CrossMark

# A note on the convergence of ADMM for linearly constrained convex optimization problems

**Liang Chen[1]** · **Defeng Sun[2]** · **Kim-Chuan Toh[3]**

**Abstract** This note serves two purposes. Firstly, we construct a counterexample to show that the statement on the convergence of the alternating direction method of multipliers (ADMM) for solving linearly constrained convex optimization problems in a highly influential paper by Boyd et al. (Found Trends Mach Learn 3(1):1–122, 2011) can be false if no prior condition on the existence of solutions to all the subproblems involved is assumed to hold. Secondly, we present fairly mild conditions to guarantee the existence of solutions to all the subproblems of the ADMM and provide a rigorous convergence analysis on the ADMM with a computationally more attractive large step-length that can even exceed the practically much preferred golden ratio of $(1 + \sqrt{5})/2$.

✉ Liang Chen
  chl@hnu.edu.cn

  Defeng Sun
  matsundf@nus.edu.sg

  Kim-Chuan Toh
  mattohkc@nus.edu.sg

[1]  College of Mathematics and Econometrics, Hunan University, Changsha 410082, China

[2]  Department of Mathematics and Risk Management Institute, National University of Singapore, 10 Lower Kent Ridge Road, Singapore, Singapore

[3]  Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore, Singapore

⌐ Springer

## 1 Introduction

Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be three finite-dimensional real Euclidean spaces each endowed with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$. Let $f : \mathcal{Y} \to (-\infty, +\infty]$ and $g : \mathcal{Z} \to (-\infty, +\infty]$ be two closed proper convex functions and $\mathcal{A} : \mathcal{X} \to \mathcal{Y}$ and $\mathcal{B} : \mathcal{X} \to \mathcal{Z}$ be two linear maps. The effective domains of $f$ and $g$ are denoted by dom $f$ and dom $g$, respectively. Consider the following 2-block separable convex optimization problem:

$$\min_{y \in \mathcal{Y}, z \in \mathcal{Z}} \left\{ f(y) + g(z) \quad \text{s.t.} \quad \mathcal{A}^* y + \mathcal{B}^* z = c \right\}, \tag{1}$$

where $c \in \mathcal{X}$ is the given data and the linear maps $\mathcal{A}^*$ and $\mathcal{B}^*$ are the adjoints of $\mathcal{A}$ and $\mathcal{B}$, respectively.

Let $\sigma > 0$ be a given penalty parameter. The augmented Lagrangian function of problem (1) is defined by, for any $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$,

$$\mathcal{L}_\sigma(y, z; x) := f(y) + g(z) + \langle x, \mathcal{A}^* y + \mathcal{B}^* z - c \rangle + \frac{\sigma}{2} \|\mathcal{A}^* y + \mathcal{B}^* z - c\|^2. \tag{2}$$

Choose an initial point $(x^0, y^0, z^0) \in \mathcal{X} \times \text{dom } f \times \text{dom } g$ and a step-length $\tau \in (0, +\infty)$. The classical alternating direction method of multipliers (ADMM) of Glowinski and Marroco [9] and Gabay and Mercier [6] then takes the following scheme for $k = 0, 1, \ldots,$

$$\begin{cases} y^{k+1} = \arg\min_y \mathcal{L}_\sigma(y, z^k; x^k), \\ z^{k+1} = \arg\min_z \mathcal{L}_\sigma(y^{k+1}, z; x^k), \\ x^{k+1} = x^k + \tau\sigma(\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^{k+1} - c). \end{cases} \tag{3}$$

The convergence analysis for the ADMM scheme (3) under certain settings was first conducted by Gabay and Mercier [6], Glowinski [7] and Fortin and Glowinski [5]. One may refer to [1] and [3] for recent surveys on this topic and to [8] for a note on the historical development of the ADMM.

In a highly influential paper[1] written by Boyd et al. [1], it was asserted [Section 3.2.1, Page 17] that if $f$ and $g$ are closed proper convex functions [1, Assumption 1] and the Lagrangian function of problem (1) has a saddle point [1, Assumption 2], then the ADMM scheme (3) converges for $\tau = 1$. This, however, turns to be false without imposing the prior condition that all the subproblems involved have solutions.

---

[1] It has been cited 2,229 times as captured by Google Scholar as of July 8, 2015.

To demonstrate our claim, in this note we shall provide a simple example (see Sect. 3) with the following four nice properties:

(P1) Both $f$ and $g$ are closed proper convex functions;
(P2) The Lagrangian function has infinitely many saddle points;
(P3) The Slater constraint qualification (CQ) holds; and
(P4) The linear operator $\mathcal{B}$ is nonsingular.

Note that our example to be constructed satisfies the two assumptions made in [1], i.e., (P1) and (P2), and the two additional favorable properties (P3) and (P4). Yet, the ADMM scheme (3) even with $\tau = 1$ may not be well-defined for solving problem (1). A closer examination of the proofs given in [1] reveals that the authors mistakenly took for granted the existence of solutions to all the subproblems in (3) under (P1) and (P2) only. Here we will fix this gap by presenting fairly mild conditions to guarantee the existence of solutions to all the subproblems in (3). Moreover we shall analyze the convergence of the ADMM with a computationally more attractive large step-length that can even be bigger than the golden ratio of $(1 + \sqrt{5})/2$.

The remaining parts of this note are organized as follows. In Sect. 2, we first present some necessary preliminary results from convex analysis for later discussions and then provide conditions under which the subproblems in the ADMM scheme (3) are solvable, or even admit bounded solution sets, so that this scheme is well-defined. In Sect. 3, based on several results established in Sect. 2, we construct a counterexample that satisfies (P1)–(P4) to show that the conclusion on the convergence of ADMM scheme (3) in [1, Section 3.2.1] can be false without making further assumptions. In Sect. 4, we establish some satisfactory convergence properties for the ADMM scheme (3) with a computationally more attractive large step-length that can even exceed the golden ratio of $(1 + \sqrt{5})/2$, under fairly weak assumptions. We conclude this note in Sect. 5.

## 2 Preliminaries

Let $\mathcal{U}$ be a finite dimensional real Euclidean space endowed with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\|\cdot\|$. Let $\mathcal{O} : \mathcal{U} \to \mathcal{U}$ be any self-adjoint positive semidefinite linear operator. For any $u, u' \in \mathcal{U}$, define $\langle u, u' \rangle_{\mathcal{O}} := \langle u, \mathcal{O}u' \rangle$ and $\|u\|_{\mathcal{O}} := \sqrt{\langle u, \mathcal{O}u \rangle}$ so that

$$\langle u, u' \rangle_{\mathcal{O}} = \tfrac{1}{2} \left( \|u\|_{\mathcal{O}}^2 + \|u'\|_{\mathcal{O}}^2 - \|u - u'\|_{\mathcal{O}}^2 \right) = \tfrac{1}{2} \left( \|u + u'\|_{\mathcal{O}}^2 - \|u\|_{\mathcal{O}}^2 - \|u'\|_{\mathcal{O}}^2 \right). \tag{4}$$

For any given set $U \subseteq \mathcal{U}$, we denote its relative interior by $\mathrm{ri}(U)$ and define its indicator function $\delta_U : \mathcal{U} \to (-\infty, +\infty]$ by

$$\delta_U(u) := \begin{cases} 0, & \text{if } u \in U, \\ +\infty, & \text{if } u \notin U. \end{cases}$$

