A Three-Operator Splitting Perspective of a Three-Block ADMM for Convex Quadratic Semidefinite Programming and Extensions

X. K. Chang · L. Chen · S. Y. Liu

2017/10/23, revised on 2018/07/05

Abstract In recent years, several convergent multi-block variants of the alternating direction method of multipliers (ADMM) have been proposed for solving the convex quadratic semidefinite programming via its dual, which is naturally a 3-block separable convex optimization problem with one coupled linear equality constraint. Among of these ADMM-type algorithms, the modified 3-block ADMM in [Chang et al., Neurocomput. 214: 575–586 (2016)] bears a peculiar feature that the augmented Lagrangian function is not necessarily to be minimized with respect to the block-variable corresponding to the quadratic term of the objective function. In this paper, we lay the theoretical foundation of this phenomena by interpreting this modified 3-block ADMM as a realization of a 3-operator splitting framework. Based on this perspective, we are able to extend this modified 3-block ADMM to a generalized 3-block ADMM, which not only applies to the more general convex composite quadratic programming setting but also admits the potential of achieving even a better numerical performance.

Keywords Convex composite quadratic programming · Convex quadratic semidefinite programming · Multi-Block · Alternating direction method of multipliers (ADMM) · Operator splitting · Generalized ADMM

Mathematics Subject Classification (2010) 90C25 · 90C22 · 65K05 · 47H05

Xiaokai Chang
School of Mathematics and Statistics, Xidian University, Xian 710071, P.R. China.
E-mail: xkchang@lut.cn

Liang Chen
College of Mathematics and Econometrics, Hunan University, Changsha 4100082, P.R. China
E-mail: chl@hnu.edu.cn

Sanyang Liu
School of Mathematics and Statistics, Xidian University, Xi’an 710071, P.R. China.
E-mail: liusanyang@126.com
1 Introduction

The convex quadratic semidefinite programming (CQSDP) has found many concrete applications in economics and engineering, and captures many well-studied problems, including the linearly constrained semidefinite least squares problem, the nearest Euclidean distance matrix (EDM) problem and the nearest correlation matrix problem.

To solve CQSDP problems, several algorithms have been proposed from different angles and here we only mention a few typical and relevant approaches. Based on certain perturbations of the Kuhn-Krash-Tucker (KKT) system of the CQSDP problem and its dual, Toh and his coauthors have proposed several inexact interior-point methods for solving them \cite{12, 25, 24}. By using the generalized Newton method together with the conjugate gradient method, many efficient methods were proposed for solving CQSDP problems \cite{18, 19}. For an important class of the CQSDP problem, i.e. the nearest correlation matrix problem, the quadratic convergence of the Newton-CG method has been obtained by Qi and Sun \cite{18}. For general CQSDP problems, the most recently developed solver QSDPNAL in Li, Sun and Toh \cite{16} has demonstrated that a two-phase augmented Lagrangian method, which properly combines both first-order and second-order algorithms, possesses a pretty promising numerical performance. A closer look of this solver shows that the inexact multi-block proximal ADMM studied in \cite{4}, in which the inexact block symmetric Gauss-Seidel iteration technique elaborated by Li, Sun and Toh \cite{15} was tightly incorporated, has been utilized to generate an approximate solution with a low to medium accuracy to warm-start an augmented Lagrangian method, whose subproblems are solved by a semismooth Newton method.

Just as in QSDPNAL, first-order ADMM-type algorithms are of their own importance for solving CQSDP problems. In fact, many extensions and modifications of the classic ADMM of Glowinski and Marrocco \cite{11} and Gabay and Mercier \cite{10} have been considered in recent years for solving the CQSDP problem via its dual, which is innately a 3-block separable convex optimization problem with one coupled linear constraint. Indeed, the most intuitive idea is to directly extend the classic ADMM to 3-block problems and the corresponding numerical performance is pretty good for many instances of problems \cite{2, 4, 14, 22}. However, the direct extension of ADMM (ADMMe) to problems with more than two blocks of variables is not guaranteed to be convergent (c.f. \cite{3} for a concrete example). Therefore, attentions have been paid to the design of multi-block ADMM-type algorithms and, fortunately, several algorithms of this type have been successfully applied to solving the CQSDP problem via its dual with a satisfactory numerical efficiency and a theoretical guarantee of convergence \cite{1, 4, 14}.

Among of these ADMM-type algorithms for CQSDP problems, the modified 3-block ADMM by Chang et al. \cite{1} has a distinct feature that one of the subproblems, i.e., the minimization of the augmented Lagrangian function with respect to a certain block of variables, can always be skipped. This saves both
the computational cost and the memory for variable storage, and, more importantly, the convergence is guaranteed under only one extra condition on the penalty parameter $\sigma$, while the proof for the convergence is much more involved. The peculiar feature of this method inspired us to get a further understanding of its theoretical foundation. Moreover, we are also concerned with the question that to what extent can this algorithm be improved or generalized, as well as whether this ADMM-type algorithm can be applied to more general problems.

In order to conduct the theoretical analysis to address the concerns mentioned above, an indispensable tool is the concept of the maximal monotone operator as well as the corresponding operator splitting methods [17, 23] which are designed to find its roots. The interconnection between the operator splitting methods and ADMM-type algorithms was first established by Gabay [9], who showed that the classic ADMM with unit step-length can be explained as the well-known Douglas-Rachford operator splitting method. Based on this observation, Eckstein and Bertsekas [7] presented a generalized ADMM for the purpose of improving the performance of the classic ADMM (with unit step-length) via an over-relaxation step. We mention that for a recent survey on this topic one may refer to [8], and one also can refer to [26] and references therein for more details and recent progresses on generalized ADMM. Consequently, we are interested if one can also interpret the modified 3-block ADMM in [1] via a certain operator splitting scheme and get further improvements on this algorithm via certain over-relaxation steps.

