Journal of Optimization Theory and Applications manuscript No.
(will be inserted by the editor)

Pointwise and Ergodic Convergence Rates of a Variable Metric Proximal Alternating Direction Method of Multipliers

Max L.N. Gonçalves · Maicon Marques Alves · Jefferson G. Melo

Communicated by Hedy Attouch

Received: date / Accepted: date

Abstract In this paper, we obtain global pointwise and ergodic convergence rates for a variable metric proximal alternating direction method of multipliers for solving linearly constrained convex optimization problems. We first propose and study nonasymptotic convergence rates of a variable metric hybrid proximal extragradient framework for solving monotone inclusions. Then, the convergence rates for the former method are obtained essentially by showing that it falls within the latter framework. To the best of our knowledge, this is the first time that global pointwise (resp. pointwise and ergodic) convergence rates are obtained for the variable metric proximal alternating direction method of multipliers (resp. variable metric hybrid proximal extragradient framework).

Max L.N. Gonçalves (**Corresponding Author**)

IME, Universidade Federal de Goiás, Goiânia, GO 74001-970, BR. E-mail: maxlng@ufg.br

Maicon Marques Alves

Departamento de Matemática, Universidade Federal de Santa Catarina, Florianópolis, SC 88040-900, BR. E-mail: maicon.alves@ufsc.br

Jefferson G. Melo

IME, Universidade Federal de Goiás, Goiânia, GO 74001-970, BR. E-mail: jefferson@ufg.br

Keywords alternating direction method of multipliers · variable metric · pointwise and ergodic convergence rates · hybrid proximal extragradient method · convex program

AMS subject classification: 90C25 · 90C60 · 49M27 · 47H05 · 47J22 · 65K10.

1 Introduction

We consider two-block structured linearly constrained convex optimization problems. Problems of this type appear in different branches of applied mathematics, including machine learning, imaging and inverse problems. One of the most popular methods in nowadays research for finding approximate solutions of such problems is the alternating direction method of multipliers (ADMM) [1–3], for which many variants have been proposed and studied in the literature; see, e.g., [4–19].

In this paper, we obtain global ergodic and pointwise convergence rates for a variable metric proximal ADMM, which encompasses several recently studied ADMM variants. This variable metric version of the ADMM allows the use of variable metrics in both proximal and penalty terms, induced by self adjoint semipositive and positive definite linear operators, respectively.

Our study is done by first establishing global ergodic and pointwise convergence rates for a variable metric hybrid proximal extragradient (HPE) framework for finding zeroes of maximal monotone operators, and then by showing that the variable metric proximal ADMM can be seen as an instance of the latter framework. To the best of our knowledge, this is the first time that global pointwise (resp. pointwise and ergodic) convergence rates are obtained for the variable metric proximal ADMM (resp. variable metric HPE framework). Besides, our analysis allows degenerate metrics (induced by positive semidefinite linear operators) which makes the variable metric proximal ADMM (and variable metric HPE framework) more suitable for applications.

This paper is organized as follows. Section 2 contains four subsections. Subsection 2.1 contains our notation and basic results. Subsection 2.2 presents the problem of interest in this paper as well as the variable metric proximal ADMM and discusses some related works. The third and fourth subsections are devoted to discuss the hybrid proximal extragradient frameworks and present the main contributions of this paper, respectively. Section 3 introduces the variable metric HPE framework and presents its nonasymptotic pointwise and ergodic convergence rates, whose proofs are postponed to Appendix A. Section 4 contains two subsections. In Subsection 4.1, we formally state the variable metric proximal ADMM (5)–(7) and present its nonasymptotic pointwise and ergodic convergence rates. In Subsection 4.2, we prove the convergence rates of the variable metric proximal ADMM by viewing it as an instance of the variable metric HPE framework.

2 Preliminaries and the Main Contributions of this Paper

This section contains four subsections. The first subsection contains our notation and basic results. The second one presents the problem of interest in this paper as well as the variable metric proximal ADMM and discusses some related works. The third and fourth subsections are devoted to discuss the hybrid proximal extragradient frameworks and present the main contributions of this paper, respectively.

2.1 Basic Results and Notation

Let $\mathscr{Z}$ be a finite-dimensional real vector space endowed with inner product $\langle \cdot,\cdot\rangle_{\mathscr{Z}}$ and induced norm $\|\cdot\|_{\mathscr{Z}} := \sqrt{\langle \cdot,\cdot\rangle_{\mathscr{Z}}}$. Denote by $\mathscr{M}_+^{\mathscr{Z}}$ (resp. $\mathscr{M}_{++}^{\mathscr{Z}}$) the space of selfadjoint positive semidefinite (resp. definite) linear operators on $\mathscr{Z}$. Each element $M \in \mathscr{M}_+^{\mathscr{Z}}$ induces a symmetric bilinear form $\langle M(\cdot),\cdot\rangle_{\mathscr{Z}}$ on $\mathscr{Z}\times\mathscr{Z}$ and a seminorm $\|\cdot\|_{\mathscr{Z},M} := \sqrt{\langle M(\cdot),\cdot\rangle_{\mathscr{Z}}}$ on $\mathscr{Z}$. Since $\langle M(\cdot),\cdot\rangle_{\mathscr{Z}}$ is symmetric and bilinear, the following hold, for all $z,z' \in \mathscr{Z}$,

$$\langle z, Mz'\rangle \le \frac{1}{2}\|z\|^2_{\mathscr{Z},M} + \frac{1}{2}\|z'\|^2_{\mathscr{Z},M}, \tag{1}$$

$$\|z+z'\|^2_{\mathscr{Z},M} \le 2\left(\|z\|^2_{\mathscr{Z},M} + \|z'\|^2_{\mathscr{Z},M}\right). \tag{2}$$

On the other hand, each element $M \in \mathscr{M}_{++}^{\mathscr{Z}}$ induces an inner product $\langle M(\cdot),\cdot\rangle_{\mathscr{Z}}$ and a associated norm $\|\cdot\|_{\mathscr{Z},M} := \sqrt{\langle M(\cdot),\cdot\rangle_{\mathscr{Z}}}$ on $\mathscr{Z}$, etc. Let the partial order $\preceq$ on $\mathscr{M}_{+}^{\mathscr{Z}}$ be defined by

$$M \preceq N \iff N - M \in \mathscr{M}_{+}^{\mathscr{Z}}.$$

Next proposition, whose proof is omitted, will be useful in this paper.

Proposition 2.1 *Let $M,N \in \mathscr{M}_{+}^{\mathscr{Z}}$ and $c > 0$. If $M \preceq cN$, then*

$$\|\cdot\|_{\mathscr{Z},M} \leq \sqrt{c}\,\|\cdot\|_{\mathscr{Z},N} \;\text{ and }\; \|M(\cdot)\|_{\mathscr{Z}} \leq \sqrt{c\|N\|}\,\|\cdot\|_{\mathscr{Z},M}. \tag{3}$$

A set-valued mapping $T : \mathscr{Z} \rightrightarrows \mathscr{Z}$ is said to be *monotone* iff

$$\langle v - v', z - z'\rangle \geq 0 \quad \forall\, z, z' \in \mathscr{Z}, \forall\, v \in T(z), \forall\, v' \in T(z').$$

Moreover, T is *maximal monotone* iff it is monotone and, additionally, if S is a monotone operator such that $T(z) \subset S(z)$ for every $z \in \mathscr{Z}$, then $T = S$. The *inverse* operator $T^{-1} : \mathscr{Z} \rightrightarrows \mathscr{Z}$ of T is given by $T^{-1}(v) := \{z \in \mathscr{Z} \,:\, v \in T(z)\}$. Given $\varepsilon \geq 0$, the ε-enlargement $T^{\varepsilon} : \mathscr{Z} \rightrightarrows \mathscr{Z}$ of a set-valued mapping $T : \mathscr{Z} \rightrightarrows \mathscr{Z}$ is defined as

$$T^{\varepsilon}(z) := \{v \in \mathscr{Z} \,:\, \langle v - v', z - z'\rangle \geq -\varepsilon,\ \forall z' \in \mathscr{Z}, \forall\, v' \in T(z')\} \quad \forall z \in \mathscr{Z}.$$

Recall that the ε-subdifferential of a convex function $f : \mathscr{Z} \to \overline{\mathbb{R}}$ is defined by

$$\partial_{\varepsilon} f(z) := \{v \in \mathscr{Z} \,:\, f(z') \geq f(z) + \langle v, z' - z\rangle - \varepsilon\ \ \forall z' \in \mathscr{Z}\}$$

for every $z \in \mathscr{Z}$. When $\varepsilon = 0$, then $\partial_0 f(z)$ is denoted by $\partial f(z)$ and is called the *subdifferential* of f at z. If f is a proper, closed and convex function, then ∂f is maximal monotone [37].

The following result is a particular case of the *weak transportation formula* in [38, Theorem 2.3] combined with [39, Proposition 2(i)].

Theorem 2.1 *Suppose $T : \mathscr{Z} \rightrightarrows \mathscr{Z}$ is maximal monotone and let $\tilde{z}_i, r_i \in \mathscr{Z}$, for $i = 1,\dots,k$, be such that $r_i \in T(\tilde{z}_i)$ and define*

$$\tilde{z}_k^a := \frac{1}{k}\sum_{i=1}^{k} \tilde{z}_i\,, \quad r_k^a := \frac{1}{k}\sum_{i=1}^{k} r_i\,, \quad \varepsilon_k^a := \frac{1}{k}\sum_{i=1}^{k}\langle r_i, \tilde{z}_i - \tilde{z}_k^a\rangle.$$

Then, the following hold:

(a) $\varepsilon_k^a \geq 0$ *and* $r_k^a \in T^{\varepsilon_k^a}(\tilde{z}_k^a)$;

(b) *if, in addition, $T = \partial f$ for some proper, closed and convex function f, then $r_k^a \in \partial_{\varepsilon_k^a} f(\tilde{z}_k^a)$.*

2.2 Variable Metric Proximal Alternating Direction Method of Multipliers and Related Works

Consider the linearly constrained convex optimization problem

$$\min\ \{f(x) + g(y) :\ Ax + By = b\}, \tag{4}$$

where $f : \mathscr{X} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ and $g : \mathscr{Y} \to \overline{\mathbb{R}}$ are extended-real-valued proper, closed and convex functions, $\mathscr{X}, \mathscr{Y}$ and Γ are finite-dimensional real vector spaces, $A : \mathscr{X} \to \Gamma$ and $B : \mathscr{Y} \to \Gamma$ are linear operators, and $b \in \Gamma$.

In this paper, we obtain global ergodic and pointwise convergence rates for a variable metric proximal ADMM which can be described as follows: given an initial point $(x_0, y_0, \gamma_0) \in \mathscr{X} \times \mathscr{Y} \times \Gamma$ and a stepsize $\theta > 0$, compute a sequence $\{(x_k, y_k, \gamma_k)\}$, recursively, by

$$x_k \in \arg\min_{x\in\mathscr{X}} \left\{ f(x) - \langle \gamma_{k-1}, Ax\rangle_{\mathscr{X}} + \frac{1}{2}\|Ax + By_{k-1} - b\|^2_{\Gamma,H_k} + \frac{1}{2}\|x - x_{k-1}\|^2_{\mathscr{X},R_k} \right\}, \tag{5}$$

$$y_k \in \arg\min_{y\in\mathscr{Y}} \left\{ g(y) - \langle \gamma_{k-1}, By\rangle_{\mathscr{Y}} + \frac{1}{2}\|Ax_k + By - b\|^2_{\Gamma,H_k} + \frac{1}{2}\|y - y_{k-1}\|^2_{\mathscr{Y},S_k} \right\}, \tag{6}$$

$$\gamma_k = \gamma_{k-1} - \theta H_k\left(Ax_k + By_k - b\right), \tag{7}$$

where H_k, R_k and S_k are selfadjoint linear operators such that H_k is positive definite and R_k and S_k are positive semidefinite, and $\|\cdot\|^2_{\Gamma,H_k} := \langle H_k(\cdot), \cdot\rangle_{\Gamma}$, etc.

Remark 2.1 (i) An usual choice for the linear operator H_k is $\beta_k I$, where $\beta_k > 0$ plays the role of a penalty parameter.

(ii) The proximal terms in (5) and (6) defined by R_k and S_k, respectively, may have different roles. Namely, they can be used to regularize the subproblems in (5) and (6), making them strongly convex (when R_k and S_k are positive definite operators) and hence admitting unique solutions. Moreover, by a careful choice of these operators, subproblems (5) and (6) may become much easier to solve or even have closed-form solutions; for instance, similarly to [18], if $H_k = \beta_k I$, then the choices $R_k = \alpha_k I - \beta_k A^*A$ with $\alpha_k \geq \beta_k \|A^*A\|$ and $S_k = s_k I - \beta_k B^*B$ with $s_k \geq \beta_k \|B^*B\|$ eliminate the presence of quadratic forms associated to A^*A and B^*B in (5) and (6), respectively, and hence these subproblems become proximal type.

