ 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 $\operatorname{rank}(B) = 0$ and \mathcal{K} is a polyhedral cone, problem (109) reduces to the SDP problem considered in [17]. In our experiments, we test both the cases where $\operatorname{rank}(B) = 5$ and $\operatorname{rank}(B) = 10$. All the linear constraints are extracted from the numerical test examples in [17] (Section 4.1). For instance, we construct QSDP-BIQ problem sets based on the formulation in [17] as follows:

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

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 spirit, 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 [9] (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 quad-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 (109) and its dual (111) by using the following relative residual obtained from the general optimality condition (107):

$$\eta_{\text{qsdp}} = \max\{\eta_P, \eta_D, \eta_Z, \eta_{S_1}, \eta_{S_2}\}, \quad (113)$$

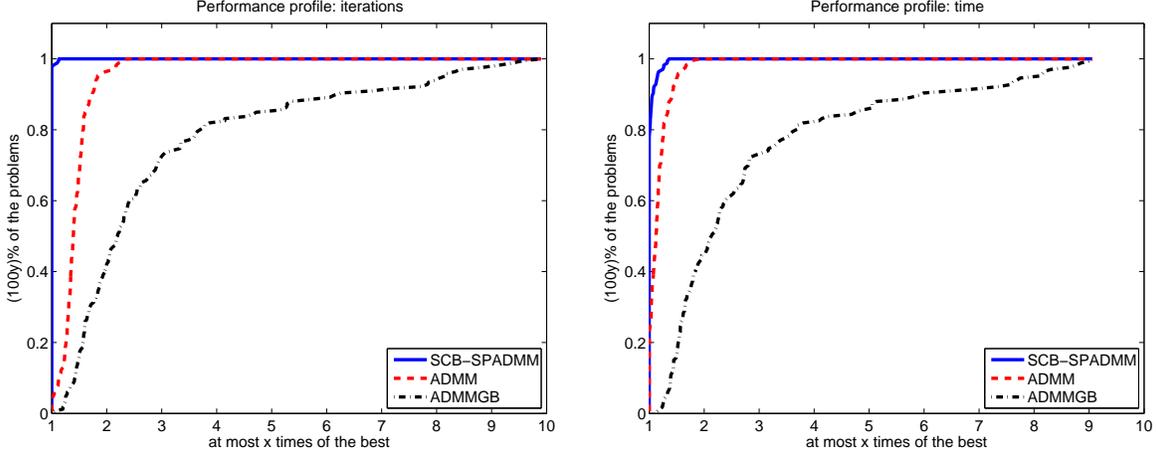


Figure 1: Performance profiles of SCB-SPADMM, ADMM and ADMMGB for the tested large scale QSDP.

where

$$\eta_P = \frac{\|\mathcal{A}_E X - b_E\|}{1 + \|b_E\|}, \quad \eta_D = \frac{\|Z + \mathcal{B}^* \Xi + S + \mathcal{A}_E^* y_E - C\|}{1 + \|C\|}, \quad \eta_Z = \frac{\|X - \Pi_{\mathcal{K}}(X - Z)\|}{1 + \|X\| + \|Z\|},$$

$$\eta_{S_1} = \frac{|\langle S, X \rangle|}{1 + \|S\| + \|X\|}, \quad \eta_{S_2} = \frac{\|X - \Pi_{\mathcal{S}_+^n}(X)\|}{1 + \|X\|}.$$

We terminate the solvers SCB-SPADMM, ADMM and ADMMGB when $\eta_{\text{qsdp}} < 10^{-6}$ with the maximum number of iterations set at 25000.

Table 4 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 $\text{rank}(B) = 10$, since we obtain similar results for the case of $\text{rank}(B) = 5$. 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 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 $G \in \mathcal{S}^n$:

$$\begin{aligned} \min \quad & \frac{1}{2} \|H \circ (X - G)\|_F^2 + \langle C, X \rangle \\ \text{s.t.} \quad & \mathcal{A}_E X = b_E, \quad X \in \mathcal{S}_+^n \cap \mathcal{K}, \end{aligned} \tag{114}$$

where $H \in \mathcal{S}^n$ is a nonnegative weight matrix, $\mathcal{A}_E : \mathcal{S}^n \rightarrow \mathfrak{R}^{m_E}$ is a linear map, $G \in \mathcal{S}^n$, $C \in \mathcal{S}^n$ and $b_E \in \mathfrak{R}^{m_E}$ 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 (109) 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:

$$\begin{aligned} \min \quad & \|H \circ (X - G)\|_2 + \langle C, X \rangle \\ \text{s.t.} \quad & \mathcal{A}_E X = b_E, \quad X \in \mathcal{S}_+^n \cap \mathcal{K}. \end{aligned} \tag{115}$$

Note, in (115), instead of the Frobenius norm, we use the spectral norm. By introducing a slack variable Y , we can reformulate problem (115) as

$$\begin{aligned} \min \quad & \|Y\|_2 + \langle C, X \rangle \\ \text{s.t.} \quad & H \circ (X - G) = Y, \quad \mathcal{A}_E X = b_E, \quad X \in \mathcal{S}_+^n \cap \mathcal{K}. \end{aligned} \tag{116}$$