Let $\theta : \mathcal{U} \to (-\infty, +\infty]$ be a closed proper convex function. We use $\mathrm{dom}\,\theta$ and $\mathrm{epi}(\theta)$ to denote its effective domain and its epigraph, respectively. Moreover, we use

$\partial\theta(\cdot)$ to denote the subdifferential mapping [11, Section 23] of $\theta(\cdot)$, which is defined by

$$\partial\theta(u) := \{v \in \mathcal{U} \mid \theta(u') \geq \theta(u) + \langle v, u' - u\rangle \, \forall \, u' \in \mathcal{U}\}, \quad \forall \, u \in \mathcal{U}. \qquad (5)$$

It holds that there exists a self-adjoint positive semidefinite linear operator $\Sigma_\theta : \mathcal{U} \to \mathcal{U}$ such that for any $u, u'$ with $v \in \partial\theta(u)$ and $v' \in \partial\theta(u')$,

$$\langle v - v', u - u'\rangle \geq \|u - u'\|_{\Sigma_\theta}^2. \qquad (6)$$

Since $\theta$ is closed, proper and convex, by [11, Theorem 8.5] we know that the recession function [11, Section 8] of $\theta$, denoted by $\theta 0^+$, is a positively homogeneous closed proper convex function that can be written as, for an arbitrary $u' \in \mathrm{dom}\,\theta$,

$$\theta 0^+(u) = \lim_{\rho \to +\infty} \frac{\theta(u' + \rho u) - \theta(u')}{\rho}, \quad \forall \, u \in \mathcal{U}.$$

The Fenchel conjugate $\theta^*(\cdot)$ of $\theta$ is a closed proper convex function defined by

$$\theta^*(v) := \sup_{u \in \mathcal{U}} \{\langle u, v\rangle - \theta(u)\}, \quad \forall \, v \in \mathcal{U}.$$

Since $\theta$ is closed, by [11, Theorem 23.5] we know that

$$v \in \partial\theta(u) \Leftrightarrow u \in \partial\theta^*(v). \qquad (7)$$

The dual of problem (1) takes the form of

$$\max_{x \in \mathcal{X}} \{h(x) := -f^*(-\mathcal{A}x) - g^*(-\mathcal{B}x) - \langle c, x\rangle\}. \qquad (8)$$

The Lagrangian function of problem (1) is defined by

$$\mathcal{L}(y, z; x) := f(y) + g(z) + \langle x, \mathcal{A}^*y + \mathcal{B}^*z - c\rangle, \quad \forall \, (y, z, x) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}, \qquad (9)$$

which is convex in $(y, z) \in \mathcal{Y} \times \mathcal{Z}$ and concave in $x \in \mathcal{X}$. Recall that we say that the Slater CQ for problem (1) holds if

$$\{(y, z) \mid y \in \mathrm{ri}(\mathrm{dom}\, f), \, z \in \mathrm{ri}(\mathrm{dom}\, g), \, \mathcal{A}^*y + \mathcal{B}^*z = c\} \neq \emptyset.$$

Under the above Slater CQ, from [11, Corollaries 28.2.2 and 28.3.1] we know that $(\bar{y}, \bar{z}) \in \mathrm{dom}\, f \times \mathrm{dom}\, g$ is a solution to problem (1) if and only if there exists a Lagrangian multiplier $\bar{x} \in \mathcal{X}$ such that $(\bar{x}, \bar{y}, \bar{z})$ is a saddle point to the Lagrangian function (9), or, equivalently, $(\bar{x}, \bar{y}, \bar{z})$ is a solution to the following Karush-Kuhn-Tucker (KKT) system

$$-\mathcal{A}x \in \partial f(y), \quad -\mathcal{B}x \in \partial g(z) \quad \text{and} \quad \mathcal{A}^*y + \mathcal{B}^*z = c. \qquad (10)$$

Furthermore, if the solution set to the KKT system (10) is nonempty, by [11, Theorem 30.4 and Corollary 30.5.1] we know that a vector $(\bar{x}, \bar{y}, \bar{z}) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ is a solution to (10) if and only if $(\bar{y}, \bar{z})$ is an optimal solution to problem (1) and $\bar{x}$ is an optimal solution to problem (8).

In the following, we shall conduct discussions on the existence of solutions to the subproblems in the ADMM scheme (3). Let the augmented Lagrangian function $\mathcal{L}_\sigma$ be defined by (2) and $(x', y', z') \in \mathcal{X} \times \text{dom } f \times \text{dom } g$ be an arbitrarily given point. Consider the following two auxiliary optimization problems:

$$\min_{y \in \mathcal{Y}} \left\{ F(y) := \mathcal{L}_\sigma(y, z'; x') \right\} \tag{11}$$

and

$$\min_{z \in \mathcal{Z}} \left\{ G(z) := \mathcal{L}_\sigma(y', z; x') \right\}. \tag{12}$$

Note that since $z' \in \text{dom } g$, problem (11) is equivalent to

$$\min_{y \in \mathcal{Y}} \left\{ \widehat{F}(y) := f(y) + \tfrac{\sigma}{2} \| \mathcal{A}^* y + (\mathcal{B}^* z' - c + x'/\sigma) \|^2 \right\}. \tag{13}$$

We now study under what conditions the problems (11) and (12) are solvable or have bounded solution sets. For this purpose, we consider the following assumptions:

**Assumption 1** $f0^+(y) > 0$ for any $y \in \mathcal{M}$, where

$$\mathcal{M} := \{y \in \mathcal{Y} \mid \mathcal{A}^* y = 0\} \backslash \{y \in \mathcal{Y} \mid f0^+(-y) = -f0^+(y) = 0\}.$$

**Assumption 2** $g0^+(z) > 0$ for any $z \in \mathcal{N}$, where

$$\mathcal{N} := \{z \in \mathcal{Z} \mid \mathcal{B}^* z = 0\} \backslash \{z \in \mathcal{Z} \mid g0^+(-z) = -g0^+(z) = 0\}.$$

**Assumption 3** $f0^+(y) > 0$ for any $0 \neq y \in \{y \in \mathcal{Y} \mid \mathcal{A}^* y = 0\}$.

**Assumption 4** $g0^+(z) > 0$ for any $0 \neq z \in \{z \in \mathcal{Z} \mid \mathcal{B}^* z = 0\}$.

Note that Assumptions 1–4 are not very restrictive. For example, if both $f$ and $g$ are coercive, in particular if they are norm functions, all the four assumptions hold automatically without any other conditions.

