SIAM J. NUMER. ANAL.
Vol. 56, No. 4, pp. 2095–2123

PARTIAL ERROR BOUND CONDITIONS AND THE LINEAR CONVERGENCE RATE OF THE ALTERNATING DIRECTION METHOD OF MULTIPLIERS*

YONGCHAO LIU[†], XIAOMING YUAN[‡], SHANGZHI ZENG[‡], AND JIN ZHANG[§]

Abstract. In the literature, error bound conditions have been widely used to study the linear convergence rates of various first-order algorithms. Most of the literature focuses on how to ensure these error bound conditions, usually posing numerous assumptions or special structures on the model under discussion. In this paper, we focus on the alternating direction method of multipliers (ADMM) and show that the known error bound conditions for studying the ADMM's linear convergence rate can indeed be further weakened if the error bound is studied over the specific iterative sequence it generates. An error bound condition based on ADMM's iterations is thus proposed, and linear convergence under this condition is proved. Furthermore, taking advantage of a specific feature of ADMM's iterative scheme by which part of the perturbation is automatically zero, we propose the so-called partial error bound condition, which is weaker than known error bound conditions in the literature, and we derive the linear convergence rate of ADMM. We further show that this partial error bound condition is useful for interpreting the difference if the two primal variables are updated in different orders when implementing the ADMM. This has been empirically observed in the literature, yet no theory is known.

Key words. convex programming, alternating direction method of multipliers, calmness, partial error bound, linear convergence rate

AMS subject classifications. 90C25, 90C33, 90C22

DOI. 10.1137/17M1144623

1. Introduction.

In this paper we study error bound conditions to ensure the linear convergence rate of the alternating direction method of multipliers (ADMM). The ADMM was originally proposed [5, 19] for solving some nonlinear elliptic equations, and it has recently found a broad spectrum of applications in various fields. To recall the ADMM, we focus on the convex minimization model with linear constraints and an objective function that is the sum of two functions without coupled variables:

$$\begin{aligned} \min_{x\in\mathcal{X},y\in\mathcal{Y}} \quad & f(x)+g(y) \\ \text{s.t.} \quad & Ax+By=b, \end{aligned} \tag{1.1}$$

where $f:\mathbb{R}^{n_1}\to\mathbb{R}$ and $g:\mathbb{R}^{n_2}\to\mathbb{R}$ are both convex (not necessarily smooth) functions; $A\in\mathbb{R}^{m\times n_1}$ and $B\in\mathbb{R}^{m\times n_2}$ are given matrices; $\mathcal{X}\subset\mathbb{R}^{n_1}$ and $\mathcal{Y}\subset\mathbb{R}^{n_2}$ are

*Received by the editors August 23, 2017; accepted for publication (in revised form) April 20, 2018; published electronically July 17, 2018.
http://www.siam.org/journals/sinum/56-4/M114462.html
Funding: The work of the first author was supported by NSFC 11571056. The work of the second author was supported by the General Research Fund from the Hong Kong Research Grants Council, 12302318. The work of the fourth author was supported by NSFC 11601458.
[†]School of Mathematic Science, Dalian University of Technology, Dalian, Liaoning 116024, China (lyc@dlut.edu.cn).
[‡]Department of Mathematics, The University of Hong Kong, Hong Kong, China (xmyuan@hku.hk, zengsz@connect.hku.hk).
[§]Corresponding author. Department of Mathematics, Hong Kong Baptist University, Hong Kong, China, and HKBU Institute of Research and Continuing Education, Shenzhen, China (zhangjin@hkbu.edu.hk).

convex sets; and $b \in \mathbb{R}^m$. The iterative scheme of ADMM for (1.1) reads as

(1.2)
$$\begin{cases} x^{k+1} = \arg\min_{x\in\mathcal{X}} \left\{ f(x) - (\lambda^k)^T(Ax + By^k - b) + \frac{\beta}{2}\|Ax + By^k - b\|^2 \right\}, \\ y^{k+1} = \arg\min_{y\in\mathcal{Y}} \left\{ g(y) - (\lambda^k)^T(Ax^{k+1} + By - b) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \right\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), \end{cases}$$

where λ is the Lagrange multiplier and $\beta > 0$ is a penalty parameter. The subproblems arising in ADMM's iterations may be much easier to solve than the original problem (1.1), and indeed they may have closed-form solutions when f and g are special enough. This feature makes the implementation of ADMM extremely easy for some applications arising in areas such as compressive sensing, image processing, statistical learning, sparse and low-rank optimization problems, etc., and it explains the popularity of ADMM in various areas. We refer to [4, 13, 18] for some review papers on ADMM.

Under some mild conditions such as the nonemptiness of the solution set of problem (1.1), the convergence of ADMM has been well studied in the literature, e.g., [11, 12, 16, 17, 19, 20, 30, 38]. Because of applications recently found in various areas, research on the convergence analysis of the ADMM has regained attention, and more efforts have been devoted to convergence-rate analysis. In [32, 33, 41], the worst-case $O(1/k)$ convergence rate measured by the iteration complexity has been established for ADMM in both the ergodic and nonergodic senses, where k is the iteration counter. Such a convergence rate is sublinear. Consequently, some results for the linear convergence rate of ADMM have been established either for special cases of the generic model (1.1) or for scenarios where more assumptions are imposed on the model (1.1). For example, it is shown in [3, Theorem 6.4] that the local linear convergence rate of ADMM can be guaranteed for special linear and quadratic cases of (1.1) if it is assumed that both minimization subproblems in (1.2) have unique optimal solutions, and if some strict complementarity conditions hold. Moreover, if f and/or g is strongly convex, and one of them is differentiable with a Lipschitz continuous gradient, and if the generated iterative sequence is assumed to be bounded, together with some full-rank conditions of the coefficient matrices, then the global linear convergence rate of ADMM is proved in [8, 42].

Error bound conditions turn out to play an important role in studying the linear convergence rate of the ADMM. To elucidate error bound conditions, we first mention the Karush–Kuhn–Tucker (KKT) system of the problem (1.1):

$$\begin{cases} 0 \in \partial f(x) - A^T\lambda + \mathcal{N}_{\mathcal{X}}(x), \\ 0 \in \partial g(y) - B^T\lambda + \mathcal{N}_{\mathcal{Y}}(y), \\ 0 = Ax + By - b, \end{cases} \tag{1.3}$$

where $\partial f(x)$ denotes the subgradient of convex function f at x and $\mathcal{N}_{\mathcal{C}}(c) := \{\xi : \langle \xi, \zeta - c\rangle \le 0 \ \forall \zeta \in \mathcal{C}\}$ denotes the normal cone at c to a given convex set $\mathcal{C}$. Let S^* be the solution set of the KKT system (1.3) and assume it is nonempty. Furthermore, let $r : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m \to \mathbb{R}_+$ be a residual error function satisfying $r(x, y, \lambda) = 0$ iff $(x, y, \lambda) \in S^*$. We say that the KKT system (1.3) admits a local error bound around a given point $(x^*, y^*, \lambda^*) \in S^*$ with the residual error function $r(x, y, \lambda)$ if there exists a neighborhood

$$\mathcal{B}_\epsilon(x^*, y^*, \lambda^*) := \{(x, y, \lambda) : \|(x, y, \lambda) - (x^*, y^*, \lambda^*)\| \le \epsilon\}$$

of the point (x^*, y^*, λ^*) and a constant $\kappa > 0$ such that

$$\text{(1.4)} \qquad [\text{EB}^{\text{r}}] \qquad dist((x,y,\lambda), S^*) \leq \kappa \cdot r(x,y,\lambda) \quad \forall (x,y,\lambda) \in \mathcal{B}_\epsilon(x^*, y^*, \lambda^*).$$

Throughout, we define $dist(c, \mathcal{C}) := \inf\{\|c - c'\| \mid c' \in \mathcal{C}\}$ for a given subset $\mathcal{C}$ and vector c in the same space, and $\|\cdot\|$ is the 2-norm unless otherwise specified. If this estimate is valid for every $(x,y,\lambda) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m$, rather than merely $(x,y,\lambda) \in \mathcal{B}_\epsilon(x^*, y^*, \lambda^*)$, we say that the KKT system (1.3) admits a global error bound.

Error bound conditions of the KKT system (1.3) with various choices of the residual error function $r(x,y,\lambda)$ have inspired some work that has studied the linear convergence rate of the ADMM (see, e.g., [28, 29, 54]). According to (1.3), it is natural to define a mapping $\phi : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m \rightrightarrows \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m$ as

$$\text{(1.5)} \qquad \phi(x,y,\lambda) = \begin{pmatrix} \partial f(x) - A^T\lambda + \mathcal{N}_{\mathcal{X}}(x) \\ \partial g(y) - B^T\lambda + \mathcal{N}_{\mathcal{Y}}(y) \\ Ax + By - b \end{pmatrix}$$

and then a residual error function as $r(x,y,\lambda) = dist(0, \phi(x,y,\lambda))$. We call ϕ, as defined by (1.5), the KKT mapping for obvious reasons, and the KKT system (1.3) can be written as $0 \in \phi(x,y,\lambda)$. With $\phi(x,y,\lambda)$ given in (1.5), let us define $S : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m \rightrightarrows \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m$ as

$$\text{(1.6)} \qquad S(p) := \{(x,y,\lambda) \mid p \in \phi(x,y,\lambda)\}$$

with $p = (p_1, p_2, p_3) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m$. Obviously, $S(0) = S^*$. Recall that we are interested in finding $0 \in \phi(x,y,\lambda)$, i.e., $(x,y,\lambda) \in S(0) = \{(x,y,\lambda) \mid 0 \in \phi(x,y,\lambda)\}$. Hence, p in (1.6) plays the role of a perturbation parameter and (1.6) can be regarded as a perturbed system of the KKT system (1.3). This is also why we purposely use the same letter S to define the mapping in (1.6) in addition to the notation S^* for the solution set of the KKT system (1.3).

Let us use the notation $w = (x,y,\lambda)$ for a more compact presentation. Now, using $dist(0, \phi(w))$ as the residual error function, the KKT system (1.3) is said to admit a local error bound around a feasible point $\bar{w} = (\bar{x}, \bar{y}, \bar{\lambda})$ if there exists a neighborhood $\mathcal{B}_\epsilon(\bar{w})$ of $\bar{w}$ and some constant $\kappa > 0$ such that

$$\text{(1.7)} \qquad [\text{FEB}] \qquad dist(w, S(0)) \leq \kappa \cdot dist(0, \phi(w)) \qquad \forall w \in \mathcal{B}_\epsilon(\bar{w}).$$

Indeed, in terms of variational analysis, the existence of an error bound around the reference point $\bar{w}$ with the residual error function $r(w) = dist(0, \phi(w))$ is exactly the metric subregularity of the KKT mapping $\phi(w)$ at $(\bar{w}, 0)$. The set-valued mapping $\phi(w)$ is called metrically subregular around $(\bar{w}, 0)$ if there exists a neighborhood $\mathcal{B}_\epsilon(\bar{w})$ of $\bar{w}$ and $\kappa \geq 0$ such that

$$dist\left(w, \phi^{-1}(0)\right) \leq \kappa \cdot dist\left(0, \phi(w)\right) \quad \forall w \in \mathcal{B}_\epsilon(\bar{w}).$$

Equivalently, $\phi(w)$ is metrically subregular around $(\bar{w}, 0)$ if there exists a neighborhood $\mathcal{B}_\epsilon(\bar{w})$ of $\bar{w}$ and $\kappa \geq 0$ such that

$$\text{(1.8)} \qquad S(p) \cap \mathcal{B}_\epsilon(\bar{w}) \subset S(0) + \kappa \|p\| \cdot \mathcal{B}_1(0) \quad \forall p,$$

i.e., the set-valued mapping $S(p)$ is calm around $(0, \bar{w})$. We refer to [9, 46] for more details on the concepts of metric subregularity, calmness, and their relationship. Note that calmness was first introduced as the pseudo-upper-Lipschitz continuity in [55], taking into account that it is weaker than both the upper-Lipschitz continuity of Robinson [43, 44] and the pseudo-Lipschitz continuity of Aubin [2]. The set-valued mapping $S(p)$ is called upper-Lipschitz continuous around $(0,\bar{w})$ if there exists a neighborhood $\mathcal{B}_\epsilon(0)$ of 0 and $\kappa \geq 0$ such that

$$S(p) \subset S(p') + \kappa \|p - p'\| \cdot \mathcal{B}_1(0) \quad \forall p, p' \in \mathcal{B}_\epsilon(0) \tag{1.9}$$

and is called pseudo-Lipschitz continuous around $(0,\bar{w})$ if there exist neighborhoods $\mathcal{B}_\epsilon(\bar{w})$ of $\bar{w}$ and $\mathcal{B}_\epsilon(0)$ of 0, and $\kappa \geq 0$ such that

$$S(p) \cap \mathcal{B}_\epsilon(\bar{w}) \subset S(p') + \kappa \|p - p'\| \cdot \mathcal{B}_1(0) \quad \forall p, p' \in \mathcal{B}_\epsilon(0). \tag{1.10}$$

Moreover, $S(p)$ in (1.6) considers the canonical perturbation p of $\phi(w)$. From now on, we call (1.7) a full error bound (FEB) condition since p fully perturbs ϕ in (1.6).