In this paper, we fulfil our objective by showing that the modified 3-block ADMM in [1] can be explained as an application of the 3-operator splitting framework studied in Davis and Yin [6]. We conduct our analysis in a much general setting in which the model that we will consider contains the CQSDP problem as a special case. Moreover, based on this operator splitting perspective, we present a generalized 3-block ADMM, in the sense of Eckstein and Bertsekas [7], in which an over-relaxation step is incorporated. We mention that, such as in Xiao et al. [26], this kind of over-relaxation can lead to an obvious improvement on the numerical efficiency of ADMM-type algorithms.

The remaining parts of this paper are organized as follows. In Section 2, we give a quick review of the CQSDP problem and the modified 3-block ADMM algorithm proposed by Chang et al. [1]. Section 3 is devoted to the operator-splitting perspective of this modified 3-block ADMM for the CQSDP problem. In Section 4, we introduce the convex composite quadratic optimization model and present a generalized version of the modified 3-block ADMM in [1] for solving this problem. With the result established in Section 3, the convergence analysis of the proposed algorithm can be conducted in a very concise manner. We conclude this paper in Section 5.
2 Preliminaries

2.1 Basic Concepts

Let \( H \) be an arbitrary finite dimensional real Hilbert space endowed with an inner product denoted by \( \langle \cdot, \cdot \rangle \) and its induced norm \( \| \cdot \| \). Let \( F : H \to H \) be an arbitrary set-valued mapping. If \( F \) is single-valued, it is called \( \beta \)-cocoercive (or \( \beta \)-inverse-strongly monotone) for a certain constant \( \beta > 0 \), if

\[
\langle F(x) - F(x'), x - x' \rangle \geq \beta \| F(x) - F(x') \|^2, \quad \forall x, x' \in H;
\]

If \( F \) is a self-adjoint positive semidefinite linear operator, we use \( \lambda_{\text{max}}(F) \) to denote its largest eigenvalue, i.e.,

\[
\lambda_{\text{max}}(F) : = \max_{\|x\| = 1} \langle x, Fx \rangle.
\]

In this case, it is easy to verify that \( F \) is \( \frac{1}{\lambda_{\text{max}}(F)} \)-coercive. If \( F \) is a multi-valued maximal monotone operator and \( \sigma > 0 \) is a constant, the mapping \( J_{\sigma F} := (I + \sigma F)^{-1} \), which is called the Minty resolvant of \( F \), is a single valued mapping, and this mapping is also nonexpansive [20, Theorem 12.12]. Here, \( I \) denotes the identity operator from \( H \) to itself and it will be kept as the notation for the identity operator from any space to itself, if no ambiguity is caused.

Let \( f : H \to (-\infty, \infty] \) be a closed proper convex function. The subdifferential mapping \( \partial f \) of \( f \) is then a maximal monotone operator and in this case

\[
J_{\sigma \partial f}(x) = (I + \sigma \partial f)^{-1}(x) = \arg\min_z \left\{ f(z) + \frac{1}{2\sigma} \| x - z \|^2 \right\}, \quad \forall x \in H.
\]

For any set \( C \subset H \), the indicator function \( \delta_C : H \to (-\infty, \infty] \) is defined by \( \delta_C(x) = 0 \) if \( x \in C \) and \( \delta_C(x) = +\infty \) otherwise. If \( C \) is a closed convex set, \( \delta_C \) is therefore a closed proper convex function and, in this case, \( J_{\sigma \partial \delta_C}(x) = \Pi_C(x) \), i.e., the metric projection of \( x \) onto \( C \), and

\[
\partial \delta_C(x) = \mathcal{N}_C(x) := \{ z | \langle z, x' - x \rangle \leq 0, \forall x' \in H \}.
\]

Here, the mapping \( \mathcal{N}_C \) is called the normal cone mapping of the set \( C \) and \( \mathcal{N}_C(x) \) is called the normal cone of \( C \) at \( x \), which is a closed convex cone.

Let \( S^n \) be the space of \( n \times n \) real symmetric matrices endowed with the standard trace inner product \( \langle \cdot, \cdot \rangle \) and the Frobenius norm \( \| \cdot \| \). We use \( S^n_+ \) and \( S^n_{++} \) to denote the sets of symmetric positive semidefinite and positive definite matrices in \( S^n \), respectively.

2.2 The CQSDP Problem

The CQSDP problem takes the following standard form:

\[
\min_X \frac{1}{2} \langle X, \varphi(X) \rangle + \langle C, X \rangle \quad \text{s.t.} \quad A(X) = b, \ X \in S^+_n,
\]
where $\varphi : S^n \rightarrow S^n$ is a given self-adjoint positive semidefinite linear operator, $A : S^n \rightarrow \mathbb{R}^m$ is a linear map defined by

$$A(X) := \begin{pmatrix} \langle A_1, X \rangle \\ \vdots \\ \langle A_m, X \rangle \end{pmatrix} \in \mathbb{R}^m, \quad \forall X \in S^n$$

with the given $A_i \in S^n$, $i = 1, \ldots, m$, and $b \in \mathbb{R}^m$. The adjoint of $A$, with respect to the standard inner products in $S^n$ and $\mathbb{R}^m$, is denoted by $A^*$. Note that $S^n_+$ is a closed convex self-dual cone. Then, the dual of problem (1) can be equivalently formulated in minimization form as follows

$$\min_{W, y, Z} \frac{1}{2} \langle W, \varphi(W) \rangle - b^T y + \delta_{S^n_+}(Z)$$

