Pointwise and ergodic convergence rates of a variable metric proximal ADMM

Max L.N. Gonçalves [*] M. Marques Alves [†] Jefferson G. Melo [*]

January 25, 2017

Abstract

In this paper, we obtain global $\mathcal{O}(1/\sqrt{k})$ pointwise and $\mathcal{O}(1/k)$ ergodic convergence rates for a variable metric proximal alternating direction method of multipliers (VM-PADMM) for solving linearly constrained convex optimization problems. The VM-PADMM can be seen as a class of ADMM variants, allowing the use of degenerate metrics (defined by noninvertible linear operators). We first propose and study nonasymptotic convergence rates of a variable metric hybrid proximal extragradient (VM-HPE) framework for solving monotone inclusions. Then, the above-mentioned convergence rates for the VM-PADMM 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 VM-PADMM (resp. VM-HPE framework).

2000 Mathematics Subject Classification: 90C25, 90C60, 49M27, 47H05, 47J22, 65K10.

Key words: alternating direction method of multipliers, variable metric, pointwise and ergodic convergence rates, hybrid proximal extragradient method, convex program.

1 Introduction

We consider the linearly constrained convex optimization problem

$$\begin{aligned} \text{minimize} \quad & f(x) + g(y) \\ \text{subject to} \quad & Ax + By = b, \end{aligned} \tag{1}$$

where $f : \mathcal{X} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ and $g : \mathcal{Y} \to \overline{\mathbb{R}}$ are extended-real-valued proper closed and convex functions, $\mathcal{X}, \mathcal{Y}$ and Γ are finite-dimensional real vector spaces, and $A : \mathcal{X} \to \Gamma$ and $B : \mathcal{Y} \to \Gamma$ are linear operators. One of the most popular methods for solving (1) is the alternating direction method of multipliers (ADMM) [4, 14, 15], for which many variants have been proposed and studied in the literature; see, e.g., [1, 3, 7, 9, 10, 11, 12, 13, 17, 18, 19, 21, 25, 31].

[*]Instituto de Matemática e Estatística, Universidade Federal de Goiás, Campus II- Caixa Postal 131, CEP 74001-970, Goiânia-GO, Brazil. (E-mails: maxlng@ufg.br and jefferson@ufg.br). The work of these authors was supported in part by CNPq Grants 406250/2013-8, 444134/2014-0 and 309370/2014-0.

[†]Departamento de Matemática, Universidade Federal de Santa Catarina, Florianópolis, Brazil, 88040-900 (maicon.alves@ufsc.br). The work of this author was partially supported by CNPq grants no. 406250/2013-8, 306317/2014-1 and 405214/2016-2.

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

$$x_k \in \operatorname{argmin}_{x\in\mathcal{X}} \left\{ f(x) - \langle \gamma_{k-1}, Ax\rangle_{\mathcal{X}} + \frac{1}{2}\|Ax + By_{k-1} - b\|^2_{\Gamma,H_k} + \frac{1}{2}\|x - x_{k-1}\|^2_{\mathcal{X},R_k} \right\}, \quad (2)$$

$$y_k \in \operatorname{argmin}_{y\in\mathcal{Y}} \left\{ g(y) - \langle \gamma_{k-1}, By\rangle_{\mathcal{Y}} + \frac{1}{2}\|Ax_k + By - b\|^2_{\Gamma,H_k} + \frac{1}{2}\|y - y_{k-1}\|^2_{\mathcal{Y},S_k} \right\}, \quad (3)$$

$$\gamma_k = \gamma_{k-1} - H_k\left(Ax_k + By_k - b\right), \quad (4)$$

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. We start by reviewing some existing methods and works related to the above method.

VM-PADMM and some variants. The VM-PADMM (2)–(4) 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,

- by taking $H_k = \beta I$ with $\beta > 0$, $R_k = 0$ and $S_k = 0$, it reduces to the standard ADMM, whose the ergodic convergence rate was established in [30];
- the ADMM in [21] (related to the Uzawa method [37]) consists of taking $H_k = \beta I$ with $\beta > 0$, R_k constant and $S_k = 0$. Pointwise and ergodic convergence rates for this variant were obtained in [21, 22];
- 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, 10, 16], where convergence rates are analyzed;
- by choosing $H_k = \beta_k I$, $R_k = 0$ and $S_k = 0$, it corresponds to a variable penalty parameter ADMM, for which asymptotic convergence analysis was considered in [20, 23, 35];
- the VM-PADMM (2)–(4) with R_k and S_k positive definite is closely related to the method studied in [19, 26] for solving (point-to-point) continuous monotone variational inequality problems (in the setting of problem (1), it demands f and g to be continuously differentiable). We mention that, contrary to our analysis, the latter references do not present nonasymptotic convergence rates;
- by letting $H_k = \beta I$, $\beta > 0$, the resulting method becomes similar to Algorithm 7 in [2], where a composite structure of f is considered and ergodic convergence rates were obtained under the additional conditions that $B = I$ in (1) and the dual solution set of (1) be bounded.