In the literature, other error bound conditions have been defined as well for studying the linear convergence rate of the ADMM and its variants. For instance, based on the so-called natural map (see [15, p. 83]) in terms of the Moreau–Yosida proximal mapping, the following mapping is used in [29]:

$$R_1(w) = \begin{pmatrix} x - \mathrm{Prox}_{f+\delta_\mathcal{X}}(x + A^T\lambda) \\ y - \mathrm{Prox}_{g+\delta_\mathcal{Y}}(y + B^T\lambda) \\ Ax + By - b \end{pmatrix}, \tag{1.11}$$

where δ is the indicator function of a convex set and Prox_h is the proximal mapping associated with the function h, i.e.,

$$\mathrm{Prox}_h(a) := \arg\min_{t\in\mathbb{R}^n} \left\{ h(t) + \frac{1}{2}\|t - a\|^2 \right\}.$$

The mapping defined by (1.11) is also called the proximal KKT mapping. Then a residual error function is defined as $r(w) = dist(0, R_1(w))$ in [29]. Accordingly, the KKT system (1.3) is said to admit a local proximal error bound around $\bar{w}$ if there exists a neighborhood $\mathcal{B}_\epsilon(\bar{w})$ of $\bar{w}$ and some $\kappa > 0$ such that

$$[\text{Proximal EB} - \text{I}] \quad dist(w, S^*) \leq \kappa \cdot \|R_1(w)\| \qquad \forall w \in \mathcal{B}_\epsilon(\bar{w}). \tag{1.12}$$

Note that (1.12) coincides with the metric subregularity of $R_1(w)$ at $(\bar{w}, 0)$. Under the proximal error bound condition (1.12), the linear convergence rate of the ADMM (1.2) (and its variant with a relaxation factor) is obtained in [29] for the special case of the problem (1.1) where the objective function is quadratic. The conditions used in [3], such as the uniqueness of optimal solutions to the subproblems, and strict complementarity are not needed by the analysis in [29].

Later, an alternative form of (1.11) is considered in [54]:

$$R_2(w) = \begin{pmatrix} x - \mathrm{Proj}_\mathcal{X}(x - \partial f(x) + A^T\lambda) \\ y - \mathrm{Proj}_\mathcal{Y}(y - \partial g(y) + B^T\lambda) \\ Ax + By - b \end{pmatrix}, \tag{1.13}$$

where $\mathrm{Proj}_\mathcal{C}(\zeta) := \arg\min_{\xi\in\mathcal{C}} \{\|\xi - \zeta\|\}$ is the canonical projection operator onto a given convex set $\mathcal{C}$. Accordingly, the residual error function is defined as $r(w) = dist(0, R_2(w))$, and the KKT system (1.3) is said to admit a local error bound around $\bar{w}$ if there exists a neighborhood $\mathcal{B}_\epsilon(\bar{w})$ of $\bar{w}$ and some $\kappa > 0$ such that

$$[\text{Proximal EB} - \text{II}] \quad dist(w, S^*) \leq \kappa \cdot dist(0, R_2(w)) \quad \forall w \in \mathcal{B}_\epsilon(\bar{w}). \tag{1.14}$$

Under the error bound condition (1.14), which also reads as the metric subregularity of $R_2(w)$ at $(\bar{w}, 0)$, the linear convergence rate of the ADMM (1.2) and its linearized variant is established in [54] for the special case of (1.1) where ∂f and ∂g are both polyhedral multifunctions. Recall that a set-valued mapping is called a polyhedral multifunction if its graph is the union of finitely many convex polyhedra. Note that the projection operator onto a closed convex set $\mathcal{C}$ can be regarded as the proximal operator associated with the indicator function over $\mathcal{C}$. We also call (1.14) a local proximal error bound of the KKT system (1.3).

In [1], the linear convergence of the Douglas–Rachford splitting method is studied under the metric subregularity assumption of the involved operator, and this result includes the linear convergence rate of the ADMM (1.2) for the model (1.1) as a special case. As analyzed in [1], when ∂f and ∂g in (1.1) are both polyhedral multifunctions, the required metric subregularity is satisfied and thus the linear convergence of ADMM can be ensured.

As in [32], instead of the original ADMM scheme (1.2), our analysis is for the slightly generalized proximal version of the ADMM (PADMM),

(1.15)
$$\begin{cases} x^{k+1} = \arg\min\limits_{x\in\mathcal{X}} \left\{ f(x) - (\lambda^k)^T(Ax + By^k - b) + \dfrac{\beta}{2}\|Ax + By^k - b\|^2 + \dfrac{1}{2}\|x - x^k\|_D^2 \right\}, \\ y^{k+1} = \arg\min\limits_{y\in\mathcal{Y}} \left\{ g(y) - (\lambda^k)^T(Ax^{k+1} + By - b) + \dfrac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \right\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), \end{cases}$$

where $D \in \mathbb{R}^{n_1\times n_1}$ is a symmetric and positive semidefinite matrix. Here, we slightly abuse the notation $\|x\|_D^2$ for the number $x^T Dx$ even though D may be only positive semidefinite. The penalty parameter β is fixed throughout our discussion. This scheme includes the original ADMM scheme (1.2) and the linearized ADMM (or split inexact Uzawa method in [56]) as special cases with $D = 0$ and $D = (\sigma I_{n_1} - \beta A^T A)$ with $\sigma > \beta\|A^T A\|$, respectively. As analyzed in [47], the linearized ADMM (see, e.g., [37, 51, 53]) is highly relevant to the primal-dual hybrid gradient (PDHG) method which is popularly used in many areas such as machine learning and image processing; see, e.g., [6, 50], to mention a few. We also refer to, e.g., [31, 49], for some revisits to the PDHG method from the perspective of the proximal point algorithm in [40]. Hence, we include this case in our discussion and consider the PADMM (1.15).

On the other hand, for our desired linear convergence of PADMM (1.15), previously introduced error bound conditions (1.7), (1.12), and (1.14) in the mentioned literature are all proposed on the basis of the KKT system (1.3). Generally, they are assumed only dependently on the model (1.1) but irrelevant to any specific algorithm under discussion. We thus call them generic error bound conditions. Obviously, they are somehow too "sufficient" for studying the convergence rate of a specific algorithm. Indeed, most of the efforts, e.g., [3, 28, 29, 54], have been focused on how to ensure these error bound conditions, usually by imposing more assumptions or special structures on the model (1.1), so that the linear convergence rate of ADMM can be guaranteed. In other words, the structures and features of a specific algorithm are ignored when its linear convergence rate is studied via error bound conditions, and thus directly using these generic error bound conditions indeed shrinks the range that validates the linear convergence rate of ADMM.

In this paper, we are more interested in weakening error bound conditions (1.7), (1.12), and (1.14) for the purpose of ensuring the linear convergence rate of the specific

PADMM (1.15). We investigate certain error bound conditions that are specifically based on the iterative scheme (1.15). Our only interest is to estimate the error bound for the specific sequence $\{w^k\}$ generated by (1.15) to the solution set S^* of the KKT system (1.3), rather than the generic error bound conditions in the forms of (1.7), (1.12), or (1.14).

In section 3, we propose an error bound condition based on the specific iterative scheme (1.15) and prove that it suffices to ensure the linear convergence of the PADMM (1.15). We also show that the generic FEB (1.7) is sufficient to ensure this iteration based error bound condition.

In section 4, we clarify the equivalence between the FEB (1.7) and the proximal EB-I (1.12) and proximal EB-II (1.14). Because of the equivalence, theoretically we can choose from (1.7), (1.12), and (1.14). We further explain why we choose the FEB (1.7) to conduct the convergence analysis for the PADMM (1.15) from a perturbation analysis perspective.

With the purpose of studying the error bound condition based on PADMM iteration, we find that a more meticulous analysis for the sequence generated by (1.15) immediately provides an insight and helps to further weaken the mentioned error bound conditions but still ensure the linear convergence rate of the PADMM (1.15). More specifically, for the sequence $\{w^k\}$ generated by (1.15), the second part of the KKT system (1.3), i.e., $0 \in \partial g(y^k) - B^T\lambda^k + \mathcal{N}_{\mathcal{Y}}(y^k)$, holds for all iterations. In the language of perturbation analysis, the sequence $\{w^k\}$ generated by the PADMM (1.15) introduces no perturbation to the part $0 \in \partial g(y) - B^T\lambda + \mathcal{N}_{\mathcal{Y}}(y)$ in (1.6). This interesting observation suggests that there is no need to fully satisfy a general error bound condition that is derived based on the KKT system (1.3), and a partial error bound (PEB) condition without consideration of the perturbation to the part $\partial g(y) - B^T\lambda + \mathcal{N}_{\mathcal{Y}}(y)$ is sufficient to study the linear convergence rate of the PADMM (1.15). We will provide details in section 5. In particular, an example is constructed to illustrate that the PEB condition is indeed weaker than the known full counterparts.

A byproduct of our analysis, which has its own interest, especially in implementation perspectives, is that our theory for deriving a PEB condition based on PADMM iteration can interpret the difference of efficiency caused by changing the order of updating the primal variables x and y in the ADMM (1.2). It has been empirically observed that the convergence speed may change if we swap the order of x and y in the ADMM (1.2), despite there being no difference from the theoretical convergence-proof point of view. No rigorous theory is currently known to explain this difference. We will show by an example that swapping the order of x and y in (1.2) does make a difference in satisfying the PEB condition for the ADMM (1.2). This theoretical justification gives hints to users to decide a more appropriate order of updating the primal variables for a specific application of the problem (1.1) so that the associated PEB can be met more easily, and hence the linear convergence rate of ADMM can be found. We focus on this analysis in section 6.

The discussion starts with some preliminaries in section 2 and ends with some conclusions and possible future works in section 7.

2. Preliminaries.

In this section, we state assumptions under which our further analysis will be conducted, recall the variational inequality characterization of the problem (1.1), and provide some known or obvious convergence results of the PADMM (1.15).

2.1. Assumptions. To characterize the solution set of the problem (1.1) by the first-order optimality conditions, we need certain constraint qualifications such as the strong conical hull intersection property (CHIP) for the sets $\mathcal{X} \times \mathcal{Y}$ and $\mathcal{F}$ defined by

$$\mathcal{F} := \{(x, y) \mid Ax + By = b\}. \tag{2.1}$$

In particular, for any (x, y) feasible for the problem (1.1), there holds

$$\mathcal{N}_{\mathcal{F}\cap\mathcal{X}\times\mathcal{Y}}(x, y) := \mathcal{N}_{\mathcal{F}}(x, y) + \mathcal{N}_{\mathcal{X}}(x) \times \mathcal{N}_{\mathcal{Y}}(y).$$

The strong CHIP plays a similar role as the Abadie constraint qualification, which is regarded as not restrictive. Throughout, to avoid triviality, the following nonemptyness assumption is assumed.

Assumption 2.1. The optimal solution set of problem (1.1) is nonempty.

Under Assumption 2.1 and strong CHIP, $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$ is an optimal solution point of the problem (1.1) iff there exists a Lagrange multiplier $\lambda^* \in \mathbb{R}^m$ such that (x^*, y^*, λ^*) solves the KKT system (1.3).

2.2. Variational inequality characterization of (1.1). As analyzed in [32], the problem (1.1) can be characterized by the variational inequality: finding $w^* = (x^*, y^*, \lambda^*) \in \Omega := \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m$ such that

$$\mathrm{VI}(\Omega, F, \theta): \qquad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0 \quad \forall\, w \in \Omega, \tag{2.2}$$

where

$$u = (x, y), \quad \theta(u) = f(x) + g(y) \ \text{ and } \ F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}. \tag{2.3}$$

Note that the mapping $F(w)$ defined by (2.3) is monotone as it is affine with a skew-symmetric matrix. Since S^* is assumed to be nonempty, the solution set of $\mathrm{VI}(\Omega, F, \theta)$, denoted by Ω^*, is also nonempty.

2.3. Convergence of (1.15). Our main purpose is discussing error bound conditions that can ensure the linear convergence rate of the PADMM (1.15) under the by-default assumption that the convergence of (1.15) is given. As a prerequisite of the analysis to be delineated, the convergence of (1.15) can be easily given by various results in the literature. In this subsection, we briefly mention the convergence of (1.15) and give a particular sufficient condition to ensure it.

With the given model (1.1) and the iterative scheme of the PADMM (1.15), let us define the matrix H and its submatrix H_0 as follows to simplify the notation in our analysis:

$$H = \begin{pmatrix} D & 0 & 0 \\ 0 & \beta B^T B & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix} \quad \text{and} \quad H_0 = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}. \tag{2.4}$$

Moreover, let us make the following assumption.

Assumption 2.2. One of the following conditions is satisfied:

(1) $D \succeq 0$, and both A and B are of full column rank; or

(2) $D \succ 0$, and B is of full column rank.

Obviously, $H \succeq 0$ and $H_0 \succ 0$ for either of the cases in Assumption 2.2. In particular, $H \succ 0$ if Case (2) of Assumption 2.2 holds. Hereafter, we also slightly abuse the notation $\|w\|_H$ for $\sqrt{w^T H w}$ even though H might only be positive semidefinite. Moreover, there exists a constant $L_H > 0$ such that

$$\|w\|_H \le L_H \|w\| \qquad \forall w \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m.$$

To derive the convergence of (1.15), first notice that the iterative scheme (1.15) can be written as

$$\text{(2.5)}\quad \begin{cases} 0 \in \partial f(x^{k+1}) - A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1}) + D(x^{k+1} - x^k) + \mathcal{N}_{\mathcal{X}}(x^{k+1}), \\ 0 \in \partial g(y^{k+1}) - B^T\lambda^{k+1} + \mathcal{N}_{\mathcal{Y}}(y^{k+1}), \\ 0 = Ax^{k+1} + By^{k+1} - b + \dfrac{1}{\beta}(\lambda^{k+1} - \lambda^k). \end{cases}$$

We recall some inequalities established in the literature (see., e.g., [14, 29, 33, 54]) for deriving the convergence of the ADMM (1.2), the PADMM (1.15), and their variants. Some of the proofs are omitted.

Lemma 2.1 (see [33, Lemma 3.1]). *Let $\{w^k = (x^k, y^k, \lambda^k)\}$ be the sequence generated by the PADMM* (1.15)*; then we have*

(2.6)

$$\theta(w) - \theta(w^{k+1}) + \left(w - w^{k+1}\right)^T \left\{F(w) + \eta(y^k, y^{k+1}) + H(w^{k+1} - w^k)\right\} \ge 0 \ \ \forall\, w \in \Omega,$$

where

$$\eta(y^k, y^{k+1}) := \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - y^{k+1}).$$

The next proposition gives some important inequalities for the sequence $\{w^k\}$ generated by the PADMM (1.15).