s.t. $-\varphi(W) + A^*(y) + Z = C,$

$$\forall (W, y, Z) \in S^n \times \mathbb{R}^m \times S^n. \tag{2}$$

where $\delta_{S^n_+}$ is the indicator function of $S^n_+, W \in S^n, Z \in S^n$ and $y \in \mathbb{R}^m$. The Lagrangian function of problem (2) is defined by

$$l(W, y, Z; X) := \frac{1}{2} \langle W, \varphi(W) \rangle - b^T y + \delta_{S^n_+}(Z) + \langle -\varphi(W) + A^*(y) + Z - C, X \rangle,$$

$$\forall (W, y, Z, X) \in S^n \times \mathbb{R}^m \times S^n \times S^n.$$

Therefore, the KKT system of problem (2) is given by

$$\begin{cases} 
\varphi(W) = \varphi(X), \quad A(X) = b, \\
-\varphi(W) + A^*(y) + Z = C, \\
X \in S^n_+, \quad Z \in S^n_+, \quad \langle Z, X \rangle = 0.
\end{cases} \tag{3}$$

For any $(W, y, Z, X) \in S^n \times \mathbb{R}^m \times S^n \times S^n$ satisfying the KKT system (3), $(W, y, Z)$ is a solution to problem (2) while $X$ is a solution to problem (1).

2.3 A Modified 3-Block ADMM for Problem (2)

Let $\sigma > 0$ be the penalty parameter. The augmented Lagrangian function of problem (2) can be defined by

$$L_\sigma(W, y, Z; X) := l(W, y, Z; X) + \frac{\sigma}{2} \| -\varphi(W) + A^*(y) + Z - C \|^2,$$

$$\forall (W, y, Z, X) \in S^n \times \mathbb{R}^m \times S^n \times S^n. \tag{4}$$

Choose an initial points \((y^0, Z^0, X^0) \in \mathbb{R}^m \times S^n \times \mathcal{S}^n\). A direct extension of the classic ADMM to the 3-block problem (2) consists of the following steps, for \(k = 0, 1, \ldots, \)

\[
\begin{aligned}
W^{k+1} &:= \arg\min_W L_\sigma(W, y^k, Z^k; X^k), \\
y^{k+1} &:= \arg\min_y L_\sigma(W^{k+1}, y, Z^k; X^k), \\
Z^{k+1} &:= \arg\min_Z L_\sigma(W^{k+1}, y^{k+1}, Z; X^k), \\
X^{k+1} &:= X^k + \tau \sigma \left( A^*(y^{k+1}) + Z^{k+1} - \varphi(W^{k+1}) - C \right),
\end{aligned}
\tag{5}
\]

where \(\tau > 0\) is the step-length. Generally, the convergence of the iteration scheme (5) can not be guaranteed.

In Chang et al. [1], the authors have proposed the following algorithm to solve problem (2), by modifying the iteration scheme (5).

**Algorithm 1 (A Modified ADMM for the CQSDP problem (2))**

Let \(\sigma > 0\) be the given parameter. Choose \(Z^0 \in S^n\) and \(X^0 \in S^n\). For \(k = 0, 1, \ldots, \)

1. Set \(W^{k+1} := X^k\).
2. Compute \(y^{k+1} := \arg\min_y L_\sigma(W^{k+1}, y, Z^k; X^k)\).
3. Compute \(Z^{k+1} := \arg\min_Z L_\sigma(W^{k+1}, y^{k+1}, Z; X^k)\).
4. Update \(X^{k+1} := X^k + \sigma \left( A^*(y^{k+1}) + Z^{k+1} - \varphi(W^{k+1}) - C \right)\).

Compared with the directly extended 3-block ADMM scheme (5), Algorithm 1 always set \(W^{k+1}\) as \(X^k\) instead of minimizing the augmented Lagrangian function with respect to the block-variable \(W\). The original idea of this 3-block ADMM is quite intuitive since that for any \((W, y, Z, X) \in S^n \times \mathbb{R}^m \times S^n \times S^n\) being a solution to the KKT system of the CQSDP problem, \((X, y, Z, X)\) is also a solution to it. Therefore, whenever dealing with a subproblem with respect to \(W\), one may directly use the value of \(X\) to substitute \(W\) instead of solving this subproblem. However, the convergence analysis of Algorithm 1 in [1] is very complicated.

The following Assumption was used in [1] for analyzing the convergence of Algorithm 1.

**Assumption 1** The linear operator \(A\) is surjective and the Slater’s constraint qualification holds for problem (1), i.e. there exists a matrix \(\bar{X} \in \mathcal{S}^n\) satisfying \(A(\bar{X}) = b\).

**Remark 1** The first part of Assumption 1 implies that the linear operator \(AA^*\) is nonsingular. Consequently, Step 2 of Algorithm 1 is well-defined. Moreover, under Assumption 1, we know from [21, Corollaries 28.2.2 & 28.3.1] that \(X \in \mathcal{S}^n\) is a solution to problem (1) if and only if there exists a vector \((W, y, Z) \in S^n \times \mathbb{R}^m \times S^n\) such that \((W, y, Z, X)\) is a solution to the KKT system (3). Additionally, for any
(W, y, Z, X) that satisfies (3), from [21, Corollary 30.5.1] we know that X is an optimal solution to problem (1) and (W, y, Z) is an optimal solution to problem (2).

The convergence properties of Algorithm 1 have been given in [1, Theorem 1]. We summarize these results as follows.

**Proposition 1** Suppose that the solutions set to problem (1) is nonempty, Assumption 1 holds and \( \sigma \in \left(0, \frac{1}{\lambda_{\max}(\varphi)}\right) \). Then, the sequence \( \{(W_k^k, y^k, Z_k^k, X_k^k)\} \) generated by Algorithm 1 converges to a point which is a solution to the KKT system (3).

