Global Convergence of Unmodified 3-Block ADMM for a Class of Convex Minimization Problems

Tianyi Lin *   Shiqian Ma †   Shuzhong Zhang ‡

May 16, 2015

Abstract

The alternating direction method of multipliers (ADMM) has been successfully applied to solve structured convex optimization problems due to its superior practical performance. The convergence properties of the 2-block ADMM have been studied extensively in the literature. Specifically, it has been proven that the 2-block ADMM globally converges for any penalty parameter $\gamma > 0$. In this sense, the 2-block ADMM allows the parameter to be free, i.e., there is no need to restrict the value for the parameter when implementing this algorithm in order to ensure convergence. However, for the 3-block ADMM, Chen et al. [4] recently constructed a counter-example showing that it can diverge if no further condition is imposed. The existing results on studying further sufficient conditions on guaranteeing the convergence of the 3-block ADMM usually require γ to be smaller than a certain bound, which is usually either difficult to compute or too small to make it a practical algorithm. In this paper, we show that the 3-block ADMM still globally converges with any penalty parameter $\gamma > 0$ when applied to solve a class of commonly encountered problems to be called regularized least squares decomposition (RLSD) in this paper, which covers many important applications in practice.

Keywords: ADMM, Global Convergence, Convex Minimization, Regularized Least Squares Decomposition.

*Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong. Email: tylin@se.cuhk.edu.hk.

†Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong. Email: sqma@se.cuhk.edu.hk. Research of this author was supported in part by the Hong Kong Research Grants Council General Research Fund (Grant 14205314).

‡Department of Industrial and Systems Engineering, University of Minnesota, Minneapolis, MN 55455. Email: zhangs@umn.edu. Research of this author was supported in part by the NSF Grant CMMI-1161242.

1 Introduction

The alternating direction method of multipliers (ADMM) has been very successfully applied to solve many structured convex optimization problems arising from machine learning, image processing, statistics, computer vision and so on; see the recent survey paper [2]. The ADMM is particularly efficient when the problem has a separable structure in functions and variables. For example, the following convex minimization problem with 2-block variables can usually be solved by ADMM, provided that a certain structure of the problem is in place:

$$\begin{array}{ll} \min & f_1(x_1) + f_2(x_2) \\ \text{s.t.} & A_1x_1 + A_2x_2 = b \\ & x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, \end{array} \tag{1}$$

where $f_i(x_i) : \mathbb{R}^{n_i} \to \mathbb{R}^p, i = 1, 2$, are proper closed convex functions, $A_i \in \mathbb{R}^{p\times n_i}, i = 1, 2$, $b \in \mathbb{R}^p$ and $\mathcal{X}_i, i = 1, 2$, are closed convex sets. A typical iteration of the 2-block ADMM (with given (x_2^k, λ^k)) for solving (1) can be described as

$$\left\{ \begin{array}{ll} x_1^{k+1} & := \text{argmin}_{x_1\in\mathcal{X}_1} \ \bar{\mathcal{L}}_\gamma(x_1, x_2^k; \lambda^k) \\ x_2^{k+1} & := \text{argmin}_{x_2\in\mathcal{X}_2} \ \bar{\mathcal{L}}_\gamma(x_1^{k+1}, x_2; \lambda^k) \\ \lambda^{k+1} & := \lambda^k - \gamma(A_1x_1^{k+1} + A_2x_2^{k+1} - b), \end{array} \right. \tag{2}$$

where the augmented Lagrangian function $\bar{\mathcal{L}}_\gamma$ is defined as

$$\bar{\mathcal{L}}_\gamma(x_1, x_2; \lambda) := f_1(x_1) + f_2(x_2) - \langle \lambda, A_1x_1 + A_2x_2 - b\rangle + \frac{\gamma}{2}\|A_1x_1 + A_2x_2 - b\|_2^2,$$

where λ is the Lagrange multiplier and $\gamma > 0$ is a penalty parameter, which can also be viewed as a step size on the dual update. The convergence properties of 2-block ADMM (2) have been studied extensively in the literature; see for example [27, 11, 10, 12, 9, 18, 29, 8, 1]. A very nice property of the 2-block ADMM is that it is *parameter restriction-free*: it has been proven that the 2-block ADMM (2) is globally convergent for any parameter $\gamma > 0$, starting from anywhere. This property makes the 2-block ADMM particularly attractive for solving structured convex optimization problems in the form of (1).

However, this is not the case when ADMM is applied to solve convex problems with 3-block variables:

$$\begin{array}{ll} \min & f_1(x_1) + f_2(x_2) + f_3(x_3) \\ \text{s.t.} & A_1x_1 + A_2x_2 + A_3x_3 = b \\ & x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, x_3 \in \mathcal{X}_3. \end{array} \tag{3}$$

Note that the 3-block ADMM for solving (3) can be described as

$$\begin{cases} x_1^{k+1} & := \text{argmin}_{x_1 \in \mathcal{X}_1} \; \mathcal{L}_\gamma(x_1, x_2^k, x_3^k; \lambda^k) \\ x_2^{k+1} & := \text{argmin}_{x_2 \in \mathcal{X}_2} \; \mathcal{L}_\gamma(x_1^{k+1}, x_2, x_3^k; \lambda^k) \\ x_3^{k+1} & := \text{argmin}_{x_3 \in \mathcal{X}_3} \; \mathcal{L}_\gamma(x_1^{k+1}, x_2^{k+1}, x_3; \lambda^k) \\ \lambda^{k+1} & := \lambda^k - \gamma(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b), \end{cases} \tag{4}$$

where the augmented Lagrangian function is defined as

$$\mathcal{L}_\gamma(x_1, x_2, x_3; \lambda) := f_1(x_1) + f_2(x_2) + f_3(x_3) - \langle \lambda, A_1 x_1 + A_2 x_2 + A_3 x_3 - b \rangle + \frac{\gamma}{2} \| A_1 x_1 + A_2 x_2 + A_3 x_3 - b \|_2^2.$$

Regarding its general convergence however, Chen et al. constructed a counterexample in [4] showing that the 3-block ADMM (4) can diverge if no further condition is imposed. On the other hand, the 3-block ADMM (4) has been successfully used in many important applications such as the robust and stable principal component pursuit problem [33, 40], the robust image alignment problem [30], Semidefinite Programming [37], and so on. It is therefore of great interest to further study sufficient conditions to guarantee the convergence of 3-block ADMM (4). Han and Yuan [13] showed that the 3-block ADMM (4) converges if all the functions f_1, f_2, f_3 are strongly convex and γ is restricted to be smaller than a certain bound. This condition is relaxed in Chen, Shen and You [5] and Lin, Ma and Zhang [24] to allow only f_2 and f_3 to be strongly convex and γ is restricted to be smaller than a certain bound. Moreover, the first sublinear convergence rate result of multi-block ADMM is established in [24]. Closely related to [5, 24], Cai, Han and Yuan [3] and Li, Sun and Toh [22] proved the convergence of the 3-block ADMM (4) under the assumption that only one of the functions f_1, f_2 and f_3 is strongly convex, and γ is restricted to be smaller than a certain bound. Davis and Yin [6] studied a variant of the 3-block ADMM (see Algorithm 8 in [6]) which requires that f_1 is strongly convex and γ is smaller than a certain bound to guarantee the convergence. In addition to strong convexity of f_2 and f_3, and the boundedness of γ, by assuming further conditions on the smoothness of the functions and some rank conditions on the matrices in the linear constraints, Lin, Ma and Zhang [26] proved the globally linear convergence of 3-block ADMM (4). More recently, Lin, Ma and Zhang [25] further proposed several alternative approaches to ensure the sublinear convergence rate of (4) without requiring any function to be strongly convex. Remark that in all these works, to trade for a convergence guarantee the penalty parameter γ is required to be small, which potentially affects the practical effectiveness of the 3-block ADMM (4), while the 2-block ADMM (2) does not suffer from such compromises.

Alternatively, one may opt to modify the 3-block ADMM (4) to achieve convergence, with similar per-iteration computational complexity as (4). The existing methods in the literature along this line can be classified into the following three main categories. (i) The first class of algorithms requires a correction step in the updates (see, e.g., [15, 16, 17, 14]). (ii) The second class of algorithms adds proximal terms and/or dual step size to the ADMM updates, i.e., these algorithms change (4) to

$$\begin{cases} x_1^{k+1} & := \text{argmin}_{x_1 \in \mathcal{X}_1} \ \mathcal{L}_\gamma(x_1, x_2^k, x_3^k; \lambda^k) + \frac{1}{2}\|x - x_1^k\|_{P_1} \\ x_2^{k+1} & := \text{argmin}_{x_2 \in \mathcal{X}_2} \ \mathcal{L}_\gamma(x_1^{k+1}, x_2, x_3^k; \lambda^k) + \frac{1}{2}\|x - x_2^k\|_{P_2} \\ x_3^{k+1} & := \text{argmin}_{x_3 \in \mathcal{X}_3} \ \mathcal{L}_\gamma(x_1^{k+1}, x_2^{k+1}, x_3; \lambda^k) + \frac{1}{2}\|x - x_3^k\|_{P_3} \\ \lambda^{k+1} & := \lambda^k - \alpha\gamma(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b), \end{cases} \tag{5}$$

where matrices $P_i \succeq 0$ and $\alpha > 0$ denotes a step size for the dual update. Global convergence and convergence rate for (5) and its variants (for example, allowing to update x_1, x_2, x_3 in a Jacobian manner instead of a Gauss-Seidel manner) are analyzed under various conditions (see, e.g., [20, 7, 19, 31, 22]). Note that these works usually require restrictive conditions on P_i, α and γ that may also affect the performance of solving large-scale problems arising from practice. Notwithstanding all these efforts, many authors acknowledge that the unmodified 3-block ADMM (4) usually outperforms its variants (5) and the ones with correction step in practice (see, e.g., the discussions in [31, 35]). (iii) The recent work by Sun, Luo and Ye [32] on a randomly permuted ADMM is probably the only variant of 3-block ADMM which does not restrict the γ value, but its convergence is now only guaranteed for solving a squared and nonsingular linear system.

Motivated by the fact that the 2-block ADMM (2) allows the parameter to be free, in this paper we set out to explore the structures of 3-block model for which the unmodified 3-block ADMM (4) converges for all parameter values. Given the superior performance of (4), such property is of great practical importance. In this paper, we show that the 3-block ADMM (4) is *globally convergent for any fixed* $\gamma > 0$ when it is applied to solving a class of convex problems, termed the Regularized Least Squares Decomposition (RLSD) in this paper, which covers many important applications in practice as we shall discuss next.

2 Regularized Least Squares Decomposition

Let us consider the following problem, to be called regularized least squares decomposition (RLSD):

$$\begin{array}{ll} \min & f_1(x_1) + f_2(x_2) + \frac{1}{2}\|A_1 x_1 + A_2 x_2 - b\|^2 \\ \text{s.t.} & x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, \end{array} \tag{6}$$

where one seeks to decompose the observed data b into two components $A_1 x_1$ and $A_2 x_2$, and f_1 and f_2 denote some regularization functions that promote certain structures of x_1 and x_2 in the decomposed terms. One may also view (6) as a data fitting problem with two regularization terms, where $\|A_1 x_1 + A_2 x_2 - b\|^2$ denotes a least squares loss function on the data fitting term. One way to solve (6) is to apply the 3-block ADMM (4) to solve its equivalent reformulation:

$$\begin{array}{ll} \min & f_1(x_1) + f_2(x_2) + f_3(x_3) \\ \text{s.t.} & A_1 x_1 + A_2 x_2 + x_3 = b, \quad x_i \in \mathcal{X}_i, \ i = 1, 2, \end{array} \tag{7}$$

where $f_3(x_3) = \frac{1}{2}\|x_3\|_2^2$. Many works in the literature (including Boyd et al. [2] and Hong, Luo and Razaviyayn [21]) have suggested to apply ADMM to solve (6) by reformulating it as (7). The advantage of using ADMM to solve (7) is that the subproblems are usually easy to solve. Especially, the subproblem for x_3 has a closed-form solution. Yang and Zhang [39] applied the 2-block ADMM to solve the following ℓ_1-norm regularized least squares problem (or the so-called Lasso problem [34] in statistics):

$$\min_x \ \beta\|x\|_1 + \frac{1}{2}\|Ax - b\|^2, \tag{8}$$

where $\beta > 0$ is a weighting parameter. Therefore, the Lasso problem is in fact RLSD with one block of variables (more on this later). In order to use ADMM, Yang and Zhang [39] reformulated (8) as

$$\begin{array}{ll} \min_{x,r} & \beta\|x\|_1 + \frac{1}{2}\|r\|^2 \\ \text{s.t.} & Ax - r = b, \end{array} \tag{9}$$

in which the two-block variables x and r are associated with two structured functions $\|x\|_1$ and $\|r\|^2$, respectively. Numerical experiments conducted in [39] showed that the 2-block ADMM greatly outperforms other state-of-the-art solvers on this problem. It is noted that the problem RLSD (6) reduces to the Lasso problem (8) when f_2 and x_2 vanish and f_1 is the ℓ_1 norm. Problem RLSD (6) actually covers many interesting applications in practice, and in the following we will discuss a few examples. RLSD (6) is sometimes also known as sharing problem in the literature, and we refer the interested readers to [2] and [21] for more examples of this problem.