Proposition 2.2 (see [33, Theorem 4.1]). *Let $\{w^k = (x^k, y^k, \lambda^k)\}$ be the sequence generated by the PADMM* (1.15)*. For any point $w^* = (x^*, y^*, \lambda^*)$ in S^*, we have*

$$\text{(2.7)}\qquad \|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 - \|w^{k+1} - w^k\|_H^2,$$

and consequently it holds that

$$\text{(2.8)}\qquad \sum_{k=0}^{\infty} \|w^{k+1} - w^k\|_H^2 \le \infty.$$

Then, we show that Assumptions 2.1 and 2.2 and strong CHIP are sufficient to ensure the convergence of the PADMM (1.15).

Theorem 2.3. *Let $\{w^k\}$ be the sequence generated by the PADMM* (1.15)*. If Assumptions* 2.1 *and* 2.2 *and strong CHIP are all satisfied, then $\{w^k\}$ converges to a solution point $w^* \in S^*$.*

Proof. We first consider case (1) of Assumptions 2.2. For this case, $H \succeq 0$ but both A and B are of full column rank. It follows from (2.7) that the sequence $\{v^k = (y^k, \lambda^k)\}$ is bounded. Moreover, (2.8) in Proposition 2.2 implies that $\|w^{k+1} - w^k\|_H \to$

0 and hence the boundedness of the sequences $\{\frac{1}{\beta}(\lambda^k - \lambda^{k+1})\}$ and $\{B(y^k - y^{k+1})\}$, by the definition of H in (2.4). We thus know the sequence $\{Ax^{k+1} + By^{k+1} - b\}$ is also bounded because of the identity

$$Ax^{k+1} + By^{k+1} - b = \frac{1}{\beta}(\lambda^k - \lambda^{k+1}),$$

which is obvious from the update scheme of the scheme (1.15). Therefore, the boundedness of $\{v^k\}$ ensures that the sequence $\{Ax^k\}$ is bounded. Since A is assumed to be of full column rank, $\{x^k\}$ is bounded. Overall, we prove that the sequence $\{w^k\}$ is bounded. Let $\{w^{k_j}\}$ be a subsequence of $\{w^k\}$ converging to w^*. Then for any fixed $w \in \Omega$, considering the inequality (2.6) for the subsequence $\{w^{k_j}\}$ and taking $j \to \infty$, and using the fact $\|w^{k_j+1} - w^{k_j}\|_H \to 0$ implied by (2.8), we conclude that $w^* \in S^*$. Now we need to prove that $w^k \to w^*$ as $k \to \infty$. It follows from (2.7) that $\|w^k - w^*\|_H \to 0$, which implies that $\|v^k - v^*\| \to 0$ because B is of full column rank and hence $H_0 \succ 0$. We thus have $y^k \to y^*$ and $\lambda^k \to \lambda^*$. Notice that

$$A(x^k - x^*) + B(y^k - y^*) = Ax^k + By^k - b = \frac{1}{\beta}(\lambda^{k+1} - \lambda^k),$$

where the first equality follows from the optimality of (x^*, y^*), and the second equality is a direct consequence of the definition of λ^{k+1} in (1.15). Since $\|w^{k+1} - w^k\|_H \to 0$ implies $\lambda^{k+1} - \lambda^k \to 0$, we have $A(x^k - x^*) + B(y^k - y^*) \to 0$. Because $y^k \to y^*$ and A is of full column rank, we immediately have $x^k \to x^*$, and hence $w^k \to w^*$ as $k \to \infty$.

Now, we consider case (2) of Assumption 2.2. For this case, we have $H \succ 0$. Then, by (2.7), we know that the sequence $\{w^k\}$ is bounded and let $\{w^{k_j}\}$ be a subsequence of $\{w^k\}$ converging to w^*. Similar to the discussion above, for any fixed $w \in \Omega$, considering the inequality (2.6) for the subsequence $\{w^{k_j}\}$, taking the limit over j, and using the fact that $\|w^{k_j+1} - w^{k_j}\|_H \to 0$, we obtain $w^* \in S^*$. Then, using (2.7), we have $\|w^k - w^*\|_H \to 0$. Since $H \succ 0$ for this case, we immediately have $w^k \to w^*$ as $k \to \infty$ and the proof is complete. ∎

Note that Assumptions 2.1 and 2.2, and strong CHIP are sufficient to ensure the convergence of the PADMM (1.15); however, they are not necessary.

3. Iteration-based error bound condition. In this section, with the by-default given convergence of the sequence $\{w^k\}$ generated by the PADMM (1.15) to $w^* \in S^*$, we focus on the discussion of its linear convergence rate. Note that it is not necessary to assume Assumption 2.2 in the analysis.

As mentioned, in the literature, some generic error bound conditions depending only on the model have been studied for the linear convergence rate of the ADMM (1.2) and its variants, and in the literature, it is focused on how to ensure the proximal EB-I and proximal EB-II conditions by posing assumptions or requiring special structures in the model (1.1). These error bound conditions or related studies are usually too restrictive, and they do not take into consideration the specific structures and properties of the algorithm under discussion. Meanwhile, it seems beneficial to estimate the error only for the specific iterative sequence, instead of arbitrary points within a region, when the convergence rate of a particular algorithm is studied. We hence study the linear convergence rate of the PADMM (1.15) under some error bound condition that is based on the specific iterative scheme of (1.15). We shall show that this PADMM-iteration-based consideration can indeed weaken the mentioned generic error bound conditions.

We first make some notation clear. Recall the definition of H in (2.4). We shall use the notation

$$dist_H(w, \mathcal{C}) := \inf\{\|w - w'\|_H \,\big|\, w' \in \mathcal{C}\} \tag{3.1}$$

for a given subset $\mathcal{C}$ and vector w in the same space. As mentioned, $H \succeq 0$ under Assumption 2.2. When $dist_H(\cdot, S^*)$ and $dist(\cdot, S^*)$ are considered, it follows from (3.1) that

$$dist_H(w, S^*) \leq L_H \cdot dist(w, S^*) \qquad \forall w \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m. \tag{3.2}$$

Moreover, notice that the variable x is intermediate and it is not involved in the iteration of the original ADMM (1.2); see, e.g., [4]. When our analysis generally conducted for the PADMM (1.15) is specified for the original ADMM (1.2), i.e., $D = 0$, we also need the notation $v = (y, \lambda)$ to exclude the intermediate variable x and $S_v^* := \{(y^*, \lambda^*) \mid (x^*, y^*, \lambda^*) \in S^* \text{ for some } x^*\}$. Accordingly, H_0 is needed to present the analysis for (1.2) compactly and instead of $dist_H(w, S^*)$, we use

$$dist_{H_0}(v, S_v^*) := \inf\{\|v - v'\|_{H_0} \,\big|\, v' \in S_v^*\} \tag{3.3}$$

when the original ADMM (1.2) is considered in our analysis. Also, we use the notation

$$S_\lambda^* := \{\lambda^* \mid (x^*, y^*, \lambda^*) \in S^* \text{ for some } (x^*, y^*)\} \tag{3.4}$$

when the convergence of the sequence of Lagrange multiplier $\{\lambda^k\}$ is highlighted.

3.1. PADMM-iteration-based error bound condition for the linear convergence rate. We first present a PADMM-iteration-based error bound condition and show that it suffices to guarantee the linear convergence rate of the sequence generated by the PADMM (1.15). We refer to more literature, e.g., [48, 52], for some preliminary studies of algorithm-based error bound conditions for other algorithms.

DEFINITION 3.1 (PADMM-iteration-based error bound condition). *Let $\{w^k\}$ be the sequence generated by the PADMM* (1.15). *If there exists $\kappa > 0$ and $\epsilon > 0$ such that*

$$dist_H(w^{k+1}, S^*) \leq \kappa \cdot \|w^{k+1} - w^k\|_H \qquad \text{when } w^{k+1} \in \mathcal{B}_\epsilon(w^*), \tag{3.5}$$

then $\{w^k\}$ is said to satisfy a PADMM-iteration-based error bound condition.

With (3.5), it is easy to prove the local linear convergence rate for the PADMM (1.15). We need one more theorem for preparation.

THEOREM 3.2. *Let $\{w^k\}$ be the sequence generated by the PADMM* (1.15) *and it converges to w^*. If Assumption* 2.1 *and strong CHIP are both satisfied, for any $\epsilon > 0$, there exists $\tilde{\epsilon} > 0$ such that*

$$\|w^{k+1} - w^k\|_H < \tilde{\epsilon} \Longrightarrow w^{k+1} \in \mathcal{B}_\epsilon(w^*).$$

Proof. It follows from the convergence of $\{w^k\}$ that, for any $\epsilon > 0$, there exists an integer $K > 0$ such that

$$w^{k+1} \in \mathcal{B}_\epsilon(w^*) \quad \forall\, k \geq K.$$

Taking $\tilde{\epsilon} := \min_{0 \leq k < K}\{\|w^{k+1} - w^k\|_H\} > 0$, we have

$$\|w^{k+1} - w^k\|_H < \tilde{\epsilon} \Longrightarrow k \geq K \Longrightarrow w^{k+1} \in \mathcal{B}_\epsilon(w^*),$$

and the proof is complete. ∎

We first prove a local property for the sequence $\{dist_H^2(w^{k+1}, S^*)\}$.

THEOREM 3.3. *Assume that Assumption* 2.1 *and strong CHIP are both satisfied. If the sequence* $\{w^k\}$ *generated by the PADMM* (1.15) *converges to* w^* *and it satisfies the PADMM-iteration-based error bound condition* (3.5)*, then there exists* $\kappa > 0$ *and* $\epsilon > 0$ *such that*

$$dist_H^2(w^{k+1}, S^*) \leq \left(1 + \frac{1}{\kappa^2}\right)^{-1} \cdot dist_H^2(w^k, S^*) \qquad \text{when } \|w^{k+1} - w^k\|_H < \epsilon.$$

Proof. First, it follows from (2.7) that

$$dist_H^2(w^{k+1}, S^*) \leq dist_H^2(w^k, S^*) - \|w^{k+1} - w^k\|_H^2 \qquad \forall k = 1, 2, \dots.$$

By virtue of Theorem 3.2 and (3.5), there exists $\kappa > 0$ and $\epsilon > 0$ such that

$$dist_H(w^{k+1}, S^*) \leq \kappa \cdot \|w^{k+1} - w^k\|_H^2 \qquad \text{when } \|w^{k+1} - w^k\|_H < \epsilon.$$

Subsequently, we have

$$dist_H^2(w^{k+1}, S^*) \leq dist_H^2(w^k, S^*) - \frac{1}{\kappa^2} dist_H^2(w^{k+1}, S^*) \quad \text{when } \|w^{k+1} - w^k\|_H < \epsilon,$$

and the proof is complete. ∎

Moreover, we observe that the local property of the sequence $\{dist_H^2(w^{k+1}, S^*)\}$ established in Theorem 3.3 is essentially global. Hence, there is no difference in studying the local or global property for the sequence $\{dist_H^2(w^{k+1}, S^*)\}$ under the PADMM-iteration-based error bound condition (3.5). The following theorem is inspired by [15, Proposition 6.1.2].

THEOREM 3.4. *Assume that Assumption* 2.1 *and strong CHIP are both satisfied. If the sequence* $\{w^k\}$ *generated by the PADMM* (1.15) *converges to* w^* *and it satisfies the PADMM-iteration-based error bound condition* (3.5)*, then there exists* $\tilde{\kappa} > 0$ *such that*

$$dist_H^2(w^{k+1}, S^*) \leq \left(1 + \frac{1}{\tilde{\kappa}^2}\right)^{-1} \cdot dist_H^2(w^k, S^*) \qquad \forall\ k \geq 0. \tag{3.6}$$

Proof. According to Theorem 3.3, there exists $\kappa > 0$ and $\epsilon > 0$ such that

$$dist_H(w^{k+1}, S^*) \leq \kappa \cdot \|w^{k+1} - w^k\|_H \qquad \text{when } \|w^{k+1} - w^k\|_H < \epsilon.$$

Thus, we only need to consider indices k such that $\|w^{k+1} - w^k\|_H \geq \epsilon$. According to (2.7), there is a constant $M > 0$ such that $\|w^k - w^*\|_H \leq M\ \forall\ k \geq 0$. We immediately have

$$dist_H(w^{k+1}, S^*) \leq \|w^{k+1} - w^*\|_H \leq M/\epsilon \cdot \|w^{k+1} - w^k\|_H \quad \text{when } \|w^{k+1} - w^k\|_H \geq \epsilon.$$

Letting $\tilde{\kappa} := \max\{\kappa, M/\epsilon\}$, we obtain the result

$$dist_H(w^{k+1}, S^*) \leq \tilde{\kappa} \cdot \|w^{k+1} - w^k\|_H \qquad \forall\ k \geq 0.$$

Together with (2.7), we have

$$dist_H^2(w^{k+1}, S^*) \leq \left(1 + \frac{1}{\tilde{\kappa}^2}\right)^{-1} \cdot dist_H^2(w^k, S^*) \qquad \forall\ k \geq 0,$$

and the proof is complete. ∎

Based on Theorem 3.4, the linear convergence rate of the sequence $\{\lambda^k\}$ generated by the PADMM (1.15) can be immediately derived. We summarize it in the following theorem.