**Remark 2** If \( \varphi = 0 \), problem (1) is then a standard linear semidefinite programming problem and its dual will be a 2-block convex optimization problem. In this case, the requirement that \( \sigma \in \left(0, \frac{1}{\lambda_{\max}(\varphi)}\right) \) is no longer necessary and Algorithm 1 is then automatically the classic 2-block ADMM with the unit step-length.

### 3 A Three-Operator Splitting Perspective

In this section, we narrate Algorithm 1 from a 3-operator splitting perspective. Note that the problem (1) can be written as

\[
\min_{X \in S^n} f(X) + g(X) + h(X),
\]

where

\[
\begin{align*}
    f(X) &:= \delta_K(X) \quad \text{with} \quad K := \{X \in S^n | A(X) = b\}, \\
    g(X) &:= \delta_{\mathcal{S}_n}(X), \\
    h(X) &:= \frac{1}{2} \langle X, \varphi(X) \rangle + \langle C, X \rangle.
\end{align*}
\]

Under Assumption 1, we know from [21, Theorem 23.8] that X is a solution to problem (1) if and only if

\[
0 \in \mathcal{N}_K(X) + \mathcal{N}_{\mathcal{S}_n}(X) + (\varphi(X) + C) = \partial f(X) + \partial g(X) + \nabla h(X).
\]

Therefore, one can try to solve problem (1) via solving the above inclusion problem. In fact, Algorithm 1 can be interpreted as an operator splitting algorithm applied to solve this inclusion problem. This will be explained in details as follows.

Let \( \{X^k\}, \{y^k\} \) and \( \{Z^k\} \) be the sequences generated by Algorithm 1. We define for \( k \geq 0 \),

\[
U^k := X^k - \sigma(\varphi(X^k) + C - A^*y^{k+1}).
\]

Moreover, just as [6, Equation (1.2)], we define the mapping \( \mathcal{T} : S^n \to S^n \) by

\[
\mathcal{T} := I - \mathcal{J}_{\sigma \mathcal{N}_{\mathcal{S}_n}} + \mathcal{J}_{\sigma \mathcal{N}_K} \circ (2\mathcal{J}_{\sigma \mathcal{N}_{\mathcal{S}_n}} - I - \sigma \nabla h(\mathcal{J}_{\sigma \mathcal{N}_{\mathcal{S}_n}})).
\]

Then, we have the following result.
**Theorem 1** Let \( \{X^k\}, \{y^k\} \) and \( \{Z^k\} \) be the sequences generated by Algorithm 1. Then the sequence \( \{U^k\} \) defined in (6) satisfies

\[
U^{k+1} = T(U^k).
\]

**Proof** Note that for any \( k \geq 0 \), steps 1 to 4 of Algorithm 1 can be reorganized as follows

\[
\begin{align*}
\begin{cases}
y^{k+1} := -(AA^*)^{-1}((AX^k - b)/\sigma + A(Z^k - \varphi(X^k) - C)), \\
X^{k+1} := \Pi_{\mathcal{S}_1}(U^k), \\
Z^{k+1} := (X^{k+1} - U^k)/\sigma.
\end{cases}
\end{align*}
\]  

(7)

Therefore, it holds that

\[
X^{k+1} = (I + \sigma \mathcal{N}_{\mathcal{S}_1}^{-1})(U^k) = \mathcal{J}_{\sigma \mathcal{S}_1}(U^k).
\]  

(8)

Moreover, one can readily obtain that

\[
y^{k+2} = -\frac{1}{\sigma}((AA^*)^{-1}(AX^{k+1} - b + \sigma A(Z^{k+1} - \varphi(X^{k+1}) - C)) - \frac{1}{\sigma}(X^{k+1} - U^k))
\]

\[
= -\frac{1}{\sigma}((AA^*)^{-1}(AX^{k+1} - b + \sigma A(X^{k+1} - U^k) - \varphi(X^{k+1}) - C))
\]

\[
= -\frac{1}{\sigma}((AA^*)^{-1}(AX^{k+1} - b + \sigma A(X^{k+1} - U^k) - \varphi(X^{k+1}) + C)) - \frac{1}{\sigma}(X^{k+1} - U^k)
\]

Note that

\[
\Pi_k(X) = X - A^*(AA^*)^{-1}(AX - b), \quad \forall X \in \mathbb{R}^n.
\]

Then, by using (8) we can get

\[
\begin{align*}
&\mathcal{J}_{\sigma \mathcal{N}_k} \circ (2\mathcal{J}_{\sigma \mathcal{N}_{\mathcal{S}_1}} - I - \sigma \nabla h(\mathcal{J}_{\sigma \mathcal{N}_{\mathcal{S}_1}}))(U^k) \\
&= \Pi_k(2X^{k+1} - U^k - \sigma[\varphi(X^{k+1}) + C])
\end{align*}
\]

\[
= (2X^{k+1} - U^k - \sigma[\varphi(X^{k+1}) + C]) + \sigma A^*y^{k+2}.
\]