**Proposition 2.1** *It holds that*

(a) *Problem (11) is solvable if Assumption 1 holds, and problem (12) is solvable if Assumption 2 holds.*
(b) *The solution set to problem (11) is nonempty and bounded if and only if Assumption 3 holds, and the solution set to problem (12) is nonempty and bounded if and only if Assumption 4 holds.*

*Proof* (a) We first show that when Assumption 1 holds, the solution set to problem (13) is not empty. Consider the recession function $\widehat{F}0^+$ of $\widehat{F}$. On the one hand, by using [11, Theorem 9.3] and the second example given in [11, Pages 67–68], we know that for any $y \in \mathcal{Y}$ such that $\mathcal{A}^* y \neq 0$, one must have $\widehat{F}0^+(y) = +\infty$. On the other hand, for any $y \in \mathcal{Y}$ such that $\mathcal{A}^* y = 0$, by the definition of $\widehat{F}(y)$ in (13) we have

$$\widehat{F}0^+(y) = f0^+(y) + \sigma \langle \mathcal{A}(\mathcal{B}^* z' - c + x'/\sigma), y \rangle = f0^+(y).$$

Hence, by Assumption 1 we know that $\widehat{F}0^+(y) > 0$ for all $y \in \mathcal{Y}$ except for those satisfying $\widehat{F}0^+(-y) = -\widehat{F}0^+(y) = 0$. Then, by [11, part (b) in Corollary 13.3.4] and the closedness of $\widehat{F}$, it holds that $0 \in \mathrm{ri}(\mathrm{dom}\ \widehat{F}^*)$. Furthermore, by [11, Theorem 23.4] we know that $\partial \widehat{F}^*(0)$ is a nonempty set, i.e., there exists a $\hat{y} \in \mathcal{Y}$ such that $\hat{y} \in \partial \widehat{F}^*(0)$. By noting that $\widehat{F}$ is closed and using (7), we then have $0 \in \partial \widehat{F}(\hat{y})$, which implies that $\hat{y}$ is an optimal solution to problem (13) and hence to problem (11).

By repeating the above discussions we know that problem (12) is also solvable if Assumption 2 holds.

(b) By reorganizing the proofs for part (a), we can see that Assumption 3 holds if and only if $\widehat{F}0^+(y) > 0$ for all $0 \neq y \in \mathcal{Y}$. As a result, if Assumption 3 holds, from [11, Theorem 27.2] we know that problem (13) has a nonempty and bounded solution set. Conversely, if the solution set to problem (13) is nonempty and bounded, by [11, Corollary 8.7.1] we know that there does not exist any $0 \neq y \in \mathcal{Y}$ such that $\widehat{F}0^+(y) \leq 0$, so that Assumption 3 holds. Similarly, we can prove the remaining results of part (b). This completes the proof of the proposition. □

Based on Proposition 2.1 and its proof, we have the following result.

**Corollary 2.1** *If problem (1) has a nonempty and bounded solution set, then both problems (11) and (12) have nonempty and bounded solution sets.*

*Proof* Since problem (1) has a nonempty and bounded solution set, there does not exist any $0 \neq y \in \mathcal{Y}$ with $\mathcal{A}^* y = 0$ such that $f0^+(y) \leq 0$, or $0 \neq z \in \mathcal{Z}$ with $\mathcal{B}^* z = 0$ such that $g0^+(z) \leq 0$. Thus, Assumptions 3 and 4 hold. Then, by part (b) in Proposition 2.1 we know that the conclusion of Corollary 2.1 holds. □

A function $\varphi : \mathcal{U} \to (-\infty, +\infty]$ is called piecewise linear-quadratic [12, Definition 10.20] if its effective domain can be represented as the union of finitely many polyhedral sets, relative to each of which this function is given by an expression of $\frac{1}{2}\langle u, \mathcal{Q}u \rangle + \langle a, u \rangle + \alpha$ for some scalar $\alpha \in \mathfrak{R}$, vector $a \in \mathcal{U}$, and symmetric linear operator $\mathcal{Q} : \mathcal{U} \to \mathcal{U}$.

**Proposition 2.2** *If $f$ (or $g$) is a closed proper piecewise linear-quadratic convex function, especially a polyhedral convex function, we can replace the " $>$ " in Assumption 1 ( or 2) by " $\geq$ " and the corresponding sufficient condition in part (a) of Proposition 2.1 is also necessary.*

*Proof* Note that when $f$ is a closed piecewise linear-quadratic convex function, the function $\widehat{F}$ defined in problem (13) is a piecewise linear-quadratic convex function

with dom $\widehat{F} = \text{dom } f$ being a closed convex polyhedral set. Then by [12, Theorem 11.14 (b)] we know that $\widehat{F}^*$ is also a piecewise linear-quadratic convex function whose effective domain is a closed convex polyhedral set. By repeating the discussions for proving part (a) of Proposition 2.1 and using [11, part (a) in Corollary 13.3.4] we can obtain that Assumption 1 with ">" being replaced by "$\geq$" holds if and only if $0 \in \text{dom } \widehat{F}^*$, or $\partial \widehat{F}^*(0)$ is a nonempty set [12, Proposition 10.21], which is equivalent to saying that arg min $\widehat{F}$ is a nonempty set. If $g$ is piecewise linear-quadratic we can get a similar result. □

Finally, we need the following easy-to-verify result on the convergence of quasi-Fejér monotone sequences.

**Lemma 2.1** *Let $\{a_k\}_{k \geq 0}$ be a nonnegative sequence of real numbers satisfying $a_{k+1} \leq a_k + \varepsilon_k$ for all $k \geq 0$, where $\{\varepsilon_k\}_{k \geq 0}$ is a nonnegative and summable sequence of real numbers. Then the quasi-Fejér monotone sequence $\{a_k\}$ converges to a unique limit point.*

## 3 A Counterexample

In this section, we shall provide an example that satisfies all the properties (P1)-(P4) stated in Sect. 1 to show that the solution set to a certain subproblem in the ADMM scheme (3) can be empty if no further assumptions on $f$, $g$ or $\mathcal{A}$ are made. This means that the convergence analysis for the ADMM stated in [1] can be false. The construction of this example relies on Proposition 2.1. The parameter $\sigma$ and the initial point $(x^0, y^0, z^0)$ in the counterexample are just selected for the convenience of computations and one can construct similar examples for arbitrary penalty parameters and initial points.

We now present this example, which is a 3-dimensional 2-block convex optimization problem.

---

*Example 3.1* Let $\delta_{\geq 0}(\cdot)$ be the indicator function of the nonnegative real numbers. Consider problem (1) with $f(y_1, y_2) := \max(e^{-y_1} + y_2, y_2^2)$, $g(z) := \delta_{\geq 0}(z)$, $\mathcal{A}^* = (0, 1)$, $\mathcal{B}^* = -1$, and $c = 2$, i.e.,

$$\min_{(y_1, y_2, z) \in \mathfrak{R}^3} \left\{ \max(e^{-y_1} + y_2, y_2^2) + \delta_{\geq 0}(z) \mid 0y_1 + y_2 - z = 2 \right\}. \qquad (14)$$

---

In this example, $f$ and $g$ are closed proper convex functions with ri(dom $f$) = dom $f = \mathfrak{R}^2$ and ri(dom $g$) = $\{z \mid z > 0\} \subset$ dom $g$. The vector $(0, 3, 1) \in \mathfrak{R}^3$ lies in ri(dom $f$) $\times$ ri(dom $g$) and satisfies the constraint in problem (14). Hence, for problem (14), the Slater CQ holds. It is easy to check that the optimal solution set to problem (14) is given by

$$\{(y_1, y_2, z) \in \mathfrak{R}^3 \mid y_1 \geq -\log_e 2, \; y_2 = 2, \; z = 0\}$$

and the corresponding optimal objective value is 4. The Lagrangian function of problem (14) is given by

$$\mathcal{L}(y_1, y_2, z; x) = \max(e^{-y_1} + y_2, y_2^2) + \delta_{\geq 0}(z) + x(y_2 - z - 2),$$
$$\forall (y_1, y_2, z, x) \in \mathfrak{R}^4.$$

We now compute the dual of problem (14) based on this Lagrangian function.