The dual of problem (116) is given by

$$\begin{aligned} \max \quad & -\delta_{\mathcal{K}}^*(-Z) + \langle H \circ G, \Xi \rangle + \langle b_E, y_E \rangle \\ \text{s.t.} \quad & Z + H \circ \Xi + S + \mathcal{A}_E^* y_E = C, \quad \|\Xi\|_* \leq 1, \quad S \in \mathcal{S}_+^n, \end{aligned} \tag{117}$$

which is obviously equivalent to the following problem

$$\begin{aligned} \max \quad & -\delta_{\mathcal{K}}^*(-Z) + \langle H \circ G, \Xi \rangle + \langle b_E, y_E \rangle \\ \text{s.t.} \quad & Z + H \circ \Xi + S + \mathcal{A}_E^* y_E = C, \quad \|\Gamma\|_* \leq 1, \quad S \in \mathcal{S}_+^n, \\ & \mathcal{D}^* \Gamma - \mathcal{D}^* \Xi = 0, \end{aligned} \tag{118}$$

where $\mathcal{D} : \mathcal{S}^n \rightarrow \mathcal{S}^n$ is a nonsingular linear operator. Note that SCB-SPADMM can not be directly applied to solve the problem (117) while the equivalent reformulation (118) fits our model nicely.

In our numerical test, matrix \widehat{G} is the gene correlation matrix from [13]. 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 $G_{ii} = 1$, $i = 1, \dots, n$. 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

$$\text{tmp} = \text{kron}(\text{ones}(25,25), H_0); H = \text{tmp}(1:n, 1:n); H = (H' + H) / 2.$$

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

Table 1: The performance of SCB-SPADMM, ADMM, ADMMGB on Frobenius norm H-weighted NCM problems (dual of (114)) (accuracy = 10^{-6}). 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”.

problem	n_s	α	iteration			η_{qsdp}			η_{gap}			time		
			scb	admm	gb	scb	admm	gb	scb	admm	gb	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
Arabidopsis	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	1255	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
	1255	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
hereditarybc	1869	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
	1869	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

G and H are generated in the same way as in [11]. 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 (116). Here, with reference to the general relative residue (108), we measure the accuracy of an approximate optimal solution (X, Z, Ξ, S, y_E) for spectral norm H-weighted NCM problem (115) (equivalently (116)) and its dual (117) (equivalently (118)) by using the following relative residual derived from the general optimality condition (107):

$$\eta_{\text{sncm}} = \max\{\eta_P, \eta_D, \eta_Z, \eta_{S_1}, \eta_{S_2}, \eta_\Xi\}, \quad (119)$$

where

$$\eta_P = \frac{\|\mathcal{A}_E X - b_E\|}{1 + \|b_E\|}, \quad \eta_D = \frac{\|Z + H \circ \Xi + S + \mathcal{A}_E^* y_E\|}{1 + \|Z\| + \|S\|}, \quad \eta_Z = \frac{\|X - \Pi_{\mathcal{K}}(X - Z)\|}{1 + \|X\| + \|Z\|},$$

$$\eta_{S_1} = \frac{|\langle S, X \rangle|}{1 + \|S\| + \|X\|}, \quad \eta_{S_2} = \frac{\|X - \Pi_{S_+^n}(X)\|}{1 + \|X\|}, \quad \eta_\Xi = \frac{\|\Xi - \Pi_{\{X \in \mathbb{R}^{n \times n} : \|X\|_* \leq 1\}}(\Xi - H \circ (X - G))\|}{1 + \|\Xi\| + \|H \circ (X - G)\|}.$$

Firstly, numerical results for solving F-norm H-weighted NCM problems (115) are reported. We compare all three algorithms, namely SCB-SPADMM, ADMM, ADMMGB using the relative residue (113). We terminate the solvers when $\eta_{\text{qsdp}} < 10^{-6}$ with the maximum number of iterations set at 25000.

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 (115). As mentioned before, SCB-SPADMM is applied to solve the problem (118) rather than (117). We implemented all the algorithms for solving problem (118) using the relative residue (119). We terminate the solvers when $\eta_{\text{sncm}} < 10^{-5}$ with the maximum number of iterations set at 25000. In Table 2, we report detailed numerical results for SCB-SPADMM, ADMM and

Table 2: The performance of SCB-SPADMM, ADMM, ADMMGB on spectral norm H-weighted NCM problem (118) (accuracy = 10^{-5}). 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”.

problem	n_s	α	iteration	η_{sncm}			η_{gap}			time		
			scb admm gb	scb admm gb	scb admm gb	scb admm gb	scb admm gb	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
Arabidopsis	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	1255	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
	1255	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

Table 3: The performance of LADMM, LADMMGB on spectral norm H-weighted NCM problem(117) (accuracy = 10^{-5}). In the table, “lgb” stands for LADMMGB. The computation time is in the format of “hours:minutes:seconds”.

problem	n_s	α	iteration	η_{sncm}		η_{gap}		time	
			ladmm lgb	ladmm lgb	ladmm lgb	ladmm lgb	ladmm lgb		
Lymph	587	0.10	8401 25000	9.9-6	1.4-5	-1.6-5	-2.1-5	23:59	1:22:58
Lymph	587	0.05	13609 25000	9.9-6	2.3-5	-1.6-5	-4.2-5	39:29	1:18:50

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.

Observe that although there is no convergence guarantee, one may still apply the directly extended ADMM with 4 blocks to the original dual problem (117) 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 Gaussian back substitution proposed in [10] (we call the method LADMMGB here and use the parameter $\alpha = 0.99$ in the Gaussian back substitution step) can also be applied to the original problem (117). We have also implemented LADMM and LADMMGB in MATLAB. Our experiments show that solving the problem (117) directly is much slower than solving the equivalent problem (118). Thus, the reformulation of (117) to (118) 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, Sun and Toh in [22] and Yang, Sun and Toh in [21], respectively. Naturally, our next target is to extend the approach of [22, 21] to solve QSDP with an initial point generated by SCB-SPADMM. We will report our corresponding findings in subsequent works.