Contributions of the paper. We obtain an $\mathcal{O}(1/k)$ global convergence rate for an ergodic sequence associated to the VM-PADMM (2)–(4), 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

$$\begin{aligned} &r_x \in \partial_{\varepsilon_x} f(x) - A^*\tilde{\gamma}, \quad r_y \in \partial_{\varepsilon_y} g(y) - B^*\tilde{\gamma}, \quad r_\gamma = Ax + By - b, \\ &\max\left\{\|r_x\|^*_x,\ \|r_y\|^*_y,\ \|r_\gamma\|^*_\gamma\right\} \leq \rho, \\ &\varepsilon_x + \varepsilon_y \leq \varepsilon, \end{aligned} \quad (5)$$

in at most $\mathcal{O}\left(\max\left\{\lceil d_0/\rho\rceil, \lceil d_0^2/\varepsilon\rceil\right\}\right)$ iterations, where $\|\cdot\|_x^*$, $\|\cdot\|_y^*$ and $\|\cdot\|_\gamma^*$ denote dual seminorms associated to the linear operators H_k, R_k and S_k, and d_0 is a scalar measuring the quality of the initial point. Moreover, we establish an $\mathcal{O}(1/\sqrt{k})$ pointwise convergence rate in which the inclusions in (5) are strengthened, in the sense that $\varepsilon_x = \varepsilon_y = 0$, and the bound on the number of iterations becomes $\mathcal{O}\left(\lceil d_0^2/\rho^2\rceil\right)$. Our study is done by first establishing global ergodic and pointwise convergence rates for a variable metric hybrid proximal extragradient (VM-HPE) framework for finding zeroes of maximal monotone operators, and then by showing that the VM-PADMM (2)–(4) 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 VM-PADMM (2)–(4) (resp. VM-HPE framework). Besides, our analysis allows degenerate metrics (induced by positive semidefinite linear operators) which makes the VM-PADMM (2)–(4) (and the VM-HPE framework) more suitable for applications. We next briefly review some related works to the VM-HPE framework.

VM-HPE type frameworks. The VM-HPE framework proposed in this work is a generalization of a special instance of the HPE framework [36] 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 [28] and subsequently applied to the study of several methods; see, for example, [24, 27, 29, 30]. An inexact variable metric proximal point type method was proposed in [32] but, contrary to our VM-HPE framework, it demands the metrics to be nondegenerate (induced by invertible linear operators). Moreover, the convergence analysis presented in [32] does not include nonasymptotic convergence rates.

Outline of the paper. Subsection 1.1 presents our notation and basic results. Section 2 introduces the VM-HPE framework and presents its nonasymptotic pointwise and ergodic convergence rates, whose proofs are postponed to Appendix A. Section 3 contains two subsections. In Subsection 3.1, we formally state the VM-ADMM (2)–(4) and presents its nonasymptotic pointwise and ergodic convergence rates. In Subsection 3.2, we obtain the convergence rates of the VM-ADMM by viewing it as an instance of the VM-HPE framework.

1.1 Basic results and notation

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

$$\langle z, Mz'\rangle \leq \frac{1}{2}\|z\|_{\mathcal{Z},M}^2 + \frac{1}{2}\|z'\|_{\mathcal{Z},M}^2, \tag{6}$$

$$\|z+z'\|_{\mathcal{Z},M}^2 \leq 2\left(\|z\|_{\mathcal{Z},M}^2 + \|z'\|_{\mathcal{Z},M}^2\right). \tag{7}$$

Moreover, each $M \in \mathcal{M}_+^{\mathcal{Z}}$ also induces a (extended) dual seminorm on $\mathcal{Z}$ defined by

$$\|z\|_{\mathcal{Z},M}^* := \sup_{\|z'\|_{\mathcal{Z},M}\leq 1} \langle z, z'\rangle_{\mathcal{Z}} \qquad (z \in \mathcal{Z}).$$

On the other hand, each $M \in \mathcal{M}_{++}^{\mathcal{Z}}$ induces an inner product $\langle M(\cdot),\cdot\rangle_{\mathcal{Z}}$ and a norm $\|\cdot\|_{\mathcal{Z},M} := \sqrt{\langle M(\cdot),\cdot\rangle_{\mathcal{Z}}}$ on $\mathcal{Z}$, etc.

Next two propositions, whose proofs are omitted, will be useful in this paper.

Proposition 1.1. *For every $M \in \mathcal{M}_+^{\mathcal{Z}}$, we have* $\operatorname{dom} \|\cdot\|_{\mathcal{Z},M}^* = \mathcal{R}(M)$ *and* $\|M(\cdot)\|_{\mathcal{Z},M}^* = \|\cdot\|_{\mathcal{Z},M}$, *where $\mathcal{R}(M)$ denotes the range of M.*

Let the partial order $\preceq$ on $\mathcal{M}_+^{\mathcal{Z}}$ be defined by

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

Proposition 1.2. *Let $M, N \in \mathcal{M}_+^{\mathcal{Z}}$ and $c > 0$. If $M \preceq cN$, then*

$$\|\cdot\|_{\mathcal{Z},M} \leq \sqrt{c}\,\|\cdot\|_{\mathcal{Z},N} \;\text{ and }\; \|\cdot\|_{\mathcal{Z},N}^* \leq \sqrt{c}\,\|\cdot\|_{\mathcal{Z},M}^*. \tag{8}$$

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

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

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

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

Recall that the ε-subdifferential of a convex function $f : \mathcal{Z} \to \overline{\mathbb{R}}$ is defined by $\partial_{\varepsilon} f(z) := \{v \in \mathcal{Z} \mid f(z') \geq f(z) + \langle v, z' - z \rangle - \varepsilon \;\; \forall z' \in \mathcal{Z}\}$ for every $z \in \mathcal{Z}$. When $\varepsilon = 0$, then $\partial_0 f(z)$ is denoted by $\partial f(z)$ and is called the *subdifferential* of f at z. The operator ∂f is trivially monotone if f is proper. If f is a proper closed and convex function, then ∂f is also maximal monotone [34].

The following result is a particular case of the *weak transportation formula* in [6, Theorem 2.3] combined with [5, Proposition 2(i)].