THEOREM 3.5. *Assume that Assumption* 2.1 *and strong CHIP are both satisfied. If the sequence* $\{w^k\}$ *generated by the PADMM* (1.15) *converges to* w^* *and it satisfies the PADMM-iteration-based error bound condition* (3.5)*, then there exists* $\tilde{\kappa} > 0$ *such that*

$$dist(\lambda^k, S_\lambda^*) \le \left(1 + \frac{1}{\tilde{\kappa}^2}\right)^{-\frac{k}{2}} \cdot dist_H(w^0, S^*) \qquad \forall\ k \ge 0,$$

where S_λ^* *is defined by* (3.4)*. That is, the sequence* $\{\lambda^k\}$ *generated by the PADMM* (1.15) *converges linearly.*

If the convergence of PADMM (1.15) is guaranteed specifically by Assumption 2.2 as discussed in subsection 2.3, then accordingly we can further specify the linear convergence rate of the PADMM (1.15) in the following two theorems. Note that the linear convergence results established below are both global, according to Theorem 3.4.

THEOREM 3.6 (global linear convergence rate of $\{v^k\}$). *Let assumptions in Theorem* 3.4 *hold, and additionally if case* (1) *of Assumption* 2.2 *holds, then it follows that*

$$dist_{H_0}(v^k, S_v^*) \le \left(1 + \frac{1}{\tilde{\kappa}^2}\right)^{-\frac{k}{2}} \cdot dist_H(w^0, S^*) \qquad \forall\ k \ge 0.$$

That is, the sequence $\{v^k\}$ *generated by the PADMM* (1.15) *converges linearly.*

THEOREM 3.7 (global linear convergence rate of $\{w^k\}$). *Let assumptions in Theorem* 3.4 *hold, and additionally if case* (2) *of Assumption* 2.2 *holds, then it follows that*

$$dist_H(w^k, S^*) \le \left(1 + \frac{1}{\tilde{\kappa}^2}\right)^{-\frac{k}{2}} \cdot dist_H(w^0, S^*) \qquad \forall\ k \ge 0.$$

That is, the sequence $\{w^k\}$ *generated by the PADMM* (1.15) *converges linearly.*

For the special case where $D = 0$, the PADMM (1.15) reduces to the original ADMM (1.2). Theorem 3.6 indicates the linear convergence rate of the ADMM (1.2) in the sense of $\{v^k\}$ under case (1) of Assumption 2.2, which is consistent with the analysis in the ADMM literature. Recall that the variable x is intermediate and it is not involved in the iteration performed by (1.2); hence convergence results of the ADMM (1.2) are measured only by the variables y and λ, and x does not appear.

3.2. FEB (1.7) is sufficient to ensure (3.5). In the last subsection, we have proved the linear convergence rate of PADMM (1.15) under the PADMM-iteration-based error bound condition (3.5). Generally this condition cannot be checked directly. But we shall show that the FEB (1.7) suffices to ensure (3.5); hence (3.5) is theoretically weaker than (1.7).

Let us start with presenting a lemma which will be often used in the analysis later. The proof is trivial by using the characterization of an iterate of the PADMM (1.15) given in (2.5); it is thus omitted. We need one more matrix to simplify the notation in the analysis:

$$\hat{H} := \begin{pmatrix} D & -\beta A^T B & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix}. \tag{3.7}$$

LEMMA 3.8. *Let $\{w^k\}$ be the sequence generated by the PADMM* (1.15)*, $\phi(\cdot)$ be defined by* (1.5)*, and $\hat{H}$ as in* (3.7)*. Then, we have*

$$\begin{pmatrix} D(x^k - x^{k+1}) - \beta A^T B(y^k - y^{k+1}) \\ 0 \\ \frac{1}{\beta}(\lambda^k - \lambda^{k+1}) \end{pmatrix} \in \phi(x^{k+1}, y^{k+1}, \lambda^{k+1}), \tag{3.8}$$

or equivalently,

$$\hat{H}(w^k - w^{k+1}) \in \phi(x^{k+1}, y^{k+1}, \lambda^{k+1}). \tag{3.9}$$

Based on (3.8), we immediately find that $dist(0, \phi(w^{k+1}))$ can be bounded by $\|w^{k+1} - w^k\|_H$. This is shown in the following lemma.

LEMMA 3.9. *Let $\{w^k\}$ be the sequence generated by the PADMM* (1.15) *and $\phi(\cdot)$ be defined by* (1.5)*. There exists $L_1 > 0$ such that*

$$dist(0, \phi(w^{k+1})) \leq L_1 \|w^{k+1} - w^k\|_H. \tag{3.10}$$

Proof. It follows from (3.8) that

$$\begin{aligned} dist(0, \phi(w^{k+1})) &\leq \left(\|D(x^k - x^{k+1}) - \beta A^T B(y^k - y^{k+1})\|^2 + \left\| \frac{1}{\beta}(\lambda^k - \lambda^{k+1}) \right\|^2 \right)^{\frac{1}{2}} \\ &\leq \|D(x^k - x^{k+1}) - \beta A^T B(y^k - y^{k+1})\| + \left\| \frac{1}{\beta}(\lambda^k - \lambda^{k+1}) \right\| \\ &\leq \|D(x^{k+1} - x^k)\| + \rho(A)\sqrt{\beta} \|\sqrt{\beta} B(y^{k+1} - y^k)\| \\ &\quad + \frac{1}{\sqrt{\beta}} \left\| \frac{1}{\sqrt{\beta}}(\lambda^{k+1} - \lambda^k) \right\| \\ &\leq \left(\sqrt{\rho(D)} + \rho(A)\sqrt{\beta} + \frac{1}{\sqrt{\beta}} \right) \|w^{k+1} - w^k\|_H, \end{aligned}$$

where $\rho(D) \geq 0$ and $\rho(A) \geq 0$ are the spectral radii of the matrices D and A, respectively. Therefore, the assertion (3.10) is proved with $L_1 := \sqrt{\rho(D)} + \rho(A)\sqrt{\beta} + \frac{1}{\sqrt{\beta}} > 0$. ∎

Now, it becomes clear that the FEB (1.7) gives the relationship between the terms $dist(w^{k+1}, S^*)$ and $dist(0, \phi(w^{k+1}))$ and thus effectively bridges the inequalities (3.2) and (3.10) and eventually ensures the PADMM-iteration-based error bound condition (3.5). We give the full description in the following lemma. Our motivation for studying the FEB (1.7) for the linear convergence rate of the PADMM (1.15) is indeed justified; more details will be given in section 4.

LEMMA 3.10. *Let $\{w^k\}$ be the sequence generated by the PADMM* (1.15) *and it converges to w^*. Then the FEB* (1.7) *around w^* ensures the PADMM-iteration-based error bound condition* (3.5).

Proof. It follows from (3.10) in Lemma 3.9 and the FEB (1.7) that there exists $\kappa > 0$ and $\epsilon > 0$ such that

$$dist(w^{k+1}, S(0)) \leq \kappa dist(0, \phi(w^{k+1})) \leq L_1 \kappa \|w^{k+1} - w^k\|_H \qquad \text{when } w^{k+1} \in \mathcal{B}_\epsilon(w^*).$$

According to (3.2), we know that $dist_H(\cdot, S^*) \leq L_H \cdot dist(\cdot, S^*)$ holds for $L_H > 0$. Thus, we have

$$dist_H(w^{k+1}, S^*) \leq L_H \cdot dist(w^{k+1}, S(0)) \leq L_H L_1 \kappa \|w^{k+1} - w^k\|_H \text{ when } w^{k+1} \in \mathcal{B}_\epsilon(w^*),$$

and the proof is complete. ∎

Remark 3.1. Recall the definitions of $\phi(\cdot)$ in (1.5) and $S(p)$ in (1.6) and also note that the sequence $\{w^k\}$ generated by the PADMM (1.15) ensures (3.8). Hence, the term $\hat{H}(w^k - w^{k+1})$ in (3.9) can be regarded as a perturbation p of $S(p)$. Moreover, it follows from (1.8) that the set-valued map $S(p)$ is calm around $(0, \bar{w})$ iff there exists $\kappa > 0, \sigma > 0$ and a neighborhood $\mathcal{B}_\epsilon(\bar{w})$ of $\bar{w}$ such that

$$dist(w, S(0)) \leq k\|p\| \qquad \forall\ w \in \mathcal{B}_\epsilon(\bar{w}) \cap S(p),\ \|p\| < \sigma.$$

Then, according to Lemma 3.9, it is clear that the calmness of $S(p)$, which is independent of the iterative sequence $\{w^k\}$ generated by the PADMM (1.15), suffices to ensure the PADMM-iteration-based error bound condition (3.5). Also, notice that the calmness of $S(p)$ at $(0, \bar{w})$ is equivalent to the FEB (1.7) around $\bar{w}$. Hence, it is reasonable to study the FEB (1.7) to ensure (3.5) for the PADMM (1.15). We refer to [52] for a more general study, in which a unified framework is proposed to develop appropriate sufficient conditions for ensuring various error bound conditions that are based on some algorithmic iterations.

Using Theorem 3.4 and Lemma 3.10, we immediately have the following theorem and its proof is omitted.

THEOREM 3.11. *Let $\{w^k\}$ be the sequence generated by the PADMM* (1.15) *and it converges to w^*. If Assumption* 2.1 *and strong CHIP are both satisfied, and the FEB* (1.7) *is fulfilled around w^*, then there exists $\tilde{\kappa}$ such that*

$$dist_H^2(w^{k+1}, S^*) \leq \left(1 + \frac{1}{\tilde{\kappa}^2}\right)^{-1} \cdot dist_H^2(w^k, S^*) \qquad \forall\ k \geq 0.$$

Then, we can elaborate on the global linear convergence rate of the sequence generated by the PADMM (1.15) under different scenarios. We summarize the results in the following theorem and skip the proof.

THEOREM 3.12 (global linear convergence rate under FEB (1.7)). *Let the assumptions of Theorem* 3.11 *hold. Then we have*

$$dist(\lambda^k, S_\lambda^*) \leq \left(1 + \frac{1}{\tilde{\kappa}^2}\right)^{-\frac{k}{2}} \cdot dist_H(w^0, S^*) \qquad \forall\ k \geq 0,$$

where S_λ^* is defined by (3.4). *That is, the sequence $\{\lambda^k\}$ generated by the PADMM* (1.15) *converges linearly. In addition, if Assumption* 2.2 *is assumed, then we have the following assertions:*

(1) *If case* (1) *of Assumption* 2.2 *holds, it follows that*

$$dist_{H_0}(v^k, S_v^*) \leq \left(1 + \frac{1}{\tilde{\kappa}^2}\right)^{-\frac{k}{2}} \cdot dist_H(w^0, S^*) \qquad \forall\ k \geq 0.$$

That is, the sequence $\{v^k\}$ generated by the PADMM (1.15) *converges linearly.*

(2) *If case* (2) *of Assumption* 2.2 *holds, it follows that*

$$dist_H(w^k, S^*) \leq \left(1 + \frac{1}{\tilde{\kappa}^2}\right)^{-\frac{k}{2}} \cdot dist_H(w^0, S^*) \qquad \forall\ k \geq 0.$$

That is, the sequence $\{w^k\}$ generated by the PADMM (1.15) *converges linearly.*

In general, the FEB (1.7) may not hold (see the next section for such an example). The following corollary suggests some interesting cases with practical interests where the validation of the FEB (1.7) can be easily verified.

COROLLARY 3.13. *In the model* (1.1), *suppose that both* ∂f *and* ∂g *are polyhedral multifunctions, and* $\mathcal{X}$ *and* $\mathcal{Y}$ *are polyhedral sets. Then, the FEB* (1.7) *is fulfilled around any point in* S^*.

Proof. Note first $\mathcal{F}$ defined by (2.1) is a polyhedra. Since the graph of $\mathcal{N}_{\mathcal{X}}$ is a finite union of polyhedral convex sets, $\mathcal{N}_{\mathcal{X}}$ is polyhedral. Hence, the sum of polyhedral maps $\partial f + \mathcal{N}_{\mathcal{X}}$ is polyhedral. Similarly, $\partial g + \mathcal{N}_{\mathcal{Y}}$ is polyhedral as well, and so is the inverse map

$$S(p) := \{(x, y, \lambda) : p \in \phi(x, y, \lambda)\}.$$

By [45, Proposition 1], $S(\cdot)$ is upper-Lipschitz. Hence, FEB (1.7) is fulfilled around any KKT point. ∎

4. More discussions on various error bound conditions. In the preceding section, we have shown the linear convergence of PADMM (1.15) under the FEB (1.7). In fact, though different in form, the FEB (1.7), the proximal EB-I (1.12) and the proximal EB-II (1.14) are in essence equivalent (see the proofs in Appendix A). In this section, we provide more details on why we prefer the FEB (1.7) to the proximal EB-I (1.12) and proximal EB-II (1.14) for analyzing the linear convergence rate of PADMM (1.15) despite their theoretical equivalence.

As briefly mentioned preceding Lemma 3.10, to meet the PADMM-iteration-based error bound condition (3.5), we need to bound the term $dist_H(w^{k+1}, S^*)$ by $\|w^{k+1} - w^k\|_H$. On the other hand, the inequalities in (3.2) and (3.10) give us

$$dist_H(w^{k+1}, S^*) \le L_H \cdot dist(w^{k+1}, S^*), \quad dist(0, \phi(w^{k+1})) \le L_1 \|w^{k+1} - w^k\|_H.$$

Hence, essentially we need to build up the link between the terms $dist(w^{k+1}, S^*)$ and $dist(0, \phi(w^{k+1}))$. This is perfectly achieved by the FEB (1.7).

For the proximal EB-I (1.12), however, it facilitates bridging the terms $dist(w^{k+1}, S^*)$ and $\|R_1(w^{k+1})\|$, or the terms $dist(w^{k+1}, S^*)$ and $dist(0, R_2(w^{k+1}))$ by the proximal EB-II (1.14). In other words, neither (1.12) nor (1.14) can be directly used for bridging the terms $dist_H(w^{k+1}, S^*)$ and $\|w^{k+1} - w^k\|_H$ and hence ensuring (3.5). Additional and more complicated manipulations are needed if (1.12) or (1.14) is used.