6 Acknowledgements

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.

References

- [1] C. CHEN, B. HE, Y. YE, AND X. YUAN, *The direct extension of admm for multi-block convex minimization problems is not necessarily convergent*, Optimization Online, (2013).
- [2] J. ECKSTEIN AND D. P. BERTSEKAS, *On the douglasrachford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming, 55 (1992), pp. 293–318.
- [3] M. FAZEL, T. K. PONG, D. F. SUN, AND P. TSENG, *Hankel matrix rank minimization with applications to system identification and realization*, SIAM Journal on Matrix Analysis and Applications, 34 (2013), pp. 946–977.
- [4] M. FORTIN AND R. GLOWINSKI, *Augmented Lagrangian methods*, vol. 15 of Studies in Mathematics and its Applications, North-Holland Publishing Co., Amsterdam, 1983. Applications to the numerical solution of boundary value problems, Translated from the French by B. Hunt and D. C. Spicer.
- [5] D. GABAY, *Applications of the method of multipliers to variational inequalities*, in Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, M. Fortin and R. Glowinski, eds., vol. 15 of Studies in Mathematics and Its Applications, Elsevier, 1983, pp. 299 – 331.
- [6] D. GABAY AND B. MERCIER, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Computers and Mathematics with Applications, 2 (1976), pp. 17–40.
- [7] R. GLOWINSKI, *Lectures on numerical methods for nonlinear variational problems*, vol. 65 of Tata Institute of Fundamental Research Lectures on Mathematics and Physics, Tata Institute of Fundamental Research, Bombay, 1980. Notes by M. G. Vijayasundaram and M. Adimurthi.

- [8] R. GLOWINSKI AND A. MARROCCO, *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de dirichlet non linéaires*, Revue Française d'Automatique, Informatique et Recherche Opérationnelle, 9 (1975), pp. 41–76.
- [9] B. HE, M. TAO, AND X. YUAN, *Alternating direction method with gaussian back substitution for separable convex programming*, SIAM Journal on Optimization, 22 (2012), pp. 313–340.
- [10] B. HE AND X. YUAN, *Linearized alternating direction method with gaussian back substitution for separable convex programming*, preprint, (2011).
- [11] K. JIANG, D. SUN, AND K.-C. TOH, *An inexact accelerated proximal gradient method for large scale linearly constrained convex sdp*, SIAM Journal on Optimization, 22 (2012), pp. 1042–1064.
- [12] O. KLOPP, *Noisy low-rank matrix completion with general sampling distribution*, Bernoulli, 20 (2014), pp. 282–303.
- [13] L. LI AND K.-C. TOH, *An inexact interior point method for l_1 -regularized sparse covariance selection*, Mathematical Programming Computation, 2 (2010), pp. 291–315.
- [14] W. MIAO, *Matrix Completion Models with Fixed Basis Coefficients and Rank Regularized Problems with Hard Constraints*, PhD thesis, National University of Singapore, 2013.
- [15] W. MIAO, S. PAN, AND D. F. SUN, *A rank-corrected procedure for matrix completion with fixed basis coefficients*, Technical Report, National University of Singapore, (2014).
- [16] S. NEGAHBAN AND M. J. WAINWRIGHT, *Restricted strong convexity and weighted matrix completion: Optimal bounds with noise*, The Journal of Machine Learning Research, 13 (2012), pp. 1665–1697.
- [17] D. F. SUN, K.-C. TOH, AND L. YANG, *A convergent proximal alternating direction method of multipliers for conic programming with 4-block constraints*, arXiv preprint arXiv:1404.5378, (2014).
- [18] M. TAO AND X. YUAN, *Recovering low-rank and sparse components of matrices from incomplete and noisy observations*, SIAM Journal on Optimization, 21 (2011), pp. 57–81.
- [19] J. WRIGHT, A. GANESH, S. RAO, Y. PENG, AND Y. MA, *Robust principal component analysis: Exact recovery of corrupted low-rank matrices by convex optimization*, in Proc. of Neural Information Processing Systems, vol. 3, 2009.
- [20] B. WU, *High-Dimensional Analysis on Matrix Decomposition with Application to Correlation Matrix Estimation in Factor Models*, PhD thesis, National University of Singapore, 2014.
- [21] L. YANG, D. F. SUN, AND K.-C. TOH, *Sdpnal +: A majorized semismooth newton-cg augmented lagrangian method for semidefinite programming with nonnegative constraints*, arXiv preprint arXiv:1406.0942, (2014).
- [22] X.-Y. ZHAO, D. F. SUN, AND K.-C. TOH, *A newton-cg augmented lagrangian method for semidefinite programming*, SIAM Journal on Optimization, 20 (2010), pp. 1737–1765.