Theorem 1.3. *Suppose $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ is maximal monotone and let $\tilde{z}_i, r_i \in \mathcal{Z}$, for $i = 1, \ldots, k$, 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 A variable metric HPE framework

Consider the monotone inclusion problem

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

where $\mathcal{Z}$ is a finite-dimensional inner product real vector space and $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ 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 (VM-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 [36], for which the iteration complexity was recently obtained in [28]. Our main results on pointwise and ergodic convergence rates for the VM-HPE framework are presented in Theorems 2.2 and 2.3, respectively. In Section 3, we will show how the VM-HPE framework can be used to analyze the nonasymptotic convergence of a VM-PADMM for solving linearly constrained convex optimization problems.

We begin by stating the VM-HPE framework.

A variable metric hybrid proximal extragradient (VM-HPE) framework

(0) Let $z_0 \in \mathcal{Z}$, $\eta_0 \in \mathbb{R}_+$ and $\sigma \in [0,1)$ be given, and set $k = 1$.

(1) Choose $M_k \in \mathcal{M}_+^{\mathcal{Z}}$ and find $(z_k, \tilde{z}_k, \eta_k) \in \mathcal{Z} \times \mathcal{Z} \times \mathbb{R}_+$ such that

$$r_k := M_k(z_{k-1} - z_k) \in T(\tilde{z}_k), \tag{10}$$

$$\|z_k - \tilde{z}_k\|^2_{\mathcal{Z},M_k} + \eta_k \leq \sigma \|z_{k-1} - \tilde{z}_k\|^2_{\mathcal{Z},M_k} + \eta_{k-1}. \tag{11}$$

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

end

Remarks. 1) Letting $M_k \equiv I$ and $\eta_k \equiv 0$ in (10) and (11), respectively, we find that the sequences $\{z_k\}$, $\{\tilde{z}_k\}$ and $\{r_k\}$ satisfy

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

which is to say that in this case the VM-HPE framework reduces to a special case of the HPE framework (see pp. 2763 in [28]) with $\lambda_k \equiv 1$ (in the notation of [28]) or, in other words, the VM-HPE framework is a generalization of a special case of the HPE framework in which variations in the metric are allowed along the iterations. 2) If the sequence $\{M_k\}_{k \geq 0}$ is taken to be constant, then the VM-HPE framework reduces to a special case of the NE-HPE framework studied in [16]. 3) We also mention that a variable metric inexact proximal point method with relative error tolerance was proposed in [32] but, contrary to our framework, the method of [32] demands that every operator M_k must be positive definite. Moreover, the convergence analysis presented in [32] does not include nonasymptotic convergence rates. The fact that the VM-HPE framework allows positive semidefinite operators M_k will be crucial for viewing the VM-PADMM of Section 3 as a special instance of it.

From now on in this section, we assume the following condition to hold:

Assumption 2.1. *For the sequence* $\{M_k\}_{k \geq 1}$ *generated by the* VM-HPE framework, *there exist* $M_0 \in \mathcal{M}_+^{\mathcal{Z}}$, $0 \leq C_S < \infty$ *and, for each* $k \geq 0$, $c_k \geq 0$ *such that* $\{c_k\}_{k \geq 0}$ *and* $\{M_k\}_{k \geq 0}$ *satisfy*

$$\sum_{i=0}^{k} c_i \leq C_S, \qquad \frac{1}{1+c_k} M_k \preceq M_{k+1} \preceq (1+c_k) M_k \quad \forall\, k \geq 0. \tag{12}$$

Remark. The above assumption (which is similar to condition (1.4) in [32]) 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 Assumption 2.1 implies the existence of a constant $C_P > 0$ such that $\{c_k\}_{k\geq 0}$ and $\{M_k\}_{k\geq 0}$ satisfy

$$\prod_{i=0}^{k}(1+c_i) \leq C_P \quad \text{and} \quad M_j \preceq C_P M_k, \quad \forall\, j,k \geq 0. \tag{13}$$

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

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

which measures the "quality" of the initial guess $z_0 \in \mathcal{Z}$ in the VM-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.

Theorem 2.2. (Pointwise convergence rate of the VM-HPE framework) *Let $\{\tilde{z}_k\}$, $\{r_k\}$ and $\{M_k\}$ be generated by the* VM-HPE framework. *Let also C_P and d_0 be as in* (13) *and* (14), *respectively. Then, for every $k \geq 1$, $r_k \in T(\tilde{z}_k)$ and there exists $i \leq k$ such that*

$$\|r_i\|^*_{\mathcal{Z},M_i} \leq \left(\frac{2(1+\sigma)C_P(d_0^2+\eta_0) + 2(1-\sigma)\eta_0}{(1-\sigma)k}\right)^{1/2}. \tag{15}$$

Remarks. 1) If $c_k \equiv 0$ in Assumption 2.1 (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 [16, 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 [28, Theorem 4.4(a)]. 2) For a given tolerance $\rho > 0$, Theorem 2.2 ensures that there exists an index

$$i = \mathcal{O}\left(\left\lceil \frac{C_p(d_0^2+\eta_0)}{\rho^2} \right\rceil\right) \tag{16}$$

such that

$$r_i \in T(\tilde{z}_i) \;\; \text{and} \;\; \|r_i\|^*_{\mathcal{Z},M_i} \leq \rho. \tag{17}$$

In this case, $\tilde{z}_i \in \mathcal{Z}$ can be interpreted as a ρ-approximate solution of (9) with residual $r_i \in \mathcal{Z}$ (see, e.g., [28] for the definition of a related concept).