Example 2.1 *Stable principal component pursuit [40]. This problem aims to recover a low-rank matrix (the principal components) from a high dimensional data matrix despite both small entry-wise noise and gross sparse errors. This problem can be formulated as (see Eq. (15) of [40]):*

$$\min_{L,S} \ \beta_1\|L\|_* + \beta_2\|S\|_1 + \frac{1}{2}\|M - L - S\|_F^2, \tag{10}$$

where $M \in \mathbb{R}^{m\times n}$ is the given corrupted data matrix, L and S are respectively low-rank and sparse component of M. It is obvious that this problem is in the form of (6) with $\mathcal{X}_1 = \mathcal{X}_2 = \mathbb{R}^{m\times n}$. For solving (10) using the 3-block ADMM (4), see [33].

Example 2.2 *Static background extraction from surveillance video [23, 28]. This problem aims to extract the static background from a surveillance video. Given a sequence of frames of a surveillance video $M \in \mathbb{R}^{m\times n}$, this problem finds a decomposition of M in the form of $M = ue^\top + S$, where $u \in \mathbb{R}^m$ denotes the static background of the video, e is the all-ones vector, and S denotes the sparse moving foreground in the video. Since the components of u represent the pixel values of the background image, we can restrict u as $b_\ell \le u \le b_u$, with $b_\ell = 0$ and $b_u = 255$. This problem can then be formulated as*

$$\begin{array}{ll} \min_{u,S} & \beta\|S\|_1 + \frac{1}{2}\|M - ue^\top - S\|_F^2 \\ \text{s.t.} & b_\ell \le u \le b_u. \end{array} \tag{11}$$

Note that (11) *is a slight modification of Eq. (1.9) in [23] with the bounded constraints added to* u *in order to get a background image with more physical meanings. A similar model was considered by Ma et al. in [28] for molecular pattern discovery and cancer gene identification. We refer the interested readers to [23] and [28] for more details of this problem.*

Example 2.3 *Compressive Principal Component Pursuit [38]. This problem also considers decomposing a matrix* M *into a low-rank part and a sparse part as* (10)*. The difference is that* M *is observed via a small set of linear measurements. This problem can thus be formulated as*

$$\min_{L,S} \ \beta_1\|L\|_* + \beta_2\|S\|_1 + \frac{1}{2}\|M - \mathcal{A}(L) - \mathcal{A}(S)\|_F^2, \tag{12}$$

where $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ *is a linear mapping. Note that* (12) *is an unconstrained version of Eq. (1.7) in [38], and* (12) *is particularly interesting when there are noises in the compressive measurements* M*. Similar problem has also been considered in [36].*

In this paper, we prove that the unmodified 3-block ADMM (4) globally converges with any parameter $\gamma > 0$, when it is applied to solve the RLSD problem (7). This result provides theoretical foundations for using the unmodified 3-block ADMM with *a free choice of any parameter* $\gamma > 0$.

The following assumptions are made throughout this paper.

Assumption 2.4 *The optimal set* Ω^* *for problem* (7) *is non-empty.*

According to the first-order optimality conditions for (7), solving (7) is equivalent to finding

$$(x_1^*, x_2^*, x_3^*, \lambda^*) \in \Omega^*$$

such that the following holds:

$$\begin{cases} f_1(x_1) - f_1(x_1^*) - (x_1 - x_1^*)^\top \left(A_1^\top \lambda^*\right) \geq 0, & \forall x_1 \in \mathcal{X}_1, \\ f_2(x_2) - f_2(x_2^*) - (x_2 - x_2^*)^\top \left(A_2^\top \lambda^*\right) \geq 0, & \forall x_2 \in \mathcal{X}_2, \\ \nabla f_3(x_3^*) - \lambda^* = 0, \\ A_1 x_1^* + A_2 x_2^* + x_3^* = b. \end{cases} \tag{13}$$

Assumption 2.5 *We assume the following conditions hold.*

1. A_1 *and* A_2 *have full column rank.*

2. *The objective functions* f_1 *and* f_2 *are lower semi-continuous, and proper closed convex functions.*

3. *$f_i + \boldsymbol{1}_{\mathcal{X}_i}, i = 1, 2$, are both coercive functions, where $\boldsymbol{1}_{\mathcal{X}_i}$ denotes the indicator function of $\mathcal{X}_i$, i.e.,*

$$\boldsymbol{1}_{\mathcal{X}_i}(x_i) = \begin{cases} 0, & \text{if } x_i \in \mathcal{X}_i \\ +\infty, & \text{otherwise.} \end{cases}$$

Note that this assumption implies that f_1 and f_2 have finite lower bounds on $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively, i.e.,

$$\inf_{x_1 \in \mathcal{X}_1} f_1(x_1) > f_1^* > -\infty, \qquad \inf_{x_2 \in \mathcal{X}_2} f_2(x_2) > f_2^* > -\infty.$$

Remark 2.6 *We remark here that requiring $f_i + \boldsymbol{1}_{\mathcal{X}_i}$ to be a coercive function is not a restrictive assumption. Many functions used as regularization terms including ℓ_1-norm, ℓ_2-norm, ℓ_∞-norm for vectors and nuclear norm for matrices are all coercive functions; assuming the compactness of $\mathcal{X}_i$ also leads to the coerciveness of $f_i + \boldsymbol{1}_{\mathcal{X}_i}$. For instance, problems considered in Examples 2.1-2.3 all satisfy this assumption.*

Our main result in this paper is summarized in the following theorem, whose proof will be given in Section 3.

Theorem 2.7 *Assume that Assumptions 2.4 and 2.5 hold. For any given $\gamma > 0$, let $(x_1^k, x_2^k, x_3^k; \lambda^k)$ be the sequence generated by the 3-block ADMM (4) for solving (7). Then any limit point of $(x_1^k, x_2^k, x_3^k; \lambda^k)$ is an optimal solution to problem (7). Moreover, the objective function value converges to the optimal value and the constraint violation converges to zero, i.e.,*

$$\lim_{k \to \infty} \left| f(x_1^k) + f_2(x_2^k) + f_3(x_3^k) - f^* \right| = 0, \quad \text{and} \quad \lim_{k \to \infty} \left\| A_1 x_1^k + A_2 x_2^k + x_3^k - b \right\| = 0, \tag{14}$$

where f^* denotes the optimal objective value of problem (7).

In our analysis, the following well-known identity and inequality are used frequently,

$$(w_1 - w_2)^\top (w_3 - w_1) = \frac{1}{2} \left(\|w_2 - w_3\|^2 - \|w_1 - w_2\|^2 - \|w_1 - w_3\|^2 \right), \tag{15}$$

$$w_1^\top w_2 \geq -\frac{1}{2\xi} \|w_1\|^2 - \frac{\xi}{2} \|w_2\|^2, \quad \forall \xi > 0. \tag{16}$$

Notations. We denote by $f(u) \equiv \sum_{i=1}^{3} f_i(x_i)$ the sum of the separable functions. We will use the following notations to simplify the presentation

$$u := \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, w := \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda \end{pmatrix}, F(w) := \begin{pmatrix} -A_1^\top \lambda \\ -A_2^\top \lambda \\ -\lambda \\ A_1 x_1 + A_2 x_2 + x_3 - b \end{pmatrix}. \tag{17}$$

When there is no ambiguity, we often use $\| \cdot \|$ to denote the Euclidean norm $\| \cdot \|_2$.

3 Convergence Analysis

In this section, we shall prove Theorem 2.7. We will divide the proof into three parts: Theorems 3.1, 3.2 and 3.3 show that the conclusion of Theorem 2.7 holds true if $\gamma \in (1, +\infty)$, $\gamma \in \left(\sqrt{2}-1, 1\right]$ and $\gamma \in \left(0, \frac{1}{2}\right]$, respectively. As a result, combining Theorems 3.1, 3.2 and 3.3 the conclusion of Theorem 2.7 follows for any $\gamma > 0$.

Since $f_3(x_3) = \frac{1}{2}\|x_3\|_2^2$ in (7), the 3-block ADMM (4) for solving (7) reduces to

$$x_1^{k+1} := \underset{x_1 \in \mathcal{X}_1}{\operatorname{argmin}} \; f(x_1) + \frac{\gamma}{2}\left\| A_1 x_1 + A_2 x_2^k + x_3^k - b - \frac{1}{\gamma}\lambda^k \right\|^2, \tag{18}$$

$$x_2^{k+1} := \underset{x_2 \in \mathcal{X}_2}{\operatorname{argmin}} \; f(x_2) + \frac{\gamma}{2}\left\| A_1 x_1^{k+1} + A_2 x_2 + x_3^k - b - \frac{1}{\gamma}\lambda^k \right\|^2, \tag{19}$$

$$x_3^{k+1} := \frac{1}{\gamma+1}\left[\lambda^k - \gamma\left(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b\right)\right], \tag{20}$$

$$\lambda^{k+1} := \lambda^k - \gamma\left(A_1 x_1^{k+1} + A_2 x_2^{k+1} + x_3^{k+1} - b\right). \tag{21}$$

An immediate observation from (20) and (21) is $x_3^k = \lambda^k$ for any $k > 0$.

The first-order optimality conditions for (18)-(19) are given by $x_i^{k+1} \in \mathcal{X}_i, i = 1, 2$ and

$$\left(x_1 - x_1^{k+1}\right)^\top \left[g_1(x_1^{k+1}) - A_1^\top \lambda^k + \gamma A_1^\top \left(A_1 x_1^{k+1} + A_2 x_2^k + x_3^k - b\right)\right] \geq 0, \quad \forall x_1 \in \mathcal{X}_1, \tag{22}$$

$$\left(x_2 - x_2^{k+1}\right)^\top \left[g_2(x_2^{k+1}) - A_2^\top \lambda^k + \gamma A_2^\top \left(A_1 x_1^{k+1} + A_2 x_2^{k+1} + x_3^k - b\right)\right] \geq 0, \quad \forall x_2 \in \mathcal{X}_2, \tag{23}$$

where $g_i \in \partial f_i$ is the subgradient of f_i for $i = 1, 2$. Moreover, by combining with (21), (22) and (23) can be rewritten as

$$\left(x_1 - x_1^{k+1}\right)^\top \left[g_1(x_1^{k+1}) - A_1^\top \lambda^{k+1} + \gamma A_1^\top \left(A_2(x_2^k - x_2^{k+1}) + (x_3^k - x_3^{k+1})\right)\right] \geq 0, \tag{24}$$

$$\left(x_2 - x_2^{k+1}\right)^\top \left[g_2(x_2^{k+1}) - A_2^\top \lambda^{k+1} + \gamma A_2^\top \left(x_3^k - x_3^{k+1}\right)\right] \geq 0. \tag{25}$$

3.1 Proof for $\gamma \in (1, +\infty)$

In this subsection, we prove that the 3-block ADMM (18)-(21) is convergent for any $\gamma \in (1, +\infty)$.