Moreover, it is easy to see from (6) that

\[
U^{k+1} = X^{k+1} - \sigma(\varphi(X^{k+1}) + C - A^*y^{k+2})
\]

\[
= X^{k+1} + \sigma A^*y^{k+2} - \sigma[\varphi(X^{k+1}) + C]
\]

\[
= U^k + (X^{k+1} - U^k - \sigma[\varphi(X^{k+1}) + C]) + \sigma A^*y^{k+2}
\]

\[
= U^k - X^{k+1} + (2X^{k+1} - U^k - \sigma[\varphi(X^{k+1}) + C]) + \sigma A^*y^{k+2}
\]

\[
= U^k - \mathcal{J}_{\sigma \mathcal{N}_{\mathcal{S}_1}}(U^k) + (\mathcal{J}_{\sigma \mathcal{N}_k} \circ (2\mathcal{J}_{\sigma \mathcal{N}_{\mathcal{S}_1}} - I - \sigma \nabla h(\mathcal{J}_{\sigma \mathcal{N}_{\mathcal{S}_1}})))(U^k),
\]

which, together with the definition of \( T \), completes the proof. \[\square\]
Remark 3 The definition of the operator $T$ was introduced by Davis and Yin [6, Equation (1.2)]. This operator was regarded as a combination of the well-known Douglas-Rachford splitting and the forward-backward splitting. As will be seen in the next section, based on the properties of $T$, the global convergence of Algorithm 1, with the looser requirement $\sigma \in (0, \frac{2}{\lambda_{\max}(\phi)})$, can be alternatively proved by using [6, Theorem 2.1] together with Theorem 1.

4 Generalizations and Extensions

The successful application of Algorithm 1 to the CQSDP problem (1) via its dual (2) and the explanation from the operator splitting perspective made in Section 3 inspired us to consider extending this algorithm to much general problems. In this section, we consider the following convex composite quadratic programming [14] problem

$$\min_{x \in \mathcal{X}} \theta^*(x) + \frac{1}{2} \langle x, Q(x) \rangle + \langle c, x \rangle$$

s.t. $Ax = b,$

(9)

where $\mathcal{A} : \mathcal{X} \to \mathcal{Y}$ is a linear map, $\mathcal{X}$ and $\mathcal{Y}$ are finite dimensional Euclidean spaces each endowed with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$. $\theta^*$ is the Fenchel conjugate function of the closed proper convex (possibly nonsmooth) function $\theta : \mathcal{X} \to (-\infty, \infty]$, $Q : \mathcal{X} \to \mathcal{X}$ is a self-adjoint positive semidefinite linear operator, and $c \in \mathcal{X}$ and $b \in \mathcal{Y}$ are the given data. Obviously, problem (2) is an instance of problem (11) in which $\mathcal{X} = \mathbb{S}^n$, $\mathcal{Y} = \mathbb{R}^m$ and $\theta^*$ being the indicator function of $\mathbb{S}^n$. We make the following assumption on problem (9).

Assumption 2 The linear operator $\mathcal{A}$ is surjective and there exists a point $x \in \text{ri}(\text{dom} \theta^*)$ such that $Ax = b$.

Under Assumption 2 we know that $x$ is a solution to problem (9) if and only if there exists a vector $(w, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{X}$ such that $(x, w, y, z)$ solves the following KKT system of problem (11)

$$\begin{cases} Qw = Qx, \\
Ax - b = 0, \\
0 \in x - \partial \theta(-z), \\
A^*y + z - Qw - c = 0.
\end{cases}$$

(10)

Moreover, such a vector $(w, y, z)$ is a solution to the dual of problem (9), which can equivalently be recast in minimization form as

$$\min_{w, y, z} \left\{ \frac{1}{2} \langle w, Qw \rangle - \langle b, y \rangle + \theta(-z) \mid - Qw + A^*y + z = c \right\},$$

(11)
where the decision variables \( w \in \mathcal{X}, y \in \mathcal{Y} \) and \( z \in \mathcal{X} \).

Let \( \sigma > 0 \) be the penalty parameter. The augmented Lagrangian function of problem (11) is defined by

\[
L_\sigma(w, y, z; x) := \frac{1}{2}\langle w, Qw \rangle - \langle b, y \rangle + \theta(-z) + \langle A^*y + z - Qw - c, x \rangle + \frac{\sigma}{2}\|A^*y + z - Qw - c\|^2, \\
\forall (w, y, z; x) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{X}.
\]

In sequel, we will extend and generalize Algorithm 1 to problem (11), and prove its convergence via the existing convergence theorem of the 3-operator splitting method in [6].

**Algorithm 2 (A Generalized Modified ADMM for problem (11))**

Let \( \sigma > 0 \) and \( \rho \in (0, 2) \). Choose initial variables \( z^0 \) such that \(-z^0 \in \text{dom} \theta\) and \( x^0 \in \mathcal{X} \). For \( k = 0, 1, \ldots \),

**Step 1.** Set \( w^{k+1} := x^k \);

**Step 2.** Compute \( y^{k+1} := \arg\min_y L_\sigma(w^{k+1}, y, z^k; x) \);

**Step 3.** Compute

\[
\hat{z}^{k+1} := \arg\min_z \left\{ \theta(-z) + \langle z, x^k \rangle + \frac{\rho}{2}\|A^*y^{k+1} - (1 - \rho)z^k + z - \rho Qw^{k+1} - \rho c\|^2 \right\};
\]

**Step 4.** Update \( x^{k+1} := x^k + \sigma(A^*y^{k+1} - (1 - \rho)z^k + \hat{z}^{k+1} - \rho Qw^{k+1} - \rho c) \).

**Remark 4** The above algorithm is called a generalized modified ADMM since that it can viewed as a direct extension of the generalized 2-block ADMM [7] to problem (11). Moreover, we should mention that, generally, since that \( A \) is surjective, all the subproblems are well-defined and admit unique solutions. The well-definedness of subproblems is very essential for ADMM-type algorithm. On this part, one may refer to a counterexample by Chen et al. [5, Section 3].