(iii) The variable metric proximal ADMM (5)–(7) can be seen as a class of ADMM variants, depending on the choices of the linear operators H_k, R_k and S_k. Namely,

(1) by taking $H_k = \beta I$ with $\beta > 0$, $R_k = 0$, $S_k = 0$ and $\theta = 1$, it reduces to the standard ADMM, whose ergodic convergence rate was established in [20] by showing that it is a special instance of the HPE framework [21];

(2) the ADMM in [15] (related to the Uzawa method [22]) consists of taking $H_k = \beta I$ with $\beta > 0$, R_k constant, $S_k = 0$ and $\theta = 1$. Pointwise and ergodic convergence rates for this variant were obtained in [15, 23];

(3) the AD-PMM proposed in [18] for solving (4) with $B = -I$ and $b = 0$ corresponds to taking $H_k = \beta I$ with $\beta > 0$, R_k and S_k constant, and $\theta = 1$. Pointwise and ergodic convergence rates for the AD-PMM were established in [18]. It is worth pointing out that the well-known Chambolle-Pock primal-dual algorithm proposed in [6] is a special case of AD-PMM (see [18, Proposition 3.1]);

(4) the proximal ADMM consists of choosing $H_k = \beta I$ with $\beta > 0$, R_k and S_k constant. This method has been studied by many authors; see, for instance [8, 24–26], where convergence rates are analyzed;

(5) by choosing $H_k = \beta_k I$, $R_k = 0$ and $S_k = 0$, it corresponds to a variable penalty parameter ADMM, for which an asymptotic convergence analysis was considered in [27–29];

(6) the variable metric proximal ADMM (5)–(7) with R_k and S_k positive definite is closely related to the method studied in [14,30] for solving (point-to-point) continuous and monotone variational inequalities (in the setting of problem (4), it demands f and g to be continuously differentiable). We mention that, contrary to our analysis, the latter references consider the stepsize $\theta = 1$ and do not present nonasymptotic convergence rates;

(7) by letting $H_k = \beta I$, $\beta > 0$, and $\theta = 1$, the resulting method becomes similar to Algorithm 7 in [31] with $h = 0$, where a composite structure of f is considered and ergodic convergence rates were obtained under the additional conditions that $B = I$ in (4) and the dual solution set of (4) be bounded;

(8) the instance of the variable metric proximal ADMM for solving (4) with $B = -I$ and $b = 0$ consisting of choosing $\theta = 1$, $H_k = \beta_k I$, $R_k = \tau_k^{-1} I - \beta_k A^* A$ with $\tau_k \beta_k \|A^* A\| \leq 1$, and $S_k = 0$ can be seen as a variant of the Chambolle-Pock primal-dual algorithm [6, Algorithm 1] in which the parameters τ_k and β_k can vary along the iterations. The proof of this fact is similar to [18, Proposition 3.1], where it is proved that [6, Algorithm 1] is an instance of the AD-PMM (see the third comment above). Other variants of the Chambolle-Pock algorithm in which the parameters can vary along the iterations can be found, for instance, in [6, 19], where the authors showed that, under some additional assumptions, careful choices of these parameters lead to accelerated versions of the method. However, the assumption on the variable metrics considered here (see (10)) seems to be quite restrictive in order to include the aforementioned accelerated schemes in our setting.

2.3 Variable Metric Hybrid Proximal Extragradient Frameworks

The variable metric HPE framework proposed in this work is a generalization of a special instance of the HPE framework [21] allowing variations in the metric (induced by positive semidefinite linear operators) along the iterations. The iteration complexity of the HPE framework was first analyzed in [32] and subsequently applied to the study of several methods; see, for example, [20, 33–35]. An inexact variable metric proximal point type method was proposed in [36] but, contrary to our variable metric HPE framework, it demands the metrics to be nondegenerate (induced by invertible linear operators). Moreover, the convergence analysis presented in [36] does not include nonasymptotic convergence rates.

2.4 The Main Contributions of this Paper

We obtain an $\mathscr{O}(1/k)$ global convergence rate for an ergodic sequence associated to the variable metric proximal ADMM (5)–(7) with $\theta \in]0,(\sqrt{5}+1)/2[$, which provides, for given tolerances $\rho,\varepsilon > 0$, triples $(x,y,\tilde{\gamma})$, (r_x,r_y,r_γ) and scalars $\varepsilon_x,\varepsilon_y \geq 0$ such that

$$\begin{aligned} r_x \in \partial_{\varepsilon_x} f(x) - A^*\tilde{\gamma}, \quad r_y \in \partial_{\varepsilon_y} g(y) - B^*\tilde{\gamma}, \quad r_\gamma = Ax + By - b, \\ \sqrt{\|r_x\|^2_{\mathscr{X}} + \|r_y\|^2_{\mathscr{Y}} + \|r_\gamma\|^2_{\Gamma}\}} \leq \rho, \qquad \varepsilon_x + \varepsilon_y \leq \varepsilon, \end{aligned} \tag{8}$$

in at most $\mathscr{O}\left(\max\left\{\lceil d_0/\rho\rceil, \lceil d_0^2/\varepsilon\rceil\right\}\right)$ iterations, where d_0 is a scalar measuring the quality of the initial point. Moreover, we establish an $\mathscr{O}(1/\sqrt{k})$ pointwise convergence rate in which the inclusions in (8) are strengthened, in the sense that $\varepsilon_x = \varepsilon_y = 0$, and the bound on the number of iterations becomes $\mathscr{O}\left(\lceil d_0^2/\rho^2\rceil\right)$.

3 A Variable Metric HPE Framework

Consider the monotone inclusion problem

$$0 \in T(z), \tag{9}$$

where $\mathscr{Z}$ is a finite-dimensional inner product real vector space and $T : \mathscr{Z} \rightrightarrows \mathscr{Z}$ is maximal monotone. Assume that the solution set $T^{-1}(0)$ of (9) is nonempty.

In this section, we propose a variable metric hybrid proximal extragradient (HPE) framework for solving (9) and analyze its nonasymptotic convergence rates. The proposed framework finds its roots in the hybrid proximal extragradient (HPE) framework of [21], for which the iteration complexity was recently obtained in [32]. Our main results on pointwise and ergodic convergence rates for the variable metric HPE framework are presented in Theorems 3.1 and 3.2, respectively. In Section 4, we will show how the variable metric HPE framework can be used to analyze the nonasymptotic convergence of a variable metric proximal ADMM for solving linearly constrained convex optimization problems. This technique was first considered in [20] and subsequently in [12, 25].

The metrics used in the variable metric HPE are defined by a nonnull sequence $\{M_k\}_{k\geq 0} \subset \mathscr{M}_+^{\mathscr{Z}}$ satisfying the following condition:

(C1) there exist $0 \leq C_S < \infty$ and $\{c_k\} \subset [0,\infty[$ such that

$$\sum_{i=0}^{k} c_i \leq C_S, \qquad \frac{1}{1+c_k} M_k \preceq M_{k+1} \preceq (1+c_k) M_k \quad \forall\, k \geq 0. \tag{10}$$

Remark 3.1 The above assumption (which is similar to condition (1.4) in [36]) is satisfied, for instance, if the sequence $\{M_k\}_{k\geq 0}$ is taken to be constant and $c_k \equiv 0$, in which case one can choose $C_S = 0$.

It is easy to check that condition **C1** implies the existence of a constant $C_P > 0$ such that $\{c_k\}_{k\geq 0}$ and $\{M_k\}_{k\geq 0}$ satisfy

$$\prod_{i=0}^{k} (1+c_i) \leq C_P \quad \text{and} \quad M_j \preceq C_P M_k, \quad \forall\, j,k \geq 0. \tag{11}$$

We now state the variable metric HPE framework.

Variable metric HPE framework

(0) Let $z_0 \in \mathscr{Z}$, $\eta_0 \in \mathbb{R}_+$, and $\sigma \in [0,1[$ be given. Let $\{M_k\}_{k\geq 0} \subset \mathscr{M}_+^{\mathscr{Z}}$ be a nonnull sequence satisfying condition **C1**, and set $k = 1$.

(1) Find $(z_k, \tilde{z}_k, \eta_k) \in \mathscr{Z} \times \mathscr{Z} \times \mathbb{R}_+$ such that

$$r_k := M_k(z_{k-1} - z_k) \in T(\tilde{z}_k), \tag{12}$$

$$\|z_k - \tilde{z}_k\|^2_{\mathscr{Z},M_k} + \eta_k \leq \sigma \|z_{k-1} - \tilde{z}_k\|^2_{\mathscr{Z},M_k} + \eta_{k-1}. \tag{13}$$

(2) Set $k \leftarrow k+1$ and go to step 1.

Remark 3.2 (i) Letting $M_k \equiv I$ and $\eta_k \equiv 0$ in (12) and (13), respectively, we find that the sequences $\{z_k\}$, $\{\tilde{z}_k\}$ and $\{r_k\}$ satisfy

$$r_k \in T(\tilde{z}_k), \quad \|r_k + \tilde{z}_k - z_{k-1}\|^2_{\mathscr{Z}} \leq \sigma \|\tilde{z}_k - z_{k-1}\|^2_{\mathscr{Z}},$$
$$z_k = z_{k-1} - r_k,$$

which is to say that in this case the variable metric HPE framework reduces to a special case of the HPE framework (see pp. 2763 in [32]) with $\lambda_k \equiv 1$ (in the notation of [32]) or, in other words, the variable metric HPE framework is a generalization of a special case of the HPE framework in which variations in the metric are allowed along the iterations.

(ii) If the sequence $\{M_k\}_{k\geq 0}$ is taken to be constant, then the variable metric HPE framework reduces to a special case of the NE-HPE framework studied in [25].

(iii) We also mention that a variable metric inexact proximal point method with relative error tolerance was proposed in [36] but, contrary to our framework, the method of [36] demands that every operator M_k must be positive definite. Moreover, the convergence analysis presented in [36] does not include nonasymptotic convergence rates. The fact that the variable metric HPE framework allows positive semidefinite operators M_k will be crucial for viewing the variable metric proximal ADMM of Section 4 as a special instance of it.

In the remaining part of this section, we present pointwise and ergodic convergence rates for the variable metric HPE framework. These results will depend on the quantity:

$$d_0 := \inf\{\|z^* - z_0\|_{\mathscr{Z},M_0} \,:\, z^* \in T^{-1}(0)\}, \tag{14}$$

which measures the “quality” of the initial guess $z_0 \in \mathscr{Z}$ in the variable metric HPE framework with respect to the solution set $T^{-1}(0)$.

For technical reasons and for the convenience of the reader, the proofs of the next two theorems will be given in Appendix A. We mention that these proofs follow the same lines (although they are far from being a direct consequence) of [25].

Theorem 3.1 (Pointwise convergence rate of the variable metric HPE framework) *Let $\{\tilde{z}_k\}$ and $\{r_k\}$ be generated by the* variable metric HPE framework. *Then, for every $k \geq 1$, $r_k \in T(\tilde{z}_k)$ and there exists $i \leq k$ such that*

$$\|r_i\|_{\mathscr{X}} \leq \left(\frac{\left[2(1+\sigma)C_P(d_0^2+\eta_0)+2(1-\sigma)\eta_0\right] C_P\|M_0\|}{(1-\sigma)k} \right)^{1/2}, \tag{15}$$

where M_0, C_P and d_0 are as in step 0 of the variable metric HPE framework, (11) *and* (14)*, respectively.*

Remark 3.3 (i) If $c_k \equiv 0$ in condition **C1** (in which case $M_k \equiv M_0$), then the upper bound in (15) with $C_S = 0$ and $C_P = 1$ reduces essentially to a special case of [25, Theorem 3.3(a)] (with $\lambda_k \equiv 1, \varepsilon_k \equiv 0$ and $d(w)_z(z') = (1/2)\|z-z'\|^2$). Additionally, if $M_0 = I$ and $\eta_0 = 0$, then the bound (15) becomes similar to the corresponding one in [32, Theorem 4.4(a)].

(ii) For a given tolerance $\rho > 0$, Theorem 3.1 ensures that there exists an index

$$i = \mathscr{O}\left(\left\lceil \frac{C_P^2\|M_0\|(d_0^2+\eta_0)}{\rho^2} \right\rceil \right) \tag{16}$$

such that

$$r_i \in T(\tilde{z}_i) \text{ and } \|r_i\|_{\mathscr{X}} \leq \rho. \tag{17}$$

In this case, $\tilde{z}_i \in \mathscr{X}$ can be interpreted as a ρ-approximate solution of (9) with residual $r_i \in \mathscr{X}$ (see, e.g., [32] for the definition of a related concept).