Let us further explain the differences among these error bound conditions in studying the linear convergence rate of the particular PADMM (1.15) from the perturbation perspective. As mentioned, the FEB (1.7) around a reference point $\bar{w}$ is equivalent to the calmness of $S(p)$ at $(0, \bar{w})$. On the other hand, if we define $S_{prox-I} : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m \rightrightarrows \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m$ as

$$S_{prox-I}(p) := \left\{ (x, y, \lambda) \mid \begin{pmatrix} p_1 \in (\partial f + \mathcal{N}_{\mathcal{X}})(x - p_1) - A^T\lambda \\ p_2 \in (\partial g + \mathcal{N}_{\mathcal{Y}})(y - p_2) - B^T\lambda \\ p_3 = Ax + By - b \end{pmatrix} \right\} \tag{4.1}$$

with $p = (p_1, p_2, p_3) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m$, then we have $S_{prox-I}(0) = S^*$ and hence the proximal EB-I (1.12) around a reference point $\bar{w}$ is equivalent to the calmness of $S_{prox-I}(p)$ at $(0, \bar{w})$. That is, there exists $\kappa > 0, \sigma > 0$ and a neighborhood $\mathcal{B}_\epsilon(\bar{w})$ of $\bar{w}$ such that

$$dist(w, S_{prox-I}(0)) \le k\|p\| \qquad \forall\ w \in \mathcal{B}_\epsilon(\bar{w}) \cap S_{prox-I}(p),\ \|p\| < \sigma.$$

Let us further define $S_{prox-II} : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m \rightrightarrows \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m$ as

$$S_{prox-II}(p) := \left\{ (x, y, \lambda) \mid \begin{pmatrix} p_1 \in \partial f(x) - A^T\lambda + \mathcal{N}_{\mathcal{X}}(x - p_1) \\ p_2 \in \partial g(y) - B^T\lambda + \mathcal{N}_{\mathcal{Y}}(y - p_2) \\ p_3 = Ax + By - b \end{pmatrix} \right\}$$

with $p = (p_1, p_2, p_3) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m$. It is easy to see that $S_{prox-II}(0) = S^*$ and the proximal EB-II (1.14) around a reference point $\bar{w}$ is equivalent to the calmness of $S_{prox-II}(p)$ at $(0, \bar{w})$. That is, there exists $\kappa > 0, \sigma > 0$ and a neighborhood $\mathcal{B}_\epsilon(\bar{w})$ of $\bar{w}$ such that

$$dist(w, S_{prox-II}(0)) \leq k\|p\| \qquad \forall\, w \in \mathcal{B}_\epsilon(\bar{w}) \cap S_{prox-II}(p),\ \|p\| < \sigma.$$

According to [52], for the sequence $\{\mathbb{w}^k\}$ generated by an algorithm, if $\hat{H}(\mathbb{w}^{k+1} - \mathbb{w}^k)$ with an appropriate $\hat{H}$ is regarded as the perturbation of the corresponding optimality system, then the calmness of the induced set-valued mapping straightforwardly implies the desirable error bound that is tailored for the algorithm under investigation. When the PADMM (1.15) is considered, as shown by (3.8) in Lemma 3.8, $\hat{H}(w^k - w^{k+1})$ corresponds to the canonical perturbation of the KKT system $0 \in \phi(w)$. Hence, it motivates us to consider the perturbed multifunction $S(p)$ defined by (1.6), instead of $S_{prox-I}(p)$ or $S_{prox-II}(p)$. That is, we prefer the FEB (1.7), rather than the proximal EB-I (1.12) or proximal EB-II (1.14), for analyzing the linear convergence of the PADMM (1.15).

In addition to the superiority of yielding an easier analysis for the linear convergence rate of the problem (1.1), studying the calmness of $S(p)$, rather than $S_{prox-I}(p)$ or $S_{prox-II}(p)$, may lead to some interesting future work, as we shall mention in section 7. Also, as we shall show soon in the next section, considering the perturbed mapping $S(p)$ in (1.6) enables us discern that the second part of the left-hand side of (3.8) remains zero for each iteration of the PADMM (1.15). This insight inspires us to study a PEB condition to ensure the linear convergence of the PADMM (1.15), which seems to be novel in the literature.

5. Partial error bound for the linear convergence of PADMM (1.15).

We have established the linear convergence rate for the PADMM (1.15) under the PADMM-iteration-based error bound condition (3.5) and shown that the FEB (1.7) sufficiently ensures (3.5). In this section, we show that the FEB (1.7) can be further weakened if the specific iterative scheme (1.15) is fully considered. As mentioned, this is accomplished by the observation that there is no perturbation to the second part of the perturbed mapping $S(p)$ in (1.6). Hence, taking into consideration the specific iterative scheme enables us to weaken the FEB (1.7) to guarantee (3.5) and hence the linear convergence rate of the PADMM (1.15).

5.1. Partial error bound conditions and linear convergence.

Recall the KKT system (1.3) and the definition of the generic error bound condition (1.4). Using the terminology initiated in [39], we can also define the so-called local PEB for (1.3).

DEFINITION 5.1. *Assume S is represented as the intersection of two closed sets, i.e., $S = \mathcal{C} \cap \mathcal{D}$. The KKT system* (1.3) *is said to admit a PEB on the set $\mathcal{C}$ around $w^* \in S$ if there exists a nonnegative function $\bar{r} : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m \to \mathbb{R}_+$ satisfying $\bar{r}(w) = 0$ for $w \in \mathcal{D}$, a neighborhood $\mathcal{B}_\epsilon(w^*)$ of the point w^* and a constant $\kappa > 0$ such that*

$$[\text{PEB}^{\bar{\text{r}}} \text{ on } \mathcal{C}] \quad dist(w, S) \leq \kappa \cdot \bar{r}(w) \quad \forall\, w \in \mathcal{B}_\epsilon(w^*) \cap \mathcal{C}.$$

Obviously, from the definition, for any given closed set $\mathcal{C}$, PEB$^{\bar{r}}$ on $\mathcal{C}$ is weaker than EB^r defined by (1.4). Taking a closer look at (2.5) and (3.8), we notice that the optimality condition with respect to y, i.e., $0 \in \partial g(y) - B^T\lambda + \mathcal{N}_{\mathcal{Y}}(y)$, is satisfied for all iterates of (1.15). Let us define

$$S_g := \{w \mid 0 \in \partial g(y) - B^T\lambda + \mathcal{N}_{\mathcal{Y}}(y)\}.$$

Then, this observation motivates us to consider a partially perturbed KKT mapping $S_P : \mathbb{R}^{n_1} \times \mathbb{R}^m \rightrightarrows \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m$ as

$$S_P(p) := \{w \in S_g \mid p \in \phi_P(x, y, \lambda)\},$$

where $\phi_P : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m \rightrightarrows \mathbb{R}^{n_1} \times \mathbb{R}^m$ is defined as

$$\phi_P(w) = \begin{pmatrix} \partial f(x) - A^T\lambda + \mathcal{N}_{\mathcal{X}}(x) \\ Ax + By - b \end{pmatrix}. \tag{5.1}$$

Hence, we define a PEB that is particularly based on the specific sequence $\{w^k\}$ generated by the PADMM (1.15). Note that this definition may not be extended to other algorithms evidently.

Definition 5.2 (partial error bound). *Let $\{w^k\}$ be the sequence generated by the PADMM* (1.15) *and it converges to w^*. The KKT system* (1.3) *is said to admit a partial local error bound around w^* if there exists a neighborhood $\mathcal{B}_\epsilon(w^*)$ of w^* and some $\kappa > 0$ such that*

$$\text{[PEB]} \qquad dist(w, S_P(0)) \le \kappa \cdot dist(0, \phi_P(w)) \qquad \forall\ w \in S_g \cap \mathcal{B}_\epsilon(w^*). \tag{5.2}$$

Apparently, it holds that $S_P(0) = S(0) = S^*$, and the following relationship is easy to obtain: the proximal EB-I (1.12) and proximal EB-II (1.14) $\Leftrightarrow$ the FEB (1.7) $\Rightarrow$ PEB (5.2). That is, the PEB (5.2) is the weakest one.

We next show that the PEB (5.2) suffices to imply (3.5) and hence to ensure the linear convergence for the PADMM (1.15).

Lemma 5.3. *Let $\{w^k\}$ be the sequence generated by the PADMM* (1.15). *If the PEB* (5.2) *holds, then the PADMM-iteration-based error bound condition* (3.5) *holds as well.*

Proof. First, by virtue of (3.8) in Lemma 3.8, there always holds

$$0 \in \partial g(y^{k+1}) - B^T\lambda^{k+1} + \mathcal{N}_{\mathcal{Y}}(y^{k+1}),$$

which indicates that $w^{k+1} \in S_g$. Then by Lemma 3.9, (3.10), there is $L_1 > 0$ such that

$$dist(0, \phi_P(w^{k+1})) \le L_1\|w^{k+1} - w^k\|_H.$$

Furthermore, according to the PEB (5.2), it follows from $w^{k+1} \in S_g$ that there is $\kappa > 0$ and $\epsilon > 0$ such that

$$dist(w^{k+1}, S^*) = dist(w^{k+1}, S_P(0)) \le L_1\kappa\|w^{k+1} - w^k\|_H \quad \text{when } w^{k+1} \in \mathcal{B}_\epsilon(w^*).$$

Then, it follows from (3.2) that $dist_H(\cdot, S^*) \le L_H \cdot dist(\cdot, S^*)$ holds. Subsequently the desired estimate follows

$$dist_H(w^{k+1}, S^*) \le L_H \cdot dist(w^{k+1}, S^*) \le L_H L_1 \kappa \|w^{k+1} - w^k\|_H \quad \text{when } w^{k+1} \in \mathcal{B}_\epsilon(w^*),$$

and the proof is complete. ∎

Similar to the proof of Theorem 3.4, we can derive an important equality for the sequence $\{dist_H^2(w^{k+1}, S^*)\}$ under (5.2).

THEOREM 5.1. *Let $\{w^k\}$ be the sequence generated by the PADMM (1.15) and it converges to w^*. If Assumption 2.1 and strong CHIP are both satisfied, and the PEB (5.2) is fulfilled around w^*, then there exists $\tilde{\kappa}$ such that*

$$dist_H^2(w^{k+1}, S^*) \leq \left(1 + \frac{1}{\tilde{\kappa}^2}\right)^{-1} \cdot dist_H^2(w^k, S^*) \qquad \forall\ k \geq 0.$$

Similar to Theorem 3.12, we can further specify Theorem 5.1 as the global linear convergence rate of the PADMM (1.15) under various scenarios. We present them in the following theorem and skip the proof.

THEOREM 5.2. *Let $\{w^k\}$ be the sequence generated by the PADMM (1.15) and it converges to w^*. If Assumption 2.1 and strong CHIP are both satisfied, and the PEB (5.2) is fulfilled around w^*, then there exists $\tilde{\kappa}$ such that*

$$dist(\lambda^k, S_\lambda^*) \leq \left(1 + \frac{1}{\tilde{\kappa}^2}\right)^{-\frac{k}{2}} \cdot dist_H(w^0, S^*) \qquad \forall\ k \geq 0,$$

where S_λ^* is defined by (3.4). That is, the sequence $\{\lambda^k\}$ generated by the PADMM (1.15) converges linearly. In addition, if Assumption 2.2 is assumed, then we have the following assertions:

(1) *If case (1) of Assumption 2.2 holds, it follows that*

$$dist_{H_0}(v^k, S_v^*) \leq \left(1 + \frac{1}{\tilde{\kappa}^2}\right)^{-\frac{k}{2}} \cdot dist_H(w^0, S^*) \qquad \forall\ k \geq 0. \tag{5.3}$$

That is, the sequence $\{v^k\}$ generated by the PADMM (1.15) converges linearly.

(2) *If case (2) of Assumption 2.2 holds, it follows that*

$$dist_H(w^k, S^*) \leq \left(1 + \frac{1}{\tilde{\kappa}^2}\right)^{-\frac{k}{2}} \cdot dist_H(w^0, S^*) \qquad \forall\ k \geq 0. \tag{5.4}$$

That is, the sequence $\{w^k\}$ generated by the PADMM (1.15) converges linearly.

5.2. Example. It is interesting to compare the FEB (1.7) and the PEB (5.2). We next present an example which ensures the PEB (5.2) but fails to guarantee the FEB (1.7) at its optimal solution point. Hence, together with the fact that the FEB (1.7) is sufficient to guarantee the PEB (5.2), we show that the PEB (5.2) is weaker than the FEB (1.7).

Example 5.3. Consider a special case of the model (1.1) as

$$\begin{aligned} \min_{x,y} \quad & \frac{1}{2}x^2 + \frac{1}{2}y_1^2 + \frac{1}{4}y_2^4 \\ \text{s.t.} \quad & Ax + By = 0, \end{aligned} \tag{5.5}$$

where

$$A = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{5.6}$$

with $x \in \mathbb{R}, y = (y_1, y_2, y_3, y_4) \in \mathbb{R}^4$. Let $w = (x, y, \lambda) \in \mathbb{R}^8$. The strong CHIP follows trivially from the linearity. The KKT residual mapping in (1.5) can be specified as $\phi : \mathbb{R} \times \mathbb{R}^4 \times \mathbb{R}^3 \to \mathbb{R} \times \mathbb{R}^4 \times \mathbb{R}^3$ given by

$$\phi(w) = \begin{pmatrix} x - \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \\ \begin{pmatrix} y_1 \\ 0 \\ 0 \\ y_4^3 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} x + \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} y \end{pmatrix}.$$

It is easy to see that the optimal solution point is $w^* = (0, 0, 0, 0, 0, 0, 0, 0)$ and the solution set $S^* = \{w^*\}$. Hence, we have $dist(w, S^*) = \|w\|$. Here, all error bound conditions use l_1 norm for this example.