Before presenting the ergodic convergence of the VM-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 2.3. (Ergodic convergence rate of the VM-HPE framework) *Let* $\{\tilde{z}_k^a\}$, $\{r_k^a\}$ *and* $\{\varepsilon_k^a\}$ *be given as in* (18) *and* $\{M_k\}$ *be generated by the* VM-HPE framework. *Let also* C_S, C_P *and* d_0 *be as in* (12), (13) *and* (14), *respectively. Then, for every* $k \geq 1$, *we have* $r_k^a \in T^{\varepsilon_k^a}(\tilde{z}_k^a)$ *and*

$$\|r_k^a\|_{\mathcal{Z},M_k}^* \leq \frac{\mathcal{E}\sqrt{d_0^2+\eta_0}}{k}, \tag{19}$$

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

where $\mathcal{E} := (1+C_P)\left(\sqrt{C_P}+C_SC_P\right)+C_SC_P^{3/2}$ *and* $\widehat{\mathcal{E}} := 2C_P(1+C_S)\left[\sigma C_P/(1-\sigma)+2(1+C_P)\right]$.

Remarks. 1) Similarly to the first remark after Theorem 2.2, Theorem 2.3 is also related to [16, Theorem 3.4] and [28, Theorem 4.7]. 2) For given tolerances $\rho, \varepsilon > 0$, Theorem 2.3 ensures that in at most

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

iterations there hold

$$r_k^a \in T^{\varepsilon_k^a}(\tilde{z}_k^a), \quad \|r_k^a\|_{\mathcal{Z},M_k}^* \leq \rho \ \text{ and } \ \varepsilon_k^a \leq \varepsilon. \tag{22}$$

Note that (21), in terms of the dependence on $\rho > 0$, is better than the bound in (16) by a factor of $\mathcal{O}(\rho)$ 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).

3 A variable metric proximal alternating direction method of multipliers

This section contains two subsections. In Subsection 3.1, we formally state the VM-PADMM (2)–(4) and present its nonasymptotic convergence rates. The main results are Theorems 3.2 and 3.3 in which pointwise and ergodic convergence rates are obtained, respectively. The proofs of the latter theorems are discussed separately in Subsection 3.2 by viewing the method as an instance of the VM-HPE framework and by applying the results of Section 2.

3.1 VM-PADMM and its convergence rates

Let $\mathcal{X}$, $\mathcal{Y}$ and Γ be finite-dimensional real inner product vector spaces. Consider the convex optimization problem (1), i.e.,

$$\begin{aligned} &\text{minimize} \quad f(x)+g(y) \\ &\text{subject to} \quad Ax+By=b, \end{aligned} \tag{23}$$

where the following assumptions are assumed to hold:

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

(O2) $A : \mathcal{X} \to \Gamma$ and $B : \mathcal{Y} \to \Gamma$ 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.,[33, Corollaries 28.2.2 and 28.3.1]), a vector $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$ is a solution of (23) if and only if 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

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

which is assumed to be nonempty.

The convergence rates of the VM-PADMM (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 (45)) in $\mathcal{X} \times \mathcal{Y} \times \Gamma$, and by applying the results of the previous section.

Variable metric proximal alternating direction method of multipliers (VM-PADMM).

(0) Let $(x_0, y_0, \gamma_0) \in \mathcal{X} \times \mathcal{Y} \times \Gamma$ be given, and set $k = 1$.

(1) Choose $R_k \in \mathcal{M}_+^{\mathcal{X}}$, $S_k \in \mathcal{M}_+^{\mathcal{Y}}$ and $H_k \in \mathcal{M}_{++}^{\Gamma}$ and compute an optimal solution $x_k \in \mathcal{X}$ of the subproblem

$$\min_{x \in \mathcal{X}} \left\{ f(x) - \langle \gamma_{k-1}, Ax \rangle_{\mathcal{X}} + \frac{1}{2}\|Ax + By_{k-1} - b\|^2_{\Gamma, H_k} + \frac{1}{2}\|x - x_{k-1}\|^2_{\mathcal{X}, R_k} \right\} \tag{26}$$

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

$$\min_{y \in \mathcal{Y}} \left\{ g(y) - \langle \gamma_{k-1}, By \rangle_{\mathcal{Y}} + \frac{1}{2}\|Ax_k + By - b\|^2_{\Gamma, H_k} + \frac{1}{2}\|y - y_{k-1}\|^2_{\mathcal{Y}, S_k} \right\}. \tag{27}$$

(2) Set

$$\gamma_k = \gamma_{k-1} - H_k\left(Ax_k + By_k - b\right), \tag{28}$$

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

end

Remarks. 1) As already mentioned in Section 1, the VM-PADMM can be regarded as a class of ADMM instances, allowing a unified study of different variants of ADMM. 2) An usual choice for the linear operator H_k is $\beta_k I$, where $\beta_k > 0$ plays the role of a penalty parameter. 3) The proximal terms in (26) and (27) defined by R_k and S_k, respectively, may have different roles. Namely, they can be used to regularize the subproblems in (26) and (27), 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 (26) and (27) may become much easier to solve; for instance, if $H_k = \beta_k I$, then $R_k = \tau_k I - \beta_k A^*A$ with $\tau_k > \beta_k\|A^*A\|$ and $S_k = \theta_k I - \beta_k B^*B$ with $\theta_k > \beta_k\|B^*B\|$ eliminate the presence of quadratic forms associated to A^*A and B^*B in (26) and (27), respectively.