Theorem 3.1 *Let $\left(x_1^k, x_2^k, x_3^k, \lambda^k\right)$ be generated by the 3-block ADMM* (18)-(21)*, and $\gamma \in (1, +\infty)$. Then $\left(x_1^k, x_2^k, x_3^k, \lambda^k\right)$ is bounded, and any of its cluster point $(x_1^*, x_2^*, x_3^*, \lambda^*)$ is an optimal solution of* (7)*. Moreover,* (14) *holds.*

Proof. Note that the augmented Lagrangian function is

$$\mathcal{L}_\gamma(x_1, x_2, x_3; \lambda) = f_1(x_1) + f_2(x_2) + \frac{1}{2}\|x_3\|_2^2 - \langle \lambda, A_1 x_1 + A_2 x_2 + x_3 - b\rangle + \frac{\gamma}{2}\|A_1 x_1 + A_2 x_2 + x_3 - b\|_2^2.$$

The following inequalities hold:

$$\begin{array}{rl} & \mathcal{L}_\gamma(x_1^k, x_2^k, x_3^k; \lambda^k) - \mathcal{L}_\gamma(x_1^{k+1}, x_2^k, x_3^k; \lambda^k) \\ = & f_1(x_1^k) - f_1(x_1^{k+1}) - \langle \lambda^k, A_1 x_1^k - A_1 x_1^{k+1} \rangle \\ & + \frac{\gamma}{2}\|A_1 x_1^k + A_2 x_2^k + x_3^k - b\|_2^2 - \frac{\gamma}{2}\|A_1 x_1^{k+1} + A_2 x_2^k + x_3^k - b\|_2^2 \\ \geq & g_1(x_1^{k+1})^\top (x_1^k - x_1^{k+1}) - \langle \lambda^k, A_1 x_1^k - A_1 x_1^{k+1} \rangle \\ & + \gamma (A_1 x_1^k - A_1 x_1^{k+1})^\top (A_1 x_1^{k+1} + A_2 x_2^k + x_3^k - b) + \frac{\gamma}{2}\|A_1 x_1^k - A_1 x_1^{k+1}\|_2^2 \\ \geq & \frac{\gamma}{2}\|A_1 x_1^k - A_1 x_1^{k+1}\|_2^2, \end{array} \tag{26}$$

where the first inequality is due to the convexity of f_1 and the identity (15), and the second inequality is obtained by setting $x_1 = x_1^k$ in (22). Similarly,

$$\begin{array}{rl} & \mathcal{L}_\gamma(x_1^{k+1}, x_2^k, x_3^k; \lambda^k) - \mathcal{L}_\gamma(x_1^{k+1}, x_2^{k+1}, x_3^k; \lambda^k) \\ = & f_2(x_2^k) - f_2(x_2^{k+1}) - \langle \lambda^k, A_2 x_2^k - A_2 x_2^{k+1} \rangle \\ & + \frac{\gamma}{2}\|A_1 x_1^{k+1} + A_2 x_2^k + x_3^k - b\|_2^2 - \frac{\gamma}{2}\|A_1 x_1^{k+1} + A_2 x_2^{k+1} + x_3^k - b\|_2^2 \\ \geq & g_2(x_2^{k+1})^\top (x_2^k - x_2^{k+1}) - \langle \lambda^k, A_2 x_2^k - A_2 x_2^{k+1} \rangle \\ & + \gamma (A_2 x_2^k - A_2 x_2^{k+1})^\top (A_1 x_1^{k+1} + A_2 x_2^{k+1} + x_3^k - b) + \frac{\gamma}{2}\|A_2 x_2^k - A_2 x_2^{k+1}\|_2^2 \\ \geq & \frac{\gamma}{2}\|A_2 x_2^k - A_2 x_2^{k+1}\|_2^2, \end{array} \tag{27}$$

where the first inequality is due to the convexity of f_2 and the identity (15), and the second inequality is obtained by setting $x_2 = x_2^k$ in (23). By (20), it is easy to show that

$$\mathcal{L}_\gamma(x_1^{k+1}, x_2^{k+1}, x_3^k; \lambda^k) - \mathcal{L}_\gamma(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}; \lambda^k) \geq \frac{\gamma+1}{2}\left\|x_3^k - x_3^{k+1}\right\|^2. \tag{28}$$

Combining (26), (27) and (28) yields

$$\begin{array}{rl} & \mathcal{L}_\gamma(x_1^k, x_2^k, x_3^k, \lambda^k) - \mathcal{L}_\gamma(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^k) \\ \geq & \frac{\gamma}{2}\|A_1 x_1^k - A_1 x_1^{k+1}\|^2 + \frac{\gamma}{2}\|A_2 x_2^k - A_2 x_2^{k+1}\|^2 + \frac{\gamma+1}{2}\|x_3^k - x_3^{k+1}\|^2. \end{array} \tag{29}$$

By (20) and (21), it is not difficult to get $\lambda^{k+1} = x_3^{k+1}$, and

$$\mathcal{L}_\gamma(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^k) - \mathcal{L}_\gamma(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1}) = -\frac{1}{\gamma}\|x_3^{k+1} - x_3^k\|^2. \tag{30}$$

Combining (29) and (30) yields,

$$\begin{array}{rl} & \mathcal{L}_\gamma(x_1^k, x_2^k, x_3^k, \lambda^k) - \mathcal{L}_\gamma(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1}) \\ \geq & \frac{\gamma}{2}\|A_1 x_1^k - A_1 x_1^{k+1}\|^2 + \frac{\gamma}{2}\|A_2 x_2^k - A_2 x_2^{k+1}\|^2 + \left(\frac{\gamma+1}{2} - \frac{1}{\gamma}\right)\|x_3^k - x_3^{k+1}\|^2 \\ \geq & M(\|A_1 x_1^k - A_1 x_1^{k+1}\|^2 + \|A_2 x_2^k - A_2 x_2^{k+1}\|^2 + \|x_3^k - x_3^{k+1}\|^2), \end{array} \tag{31}$$

where

$$M := \min\left\{\frac{\gamma}{2}, \frac{\gamma+1}{2} - \frac{1}{\gamma}\right\} > 0,$$

because of the fact that $\gamma > 1$. Therefore we know that $\mathcal{L}_\gamma(x_1^k, x_2^k, x_3^k, \lambda^k)$ is monotonically decreasing. Now we show that the augmented Lagrangian function has a uniform lower bound $L^* := f_1^* + f_2^*$. In fact, we have the following inequality:

$$
\begin{aligned}
& \mathcal{L}_\gamma\left(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1}\right) \\
= \; & f_1(x_1^{k+1}) + f_2(x_2^{k+1}) + \frac{1}{2}\left\|x_3^{k+1}\right\|^2 - \left\langle \lambda^{k+1}, \sum_{i=1}^2 A_i x_i^{k+1} + x_3^{k+1} - b \right\rangle + \frac{\gamma}{2}\left\|\sum_{i=1}^2 A_i x_i^{k+1} + x_3^{k+1} - b\right\|^2 \\
= \; & f_1(x_1^{k+1}) + f_2(x_2^{k+1}) + \frac{1}{2}\left\|\sum_{i=1}^2 A_i x_i^{k+1} - b\right\|^2 + \frac{\gamma - 1}{2}\left\|\sum_{i=1}^2 A_i x_i^{k+1} + x_3^{k+1} - b\right\|^2 \\
\geq \; & f_1^* + f_2^* = L^*,
\end{aligned} \tag{32}
$$

where in the second equality we used the fact that $x_3^{k+1} = \lambda^{k+1}$. Note that (31) and (32) imply that $\{(x_1^k, x_2^k) : k = 0, 1, \ldots\}$ is bounded by using the facts that $x_1^k \in \mathcal{X}_1$, $x_2^k \in \mathcal{X}_2$ and $f_1 + \mathbf{1}_{\mathcal{X}_1}$ and $f_2 + \mathbf{1}_{\mathcal{X}_2}$ are coercive. Note that (31) and (32) also imply that $\mathcal{L}_\gamma(x_1^k, x_2^k, x_3^k; \lambda^k)$ is convergent.

By combining (31) and (32) we know that the following holds for any integer $K > 0$:

$$
\begin{aligned}
& \sum_{k=0}^K \left(\left\|A_1 x_1^k - A_1 x_1^{k+1}\right\|^2 + \left\|A_2 x_2^k - A_2 x_2^{k+1}\right\|^2 + \left\|x_3^k - x_3^{k+1}\right\|^2\right) \\
\leq \; & \frac{1}{M}\sum_{k=0}^K \left(\mathcal{L}_\gamma(x_1^k, x_2^k, x_3^k, \lambda^k) - \mathcal{L}_\gamma(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1})\right) \\
= \; & \frac{1}{M}\left(\mathcal{L}_\gamma(x_1^0, x_2^0, x_3^0, \lambda^0) - \mathcal{L}_\gamma(x_1^{K+1}, x_2^{K+1}, x_3^{K+1}, \lambda^{K+1})\right) \\
\leq \; & \frac{1}{M}\left(\mathcal{L}_\gamma(x_1^0, x_2^0, x_3^0, \lambda^0) - L^*\right).
\end{aligned}
$$

By letting $K \to +\infty$ we obtain

$$
\sum_{k=0}^\infty \left(\left\|A_1 x_1^k - x_1^{k+1}\right\|^2 + \left\|A_2 x_2^k - A_2 x_2^{k+1}\right\|^2 + \left\|x_3^k - x_3^{k+1}\right\|^2\right) \leq \frac{1}{M}\left(\mathcal{L}_\gamma(x_1^0, x_2^0, x_3^0, \lambda^0) - L^*\right) < \infty,
$$

and hence

$$
\lim_{k\to\infty} \left(\|A_1 x_1^k - A_1 x_1^{k+1}\| + \|A_2 x_2^k - A_2 x_2^{k+1}\| + \|x_3^k - x_3^{k+1}\|\right) = 0. \tag{33}
$$

By using (21), $\lambda^k = x_3^k$, and the boundedness of $\{(x_1^k, x_2^k) : k = 0, 1, \ldots\}$, we can conclude that $\{(x_1^k, x_2^k, x_3^k, \lambda^k) : k = 0, 1, \ldots\}$ is a bounded sequence. Therefore, there exists a limit point $(x_1^*, x_2^*, x_3^*, \lambda^*)$ and a subsequence $\{k_q\}$ such that

$$
\lim_{q\to\infty} x_i^{k_q} = x_i^*, i = 1, 2, 3, \quad \lim_{q\to\infty} \lambda^{k_q} = \lambda^*.
$$

By using (33), we have

$$\lim_{q\to\infty} x_i^{k_q+1} = x_i^*, i = 1, 2, 3, \quad \lim_{q\to\infty} \lambda^{k_q+1} = \lambda^*.$$

Since $\mathcal{L}_\gamma(x_1^k, x_2^k, x_3^k; \lambda^k)$ is convergent, we know that

$$\lim_{k\to\infty} \mathcal{L}_\gamma(x_1^k, x_2^k, x_3^k; \lambda^k) = \mathcal{L}_\gamma(x_1^*, x_2^*, x_3^*; \lambda^*). \tag{34}$$

By combining (20), (21), (24) and (25), we know the following relations for any $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$:

$$\begin{aligned} f_1(x_1) - f_1(x_1^{k_q+1}) + \left(x_1 - x_1^{k_q+1}\right)^\top \left[-A_1^\top \lambda^{k_q+1} + \gamma A_1^\top \left(A_2(x_2^{k_q} - x_2^{k_q+1}) + (x_3^{k_q} - x_3^{k_q+1})\right)\right] &\geq 0, \\ f_2(x_2) - f_2(x_2^{k_q+1}) + \left(x_2 - x_2^{k_q+1}\right)^\top \left[-A_2^\top \lambda^{k_q+1} + \gamma A_2^\top \left(x_3^{k_q} - x_3^{k_q+1}\right)\right] &\geq 0, \\ x_3^{k_q+1} - \lambda^{k_q+1} &= 0, \\ A_1 x_1^{k_q+1} + A_2 x_2^{k_q+1} + x_3^{k_q+1} - b - \frac{1}{\gamma}\left(\lambda^{k_q} - \lambda^{k_q+1}\right) &= 0. \end{aligned}$$

Letting $q \to +\infty$, and using (33) and the lower semi-continuity of f_1 and f_2, we have the following relations for any $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$:

$$\begin{aligned} f_1(x_1) - f_1(x_1^*) - (x_1 - x_1^*)^\top (A_1^\top \lambda^*) &\geq 0, \\ f_2(x_2) - f_2(x_2^*) - (x_2 - x_2^*)^\top (A_2^\top \lambda^*) &\geq 0, \\ x_3^* - \lambda^* &= 0, \\ A_1 x_1^* + A_2 x_2^* + x_3^* - b &= 0. \end{aligned}$$

Therefore, $(x_1^*, x_2^*, x_3^*, \lambda^*)$ satisfies the optimality conditions of problem (7) and is an optimal solution of problem (7).

Moreover, we have

$$\|A_1 x_1^k + A_2 x_2^k + x_3^k - b\| = \frac{1}{\gamma}\|\lambda^{k-1} - \lambda^k\| \to 0, \quad \text{when } k \to +\infty,$$

and

$$\begin{aligned} & \left| f(x_1^k) + f_2(x_2^k) + \frac{1}{2}\|x_3^k\|^2 - f^* \right| \\ \leq & \left| \mathcal{L}_\gamma(x_1^k, x_2^k, x_3^k, \lambda^k) - \mathcal{L}_\gamma(x_1^*, x_2^*, x_3^*, \lambda^*) \right| + \|\lambda^k\| \cdot \|A_1 x_1^k + A_2 x_2^k + x_3^k - b\| \\ & + \frac{\gamma}{2}\|A_1 x_1^k + A_2 x_2^k + x_3^k - b\|^2 \to 0, \quad \text{when } k \to \infty, \end{aligned}$$

where we used (34). Therefore, (14) is proven. □

3.2 Proof for $\gamma \in (\sqrt{2}-1, 1]$

In this subsection, we prove that the 3-block ADMM (18)-(21) is convergent for any $\gamma \in (\sqrt{2}-1, 1]$.