For simplicity, let us take $D = 0$ in (1.15) and consider the original ADMM scheme (1.2). It is easy to see that the exact expression of the iterative scheme (1.2) for this example is explicitly written as

$$w^{k+1} = \left(\frac{\lambda_3^k - \beta y_2^k}{1 + 2\beta}, 0, \frac{\beta y_2^k - \lambda_3^k}{1 + 2\beta}, 0, 0, 0, 0, \frac{\beta^2 y_2^k + (1 + \beta)\lambda_3^k}{1 + 2\beta} \right) \qquad \forall k \geq 1.$$

Recall that the variable x is intermediate and we only need to focus on the nonzero y- and λ-variables of the iteration, *i.e.*, $(y_2^{k+1}, \lambda_3^{k+1})$. Accordingly, we define the matrix T as

$$T = \begin{pmatrix} \frac{\beta}{1+2\beta} & \frac{-1}{1+2\beta} \\ \frac{\beta^2}{1+2\beta} & \frac{1+\beta}{1+2\beta} \end{pmatrix}, \tag{5.7}$$

and the iteration is essentially executed by the recursion:

$$(y_2^{k+1}, \lambda_3^{k+1}) = T(y_2^k, \lambda_3^k) \ \forall k \geq 1.$$

By straightforward calculation, the eigenvalues of the matrix T in (5.7) are $(1 + 2\beta \pm \sqrt{1 - 4\beta^2})/(2 + 4\beta)$. Therefore, $\rho(T)$, the spectral radius of T, is strictly smaller than 1 for any $\beta > 0$. Thus, we know that $\{w^k\}$ converges to $w^* = (0, 0, 0, 0, 0, 0, 0, 0)$ linearly.

We next show by contradiction that the KKT system $\phi(x, y, \lambda) = 0$ fails to admit the FEB (1.7) around w^*. Let $\{\delta_k\}$ be a sequence such that $\delta_k \searrow 0$ and define the sequence $\{w^k\}$ by $w^k = (0, 0, -\delta_k, -\delta_k, \delta_k, 0, 0, 0) \ \forall k \geq 0$. It is clear from the construction that $w^k \to w^*$ and $dist(w^k, S^*) = 3\|\delta_k\|$. On the other hand, $dist(0, \phi(w^k)) = \|\phi(w^k)\| = \|\delta_k^3\| \ \forall k \geq 0$. This leads to $dist(0, \phi(w^k)) = o(dist(w^k, S^*))$. Consequently, the KKT system $\phi(x, y, \lambda) = 0$ does not possess the FEB (1.7) around w^*.

For analyzing the PEB (5.2) around w^*, in this example, we specify the partial KKT residual mapping $\phi_P : \mathbb{R} \times \mathbb{R}^4 \times \mathbb{R}^3 \to \mathbb{R} \times \mathbb{R}^3$ defined by (5.1):

$$\phi_P(w) = \begin{pmatrix} x - \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} x + \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} y \end{pmatrix}.$$

Let us further define

$$S_y := \left\{ (x, y, \lambda) \mid 0 = \begin{pmatrix} y_1 \\ 0 \\ 0 \\ y_4^3 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \right\},$$

which can be simplified as

$$S_y := \{(x, y, \lambda) \mid y_1 - \lambda_1 = 0,\ \lambda_2 = 0,\ \lambda_1 = 0,\ \ y_4^3 - \lambda_1 - \lambda_2 = 0\}.$$

Therefore, for any sequence $\{w^k\} \subseteq S_y \cap \mathcal{B}_{\frac{1}{2}}(w^*)$ and $\|w^k\| \to 0$, the following equality holds:

$$\begin{aligned} dist(0, \phi_P(w^k)) = \|\phi_P(w^k)\| &= |x^k - \lambda_2^k - \lambda_3^k| + |y_1^k + y_3^k + y_4^k| + |x^k + y_2^k + y_4^k| + |x^k| \\ &= |x^k| + |x^k - \lambda_3^k| + |y_3^k| + |x^k + y_2^k|. \end{aligned}$$

It is clear that

$$|y_2| \leq |x + y_2| + |x|$$

and

$$|\lambda_3| \leq |x - \lambda_3| + |x|.$$

Consequently, for the sequence $\{w^k\} \subseteq S_y \cap \mathcal{B}_{\frac{1}{2}}(w^*)$, we have the following estimate:

$$\begin{aligned} dist(w^k, S^*) = \|w^k\| &= |x^k| + |y_1^k| + |y_2^k| + |y_3^k| + |y_4^k| + |\lambda_1^k| + |\lambda_2^k| + |\lambda_3^k| \\ &\leq |x^k| + |x^k + y_2^k| + |x^k| + |y_3^k| + |x^k - \lambda_3^k| + |x^k| \\ &\leq 3(|x^k| + |x^k - \lambda_3^k| + |y_3^k| + |x^k + y_2^k|) \\ &= 3\, dist(0, \phi_P(w^k)). \end{aligned}$$

Therefore, the KKT system $\phi(x, y, \lambda) = 0$ admits the PEB (5.2) around w^*.

Remark 5.4. Example 5.3 with a few variables is sufficient to prove the advantage of considering the PEB (5.2) for studying the linear convergence rate of the ADMM (1.15). It is analogous to constructing convex polynomial optimization problems in higher dimension so that only the PEB (5.2) holds while the FEB (1.7) does not.

Note that the matrix B given in (5.6) is not of full column rank; hence Assumption 2.2 is not satisfied and this reflects that Assumption 2.2 is sufficient, instead of necessary, to ensure the convergence of the PADMM (1.15). Moreover, it is verified that the FEB (1.7) fails and thus it is invalid to explain the linear convergence rate of the application of the ADMM (1.2) to this specific example. Instead, the PEB (5.2) is satisfied for this example and hence the linear convergence rate of the ADMM is theoretically explained.

6. Difference of updating the primal variables in ADMM (1.2). Despite the main purpose of studying the linear convergence rate of the PADMM (1.15) under weaker error bound conditions, an interesting byproduct of this paper is a theoretical explanation for the changes that occur when the primal variables are updated in a different order. For simplicity, let us focus on the original ADMM (1.2) in this section.

6.1. Another form of ADMM. If we swap the order of the primal variables x and y in (1.2), another form of the ADMM is obtained:

$$\begin{cases} y^{k+1} = \arg\min\limits_{y\in\mathcal{Y}} \left\{ g(y) - \langle \lambda^k, Ax^k + By - b\rangle + \dfrac{\beta}{2}\|Ax^k + By - b\|^2 \right\}, \\ x^{k+1} = \arg\min\limits_{x\in\mathcal{X}} \left\{ f(x) - \langle \lambda^k, Ax + By^{k+1} - b\rangle + \dfrac{\beta}{2}\|Ax + By^{k+1} - b\|^2 \right\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{cases} \tag{6.1}$$

Obviously, the iterative scheme (6.1) can be written as

$$\begin{cases} 0 \in \partial g(y^{k+1}) - B^T\lambda^{k+1} + \beta B^T A(x^k - x^{k+1}) + \mathcal{N}_{\mathcal{Y}}(y^{k+1}), \\ 0 \in \partial f(x^{k+1}) - A^T\lambda^{k+1} + \mathcal{N}_{\mathcal{X}}(x^{k+1}), \\ 0 = Ax^{k+1} + By^{k+1} - b + \dfrac{1}{\beta}(\lambda^{k+1} - \lambda^k). \end{cases} \tag{6.2}$$

The convergence of (6.1) certainly holds given the convergence of (1.2). But these two schemes differ in the intermediate variables and the order of updating the primal variables: x and y. Numerically, it does make a difference whether x or y is placed as the first variable to be updated. An immediate explanation is that if the x-subproblem is significantly more complicated than the y-subproblem, it seems smarter to update y first so as to avoid the possible transmission of error caused by solving the x-subproblem inexactly. Such situations arise in the case where, e.g., one subproblem is in a higher dimension or of a more complicated nature than the other one. Representative examples are the so-called sparse and low-rank optimization models which at each iteration require one to solve a subproblem involving the singular value decomposition of a large matrix and thus inner iterations with accumulative errors are inevitable, and the other subproblem which usually has the closed-form solution and hence no inner iteration is needed. For such problems, it is highly suggested to update the easier subproblem first and this makes a significant difference in the eventual numerical performance; see, e.g., [36, 53]. Meanwhile, it seems no theory is known to explain this difference caused by different orders of updating the primal variables. We next show that the two schemes may admit different convergence rates in sense of different PEB assumptions and thus provide a theoretical explanation for this issue.

6.2. Partial error bound condition for (6.1). We need the matrix to simplify the notation in the analysis:

$$\tilde{H} := \begin{pmatrix} 0 & 0 & 0 \\ -\beta B^T A & 0 & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix}. \tag{6.3}$$

Similar to Lemma 3.8, we present the following lemma, which follows directly from the characterization of an iterate of (6.1) given in (6.2).

LEMMA 6.1. *Let $\{w^k\}$ be the sequence generated by* (6.1) *and $\phi(\cdot)$ be defined by* (1.5) *and $\tilde{H}$ in* (6.3). *Then, we have*

$$\begin{pmatrix} 0 \\ -\beta B^T A(x^k - x^{k+1}) \\ \frac{1}{\beta}(\lambda^k - \lambda^{k+1}) \end{pmatrix} \in \phi(x^{k+1}, y^{k+1}, \lambda^{k+1}). \tag{6.4}$$

or equivalently,

$$\tilde{H}(w^k - w^{k+1}) \in \phi(x^{k+1}, y^{k+1}, \lambda^{k+1}).$$

Taking a closer look at (6.2) and (6.4), we notice that the optimality condition with respect to x, i.e., $0 \in \partial f(x) - A^T\lambda + \mathcal{N}_{\mathcal{X}}(x)$, is satisfied for all iterates of (6.1). Following our discussion in the preceding section, by letting

$$S_f := \{(x, y, \lambda) \mid 0 \in \partial f(x) - A^T\lambda + \mathcal{N}_{\mathcal{X}}(x)\},$$

we can define a PEB for the ADMM scheme (6.1) as follows.

Definition 6.2 (partial error bound-yx). *Let $\{w^k\}$ be the sequence generated by* (6.1) *and it converges to w^*. The KKT system* (1.3) *is said to admit a local PEB-yx around w^* if there exists a neighborhood $\mathcal{B}_\epsilon(w^*)$ of w^* and $\kappa > 0$ such that*

$$\text{[PEB} - \text{yx]} \qquad dist(w, S^*) \le \kappa \cdot dist(0, \bar{\phi}_P(w)) \quad \forall\, w \in S_f \cap \mathcal{B}_\epsilon(w^*), \tag{6.5}$$

where $\bar{\phi}_P : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m \rightrightarrows \mathbb{R}^{n_1} \times \mathbb{R}^m$ is defined as

$$\bar{\phi}_P(w) = \begin{pmatrix} \partial g(y) - B^T\lambda + \mathcal{N}_{\mathcal{Y}}(y) \\ Ax + By - b \end{pmatrix}.$$

We define the matrix $H_{x\lambda}$ and its submatrix $H^0_{x\lambda}$ as follows to simplify the notation in our analysis:

$$H_{x\lambda} = \begin{pmatrix} \beta A^T A & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix} \quad \text{and} \quad H^0_{x\lambda} = \begin{pmatrix} \beta A^T A & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}.$$

Also, we use the notation

$$S^*_{x\lambda} := \{(x^*, \lambda^*) \mid (x^*, y^*, \lambda^*) \in S^* \text{ for some } y^*\} \tag{6.6}$$

when the convergence of the sequence of $\{x^k, \lambda^k\}$ is highlighted. Consequently, we can prove the global linear convergence rate of the scheme (6.1) if the PEB-yx (6.5) is assumed. The details are omitted.

Proposition 6.3. *Let the sequence $\{w^k\}$ be generated by* (6.1) *and it converges to w^*. If Assumption* 2.1 *and strong CHIP are both satisfied, and the PEB-yx condition* (6.5) *is fulfilled around w^*, then there exists $\tilde{\kappa}$ such that*

$$dist^2_{H_{x\lambda}}(w^{k+1}, S^*) \le \left(1 + \frac{1}{\tilde{\kappa}^2}\right)^{-1} \cdot dist^2_{H_{x\lambda}}(w^k, S^*) \qquad \forall\, k \ge 0.$$

Moreover, we have

$$dist(\lambda^k, S^*_\lambda) \le \left(1 + \frac{1}{\tilde{\kappa}^2}\right)^{-\frac{k}{2}} \cdot dist_H(w^0, S^*) \qquad \forall\, k \ge 0,$$

where S^*_λ is defined by (3.4)*. That is, the sequence $\{\lambda^k\}$ generated by* (6.1) *converges linearly. In addition, if A is of full column rank, then $H^0_{x\lambda} \succ 0$ and it follows that*

$$dist_{H^0_{x\lambda}}\left((x^k, \lambda^k), S^*_{x\lambda}\right) \le \left(1 + \frac{1}{\tilde{\kappa}^2}\right)^{-\frac{k}{2}} \cdot dist_H(w^0, S^*) \qquad \forall\, k \ge 0,$$

where $S^*_{x\lambda}$ is defined by (6.6)*. That is, the sequence $\{(x^k, \lambda^k)\}$ generated by* (6.1) *converges linearly.*

6.3. Difference between PEB (5.2) and PEB-yx (6.5). By comparing Corollary 5.1 and Proposition 6.3, a clear conclusion can be drawn from the difference in the PEB conditions (5.2) and (6.5). Let us reconsider Example 5.3 for an illustration of the difference. In particular, we will show that Example 5.3 does not admit PEB-yx around the optimal solution. The PEB (5.2), on the other hand, is satisfied according to the analysis for Example 5.3 in section 5.2.