From now on in this section, the following conditions are assumed to hold:

Assumption 3.1. *For the sequences* $\{R_k\}_{k\geq 1}$, $\{S_k\}_{k\geq 1}$ *and* $\{H_k\}_{k\geq 1}$ *generated by the* VM-PADMM*, there exist* $R_0 \in \mathcal{M}_+^{\mathcal{X}}$, $S_0 \in \mathcal{M}_+^{\mathcal{Y}}$, $H_0 \in \mathcal{M}_{++}^{\Gamma}$, $0 \leq C_S < \infty$ *and, for each* $k \geq 0$, $c_k \in [0,1]$ *such that* $\{c_k\}_{k\geq 0}$, $\{Q_{k,1} := R_k\}_{k\geq 0}$, $\{Q_{k,2} := S_k\}_{k\geq 0}$ *and* $\{Q_{k,3} := H_k\}_{k\geq 0}$ *satisfy*

$$\sum_{i=0}^{k} c_i \leq C_S, \quad \frac{1}{1+c_k} Q_{k,j} \preceq Q_{k+1,j} \preceq (1+c_k) Q_{k,j} \quad \forall k \geq 0, \;\; j = 1,2,3. \tag{29}$$

Analogously to condition (13), assumption 3.1 implies the existence of $C_P > 0$ such that $\{c_k\}_{k\geq 0}$ satisfies

$$\prod_{i=0}^{k} (1 + c_i) \leq C_P \qquad \forall k \geq 0. \tag{30}$$

We mention that Assumption 3.1 is similar to Condition C in [19] but, contrary to the latter reference, none of the operators R_k and S_k is assumed to be positive definite.

Similarly to the previous section, the following quantity will be needed:

$$d_0 := \inf \left\{ \left(\|x_0 - x^*\|^2_{\mathcal{X},R_0} + \|y_0 - y^*\|^2_{\mathcal{Y},(B^* H_0 B + S_0)} + \|\gamma_0 - \gamma^*\|^2_{\Gamma, H_0^{-1}} \right)^{1/2} \mid (x^*, y^*, \gamma^*) \in \Omega^* \right\}, \tag{31}$$

where (x_0, y_0, γ_0) is given in Step (0) of the VM-PADMM, $R_0 \in \mathcal{M}_+^{\mathcal{X}}$, $S_0 \in \mathcal{M}_+^{\mathcal{Y}}$ and $H_0 \in \mathcal{M}_{++}^{\Gamma}$ are given in Assumption 3.1, and Ω^* is defined in (25).

Next we present the two main results of this paper, whose proofs are given in Subsection 3.2.

Theorem 3.2. (Pointwise convergence rate of the VM-PADMM) *Let* $\{(x_k, y_k, \gamma_k)\}$, $\{R_k\}$, $\{S_k\}$ *and* $\{H_k\}$ *be generated by the* VM-PADMM *and let*

$$\tilde{\gamma}_k := \gamma_{k-1} - H_k(Ax_k + By_{k-1} - b) \quad \forall\, k \geq 1. \tag{32}$$

Let also C_P *and* d_0 *be as in* (30) *and* (31)*, respectively. Then, for all* $k \geq 1$,

$$\begin{pmatrix} r_{k,x} \\ r_{k,y} \\ r_{k,\gamma} \end{pmatrix} := \begin{pmatrix} R_k(x_{k-1} - x_k) \\ (B^* H_k B + S_k)(y_{k-1} - y_k) \\ H_k^{-1}(\gamma_{k-1} - \gamma_k) \end{pmatrix} \in \begin{pmatrix} \partial f(x_k) - A^* \tilde{\gamma}_k \\ \partial g(y_k) - B^* \tilde{\gamma}_k \\ Ax_k + By_k - b \end{pmatrix} \tag{33}$$

and there exists $i \leq k$ *such that*

$$\max \left\{ \|r_{i,x}\|^*_{\mathcal{X},R_i}\,,\; \|r_{i,y}\|^*_{\mathcal{Y},(B^* H_i B + S_i)}\,,\; \|r_{i,\gamma}\|^*_{\Gamma, H_i^{-1}} \right\} \leq \frac{d_0}{\sqrt{k}} \sqrt{2(15 C_p + 4)}. \tag{34}$$

Remark. For a given tolerance $\rho > 0$, Theorem 3.2 guarantees the existence of triples $(x, y, \tilde{\gamma})$, (r_x, r_y, r_γ) and operators $R \in \mathcal{M}_+^{\mathcal{X}}$, $S \in \mathcal{M}_+^{\mathcal{Y}}$ and $H \in \mathcal{M}_{++}^{\Gamma}$ (generated by the VM-PADMM) such that

$$\begin{aligned} &r_x \in \partial f(x) - A^* \tilde{\gamma}, \quad r_y \in \partial g(y) - B^* \tilde{\gamma}, \quad r_\gamma = Ax + By - b, \\ &\max \left\{ \|r_x\|^*_{\mathcal{X},R}\,,\; \|r_y\|^*_{\mathcal{Y},(B^* H B + S)}\,,\; \|r_\gamma\|^*_{\Gamma, H^{-1}} \right\} \leq \rho, \end{aligned} \tag{35}$$

in at most

$$\mathcal{O}\left(\left\lceil \frac{C_p\, d_0^2}{\rho^2} \right\rceil\right) \tag{36}$$

iterations, where C_P and d_0 are as in (30) and (31), respectively. The triple $(x, y, \tilde{\gamma})$ in (35) can be seen as a ρ-approximate solution of the KKT system (24) with residual (r_x, r_y, r_γ).

Before proceeding to present the ergodic convergence of the VM-PADMM we need to introduce its associated ergodic sequences. Let $\{(x_k, y_k, \gamma_k)\}$ be generated by the VM-PADMM, let $\{\tilde{\gamma}_k\}$ and $\{(r_{k,x}, r_{k,y}, r_{k,\gamma})\}$ be defined as in (32) and (33), respectively, and let the *ergodic* sequences associated to them be defined by

$$(x_k^a, y_k^a) := \frac{1}{k}\sum_{i=1}^{k}(x_i, y_i)\,, \quad \tilde{\gamma}_k^a := \frac{1}{k}\sum_{i=1}^{k}\tilde{\gamma}_i, \tag{37}$$

$$(r_{k,x}^a, r_{k,y}^a, r_{k,\gamma}^a) := \frac{1}{k}\sum_{i=1}^{k}(r_{i,x}, r_{i,y}, r_{i,\gamma}), \tag{38}$$

$$(\varepsilon_{k,x}^a, \varepsilon_{k,y}^a) := \frac{1}{k}\sum_{i=1}^{k}\left(\langle r_{i,x} + A^*\tilde{\gamma}_i, x_i - x_k^a\rangle_{\mathcal{X}},\ \langle r_{i,y} + B^*\tilde{\gamma}_i, y_i - y_k^a\rangle_{\mathcal{Y}}\right). \tag{39}$$