Theorem 3.2 *Let* $\left(x_1^k, x_2^k, x_3^k, \lambda^k\right)$ *be generated by 3-block ADMM* (18)-(21)*, and* $\gamma \in (\sqrt{2}-1, 1]$*. Then* $\left(x_1^k, x_2^k, x_3^k, \lambda^k\right)$ *is bounded, and it converges to an optimal solution of* (7)*, which further implies that* (14) *holds.*

Proof. Let $(x_1^*, x_2^*, x_3^*, \lambda^*) \in \Omega^*$. By setting $x_1 = x_1^*$ in (24) and $x_2 = x_2^*$ in (25), we get,

$$\left(x_1^* - x_1^{k+1}\right)^\top \left[g_1(x_1^{k+1}) - A_1^\top \lambda^{k+1} + \gamma A_1^\top \left(A_2(x_2^k - x_2^{k+1}) + (x_3^k - x_3^{k+1})\right)\right] \geq 0, \tag{35}$$

$$\left(x_2^* - x_2^{k+1}\right)^\top \left[g_2(x_2^{k+1}) - A_2^\top \lambda^{k+1} + \gamma A_2^\top \left(x_3^k - x_3^{k+1}\right)\right] \geq 0. \tag{36}$$

From the optimality conditions (13), and (21), we can get

$$\begin{aligned}
& \frac{1}{\gamma}(\lambda^k - \lambda^{k+1})^\top(\lambda^{k+1} - \lambda^*) - (\lambda^k - \lambda^{k+1})^\top((A_2 x_2^k - A_2 x_2^{k+1}) + (x_3^k - x_3^{k+1})) \\
& +\gamma(A_2 x_2^{k+1} - A_2 x_2^*)^\top(A_2 x_2^k - A_2 x_2^{k+1}) + \gamma(x_3^{k+1} - x_3^*)^\top((A_2 x_2^k - A_2 x_2^{k+1}) + (x_3^k - x_3^{k+1})) \\
= \; & \frac{1}{\gamma}(\lambda^k - \lambda^{k+1})^\top(\lambda^{k+1} - \lambda^*) - \gamma(A_2 x_2^{k+1} - A_2 x_2^*)^\top(x_3^k - x_3^{k+1}) \\
& -\gamma(A_1 x_1^{k+1} - A_1 x_1^*)^\top((A_2 x_2^k - A_2 x_2^{k+1}) + (x_3^k - x_3^{k+1})) \\
= \; & (A_1 x_1^{k+1} + A_2 x_2^{k+1} + x_3^{k+1} - b)^\top(\lambda^{k+1} - \lambda^*) - \gamma(A_2 x_2^{k+1} - A_2 x_2^*)^\top(x_3^k - x_3^{k+1}) \\
& -\gamma(A_1 x_1^{k+1} - A_1 x_1^*)^\top((A_2 x_2^k - A_2 x_2^{k+1}) + (x_3^k - x_3^{k+1})) \\
= \; & \left[A_1(x_1^{k+1} - x_1^*) + A_2(x_2^{k+1} - x_2^*) + (x_3^{k+1} - x_3^*)\right]^\top(\lambda^{k+1} - \lambda^*) - \gamma(A_2 x_2^{k+1} - A_2 x_2^*)^\top(x_3^k - x_3^{k+1}) \\
& -\gamma(A_1 x_1^{k+1} - A_1 x_1^*)^\top((A_2 x_2^k - A_2 x_2^{k+1}) + (x_3^k - x_3^{k+1})) \\
= \; & (A_1 x_1^{k+1} - A_1 x_1^*)^\top \left[(\lambda^{k+1} - \lambda^*) - \gamma((A_2 x_2^k - A_2 x_2^{k+1}) + (x_3^k - x_3^{k+1}))\right] + (x_3^{k+1} - x_3^*)^\top(\lambda^{k+1} - \lambda^*) \\
& +(A_2 x_2^{k+1} - A_2 x_2^*)^\top \left[(\lambda^{k+1} - \lambda^*) - \gamma(x_3^k - x_3^{k+1})\right] \\
\geq \; & (x_1^{k+1} - x_1^*)^\top(g_1(x_1^{k+1}) - g_1(x_1^*)) + (x_2^{k+1} - x_2^*)^\top(g_2(x_1^{k+1}) - g_2(x_1^*)) + \|x_3^{k+1} - x_3^*\|^2 \\
\geq \; & \|x_3^{k+1} - x_3^*\|^2,
\end{aligned} \tag{37}$$

where the first inequality holds by adding (35) and (36), and the second inequality holds because of the monotonicity of g_1 and g_2. By using the fact that $x_3^k = \lambda^k$, (37) can be reduced to

$$\begin{aligned}
& \frac{1}{\gamma}(\lambda^k - \lambda^{k+1})^\top(\lambda^{k+1} - \lambda^*) + \gamma(A_2 x_2^{k+1} - A_2 x_2^*)^\top(A_2 x_2^k - A_2 x_2^{k+1}) + \gamma(x_3^{k+1} - x_3^*)^\top(x_3^k - x_3^{k+1}) \\
\geq \; & \|x_3^{k+1} - x_3^*\|^2 + \|x_3^k - x_3^{k+1}\|^2 + (\lambda^k - \lambda^{k+1})^\top(A_2 x_2^k - A_2 x_2^{k+1}) \\
& -\gamma(x_3^{k+1} - x_3^*)^\top(A_2 x_2^k - A_2 x_2^{k+1}).
\end{aligned} \tag{38}$$

Now by applying (15) to the three terms on the left hand side of (38) we get,

$$\begin{aligned}
&\left[\frac{1}{2\gamma}\|\lambda^k-\lambda^*\|^2+\frac{\gamma}{2}\|A_2x_2^k-A_2x_2^*\|^2+\frac{\gamma}{2}\|x_3^k-x_3^*\|^2\right] \\
&-\left[\frac{1}{2\gamma}\|\lambda^{k+1}-\lambda^*\|^2+\frac{\gamma}{2}\|A_2x_2^{k+1}-A_2x_2^*\|^2+\frac{\gamma}{2}\|x_3^{k+1}-x_3^*\|^2\right] \\
\geq\ & \|x_3^{k+1}-x_3^*\|^2+\|x_3^{k+1}-x_3^k\|^2+\frac{1}{2\gamma}\|\lambda^{k+1}-\lambda^k\|^2+\frac{\gamma}{2}\|A_2x_2^{k+1}-A_2x_2^k\|^2+\frac{\gamma}{2}\|x_3^{k+1}-x_3^k\|^2 \\
&+(\lambda^k-\lambda^{k+1})^\top(A_2x_2^k-A_2x_2^{k+1})-\gamma(x_3^{k+1}-x_3^*)^\top(A_2x_2^k-A_2x_2^{k+1}). \qquad (39)
\end{aligned}$$

By applying (16), we have

$$-\gamma\left(x_3^{k+1}-x_3^*\right)^\top\left(A_2x_2^k-A_2x_2^{k+1}\right)\geq-\gamma\left\|x_3^{k+1}-x_3^*\right\|^2-\frac{\gamma}{4}\left\|A_2x_2^{k+1}-A_2x_2^k\right\|^2. \qquad (40)$$

From (39), (40) and the following identity

$$\begin{aligned}
&\frac{1}{\gamma}\left\|\lambda^{k+1}-\lambda^k\right\|^2+\left(\lambda^{k+1}-\lambda^k\right)^\top\left(A_2x_2^{k+1}-A_2x_2^k\right)+\frac{\gamma}{4}\left\|A_2x_2^{k+1}-A_2x_2^k\right\|^2 \\
=\ & \left\|\sqrt{\frac{1}{\gamma}}\left(\lambda^{k+1}-\lambda^k\right)+\sqrt{\frac{\gamma}{4}}\left(A_2x_2^{k+1}-A_2x_2^k\right)\right\|^2,
\end{aligned}$$

we have

$$\begin{aligned}
&\left[\frac{1}{2\gamma}\left\|\lambda^k-\lambda^*\right\|^2+\frac{\gamma}{2}\left\|A_2x_2^k-A_2x_2^*\right\|^2+\frac{\gamma}{2}\left\|x_3^k-x_3^*\right\|^2\right] \\
&-\left[\frac{1}{2\gamma}\left\|\lambda^{k+1}-\lambda^*\right\|^2+\frac{\gamma}{2}\left\|A_2x_2^{k+1}-A_2x_2^*\right\|^2+\frac{\gamma}{2}\left\|x_3^{k+1}-x_3^*\right\|^2\right] \\
\geq\ & (1-\gamma)\left\|x_3^{k+1}-x_3^*\right\|^2+\left(1+\frac{\gamma}{2}-\frac{1}{2\gamma}\right)\left\|x_3^{k+1}-x_3^k\right\|^2+\left\|\sqrt{\frac{1}{\gamma}}\left(\lambda^{k+1}-\lambda^k\right)+\sqrt{\frac{\gamma}{4}}\left(A_2x_2^{k+1}-A_2x_2^k\right)\right\|^2 \\
\geq\ & 0, \qquad (41)
\end{aligned}$$

where the last inequality holds since $1+\frac{\gamma}{2}-\frac{1}{2\gamma}>0$ due to the fact that $\gamma\in\left(\sqrt{2}-1,1\right]$. In other words, $\frac{1}{2\gamma}\left\|\lambda^k-\lambda^*\right\|^2+\frac{\gamma}{2}\left\|A_2x_2^k-A_2x_2^*\right\|^2+\frac{\gamma}{2}\left\|x_3^k-x_3^*\right\|^2$ is non-increasing and lower bounded, and thus it is convergent. This further implies that $\|x_3^{k+1}-x_3^k\|\to 0$ from (41). Hence, $\|\lambda^{k+1}-\lambda^k\|\to 0$. Finally, again from (41) we have $\|A_2x_2^{k+1}-A_2x_2^k\|\to 0$.

Since (41) also shows that $\frac{1}{2\gamma}\|\lambda^k-\lambda^*\|^2+\frac{\gamma}{2}\|A_2x_2^k-A_2x_2^*\|^2+\frac{\gamma}{2}\|x_3^k-x_3^*\|^2$ is upper bounded, we can conclude that $\left\{\left(x_2^k,x_3^k,\lambda^k\right):k=0,1,\ldots\right\}$ is bounded because A_2 has full column rank. It follows from (21) and the fact that A_1 has full column rank that $\left\{x_1^k:k=0,1,\ldots\right\}$ is bounded. Therefore, there exists a limit point $\left(\bar{x}_1,\bar{x}_2,\bar{x}_3,\bar{\lambda}\right)$ and a subsequence $\{k_q\}$ such that

$$\lim_{q\to\infty}x_i^{k_q}=\bar{x}_i, i=1,2,3, \quad \lim_{q\to\infty}\lambda^{k_q}=\bar{\lambda}.$$

By $\|A_2x_2^{k+1}-A_2x_2^k\|\to 0$, $\|x_3^{k+1}-x_3^k\|\to 0$ and $\|\lambda^{k+1}-\lambda^k\|\to 0$, we have

$$\lim_{q\to\infty}x_i^{k_q+1}=\bar{x}_i, i=2,3, \quad \lim_{q\to\infty}\lambda^{k_q+1}=\bar{\lambda}.$$

By the same argument as in Theorem 3.1, we conclude $\left(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{\lambda}\right)$ is an optimal solution of (7).

Finally, we prove that the whole sequence $(x_1^k, x_2^k, x_3^k, \lambda^k)$ converges to $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{\lambda})$, which gives the conclusion of Theorem 3.2 and also implies (14). It suffices to prove that $(A_1 x_1^k, A_2 x_2^k, x_3^k, \lambda^k)$ converges to $(A_1\bar{x}_1, A_2\bar{x}_2, \bar{x}_3, \bar{\lambda})$ since A_1 and A_2 both have full column rank. Note that since $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{\lambda})$ is an optimal solution of (7), (41) holds with $(x_2^*, x_3^*, \lambda^*)$ replaced by $(\bar{x}_2, \bar{x}_3, \bar{\lambda})$. Therefore, $\frac{1}{2\gamma}\|\lambda^k - \bar{\lambda}\|^2 + \frac{\gamma}{2}\|A_2 x_2^k - A_2\bar{x}_2\|^2 + \frac{\gamma}{2}\|x_3^k - \bar{x}_3\|^2$ is non-increasing. Moreover, we have $\frac{1}{2\gamma}\|\lambda^{k_q} - \bar{\lambda}\|^2 + \frac{\gamma}{2}\|A_2 x_2^{k_q} - A_2\bar{x}_2\|^2 + \frac{\gamma}{2}\|x_3^{k_q} - \bar{x}_3\|^2 \to 0$. Therefore, it follows that

$$\frac{1}{2\gamma}\|\lambda^k - \bar{\lambda}\|^2 + \frac{\gamma}{2}\|A_2 x_2^k - A_2\bar{x}_2\|^2 + \frac{\gamma}{2}\|x_3^k - \bar{x}_3\|^2 \to 0,$$

i.e., the whole sequence of $(A_2 x_2^k, x_3^k, \lambda^k)$ converges to $(A_2\bar{x}_2, \bar{x}_3, \bar{\lambda})$. Furthermore, $\|A_1 x_1^k - A_1\bar{x}_1\| \to 0$ by using (21). This completes the proof. □

3.3 Proof for $\gamma \in (0, \frac{1}{2})$

In this subsection, we prove that the 3-block ADMM (18)-(21) is convergent for any $\gamma \in (0, \frac{1}{2})$.