**Lemma 3.1** *The objective function of the dual of problem* (14) *is given by*

$$h(x) = \begin{cases} -x^2/4 - 2x, & \text{if } x \in (-\infty, -2), \\ 1 - x, & \text{if } x \in [-2, -1), \\ -2x, & \text{if } x \in [-1, 0], \\ -\infty, & \text{if } x \in (0 + \infty). \end{cases}$$

*Proof* By the definition of the dual objective function, we have

$$
\begin{aligned}
h(x) &= \inf_{y_1, y_2, z} \mathcal{L}(y_1, y_2, z; x) \\
&= \inf_{z \geq 0, y_2} \left\{ \inf_{y_1} (\max(e^{-y_1} + y_2, y_2^2) + (y_2 - z - 2)x) \right\} \\
&= \inf_{z \geq 0, y_2} \{ \max(y_2, y_2^2) + y_2 x - zx - 2x \} \\
&= \min_{y_2} \left( \inf_{y_2 \in [0,1], z \geq 0} \{ y_2 + y_2 x - zx - 2x \}, \inf_{y_2 \notin [0,1], z \geq 0} \{ y_2^2 + y_2 x - zx - 2x \} \right).
\end{aligned}
$$

For any given $x \in \mathfrak{R}$, we have

$$
\begin{aligned}
&\inf_{y_2 \in [0,1], z \geq 0} \{ y_2 + y_2 x - zx - 2x \} \\
&= \inf_{y_2 \in [0,1]} \{ y_2(1 + x) \} + \inf_{z \geq 0} \{ -zx \} - 2x = \begin{cases} 1 - x, & \text{if } x < -1, \\ -2x, & \text{if } x \in [-1, 0], \\ -\infty, & \text{if } x > 0. \end{cases}
\end{aligned}
$$

Moreover, for any $x \in \mathfrak{R}$, it holds that

$$
\begin{aligned}
&\inf_{y_2 \notin [0,1], z \geq 0} \{ y_2^2 + y_2 x - zx - 2x \} \\
&= \inf_{y_2 \notin [0,1]} \{ y_2^2 + y_2 x + x^2/4 - x^2/4 - 2x \} + \inf_{z \geq 0} \{ -zx \} \\
&= \inf_{y_2 \notin [0,1]} \{ (y_2 + x/2)^2 \} + \inf_{z \geq 0} \{ -zx \} - x^2/4 - 2x \\
&= \begin{cases} -x^2/4 - 2x, & \text{if } x < -2, \\ 1 - x, & \text{if } x \in [-2, -1], \\ -2x, & \text{if } x \in [-1, 0], \\ -\infty, & \text{if } x > 0. \end{cases}
\end{aligned}
$$

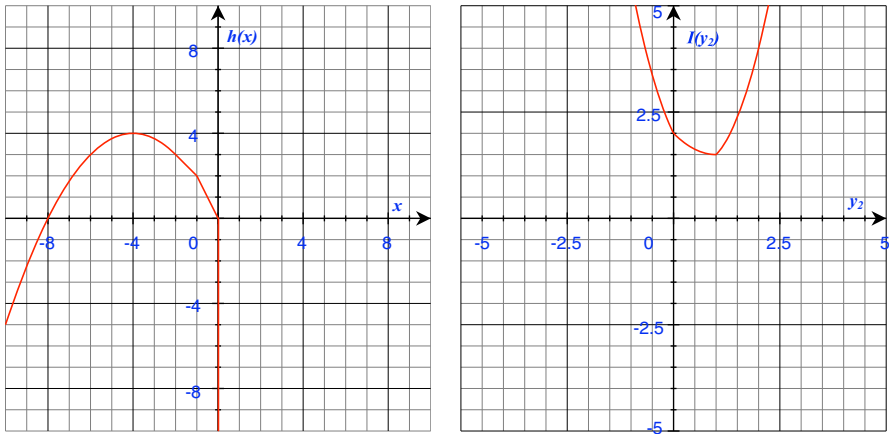Then by combining the above discussions on the two cases we obtain the conclusion of this lemma.  □

**Fig. 1** Graphs of the dual objective function $h(x)$ (*left*) and the function $I(y_2)$ (*right*)

By Lemma 3.1, one can see that the optimal solution to the dual of problem (14) is $\bar{x} = -4$ and the optimal value of the dual of problem (14) is $h(-4) = 4$ (see Fig. 1). Moreover, the set of solutions to the KKT system (10) for problem (14) is given by

$$\left\{ (y_1, y_2, z, x) \in \Re^4 \mid y_1 \geq -\log_e 2, \ y_2 = 2, \ z = 0, \ x = -4 \right\}.$$

Next, we consider solving problem (14) by using the ADMM scheme (3). For convenience, let $\sigma = 1$ and set the initial point $(x^0, y_1^0, y_2^0, z^0) = (0, 0, 0, 0)$. Now, one should compute $(y_1^1, y_2^1)$ by solving

$$\min_{y_1, y_2} \mathcal{L}_\sigma(y_1, y_2, z^0; x^0).$$

Define the function $I(\cdot) : \Re \to [-\infty, +\infty]$ by

$$
\begin{aligned}
I(y_2) : &= \inf_{y_1} \mathcal{L}_\sigma(y_1, y_2, z^0; x^0) \\
&= \inf_{y_1} \left\{ \max\left(e^{-y_1} + y_2, \ y_2^2\right) + (y_2 - 2)^2/2 \right\} \\
&= \begin{cases} \frac{3}{2}y_2^2 - 2y_2 + 2 & \text{if } y_2 \notin [0, 1], \\ \frac{1}{2}y_2^2 - y_2 + 2 & \text{if } y_2 \in [0, 1]. \end{cases}
\end{aligned}
$$

By direct calculations we can see that the above infimum is attained at $\bar{y}_2 = 1$ with $I(\bar{y}_2) = 1.5$ (see Fig. 1). However, we have for any $y_1 \in \Re$,

$$\mathcal{L}_\sigma(y_1, 1, 0; 0) = \max(e^{-y_1} + 1, 1) + 0.5 = e^{-y_1} + 1.5 > \inf_{y_1, y_2} \mathcal{L}_\sigma(y_1, y_2, z^0; x^0).$$

This means that although $\inf_{y_1, y_2} \mathcal{L}_\sigma(y_1, y_2, z^0; x^0) = 1.5$ is finite, it cannot be attained at any $(y_1, y_2) \in \Re^2$. Then the subproblem for computing $(y_1^1, y_2^1)$ is not

solvable and hence the ADMM scheme (3) is not well-defined. Note that for problem
(14), Assumption 1 fails to hold since the direction $y = (1, 0)$ satisfies $\mathcal{A}^* y = 0$ and
$f0^+(y) = 0$ but $f0^+(-y) = +\infty$.