As previously mentioned, we can easily write down the explicit recursion for the application of the ADMM scheme (6.1) to Example 5.3, and the convergence is clearly implied. We omit the details for succinctness. We further show the difference in the two PEB conditions (5.2) and (6.5) in this example. Therefore, the convergence rates of (1.2) and (6.1) may be different according to the proposed PEB theory. To this end, the associated partial KKT residual mapping $\bar{\phi}_P : \mathbb{R} \times \mathbb{R}^4 \times \mathbb{R}^3 \to \mathbb{R} \times \mathbb{R}^4$ reads as

$$\bar{\phi}_P(x, y, \lambda) = \begin{pmatrix} \begin{pmatrix} y_1 \\ 0 \\ 0 \\ y_4^3 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} x + \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} y \end{pmatrix}.$$

Let $\{\delta_k\}$ be a sequence such that $\delta_k \searrow 0$ and define the sequence $\{w^k\}$ by $w^k = (0, 0, -\delta_k, -\delta_k, \delta_k, 0, 0, 0)$, where $k = 0, 1, \ldots$. It is clear from the construction that $w^k \to w^*$, $\{w^k\} \subseteq S_x$ and $dist(w^k, S) = 3\|\delta_k\|$, where

$$S_x := \left\{ (x, y, \lambda) \mid 0 = x - \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \right\}.$$

On the other hand, $dist(0, \bar{\phi}_P(w^k)) = \|\bar{\phi}_P(w^k)\| = \|\delta_k^3\|$ for $k = 0, 1, \ldots$. This leads to $dist(0, \bar{\phi}_P(w^k)) = o(dist(w^k, S^*))$. Consequently, the PEB-$yx$ is not fulfilled around w^*. That is, the ADMM (1.2) with the updating order of $x - y$ admits the PEB and it converges linearly, but the PEB-yx (6.5) is not satisfied and there is no guarantee to the linear convergence rate for the ADMM (6.1) with the updating order of $y - x$.

6.4. More discussions. It was shown in preceding subsections that the orders of $x - y$ and $y - x$ make difference for the PADMM in sense of satisfying the PEB (5.2) and PEB-yx (6.5). This difference in the sense of PEB conditions may result in significant difference in the convergence rate of the PADMM. It is thus interesting to discuss which order should be used for updating the variables for a given particular application of the model (1.1). In this subsection, we show that one of these two PEB conditions may be weaker than the other one and thus further justify their theoretical difference. In particular, we prove that the PEB condition corresponding to one of the orders can ensure the other one under some additional assumptions. We first need to recall some well-known results in variational analysis literature.

PROPOSITION 6.4 (see [7, Proposition 2.4.3], Clarke exact penalty principle). *Let f be Lipschitz of rank K on a set U. Let x belong to a set $C \subset U$ and suppose that f attains a minimum over C at x. Then for any $\hat{K} \geq K$, the function $g(y) = f(y) + \hat{K} \cdot dist(y, C)$ attains a minimum over U at x. If $\hat{K} > K$ and C is closed, then any other point minimizing g over U must also lie in C.*

LEMMA 6.5 ([46, Exercise 9.37]). *Given A set-valued map $\mathcal{T} : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, and $(\bar{x}, \bar{p}) \in gph\mathcal{T}$, if $\mathcal{T}$ is pseudo-Lipschitz around $(\bar{x}, \bar{p})$, then there exists a neighborhood $\mathcal{B}_\epsilon(\bar{x})$ of $\bar{x}$ such that function $h(x) := dist(\bar{p}, \mathcal{T}(x))$ is Lipschitz on $\mathcal{B}_\epsilon(\bar{x})$.*

Now, we show that under certain conditions, the PEB (5.2) implies the PEB-yx (6.5).

THEOREM 6.6. *For any point $w^* \in S^*$, if the multifunction $\phi_P(w)$ defined by* (5.1) *is pseudo-Lipschitz around $(w^*, 0)$ and $\phi_g(w) := \partial g(y) - B^T\lambda + \mathcal{N}_\mathcal{Y}(y)$ is metrically subregular at $(w^*, 0)$, then the PEB-yx* (6.5) *holds at w^* if the PEB* (5.2) *is fulfilled around w^*.*

Proof. When the PEB (5.2) holds at w^*, there exist a neighborhood $\mathcal{B}_{\epsilon_1}(w^*)$ of w^* and some $\kappa_1 > 0$ such that

$$dist(w, S^*) \leq \kappa_1 \cdot dist(0, \phi_P(w)) \qquad \forall\ w \in S_g \cap \mathcal{B}_{\epsilon_1}(w^*).$$

Since $\phi_P(w)$ is pseudo-Lipschitz around $(w^*, 0)$, by Lemma 6.5, there exists $\epsilon_2 > 0$ to ensure that $h_1(w) := dist(0, \phi_P(w))$ is Lipschitz continuous on $\mathcal{B}_{\epsilon_2}(w^*)$. Because S is a convex subset, $h_2(w) := dist(w, S^*)$ is Lipschitz continuous on $\mathcal{B}_{\epsilon_2}(w^*)$, and so is $h(w) := \kappa \cdot dist(0, \phi_P(w)) - dist(w, S^*)$. Furthermore, we notice that w^* is a solution point of the following optimization problem:

$$\begin{aligned} \min_{w} \quad & \kappa_1 \cdot dist(0, \phi_P(w)) - dist(w, S^*) \\ \text{s.t.} \quad & w \in S_g \cap \mathcal{B}_{\tilde{\epsilon}}(w^*), \end{aligned}$$

where $\tilde{\epsilon} = \min\{\epsilon_1, \epsilon_2\}$. Then, it follows from Proposition 6.4 that there exists $L > 0$ such that the function

$$\tilde{h}(w) := \kappa_1 \cdot dist(0, \phi_P(w)) - dist(w, S^*) + L \cdot dist(w, S_g)$$

attains a minimum over $\mathcal{B}_{\tilde{\epsilon}}(w^*)$ at w^*. Therefore, we have

$$dist(w, S^*) \leq \max\{\kappa_1, L\} \cdot (dist(0, \phi_P(w)) + dist(w, S_g)) \qquad \forall\ w \in \mathcal{B}_{\tilde{\epsilon}}(w^*). \tag{6.7}$$

By the metric subregularity of $\phi_g(w)$ at $(w^*, 0)$, there exist $\epsilon_3 > 0$ and $\kappa_3 > 0$ such that

$$dist(w, S_g) \leq \kappa_3 \cdot dist(0, \phi_g(w)) \qquad \forall\ w \in \mathcal{B}_{\epsilon_3}(w^*). \tag{6.8}$$

Combining (6.7) and (6.8), we obtain

$$\begin{aligned} dist(w, S^*) &\leq \kappa \cdot (dist(0, \phi_P(w)) + dist(0, \phi_g(w))) \\ &\leq \sqrt{2}\kappa \cdot dist(0, \phi(w)) \qquad \forall\ w \in \mathcal{B}_\epsilon(w^*), \end{aligned}$$

where $\kappa = \max\{\kappa_1, L\} \cdot \max\{1, \kappa_3\}$ and $\epsilon = \min\{\tilde{\epsilon}, \epsilon_3\}$. We conclude that FEB (1.7), and hence PEB-yx (6.5), is fulfilled around w^*. ∎

Hence, for a specific application of the model (1.1), we prefer to choose the order of updating the variables that the corresponding PEB condition can be satisfied more likely and thus the linear convergence may be guaranteed.

7. Conclusions and future work. We studied error bound conditions to ensure the linear convergence rate for the ADMM in a convex programming context. Different from the existing literature that requires stronger assumptions or special structures on the model under discussion to ensure certain error bound conditions, we weakened these error bound conditions by considering the structures and properties of the specific algorithm under discussion. That is, algorithmic-based error bound conditions should be considered, and they can be weaker than general-purpose error bound conditions. We give both full and partial error bound conditions in accordance with the ADMM's special iterative scheme to derive its linear convergence rate; the idea of a PEB is inspired by an observation on the partially perturbed system (3.8). Furthermore, we construct an example to show that the PEB condition is weaker than the generic counterparts. The main analysis also inspires byproducts. First, a theoretical interpretation is given to explain the difference if the two primal variables are updated by different orders in ADMM's iteration. Second, the equivalence among various error bound conditions widely used in the literature is established. Our new philosophy of weakening existing error bound conditions in accordance with the specific structure of an algorithm may inspire similar research in other contexts. Moreover, we use the concepts of calmness/metric subregularity in our analysis, and the main PEB result is inspired by a perturbation perspective. We believe more deliberately drawing on the experience of these well-developed techniques in variational analysis and perturbation analysis will lead to more interesting and deeper results for the convergence analysis of other popular algorithms.

The linear convergence rates of ADMM schemes and other first-order methods via various error bound conditions have been studied in other contexts as well. For example, in [35], an extended version of the ADMM scheme (1.2) with a sufficiently small step size for updating the dual variable λ is considered for a similar but more complicated case of (1.1) where there are more than two blocks of functions in the objective; a variant of the PADMM is studied in [28] under the calmness condition of S_{prox-I} defined by (4.1). In this paper, we concentrate on a relatively simpler convex scenario where only the two-block separable convex minimization model (1.1) and the proximal version of ADMM (1.15) are considered, so that our idea can be exposed more clearly with simpler notation. Technically, we believe it is possible to extend our analysis to various more complex scenarios such as models with nonconvex function components in their objectives, and more sophisticated variants of the ADMM for two-block or even multiple-block models. Let us just mention one specific extension: it is trivial to extend our analysis to a more general version of the PADMM (1.15) considered in [12, 14]:

$$\begin{cases} x^{k+1} = \arg\min_{x\in\mathcal{X}} \left\{ f(x) - (\lambda^k)^T(Ax + By^k - b) + \frac{\beta}{2}\|Ax + By^k - b\|^2 + \frac{1}{2}\|x - x^k\|_D^2 \right\}, \\ y^{k+1} = \arg\min_{y\in\mathcal{Y}} \left\{ g(y) - (\lambda^k)^T(Ax^{k+1} + By - b) + \frac{\beta}{2}\|\alpha Ax^{k+1} \right. \\ \qquad\qquad\qquad \left. -(1-\alpha)(By^k - b) + By - b\|^2 \right\}, \\ \lambda^{k+1} = \lambda^k - \beta(\alpha Ax^{k+1} - (1-\alpha)(By^k - b) + By^{k+1} - b), \end{cases}$$

where $\alpha \in (0, 2)$ is a relaxation factor. We skip the tedious analysis for this more complex extension and present our analysis in the simplest context.

The authors of a recent paper [10] investigated some unconstrained separable convex optimization problems and illustrated that subregularity of the gradient-like mapping is equivalent to the subregularity of its subdifferential. Therefore, they employed the quadratic growth condition to characterize the error bound condition and establish the linear convergence rate for the proximal gradient method. In the literature, much interesting work provides characterizations and criteria for error bound properties in terms of various derivative-like objects in either the primal space (via directional derivatives, slopes, etc.) or the dual space (via subdifferentials, normal cones). Notice that for a given $\bar{w}$, the FEB (1.7) around $\bar{w}$ is equivalent to the metric subregularity of the KKT mapping ϕ at $(\bar{w}, 0)$ and the calmness of $S(p)$ at $(0, \bar{w})$. The proximal EB-I (1.12) and proximal EB-II (1.14) around $\bar{w}$, on the other hand, are equivalent to the calmness of $S_{prox-I}(p)$ and $S_{prox-II}(p)$, respectively, at $(0, \bar{w})$. It is known that computing the exact formula of those derivative-like objects is much simpler in absence of any proximal operators. This perspective motivates us to call on the existing extensive literature of the verifiable first- and second-order sufficient conditions for the metric subregularity of ϕ or the calmness of $S(p)$; see, e.g., [21, 22, 23, 24, 25, 26, 27, 34]. Therefore, for the particular constrained model (1.1), we may consider investigating verifiable sufficient conditions for the metric subregularity/calmness, and hence the linear convergence rates for various ADMM-type algorithms and other schemes. In particular, it is interesting to note that for the problem data with underlying polyhedral structures, the second-order sufficient condition is nearly necessary; see, e.g., [22]. This is our future work.

Appendix A. Equivalence of several error bound conditions. We show that the mentioned error bound conditions (1.7), (1.12), and (1.14) are all equivalent. First, we prove that (1.7) holds if (1.12) or (1.14) holds.