Before presenting the ergodic convergence of the variable metric HPE framework, let us define the ergodic sequences $\{\tilde{z}_k^a\}$, $\{r_k^a\}$ and $\{\varepsilon_k^a\}$ associated to $\{\tilde{z}_k\}$ and $\{r_k\}$ as follows:

$$\tilde{z}_k^a := \frac{1}{k}\sum_{i=1}^k \tilde{z}_i, \quad r_k^a := \frac{1}{k}\sum_{i=1}^k r_i, \quad \varepsilon_k^a := \frac{1}{k}\sum_{i=1}^k \langle r_i, \tilde{z}_i - \tilde{z}_k^a\rangle. \tag{18}$$

Theorem 3.2 (Ergodic convergence rate of the variable metric HPE framework) *Let* $\{\tilde{z}_k^a\}$, $\{r_k^a\}$ *and* $\{\varepsilon_k^a\}$ *be given as in* (18). *Let also* M_0, C_S, C_P *and* d_0 *be as in step 0 of the* variable metric HPE framework*,* (10)*,* (11) *and* (14)*, respectively. Then, for every* $k \geq 1$*, we have* $r_k^a \in T^{\varepsilon_k^a}(\tilde{z}_k^a)$ *and*

$$\|r_k^a\|_{\mathscr{X}} \leq \frac{\mathscr{E}\sqrt{(d_0^2+\eta_0)\|M_0\|}}{k}, \tag{19}$$

$$0 \leq \varepsilon_k^a \leq \frac{\widehat{\mathscr{E}}(d_0^2+\eta_0)}{k}, \tag{20}$$

where $\mathscr{E} := \left((1+C_S)(1+\sqrt{C_P})C_P + C_S C_P^2\right)$ *and* $\widehat{\mathscr{E}} := 2C_P(1+C_S)\left[\sigma C_P/(1-\sigma)+2(1+C_P)\right]$.

Remark 3.4 (i) Similarly to Remark 3.3(i), Theorem 3.2 is also related to [25, Theorem 3.4] and [32, Theorem 4.7].

(ii) For given tolerances $\rho, \varepsilon > 0$, Theorem 3.2 ensures that in at most

$$\mathscr{O}\left(\left\lceil (1+C_S)C_p^2 \max\left\{\frac{\|M_0\|\sqrt{d_0^2+\eta_0}}{\rho}, \frac{d_0^2+\eta_0}{\varepsilon}\right\}\right\rceil\right) \tag{21}$$

iterations there hold

$$r_k^a \in T^{\varepsilon_k^a}(\tilde{z}_k^a), \;\; \|r_k^a\|_{\mathscr{X}} \leq \rho \;\text{ and }\; \varepsilon_k^a \leq \varepsilon. \tag{22}$$

Note that (21), in terms of the dependence on $\rho > 0$, is better than the bound in (16) by a factor of $\mathscr{O}(\rho)$ but, on the other hand, since ε_k^a can be strictly positive, the inclusion in (22) is potentially weaker than the one in (17).

4 A Variable Metric Proximal Alternating Direction Method of Multipliers

This section contains two subsections. In Subsection 4.1, we formally state the variable metric proximal ADMM (5)–(7) and present its nonasymptotic convergence rates. The main results are Theorems 4.1 and 4.2, in which pointwise and ergodic convergence rates are obtained, respectively. The proofs of the latter theorems are discussed separately in Subsection 4.2 by viewing the method as an instance of the variable metric HPE framework and by applying the results of Section 3.

4.1 Variable metric proximal ADMM and its convergence rates

Let $\mathscr{X}$, $\mathscr{Y}$ and Γ be finite-dimensional real inner product vector spaces. Consider the convex optimization problem (4), i.e.,

$$\min \ \{f(x)+g(y) : \ Ax+By=b\}, \tag{23}$$

where the following assumptions are assumed to hold:

(O1) $f : \mathscr{X} \to \overline{\mathbb{R}}$ and $g : \mathscr{Y} \to \overline{\mathbb{R}}$ are proper, closed and convex functions;

(O2) $A : \mathscr{X} \to \Gamma$ and $B : \mathscr{Y} \to \Gamma$ are linear operators and $b \in \Gamma$;

(O3) the solution set of (23) is nonempty.

Under the above assumptions and standard constraint qualifications (see, e.g., [40, Corollaries 28.2.2 and 28.3.1]), a vector $(x^*,y^*) \in \mathscr{X} \times \mathscr{Y}$ is a solution of (23) iff there exists a (Lagrange multiplier) $\gamma^* \in \Gamma$ such that (x^*,y^*,γ^*) is a solution of

$$0 \in \partial f(x) - A^*\gamma, \quad 0 \in \partial g(y) - B^*\gamma, \quad Ax+By-b=0. \tag{24}$$

Motivated by the above statement, we define

$$\Omega^* := \{(x^*,y^*,\gamma^*) \in \mathscr{X} \times \mathscr{Y} \times \Gamma \ : \ (x^*,y^*,\gamma^*) \text{ is a solution of } (24)\}, \tag{25}$$

which is assumed to be nonempty.

The convergence rates of the variable metric proximal ADMM (stated below) for solving (23) will be obtained by viewing the optimization problem (23) as the monotone inclusion (24), which is associated to a certain maximal monotone operator (see (46)) in $\mathscr{X} \times \mathscr{Y} \times \Gamma$, and by applying the results of the previous section.

In order to state the variable metric proximal ADMM, we consider sequences $\{R_k\}_{k\geq 0} \subset \mathscr{M}_+^{\mathscr{X}}$, $\{S_k\}_{k\geq 0} \subset \mathscr{M}_+^{\mathscr{Y}}$ and $\{H_k\}_{k\geq 0} \subset \mathscr{M}_{++}^{\Gamma}$ satisfying the following condition:

(C2) there exist $0 \le C_S < \infty$ and $\{c_k\}_{k\ge 0} \subset [0,1]$ such that $\{c_k\}_{k\ge 0}$, $\{Q_{k,1} := R_k\}_{k\ge 0}$, $\{Q_{k,2} := S_k\}_{k\ge 0}$ and $\{Q_{k,3} := H_k\}_{k\ge 0}$ satisfy

$$\sum_{i=0}^{k} c_i \le C_S, \quad \frac{1}{1+c_k} Q_{k,j} \preceq Q_{k+1,j} \preceq (1+c_k) Q_{k,j} \quad \forall k \ge 0, \;\; j = 1,2,3. \tag{26}$$

Analogously to condition (11), condition **C2** implies the existence of $C_P > 0$ such that $\{c_k\}_{k\ge 0}$ satisfies

$$\prod_{i=0}^{k} (1+c_i) \le C_P \qquad \forall k \ge 0. \tag{27}$$

We mention that condition **C2** is similar to Condition C in [14] but, contrary to the latter reference, none of the operators R_k and S_k is assumed to be positive definite.

Variable metric proximal ADMM

(0) Let $(x_0, y_0, \gamma_0) \in \mathscr{X} \times \mathscr{Y} \times \Gamma$ and $\theta \in]0, (\sqrt{5}+1)/2[$ be given. Consider sequences $\{R_k\}_{k\ge 0} \subset \mathscr{M}_+^{\mathscr{X}}$, $\{S_k\}_{k\ge 0} \subset \mathscr{M}_+^{\mathscr{Y}}$, and $\{H_k\}_{k\ge 0} \subset \mathscr{M}_{++}^{\Gamma}$ satisfying condition **C2**, and set $k = 1$.

(1) Compute an optimal solution $x_k \in \mathscr{X}$ of the subproblem

$$\min_{x\in\mathscr{X}} \left\{ f(x) - \langle \gamma_{k-1}, Ax \rangle_{\mathscr{X}} + \frac{1}{2} \|Ax + By_{k-1} - b\|^2_{\Gamma, H_k} + \frac{1}{2} \|x - x_{k-1}\|^2_{\mathscr{X}, R_k} \right\} \tag{28}$$

and compute an optimal solution $y_k \in \mathscr{Y}$ of the subproblem

$$\min_{y\in\mathscr{Y}} \left\{ g(y) - \langle \gamma_{k-1}, By \rangle_{\mathscr{Y}} + \frac{1}{2} \|Ax_k + By - b\|^2_{\Gamma, H_k} + \frac{1}{2} \|y - y_{k-1}\|^2_{\mathscr{Y}, S_k} \right\}. \tag{29}$$

(2) Set

$$\gamma_k = \gamma_{k-1} - \theta H_k \left(Ax_k + By_k - b \right), \tag{30}$$

$k \leftarrow k+1$, and go to step (1).

In the remaining part of this section, we present pointwise and ergodic convergence rates for the variable metric proximal ADMM. For this end, the following quantities will be needed:

$$m_0 := \max \left\{ \|R_0\|, \|B^* H_0 B + S_0\|, [1/\theta] \|H_0^{-1}\| \right\} \tag{31}$$

and

$$d_0 := \inf\left\{ \left(\|x_0 - x^*\|^2_{\mathscr{X},R_0} + \|y_0 - y^*\|^2_{\mathscr{Y},(B^*H_0B+S_0)} + \|\gamma_0 - \gamma^*\|^2_{\Gamma,\theta^{-1}H_0^{-1}} \right)^{1/2} : (x^*, y^*, \gamma^*) \in \Omega^* \right\}, \quad (32)$$

where Ω^* is defined in (25).

Next we present the two main results of this paper, whose proofs are given in Subsection 4.2.

Theorem 4.1 (Pointwise convergence rate of the variable metric proximal ADMM) *Let* $\{R_k\}$*,* $\{S_k\}$ *and* $\{H_k\}$ *be as in Step 0 of the* variable metric proximal ADMM*. Let* $\{(x_k, y_k, \gamma_k)\}$ *be generated by the* variable metric proximal ADMM *and define*

$$\tilde{\gamma}_k := \gamma_{k-1} - H_k(Ax_k + By_{k-1} - b) \quad \forall k \geq 1. \quad (33)$$

Then, for all $k \geq 1$*,*

$$\begin{pmatrix} r_{k,x} \\ r_{k,y} \\ r_{k,\gamma} \end{pmatrix} := \begin{pmatrix} R_k(x_{k-1} - x_k) \\ (B^*H_kB + S_k)(y_{k-1} - y_k) \\ \theta^{-1}H_k^{-1}(\gamma_{k-1} - \gamma_k) \end{pmatrix} \in \begin{pmatrix} \partial f(x_k) - A^*\tilde{\gamma}_k \\ \partial g(y_k) - B^*\tilde{\gamma}_k \\ Ax_k + By_k - b \end{pmatrix} \quad (34)$$

and there exists a parameter $\sigma_\theta \in]0,1[$ *such that, for some* $i \leq k$*,*

$$\sqrt{\|r_{i,x}\|^2_{\mathscr{X}} + \|r_{i,y}\|^2_{\mathscr{Y}} + \|r_{i,\gamma}\|^2_{\Gamma}} \leq \frac{d_0}{\sqrt{k}} \sqrt{\frac{[2(1+\sigma_\theta)C_P(1+\tau_\theta) + 2(1-\sigma_\theta)\tau_\theta]C_P m_0}{(1-\sigma_\theta)}}, \quad (35)$$

where $\tau_\theta := (8(\sigma_\theta + \theta - 1)\max\{1, \theta/(2-\theta)\})/(\theta\sqrt{\theta})$*, and* C_P*,* m_0*, and* d_0 *are as in* (27)*,* (31)*, and* (32)*, respectively.*

Remark 4.1 For a given tolerance $\rho > 0$, Theorem 4.1 guarantees the existence of triples $(x, y, \tilde{\gamma})$ and (r_x, r_y, r_γ) generated by the variable metric proximal ADMM such that

$$\begin{aligned} &r_x \in \partial f(x) - A^*\tilde{\gamma}, \quad r_y \in \partial g(y) - B^*\tilde{\gamma}, \quad r_\gamma = Ax + By - b, \\ &\sqrt{\|r_x\|^2_{\mathscr{X}} + \|r_y\|^2_{\mathscr{Y}} + \|r_\gamma\|^2_{\Gamma}} \leq \rho, \end{aligned} \quad (36)$$

in at most

$$\mathscr{O}\left(\left\lceil \frac{C_P^2 m_0 d_0^2}{\rho^2} \right\rceil\right) \tag{37}$$

iterations, where C_P, m_0 and d_0 are as in (27), (31) and (32), respectively. The triple $(x, y, \tilde{\gamma})$ in (36) can be seen as a ρ-approximate solution of the Lagrangian system (24) with residual (r_x, r_y, r_γ).

Remark 4.2 (i) Theorem 4.1, in particular, establishes pointwise convergence rates (unknown so far, up to our knowledge) for the ADMM variants described in comments 6 and 8 of Remark 2.1(iii).