Theorem 3.3. (Ergodic convergence rate of the VM-PADMM) *Let $\{R_k\}$, $\{S_k\}$ and $\{H_k\}$ be generated by the* VM-PADMM *and 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* (37)*–*(39)*. Let also C_S, C_P, and d_0 be as in* (29)*,* (30) *and* (31)*, respectively. Then, for all $k \geq 1$, we have $\varepsilon_{k,x}^a$, $\varepsilon_{k,y}^a \geq 0$,*

$$\begin{pmatrix} r_{k,x}^a \\ r_{k,y}^a \\ r_{k,\gamma}^a \end{pmatrix} \in \begin{pmatrix} \partial f_{\varepsilon_{k,x}^a}(x_k^a) - A^*\tilde{\gamma}_k^a \\ \partial g_{\varepsilon_{k,y}^a}(y_k^a) - B^*\tilde{\gamma}_k^a \\ Ax_k^a + By_k^a - b \end{pmatrix} \tag{40}$$

and

$$\max\left\{\|r_{k,x}^a\|_{\mathcal{X},R_k}^*,\ \|r_{k,y}^a\|_{\mathcal{Y},(B^*H_kB^*+S_k)}^*,\ \|r_{k,\gamma}^a\|_{\Gamma,H_k^{-1}}^*\right\} \leq \frac{\sqrt{5}\mathcal{E}\, d_0}{k}\,, \tag{41}$$

$$\varepsilon_{k,x}^a + \varepsilon_{k,y}^a \leq \frac{\widetilde{\mathcal{E}} d_0^2}{k}, \tag{42}$$

where $\mathcal{E}$ is defined as in Theorem 2.3 and $\widetilde{\mathcal{E}} := 10C_P(1 + C_S)\,(2 + 3C_P)$.

Remark. Given tolerances $\rho, \varepsilon > 0$, Theorem 3.3 guarantees that there exist scalars $\varepsilon_x, \varepsilon_y \geq 0$, triples $(x, y, \tilde{\gamma})$, (r_x, r_y, r_γ) and operators $R \in \mathcal{M}_+^{\mathcal{X}}$, $S \in \mathcal{M}_+^{\mathcal{Y}}$ and $H \in \mathcal{M}_{++}^{\Gamma}$ (generated by the VM-PADMM) such that

$$\begin{aligned} & r_x \in \partial_{\varepsilon_x} f(x) - A^*\tilde{\gamma}, \quad r_y \in \partial_{\varepsilon_y} g(y) - B^*\tilde{\gamma}, \quad r_\gamma = Ax + By - b, \\ & \max\left\{\|r_x\|_{\mathcal{X},R}^*\,,\ \|r_y\|_{\mathcal{Y},(B^*HB+S)}^*\,,\ \|r_\gamma\|_{\Gamma,H^{-1}}^*\right\} \leq \rho, \\ & \varepsilon_x + \varepsilon_y \leq \varepsilon, \end{aligned} \tag{43}$$

in at most

$$\mathcal{O}\left((1+C_S)C_p^2 \max\left\{\left\lceil\frac{d_0}{\rho}\right\rceil, \left\lceil\frac{d_0^2}{\varepsilon}\right\rceil\right\}\right) \tag{44}$$

iterations, where C_S, C_P and d_0 are as in Assumption 3.1, (30) and (31), respectively. Note that while the dependence on the tolerance ρ in (44) is better than the corresponding one in (36) by a factor of $\mathcal{O}(\rho)$, the inclusions in (43) are potentially weaker than the corresponding ones in (35). The triple $(x, y, \tilde{\gamma})$ in (43) can be seen as a (ρ, ε)-approximate solution of the KKT system (24) with residual (r_x, r_y, r_γ).

3.2 Proof of Theorems 3.2 and 3.3

The main goal of this subsection is to prove Theorems 3.2 and 3.3 by viewing the VM-PADMM as an instance of the VM-HPE framework of Section 2 for solving (9) with $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ defined by

$$T(z) := \begin{pmatrix} \partial f(x) - A^*\gamma \\ \partial g(y) - B^*\gamma \\ Ax + By - b \end{pmatrix}, \qquad \forall z := (x, y, \gamma) \in \mathcal{Z} \tag{45}$$

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

$$\langle z, z' \rangle_{\mathcal{Z}} := \langle x, x' \rangle_{\mathcal{X}} + \langle y, y' \rangle_{\mathcal{Y}} + \langle \gamma, \gamma' \rangle_{\Gamma}. \tag{46}$$

The desired results will then follow essentially from Theorems 2.2 and 2.3, and from the identity

$$T^{-1}(0) = \Omega^*, \tag{47}$$

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:

$$M_k := \begin{pmatrix} R_k & 0 & 0 \\ 0 & B^*H_kB + S_k & 0 \\ 0 & 0 & {H_k}^{-1} \end{pmatrix} : \mathcal{Z} \to \mathcal{Z} \qquad \forall\, k \geq 0, \tag{48}$$

where $\{R_k\}_{k\geq 1}$, $\{S_k\}_{k\geq 1}$ and $\{H_k\}_{k\geq 1}$ are generated by the VM-PADMM and $R_0 \in \mathcal{M}_+^{\mathcal{X}}$, $S_0 \in \mathcal{M}_+^{\mathcal{Y}}$, $H_0 \in \mathcal{M}_{++}^{\Gamma}$ are given in Assumption 3.1.