PROPOSITION A.1. *If the KKT system* (1.3) *admits either the proximal EB-I* (1.12) *or proximal EB-II* (1.14) *around a KKT point $\bar{w}$, it also admits the FEB* (1.7) *around $\bar{w}$.*

Proof. Given w, for any $u \in \partial f(x) - A^T\lambda + \mathcal{N}_{\mathcal{X}}(x)$, it holds that

$$x = \mathrm{Prox}_{f+\delta_{\mathcal{X}}}(x + A^T\lambda + u)$$

and

$$x = \mathrm{Proj}_{\mathcal{X}}(x - \xi + A^T\lambda + u) \qquad \textit{for some } \xi \in \partial f(x).$$

Since it holds that

$$0 \in \mathcal{N}_{\mathcal{X}}(x) + x - (x - \partial f(x) + A^T\lambda + u),$$

we have

$$\|x - \mathrm{Prox}_{f+\delta_{\mathcal{X}}}(x + A^T\lambda)\| = \|\mathrm{Prox}_{f+\delta_{\mathcal{X}}}(x + A^T\lambda + u) - \mathrm{Prox}_{f+\delta_{\mathcal{X}}}(x + A^T\lambda)\| \le \|u\|,$$

and thus

$$\begin{aligned} &dist(0, x - \mathrm{Proj}_{\mathcal{X}}(x - \partial f(x) + A^T\lambda)) \\ &\le \|x - \mathrm{Proj}_{\mathcal{X}}(x - \xi + A^T\lambda)\| \\ &= \|\mathrm{Proj}_{\mathcal{X}}(x - \xi + A^T\lambda + u) - \mathrm{Proj}_{\mathcal{X}}(x - \xi + A^T\lambda)\| \le \|u\|. \end{aligned}$$

Since u is arbitrarily chosen, we have the relations

$$\|x - \mathrm{Prox}_{f+\delta_{\mathcal{X}}}(x + A^T\lambda)\| \le dist(0, \partial f(x) - A^T\lambda + \mathcal{N}_{\mathcal{X}}(x))$$

and

$$dist(0, x - \mathrm{Proj}_{\mathcal{X}}(x - \partial f(x) + A^T\lambda)) \le dist(0, \partial f(x) - A^T\lambda + \mathcal{N}_{\mathcal{X}}(x)).$$

Similarly, we can establish the same results for $dist(0, \partial g(y) - B^T\lambda + \mathcal{N}_{\mathcal{Y}}(y))$. That is, we have

$$\|y - \mathrm{Prox}_{g+\delta_{\mathcal{Y}}}(y + B^T\lambda)\| \le dist(0, \partial g(y) - B^T\lambda + \mathcal{N}_{\mathcal{Y}}(y))$$

and

$$dist(0, y - \mathrm{Proj}_{\mathcal{Y}}(y - \partial g(y) + B^T\lambda)) \le dist(0, \partial g(y) - B^T\lambda + \mathcal{N}_{\mathcal{Y}}(y)).$$

Using these inequalities, it is easy to see that

$$\|R_1(w)\| \le dist(0, \phi(w)), \qquad dist(0, R_2(w)) \le dist(0, \phi(w)),$$

and the proof is complete. ∎

Notice the equality $\mathrm{Prox}_{th} = (I + t\partial h)^{-1}$. It is easy to show that (1.7) implies either (1.12) or (1.14) as well. We summarize this result in the following proposition.

PROPOSITION A.2. *If the KKT system* (1.3) *admits the FEB* (1.7) *around a KKT point $\bar{w}$, it admits the proximal EB-I* (1.12) *and proximal EB-II* (1.14) *around $\bar{w}$ as well.*

Proof. First, by virtue of

$$\begin{aligned} x + A^T\lambda - \mathrm{Prox}_{f+\delta_{\mathcal{X}}}(x + A^T\lambda) &\in \left(\partial f + \mathcal{N}_{\mathcal{X}}\right)\left(\mathrm{Prox}_{f+\delta_{\mathcal{X}}}(x + A^T\lambda)\right), \\ y + B^T\lambda - \mathrm{Prox}_{g+\delta_{\mathcal{Y}}}(y + B^T\lambda) &\in \left(\partial g + \mathcal{N}_{\mathcal{Y}}\right)\left(\mathrm{Prox}_{g+\delta_{\mathcal{Y}}}(y + B^T\lambda)\right), \end{aligned}$$

we conclude that

$$dist\left(0, \phi\left(\mathrm{Prox}_{f+\delta_{\mathcal{X}}}(x + A^T\lambda), \mathrm{Prox}_{g+\delta_{\mathcal{Y}}}(y + B^T\lambda), \lambda\right)\right) \le \|R_1(w)\|. \tag{A.1}$$

Therefore, for any $w \in \mathcal{B}_\epsilon(\bar{w})$, we have

$$\begin{aligned} dist(w, S(0)) &\le c_1(\|x - \mathrm{Prox}_{f+\delta_{\mathcal{X}}}(x + A^T\lambda)\| + \|y - \mathrm{Prox}_{g+\delta_{\mathcal{Y}}}(y + B^T\lambda)\|) \\ &\quad + dist\left(\left(\mathrm{Prox}_{f+\delta_{\mathcal{X}}}(x + A^T\lambda), \mathrm{Prox}_{g+\delta_{\mathcal{Y}}}(y + B^T\lambda), \lambda\right), S(0)\right) \\ &\le c_1(\|x - \mathrm{Prox}_{f+\delta_{\mathcal{X}}}(x + A^T\lambda)\| + \|y - \mathrm{Prox}_{g+\delta_{\mathcal{Y}}}(y + B^T\lambda)\|) \\ &\quad + \kappa \cdot dist\left(0, \phi\left(\mathrm{Prox}_{f+\delta_{\mathcal{X}}}(x + A^T\lambda), \mathrm{Prox}_{g+\delta_{\mathcal{Y}}}(y + B^T\lambda), \lambda\right)\right), \\ &\le (2c_1 + \kappa) \cdot \|R_1(w)\|, \end{aligned}$$

where the second inequality follows from the FEB (1.7), and the third inequality is a direct consequence of (A.1). Thus we get the proximal EB-I (1.12) around $\bar{w}$. We can obtain the proximal EB-II (1.14) similarly. The proof is complete. ∎

With Propositions A.1 and A.2, the equivalence between the FEB (1.7) and the proximal EB-I (1.12) or proximal EB-II (1.14) is established.

Acknowledgments. We are grateful to two anonymous referees for their very helpful suggestions and comments, which have helped us improve the presentation of this paper substantially. We also would like to thank Professor Jane J. Ye for her very intensive discussions, which have helped us improve section 6.4 greatly.

REFERENCES

[1] T. Aspelmeier, C. Charitha, and D. R. Luke, *Local linear convergence of the ADMM/Douglas–Rachford algorithms without strong convexity and application to statistical imaging*, SIAM J. Imaging Sci., 9 (2016), pp. 842–868.

[2] J. P. Aubin, *Lipschitz behavior of solutions to convex minimization problems*, Math. Oper. Res., 9 (1984), pp. 87–111.

[3] D. Boley, *Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs*, SIAM J. Optim., 23 (2013), pp. 2183–2207.

[4] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2010), pp. 1–122.

[5] T. F. Chan and R. Glowinski, *Finite Element Approximation and Iterative Solution of a Class of Mildly Non-Linear Elliptic Equations*, Technical report, Stanford University, 1978.

[6] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problem with applications to imaging*, J. Math. Imaging Vision 40 (2011), pp. 120–145.

[7] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Classics in Appl. Math. 5, SIAM, Philadelphia, 1990.

[8] W. Deng and W. Yin, *On the global and linear convergence of the generalized alternating direction method of multipliers*, J. Sci. Comput., 66 (2016), pp. 889–916.

[9] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings*, Springer Monogr. Math., Springer, Berlin, 2009.

[10] D. Drusvyatskiy and A. Lewis, *Error Bounds, Quadratic Growth, and Linear Convergence of Proximal Methods*, preprint, arXiv:1602.06661, 2016.

[11] J. Eckstein and D. P. Bertsekas, *An Alternating Direction Method for Linear Programming*, LIDS-P, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, 1990.

[12] J. Eckstein and D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293–318.

[13] J. Eckstein and W. Yao, *Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives*, Pacific J. Optim., 11 (2015), pp. 619–644.

[14] E. X. Fang, B. S. He, H. Liu, and X. M. Yuan, *Generalized alternating direction method of multipliers: New theoretical insights and applications*, Math. Program. Comput., 7 (2015), pp. 149–187.

[15] F. Francisco and J. S. Pang, *Finite-dimensional Variational Inequalities and Complementarity Problems*, Springer, New York, 2007.

[16] M. Fortin and R. Glowinski, *Augmented Lagrangian Methods: Applications to the Numerical Solutions of Boundary Value Problems*, Stud. Math. Appl. 15, North-Holland, Amsterdam, 1983.

[17] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite-element approximations*, Comput. Math. Appl., 2 (1976), pp. 17–40.

[18] R. Glowinski, *On alternating direction methods of multipliers: A historical perspective*, in Modeling, Simulation and Optimization for Science and Technology, W. Fitzgibbon, Y. A. Kuznetsov, P. Neittaanmäki, and O. Pironneau, Springer, New York, 2014, pp. 59–82.

[19] R. Glowinski and A. Marrocco, *Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problèmes non linéaires*, RAIRO, R2 (1975), pp. 41–76.

[20] R. Glowinski and P. L. Tallec, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*, Stud. Appl. Math., SIAM, Philadelphia, 1989.

[21] H. Gfrerer, *First order and second order characterizations of metric subregularity and calmness of constraint set mappings*, SIAM J. Optim., 21 (2011), pp. 1439–1474.

[22] H. Gfrerer, *On directional metric regularity, subregularity and optimality conditions for nonsmooth mathematical programs*, Set-Valued Var. Anal., 21 (2013), pp. 151–176.

[23] H. Gfrerer and D. Klatte, *Lipschitz and Holder stability of optimization problems and generalized equations*, Math. Program., 158 (2016), pp. 35–75.

[24] H. Gfrerer and B. S. Mordukhovich, *Complete characterizations of tilt stability in nonlinear programming under weakest qualification conditions*, SIAM J. Optim., 25 (2015), pp. 2081–2119.

[25] H. Gfrerer and B. S. Mordukhovich, *Robinson stability of parametric constraint systems via variational analysis*, SIAM J. Optim., 27 (2017), pp. 438–465.

[26] H. Gfrerer and J. V. Outrata, *On Lipschitzian properties of implicit multifunctions*, SIAM J. Optim., 26 (2016), pp. 2160–2189.

[27] H. Gfrerer and J. J. Ye, *New constraint qualifications for mathematical programs with equilibrium constraints via variational analysis*, SIAM J. Optim., 27 (2017), pp. 842–865.
[28] D. R. Han, D. F. Sun, and L. W. Zhang, *Linear rate convergence of the alternating direction method of multipliers for convex composite quadratic and semi-definite programming*, preprint, arXiv:1508.02134, 2015.
[29] D. R. Han and X. M. Yuan, *Local linear convergence of the alternating direction method of multipliers for quadratic programs*, SIAM J. Numer. Anal., 51 (2013), pp. 3446–3457.
[30] B. S. He and H. Yang, *Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities*, Oper. Res. Lett., 23 (1998), pp. 151–161.
[31] B. S. He and X. M. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective*, SIAM J. Imaging Sci. 5, (2012), pp. 119–149.
[32] B. S. He and X. M. Yuan, *On the $O(1/n)$ convergence rate of Douglas-Rachford alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700–709.
[33] B. S. He and X. M. Yuan, *On non-ergodic convergence rate of Douglas-Rachford alternating direction method of multipliers*, Numer. Math., 130 (2015), pp. 567–577.
[34] R. Henrion, A. Jourani, and J. V. Outrata, *On the calmness of a class of multifunctions*, SIAM J. Optim., 13 (2002), pp. 603–618.
[35] M. Y. Hong and Z. Q. Luo, *On the linear convergence of the alternating direction method of multipliers*, Math. Program., 162 (2017), pp. 165–199.
[36] Z. C. Lin, M. M. Chen, and Y. Ma, *The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices*, arXiv:1009.5055, 2010.
[37] Z. C. Lin, R. Liu, and H. Li, *Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning*, Mach. Learn., 99 (2015), pp. 287–325.
[38] P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979.
[39] G. S. Liu, J. J. Ye, and J. P. Zhu, *Partial exact penaty for mathematical programs with equilibrium constraints*, Set-Valued Anal., 16 (2008), pp. 785–804.
[40] B. Martinet, *Régularisation dinéquations variationnelles par approximations successives*, Rev. Française Informat. Recherche Opérationnelle, 4 (1970), pp. 154–159.
[41] R. D. Monteiro and B. F. Svaiter, *Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers*, SIAM J. Optim., 23 (2013), pp. 475–507.
[42] R. Nishihara, L. Lessard, B. Recht, A. Packard, and M. I. Jordan, *A General Analysis of the Convergence of ADMM*, arXiv:1502.02009, 2015.
[43] S. M. Robinson, *Stability theory for systems of inequalities. Part* I: *Linear systems*, SIAM J. Numer. Anal., 12 (1975), pp. 754–769.
[44] S. M. Robinson, *Stability theory for systems of inequalities, Part* II: *Differentiable nonlinear systems*, SIAM J. Numer. Anal., 13 (1976), pp. 497–513.
[45] S. M. Robinson, *Strongly regular generalized equations*, Math. Oper. Res., 5 (1980), pp. 43–62.
[46] R. T. Rockafellar and R. J. Wets, *Variational Analysis*, Springer, New York, 1998.
[47] R. Shefi, *Rate of Convergence Analysis for Convex Nonsmooth Optimization Algorithms*, Ph.D. thesis, Tel Aviv University, Israel, 2015.
[48] M. Tao and X. M. Yuan, *On the optimal linear convergence rate of a generalized proximal point algorithm*, J. Sci. Comput., 74 (2018), pp. 826–850.
[49] T. Valkonen, *Testing and Non-Linear Preconditioning of the Proximal Point Method*, arXiv:1703.05705, 2017.
[50] T. Valkonen and T. Pock, *Acceleration of the PDHGM on partially strongly convex functions*, J. Math. Imaging Vision, 59 (2017), pp. 394–414.
[51] X. F. Wang and X. M. Yuan, *The linearized alternating direction method of multipliers for Dantzig selector*, SIAM J. Sci. Comput., 34 (2012), pp. 2792–2811.
[52] X. F. Wang, J. J. Ye, X. M. Yuan, S. Z. Zeng, and J. Zhang, *Linear Convergence of the Proximal Gradient Method for Nonsmooth Nonconvex Optimization Problems via Variational Analysis*, manuscript, 2017.
[53] J. F. Yang and X. M. Yuan, *Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization*, Math. Comp., 82 (2013), pp. 301–329.
[54] W. H. Yang and D. R. Han, *Linear convergence of the alternating direction method of multipliers for a class of convex optimization problems*, SIAM J. Numer. Anal., 54 (2016), pp. 625–640.
[55] J. J. Ye and X. Y. Ye, *Necessary optimality conditions for optimization problems with variational inequality constraints*, Math. Oper. Res., 22 (1997), pp. 977–997.
[56] X. Q. Zhang, M. Burger, X. Bresson, and S. Osher, *Bregmanized nonlocal regularization for deconvolution and sparse reconstruction*, SIAM J. Imaging Sci., 3 (2010), pp. 253–276.