(ii) As mentioned in the third comment of Remark 2.1(iii), Algorithm 2 in [18] is a special case of the variable metric proximal ADMM. In this case, the pointwise iteration-complexity bound in (37) is the same as the one that can be derived from [18, Section 5.3]. Moreover, although different termination criteria and approaches are used in the literature to analyze the other ADMMs described in Remark 2.1(iii), the pointwise iteration-complexity bounds obtained for them are, basically, as in (37); see, for example, [23, 24].

Before presenting the ergodic convergence of the variable metric proximal ADMM we need to introduce its associated ergodic sequences. Let $\{(x_k, y_k, \gamma_k)\}$ be generated by the variable metric proximal ADMM, let $\{\tilde{\gamma}_k\}$ and $\{(r_{k,x}, r_{k,y}, r_{k,\gamma})\}$ be defined as in (33) and (34), respectively, and let the *ergodic* sequences associated to them be defined by

$$(x_k^a, y_k^a) := \frac{1}{k}\sum_{i=1}^{k}(x_i, y_i), \quad \tilde{\gamma}_k^a := \frac{1}{k}\sum_{i=1}^{k}\tilde{\gamma}_i, \tag{38}$$

$$(r_{k,x}^a, r_{k,y}^a, r_{k,\gamma}^a) := \frac{1}{k}\sum_{i=1}^{k}(r_{i,x}, r_{i,y}, r_{i,\gamma}), \tag{39}$$

$$(\varepsilon_{k,x}^a, \varepsilon_{k,y}^a) := \frac{1}{k}\sum_{i=1}^{k}\left(\langle r_{i,x} + A^*\tilde{\gamma}_i, x_i - x_k^a\rangle_{\mathscr{X}},\ \langle r_{i,y} + B^*\tilde{\gamma}_i, y_i - y_k^a\rangle_{\mathscr{Y}}\right). \tag{40}$$

Theorem 4.2 (Ergodic convergence rate of the variable metric proximal ADMM) *Let* $\{(x_k^a, y_k^a)\}$, $\{\tilde{\gamma}_k^a\}$, $\{(r_{k,x}^a, r_{k,y}^a, r_{k,\gamma}^a)\}$ *and* $\{(\varepsilon_{k,x}^a, \varepsilon_{k,y}^a)\}$ *be the ergodic sequences defined as in* (38)–(40). *Let also* C_S, C_P,

m_0 and d_0 be as in (26), (27), (31) *and* (32), *respectively. Then, for all $k \geq 1$, there hold $\varepsilon^a_{k,x}$, $\varepsilon^a_{k,y} \geq 0$,*

$$\begin{pmatrix} r^a_{k,x} \\ r^a_{k,y} \\ r^a_{k,\gamma} \end{pmatrix} \in \begin{pmatrix} \partial f_{\varepsilon^a_{k,x}}(x^a_k) - A^*\tilde{\gamma}^a_k \\ \partial g_{\varepsilon^a_{k,y}}(y^a_k) - B^*\tilde{\gamma}^a_k \\ Ax^a_k + By^a_k - b \end{pmatrix} \tag{41}$$

and there exists a parameter $\sigma_\theta \in]0,1[$ such that

$$\sqrt{\|r^a_{k,x}\|^2_{\mathscr{X}} + \|r^a_{k,y}\|^2_{\mathscr{Y}} + \|r^a_{k,\gamma}\|^2_{\Gamma}} \leq \frac{\sqrt{(1+\tau_\theta)m_0}\mathscr{E}\, d_0}{k}, \tag{42}$$

$$\varepsilon^a_{k,x} + \varepsilon^a_{k,y} \leq \frac{(1+\tau_\theta)\widehat{\mathscr{E}} d_0^2}{k}, \tag{43}$$

where $\mathscr{E}$ and $\widehat{\mathscr{E}}$ are as in Theorem 3.2 with $\sigma = \sigma_\theta$ and τ_θ is as in Theorem 4.1.

Remark 4.3 Given tolerances $\rho, \varepsilon > 0$, Theorem 4.2 guarantees that there exist scalars $\varepsilon_x, \varepsilon_y \geq 0$ and triples $(x,y,\tilde{\gamma})$, (r_x, r_y, r_γ) generated by the variable metric proximal ADMM such that

$$\begin{aligned} &r_x \in \partial_{\varepsilon_x} f(x) - A^*\tilde{\gamma}, \quad r_y \in \partial_{\varepsilon_y} g(y) - B^*\tilde{\gamma}, \quad r_\gamma = Ax + By - b, \\ &\sqrt{\|r_x\|^2_{\mathscr{X}} + \|r_y\|^2_{\mathscr{Y}} + \|r_\gamma\|^2_{\Gamma}} \leq \rho, \qquad \varepsilon_x + \varepsilon_y \leq \varepsilon, \end{aligned} \tag{44}$$

in at most

$$\mathscr{O}\left(\left\lceil (1+C_S)C_P^2 \max\left\{\frac{d_0\sqrt{m_0}}{\rho}, \frac{d_0^2}{\varepsilon}\right\}\right\rceil\right) \tag{45}$$

iterations, where C_S, C_P, m_0 and d_0 are as in condition **C2**, (27), (31) and (32), respectively. Note that while the dependence on the tolerance ρ in (45) is better than the corresponding one in (37) by a factor of $\mathscr{O}(\rho)$, the inclusions in (44) are potentially weaker than the corresponding ones in (36). The triple $(x,y,\tilde{\gamma})$ in (44) can be seen as a (ρ,ε)-approximate solution of the Lagrangian system (24) with residual (r_x, r_y, r_γ).

Remark 4.4 (i) Theorem 4.2, in particular, establishes ergodic convergence rates (unknown so far, up to our knowledge) for the ADMM variants described in comments 6 and 7 of Remark 2.1(iii).

(ii) It can be easily seen that Algorithm 7 in [31] with $h=0$ is an instance of the variable metric proximal ADMM (see the seventh comment of Remark 2.1(iii)). For this variant, it can be derived from [31, Theorem 12] an ergodic iteration-complexity bound $\mathscr{O}(1/\rho)$ to obtain a ρ-approximate saddle point for the Lagrangian function associated to (23) (see also [18, Theorem 5.3] for a similar result when H_k, R_k and S_k are constant). Finally, we refer the reader to [15,20,25], where similar ergodic iteration-complexity bounds were obtained for other ADMM variants.

4.2 Proof of Theorems 4.1 and 4.2

The main goal of this subsection is to prove Theorems 4.1 and 4.2 by viewing the variable metric proximal ADMM as an instance of the variable metric HPE framework of Section 3 for solving (9) with $T:\mathscr{Z}\rightrightarrows\mathscr{Z}$ defined by

$$T(z) := \begin{pmatrix} \partial f(x) - A^*\gamma \\ \partial g(y) - B^*\gamma \\ Ax+By-b \end{pmatrix}, \qquad \forall z:=(x,y,\gamma)\in\mathscr{Z} \tag{46}$$

where $\mathscr{Z}:=\mathscr{X}\times\mathscr{Y}\times\Gamma$ is endowed with the inner product of vectors $z=(x,y,\gamma)$ and $z'=(x',y',\gamma')$:

$$\langle z,z'\rangle_{\mathscr{Z}} := \langle x,x'\rangle_{\mathscr{X}} + \langle y,y'\rangle_{\mathscr{Y}} + \langle \gamma,\gamma'\rangle_{\Gamma}. \tag{47}$$

The desired results will then follow essentially from Theorems 3.1 and 3.2, and from the identity

$$T^{-1}(0) = \Omega^*, \tag{48}$$

where $T^{-1}(0)$ and Ω^* are the solution sets defined in (9) and (25), respectively. The following linear operators will be needed in our analysis:

$$M_k := \begin{pmatrix} R_k & 0 & 0 \\ 0 & B^*H_kB+S_k & 0 \\ 0 & 0 & \theta^{-1}{H_k}^{-1} \end{pmatrix} : \mathscr{Z}\to\mathscr{Z} \qquad \forall k\geq 0, \tag{49}$$

where $\{R_k\}_{k\geq 0}$, $\{S_k\}_{k\geq 0}$ and $\{H_k\}_{k\geq 0}$ are given in step 0 of the variable metric proximal ADMM.

We begin by presenting a preliminary technical result.

Proposition 4.1 *Let* $\{(x_k, y_k, \gamma_k)\}$ *be generated by the* variable metric proximal ADMM *and let* $\{\tilde{\gamma}_k\}$ *be defined as in* (33)*. Let also* $\{M_k\}$ *be defined as in* (49)*. Then,*

$$M_k \begin{pmatrix} x_{k-1} - x_k \\ y_{k-1} - y_k \\ \gamma_{k-1} - \gamma_k \end{pmatrix} \in \begin{pmatrix} \partial f(x_k) - A^* \tilde{\gamma}_k \\ \partial g(y_k) - B^* \tilde{\gamma}_k \\ Ax_k + By_k - b \end{pmatrix} \qquad \forall k \geq 1. \tag{50}$$

Proof From the first order optimality conditions for (28) and (29), we obtain, respectively,

$$0 \in \partial f(x_k) - A^* \left(\gamma_{k-1} - H_k(Ax_k + By_{k-1} - b)\right) + R_k(x_k - x_{k-1}),$$

$$0 \in \partial g(y_k) - B^* (\gamma_{k-1} - H_k(Ax_k + By_k - b)) + S_k(y_k - y_{k-1}),$$

which, combined with (33), yields

$$R_k(x_{k-1} - x_k) \in \partial f(x_k) - A^* \tilde{\gamma}_k, \quad (B^* H_k B + S_k)(y_{k-1} - y_k) \in \partial g(y_k) - B^* \tilde{\gamma}_k. \tag{51}$$

On the other hand, (30) (with the assumption $H_k \in \mathscr{M}^{\Gamma}_{++}$) gives

$$\theta^{-1} H_k^{-1} (\gamma_{k-1} - \gamma_k) = Ax_k + By_k - b. \tag{52}$$

Using (49), (51) and (52) we obtain (50). □

The next lemma will allow us to use the main results of Section 3 for analyzing the nonasymptotic convergence of the variable metric proximal ADMM.

Lemma 4.1 *The sequence* $\{M_k\}_{k\geq 0}$ *defined in* (49)*, the scalar* C_S *and the sequence* $\{c_k\}$ *given in condition* **C2***, satisfy condition* **C1**.

Proof Note that the first condition in (26) is identical to the first one in (10). Now, note that the second condition in (26) combined with the (block) diagonal structure of M_k implies the second condition in (10).

□

The following two technical results will be used to prove that the variable metric proximal ADMM is an instance of the variable metric HPE framework.

Lemma 4.2 *Let* $\{(x_k, y_k, \gamma_k)\}$ *be generated by the* variable metric proximal ADMM. *Let* d_0 *and* $\{\tilde{\gamma}_k\}$ *be as in* (32) *and* (33)*, respectively. Let also* $\{S_k\}$ *and* $\{H_k\}$ *be as in Step 0 of the* variable metric proximal ADMM. *Then, the following hold:*

(a) *for any* $k \geq 1$*, we have*

$$\tilde{\gamma}_k - \gamma_k = \frac{1-\theta}{\theta}(\gamma_k - \gamma_{k-1}) + H_k B(y_k - y_{k-1}), \quad \tilde{\gamma}_k - \gamma_{k-1} = \frac{1}{\theta}(\gamma_k - \gamma_{k-1}) + H_k B(y_k - y_{k-1});$$

(b) *we have*

$$\frac{1}{2}\|y_1 - y_0\|^2_{\mathscr{Y},S_1} - \frac{1}{\sqrt{\theta}}\langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle_\Gamma \leq 4\max\left\{1, \frac{\theta}{2-\theta}\right\} d_0^2;$$

(c) *for any* $t > 0$ *and* $k \geq 2$*, we have*

$$\frac{2}{\theta}\langle \gamma_k - \gamma_{k-1} - (1-\theta)(\gamma_{k-1} - \gamma_{k-2}), B(y_k - y_{k-1})\rangle_\Gamma \geq \frac{2t - 1 - c_{k-1}}{t}\|y_k - y_{k-1}\|^2_{\mathscr{Y},S_k} - t\|y_{k-1} - y_{k-2}\|^2_{\mathscr{Y},S_{k-1}}.$$

Proof (a) This item follows trivially from (30) and (33).