Theorem 3.3 *Let* $\left(x_1^k, x_2^k, x_3^k, \lambda^k\right)$ *be generated by 3-block ADMM* (18)-(21)*, and* $\gamma \in (0, \frac{1}{2}]$. *Then* $\left(x_1^k, x_2^k, x_3^k, \lambda^k\right)$ *is bounded, and it converges to an optimal solution of* (7)*, which further implies that* (14) *holds.*

Proof. Let $(x_1^*, x_2^*, x_3^*, \lambda^*) \in \Omega^*$. By setting $x_2 = x_2^k$ in (25), and $x_2 = x_2^{k+1}$ in (25) for the k-th iteration, we can obtain

$$(x_2^k - x_2^{k+1})^\top \left[g_2(x_2^{k+1}) - A_2^\top \lambda^{k+1} + \gamma A_2^\top \left(x_3^k - x_3^{k+1}\right)\right] \geq 0, \tag{42}$$

$$(x_2^{k+1} - x_2^k)^\top \left[g_2(x_2^k) - A_2^\top \lambda^k + \gamma A_2^\top \left(x_3^{k-1} - x_3^k\right)\right] \geq 0. \tag{43}$$

Summing (42) and (43) yields

$$\begin{aligned} & \left(A_2 x_2^{k+1} - A_2 x_2^k\right)^\top \left(\lambda^{k+1} - \lambda^k\right) \\ \geq\ & \left(x_2^{k+1} - x_2^k\right)^\top \left[g_2(x_2^{k+1}) - g_2(x_2^k)\right] + \left(A_2 x_2^{k+1} - A_2 x_2^k\right)^\top \left[\left(x_3^k - x_3^{k-1}\right) + \left(x_3^k - x_3^{k+1}\right)\right] \\ \geq\ & -\frac{\gamma}{3}\left\|A_2 x_2^{k+1} - A_2 x_2^k\right\|^2 - \frac{3\gamma}{2}\left\|x_3^k - x_3^{k-1}\right\|^2 - \frac{3\gamma}{2}\left\|x_3^{k+1} - x_3^k\right\|^2, \end{aligned} \tag{44}$$

where the second inequality follows from the monotonicity of g_2 and (16). Note that from (16) we also have the following inequality:

$$-\gamma\left(x_3^{k+1} - x_3^*\right)^\top \left(A_2 x_2^k - A_2 x_2^{k+1}\right) \geq -2\gamma\left\|x_3^{k+1} - x_3^*\right\|^2 - \frac{\gamma}{8}\left\|A_2 x_2^k - A_2 x_2^{k+1}\right\|^2. \tag{45}$$

Note from the proof of Theorem 3.2 that (39) holds for any $\gamma > 0$. By combining (44), (45) and (39), we have

$$\begin{aligned}
& \left[\frac{1}{2\gamma}\left\|\lambda^k - \lambda^*\right\|^2 + \frac{\gamma}{2}\left\|A_2x_2^k - A_2x_2^*\right\|^2 + \frac{\gamma}{2}\left\|x_3^k - x_3^*\right\|^2 + \frac{3\gamma}{2}\left\|x_3^k - x_3^{k-1}\right\|^2\right] \\
& - \left[\frac{1}{2\gamma}\left\|\lambda^{k+1} - \lambda^*\right\|^2 + \frac{\gamma}{2}\left\|A_2x_2^{k+1} - A_2x_2^*\right\|^2 + \frac{\gamma}{2}\left\|x_3^{k+1} - x_3^*\right\|^2 + \frac{3\gamma}{2}\left\|x_3^{k+1} - x_3^k\right\|^2\right] \\
\geq \; & (1-2\gamma)\left\|x_3^{k+1} - x_3^*\right\|^2 + \left(1 + \frac{\gamma}{2} + \frac{1}{2\gamma} - 3\gamma\right)\left\|x_3^{k+1} - x_3^k\right\|^2 + \frac{\gamma}{24}\left\|A_2x_2^{k+1} - A_2x_2^k\right\|^2 \\
\geq \; & \frac{1+2\gamma-5\gamma^2}{2\gamma}\left\|x_3^{k+1} - x_3^k\right\|^2 + \frac{\gamma}{24}\left\|A_2x_2^{k+1} - A_2x_2^k\right\|^2 \\
\geq \; & 0, \qquad (46)
\end{aligned}$$

where we used the facts that $\lambda^k = x_3^k$ and $\gamma \in \left(0, \frac{1}{2}\right]$. Therefore, we have $\|x_3^{k+1} - x_3^k\| \to 0$, $\|A_2x_2^{k+1} - A_2x_2^k\| \to 0$, and hence $\|\lambda^{k+1} - \lambda^k\| \to 0$. By the same arguments as in the proof of Theorem 3.2, we conclude that $(x_1^k, x_2^k, x_3^k, \lambda^k)$ is bounded, and any of its cluster point $\left(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{\lambda}\right)$ is an optimal solution of (7). Also by the same arguments as in the proof of Theorem 3.2, we can prove that the whole sequence $\left(x_1^k, x_2^k, x_3^k, \lambda^k\right)$ converges to $\left(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{\lambda}\right)$, and this completes the proof. □

4 Extensions

In this section, we give some extended results of the convergence of 3-block ADMM (4) for solving (7) that do not restrict $f_3(x_3)$ to be $\frac{1}{2}\|x_3\|^2$. Instead, we make the following assumption for f_3 in this section.

Assumption 4.1 *We assume that function f_3 is lower bounded by f_3^* and is strongly convex with parameter $\sigma > 0$ and ∇f_3 is Lipschitz continuous with Lipschitz constant $L > 0$; i.e., the following inequalities hold:*

$$\inf_{x_3 \in \mathbb{R}^p} f_3(x_3) > f_3^* > -\infty,$$

$$f_3(y) \geq f_3(x) + (y-x)^\top \nabla f_3(x) + \frac{\sigma}{2}\|y-x\|^2, \quad \forall x, y \in \mathbb{R}^p \qquad (47)$$

or equivalently,

$$(y-x)^\top (\nabla f_3(y) - \nabla f_3(x)) \geq \sigma \|y-x\|^2, \quad \forall x, y \in \mathbb{R}^p \qquad (48)$$

and

$$\|\nabla f_3(y) - \nabla f_3(x)\| \leq L\|y-x\|, \quad \forall x, y \in \mathbb{R}^p. \qquad (49)$$

For the ease of presentation, we restate the problem (7) here (with $f_3(x_3)$ not restricted as $\frac{1}{2}\|x_3\|^2$) as

$$\begin{array}{ll} \min & f_1(x_1) + f_2(x_2) + f_3(x_3) \\ \text{s.t.} & A_1x_1 + A_2x_2 + x_3 = b, \quad x_i \in \mathcal{X}_i,\ i = 1, 2, \end{array} \tag{50}$$

where f_3 satisfies Assumption 4.1. In this section, we show that the 3-block ADMM (4) converges when it is applied to solve (50), given that γ is chosen to be any value in the following range:

$$\begin{aligned} \gamma \quad \in \quad & \left(0, \min\left\{\frac{4\sigma}{\eta_2}, \frac{\sigma(\eta_2-2)}{4\eta_2} + \sqrt{\frac{\sigma^2(\eta_2-2)^2}{16\eta_2^2} + \frac{\sigma^2(\eta_2-2)}{4\eta_2}}\right\}\right) \bigcup \left(\sqrt{\sigma^2 + \frac{2L^2}{\eta_1-2}} - \sigma, \frac{4\sigma}{\eta_1}\right] \\ & \bigcup \left(\frac{\sqrt{\sigma^2+8L^2}-\sigma}{2}, +\infty\right), \end{aligned} \tag{51}$$

where η_1 and η_2 can be any value in $(2, +\infty)$. Note that if η_1 is chosen such that $\sqrt{\sigma^2 + \frac{2L^2}{\eta_1-2}} - \sigma > \frac{4\sigma}{\eta_1}$, then the second interval in (51) is empty.

Note that the 3-block ADMM for solving (50) can be written as

$$\left\{\begin{array}{ll} x_1^{k+1} & := \operatorname{argmin}_{x_1\in\mathcal{X}_1}\ f_1(x_1) + \frac{\gamma}{2}\|A_1x_1 + A_2x_2^{k+1} + x_3^{k+1} - b - \lambda^k/\gamma\|^2 \\ x_2^{k+1} & := \operatorname{argmin}_{x_2\in\mathcal{X}_2}\ f_2(x_2) + \frac{\gamma}{2}\|A_1x_1^{k+1} + A_2x_2 + x_3^{k+1} - b - \lambda^k/\gamma\|^2 \\ x_3^{k+1} & := \operatorname{argmin}_{x_3\in\mathbb{R}^p}\ f_3(x_3) + \frac{\gamma}{2}\|A_1x_1^{k+1} + A_2x_2^{k+1} + x_3 - b - \lambda^k/\gamma\|^2 \\ \lambda^{k+1} & := \lambda^k - \gamma(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b). \end{array}\right. \tag{52}$$

The first-order optimality conditions for the three subproblems in (52) are given by $x_i^{k+1} \in \mathcal{X}_i$ and $x_i \in \mathcal{X}_i$ for $i = 1, 2$, and

$$\left(x_1 - x_1^{k+1}\right)^\top \left[g_1(x_1^{k+1}) - A_1^\top\lambda^k + \gamma A_1^\top\left(A_1x_1^{k+1} + A_2x_2^k + x_3^k - b\right)\right] \geq 0, \tag{53}$$

$$\left(x_2 - x_2^{k+1}\right)^\top \left[g_2(x_2^{k+1}) - A_2^\top\lambda^k + \gamma A_2^\top\left(A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^k - b\right)\right] \geq 0, \tag{54}$$

$$\nabla f_3(x_3^{k+1}) - \lambda^k + \gamma\left(A_1x_1^{k+1} + A_2x_2^{k+1} + x_3^{k+1} - b\right) = 0, \tag{55}$$

where $g_i \in \partial f_i$ is the subgradient of f_i for $i = 1, 2$. Moreover, by combining with the updating formula for λ^{k+1}, (53)-(55) can be rewritten as

$$\left(x_1 - x_1^{k+1}\right)^\top \left[g_1(x_1^{k+1}) - A_1^\top\lambda^{k+1} + \gamma A_1^\top\left(A_2(x_2^k - x_2^{k+1}) + (x_3^k - x_3^{k+1})\right)\right] \geq 0, \tag{56}$$

$$\left(x_2 - x_2^{k+1}\right)^\top \left[g_2(x_2^{k+1}) - A_2^\top\lambda^{k+1} + \gamma A_2^\top\left(x_3^k - x_3^{k+1}\right)\right] \geq 0, \tag{57}$$

$$\nabla f_3(x_3^{k+1}) - \lambda^{k+1} = 0. \tag{58}$$

Before presenting our main result in this section, we give a technical lemma which will be used in our subsequent analysis; the proof of the lemma can be found in the appendix.