Table 4: The performance of SCB-SPADMM, ADMM, ADMMGB on QSDP- θ_+ , QSDP-QAP, QSDP-BIQ and QSDP-RCP problems (accuracy = 10^{-6}). 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”.

problem	$m_E; n_s$	rank(B)	iteration			η_{qsdp}			η_{gap}			time		
			scb	admm	gb	scb	admm	gb	scb	admm	gb	scb	admm	gb
theta6	4375 ; 300	10	311	407	549	7.9-7	9.7-7	9.9-7	2.1-6	-1.6-6	-6.2-7	08	09	14
theta62	13390 ; 300	10	153	196	229	9.6-7	9.9-7	9.6-7	-1.1-7	9.6-8	-4.5-7	04	05	06
theta8	7905 ; 400	10	314	384	616	9.5-7	9.6-7	9.5-7	2.7-6	-1.3-6	-5.4-7	17	18	33
theta82	23872 ; 400	10	158	179	234	9.5-7	9.7-7	9.9-7	-3.7-8	-2.8-7	-8.2-7	10	09	13
theta83	39862 ; 400	10	200	177	219	9.3-7	9.6-7	9.4-7	6.2-9	1.4-7	-1.2-7	11	09	14
theta10	12470 ; 500	10	329	439	614	9.0-7	8.5-7	9.7-7	-2.5-6	1.5-6	5.8-7	27	33	50
theta102	37467 ; 500	10	150	187	235	8.7-7	9.4-7	9.9-7	6.4-7	2.9-7	-9.3-7	15	15	21
theta103	62516 ; 500	10	202	184	222	9.8-7	9.5-7	9.9-7	-4.2-8	6.9-8	-1.6-7	20	15	21
theta104	87245 ; 500	10	181	181	242	9.4-7	9.5-7	9.9-7	6.9-8	2.0-7	-2.8-7	20	15	23
theta12	17979 ; 600	10	343	441	703	9.9-7	8.3-7	9.9-7	3.0-6	1.4-6	-8.8-7	42	48	1:27
theta123	90020 ; 600	10	204	205	213	9.7-7	9.8-7	9.9-7	-9.1-8	6.6-8	-1.9-7	29	25	31
san200-0.7-1	5971 ; 200	10	2150	4758	5172	9.8-7	9.9-7	9.9-7	5.1-6	2.0-6	-3.5-6	15	26	36
sanr200-0.7	6033 ; 200	10	177	223	280	9.6-7	9.7-7	9.7-7	1.9-7	-6.0-8	1.7-8	02	02	03
c-fat200-1	18367 ; 200	10	2257	3027	3268	9.9-7	9.7-7	9.9-7	-2.6-6	-2.0-6	-2.2-6	24	26	35
hamming-8-4	11777 ; 256	10	2820	2945	3517	9.9-7	9.9-7	9.9-7	-6.0-7	-6.4-7	-1.1-6	53	49	1:09
hamming-9-8	2305 ; 512	10	3891	4980	5577	9.9-7	9.9-7	9.9-7	-3.4-6	-5.8-7	9.9-7	3:54	4:12	5:50
hamming-8-3-4	16129 ; 256	10	202	220	294	4.8-7	8.9-7	9.8-7	4.5-6	5.9-7	2.2-7	04	04	06
hamming-9-5-6	53761 ; 512	10	436	535	684	8.5-7	8.7-7	9.6-7	1.1-5	-1.7-6	-1.6-7	36	37	57
brock200-1	5067 ; 200	10	198	210	291	9.7-7	9.4-7	9.8-7	9.9-8	-2.9-7	-6.9-10	02	02	03
brock200-4	6812 ; 200	10	209	186	263	9.8-7	9.9-7	9.8-7	1.2-7	-2.6-9	-1.1-7	03	02	03
brock400-1	20078 ; 400	10	168	217	275	9.0-7	9.6-7	9.7-7	8.6-7	-4.9-8	6.2-9	11	10	15
keller4	5101 ; 171	10	669	909	963	9.9-7	9.9-7	9.9-7	-1.3-8	4.6-9	-8.4-8	06	07	09
p-hat300-1	33918 ; 300	10	468	829	2501	9.9-7	9.9-7	8.3-7	-8.7-7	2.1-7	-1.0-6	14	20	1:09
be250.1	251 ; 251	10	4126	7439	25000	9.6-7	9.9-7	1.3-6	-5.8-7	-8.6-7	-1.3-8	59	1:27	5:41
be250.2	251 ; 251	10	3604	6504	16322	9.8-7	9.9-7	9.9-7	-4.9-7	-6.8-7	-7.4-9	52	1:18	3:40
be250.3	251 ; 251	10	3562	5712	8501	9.9-7	9.9-7	9.7-7	-9.2-7	-9.4-7	9.3-7	52	1:08	1:57
be250.4	251 ; 251	10	4072	7668	25000	9.9-7	9.9-7	1.4-6	-2.1-6	2.8-6	-9.4-9	57	1:32	5:41
be250.5	251 ; 251	10	3210	4635	7406	9.9-7	9.9-7	9.9-7	-8.6-7	-8.8-7	1.4-6	46	55	1:41
be250.6	251 ; 251	10	3250	5580	9812	9.9-7	9.9-7	9.9-7	-2.8-7	-3.1-7	-3.6-7	46	1:05	2:10
be250.7	251 ; 251	10	3699	6562	13501	9.9-7	9.9-7	9.9-7	-6.5-7	-3.8-7	5.4-9	52	1:17	3:03
be250.8	251 ; 251	10	3507	4712	7701	9.9-7	9.9-7	9.6-7	-9.7-7	-1.0-6	5.1-7	50	56	1:43
be250.9	251 ; 251	10	3678	7292	21001	9.9-7	9.9-7	9.9-7	-4.1-7	-7.2-7	-1.2-8	53	1:28	4:57
be250.10	251 ; 251	10	3305	5752	10500	9.9-7	9.9-7	9.9-7	-1.1-6	-8.2-7	-3.7-8	49	1:06	2:19

Table 4: The performance of SCB-SPADMM, ADMM, ADMMGB on QSDP- θ_+ , QSDP-QAP, QSDP-BIQ and QSDP-RCP problems (accuracy = 10^{-6}). 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”.