(b) First note that

$$\begin{aligned} 0 &\leq \frac{1}{2}\left\|\frac{1}{\sqrt{\theta}}(\gamma_1 - \gamma_0) + H_1 B(y_1 - y_0)\right\|^2_{\Gamma, H_1^{-1}} \\ &= \frac{1}{2}\|\gamma_1 - \gamma_0\|^2_{\Gamma,\theta^{-1}H_1^{-1}} + \frac{1}{\sqrt{\theta}}\langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle_\Gamma + \frac{1}{2}\|B(y_1 - y_0)\|^2_{\Gamma,H_1}, \end{aligned}$$

which, combined with the property (2), yields, for all $z^* := (x^*, y^*, \gamma^*) \in \Omega^*$,

$$\begin{aligned} &\frac{1}{2}\|y_1 - y_0\|^2_{\mathscr{Y},S_1} - \frac{1}{\sqrt{\theta}}\langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle_\Gamma \\ &\leq \frac{1}{2}\left(\|y_1 - y_0\|^2_{\mathscr{Y},S_1} + \|\gamma_1 - \gamma_0\|^2_{\Gamma,\theta^{-1}H_1^{-1}} + \|B(y_1 - y_0)\|^2_{\Gamma,H_1}\right) \\ &\leq \|y_1 - y^*\|^2_{\mathscr{Y},S_1} + \|y_0 - y^*\|^2_{\mathscr{Y},S_1} + \|\gamma_1 - \gamma^*\|^2_{\Gamma,\theta^{-1}H_1^{-1}} \\ &\quad + \|\gamma_0 - \gamma^*\|^2_{\Gamma,\theta^{-1}H_1^{-1}} + \|B(y_1 - y^*)\|^2_{\Gamma,H_1} + \|B(y_0 - y^*)\|^2_{\Gamma,H_1}. \end{aligned}$$

Direct use of the above inequality and (49) yields

$$\frac{1}{2}\|y_1-y_0\|^2_{\mathscr{Y},S_1}-\frac{1}{\sqrt{\theta}}\langle B(y_1-y_0),\gamma_1-\gamma_0\rangle_\Gamma \le \|z_1-z^*\|^2_{\mathscr{Z},M_1}+\|z_0-z^*\|^2_{\mathscr{Z},M_1}, \tag{53}$$

where $z_0:=(x_0,y_0,\gamma_0)$ and $z_1:=(x_1,y_1,\gamma_1)$. On the other hand, from Proposition 4.1 and (49) with $k=1$, we have $r_1:=M_1(z_0-z_1)\in T(\tilde z_1)$, where T is given in (46). Using this fact, (48) and the monotonicity of T, we obtain $\langle \tilde z_1-z^*,r_1\rangle\ge 0$ for all $z^*=(x^*,y^*,z^*)\in\Omega^*$. Hence, from the latter inequality, Lemma A.1 with $(z,z_+,\tilde z)=(z_0,z_1,\tilde z_1)$ and $M=M_1$, we have, for all $z^*=(x^*,y^*,z^*)\in\Omega^*$,

$$\|z^*-z_0\|^2_{\mathscr{Z},M_1}\ge \|z^*-z_1\|^2_{\mathscr{Z},M_1}+\|z_0-\tilde z_1\|^2_{\mathscr{Z},M_1}-\|z_1-\tilde z_1\|^2_{\mathscr{Z},M_1}. \tag{54}$$

Note now that letting $\tilde z_1:=(x_1,y_1,\tilde\gamma_1)$, it follows from (49), item (a) and some direct calculations that

$$\begin{aligned}\|z_1-\tilde z_1\|^2_{\mathscr{Z},M_1}&=\|\gamma_1-\tilde\gamma_1\|^2_{\Gamma,\theta^{-1}H_1^{-1}}=\left\|\frac{1-\theta}{\theta}(\gamma_1-\gamma_0)+H_1B(y_1-y_0)\right\|^2_{\Gamma,\theta^{-1}H_1^{-1}}\\ &=\frac{(1-\theta)^2}{\theta^2}\|\gamma_1-\gamma_0\|^2_{\Gamma,\theta^{-1}H_1^{-1}}+\frac{2(1-\theta)}{\theta^2}\langle B(y_1-y_0),\gamma_1-\gamma_0\rangle_\Gamma+\frac{1}{\theta}\|B(y_1-y_0)\|^2_{\Gamma,H_1}.\end{aligned} \tag{55}$$

Moreover, using (49) with $k=1$ and item (a), we find

$$\begin{aligned}\|z_0-\tilde z_1\|^2_{\mathscr{Z},M_1}&=\|x_0-x_1\|^2_{\mathscr{X},R_1}+\|y_0-y_1\|^2_{\mathscr{Y},(B^*H_1B+S_1)}+\|\gamma_0-\tilde\gamma_1\|^2_{\Gamma,\theta^{-1}H_1^{-1}}\\ &\ge \|B(y_1-y_0)\|^2_{\Gamma,H_1}+\left\|\frac{1}{\theta}(\gamma_1-\gamma_0)+H_kB(y_1-y_0)\right\|^2_{\Gamma,\theta^{-1}H_1^{-1}}\\ &\ge \frac{1+\theta}{\theta}\|B(y_1-y_0)\|^2_{\Gamma,H_1}+\frac{1}{\theta^2}\|\gamma_1-\gamma_0\|^2_{\Gamma,\theta^{-1}H_1^{-1}}+\frac{2}{\theta^2}\langle B(y_1-y_0),\gamma_1-\gamma_0\rangle_\Gamma.\end{aligned} \tag{56}$$

Combining the previous two estimates, we obtain

$$\begin{aligned}&\|z_0-\tilde z_1\|^2_{\mathscr{Z},M_1}-\|z_1-\tilde z_1\|^2_{\mathscr{Z},M_1}\\ &\qquad\ge \frac{2-\theta}{\theta}\|\gamma_1-\gamma_0\|^2_{\Gamma,\theta^{-1}H_1^{-1}}+\frac{2}{\theta}\langle B(y_1-y_0),\gamma_1-\gamma_0\rangle_\Gamma+\|B(y_1-y_0)\|^2_{\Gamma,H_1}\\ &\qquad=\frac{1-\theta}{\theta}\|\gamma_1-\gamma_0\|^2_{\Gamma,\theta^{-1}H_1^{-1}}+\left\|\frac{H_1^{-1/2}(\gamma_1-\gamma_0)}{\theta}+H_1^{1/2}B(y_1-y_0)\right\|^2_\Gamma\\ &\qquad\ge \frac{1-\theta}{\theta}\|\gamma_1-\gamma_0\|^2_{\Gamma,\theta^{-1}H_1^{-1}}.\end{aligned}$$

If $\theta \in]0,1]$, then the last inequality implies that

$$\|z_1 - \tilde{z}_1\|^2_{\mathscr{Z},M_1} \leq \|z_0 - \tilde{z}_1\|^2_{\mathscr{Z},M_1}. \tag{57}$$

Now, if $\theta \in]1,(\sqrt{5}+1)/2[$, we have

$$\begin{aligned}
\|z_1 - \tilde{z}_1\|^2_{\mathscr{Z},M_1} - \|z_0 - \tilde{z}_1\|^2_{\mathscr{Z},M_1} &\leq \frac{\theta-1}{\theta}\|\gamma_1 - \gamma_0\|^2_{\Gamma,\theta^{-1}H_1^{-1}} \\
&\leq \frac{2(\theta-1)}{\theta}\left(\|\gamma_1 - \gamma^*\|^2_{\Gamma,\theta^{-1}H_1^{-1}} + \|\gamma_0 - \gamma^*\|^2_{\Gamma,\theta^{-1}H_1^{-1}}\right) \\
&\leq \frac{2(\theta-1)}{\theta}\left[\|z_0 - z^*\|^2_{\mathscr{Z},M_1} + \|z_1 - z^*\|^2_{\mathscr{Z},M_1}\right]
\end{aligned}$$

where the second inequality is due to property (2), and the last inequality is due to (49) and the definitions of z_0, z_1 and z^*. Hence, combining the last estimative with (54), we obtain

$$\|z_1 - z^*\|^2_{\mathscr{Z},M_1} \leq \frac{\theta}{2-\theta}\left(1 + \frac{2(\theta-1)}{\theta}\right)\|z_0 - z^*\|^2_{\mathscr{Z},M_1} = \frac{3\theta-2}{2-\theta}\|z_0 - z^*\|^2_{\mathscr{Z},M_1}.$$

Thus, it follows from (54), (57) and the last inequality that

$$\|z_1 - z^*\|^2_{\mathscr{Z},M_1} \leq \max\left\{1, \frac{3\theta-2}{2-\theta}\right\}\|z_0 - z^*\|^2_{\mathscr{Z},M_1}. \tag{58}$$

Since, $M_1 \preceq (1+c_0)M_0 \preceq 2M_0$ (see condition **C2** and Lemma 4.1), the desired inequality follows from (53), (58), and definition of d_0 in (32).

(c) Using the first order optimality condition for (29), (33) and item (a), we find, for every $k \geq 1$,

$$\partial g(y_k) \ni B^*(\tilde{\gamma}_k - H_k B(y_k - y_{k-1})) - S_k(y_k - y_{k-1}) = \frac{1}{\theta}B^*(\gamma_k - (1-\theta)\gamma_{k-1}) - S_k(y_k - y_{k-1}).$$

For any $k \geq 2$, using the above inclusion with $k \leftarrow k$ and $k \leftarrow k-1$, the monotonicity of ∂g and the property (1), we find

$$\begin{aligned}
&\frac{1}{\theta}\langle B^*(\gamma_k - \gamma_{k-1}) - (1-\theta)B^*(\gamma_{k-1} - \gamma_{k-2}), y_k - y_{k-1}\rangle_{\mathscr{Y}} \\
&\geq \langle S_k(y_k - y_{k-1}), y_k - y_{k-1}\rangle_{\mathscr{Y}} - \langle S_{k-1}(y_{k-1} - y_{k-2}), y_k - y_{k-1}\rangle_{\mathscr{Y}} \\
&\geq \|y_k - y_{k-1}\|^2_{\mathscr{Y},S_k} - \frac{1}{2t}\|y_k - y_{k-1}\|^2_{\mathscr{Y},S_{k-1}} - \frac{t}{2}\|y_{k-1} - y_{k-2}\|^2_{\mathscr{Y},S_{k-1}}, \\
&\geq \left(1 - \frac{1+c_{k-1}}{2t}\right)\|y_k - y_{k-1}\|^2_{\mathscr{Y},S_k} - \frac{t}{2}\|y_{k-1} - y_{k-2}\|^2_{\mathscr{Y},S_{k-1}},
\end{aligned}$$

where the last inequality is due to Proposition 2.1 and condition **C2**, and so the proof of the lemma follows. □

Lemma 4.3 *For every $\theta \in]0,(\sqrt{5}+1)/2[$, there exists a scalar $\sigma_\theta \in]0,1[$ such that, for any $\sigma \in [\sigma_\theta,1)$, the matrix*

$$M_\theta(\sigma) = \begin{bmatrix} \sigma(1+\theta)-1 & (\sigma+\theta-1)(1-\theta) \\ (\sigma+\theta-1)(1-\theta) & \sigma-(1-\theta)^2 \end{bmatrix}$$

is symmetric and positive definite, and

$$\max\{(1-\theta)^2, 1-\theta, 1/(1+\theta)\} < \sigma, \quad \frac{(\sigma+\theta-1)\left(4-2\sqrt{2}\right)}{\sqrt{2}\theta} < \sigma. \tag{59}$$

Proof Since the matrix $M_\theta(\sigma)$ is symmetric, the proof is immediate by noting that for $\sigma = 1$ and for every $\theta \in]0,(\sqrt{5}+1)/2[$, $M_\theta(\sigma)$ is definite positive and (59) trivially holds. □

Next we show that the variable metric proximal ADMM can be regarded as an instance of the variable metric HPE framework.

Proposition 4.2 *Let $\{(x_k,y_k,\gamma_k)\}$ be generated by the* variable metric proximal ADMM *and let $\{\tilde{\gamma}_k\}$ and $\{M_k\}$ be defined as in* (33) *and* (49)*, respectively. Let also d_0, T, σ_θ and τ_θ be as in* (32), (46)*, Lemma 4.3, and Theorem 4.1, respectively. Define $z_0 := (x_0,y_0,\gamma_0)$, $\eta_0 := \tau_\theta d_0^2$ and, for all $k \geq 1$,*

$$z_k := (x_k,y_k,\gamma_k), \quad \tilde{z}_k := (x_k,y_k,\tilde{\gamma}_k), \quad r_k := M_k(z_{k-1}-z_k), \tag{60}$$

$$\eta_k := \frac{\sigma_\theta-(\theta-1)^2}{\theta^2}\|\gamma_k-\gamma_{k-1}\|^2_{\Gamma,\theta^{-1}H_k^{-1}} + \frac{\sqrt{2}(\sigma_\theta+\theta-1)}{\theta}\|y_k-y_{k-1}\|^2_{\mathscr{Y},S_k}. \tag{61}$$

Then, for all $k \geq 1$,

$$\begin{aligned} & r_k \in T(\tilde{z}_k), \\ & \|z_k-\tilde{z}_k\|^2_{\mathscr{Z},M_k} + \eta_k \leq \sigma_\theta\|z_{k-1}-\tilde{z}_k\|^2_{\mathscr{Z},M_k} + \eta_{k-1}. \end{aligned} \tag{62}$$

As a consequence, the variable metric proximal ADMM *falls within the* variable metric HPE *framework (with input z_0, η_0 and $\sigma = \sigma_\theta$) for solving* (9) *with T as in* (46).