*Remark 3.1* The counterexample constructed here is very simple. Yet, one may still
ask if the objective function $f$ about $(y_1, y_2)$ in problem (14) can be replaced by an
even simpler quadratic function. Actually, this is not possible as Assumption 1 holds if
$f$ is a quadratic function and the original problem has a solution. Specifically, suppose
that $\alpha \in \Re$ is a given number, $\mathcal{Q} : \mathcal{Y} \to \mathcal{Y}$ is a self-adjoint positive semidefinite linear
operator and $a \in \mathcal{Y}$ is a given vector while $f$ takes the following form

$$f(y) = \tfrac{1}{2}\langle y, \mathcal{Q}y \rangle + \langle a, y \rangle + \alpha, \quad \forall\, y \in \mathcal{Y}.$$

From [11, Pages 67–68] we know that

$$f0^+(y) = \begin{cases} \langle a, y \rangle, & \text{if } \mathcal{Q}y = 0, \\ +\infty, & \text{if } \mathcal{Q}y \neq 0. \end{cases} \tag{15}$$

If problem (1) has a solution, one must have $f0^+(y) \geq 0$ whenever $\mathcal{A}^* y = 0$. This,
together with (15), clearly implies that Assumption 1 holds.

## 4 Convergence properties of ADMM

The example presented in the previous section motivates us to reconsider the conver-
gence of the ADMM scheme (3). In the following, we will revisit the convergence
properties of the ADMM, with a computationally more attractive large step-length.

For convenience, we introduce some notations, which will be used throughout this
section. We use $\Sigma_f$ and $\Sigma_g$ to denote the two self-adjoint positive semidefinite linear
operators whose definitions, corresponding to the two functions $f$ and $g$ in problem
(1), can be drawn from (6). Let $(\bar{x}, \bar{y}, \bar{z}) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a given vector, whose
definition will be specified latter. We denote $x_e := x - \bar{x}$, $y_e := y - \bar{y}$ and $z_e := z - \bar{z}$
for any $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. If additionally the ADMM scheme (3) generates an
infinite sequence $\{(x^k, y^k, z^k)\}$, for $k \geq 0$ we denote $x_e^k := x^k - \bar{x}$, $y_e^k := y^k - \bar{y}$ and
$z_e^k := z^k - \bar{z}$, and define the following auxiliary notations

$$\begin{cases} u^k := -\mathcal{A}[x^k + (1-\tau)\sigma(\mathcal{A}^* y_e^k + \mathcal{B}^* z_e^k) + \sigma \mathcal{B}^*(z^{k-1} - z^k)], \\ v^k := -\mathcal{B}[x^k + (1-\tau)\sigma(\mathcal{A}^* y_e^k + \mathcal{B}^* z_e^k)], \\ \Psi_k := \frac{1}{\tau\sigma}\|x_e^k\|^2 + \|z_e^k\|_{\sigma \mathcal{B}\mathcal{B}^*}^2, \\ \Phi_k := \Psi_k + \max(1-\tau, 1-\tau^{-1})\sigma\|\mathcal{A}^* y_e^k + \mathcal{B}^* z_e^k\|^2 \end{cases} \tag{16}$$

with the convention $z^{-1} \equiv z^0$. Based on these notations, we have the following result.

**Proposition 4.1** *Suppose that* $(\bar{x}, \bar{y}, \bar{z}) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ *is a solution to the KKT system*
(10), *and that the ADMM scheme* (3) *generates an infinite sequence* $\{(x^k, y^k, z^k)\}$

*(which is guaranteed to be true if Assumptions 1 and 2 hold, cf. Proposition 2.1).
Then, for any $k \geq 1$,*

$$u^k \in \partial f(y^k), \quad v^k \in \partial g(z^k), \tag{17}$$

$$\Phi_k - \Phi_{k+1} \geq 2\|y_e^{k+1}\|_{\Sigma_f}^2 + \min(\tau, 1 + \tau - \tau^2)\sigma\|\mathcal{B}^*(z^{k+1} - z^k)\|^2$$
$$+ 2\|z_e^{k+1}\|_{\Sigma_g}^2 + \min(1, 1 - \tau + \tau^{-1})\sigma\|\mathcal{A}^*y_e^{k+1} + \mathcal{B}^*z_e^{k+1}\|^2 \tag{18}$$

*and*

$$\Psi_k - \Psi_{k+1} \geq 2\|y_e^{k+1}\|_{\Sigma_f}^2 + 2\|z_e^{k+1}\|_{\Sigma_g}^2 + \sigma\|\mathcal{A}^*y_e^{k+1} + \mathcal{B}^*z_e^k\|^2$$
$$+ (1 - \tau)\sigma\|\mathcal{A}^*y_e^{k+1} + \mathcal{B}^*z_e^{k+1}\|^2. \tag{19}$$

*Proof* For any $k \geq 1$, the inclusions in (17) directly follow from the first-order optimality condition of the subproblems in the ADMM scheme (3). The inequality (18) has been proved in Fazel et al.[2] [4, parts (a) and (b) in Theorem B.1]. Meanwhile, by using (B.12) in [4, Theorem B.1] and (4) we can get

$$\frac{1}{2\tau\sigma}(\|x_e^k\|^2 - \|x_e^{k+1}\|^2) - \frac{\sigma}{2}\|\mathcal{B}^*(z^{k+1} - z^k)\|^2 - \frac{\sigma}{2}\|\mathcal{B}^*z_e^{k+1}\|^2 + \frac{\sigma}{2}\|\mathcal{B}^*z_e^k\|^2$$
$$- \frac{2 - \tau}{2}\sigma\|\mathcal{A}^*y_e^{k+1} + \mathcal{B}^*z_e^{k+1}\|^2 + \sigma\langle\mathcal{B}^*(z^{k+1} - z^k), \mathcal{A}^*y_e^{k+1} + \mathcal{B}^*z_e^{k+1}\rangle$$
$$\geq \|y_e^{k+1}\|_{\Sigma_f}^2 + \|z_e^{k+1}\|_{\Sigma_g}^2,$$

which, together with the definition of $\Psi_k$ in (16), implies (19). This completes the proof. □

Now, we are ready to present several convergence properties of the ADMM scheme (3).

**Theorem 4.1** *Assume that the solution set to the KKT system (10) for problem (1) is nonempty. Suppose that the ADMM scheme (3) generates an infinite sequence $\{(x^k, y^k, z^k)\}$, which is guaranteed to be true if Assumptions 1 and 2 hold. Then, if $\tau \in (0, (1 + \sqrt{5})/2)$, one has the following results:*

(a) *the sequence $\{x^k\}$ converges to an optimal solution to the dual problem (8), and the primal objective function value sequence $\{f(y^k) + g(z^k)\}$ converges to the optimal value;*
(b) *the sequences $\{f(y^k)\}$ and $\{g(z^k)\}$ are bounded, and if Assumptions 3 and 4 hold, the sequences $\{y^k\}$ and $\{z^k\}$ are also bounded;*
(c) *any accumulation point of the sequence $\{(x^k, y^k, z^k)\}$ is a solution to the KKT system (10), and if $(x^\infty, y^\infty, z^\infty)$ is one of its accumulation points, then $\mathcal{A}^*y^k \to \mathcal{A}^*y^\infty$, $\Sigma_f y^k \to \Sigma_f y^\infty$, $\mathcal{B}^*z^k \to \mathcal{B}^*z^\infty$ and $\Sigma_g z^k \to \Sigma_g z^\infty$ as $k \to \infty$;*

---

[2] In [4] the authors studied a much general ADMM-type scheme where positive semidefinite proximal terms were added to the subproblems. The convergence properties studied in this paper can be extended to that setting with no difficulty, but in order to make this note as concise as possible we focus on ADMM only.