problem	$m_E; n_s$	rank(B)	iteration			η_{qsdp}			η_{gap}			time		
			scb	admm	gb	scb	admm	gb	scb	admm	gb	scb	admm	gb
bqp100-1	101 ; 101	10	1376	2134	3067	9.9-7	9.9-7	9.9-7	2.6-7	-1.9-7	-5.1-7	05	06	10
bqp100-2	101 ; 101	10	3109	4319	7107	9.9-7	9.9-7	9.9-7	-1.8-7	-7.2-7	-5.3-7	10	13	22
bqp100-3	101 ; 101	10	1751	2371	6276	9.9-7	9.9-7	9.9-7	-2.7-6	-3.1-6	4.7-7	06	06	20
bqp100-4	101 ; 101	10	2646	3986	13901	9.9-7	9.9-7	9.1-7	-4.0-7	-6.6-7	-3.3-8	09	11	45
bqp100-5	101 ; 101	10	1979	3001	6901	9.9-7	9.9-7	9.7-7	-3.7-7	-1.5-7	1.7-8	07	08	22
bqp100-6	101 ; 101	10	1316	2083	2937	9.4-7	9.9-7	9.9-7	1.1-7	3.3-7	-9.5-7	05	06	11
bqp100-7	101 ; 101	10	1787	2341	3664	9.9-7	9.9-7	9.9-7	-5.5-7	-5.1-7	-1.3-6	06	06	12
bqp100-8	101 ; 101	10	1820	3337	9612	9.9-7	9.9-7	9.9-7	7.3-7	8.9-8	1.1-8	06	09	32
bqp100-9	101 ; 101	10	1948	4146	15901	9.9-7	9.9-7	9.9-7	-2.2-6	-6.7-7	2.6-9	07	11	52
bqp100-10	101 ; 101	10	3207	5077	12101	9.9-7	9.9-7	9.9-7	8.0-8	4.3-7	2.7-8	10	15	38
bqp250-1	251 ; 251	10	3931	5941	11758	9.6-7	9.9-7	9.9-7	-1.2-6	-1.5-6	1.2-7	57	1:10	2:39
bqp250-2	251 ; 251	10	4007	5774	9704	9.5-7	9.9-7	9.9-7	-6.6-7	-2.3-7	-1.2-6	57	1:07	2:11
bqp250-3	251 ; 251	10	4112	5708	12202	9.9-7	9.9-7	9.9-7	-3.9-6	3.8-8	3.0-6	57	1:05	2:40
bqp250-4	251 ; 251	10	3158	4290	9671	9.9-7	9.9-7	9.9-7	-5.5-7	-2.4-6	4.5-6	45	52	2:13
bqp250-5	251 ; 251	10	4430	7349	22802	9.9-7	9.9-7	9.9-7	-2.0-6	3.6-6	-1.3-8	1:02	1:29	5:13
bqp250-6	251 ; 251	10	2871	5122	7801	9.9-7	9.9-7	9.9-7	-1.2-6	-1.3-6	-2.5-7	42	1:01	1:47
bqp250-7	251 ; 251	10	3991	5570	11508	9.9-7	9.9-7	9.9-7	-2.2-6	-2.0-6	-2.7-6	57	1:04	2:31
bqp250-8	251 ; 251	10	2882	4008	5501	9.9-7	9.8-7	9.8-7	-2.0-7	-7.1-7	-1.0-6	40	45	1:14
bqp250-9	251 ; 251	10	4127	6279	11998	9.7-7	9.9-7	9.9-7	-5.1-7	-3.9-7	3.8-6	58	1:11	2:38
bqp250-10	251 ; 251	10	3044	4185	7986	9.9-7	9.9-7	9.9-7	-9.3-7	-7.5-7	-2.5-6	43	48	1:43
bqp500-1	501 ; 501	10	6003	8391	13416	9.9-7	9.9-7	9.9-7	-3.9-7	-7.3-7	-5.4-7	6:01	7:05	13:34
bqp500-2	501 ; 501	10	6601	10203	25000	9.7-7	9.9-7	3.4-6	-4.2-7	-1.2-7	1.8-5	6:52	8:43	25:23
bqp500-3	501 ; 501	10	7450	10517	21140	9.9-7	9.9-7	9.9-7	7.6-7	-4.3-6	1.1-6	7:31	8:46	21:10
bqp500-4	501 ; 501	10	7035	9903	23551	9.6-7	9.9-7	9.9-7	-3.3-7	-1.3-6	2.6-6	7:08	8:12	23:36
bqp500-5	501 ; 501	10	6164	8406	20533	9.9-7	9.9-7	9.9-7	-8.8-7	-4.8-7	2.8-6	6:30	7:04	20:37
bqp500-6	501 ; 501	10	6905	8659	25000	9.8-7	9.9-7	1.4-4	-3.8-7	-1.5-6	-1.8-4	7:13	7:30	25:44
bqp500-7	501 ; 501	10	6587	9038	18072	9.9-7	9.9-7	9.9-7	-6.8-7	2.5-7	2.8-6	6:41	7:39	18:13
bqp500-8	501 ; 501	10	6300	8832	16496	9.9-7	9.9-7	9.9-7	1.3-6	-1.6-6	5.8-6	6:24	7:17	16:20
bqp500-9	501 ; 501	10	6532	9015	18065	9.9-7	9.9-7	9.9-7	9.9-7	-6.5-7	-3.5-6	6:39	7:37	18:10
bqp500-10	501 ; 501	10	7199	9787	24119	9.9-7	9.9-7	9.9-7	-1.9-6	2.1-6	-2.3-6	7:09	8:12	24:15
gka1d	101 ; 101	10	1600	2266	4068	9.8-7	9.9-7	9.7-7	-4.2-7	-8.8-7	7.4-7	06	06	13
gka2d	101 ; 101	10	1903	3097	5601	9.9-7	9.9-7	9.3-7	-5.9-7	-2.4-7	-3.8-8	07	09	21
gka3d	101 ; 101	10	2431	3101	5618	9.9-7	9.9-7	9.9-7	-2.6-7	-3.8-7	1.7-8	08	09	19