Proof First note that the inclusion in (62) follows from (46), (50) and the definitions of z_k, $\tilde{z}_k$ and r_k in (60). Now, using (47), (49), (60) and some direct calculations, we obtain

$$\|z_{k-1}-\tilde{z}_k\|^2_{\mathscr{Z},M_k} = \|x_{k-1}-x_k\|^2_{\mathscr{X},R_k} + \|B(y_{k-1}-y_k)\|^2_{\Gamma,H_k} + \|y_{k-1}-y_k\|^2_{\mathscr{Y},S_k} + \|\gamma_{k-1}-\tilde{\gamma}_k\|^2_{\Gamma,\theta^{-1}H_k^{-1}}. \tag{63}$$

Using the same reasoning and Lemma 4.2(a), we also find

$$\|z_k-\tilde{z}_k\|^2_{\mathscr{Z},M_k} = \|\gamma_k-\tilde{\gamma}_k\|^2_{\Gamma,\theta^{-1}H_k^{-1}} = \left\|\frac{1-\theta}{\theta}(\gamma_k-\gamma_{k-1})+H_kB(y_k-y_{k-1})\right\|^2_{\Gamma,\theta^{-1}H_k^{-1}}. \tag{64}$$

Hence, from Lemma 4.2(a) and some algebraic manipulations, we obtain

$$\begin{aligned}\sigma_\theta\|\gamma_{k-1}-\tilde{\gamma}_k\|^2_{\Gamma,\theta^{-1}H_k^{-1}} - \|\gamma_k-\tilde{\gamma}_k\|^2_{\Gamma,\theta^{-1}H_k^{-1}} &= \sigma_\theta\left\|\frac{1}{\theta}(\gamma_k-\gamma_{k-1})+H_kB(y_k-y_{k-1})\right\|^2_{\Gamma,\theta^{-1}H_k^{-1}} \\ &\quad - \left\|\frac{1-\theta}{\theta}(\gamma_k-\gamma_{k-1})+H_kB(y_k-y_{k-1})\right\|^2_{\Gamma,\theta^{-1}H_k^{-1}} \\ &= \frac{\sigma_\theta-(1-\theta)^2}{\theta^2}\|\gamma_k-\gamma_{k-1}\|^2_{\Gamma,\theta^{-1}H_k^{-1}} + \frac{\sigma_\theta-1}{\theta}\|B(y_k-y_{k-1})\|^2_{\Gamma,H_k} \\ &\quad + \frac{2(\sigma_\theta+\theta-1)}{\theta^2}\langle\gamma_k-\gamma_{k-1},B(y_k-y_{k-1})\rangle_\Gamma,\end{aligned}$$

which in turn, combined with (63) and (64), yields

$$\begin{aligned}\sigma_\theta\|z_{k-1}-\tilde{z}_k\|^2_{\mathscr{Z},M_k} - \|z_k-\tilde{z}_k\|^2_{\mathscr{Z},M_k} &= \sigma_\theta\|x_k-x_{k-1}\|^2_{\mathscr{X},R_k} + \sigma_\theta\|y_k-y_{k-1}\|^2_{\mathscr{Y},S_k} \\ &\quad + \frac{\sigma_\theta-(1-\theta)^2}{\theta^2}\|\gamma_k-\gamma_{k-1}\|^2_{\Gamma,\theta^{-1}H_k^{-1}} + \frac{\sigma_\theta(\theta+1)-1}{\theta}\|B(y_k-y_{k-1})\|^2_{\Gamma,H_k} \\ &\quad + \frac{2(\sigma_\theta+\theta-1)}{\theta^2}\langle\gamma_k-\gamma_{k-1},B(y_k-y_{k-1})\rangle_\Gamma,\end{aligned} \tag{65}$$

We will now consider two cases: $k=1$ and $k>1$. In the first case, it follows from (65) with $k=1$, Lemma 4.2(b), the first inequality in (59) with $\sigma=\sigma_\theta$, and the definitions of η_0 and η_1, that

$$\begin{aligned}\sigma_\theta\|z_0-\tilde{z}_1\|^2_{\mathscr{Z},M_1} - \|z_1-\tilde{z}_1\|^2_{\mathscr{Z},M_1} + \eta_0 - \eta_1 &\geq \left[\sigma_\theta - \frac{\sqrt{2}(\sigma_\theta+\theta-1)}{\theta} + \frac{\sigma_\theta+\theta-1}{\theta^{3/2}}\right]\|y_1-y_0\|^2_{\mathscr{Y},S_1}, \\ &\geq \left[\sigma_\theta + \frac{(\sigma_\theta+\theta-1)\left(2-3\sqrt{2}\right)}{3\theta}\right]\|y_1-y_0\|^2_{\mathscr{Y},S_1},\end{aligned}$$

where the last inequality is due to $\sqrt{\theta} \leq 3/2$. Hence, since $(2-3\sqrt{2})/3 \geq (2\sqrt{2}-4)/\sqrt{2}$, inequality (62) for $k=1$ now follows from the second inequality in (59) with $\sigma = \sigma_\theta$. On the other hand, assuming $k>1$, from inequality (65), Lemma 4.2(c) with $t=\sqrt{2}$, the first inequality in (59) with $\sigma = \sigma_\theta$, and the definition of $\{\eta_k\}$ in (61), we have

$$\begin{aligned}
\sigma_\theta \|z_{k-1} - \tilde{z}_k\|^2_{\mathscr{Z},M_k} - \|z_k - \tilde{z}_k\|^2_{\mathscr{Z},M_k} + \eta_{k-1} - \eta_k \geq \frac{\sigma_\theta(\theta+1)-1}{\theta}\|B(y_k - y_{k-1})\|^2_{\Gamma,H_k} \\
+ \frac{\sigma_\theta - (1-\theta)^2}{\theta^2}\|\gamma_{k-1} - \gamma_{k-2}\|^2_{\Gamma,\theta^{-1}H^{-1}_{k-1}} + \frac{2(\sigma_\theta+\theta-1)(1-\theta)}{\theta^2}\langle \gamma_{k-1} - \gamma_{k-2}, B(y_k - y_{k-1})\rangle_\Gamma \\
+ \left[\frac{(\sigma_\theta+\theta-1)\left(2\sqrt{2}-4+1-c_{k-1}\right)}{\sqrt{2}\theta} + \sigma_\theta\right]\|y_k - y_{k-1}\|^2_{\mathscr{Y},S_k}.
\end{aligned}$$

Since $c_{k-1} \leq 1$ (see condition **C2**), we obtain from (59) with $\sigma = \sigma_\theta$ that the term inside bracket is non-negative. Hence, inequality (62) for $k>1$ now follows from the first statement of Lemma 4.3. The last statement of the proposition follows from (62) and variable metric HPE framework's definition. □

We are now ready to prove Theorems 4.1 and 4.2.

Proof of Theorem 4.1: Due to the definitions of $\{(r_{i,x}, r_{i,y}, r_{i,\gamma})\}$ and $\{M_k\}$ in (34) and (49), respectively, it follows from Proposition 4.2 and Theorem 3.1 that, for every $k \geq 1$, (34) holds and there exists $i \leq k$ such that

$$\sqrt{\|r_{i,x}\|^2_{\mathscr{X}} + \|r_{i,y}\|^2_{\mathscr{Y}} + \|r_{i,\gamma}\|^2_{\Gamma}} \leq \frac{d_0}{\sqrt{k}}\sqrt{\frac{[2(1+\sigma_\theta)C_P(1+\tau_\theta)+2(1-\sigma_\theta)\tau_\theta]C_P\|M_0\|}{(1-\sigma_\theta)}}.$$

Therefore, the inequality in (35) now follows from the definitions of m_0 and M_0 in (31) and (49), respectively, and properties of norms. □

Proof of Theorem 4.2: First, using the definitions of m_0 and M_0 in (31) and (49), respectively, we have $\|M_0\| \leq m_0$. Hence, combining Proposition 4.2 and Theorem 3.2, and taking into account that $r^a_k = (r^a_{k,x}, r^a_{k,y}, r^a_{k,\gamma})$, we conclude that, for every $k \geq 1$,

$$\sqrt{\|r^a_{k,x}\|^2_{\mathscr{X}} + \|r^a_{k,y}\|^2_{\mathscr{Y}} + \|r^a_{k,\gamma}\|^2_{\Gamma}} \leq \frac{\sqrt{(1+\tau_\theta)m_0}\mathscr{E}\, d_0}{k}, \tag{66}$$

$$\varepsilon_k^a = \frac{1}{k}\left(\sum_{i=1}^{k}\langle r_{i,x}, x_i - x_k^a\rangle_{\mathscr{X}} + \sum_{i=1}^{k}\langle r_{i,y}, y_i - y_k^a\rangle_{\mathscr{Y}} + \sum_{i=1}^{k}\langle r_{i,\gamma}, \tilde{\gamma}_i - \tilde{\gamma}_k^a\rangle_{\Gamma}\right) \leq \frac{(1+\tau_\theta)\widehat{\mathscr{E}} d_0^2}{k}. \tag{67}$$

On the other hand, (34), (38) and (39) yield

$$Ax_k + By_k = r_{k,\gamma} + b, \quad Ax_k^a + By_k^a = r_{k,\gamma}^a + b.$$

Additionally, (38), (39) and some algebraic manipulations give

$$\sum_{i=1}^{k}\langle \tilde{\gamma}_i, r_{i,\gamma} - r_{k,\gamma}^a\rangle_{\Gamma} = \sum_{i=1}^{k}\langle \tilde{\gamma}_i - \tilde{\gamma}_k^a, r_{i,\gamma} - r_{k,\gamma}^a\rangle_{\Gamma} = \sum_{i=1}^{k}\langle \tilde{\gamma}_i - \tilde{\gamma}_k^a, r_{i,\gamma}\rangle_{\Gamma}.$$

Hence, combining the identity in (67) with the last two displayed equations, we also obtain

$$\begin{aligned}\varepsilon_k^a &= \frac{1}{k}\sum_{i=1}^{k}\Big(\langle r_{i,x}, x_i - x_k^a\rangle_{\mathscr{X}} + \langle r_{i,y}, y_i - y_k^a\rangle_{\mathscr{Y}}\Big) + \frac{1}{k}\sum_{i=1}^{k}\langle \tilde{\gamma}_i, r_{i,\gamma} - r_{k,\gamma}^a\rangle_{\Gamma} \\ &= \frac{1}{k}\sum_{i=1}^{k}\Big(\langle r_{i,x}, x_i - x_k^a\rangle_{\mathscr{X}} + \langle r_{i,y}, y_i - y_k^a\rangle_{\mathscr{Y}} + \langle \tilde{\gamma}_i, Ax_i - Ax_k^a + By_i - By_k^a\rangle_{\Gamma}\Big) \\ &= \frac{1}{k}\sum_{i=1}^{k}\langle r_{i,x} + A^*\tilde{\gamma}_i, x_i - x_k^a\rangle_{\mathscr{X}} + \frac{1}{k}\sum_{i=1}^{k}\langle r_{i,y} + B^*\tilde{\gamma}_i, y_i - y_k^a\rangle_{\mathscr{Y}} = \varepsilon_{k,x}^a + \varepsilon_{k,y}^a,\end{aligned}$$

where the last equality is due to the definitions of $\varepsilon_{k,x}^a$ and $\varepsilon_{k,y}^a$ in (40). Therefore, the inequalities in (42) and (43) now follows from (66) and (67), respectively.

To finish the proof of the theorem, note that direct use of Theorem 2.1(b) (for f and g), (34) and (38)–(40) give $\varepsilon_{k,x}^a$, $\varepsilon_{k,y}^a \geq 0$ and (41). □

5 Conclusions

We considered the linearly constrained convex optimization problem and studied a variable metric proximal alternating direction method of multipliers for solving it. We proved that this ADMM variant, which allows the use of degenerate metrics (defined by noninvertible linear operators), has $\mathscr{O}(1/\sqrt{k})$ pointwise and $\mathscr{O}(1/k)$ ergodic convergence rates. These convergence rates were obtained essentially by showing that this ADMM variant can be seen as a special case of a variable metric hybrid proximal extragradient framework for solving monotone inclusions. Convergence rates for the latter framework were also provided in this work.

Appendix A Proofs of Theorems 3.1 and 3.2

We start by presenting the following two Lemmas.