(d) *if $\Sigma_f + \mathcal{A}\mathcal{A}^* \succ 0$ and $\Sigma_g + \mathcal{B}\mathcal{B}^* \succ 0$, then each of the subproblems in the ADMM scheme* (3) *has a unique optimal solution and the whole sequence* $\{(x^k, y^k, z^k)\}$ *converges to a solution to the KKT system* (10).

*Proof* Let $(\bar{x}, \bar{y}, \bar{z}) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be an arbitrary solution to the KKT system (10) of problem (1). We first establish some basic results and then prove (a)–(d) one by one. In the following, the notations provided at the beginning of this section are used.

Note that $\|\mathcal{A}^* y_e^k\| \leq \|\mathcal{A}^* y_e^k + \mathcal{B}^* z_e^k\| + \|\mathcal{B}^* z_e^k\|$ for any $k \geq 0$. Since $\tau \in (0, (1 + \sqrt{5})/2)$, by using (16) and (18) we obtain that the sequences

$$\{\|x^k\|\}, \quad \{\|y^k\|_{\sigma\mathcal{A}\mathcal{A}^*}\} \quad \text{and} \quad \{\|z^k\|_{\sigma\mathcal{B}\mathcal{B}^*}\} \tag{20}$$

are all bounded, and

$$\sum_{k=0}^{\infty} \|y_e^k\|_{\Sigma_f}^2, \quad \sum_{k=0}^{\infty} \|z_e^k\|_{\Sigma_g}^2, \quad \sum_{k=0}^{\infty} \|\mathcal{A}^* y_e^k + \mathcal{B}^* z_e^k\|^2, \quad \sum_{k=0}^{\infty} \|\mathcal{B}^*(z^{k+1} - z^k)\|^2 < +\infty. \tag{21}$$

This, consequently, implies that $\{u^k\}$ and $\{v^k\}$ are bounded sequences. In the following, we prove (a)–(d) separately.

(a) Since $\{x^k\}$ is a bounded sequence, for any one of its accumulation points, e.g. $x^\infty \in \mathcal{X}$, it admits a subsequence, say, $\{x^{k_j}\}_{j \geq 0}$, such that $\lim_{j \to \infty} x^{k_j} = x^\infty$. By using the definitions of $\{u^k\}$ and $\{v^k\}$ in (16) we obtain that

$$u^\infty := \lim_{j \to \infty} u^{k_j} = -\mathcal{A}x^\infty \quad \text{and} \quad v^\infty := \lim_{j \to \infty} v^{k_j} = -\mathcal{B}x^\infty. \tag{22}$$

From (7) and (17) we know that for any $k \geq 1$, $y^k \in \partial f^*(u^k)$ and $z^k \in \partial g^*(v^k)$. Hence, we can get $\mathcal{A}^* y^k \in \mathcal{A}^* \partial f^*(u^k)$ and $\mathcal{B}^* z^k \in \mathcal{B}^* \partial g^*(v^k)$ so that

$$\mathcal{A}^* y^{k_j} + \mathcal{B}^* z^{k_j} \in \mathcal{A}^* \partial f^*(u^{k_j}) + \mathcal{B}^* \partial g^*(v^{k_j}), \quad \forall j \geq 0. \tag{23}$$

Then, by using (21), (22), (23) and the outer semi-continuity of subdifferential mappings of closed proper convex functions we know that

$$c \in \mathcal{A}^* \partial f^*(-\mathcal{A}x^\infty) + \mathcal{B}^* \partial g^*(-\mathcal{B}x^\infty),$$

which implies that $x^\infty$ is a solution to the dual problem (8). Therefore, we can conclude that any accumulation of $\{x^k\}$ is a solution to the dual problem (8). To finish the proof of part (a), we need to show that $\{x^k\}$ is a convergent sequence. This will be done in the following.

We define the sequence $\{\phi_k\}_{k \geq 1}$ by

$$\phi_k := \sigma \|z_e^k\|_{\mathcal{B}\mathcal{B}^*}^2 + \max(1 - \tau, 1 - \tau^{-1}) \sigma \|\mathcal{A}^* y_e^k + \mathcal{B}^* z_e^k\|^2.$$

Since $\tau \in (0, (1 + \sqrt{5})/2)$, from (18) in Proposition 4.1 and the fact that $\Phi_k \geq \phi_k$, we know that $\{\phi_k\}$ is a nonnegative and bounded sequence. Thus, there exists a

subsequence of $\{\phi_k\}$, say $\{\phi_{k_l}\}$, such that $\lim_{l\to\infty} \phi_{k_l} = \liminf_{k\to\infty} \phi_k$. Since $\{x^{k_l}\}$ is bounded, it must have a convergent subsequence, say, $\{x^{k_{l_i}}\}$, such that $\tilde{x} := \lim_{i\to\infty} x^{k_{l_i}}$ exists. Note that $(\tilde{x}, \bar{y}, \bar{z})$ is a solution to the KKT system (10). Therefore, without loss of generality, we can reset $\bar{x} = \tilde{x}$ from now on. By using (18) in Proposition 4.1 we know that the nonnegative sequence $\{\Phi_k\}$ is monotonically nonincreasing, and

$$\lim_{k\to\infty} \Phi_k = \lim_{i\to\infty} \Phi_{k_{l_i}} = \lim_{i\to\infty} \Big(\frac{1}{\tau\sigma}\|x_e^{k_{l_i}}\|^2 + \phi_{k_{l_i}}\Big) = \liminf_{k\to\infty} \phi_k. \qquad (24)$$

Since $\frac{1}{\tau\sigma}\|x_e^k\|^2 = \Phi_k - \phi_k$, we have

$$\limsup_{k\to\infty} \frac{1}{\tau\sigma}\|x_e^k\|^2 = \limsup_{k\to\infty}\{\Phi_k - \phi_k\} \leq \limsup_{k\to\infty} \Phi_k - \liminf_{k\to\infty} \phi_k = 0, \qquad (25)$$

which indicates that $\{x^k\}$ is a convergent sequence.