Table 4: The performance of SCB-SPADMM, ADMM, ADMMGB on QSDP- θ_+ , QSDP-QAP, QSDP-BIQ and QSDP-RCP problems (accuracy = 10^{-6}). 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”.

problem	$m_E; n_s$	rank(B)	iteration			η_{qsdp}			η_{gap}			time		
			scb	admm	gb	scb	admm	gb	scb	admm	gb	scb	admm	gb
gka4d	101 ; 101	10	2266	2787	6632	9.9-7	9.9-7	9.9-7	2.3-7	-4.4-7	-1.9-8	08	09	22
soybean-large-2	308 ; 307	10	1267	1717	11208	9.9-7	9.9-7	9.9-7	-5.8-8	-6.5-8	-7.9-8	20	23	2:55
soybean-large-3	308 ; 307	10	936	1362	9261	8.3-7	9.1-7	9.8-7	-5.1-8	-5.7-8	-1.7-8	17	17	2:29
soybean-large-4	308 ; 307	10	1681	2132	13401	9.9-7	9.9-7	9.9-7	-1.0-7	-1.0-7	-4.3-8	29	28	3:49
soybean-large-5	308 ; 307	10	834	1229	3937	9.9-7	9.9-7	9.9-7	-3.2-8	-1.9-8	-2.3-8	14	18	1:08
soybean-large-6	308 ; 307	10	310	475	707	9.4-7	8.9-7	8.3-7	-8.1-8	-5.8-8	-1.5-7	05	06	12
soybean-large-7	308 ; 307	10	1028	1327	3970	9.9-7	9.9-7	9.9-7	-3.6-8	-6.3-8	-1.8-8	19	20	1:12
soybean-large-8	308 ; 307	10	782	1091	2901	9.8-7	9.9-7	8.9-7	-3.7-8	-4.5-8	-1.0-8	14	15	51
soybean-large-9	308 ; 307	10	928	1187	4901	9.8-7	9.8-7	9.9-7	1.1-7	-6.0-8	-1.7-8	17	19	1:26
soybean-large-10	308 ; 307	10	309	489	518	9.9-7	9.9-7	9.7-7	2.0-7	3.1-7	1.4-7	06	07	09
soybean-large-11	308 ; 307	10	877	1605	1755	9.9-7	8.6-7	9.5-7	-2.2-7	3.5-7	-2.6-7	17	23	32
spambase-small-2	301 ; 300	10	409	610	2792	8.8-7	9.5-7	9.0-7	-3.1-7	-3.9-7	-1.1-6	06	07	40
spambase-small-3	301 ; 300	10	476	665	1201	9.6-7	9.9-7	9.6-7	7.8-9	-3.7-8	-3.3-8	09	08	17
spambase-small-4	301 ; 300	10	1305	1983	6073	9.9-7	9.9-7	9.9-7	-4.5-9	6.6-9	-1.7-8	20	28	1:36
spambase-small-5	301 ; 300	10	608	819	868	8.5-7	9.8-7	9.9-7	-7.3-7	-2.7-7	-1.4-7	11	11	14
spambase-small-6	301 ; 300	10	811	1198	1334	9.9-7	9.9-7	9.9-7	-1.5-7	-2.0-7	-1.3-7	14	17	23
spambase-small-7	301 ; 300	10	849	1240	1359	9.9-7	9.9-7	9.9-7	4.0-7	2.8-7	1.8-7	15	18	25
spambase-small-8	301 ; 300	10	1109	1244	1501	9.9-7	9.9-7	8.8-7	7.1-8	9.3-8	7.6-8	20	18	27
spambase-small-9	301 ; 300	10	1090	1415	1440	9.9-7	9.7-7	9.9-7	-1.7-7	2.9-8	-1.3-8	20	21	27
spambase-small-10	301 ; 300	10	1081	1341	1500	9.9-7	9.9-7	9.9-7	1.7-7	1.5-7	-1.5-7	20	22	27
spambase-small-11	301 ; 300	10	1319	1482	1653	9.9-7	9.9-7	9.9-7	-3.6-7	-8.3-7	-5.8-7	25	25	31
spambase-medium-2	901 ; 900	10	471	596	1201	9.9-7	9.9-7	8.9-7	-1.6-6	-1.3-6	-1.9-6	1:42	1:37	4:01
spambase-medium-3	901 ; 900	10	1205	1582	11000	9.9-7	9.9-7	9.9-7	-2.0-7	-1.8-7	-2.2-7	4:18	4:16	36:54
spambase-medium-4	901 ; 900	10	2560	2990	4045	9.7-7	9.8-7	9.9-7	-2.3-6	2.5-6	1.1-6	9:06	8:04	13:37
spambase-medium-5	901 ; 900	10	1414	1900	2901	9.9-7	9.9-7	9.0-7	7.4-8	3.8-8	-1.1-6	5:06	5:17	9:58
spambase-medium-6	901 ; 900	10	1607	2107	2698	9.9-7	9.9-7	9.9-7	-1.0-8	3.7-8	-1.3-6	6:01	6:16	9:25
spambase-medium-7	901 ; 900	10	1805	2508	2846	9.9-7	9.9-7	9.9-7	-8.7-8	-4.5-8	-1.4-6	6:55	7:36	10:00
spambase-medium-8	901 ; 900	10	1655	2309	2489	9.9-7	9.9-7	9.9-7	-2.6-8	-6.7-8	4.6-7	6:19	6:54	8:47
spambase-medium-9	901 ; 900	10	1683	2330	2687	9.9-7	9.9-7	9.9-7	2.6-8	-5.9-8	2.2-8	6:23	6:56	9:38
spambase-medium-10	901 ; 900	10	1641	2030	2617	9.9-7	9.9-7	9.8-7	-6.5-7	-4.7-7	1.9-6	6:11	5:59	9:22
spambase-medium-11	901 ; 900	10	1608	1838	3210	9.9-7	9.9-7	9.9-7	-5.0-7	5.4-7	9.0-7	6:06	5:20	11:21
abalone-medium-2	401 ; 400	10	500	682	1301	9.9-7	9.9-7	8.5-7	-7.4-8	5.8-8	3.4-8	16	17	40
abalone-medium-3	401 ; 400	10	715	1011	1679	9.9-7	9.9-7	9.9-7	-2.5-9	1.3-8	-1.1-8	24	28	56