Lemma 4.2 *Assume Assumptions 2.4 and 2.5 hold. The following results hold for the 3-block ADMM* (52) *applied to* (50) *with f_3 satisfying Assumption 4.1.*

1. *If $\gamma \in \left(\frac{\sqrt{\sigma^2+8L^2}-\sigma}{2}, +\infty\right)$, then*

$$\lim_{k\to\infty}\left\|A_2x_2^{k+1}-A_2x_2^k\right\|=0,\quad \lim_{k\to\infty}\left\|x_3^{k+1}-x_3^k\right\|=0,\quad \lim_{k\to\infty}\left\|\lambda^{k+1}-\lambda^k\right\|=0, \tag{59}$$

$\left\{\left(x_1^k,x_2^k,x_3^k,\lambda^k\right):k=0,1,2,\ldots\right\}$ *is a bounded sequence and $\mathcal{L}_\gamma(x_1^k,x_2^k,x_3^k;\lambda^k)$ converges to $\mathcal{L}_\gamma(x_1^*,x_2^*,x_3^*;\lambda^*)$.*

2. *If $\gamma \in \left(\sqrt{\sigma^2+\frac{2L^2}{\eta_1-2}}-\sigma, \frac{4\sigma}{\eta_1}\right] \bigcup \left(0, \min\left\{\frac{4\sigma}{\eta_2}, \frac{\sigma(\eta_2-2)}{4\eta_2}+\sqrt{\frac{\sigma^2(\eta_2-2)^2}{16\eta_2^2}+\frac{\sigma^2(\eta_2-2)}{4\eta_2}}\right\}\right)$ with η_1 and η_2 arbitrarily chosen in $(2,+\infty)$, then* (59) *holds, $\left\{\left(x_1^k,x_2^k,x_3^k,\lambda^k\right):k=0,1,2,\ldots\right\}$ is a bounded sequence and the whole sequence of $\left\{\left(x_1^k,x_2^k,x_3^k,\lambda^k\right):k=0,1,2,\ldots\right\}$ converges to $(x_1^*,x_2^*,x_3^*,\lambda^*)$.*

Theorem 4.3 *Assume Assumptions 2.4 and 2.5 hold. Let $\left(x_1^k,x_2^k,x_3^k,\lambda^k\right)$ be generated by the 3-block ADMM* (52) *with γ chosen as in* (51)*. Then $\left(x_1^k,x_2^k,x_3^k,\lambda^k\right)$ is bounded, and any of its cluster point $(x_1^*,x_2^*,x_3^*,\lambda^*)$ is an optimal solution of* (50)*. Moreover, we have*

$$\lim_{k\to\infty}\left|f(x_1^k)+f_2(x_2^k)+f_3(x_3^k)-f^*\right|=0,\quad \lim_{k\to\infty}\left\|A_1x_1^k+A_2x_2^k+x_3^k-b\right\|=0, \tag{60}$$

where f^* denotes the optimal objective value of problem (50)*.*

Proof. Since γ is chosen as in (51), it follows from Lemma 4.2 that $\left\{\left(x_1^k,x_2^k,x_3^k,\lambda^k\right):k=0,1,2,\ldots\right\}$ is a bounded sequence. Hence, there exists a cluster point $(x_1^*,x_2^*,x_3^*,\lambda^*)$ and a subsequence $\{k_q\}$ such that

$$\lim_{q\to\infty} x_i^{k_q}=x_i^*, i=1,2,3,\quad \lim_{q\to\infty}\lambda^{k_q}=\lambda^*.$$

By using (59), we have

$$\lim_{q\to\infty} x_i^{k_q+1}=x_i^*, i=2,3,\quad \lim_{q\to\infty}\lambda^{k_q+1}=\lambda^*.$$

For the (k_q+1)-th iteration, using the convexity of f_1 and f_2, (56)-(58) and the updating formula for λ^{k+1} can be written as

$$\begin{aligned}
f_1(x_1)-f_1(x_1^{k_q+1})+\left(x_1-x_1^{k_q+1}\right)^\top\left[-A_1^\top\lambda^{k_q+1}+\gamma A_1^\top\left(A_2(x_2^{k_q}-x_2^{k_q+1})+(x_3^{k_q}-x_3^{k_q+1})\right)\right] &\geq 0,\\
f_2(x_2)-f_2(x_2^{k_q+1})+\left(x_2-x_2^{k_q+1}\right)^\top\left[-A_2^\top\lambda^{k_q+1}+\gamma A_2^\top\left(x_3^{k_q}-x_3^{k_q+1}\right)\right] &\geq 0,\\
\nabla f_3(x_3^{k_q+1})-\lambda^{k_q+1} &= 0,\\
A_1x_1^{k_q+1}+A_2x_2^{k_q+1}+x_3^{k_q+1}-b-\frac{1}{\gamma}\left(\lambda^{k_q}-\lambda^{k_q+1}\right) &= 0,
\end{aligned}$$

where $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$. By letting $q \to +\infty$, and using (59) and the lower semi-continuity of f_1 and f_2 and the continuity of ∇f_3, we have

$$\begin{array}{rcl} f_1(x_1) - f_1(x_1^*) - (x_1 - x_1^*)^\top \left(A_1^\top \lambda^* \right) & \geq & 0, \\ f_2(x_2) - f_2(x_2^*) - (x_2 - x_2^*)^\top \left(A_2^\top \lambda^* \right) & \geq & 0, \\ \nabla f_3(x_3^*) - \lambda^* & = & 0, \\ A_1 x_1^* + A_2 x_2^* + x_3^* - b & = & 0. \end{array}$$

This implies that $(x_1^*, x_2^*, x_3^*, \lambda^*)$ is an optimal solution of problem (50). It also follows from Lemma 4.2 that

$$\lim_{k\to\infty} \left\| A_1 x_1^k + A_2 x_2^k + x_3^k - b \right\| = \lim_{k\to\infty} \frac{1}{\gamma} \left\| \lambda^k - \lambda^{k+1} \right\| = 0. \tag{61}$$

Moreover, if $\gamma \in (\frac{\sqrt{\sigma^2+8L^2}-\sigma}{2}, +\infty)$, from part 1 of Lemma 4.2 we have

$$\begin{array}{rl} & \left| f(x_1^k) + f_2(x_2^k) + f_3(x_3^k) - f^* \right| \\ \leq & \left| \mathcal{L}_\gamma \left(x_1^k, x_2^k, x_3^k, \lambda^k \right) - \mathcal{L}_\gamma \left(x_1^*, x_2^*, x_3^*, \lambda^* \right) \right| + \left\| \lambda^k \right\| \left\| A_1 x_1^k + A_2 x_2^k + x_3^k - b \right\| \\ & + \frac{\gamma}{2} \left\| A_1 x_1^k + A_2 x_2^k + x_3^k - b \right\|^2, \end{array}$$

which implies that

$$\lim_{k\to\infty} \left| f(x_1^k) + f_2(x_2^k) + f_3(x_3^k) - f^* \right| = 0, \tag{62}$$

by using part 1 of Lemma 4.2 and (61).

If γ satisfies

$$\gamma \in \left(0, \min\left\{ \frac{4\sigma}{\eta_2}, \frac{\sigma(\eta_2-2)}{4\eta_2} + \sqrt{\frac{\sigma^2(\eta_2-2)^2}{16\eta_2^2} + \frac{\sigma^2(\eta_2-2)}{4\eta_2}} \right\}\right) \bigcup \left(\sqrt{\sigma^2 + \frac{2L^2}{\eta_1 - 2}} - \sigma, \frac{4\sigma}{\eta_1} \right]$$

for arbitrarily chosen $\eta_1 > 2$ and $\eta_2 > 2$, by using part 2 of Lemma 4.2, (62) follows immediately because the whole sequence of $(x_1^k, x_2^k, x_3^k; \lambda^k)$ converges to $(x_1^*, x_2^*, x_3^*; \lambda^*)$. □

Remark 4.4 *We remark here that although the range defined in* (51) *does not cover all values in* $(0, +\infty)$*, it shows that the 3-block ADMM applied to solve* (50) *globally converges for most values of* γ*. In Table 1 we list several cases for different values of* $(\sigma, L, \eta_1, \eta_2)$*. From Table 1 we can see that in many cases the range defined in* (51) *is equal to* $(0, +\infty)$*, and in some cases although the range is not equal to* $(0, +\infty)$*, it covers most part of* $(0, +\infty)$*. In this sense, we can conclude that the choice of parameter* γ *is "relatively free" for solving* (50) *with* f_3 *satisfying Assumption 4.1.*

$(\sigma, L, \eta_1, \eta_2)$	Range in (51)
$(2, 1, 4, 4)$	$(0, +\infty)$
$(1, 0.5, 3, 5)$	$(0, +\infty)$
$(3, 2, 4, 8)$	$(0, +\infty)$
$(1, 1, 3, 4)$	$(0, 0.5) \cup (0.7321, +\infty)$

Table 1: Range of γ defined in (51)

5 Conclusions

Motivated by the fact that the 2-block ADMM globally converges for any penalty parameter $\gamma > 0$, we studied in this paper the global convergence of the 3-block ADMM. As there exists a counter-example showing that the 3-block ADMM can diverge if no further condition is imposed, it is natural to look for sufficient conditions which can guarantee the convergence of the 3-block ADMM. However, the existing results on sufficient conditions usually require γ to be smaller than a certain bound, which is usually very small and therefore not practical. In this paper, we showed that the 3-block ADMM globally converges for any $\gamma > 0$ when it is applied to solve a class of regularized least squares problems; that is, the 3-block ADMM is parameter-unrestricted for this class of problems. We also extended this result to a more general problem, and showed that the 3-block ADMM globally converges for most values of γ in $(0, +\infty)$.

References

[1] D. Boley. Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs. *SIAM Journal on Optimization*, 23(4):2183–2207, 2013.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[3] X. Cai, D. Han, and X. Yuan. The direct extension of ADMM for three-block separable convex minimization models is convergent when one function is strongly convex. *Optimization Online*, 2014.

[4] C. Chen, B. He, Y. Ye, and X. Yuan. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. *Mathematical Programming, DOI 10.1007/s10107-014-0826-5*, 2014.

[5] C. Chen, Y. Shen, and Y. You. On the convergence analysis of the alternating direction method of multipliers with three blocks. *Abstract and Applied Analysis*, Article ID 183961, 2013.

[6] D. Davis and W. Yin. A three-operator splitting scheme and its optimization applications. *Preprint*, 2015.

[7] W. Deng, M. Lai, Z. Peng, and W. Yin. Parallel multi-block ADMM with $o(1/k)$ convergence. *Preprint*, 2014.

[8] W. Deng and W. Yin. On the global and linear convergence of the generalized alternating direction method of multipliers. Technical report, Rice University CAAM, 2012.

[9] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293–318, 1992.

[10] M. Fortin and R. Glowinski. *Augmented Lagrangian methods: applications to the numerical solution of boundary-value problems*. North-Holland Pub. Co., 1983.

[11] D. Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Solution of Boundary Value Problems*. North-Holland, Amsterdam, 1983.

[12] R. Glowinski and P. Le Tallec. *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*. SIAM, Philadelphia, Pennsylvania, 1989.

[13] D. Han and X. Yuan. A note on the alternating direction method of multipliers. *Journal of Optimization Theory and Applications*, 155(1):227–238, 2012.

[14] B. He, L. Hou, and X. Yuan. On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming. *Preprint*, 2013.

[15] B. He, M. Tao, and X. Yuan. A splitting method for separate convex programming with linking linear constraints. *Optimization-online: http://www.optimization-online.org/DB_HTML/2010/06/2665.html*, 2010.

[16] B. He, M. Tao, and X. Yuan. Alternating direction method with Gaussian back substitution for separable convex programming. *SIAM Journal on Optimization*, 22:313–340, 2012.

[17] B. He, M. Tao, and X. Yuan. Convergence rate and iteration complexity on the alternating direction method of multipliers with a substitution procedure for separable convex programming. *Preprint*, 2013.

[18] B. He and X. Yuan. On the $O(1/n)$ convergence rate of Douglas-Rachford alternating direction method. *SIAM Journal on Numerical Analysis*, 50:700–709, 2012.

[19] M. Hong, T.-H. Chang, X. Wang, M. Razaviyayn, S. Ma, and Z.-Q. Luo. A block successive upper bound minimization method of multipliers for linearly constrained convex optimization. *arXiv preprint arXiv:1401.7079*, 2014.

[20] M. Hong and Z.-Q. Luo. On the linear convergence of the alternating direction method of multipliers. *Preprint*, 2012.

[21] M. Hong, Z.-Q. Luo, and M. Razaviyayn. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *Preprint available at http://arxiv.org/pdf/1410.1390v1.pdf*, 2014.

[22] M. Li, D. Sun, and K.-C. Toh. A convergent 3-block semi-proximal ADMM for convex minimization problems with one strongly convex block. *Asia-Pacific Journal of Operational Research*, 32(3):1550024 (19 pages), 2015.

[23] X. Li, M. K. Ng, and X. Yuan. Median filtering-based methods for static background extraction from surveillance video. *Numerical Linear Algebra with Applications*, 2015.

[24] T. Lin, S. Ma, and S. Zhang. On the convergence rate of multi-block ADMM. *Preprint http://arxiv.org/abs/1408.4265*, 2014.

[25] T. Lin, S. Ma, and S. Zhang. Iteration complexity analysis of multi-block ADMM for a family of convex minimization without strong convexity. *Preprint*, 2015.

[26] T. Lin, S. Ma, and S. Zhang. On the global linear convergence of the ADMM with multi-block variables. *Submitted*, May 2014.

[27] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16:964–979, 1979.

[28] S. Ma, D. Johnson, C. Ashby, D. Xiong, C. L. Cramer, J. H. Moore, S. Zhang, and X. Huang. SPARCoC: a new framework for molecular pattern discovery and cancer gene identification. *PLoS ONE*, 10(3):e0117135, 2015.

[29] R. D. C. Monteiro and B. F. Svaiter. Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. *SIAM Journal on Optimization*, 23:475–507, 2013.

[30] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2233–2246, 2012.

[31] D. Sun, K.-C. Toh, and L. Yang. A convergent 3-block semiproximal alternating direction method of multipliers for conic programming with 4-type constraints. *SIAM J. Optimization*, 25:882–915, 2015.

[32] R. Sun, Z.-Q. Luo, and Y. Ye. On the expected convergence of randomly permuted ADMM. *Preprint*, 2015.