Lemma A.1 *For any $z^*, z, z_+, \tilde{z} \in \mathscr{Z}$ and $M \in \mathscr{M}_+^{\mathscr{Z}}$, we have*

$$\|z^* - z\|^2_{\mathscr{Z},M} - \|z^* - z_+\|^2_{\mathscr{Z},M} = \|z - \tilde{z}\|^2_{\mathscr{Z},M} - \|z_+ - \tilde{z}\|^2_{\mathscr{Z},M} + 2\langle \tilde{z} - z^*, M(z - z_+)\rangle_{\mathscr{Z}}.$$

Proof Direct calculations yield

$$\begin{aligned}\|z^* - z\|^2_{\mathscr{Z},M} - \|z^* - z_+\|^2_{\mathscr{Z},M} &= 2\langle z_+ - z^*, M(z - z_+)\rangle_{\mathscr{Z}} + \|z_+ - z\|^2_{\mathscr{Z},M} \\ &= 2\langle z_+ - \tilde{z}, M(z - z_+)\rangle_{\mathscr{Z}} + 2\langle \tilde{z} - z^*, M(z - z_+)\rangle_{\mathscr{Z}} + \|z_+ - z\|^2_{\mathscr{Z},M} \\ &= 2\langle \tilde{z} - z^*, M(z - z_+)\rangle_{\mathscr{Z}} + \|\tilde{z} - z\|^2_{\mathscr{Z},M} - \|\tilde{z} - z_+\|^2_{\mathscr{Z},M}.\end{aligned}$$

□

Lemma A.2 *Let $\{z_k\}$, $\{M_k\}$, $\{\tilde{z}_k\}$ and $\{\eta_k\}$ be generated by the* variable metric HPE framework. *For every $k \geq 1$ and $z^* \in T^{-1}(0)$:*

(a) *we have*

$$\|z^* - z_k\|^2_{\mathscr{Z},M_k} \leq \|z^* - z_{k-1}\|^2_{\mathscr{Z},M_k} + \eta_{k-1} - \eta_k - (1-\sigma)\|z_{k-1} - \tilde{z}_k\|^2_{\mathscr{Z},M_k};$$

(b) *we have*

$$\|z^* - z_k\|^2_{\mathscr{Z},M_k} + \eta_k + (1-\sigma)\sum_{i=1}^{k}\|z_{i-1} - \tilde{z}_i\|^2_{\mathscr{Z},M_i} \leq C_P(\|z^* - z_0\|^2_{\mathscr{Z},M_0} + \eta_0),$$

where C_P and M_0 are as in (11) *and condition* **C1***, respectively.*

Proof (a) From Lemma A.1 with $(z, z_+, \tilde{z}) = (z_{k-1}, z_k, \tilde{z}_k)$ and $M = M_k$, (12) and (13), we obtain

$$\|z^* - z_{k-1}\|^2_{\mathscr{Z},M_k} - \|z^* - z_k\|^2_{\mathscr{Z},M_k} + \eta_{k-1} \geq (1-\sigma)\|z_{k-1} - \tilde{z}_k\|^2_{\mathscr{Z},M_k} + \eta_k + 2\langle \tilde{z}_k - z^*, r_k\rangle.$$

Hence, (a) follows from the above inequality, the fact that $0 \in T(z^*)$ and $r_k \in T(\tilde{z}_k)$ (see (12)), and the monotonicity of T.

(b) Using (a), (3) and condition **C1**, we find

$$\|z^* - z_k\|^2_{\mathscr{Z},M_k} \leq (1 + c_{k-1})\|z^* - z_{k-1}\|^2_{\mathscr{Z},M_{k-1}} + \eta_{k-1} - \eta_k - (1-\sigma)\|z_{k-1} - \tilde{z}_k\|^2_{\mathscr{Z},M_k}.$$

Thus, the result follows by applying the above inequality recursively and by using (11). □

We are now ready to prove Theorem 3.1.

Proof of Theorem 3.1: First, note that the desired inclusion holds due to (12). Now, using (2) and (13), we obtain, respectively,

$$\|z_{k-1} - z_k\|^2_{\mathscr{Z},M_k} \leq 2\left(\|z_{k-1} - \tilde{z}_k\|^2_{\mathscr{Z},M_k} + \|\tilde{z}_k - z_k\|^2_{\mathscr{Z},M_k}\right),$$

$$\|\tilde{z}_k - z_k\|^2_{\mathscr{Z},M_k} \leq \sigma\|z_{k-1} - \tilde{z}_k\|^2_{\mathscr{Z},M_k} + \eta_{k-1} - \eta_k.$$

Combining the above inequalities, we find

$$\|z_{k-1} - z_k\|^2_{\mathscr{Z},M_k} \leq 2\left[(1+\sigma)\|z_{k-1} - \tilde{z}_k\|^2_{\mathscr{Z},M_k} + \eta_{k-1} - \eta_k\right],$$

which in turn, combined with Lemma A.2(b), yields

$$\sum_{i=1}^{k} \|z_{i-1} - z_i\|^2_{\mathscr{Z},M_i} \leq \frac{2(1+\sigma)C_P(\|z^* - z_0\|^2_{\mathscr{Z},M_0} + \eta_0) + 2(1-\sigma)\eta_0}{(1-\sigma)}, \tag{68}$$

for all $z^* \in T^{-1}(0)$. Now, from (11), we obtain $M_i \preceq C_P M_0$ for every $i \geq 1$. Thus, it follows from (12) and Proposition 2.1 that

$$\sum_{i=1}^{k} \|r_i\|^2_{\mathscr{Z}} = \sum_{i=1}^{k} \|M_i(z_{i-1} - z_i)\|^2_{\mathscr{Z}} \leq C_P\|M_0\| \sum_{i=1}^{k} \|z_{i-1} - z_i\|^2_{\mathscr{Z},M_i},$$

which, combined with the fact that $\sum_{i=1}^{k} t_i \geq k \min_{i=1,\ldots,k}\{t_i\}$ and the definition in (14), proves (15). □

Before proceeding to the proof of the ergodic convergence of the variable metric HPE framework, let us first present an auxiliary result.

Proposition A.1 *Let* $\{z_k\}$, $\{M_k\}$ *and* $\{\eta_k\}$ *be generated by the* variable metric HPE framework *and consider* $\{\tilde{z}_k^a\}$ *and* $\{\varepsilon_k^a\}$ *as in* (18). *Then, for every* $k \geq 1$,

$$\varepsilon_k^a \leq \frac{1}{2k}\left(\eta_0 + \|\tilde{z}_k^a - z_0\|^2_{\mathscr{Z},M_0} + \sum_{i=1}^{k} c_{i-1}\|\tilde{z}_k^a - z_{i-1}\|^2_{\mathscr{Z},M_{i-1}}\right), \tag{69}$$

where $\{c_k\}$ *is given in condition* **C1**.

Proof Using Lemma A.1 with $(z^*, z, z_+, \tilde{z}) = (\tilde{z}_k^a, z_{i-1}, z_i, \tilde{z}_i)$ and $M = M_i$, (12) and (13), we find, for every $i = 1, \ldots, k$,

$$\begin{aligned}\|\tilde{z}_k^a - z_{i-1}\|^2_{\mathscr{Z},M_i} - \|\tilde{z}_k^a - z_i\|^2_{\mathscr{Z},M_i} + \eta_{i-1} &\geq (1-\sigma)\|\tilde{z}_i - z_{i-1}\|^2_{\mathscr{Z},M_i} + \eta_i + 2\langle r_i, \tilde{z}_i - \tilde{z}_k^a\rangle \\ &\geq \eta_i + 2\langle r_i, \tilde{z}_i - \tilde{z}_k^a\rangle,\end{aligned}$$

where the second inequality is due to the fact that $1-\sigma \geq 0$. Hence, using condition **C1** and simple calculations, we obtain

$$\|\tilde{z}_k^a - z_i\|^2_{\mathscr{Z},M_i} \leq (1+c_{i-1})\|\tilde{z}_k^a - z_{i-1}\|^2_{\mathscr{Z},M_{i-1}} + \eta_{i-1} - \eta_i - 2\langle r_i, \tilde{z}_i - \tilde{z}_k^a\rangle \quad \forall i = 1, \ldots, k.$$

Summing up the last inequality from $i = 1$ to $i = k$ and using the definition of ε_k^a in (18), we have

$$0 \leq \|\tilde{z}_k^a - z_k\|^2_{\mathscr{Z},M_k} \leq \sum_{i=1}^k c_{i-1}\|\tilde{z}_k^a - z_{i-1}\|^2_{\mathscr{Z},M_{i-1}} + \|\tilde{z}_k^a - z_0\|^2_{\mathscr{Z},M_0} + \eta_0 - 2k\varepsilon_k^a,$$

which clearly gives (69). □

Proof of Theorem 3.2: Note first that the desired inclusion and the first inequality in (20) follow from (12), (18) and Theorem 2.1(a). Take $z^* \in T^{-1}(0)$. Now, let us prove the second inequality in (20), which will follow by bounding the term in the right-hand side of (69). Note that, using the convexity of $\|\cdot\|^2_{M_{i-1}}$, inequality (2) and (18), we find

$$\|\tilde{z}_k^a - z_{i-1}\|^2_{\mathscr{Z},M_{i-1}} \leq \frac{1}{k}\sum_{j=1}^k \|\tilde{z}_j - z_{i-1}\|^2_{\mathscr{Z},M_{i-1}} \leq \frac{2}{k}\sum_{j=1}^k \left(\|\tilde{z}_j - z_j\|^2_{\mathscr{Z},M_{i-1}} + \|z_j - z_{i-1}\|^2_{\mathscr{Z},M_{i-1}}\right). \tag{70}$$

From (11), we have $M_{i-1} \preceq C_P M_j$ for all $j = 1, \ldots, k$. Hence, using Proposition 2.1, inequality (13), Lemma A.2(b) and (14), we find

$$\begin{aligned}\sum_{j=1}^k \|\tilde{z}_j - z_j\|^2_{\mathscr{Z},M_{i-1}} &\leq C_P \sum_{j=1}^k \|\tilde{z}_j - z_j\|^2_{\mathscr{Z},M_j} \\ &\leq C_P \sum_{j=1}^k \left(\sigma\|\tilde{z}_j - z_{j-1}\|^2_{\mathscr{Z},M_j} + \eta_{j-1} - \eta_j\right) \\ &\leq \frac{\sigma}{1-\sigma} C_P^2 (d_0^2 + \eta_0) + C_P \eta_0.\end{aligned} \tag{71}$$

On the other hand, using (2), $M_{i-1} \preceq C_P M_j$ for all $j=1,\ldots,k$, Proposition 2.1, Lemma A.2(b) and (14), we obtain

$$\begin{aligned}\sum_{j=1}^{k} \|z_j - z_{i-1}\|^2_{\mathscr{Z},M_{i-1}} &\leq 2\sum_{j=1}^{k}\left(\|z_j - z^*\|^2_{\mathscr{Z},M_{i-1}} + \|z^* - z_{i-1}\|^2_{\mathscr{Z},M_{i-1}}\right)\\ &\leq 2\sum_{j=1}^{k}\left(C_P\|z_j - z^*\|^2_{\mathscr{Z},M_j} + \|z^* - z_{i-1}\|^2_{\mathscr{Z},M_{i-1}}\right)\\ &\leq 2(1+C_P)C_P(d_0^2+\eta_0)k. \qquad (72)\end{aligned}$$

It follows from inequalities (70)–(72) and the fact that $k \geq 1$ that

$$\|\tilde{z}^a_k - z_{i-1}\|^2_{\mathscr{Z},M_{i-1}} \leq \left(\frac{\sigma C_P}{1-\sigma} + 2(1+C_P)\right)2C_P(d_0^2+\eta_0) + 2C_P\eta_0,$$

which, combined with Proposition A.1 and the first condition in (10), yields

$$\varepsilon^a_k \leq \frac{1}{2k}\left[2C_P(1+C_S)\left(\frac{\sigma C_P}{1-\sigma} + 2(1+C_P)\right)(d_0^2+\eta_0) + (1+2(1+C_S)C_P)\,\eta_0\right].$$

Therefore, the second inequality in (20) now follows from definition of $\widehat{\mathscr{E}}$ and simple calculations.