Now we study the convergence of the primal objective function values. On the one hand, since $(\bar{x}, \bar{y}, \bar{z})$ is a saddle point to the Lagrangian function $\mathcal{L}(\cdot)$ defined by (9), we have for any $k \geq 1, \mathcal{L}(\bar{y}, \bar{z}; \bar{x}) \leq \mathcal{L}(y^k, z^k; \bar{x})$. This, together with $\mathcal{A}^*\bar{y} + \mathcal{B}^*\bar{z} = c$, implies that for any $k \geq 1$,

$$f(\bar{y}) + g(\bar{z}) - \langle \bar{x}, \mathcal{A}^*y_e^k + \mathcal{B}^*z_e^k \rangle \leq f(y^k) + g(z^k). \qquad (26)$$

On the other hand, from (17) and (5) we know that

$$f(y^k) + \langle u^k, \bar{y} - y^k \rangle \leq f(\bar{y}) \quad \text{and} \quad g(z^k) + \langle v^k, \bar{z} - z^k \rangle \leq g(\bar{z}).$$

By combining the above two inequalities together and using (16) we can get

$$f(\bar{y}) + g(\bar{z}) - \langle x^k, \mathcal{A}^*y_e^k + \mathcal{B}^*z_e^k \rangle - \sigma \langle \mathcal{B}^*(z^{k-1} - z^k), \mathcal{A}^*y_e^k \rangle$$
$$-(1-\tau)\sigma\|\mathcal{A}^*y_e^k + \mathcal{B}^*z_e^k\|^2 \geq f(y^k) + g(z^k). \qquad (27)$$

Since the sequences in (20) are bounded, by using (21) and the fact that any nonnegative summable sequence should converge to zero we know that the left-hand-sides of both (26) and (27) converge to $f(\bar{y}) + g(\bar{z})$ when $k \to \infty$. Consequently, $\lim_{k\to\infty}\{f(y^k) + g(z^k)\} = f(\bar{y}) + g(\bar{z})$ by the squeeze theorem. Thus, part (a) is proved.
(b) From (17) we konw that for any $k \geq 1$,

$$f(y^k) \leq f(\bar{y}) - \langle u^k, \bar{y} - y^k \rangle = f(\bar{y}) - \langle u^k, \bar{y} \rangle + \langle u^k, y^k \rangle. \qquad (28)$$

On the one hand, from the boundedness of $\{u^k\}$ we know that the sequence $\{-\langle u^k, \bar{y} \rangle\}$ is bounded. On the other hand, from (21) and the boundedness of the sequences in (20) we can use

$$\langle u^k, y^k \rangle = -\langle x^k, \mathcal{A}^* y^k \rangle - (1-\tau)\sigma \langle \mathcal{A}^* y_e^k + \mathcal{B}^* z_e^k, \mathcal{A}^* y^k \rangle$$
$$-\sigma \langle \mathcal{B}^*(z^{k-1} - z^k), \mathcal{A}^* y^k \rangle$$

to get the boundedness of the sequence $\{\langle u^k, y^k \rangle\}$. Hence, from (28) we know the sequence $\{f(y^k)\}$ is bounded from above. From (10) we know

$$f(y^k) \geq f(\bar{y}) + \langle -\mathcal{A}\bar{x}, y^k - \bar{y} \rangle = f(\bar{y}) - \langle \bar{x}, \mathcal{A}^* y_e^k \rangle,$$

which, together with the fact that the sequences in (20) are bounded, implies that $\{f(y^k)\}$ is bounded from below. Consequently, $\{f(y^k)\}$ is a bounded sequence. By using similar approach, we can obtain that $\{g(z^k)\}$ is also a bounded sequence.

Next, we prove the remaining part of (b) by contradiction. Suppose that Assumption 3 holds and the sequence $\{y^k\}$ is unbounded. Note that the sequence $\{y^k/(1 + \|y^k\|)\}$ is always bounded. Thus it must have a subsequence $\{y^{k_j}/(1+\|y^{k_j}\|)\}_{j \geq 0}$, with $\{\|y^{k_j}\|\}$ being unbounded and non-decreasing, converging to a certain point $\xi \in \mathcal{Y}$. From the boundedness of the sequences in (20) we know that $\{\mathcal{A}^* y^k\}$ is bounded. Then we have

$$\mathcal{A}^* \xi = \mathcal{A}^* \left( \lim_{j \to \infty} \frac{y^{k_j}}{1 + \|y^{k_j}\|} \right) = \lim_{j \to \infty} \frac{\mathcal{A}^* y^{k_j}}{1 + \|y^{k_j}\|} = 0.$$

By noting that $\|\xi\| = 1$, one has $\xi \in \{y \in \mathcal{Y} \mid y \neq 0, \mathcal{A}^* y = 0\}$. On the other hand, define the sequence $\{d^{k_j}\}_{j \geq 0}$ by

$$d^{k_j} := \left( y^{k_j}/(1 + \|y^{k_j}\|), \ f(y^{k_j})/(1 + \|y^{k_j}\|) \right).$$

From the boundedness of the sequence $\{f(y^{k_j})\}$ and the definition of $\xi$ we know that $\lim_{j \to \infty} d^{k_j} = (\xi, 0)$. Since $(y^{k_j}, f(y^{k_j})) \in \text{epi}(f)$, by [11, Theorem 8.2] we know that $(\xi, 0)$ is a recession direction of $\text{epi}(f)$. Then from the fact that $\text{epi}(f0^+) = 0^+(\text{epi } f)$ we know that $f0^+(\xi) \leq 0$, which contradicts Assumption 3. The boundedness of $\{z^k\}$ under Assumption 4 can be similarly proved. Thus, part (b) is proved.

(c) Suppose that $(x^\infty, y^\infty, z^\infty)$ is an accumulation point of $\{(x^k, y^k, z^k)\}$. Let $\{(x^{k_j}, y^{k_j}, z^{k_j})\}_{j \geq 0}$ be a subsequence of $\{(x^k, y^k, z^k)\}$ which converges to $(x^\infty, y^\infty, z^\infty)$. By taking limits in (17) along $k_j$ for $j \to \infty$ and using (16) and (21) we can see that

$$-\mathcal{A}x^\infty \in \partial f(y^\infty), \quad -\mathcal{B}x^\infty \in \partial g(z^\infty) \quad \text{and} \quad \mathcal{A}^* y^\infty + \mathcal{B}^* z^\infty = c,$$

and this implies that $(x^\infty, y^\infty, z^\infty)$ is a solution to the KKT system (10). Now, without lose of generality we reset $(\bar{x}, \bar{y}, \bar{z}) = (x^\infty, y^\infty, z^\infty)$. Then, by part (a) we know that the sequence $\{\Phi_k\}$ defined in (16) converges to zero if $\tau \in (0, (1+\sqrt{5})/2)$. Thus, we always have

$$\lim_{k \to \infty} \|y_e^k\|_{\Sigma_f} = 0 \quad \text{and} \quad \lim_{k \to \infty} \|z_e^k\|_{\sigma \mathcal{B}\mathcal{B}^* + \Sigma_g} = 0. \tag{29}$$

As a result, it holds that $\mathcal{B}^* z^k \to \mathcal{B}^* z^\infty$, $\Sigma_f y^k \to \Sigma_f y^\infty$ and $\Sigma_g z^k \to \Sigma_g z^\infty$ as $k \to \infty$. Moreover, by using the fact that $\mathcal{A}^* y^k = (\mathcal{A}^* y^k + \mathcal{B}^* z^k) - \mathcal{B}^* z^k$ and $\mathcal{A}^* y^k + \mathcal{B}^* z^k \to \mathcal{A}^* y^\infty + \mathcal{B}^* z^\infty = c$ as $k \to \infty$, we can get $\mathcal{A}^* y^k \to \mathcal{A}^* y^\infty$ as $k \to \infty$. This completes the proof of part (c).

(d) If $\Sigma_f + \mathcal{A}\mathcal{A}^* \succ 0$ and $\Sigma_g + \mathcal{B}\mathcal{B}^* \succ 0$, then the subproblems in the ADMM scheme (3) are strongly convex, and hence each of them has a unique optimal solution. Then, by part (c) we know that $\{y^k\}$ and $\{z^k\}$ are convergent. Note that the sequence $\{x^k\}$ is convergent by part (a). Therefore, by part (c) we know that $\{(x^k, y^k, z^k)\}$ converges to a solution to the KKT system (10). Hence, part (d) is proved and this completes the proof of the theorem. □

Before concluding this note, we make the following remarks on the convergence results presented in Theorem 4.1.