Table 4: The performance of SCB-SPADMM, ADMM, ADMMGB on QSDP- θ_+ , QSDP-QAP, QSDP-BIQ and QSDP-RCP problems (accuracy = 10^{-6}). 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”.

			iteration			η_{qsdp}			η_{gap}			time		
problem	$m_E; n_s$	rank(B)	scb	admm	gb	scb	admm	gb	scb	admm	gb	scb	admm	gb
abalone-medium-4	401 ; 400	10	372	626	684	9.9-7	9.9-7	9.9-7	-5.3-8	3.6-9	6.3-9	12	16	24
abalone-medium-5	401 ; 400	10	524	779	942	9.9-7	9.9-7	9.9-7	-3.8-8	-1.4-7	-9.6-8	18	21	32
abalone-medium-6	401 ; 400	10	536	946	1162	9.7-7	9.9-7	9.9-7	-1.3-7	-2.3-7	-1.8-7	22	27	38
abalone-medium-7	401 ; 400	10	1046	1676	2013	9.9-7	9.9-7	9.9-7	-8.9-8	-4.2-8	-3.3-8	37	47	1:09
abalone-medium-8	401 ; 400	10	745	1123	1641	9.6-7	9.7-7	9.9-7	-3.9-8	-2.2-7	-9.1-8	27	32	55
abalone-medium-9	401 ; 400	10	1035	1504	1709	9.9-7	9.5-7	9.9-7	-8.3-8	7.1-8	-1.2-8	38	43	1:02
abalone-medium-10	401 ; 400	10	1349	1803	1904	9.9-7	9.4-7	9.8-7	-1.7-7	-2.0-7	-2.2-7	49	51	1:07
abalone-medium-11	401 ; 400	10	1066	1504	1704	9.9-7	9.7-7	9.5-7	-1.1-7	-1.6-7	-1.6-7	40	45	1:02
abalone-large-2	1001 ; 1000	10	594	734	909	9.9-7	9.8-7	9.9-7	4.6-7	4.5-7	1.3-7	3:16	2:35	3:54
abalone-large-3	1001 ; 1000	10	656	1014	1901	9.9-7	9.9-7	9.9-7	-1.4-8	-7.2-8	-4.4-8	3:03	3:37	8:20
abalone-large-4	1001 ; 1000	10	505	749	995	9.9-7	9.9-7	9.8-7	-1.3-9	-1.6-8	-6.6-8	2:42	2:39	4:24
abalone-large-5	1001 ; 1000	10	752	1187	1550	9.8-7	9.9-7	9.9-7	-6.8-8	-1.8-7	-1.2-7	4:11	4:16	6:53
abalone-large-6	1001 ; 1000	10	886	1364	1670	9.9-7	9.9-7	9.9-7	-9.5-8	-1.1-7	-1.2-7	4:09	4:56	7:27
abalone-large-7	1001 ; 1000	10	1206	1614	2251	9.9-7	9.9-7	9.9-7	-1.1-7	1.8-8	-7.5-8	5:40	5:47	9:59
abalone-large-8	1001 ; 1000	10	1092	1721	2046	9.9-7	9.9-7	9.9-7	-3.1-7	-1.8-7	-2.9-7	5:08	6:14	9:07
abalone-large-9	1001 ; 1000	10	1557	2407	2746	9.8-7	9.9-7	9.9-7	-3.8-7	-3.5-7	-2.8-7	8:30	8:47	12:15
abalone-large-10	1001 ; 1000	10	1682	2488	2821	9.9-7	9.9-7	9.9-7	-1.6-7	-2.6-7	-2.5-7	8:00	9:06	12:39
abalone-large-11	1001 ; 1000	10	1923	3005	3723	9.8-7	9.9-7	9.9-7	1.3-7	3.6-8	-3.5-8	9:17	11:00	16:39
segment-medium-2	701 ; 700	10	1016	1541	1880	9.7-7	9.8-7	9.9-7	1.3-6	-1.1-6	2.5-7	2:07	2:13	3:26
segment-medium-3	701 ; 700	10	713	714	1801	9.4-7	9.5-7	9.2-7	-4.0-7	-9.7-7	-8.7-7	1:24	1:03	3:20
segment-medium-4	701 ; 700	10	2282	2710	17881	9.9-7	9.9-7	9.9-7	-7.1-8	-6.5-8	-6.5-8	4:30	4:25	34:11
segment-medium-5	701 ; 700	10	2322	3100	18701	9.9-7	9.9-7	9.9-7	-1.2-7	-9.5-8	-7.3-8	4:40	5:02	35:56
segment-medium-6	701 ; 700	10	2966	3916	25000	9.9-7	9.9-7	1.4-6	-1.7-7	-1.4-7	-1.3-7	6:12	6:29	51:26
segment-medium-7	701 ; 700	10	3185	4268	25000	9.9-7	9.9-7	1.6-6	-1.7-7	-1.7-7	-1.6-7	7:03	7:34	53:28
segment-medium-8	701 ; 700	10	2998	4140	25000	9.9-7	9.9-7	1.1-6	-1.6-7	-1.7-7	-6.7-8	6:28	7:09	52:54
segment-medium-9	701 ; 700	10	2123	2635	8801	9.9-7	9.9-7	9.9-7	-1.9-7	-3.0-8	-4.3-8	4:32	4:25	18:04
segment-medium-10	701 ; 700	10	1695	2414	6101	9.9-7	9.9-7	9.8-7	-2.4-7	-1.2-7	-2.2-8	3:35	4:07	12:27
segment-medium-11	701 ; 700	10	1454	2437	2101	9.4-7	9.7-7	8.6-7	6.4-8	-6.3-7	-1.5-7	3:01	4:00	4:13
segment-large-2	1001 ; 1000	10	1348	1823	2038	9.6-7	9.9-7	9.9-7	-1.3-6	-1.3-6	-1.4-6	6:30	6:15	8:40
segment-large-3	1001 ; 1000	10	479	533	1601	9.9-7	9.9-7	8.7-7	-4.0-7	-1.0-6	-4.4-7	2:10	1:53	6:49
segment-large-4	1001 ; 1000	10	2157	2802	20226	9.9-7	9.9-7	9.9-7	-9.1-8	-9.5-8	-7.1-8	9:57	9:57	1:27:58
segment-large-5	1001 ; 1000	10	2618	3404	25000	9.9-7	9.9-7	1.0-6	-1.1-7	-9.3-8	-8.3-8	12:13	12:12	1:50:29
segment-large-6	1001 ; 1000	10	3236	4143	25000	9.9-7	9.9-7	1.4-6	-1.8-7	-1.8-7	-1.2-7	15:28	15:20	1:52:58