**Remark 5** For the case that \( \rho = 1 \), step 3 of Algorithm 2 turns to

\[
\hat{z}^{k+1} = \arg\min_z L_\sigma(w^{k+1}, y^{k+1}, z; x^k).
\]

In this case, the direct extension of the generalized ADMM is then a direct extension of the classic ADMM with unit step-length, whose \( k \)-th step takes the following form

\[
\begin{align*}
\begin{cases}
 w^{k+1} \in \text{argmin}_w L_\sigma(w, y^k, z^k, x^k), \\
y^{k+1} \in \text{argmin}_y L_\sigma(w^{k+1}, y, z^k, x^k), \\
z^{k+1} \in \text{argmin}_z L_\sigma(w^{k+1}, y^{k+1}, z, x^k), \\
x^{k+1} = x^k + \sigma(A^*y^{k+1} + z^{k+1} - Qw^{k+1} - c).
\end{cases}
\end{align*}
\]

(12)
If the order of solving the subproblems is further changed as follows

\[
\begin{align*}
    y^{k+1} &\in \arg\min_y \mathcal{L}_\sigma(w^k, y, z^k, x^k), \\
    w^{k+1} &\in \arg\min_w \mathcal{L}_\sigma(w, y^{k+1}, z^k, x^k), \\
    z^{k+1} &\in \arg\min_z \mathcal{L}_\sigma(w^{k+1}, y^{k+1}, z, x^k),
\end{align*}
\]

the convergence of this direct extension of the class ADMM has been established in Li et al. [13, Theorem 2.1] under certain conditions\(^1\). In [6, Section 4.2], the authors have considered another 3-block extension of the classic ADMM, i.e., [6, Algorithm 7]. The difference of this extension from (12) is that the subproblem for computing \(w^{k+1}\) does not contain the penalty term \(\frac{\sigma}{2}\|A^*y^k + z^k - Qw - c\|\). Moreover, in the corresponding convergence analysis, it requires \(\sigma \in \left(0, \frac{2\lambda^+}{\lambda_{\text{min}}(Q)}\right)\), where \(\lambda^+\) denotes the smallest positive eigenvalue of \(Q\). This requirement of \(\sigma\) is obviously stronger than the condition that \(\sigma \in \left(0, \frac{2}{\lambda_{\text{min}}(Q)}\right)\), which will be used in the forthcoming convergence analysis of Algorithm 2.

Next, we analyze the convergence properties of Algorithm 2. Suppose that \(\{w^k\}, \{y^k\}, \{z^k\}\) and \(\{x^k\}\) be the infinite sequences generated by Algorithm 2. Define for \(k \geq 0\)

\[
u^k := x^k + \sigma(\rho A^* y^k - (1 - \rho)z^k - \rho Qw^k - \rho c) = x^{k+1} - \sigma z^{k+1}.
\] (13)

For convenience, we define the convex set

\[
\mathcal{K} := \{x \in \mathcal{X} | Ax = b\},
\]

and the quadratic function \(q : \mathcal{X} \to (-\infty, \infty)\) by

\[
q(x) := \frac{1}{2} \langle x, Qx \rangle + \langle c, x \rangle, \quad \forall x \in \mathcal{X}.
\]

Then, the gradient of the function \(q\) is given by \(\nabla q(x) = Qx + c, \forall x \in \mathcal{X}\). Moreover, we define a single-valued mapping \(\Gamma : \mathcal{X} \to \mathcal{X}\) by

\[
\Gamma := \mathcal{I} - J_{\sigma \partial q} + J_{\sigma X_c} \circ (2J_{\sigma \partial q} - \mathcal{I} - \sigma \nabla q \circ J_{\sigma \partial q}).
\] (14)

Based on the above definitions we have the following result.

**Proposition 2** Suppose that \(\{w^k\}, \{y^k\}\) and \(\{z^k\}\) are the infinite sequences generated by Algorithm 2, and \(\{u^k\}\) is the sequence defined by (13). Then, one has that

\[
u^{k+1} = (1 - \rho)u^k + \rho \Gamma(u^k).
\]

\(^1\) In [13], the authors also have considered adding proximal terms to subproblems and using a dual step-size which can be chosen in \((0, (1 + \sqrt{5})/2)\). Since that one can restrict \(w\) always in the range space of the linear operator \(Q\) so that \(f_1(w) = \frac{1}{2}(w, Qw)\) is a strongly convex function. Hence, the results in [13] are applicable.
Proof. Note that for any $k \geq 0$

\[
0 \in -\partial \theta(-z^{k+1}) + x^k + \sigma(\rho A^* y^{k+1} - (1 - \rho)z^{k+1} + z^{k+1} - \rho w^{k+1} - \rho c)
\]

\[= -\partial \theta(-z^{k+1}) + x^{k+1}, \tag{15}\]

Since that $\theta$ is a closed proper convex function, by using \[21, \text{Theorem 23.5}\] we have that $x^{k+1} \in \partial \theta(-z^{k+1})$ so that $-z^{k+1} \in \partial \theta^*(x^{k+1})$. Therefore, it holds that

\[
0 \in \partial \theta^*(x^{k+1}) + z^{k+1} = \partial \theta^*(x^{k+1}) + \frac{1}{\sigma}(x^{k+1} - (x^{k+1} - \sigma z^{k+1}))
\]

\[= \partial \theta^*(x^{k+1}) + \frac{1}{\sigma}(x^{k+1} - u^k), \tag{16}\]

where we have used the fact that $u^k = x^{k+1} - \sigma z^{k+1}$ from (13). Thus, by using \[21, \text{Theorem 23.8 \& 23.9}\] and the above inclusion one can get that

\[
x^{k+1} = \arg\min_x \left\{ \theta^*(x) + \frac{1}{2\sigma} \|x - u^k\|^2 \right\} = (I + \sigma\partial \theta^*)^{-1}(u^k) = J_{1/\sigma \partial \theta^*}(u^k).
\]