We begin by presenting a preliminary technical result.

Proposition 3.4. *Let $\{(x_k, y_k, \gamma_k)\}$ be generated by the* VM-PADMM *and let $\{\tilde{\gamma}_k\}$ be defined as in* (32)*. Let also $\{M_k\}$ be defined as in* (48)*. Then,*

$$M_k \begin{pmatrix} x_{k-1} - x_k \\ y_{k-1} - y_k \\ \gamma_{k-1} - \gamma_k \end{pmatrix} \in \begin{pmatrix} \partial f(x_k) - A^*\tilde{\gamma}_k \\ \partial g(y_k) - B^*\tilde{\gamma}_k \\ Ax_k + By_k - b \end{pmatrix} \qquad \forall\, k \geq 1. \tag{49}$$

Proof. From the first order optimality conditions for (26) and (27), 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 (32), yields

$$R_k(x_{k-1} - x_k) \in \partial f(x_k) - A^*\tilde{\gamma}_k, \quad (B^*H_kB + S_k)(y_{k-1} - y_k) \in \partial g(y_k) - B^*\tilde{\gamma}_k. \tag{50}$$

On the other hand, (28) (and the assumption $H_k \in \mathcal{M}_{++}^{\Gamma}$) gives

$$H_k^{-1}(\gamma_{k-1} - \gamma_k) = Ax_k + By_k - b. \tag{51}$$

Using (48), (50) and (51) we obtain (49). □

The next lemma will allow us to use the main results of Section 2 for analyzing the nonasymptotic convergence of the VM-PADMM.

Lemma 3.5. *The sequence* $\{M_k\}_{k\geq 0}$ *defined in* (48)*, the scalar* C_S *and the sequence* $\{c_k\}$ *given in* Assumption 3.1 *satisfy condition* (12) *of* Assumption 2.1.

Proof. Note that the first condition in (29) is identical to the first one in (12). To finish the proof, note that the second condition in (29), which by Assumption 3.1 is assumed to hold for $\{R_k\}_{k\geq 0}$, $\{S_k\}_{k\geq 0}$ and $\{H_k\}_{k\geq 0}$, combined with the (block) diagonal structure of M_k gives the second condition in (12) for $\{c_k\}_{k\geq 0}$ and $\{M_k\}_{k\geq 0}$. □

The following technical result will be used to prove that the VM-PADMM is an instance of the VM-HPE framework.

Lemma 3.6. *Let* $\{(x_k, y_k, \gamma_k)\}$*,* $\{S_k\}$ *and* $\{H_k\}$ *be generated by the* VM-PADMM *and let* $\{\tilde{\gamma}_k\}$ *be defined as in* (32)*. Let also* d_0 *be defined as in* (31)*. Then, the following hold:*

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

$$\tilde{\gamma}_k - \gamma_k = H_kB(y_k - y_{k-1});$$

(b) *we have*

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

(c) *for any* $k \geq 2$*, we have*

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

Proof. (a) This item follows trivially from (28) and (32).

(b) First note that

$$\begin{aligned} 0 &\leq \frac{1}{2}\|\gamma_1 - \gamma_0 + H_1B(y_1 - y_0)\|^2_{\Gamma,H_1^{-1}} \\ &= \frac{1}{2}\|\gamma_1 - \gamma_0\|^2_{\Gamma,H_1^{-1}} + \langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle_{\Gamma} + \frac{1}{2}\|B(y_1 - y_0)\|^2_{\Gamma,H_1}, \end{aligned}$$

which combined with the property (7) yields, for all $z^* := (x^*, y^*, \gamma^*) \in \Omega^*$,

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

Direct use of the above inequality and (48) yields

$$\frac{1}{2}\|y_1 - y_0\|^2_{\mathcal{Y},S_1} - \langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle_\Gamma \leq \|z_1 - z^*\|^2_{\mathcal{Z},M_1} + \|z_0 - z^*\|^2_{\mathcal{Z},M_1}, \tag{52}$$

where $z_0 := (x_0, y_0, \gamma_0)$ and $z_1 := (x_1, y_1, \gamma_1)$. Note now that letting $\tilde{z}_1 := (x_1, y_1, \tilde{\gamma}_1)$, it follows from (48), item (a) and some direct calculations that

$$\|z_1 - \tilde{z}_1\|^2_{\mathcal{Z},M_1} = \|\gamma_1 - \tilde{\gamma}_1\|^2_{\Gamma,H_1^{-1}} = \|B(y_1 - y_0)\|^2_{\Gamma,H_1}. \tag{53}$$

Moreover, using (48) with $k = 1$, we find

$$\begin{aligned}\|z_0 - \tilde{z}_1\|^2_{\mathcal{Z},M_1} &= \|x_0 - x_1\|^2_{\mathcal{X},R_1} + \|y_0 - y_1\|^2_{\mathcal{Y},(B^*H_1B+S_1)} + \|\gamma_0 - \tilde{\gamma}_1\|^2_{\Gamma,H_1^{-1}}\\ &\geq \|y_0 - y_1\|^2_{\mathcal{Y},(B^*H_1B+S_1)} \geq \|B(y_1 - y_0)\|^2_{\Gamma,H_1}.\end{aligned} \tag{54}$$