[33] M. Tao and X. Yuan. Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM J. Optim.*, 21:57–81, 2011.

[34] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Royal. Statist. Soc B.*, 58(1):267–288, 1996.

[35] X. Wang, M. Hong, S. Ma, and Z.-Q. Luo. Solving multiple-block separable convex minimization problems using two-block alternating direction method of multipliers. *Preprint*, 2013.

[36] A. Waters, A. Sankaranarayanan, and R. Baraniuk. Sparcs: Recovering low-rank and sparse matrices from compressive measurements. In *NIPS*, 2011.

[37] Z. Wen, D. Goldfarb, and W. Yin. Alternating direction augmented Lagrangian methods for semidefinite programming. *Mathematical Programming Computation*, 2:203–230, 2010.

[38] J. Wright, A. Ganesh, K. Min, and Y. Ma. Compressive principal component pursuit. *Information and Inference*, 2(1):32–68, 2013.

[39] J. Yang and Y. Zhang. Alternating direction algorithms for ℓ_1 problems in compressive sensing. *SIAM Journal on Scientific Computing*, 33(1):250–278, 2011.

[40] Z. Zhou, X. Li, J. Wright, E. J. Candès, and Y. Ma. Stable principal component pursuit. *Proceedings of International Symposium on Information Theory*, 2010.

A Proof of Lemma 4.2

Proof. By (58) and the Lipschitz continuity of ∇f_3, we have

$$\|\lambda^{k+1} - \lambda^k\| \leq L\|x_3^{k+1} - x_3^k\|. \tag{63}$$

Letting $x_2 = x_2^k$ in the $(k+1)$-th iteration and $x_2 = x_2^{k+1}$ in the k-th iteration of (57) yields

$$\begin{aligned} (x_2^k - x_2^{k+1})^\top \left[g_2(x_2^{k+1}) - A_2^\top \lambda^{k+1} + \gamma A_2^\top \left(x_3^k - x_3^{k+1}\right)\right] &\geq 0, \\ (x_2^{k+1} - x_2^k)^\top \left[g_2(x_2^k) - A_2^\top \lambda^k + \gamma A_2^\top \left(x_3^{k-1} - x_3^k\right)\right] &\geq 0. \end{aligned}$$

Adding these two inequalities, using the monotonicity of g_2 and applying (16) we obtain that the following inequality holds for any $\epsilon > 0$:

$$\begin{aligned} & \left(A_2 x_2^{k+1} - A_2 x_2^k\right)^\top \left(\lambda^{k+1} - \lambda^k\right) \\ \geq \; & -\frac{\gamma}{\epsilon}\left\|A_2 x_2^{k+1} - A_2 x_2^k\right\|^2 - \frac{\gamma\epsilon}{2}\left\|x_3^k - x_3^{k-1}\right\|^2 - \frac{\gamma\epsilon}{2}\left\|x_3^{k+1} - x_3^k\right\|^2. \end{aligned} \tag{64}$$

From (58) and the strong convexity of f_3, we have

$$\left(x_3^{k+1} - x_3^k\right)^\top \left(\lambda^{k+1} - \lambda^k\right) \geq \sigma \left\|x_3^{k+1} - x_3^k\right\|^2. \tag{65}$$

Now we prove part 1 of Lemma 4.2. Firstly, we prove that $\mathcal{L}_\gamma(w^k)$ is a non-increasing sequence. By similar arguments as in (26), (27) and (28), we have the following inequalities:

$$\begin{aligned} \mathcal{L}_\gamma\left(x_1^k, x_2^k, x_3^k, \lambda^k\right) - \mathcal{L}_\gamma\left(x_1^{k+1}, x_2^k, x_3^k, \lambda^k\right) & \geq \frac{\gamma}{2}\left\|A_1 x_1^k - A_1 x_1^{k+1}\right\|^2, \\ \mathcal{L}_\gamma\left(x_1^{k+1}, x_2^k, x_3^k, \lambda^k\right) - \mathcal{L}_\gamma\left(x_1^{k+1}, x_2^{k+1}, x_3^k, \lambda^k\right) & \geq \frac{\gamma}{2}\left\|A_2 x_2^k - A_2 x_2^{k+1}\right\|^2, \\ \mathcal{L}_\gamma\left(x_1^{k+1}, x_2^{k+1}, x_3^k, \lambda^k\right) - \mathcal{L}_\gamma\left(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^k\right) & \geq \frac{\gamma+\sigma}{2}\left\|x_3^k - x_3^{k+1}\right\|^2. \end{aligned}$$

Summing these three inequalities yields

$$\begin{aligned} & \mathcal{L}_\gamma\left(x_1^k, x_2^k, x_3^k, \lambda^k\right) - \mathcal{L}_\gamma\left(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^k\right) \\ \geq \; & \frac{\gamma}{2}\left\|A_1 x_1^k - A_1 x_1^{k+1}\right\|^2 + \frac{\gamma}{2}\left\|A_2 x_2^k - A_2 x_2^{k+1}\right\|^2 + \frac{\gamma+\sigma}{2}\left\|x_3^k - x_3^{k+1}\right\|^2. \end{aligned} \tag{66}$$

By using (63), we have

$$\begin{aligned} \mathcal{L}_\gamma(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^k) - \mathcal{L}_\gamma(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1}) & = -\frac{1}{\gamma}\left\|\lambda^{k+1} - \lambda^k\right\|^2 \\ & \geq -\frac{L^2}{\gamma}\left\|x_3^k - x_3^{k+1}\right\|^2. \end{aligned} \tag{67}$$

Combining (66) and (67) yields

$$\begin{aligned} & \mathcal{L}_\gamma\left(x_1^k, x_2^k, x_3^k, \lambda^k\right) - \mathcal{L}_\gamma\left(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1}\right) \\ \geq \; & \frac{\gamma}{2}\left\|A_1 x_1^k - A_1 x_1^{k+1}\right\|^2 + \frac{\gamma}{2}\left\|A_2 x_2^k - A_2 x_2^{k+1}\right\|^2 + \left(\frac{\gamma+\sigma}{2} - \frac{L^2}{\gamma}\right)\left\|x_3^k - x_3^{k+1}\right\|^2 \\ \geq \; & M\left(\left\|A_1 x_1^k - A_1 x_1^{k+1}\right\|^2 + \left\|A_2 x_2^k - A_2 x_2^{k+1}\right\|^2 + \left\|x_3^k - x_3^{k+1}\right\|^2\right), \end{aligned} \tag{68}$$

where $M := \min\left\{\frac{\gamma}{2}, \frac{\gamma+\sigma}{2} - \frac{L^2}{\gamma}\right\}$. Since $\gamma \in \left(\frac{\sqrt{\sigma^2+8L^2}-\sigma}{2}, +\infty\right)$, we have $M > 0$.

Then we prove that $\mathcal{L}_\gamma(w^k)$ is uniformly lower bounded. Since f_1, f_2 and f_3 are all lower bounded, we have

$$\begin{aligned}
& \mathcal{L}_\gamma\left(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1}\right) \\
= \;& f_1(x_1^{k+1}) + f_2(x_2^{k+1}) + f_3(x_3^{k+1}) - \left\langle \lambda^{k+1}, \sum_{i=1}^{2} A_i x_i^{k+1} + x_3^{k+1} - b \right\rangle + \frac{\gamma}{2}\left\|\sum_{i=1}^{2} A_i x_i^{k+1} + x_3^{k+1} - b\right\|^2 \\
= \;& f_1(x_1^{k+1}) + f_2(x_2^{k+1}) + f_3(x_3^{k+1}) - \left\langle \nabla f_3(x_3^{k+1}), \sum_{i=1}^{2} A_i x_i^{k+1} + x_3^{k+1} - b \right\rangle + \frac{\gamma}{2}\left\|\sum_{i=1}^{2} A_i x_i^{k+1} + x_3^{k+1} - b\right\|^2 \\
\geq \;& f_1(x_1^{k+1}) + f_2(x_2^{k+1}) + f_3\left(b - \sum_{i=1}^{2} A_i x_i^{k+1}\right) + \frac{\gamma - L}{2}\left\|\sum_{i=1}^{2} A_i x_i^{k+1} + x_3^{k+1} - b\right\|^2 \\
> \;& f_1^* + f_2^* + f_3^* := L^*, \qquad (69)
\end{aligned}$$

where the first inequality holds from the convexity of f_3 and the Lipschitz continuity of ∇f_3. By combining (68) and (69), for any integer $K > 0$ we have

$$\begin{aligned}
& \sum_{k=0}^{K}\left(\left\|A_1 x_1^k - A_1 x_1^{k+1}\right\|^2 + \left\|A_2 x_2^k - A_2 x_2^{k+1}\right\|^2 + \left\|x_3^k - x_3^{k+1}\right\|^2\right) \\
\leq \;& \frac{1}{M}\sum_{k=0}^{K}\left(\mathcal{L}_\gamma(x_1^k, x_2^k, x_3^k, \lambda^k) - \mathcal{L}_\gamma(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1})\right) \\
= \;& \frac{1}{M}\left(\mathcal{L}_\gamma(x_1^0, x_2^0, x_3^0, \lambda^0) - \mathcal{L}_\gamma(x_1^{K+1}, x_2^{K+1}, x_3^{K+1}, \lambda^{K+1})\right) \\
\leq \;& \frac{1}{M}\left(\mathcal{L}_\gamma(x_1^0, x_2^0, x_3^0, \lambda^0) - L^*\right).
\end{aligned}$$

Letting $K \to +\infty$ yields

$$\sum_{k=0}^{\infty}\left(\left\|A_1 x_1^k - A_1 x_1^{k+1}\right\|^2 + \left\|A_2 x_2^k - A_2 x_2^{k+1}\right\|^2 + \left\|x_3^k - x_3^{k+1}\right\|^2\right) < +\infty,$$

which combining with (63) yields (59).

Note that (68) shows that $\mathcal{L}_\gamma(x_1^k, x_2^k, x_3^k; \lambda^k)$ is monotonically non-increasing. This together with (69) shows that $\mathcal{L}_\gamma(x_1^k, x_2^k, x_3^k; \lambda^k)$ converges to $\mathcal{L}_\gamma(x_1^*, x_2^*, x_3^*; \lambda^*)$. Finally, we prove that $\{\left(x_1^k, x_2^k, x_3^k, \lambda^k\right)\}$ is a bounded sequence. Note that (69) and the coerciveness of $f_1 + \mathbf{1}_{\mathcal{X}_1}$ and $f_2 + \mathbf{1}_{\mathcal{X}_2}$ imply that $\left\{\left(x_1^k, x_2^k\right) : k = 0, 1, 2, \ldots\right\}$ is a bounded sequence. This together with the updating formula of λ^{k+1} and (59) yields the boundedness of x_3^k. Moreover, this combining with (58) gives the boundedness of λ^k. Hence, $\left\{\left(x_1^k, x_2^k, x_3^k, \lambda^k\right) : k = 0, 1, 2, \ldots\right\}$ is a bounded sequence.