To finish the proof of the theorem, it remains to prove (19). Assume first that $k \geq 2$. Using (18) and simple calculations, we have

$$k\, r^a_k = \sum_{i=1}^{k} r_i = M_1(z_0 - z^*) - M_k(z_k - z^*) + \sum_{i=1}^{k-1}(M_{i+1} - M_i)(z_i - z^*). \qquad (73)$$

Since $M_k \preceq C_P M_0$ and $M_1 \preceq C_P M_0$ (see (11)), we obtain from Proposition 2.1 that

$$\|M_k(z_k - z^*)\|_{\mathscr{Z}} \leq \sqrt{C_P\|M_0\|}\,\|z_k - z^*\|_{\mathscr{Z},M_k}, \qquad (74)$$

$$\|M_1(z_0 - z^*)\|_{\mathscr{Z}} \leq \sqrt{C_P\|M_0\|}\,\|z_0 - z^*\|_{\mathscr{Z},M_1} \leq C_p\sqrt{\|M_0\|}\,\|z_0 - z^*\|_{\mathscr{Z},M_0}. \qquad (75)$$

Next step is to estimate the general term in the summation in (73). To do this, first note that using condition **C1**, we find

$$0 \preceq L_i := M_{i+1} - M_i + c_i M_{i+1} \preceq c_i(2+c_i)M_i, \qquad \forall\, i = 1,\ldots,k-1, \qquad (76)$$

and so

$$\|(M_{i+1}-M_i)(z_i-z^*)\|_{\mathscr{Z}} = \|(L_i-c_iM_{i+1})(z_i-z^*)\|_{\mathscr{Z}} \le \|L_i(z_i-z^*)\|_{\mathscr{Z}} + c_i\|M_{i+1}(z_i-z^*)\|_{\mathscr{Z}}. \quad (77)$$

It follows from the last inequality in (76) and (11) that $L_i \preceq c_i(2+c_i)M_i$ and $M_i \preceq C_P M_0$. Hence, we have

$$\begin{aligned}
\|L_i(z_i-z^*)\|^2_{\mathscr{Z}} &= \langle L_i(L_i^{1/2}(z_i-z^*)), L_i^{1/2}(z_i-z^*)\rangle \le c_i(2+c_i)\langle M_i(L_i^{1/2}(z_i-z^*)), L_i^{1/2}(z_i-z^*)\rangle \\
&\le c_i(2+c_i)C_P\langle M_0(L_i^{1/2}(z_i-z^*)), L_i^{1/2}(z_i-z^*)\rangle \\
&\le c_i(2+c_i)C_P\|M_0\|\|z_i-z^*\|^2_{\mathscr{Z},L_i} \\
&\le c_i^2(2+c_i)^2C_P\|M_0\|\|z_i-z^*\|^2_{\mathscr{Z},M_i}. \quad (78)
\end{aligned}$$

Again, using the facts that $M_{i+1} \preceq C_P M_0$ and $M_{i+1} \preceq (1+c_i)M_i$ (see (11)), and Proposition 2.1, we obtain

$$\|M_{i+1}(z_i-z^*)\|_{\mathscr{Z}} \le \sqrt{C_p\|M_0\|}\|z_i-z^*\|_{\mathscr{Z},M_{i+1}} \le \sqrt{C_P\|M_0\|(1+c_i)}\|z_i-z^*\|_{\mathscr{Z},M_i}. \quad (79)$$

Hence, using (11) and (77)–(79), we find

$$\begin{aligned}
\|(M_{i+1}-M_i)(z_i-z^*)\|_{\mathscr{Z},M_k} &\le c_i\sqrt{C_p\|M_0\|}\left(1+(1+c_i)+\sqrt{1+c_i}\right)\|z_i-z^*\|_{\mathscr{Z},M_i} \\
&\le c_i\sqrt{C_P\|M_0\|}\left(1+C_P+\sqrt{C_P}\right)\|z_i-z^*\|_{\mathscr{Z},M_i}. \quad (80)
\end{aligned}$$

Finally, using the definition of d_0 in (14), (73)–(75), (80) and Lemma A.2(b), we conclude that

$$\begin{aligned}
k\|r_k^a\|_{\mathscr{Z}} &\le \|M_1(z_0-z^*)\|_{\mathscr{Z}} + \|M_k(z_k-z^*)\|_{\mathscr{Z}} + \sum_{i=1}^{k-1}\|(M_{i+1}-M_i)(z_i-z^*)\|_{\mathscr{Z}} \\
&\le \left(C_P+\sqrt{C_P}+C_S\sqrt{C_P}\left(1+C_P+\sqrt{C_P}\right)\right)\sqrt{\|M_0\|}\max_{i=0,\dots,k}\|z_i-z^*\|_{\mathscr{Z},M_i} \\
&\le \sqrt{C_P\|M_0\|}\left(C_P+\sqrt{C_P}+C_S\sqrt{C_P}\left(1+C_P+\sqrt{C_P}\right)\right)\sqrt{d_0^2+\eta_0} \\
&\le \left((1+C_S)(1+\sqrt{C_P})C_P+C_SC_P^2\right)\sqrt{\|M_0\|}\sqrt{d_0^2+\eta_0}
\end{aligned}$$

which gives (19) for the case $k \geq 2$. Note now that by (11), we have $M_1 \preceq C_P M_0$ and so, using the second identity in (18) with $k = 1$, Proposition 2.1, Lemma A.2(b) and (14), we find

$$\begin{aligned}
\|r_1^a\|_{\mathscr{X}} = \|M_1(z_0 - z_1)\|_{\mathscr{X}} &\leq \sqrt{C_P\|M_0\|}\|z_0 - z_1\|_{\mathscr{X},M_1} \\
&\leq \sqrt{C_P\|M_0\|}(\|z_0 - z^*\|_{\mathscr{X},M_1} + \|z_1 - z^*\|_{\mathscr{X},M_1}) \\
&\leq \sqrt{C_P\|M_0\|}(\sqrt{C_P}\|z_0 - z^*\|_{\mathscr{X},M_0} + \|z_1 - z^*\|_{\mathscr{X},M_1}) \\
&\leq (C_P + \sqrt{C_P})\sqrt{\|M_0\|}\max_{i=0,1}\|z_i - z^*\|_{\mathscr{X},M_i} \\
&\leq (C_P + \sqrt{C_P})\sqrt{C_P\|M_0\|}\sqrt{d_0^2 + \eta_0},
\end{aligned}$$

which, in turn, gives (19) for $k = 1$. □

Acknowledgements The work of these authors was supported in part by CNPq Grants 406250/2013-8, 444134/2014-0, 309370/2014-0 and 406975/2016-7. We thank the reviewers for their careful reading and comments.

References

1. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**(1), 1–122 (2011)
2. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. **2**, 17–40 (1976).
3. Glowinski, R., Marroco, A.: Sur l'approximation, par éléments finis d'ordre un, et la résolution, par penalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. (1975)
4. Attouch, H., Soueycatt, M.: Augmented Lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces. Applications to games, PDE's and control. Pac. J. Optim. **5**(1), 17–37 (2008)
5. He, B., Liu, H., Wang, Z., Yuan, X.: A strictly contractive peaceman–rachford splitting method for convex programming. SIAM J. Optim. **24**(3), 1011–1040 (2014)
6. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vision **40**(1), 120–145 (2011)
7. Davis, D., Yin, W.: Convergence rate analysis of several splitting schemes. Avaliable on http://www.arxiv.org

8. Deng, W., Yin, W.: On the global and linear convergence of the generalized alternating direction method of multipliers. J. Sci. Comput. pp. 1–28 (2015).
9. Eckstein, J.: Some saddle-function splitting methods for convex programming. Optim. Method Softw. **4**(1), 75–83 (1994).
10. Fang, E.X., Bingsheng, H., Liu, H., Xiaoming, Y.: Generalized alternating direction method of multipliers: new theoretical insights and applications. Math. Prog. Comp. **7**(2), 149–187 (2015).
11. Fazel, M., Pong, T.K., Sun, D., Tseng, P.: Hankel matrix rank minimization with applications to system identification and realization. SIAM J. Matrix Anal. Appl. **34**(3), 946–977 (2013)
12. Gonçalves, M.L.N., Melo, J.G., Monteiro, R.D.C.: Improved pointwise iteration-complexity of a regularized ADMM and of a regularized non-euclidean HPE framework. SIAM J. Optim. **27**(1), 379–407 (2017).
13. Hager, W.W., Yashtini, M., Zhang, H.: An $O(1/k)$ convergence rate for the variable stepsize Bregman operator splitting algorithm. SIAM J. Numer. Anal. **54**(3), 1535–1556 (2016)
14. He, B., Liao, L.Z., Han, D., Yang, H.: A new inexact alternating directions method for monotone variational inequalities. Math. Program. **92**(1, Ser. A), 103–118 (2002)
15. He, B., Yuan, X.: On the $\mathscr{O}(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. SIAM Journal on Numer. Anal. **50**(2), 700–709 (2012).
16. Lin, T., Ma, S., Zhang, S.: An extragradient-based alternating direction method for convex minimization. Found. Comput. Math. pp. 1–25 (2015)
17. Ouyang, Y., Chen, Y., Lan, G., Pasiliao Jr., E.: An accelerated linearized alternating direction method of multipliers. SIAM J. Imaging Sci. **8**(1), 644–681 (2015)
18. Shefi, R., Teboulle, M.: Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization. SIAM J. Optim. **24**(1), 269–297 (2014)
19. BoŢ, R.I., Csetnek, E.R.: ADMM for monotone operators: convergence analysis and rates. Avaliable on http://www.arxiv.org
20. Monteiro, R.D.C., Svaiter, B.F.: Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. SIAM J. Optim. **23**(1), 475–507 (2013)
21. Solodov, M.V., Svaiter, B.F.: A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. Set-Valued Anal. **7**(4), 323–345 (1999)
22. Zhang, X., Burger, M., Bresson, X., Osher, S.: Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. SIAM J. Imaging Sci. **3**(3), 253–276 (2010)
23. He, B., Yuan, X.: On non-ergodic convergence rate of Douglas-Rachford alternating direction method of multipliers. Numer. Math. **130**(3), 567–577 (2015).
24. Cui, Y., Li, X., Sun, D., Toh, K.C.: On the convergence properties of a majorized ADMM for linearly constrained convex optimization problems with coupled objective functions. J. Optim. Theory Appl. **169**(3), 1013–1041 (2016)

25. Gonçalves, M.L.N., Melo, J.G., Monteiro, R.D.C.: Extending the ergodic convergence rate of the proximal ADMM. Avaliable on http://www.arxiv.org
26. Shen, L., Pan, S.: Weighted iteration complexity of the sPADMM on the KKT residuals for convex composite optimization. Avaliable on http://www.arxiv.org
27. He, B., Yang, H.: Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities. Oper. Res. Lett. **23**(35), 151 – 161 (1998)
28. He, B.S., Yang, H., Wang, L.: Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities. J. Optim. Theory. Appl. **106**(2), 337–356 (2000)
29. Solodov, M.V.: A class of decomposition methods for convex optimization and monotone variational inclusions via the hybrid inexact proximal point framework. Optim. Method Softw. **19**(5), 557–575 (2004).
30. Lotito, P.A., Parente, L.A., Solodov, M.V.: A class of variable metric decomposition methods for monotone variational inclusions. J. Convex Anal. **16**(3&4), 857–880 (2009)
31. Banert, S., BoŢ, R.I., Csetnek, E.R.: Fixing and extending some recent results on the ADMM algorithm. Avaliable on http://www.arxiv.org
32. Monteiro, R.D.C., Svaiter, B.F.: On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. SIAM J. Optim. **20**(6), 2755–2787 (2010)
33. He, Y., Monteiro, R.D.C.: An accelerated HPE-type algorithm for a class of composite convex-concave saddle-point problems. SIAM J. Optim. **26**(1), 29–56 (2016)
34. Marques Alves, M., Monteiro, R.D.C., Svaiter, B.F.: Regularized HPE-type methods for solving monotone inclusions with improved pointwise iteration-complexity bounds. SIAM J. Optim. **26**(4), 2730–2743 (2016).
35. Monteiro, R.D.C., Svaiter, B.F.: Complexity of variants of Tseng's modified F-B splitting and Korpelevich's methods for hemivariational inequalities with applications to saddle-point and convex optimization problems. SIAM J. Optim. **21**(4), 1688–1720 (2011)
36. Parente, L.A., Lotito, P.A., Solodov, M.V.: A class of inexact variable metric proximal point algorithms. SIAM J. Optim. **19**(1), 240–260 (2008).
37. Rockafellar, R.T.: On the maximal monotonicity of subdifferential mappings. Pacific J. Math. **33**, 209–216 (1970)
38. Burachik, R.S., Sagastizábal, C.A., Svaiter, B.F.: ε-enlargements of maximal monotone operators: theory and applications. In: Reformulation: nonsmooth, piecewise smooth, semismooth and smoothing methods (Lausanne, 1997), *Appl. Optim.*, vol. 22, pp. 25–43. Kluwer Acad. Publ., Dordrecht (1999)
39. Burachik, R.S., Iusem, A.N., Svaiter, B.F.: Enlargement of monotone operators with applications to variational inequalities. Set-Valued Anal. **5**(2), 159–180 (1997)
40. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)