On the other hand, one can readily obtain that

\[
0 = -b + Ax^{k+1} + \sigma A(A^* y^{k+2} + z^{k+1} - Q w^{k+2} - c). \tag{17}\]

Therefore,

\[
y^{k+2} = -[\sigma AA^*]^{-1} \left( (Ax^{k+1} - b) + \sigma A(z^{k+1} - Q x^{k+1} - c) \right)
\]

\[= -[\sigma AA^*]^{-1} \left( A(2x^{k+1} - u^k - \sigma (Q x^{k+1} + c)) - b \right). \tag{18}\]

Note that for any $\xi \in \mathcal{X}$ one has $\Pi_{\mathcal{X}}(\xi) = \xi - A^*(AA^*)^{-1}(A\xi - b)$. Consequently, by using (16) and (18) we can get that

\[
J_{\sigma N_{\mathcal{X}} \left( 2J_{\sigma \partial \theta^*} - I - \sigma \nabla f \circ J_{\sigma \partial \theta^*} \right)}(u^k)
\]

\[= \Pi_{\mathcal{X}} \left( 2x^{k+1} - u^k - \sigma (Q x^{k+1} + c) \right)
\]

\[= 2x^{k+1} - u^k - \sigma (Q x^{k+1} + c) + \sigma A^* y^{k+2}.
\]

From (13) and the fact that $w^{k+1} = x^k$ one has that

\[
u^{k+1} = x^{k+1} + \sigma(\rho A^* y^{k+2} - (1 - \rho)z^{k+1} - \rho Q x^{k+1} - \rho c)
\]

\[= u^k - \rho x^{k+1} + (1 + \rho)x^{k+1} - u^k + \sigma \rho A^* y^{k+2}
\]

\[= \sigma(1 - \rho) z^{k+1} - \sigma \rho (Q x^{k+1} + c)
\]

\[= u^k - \rho x^{k+1} + \rho (2x^{k+1} - u^k - \sigma (Q x^{k+1} + c) + \sigma A^* y^{k+2})
\]

\[+(1 - \rho)(x^{k+1} - u^k - \sigma z^{k+1}).
\]
Note that (13) tells that $u^k = x^{k+1} - \sigma x^{k+1}$. Therefore, we can readily get
\[
\begin{align*}
u^{k+1} &= u^k - \rho z^{k+1} + \rho (2x^{k+1} - u^k - \sigma (Qx^{k+1} + c) + \sigma A^+ y^{k+2}) \\
&= (1 - \rho)u^k + \rho (u^k - x^{k+1} + 2x^{k+1} - u^k - \sigma (Qx^{k+1} + c) + \sigma A^+ y^{k+2}) \\
&= (1 - \rho)u^k + \rho \Gamma(u^k),
\end{align*}
\]
which completes the proof. \hfill \Box

According to Proposition 2, Algorithm 2 can also be viewed as a realization of the 3-operator splitting scheme proposed in [6] applied to the following problem
\[
\min_{x} \left\{ \theta^{*}(x) + \delta_{C}(x) + \frac{1}{2} \langle x, Qx \rangle + \langle c, x \rangle \right\}.
\]
(19)

Therefore, by using Proposition 2, a part of the convergence properties of Algorithm 2 can be deduced directly from [6, Theorem 1.1]. We summarize it as follows.

**Proposition 3** Suppose that the solution set to problem (19) is nonempty and Assumption 2 holds. Let the infinite sequences $\{w^k\}, \{y^k\}, \{z^k\}$ and $\{x^k\}$ be generated by Algorithm 2 with $\sigma \in (0, \frac{2}{\delta_{\text{aux}}(Q)})$ and $\rho \in \left(0, \frac{4 - \sigma_{\text{aux}}(Q)}{2}\right)$. Then, $\{x^k\}$ converges to a solution to problem (19). Moreover, the sequence $\{u^k\}$ defined by (13) converges to a unique point, say $u^\infty$, such that $0 \in \Gamma(u^\infty)$.

Since that Algorithm 2 is intentionally designed for problem (11), Proposition 3 is still not enough for this algorithm. Therefore, we need to further analyze its convergence properties. The following theorem fulfills this objective.

**Theorem 2** Suppose that the solution set to problem (19) is nonempty and Assumption 2 holds. Let $\sigma \in (0, \frac{2}{\delta_{\text{aux}}(Q)})$ and $\rho \in \left(0, \frac{4 - \sigma_{\text{aux}}(Q)}{2}\right)$. Then, the infinite sequences $\{w^k\}, \{y^k\}, \{z^k\}$ and $\{x^k\}$ can be generated by Algorithm 2, and the sequence $\{(w^k, y^k, z^k)\}$ converges to a solution to problem (11) while the sequence $\{x^k\}$ converges to a solution to problem (9).

**Proof** According to Proposition 3 we know that both sequences $\{x^k\}$ and $\{u^k\}$ are convergent, especially that $\{x^k\}$ converges to a solution of problem (9). Define $x^\infty := \lim_{k \to \infty} x^k$ and $u^\infty := \lim_{k \to \infty} u^k$. Then, by (18) we know that the sequence $\{y^k\}$ is convergent. Moreover, by (13) we know that $\{z^k\}$ is also convergent. We define $y^\infty := \lim_{k \to \infty} y^k$ and $z^\infty := \lim_{k \to \infty} z^k$. Note that
\[
\lim_{k \to \infty} (\rho A^* x^{k+1} - (1 - \rho)z^k + z^{k+1} - \rho Q u^{k+1} - \rho c) = 0,
\]
which implies that $A^* y^\infty + z^\infty - Q x^\infty - c = 0$. Then, by taking limits on both sides of (17), one has that $A^* x^\infty - b = 0$. Also, one can take limits in (15) and obtains that $0 \in -\partial \theta(-z^\infty) + x^\infty$. Therefore, by denoting $w^\infty = x^\infty$ we can conclude that $(w^\infty, y^\infty, z^\infty, x^\infty)$ is a solution to the KKT system (10), so that $\{(w^k, y^k, z^k)\}$ converges to a solution to problem (11). This completes the proof. \hfill \Box
5 Conclusions

In this paper, we have shown that the modified 3-block ADMM in Chang et al. [1] is an instance of the 3-operator splitting scheme in [6]. Based on this observation, we considered a generalized modified 3-block ADMM applied to the more general convex composite quadratic programming model, and derived its convergence via a very concise approach. The obtained results paved the way for further study of the proposed generalized modified ADMM for convex composite quadratic programming such as the iteration complexity and the local or global convergence rate, which we leave as our future work.