Table 4: The performance of SCB-SPADMM, ADMM, ADMMGB on QSDP- θ_+ , QSDP-QAP, QSDP-BIQ and QSDP-RCP problems (accuracy = 10^{-6}). 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”.

			iteration			η_{qsdp}			η_{gap}			time		
problem	$m_E; n_s$	rank(B)	scb	admm	gb	scb	admm	gb	scb	admm	gb	scb	admm	gb
segment-large-7	1001 ; 1000	10	3505	4318	25000	9.9-7	9.9-7	1.8-6	-1.8-7	-1.7-7	-1.9-7	17:07	16:39	1:56:00
segment-large-8	1001 ; 1000	10	3063	3749	25000	9.9-7	9.9-7	1.2-6	-9.3-8	-7.8-8	-1.0-7	14:55	14:18	1:56:05
segment-large-9	1001 ; 1000	10	2497	3248	15649	9.9-7	9.9-7	9.9-7	-1.4-7	-1.2-7	-5.1-8	12:05	13:16	1:11:25
segment-large-10	1001 ; 1000	10	1723	2226	4901	9.9-7	9.9-7	9.9-7	7.4-9	1.4-8	-2.1-8	8:00	8:12	21:45
segment-large-11	1001 ; 1000	10	1571	2331	3417	9.9-7	9.7-7	9.9-7	1.9-7	-5.1-7	-1.7-8	7:20	8:30	15:23
housing-2	507 ; 506	10	3183	5358	4689	9.4-7	9.7-7	9.7-7	-1.9-7	1.8-7	2.0-7	2:54	3:22	3:48
housing-3	507 ; 506	10	845	1970	1714	9.9-7	9.9-7	9.9-7	-1.5-7	1.2-7	-2.2-8	48	1:16	1:24
housing-4	507 ; 506	10	805	1742	2057	9.4-7	9.9-7	9.9-7	-2.5-8	-4.8-8	-3.4-8	45	1:09	1:45
housing-5	507 ; 506	10	874	1262	1774	9.9-7	9.9-7	9.9-7	2.4-7	-2.3-7	-2.6-7	1:10	1:14	3:08
housing-6	507 ; 506	10	586	826	1005	9.9-7	9.9-7	9.9-7	-1.9-8	2.9-9	-8.6-8	1:41	1:26	1:39
housing-7	507 ; 506	10	583	906	1069	9.9-7	9.9-7	9.9-7	-1.3-7	-2.7-7	-1.7-7	32	37	56
housing-8	507 ; 506	10	682	904	1074	9.9-7	9.3-7	9.9-7	-1.1-7	-6.9-9	-6.6-8	39	38	59
housing-9	507 ; 506	10	765	1208	1590	8.5-7	9.9-7	9.8-7	-1.5-7	-1.3-8	8.5-8	44	53	1:26
housing-10	507 ; 506	10	1027	1381	1541	9.9-7	9.9-7	9.9-7	-6.4-8	-1.6-7	-1.0-7	58	1:02	1:27
housing-11	507 ; 506	10	867	1327	1359	9.9-7	9.9-7	9.9-7	-1.0-7	-9.0-8	-9.2-8	49	1:01	1:19