Now we prove part 2 of Lemma 4.2. We first assume that $\gamma \in \left(\sqrt{\sigma^2 + \frac{2L^2}{\eta_1 - 2}} - \sigma, \frac{4\sigma}{\eta_1}\right]$ for some

$\eta_1 > 2$ such that $\sqrt{\sigma^2 + \frac{2L^2}{\eta_1 - 2}} - \sigma < \frac{4\sigma}{\eta_1}$. Combining (56)-(58) with (13) yields

$$\begin{aligned}
& \frac{1}{\gamma}\left(\lambda^k - \lambda^{k+1}\right)^\top \left(\lambda^{k+1} - \lambda^*\right) - \gamma\left(A_1 x_1^{k+1} - A_1 x_1^*\right)^\top \left((A_2 x_2^k - A_2 x_2^{k+1}) + (x_3^k - x_3^{k+1})\right) \\
& -\gamma\left(A_2 x_2^{k+1} - A_2 x_2^*\right)^\top \left(x_3^k - x_3^{k+1}\right) \\
\geq\ & \sigma\left\|x_3^{k+1} - x_3^*\right\|^2,
\end{aligned}$$

which can be reduced to

$$\begin{aligned}
& \frac{1}{\gamma}\left(\lambda^k - \lambda^{k+1}\right)^\top \left(\lambda^{k+1} - \lambda^*\right) - \left(\lambda^k - \lambda^{k+1}\right)^\top \left((A_2 x_2^k - A_2 x_2^{k+1}) + (x_3^k - x_3^{k+1})\right) \\
& +\gamma\left(A_2 x_2^{k+1} - A_2 x_2^*\right)^\top \left(A_2 x_2^k - A_2 x_2^{k+1}\right) + \gamma\left(x_3^{k+1} - x_3^*\right)^\top \left((A_2 x_2^k - A_2 x_2^{k+1}) + (x_3^k - x_3^{k+1})\right) \\
\geq\ & \sigma\left\|x_3^{k+1} - x_3^*\right\|^2.
\end{aligned}$$

Combining this with (65) yields

$$\begin{aligned}
& \frac{1}{\gamma}\left(\lambda^k - \lambda^{k+1}\right)^\top \left(\lambda^{k+1} - \lambda^*\right) + \gamma\left(A_2 x_2^{k+1} - A_2 x_2^*\right)^\top \left(A_2 x_2^k - A_2 x_2^{k+1}\right) \\
& +\gamma\left(x_3^{k+1} - x_3^*\right)^\top \left(x_3^k - x_3^{k+1}\right) \\
\geq\ & \sigma\left\|x_3^{k+1} - x_3^*\right\|^2 + \sigma\left\|x_3^k - x_3^{k+1}\right\|^2 + \left(\lambda^k - \lambda^{k+1}\right)^\top \left(A_2 x_2^k - A_2 x_2^{k+1}\right) \\
& -\gamma\left(x_3^{k+1} - x_3^*\right)^\top \left(A_2 x_2^k - A_2 x_2^{k+1}\right). \qquad (70)
\end{aligned}$$

Now by applying (15) to the three terms on the left hand side of (70) we get

$$\begin{aligned}
& \left[\frac{1}{2\gamma}\left\|\lambda^k - \lambda^*\right\|^2 + \frac{\gamma}{2}\left\|A_2 x_2^k - A_2 x_2^*\right\|^2 + \frac{\gamma}{2}\left\|x_3^k - x_3^*\right\|^2\right] \\
& -\left[\frac{1}{2\gamma}\left\|\lambda^{k+1} - \lambda^*\right\|^2 + \frac{\gamma}{2}\left\|A_2 x_2^{k+1} - A_2 x_2^*\right\|^2 + \frac{\gamma}{2}\left\|x_3^{k+1} - x_3^*\right\|^2\right] \\
\geq\ & \sigma\left\|x_3^{k+1} - x_3^*\right\|^2 + \sigma\left\|x_3^{k+1} - x_3^k\right\|^2 + \frac{1}{2\gamma}\left\|\lambda^{k+1} - \lambda^k\right\|^2 + \frac{\gamma}{2}\left\|A_2 x_2^{k+1} - A_2 x_2^k\right\|^2 + \frac{\gamma}{2}\left\|x_3^{k+1} - x_3^k\right\|^2 \\
& +\left(\lambda^k - \lambda^{k+1}\right)^\top \left(A_2 x_2^k - A_2 x_2^{k+1}\right) - \gamma\left(x_3^{k+1} - x_3^*\right)^\top \left(A_2 x_2^k - A_2 x_2^{k+1}\right). \qquad (71)
\end{aligned}$$

For any given $\eta_1 > 2$, we have

$$-\gamma\left(x_3^{k+1} - x_3^*\right)^\top \left(A_2 x_2^k - A_2 x_2^{k+1}\right) \geq -\frac{\gamma\eta_1}{4}\left\|x_3^{k+1} - x_3^*\right\|^2 - \frac{\gamma}{\eta_1}\left\|A_2 x_2^{k+1} - A_2 x_2^k\right\|^2, \qquad (72)$$

and

$$\begin{aligned}
& \frac{\eta_1}{2\gamma(\eta_1 - 2)}\left\|\lambda^{k+1} - \lambda^k\right\|^2 + \left(\lambda^{k+1} - \lambda^k\right)^\top \left(A_2 x_2^{k+1} - A_2 x_2^k\right) + \frac{\gamma(\eta_1 - 2)}{2\eta_1}\left\|A_2 x_2^{k+1} - A_2 x_2^k\right\|^2 \\
=\ & \left\|\sqrt{\frac{\eta_1}{2\gamma(\eta_1 - 2)}}\left(\lambda^{k+1} - \lambda^k\right) + \sqrt{\frac{\gamma(\eta_1 - 2)}{2\eta_1}}\left(A_2 x_2^{k+1} - A_2 x_2^k\right)\right\|^2. \qquad (73)
\end{aligned}$$

By combining (63), (72), (73) and (71), we get

$$\begin{aligned}
& \left[\frac{1}{2\gamma}\left\|\lambda^k-\lambda^*\right\|^2+\frac{\gamma}{2}\left\|A_2x_2^k-A_2x_2^*\right\|^2+\frac{\gamma}{2}\left\|x_3^k-x_3^*\right\|^2\right] \\
& -\left[\frac{1}{2\gamma}\left\|\lambda^{k+1}-\lambda^*\right\|^2+\frac{\gamma}{2}\left\|A_2x_2^{k+1}-A_2x_2^*\right\|^2+\frac{\gamma}{2}\left\|x_3^{k+1}-x_3^*\right\|^2\right] \\
\geq\ & \left(\sigma-\frac{\gamma\eta_1}{4}\right)\left\|x_3^{k+1}-x_3^*\right\|^2+\left(\sigma+\frac{\gamma}{2}-\frac{L^2}{\gamma(\eta_1-2)}\right)\left\|x_3^{k+1}-x_3^k\right\|^2 \\
& +\left\|\sqrt{\frac{\eta_1}{2\gamma(\eta_1-2)}}\left(\lambda^{k+1}-\lambda^k\right)+\sqrt{\frac{\gamma(\eta_1-2)}{2\eta_1}}\left(A_2x_2^{k+1}-A_2x_2^k\right)\right\|^2 \\
\geq\ & 0, \qquad (74)
\end{aligned}$$

where the second inequality holds because $\gamma\in\left(\sqrt{\sigma^2+\frac{2L^2}{\eta_1-2}}-\sigma,\frac{4\sigma}{\eta_1}\right]$ implies that

$$\sigma-\frac{\eta_1\gamma}{4}\geq 0,\quad \sigma+\frac{\gamma}{2}-\frac{L^2}{\gamma(\eta_1-2)}>0.$$

Furthermore, (74) implies $\|x_3^{k+1}-x_3^k\|\to 0$ and hence $\|\lambda^{k+1}-\lambda^k\|\to 0$ and $\|A_2x_2^{k+1}-A_2x_2^k\|\to 0$ since

$$\left\|\sqrt{\frac{\eta_1}{2\gamma(\eta_1-2)}}\left(\lambda^{k+1}-\lambda^k\right)+\sqrt{\frac{\gamma(\eta_1-2)}{2\eta_1}}\left(A_2x_2^{k+1}-A_2x_2^k\right)\right\|\to 0.$$

Moreover, the sequence $\frac{1}{2\gamma}\left\|\lambda^k-\lambda^*\right\|^2+\frac{\gamma}{2}\left\|A_2x_2^k-A_2x_2^*\right\|^2+\frac{\gamma}{2}\left\|x_3^k-x_3^*\right\|^2$ is non-increasing, and this implies that $\left\{\left(A_2x_2^k,x_3^k,\lambda^k\right):k=0,1,2,\ldots\right\}$ is bounded. Since A_1 and A_2 both have full column rank, we conclude that $\left\{\left(x_1^k,x_2^k,x_3^k,\lambda^k\right):k=0,1,2,\ldots\right\}$ is a bounded sequence.

Now we assume $\gamma\in\left(0,\min\left\{\frac{4\sigma}{\eta_2},\frac{\sigma(\eta_2-2)}{4\eta_2}+\sqrt{\frac{\sigma^2(\eta_2-2)^2}{16\eta_2^2}+\frac{\sigma^2(\eta_2-2)}{4\eta_2}}\right\}\right)$ for arbitrarily chosen $\eta_2>2$. Using similar arguments as in the case $\gamma\in\left(\sqrt{\sigma^2+\frac{2L^2}{\eta_1-2}}-\sigma,\frac{4\sigma}{\eta_1}\right]$, the following inequalities hold for any given $\eta_2>2$ and $\epsilon>\frac{2\eta_2}{\eta_2-2}$:

$$-\gamma\left(x_3^{k+1}-x_3^*\right)^\top\left(A_2x_2^k-A_2x_2^{k+1}\right)\geq-\frac{\gamma\eta_2}{4}\left\|x_3^{k+1}-x_3^*\right\|^2-\frac{\gamma}{\eta_2}\left\|A_2x_2^k-A_2x_2^{k+1}\right\|^2, \qquad (75)$$

and

$$\begin{aligned}
& \left(\lambda^k-\lambda^{k+1}\right)\left(A_2x_2^k-A_2x_2^{k+1}\right)^\top \\
\geq\ & -\frac{\gamma}{\epsilon}\left\|A_2x_2^{k+1}-A_2x_2^k\right\|^2-\frac{\gamma\epsilon}{2}\left\|x_3^k-x_3^{k-1}\right\|^2-\frac{\gamma\epsilon}{2}\left\|x_3^{k+1}-x_3^k\right\|^2. \qquad (76)
\end{aligned}$$

It follows from (48) and (58) that

$$\left\|\lambda^{k+1}-\lambda^*\right\|\geq\sigma\left\|x_3^{k+1}-x_3^*\right\|. \qquad (77)$$

Therefore, we conclude from (75)-(77) and (71) that

$$\begin{aligned}
& \left[\frac{1}{2\gamma}\left\|\lambda^k-\lambda^*\right\|^2+\frac{\gamma}{2}\left\|A_2x_2^k-A_2x_2^*\right\|^2+\frac{\gamma}{2}\left\|x_3^k-x_3^*\right\|^2+\frac{\gamma\epsilon}{2}\left\|x_3^k-x_3^{k-1}\right\|^2\right] \\
& -\left[\frac{1}{2\gamma}\left\|\lambda^{k+1}-\lambda^*\right\|^2+\frac{\gamma}{2}\left\|A_2x_2^{k+1}-A_2x_2^*\right\|^2+\frac{\gamma}{2}\left\|x_3^{k+1}-x_3^*\right\|^2+\frac{\gamma\epsilon}{2}\left\|x_3^{k+1}-x_3^k\right\|^2\right] \\
\geq\ & \left(\sigma-\frac{\eta_2\gamma}{4}\right)\left\|x_3^{k+1}-x_3^*\right\|^2+\left(\sigma+\frac{\sigma^2}{2\gamma}-\gamma\epsilon\right)\left\|x_3^{k+1}-x_3^k\right\|^2+\left(\frac{\gamma}{2}-\frac{\gamma}{\eta_2}-\frac{\gamma}{\epsilon}\right)\left\|A_2x_2^{k+1}-A_2x_2^k\right\|^2 \\
\geq\ & \left(\sigma+\frac{\sigma^2}{2\gamma}-\gamma\epsilon\right)\left\|x_3^{k+1}-x_3^k\right\|^2+\left(\frac{\gamma}{2}-\frac{\gamma}{\eta_2}-\frac{\gamma}{\epsilon}\right)\left\|A_2x_2^{k+1}-A_2x_2^k\right\|^2 \\
\geq\ & 0,
\end{aligned}$$

where the second and third inequalities hold because $\gamma \in \left(0, \min\left\{\frac{4\sigma}{\eta_2}, \frac{\sigma(\eta_2-2)}{4\eta_2}+\sqrt{\frac{\sigma^2(\eta_2-2)^2}{16\eta_2^2}+\frac{\sigma^2(\eta_2-2)}{4\eta_2}}\right\}\right)$ for any $\eta_2 > 2$ implies

$$0<\gamma\leq\frac{4\sigma}{\eta_2},\quad \frac{\gamma}{2}-\frac{\gamma}{\eta_2}-\frac{\gamma}{\epsilon}>0,\quad \sigma+\frac{\sigma^2}{2\gamma}-\gamma\epsilon>0.$$

This implies $\|x_3^{k+1}-x_3^k\| \to 0$, $\|A_2x_2^{k+1}-A_2x_2^k\| \to 0$, and hence $\|\lambda^{k+1}-\lambda^k\| \to 0$. This also implies the sequence $\frac{1}{2\gamma}\left\|\lambda^k-\lambda^*\right\|^2+\frac{\gamma}{2}\left\|A_2x_2^k-A_2x_2^*\right\|^2+\frac{\gamma}{2}\left\|x_3^k-x_3^*\right\|^2+\frac{\gamma\epsilon}{2}\left\|x_3^k-x_3^{k-1}\right\|^2$ is non-increasing, which further implies that $\left\{\left(A_2x_2^k, x_3^k, \lambda^k\right) : k=0,1,2,\ldots\right\}$ is bounded. Since A_1 and A_2 both have full column rank, we conclude that $\left\{\left(x_1^k, x_2^k, x_3^k, \lambda^k\right) : k=0,1,2,\ldots\right\}$ is a bounded sequence.

Finally, using similar arguments as in Theorem 3.2 it is easy to prove that the whole sequence of $\left\{\left(x_1^k, x_2^k, x_3^k, \lambda^k\right) : k=0,1,2,\ldots\right\}$ converges to $(x_1^*, x_2^*, x_3^*, \lambda^*)$. We omit the details here for succinctness. □