*Remark 4.1* The corresponding results in part (a) of Theorem 4.1 for the ADMM scheme (3) with $\tau = 1$ have been stated in Boyd et al. [1]. However, as indicated by the counterexample constructed in Sect. 3, the proofs in [1] need to be revised with proper additional assumptions. Actually, no proof on the convergence of $\{x^k\}$ has been given in [1] at all. Nevertheless, one may view the results in part (a) as extensions of those in Boyd et al. [1] for the ADMM scheme (3) with $\tau = 1$ to a computationally more attractive ADMM scheme (3) with a rigorous proof.

*Remark 4.2* Note that, numerically, the boundedness of the sequences generated by a certain algorithm is a desirable property and Assumptions 3 and 4 can fulfill this purpose. Assumption 3 is rather mild in the sense that it holds automatically for many practical problems where $f$ has bounded level sets. Of course, the same comment can also be applied to Assumption 4.

*Remark 4.3* All the results of Theorem 4.1 are also valid if the step-length $\tau$ and the sequence generated by the ADMM scheme (3) satisfy the condition that

$$\tau \geq (1 + \sqrt{5})/2 \quad \text{but} \quad \sum_{k=1}^{\infty} \|x^{k+1} - x^k\|^2 < +\infty. \tag{30}$$

To prove this argument, one can first use $x^{k+1} - x^k = \tau\sigma(\mathcal{A}^* y_e^{k+1} + \mathcal{B}^* z_e^{k+1})$ to get $\sum_{k=0}^{\infty} \|\mathcal{A}^* y_e^k + \mathcal{B}^* z_e^k\|^2 < +\infty$. Then, by using (19) and the fact that $\|\mathcal{A}^* y_e^k\| \leq \|\mathcal{A}^* y_e^k + \mathcal{B}^* z_e^k\| + \|\mathcal{B}^* z_e^k\|$ we know the sequences in (20) are bounded. Note that $\|\mathcal{B}^*(z^{k+1} - z^k)\|^2 \leq 2\|\mathcal{A}^* y_e^{k+1} + \mathcal{B}^* z_e^{k+1}\|^2 + 2\|\mathcal{A}^* y_e^k + \mathcal{B}^* z_e^k\|^2$. This, together with (19), implies that (21) holds. The remaining procedure is similar to that for Theorem 4.1 except the following two key steps.

The first one is to show that $\{x^k\}$ is convergent. To do this, we define the nonnegative sequence $\{\psi_k\}$ by $\psi_k := \sigma\|z_e^k\|_{\mathcal{B}\mathcal{B}^*}^2$. By using (19), (21), Lemma 2.1 and fact that $1 - \tau < 0$ one can show that the sequence $\{\Psi_k\}$ is convergent. Hence, the sequence $\{\psi_k\}$ is nonnegative and bounded. Then, by similar discussions for getting (24) and (25) with $\phi_k$ and $\Phi_k$ being replaced by $\psi_k$ and $\Psi_k$, one can get $\lim_{k\to\infty} \Psi_k = \liminf_{k\to\infty} \psi_k$.

Hence, $\{x^k\}$ is convergent.

The second one is to show that (29) still holds for this case. This can be done by verifying that the sequence $\{\Psi_k\}$ defined in (16) converges to zero if $\tau \geq (1+\sqrt{5})/2$ but $\sum_{k=0}^{\infty} \|x^{k+1} - x^k\|^2 < +\infty$.

The condition (30) simplifies the condition proposed by Sun et al. [13, Theorem 2.2], in which one need $\sum_{k=1}^{\infty}\{\|\mathcal{B}^*(z^{k+1} - z^k)\|^2 + \sigma\|x^{k+1} - x^k\|^2\} < +\infty$ if $\tau \geq (1+\sqrt{5})/2$. This was used for the purpose of achieving better numerical performance. The advantage of taking the step-length $\tau \geq (1+\sqrt{5})/2$ has been observed in [2,10,13] for solving high-dimensional linear and convex quadratic semidefinite programming problems. In numerical computations, one can start with a larger $\tau$, e.g. $\tau = 1.95$, and reset it as $\tau := \max(\gamma\tau, 1.618)$ for some $\gamma \in (0, 1)$, e.g. $\gamma = 0.95$, if at the $k$-th iteration one observes that $\|x^{k+1} - x^k\|^2 > c_0/k^{1.2}$, where $c_0$ is a given positive constant. Since $\tau$ can be reset at most a finite number of times, our convergence analysis is valid for such a strategy. One may refer to [13, Remark 2.3] for more discussions on this computational issue.

## 5 Conclusions

In this note, we have constructed a simple example possessing several nice properties to illustrate that the convergence theorem of the ADMM scheme (3) with the unit step-length stated in Boyd et al. [1] can be false if no prior condition that guarantees the existence of solutions to all the subproblems involved is made. In order to correct this mistake we have presented fairly mild conditions under which all the subproblems are solvable by using standard knowledge in convex analysis. Based on these conditions, we have further established some satisfactory convergence properties of the ADMM with a computationally more attractive large step-length that can even exceed the golden ratio of $(1+\sqrt{5})/2$. In conclusion, this note has (i) clarified some confusions on the convergence results of the popular ADMM; (ii) opened the potential for designing computationally more efficient ADMM based solvers in the future.

## References

1. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**(1), 1–122 (2011)
2. Chen, L., Sun, D.F., Toh, K.-C.: An effcient inexact symmetric Gauss-Seidel based majorized ADMM for high-dimensional convex composite conic programming. Math. Program. (2016). doi:10.1007/s10107-016-1007-5
3. 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 (2015)
4. Fazel, M., Pong, T.K., Sun, D.F., Tseng, P.: Hankel matrix rank minimization with applications to system identification and realization. SIAM J. Matrix Anal. Appl. **34**(3), 946–977 (2013)
5. Fortin, M., Glowinski, R.: Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary Value Problems. Studies in Mathematics and its Applications, vol. 15. Translated from French by Hunt, B. and Spicer, D.C. Elsevier Science Publishers B.V. (1983)

6. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. **2**(1), 17–40 (1976)
7. Glowinski, R.: Lectures on numerical methods for non-linear variational problems. Published for the Tata Institute of Fundamental Research, Bombay [by] Springer-Verlag (1980)
8. Glowinski, R.: On alternating direction methods of multipliers: a historical perspective. In: Fitzgibbon, W., Kuznetsov, Y.A., Neittaanmaki, P., Pironneau, O. (eds.) Modeling, Simulation and Optimization for Science and Technology, pp. 59–82. Springer, Netherlands (2014)
9. 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'atomatique, Informatique Recherche Opérationelle. Anal. Num. **9**(2), 41–76 (1975)
10. Li, X.D., Sun, D.F., Toh, K.-C.: A Schur complement based semi-proximal ADMM for convex quadratic conic programming and extensions. Math. Program. **155**, 333–373 (2016)
11. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
12. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer, Verlag Berlin Heidelberg (2009)
13. 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 constraints. SIAM J. Optim. **25**, 882–915 (2015)