Acknowledgements The authors would like to thank Prof. Defeng Sun at The Hong Kong Polytechnic University for insightful comments and suggestions on this manuscript. The research of X. Chang was supported by the Fundamental Research Funds for Central Universities and the Innovation Fund of Xidian University. The research of L. Chen was supported by the Fundamental Research Funds for Central Universities.

References

1. Chang, X., Liu, S.Y. and Li, X.: Modified alternating direction method of multipliers for convex quadratic semidefinite programming. Neurocomput. 214: 575-586 (2016)
2. Chang X., Liu S.Y. and Zhao P.: A note on the sufficient initial condition ensuring the convergence of directly extended 3-block ADMM for special semidefinite programming. Optimization. (2018) https://doi.org/10.1080/02331934.2018.1490956
3. Chen, C., He, B., Ye, Y., Yuan, X. and Ye Y.: The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. Math. Program. 155(1): 57-79 (2016)
4. Chen, L., Sun, D.F. and Toh, K.-C.: An efficient inexact symmetric Gauss-Seidel based majorized ADMM for high-dimensional convex composite conic programming. Math. Program. 161(1): 237–270 (2017)
5. Chen, L., Sun, D.F. and Toh, K.-C.: A note on the convergence of ADMM for linearly constrained convex optimization problems. Comput. Optim. Appl. 66(2): 327–343 (2017)
6. Davis, D., Yin, W.: A three-operator splitting scheme and its optimization applications. Set-Valued Var. Anal. 25(4): 829–858 (2017)
7. Eckstein, J., Bertsekas, D.P.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. 55: 293–318 (1992)
8. Eckstein, J., Yao, W.: Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives. Pac. J. Optim. 11(4): 619–644 (2014)
9. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin, M. and Glowinski, R. (eds.): Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems. Studies in Mathematics and Its Applications, vol. 15, pp. 299–331. Elsevier (1983)
10. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. 2: 17–40 (1976)
11. Glowinski, R., Marroco, A.: Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation-dualité d’une classe de problèmes de Dirichlet non linéaires. Revue française d’automatique, Informatique Recherche Opérationelle. Analyse Numérique 9(2): 41–76 (1975)
12. Li, L., Toh, K.-C.: A polynomial-time inexact primal-dual infeasible path-following algorithm for convex quadratic SDP. Pac. J. Optim. 7: 43-61 (2011)
13. Li M., Sun, D.F. and Toh, K.-C.: A convergent 3-block semi-proximal ADMM for convex minimization problems with one strongly convex block. Asia-Pac. J. of Oper. Res. 32: 1550024(19 pages) (2015)
14. Li, X.D., Sun, D.F. and Toh, K.-C.: A Schur complement based semi-proximal ADMM for convex quadratic conic programming and extensions. Math. Program. 255(1-2): 333–373 (2016)
15. Li, X.D., Sun, D.F. and Toh, K.-C.: A block symmetric Gauss-Seidel decomposition theorem for convex composite quadratic programming and its applications. Math. Program. (2018) https://doi.org/10.1007/s10107-018-1247-7
16. Li, X.D., Sun, D.F. and Toh, K.-C.: QSDPNAL: A two-phase augmented Lagrangian method for convex quadratic semidefinite programming. Math. Prog. Comp. (2018) https://doi.org/10.1007/s12532-018-0137-6
17. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. 16(6): 964-979 (1979)
18. Qi, H.D., Sun, D.F.: A quadratically convergent Newton method for computing the nearest correlation matrix. SIAM J. Matrix Anal. Appl. 28: 360-385 (2006)
19. Qi, H.D., Sun, D.F.: An augmented Lagrangian dual approach for the $H$-weighted nearest correlation matrix problem. IMA J. Numer. Anal. 31: 491-511 (2011)
20. Rockafellar, R. T., Wets, R. J.-B.: Variational Analysis. Springer (1998)
21. Rockafellar, R. T.: Convex Analysis. Princeton (1976)
22. Sun, D.F., Toh, K.-C., Yang, L.Q.: A convergent 3-block semi-proximal alternating direction method of multipliers for conic programming with 4-type of constraints, SIAM J. Optim. 25: 882-915 (2015)
23. Tseng, P.: A modified forward-backward splitting method for maximal monotone mappings. SIAM J. Control Optim. 38(2): 431-446 (2000)
24. Toh, K.-C.: An inexact primal-dual path-following algorithm for convex quadratic SDP. Math. Program. 112: 221-254 (2008)
25. Toh, K.-C., Tutuncu, R. H. and Todd, M. J.: Inexact primal-dual path-following algorithms for a special class of convex quadratic SDP and related problems. Pac. J. Optim. 3: 135-164 (2007)
26. Xiao, Y.H., Chen, L. and Li, D.H.: A generalized alternating direction method of multipliers with semi-proximal terms for convex composite conic programming. Math. Prog. Comp. (2018) https://doi.org/10.1007/s12532-018-0134-9