From Proposition 3.4 and (48) with $k = 1$, we have $r_1 := M_1(z_0 - z_1) \in T(\tilde{z}_1)$, where T is given in (45). Using this fact, (47) and the monotonicity of T we obtain $\langle \tilde{z}_1 - z^*, r_1\rangle \geq 0$ 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 $M = M_1$, (53) and (54) we have, for all $z^* = (x^*, y^*, z^*) \in \Omega^*$,

$$\begin{aligned}\|z^* - z_0\|^2_{\mathcal{Z},M_1} &\geq \|z^* - z_1\|^2_{\mathcal{Z},M_1} + \|z_0 - \tilde{z}_1\|^2_{\mathcal{Z},M_1} - \|z_1 - \tilde{z}_1\|^2_{\mathcal{Z},M_1}\\ &\geq \|z^* - z_1\|^2_{\mathcal{Z},M_1} + \|B(y_1 - y_0)\|^2_{\Gamma,H_1} - \|B(y_1 - y_0)\|^2_{\Gamma,H_1} = \|z^* - z_1\|^2_{\mathcal{Z},M_1}.\end{aligned} \tag{55}$$

From Assumption 3.1 and Lemma 3.5 we know that $M_1 \preceq (1 + c_0)M_0 \preceq 2M_0$, which combined with Proposition 1.2 and (55) yields

$$\|z^* - z_1\|^2_{\mathcal{Z},M_1} \leq \|z^* - z_0\|^2_{\mathcal{Z},M_1} \leq 2\|z^* - z_0\|^2_{\mathcal{Z},M_0}. \tag{56}$$

Combining (31), (52), (55), (56) and taking the infimum over all $z^* \in \Omega^*$, we find the desired inequality, whence item (b).

(c) Using the first order optimality condition for (27), and (28), we obtain

$$B^*\gamma_k - S_k(y_k - y_{k-1}) \in \partial g(y_k) \quad \forall\, k \geq 1.$$

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 (6), we find

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

where the last inequality is due to Proposition 1.2 and Assumption 3.1, and so the proof of the lemma follows. □

Next we show that the VM-PADMM can be regarded as an instance of the VM-HPE framework.

Proposition 3.7. *Let* $\{(x_k, y_k, \gamma_k)\}$ *be generated by the* VM-PADMM *and let* $\{\tilde{\gamma}_k\}$ *and* $\{M_k\}$ *be defined as in* (32) *and* (48)*, respectively. Let also* d_0 *and* T *be defined as in* (31) *and* (45)*, respectively. Define* $z_0 := (x_0, y_0, \gamma_0)$*,* $\eta_0 := 4d_0^2$ *and, for all* $k \geq 1$*,*

$$z_k := (x_k, y_k, \gamma_k), \quad \tilde{z}_k := (x_k, y_k, \tilde{\gamma}_k), \quad r_k := M_k(z_{k-1} - z_k), \quad \eta_k := \frac{1}{2}\|y_{k-1} - y_k\|^2_{\mathcal{Y},S_k}. \tag{57}$$

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

$$\begin{aligned} &r_k \in T(\tilde{z}_k), \\ &\|z_k - \tilde{z}_k\|^2_{\mathcal{Z},M_k} + \eta_k \leq \frac{1}{2}\|z_{k-1} - \tilde{z}_k\|^2_{\mathcal{Z},M_k} + \eta_{k-1}. \end{aligned} \tag{58}$$

As a consequence, the VM-PADMM *falls within the* VM-HPE *framework (with input* z_0*,* η_0 *and* $\sigma = 1/2$*) for solving* (9) *with* T *as in* (45)*.*

Proof. First note that the inclusion in (58) follows from (45), (49) and the definitions of z_k, $\tilde{z}_k$ and r_k in (57). Now, using (46), (48), (57) and some direct calculations, we obtain

$$\begin{aligned} \|z_{k-1} - \tilde{z}_k\|^2_{\mathcal{Z},M_k} = &\|x_{k-1} - x_k\|^2_{\mathcal{X},R_k} + \|B(y_{k-1} - y_k)\|^2_{\Gamma,H_k} + \|y_{k-1} - y_k\|^2_{\mathcal{Y},S_k} \\ &+ \|\gamma_{k-1} - \tilde{\gamma}_k\|^2_{\Gamma,H_k^{-1}}. \end{aligned} \tag{59}$$

Using the same reasoning and Lemma 3.6(a), we also find

$$\|z_k - \tilde{z}_k\|^2_{\mathcal{Z},M_k} = \|\gamma_k - \tilde{\gamma}_k\|^2_{\Gamma,H_k^{-1}} = \|B(y_{k-1} - y_k)\|^2_{\Gamma,H_k}. \tag{60}$$

Hence, from the first identity in (60), Lemma 3.6(a) and some algebraic manipulations, we obtain

$$\begin{aligned} \frac{1}{2}\|\gamma_{k-1} - \tilde{\gamma}_k\|^2_{\Gamma,H_k^{-1}} - \|z_k - \tilde{z}_k\|^2_{\mathcal{Z},M_k} &= \frac{1}{2}\|\gamma_{k-1} - \gamma_k\|^2_{\Gamma,H_k^{-1}} + \frac{1}{2}\|\gamma_k - \tilde{\gamma}_k\|^2_{\Gamma,H_k^{-1}} \\ &\quad + \langle \gamma_{k-1} - \gamma_k, H_k^{-1}(\gamma_k - \tilde{\gamma}_k)\rangle_\Gamma - \|z_k - \tilde{z}_k\|^2_{\mathcal{Z},M_k} \\ &= \frac{1}{2}\|\gamma_{k-1} - \gamma_k\|^2_{\Gamma,H_k^{-1}} - \frac{1}{2}\|\gamma_k - \tilde{\gamma}_k\|^2_{\Gamma,H_k^{-1}} \\ &\quad - \langle \gamma_{k-1} - \gamma_k, B(y_k - y_{k-1})\rangle_\Gamma, \end{aligned}$$

which in turn, combined with (59) and (60), yields

$$\begin{aligned} \frac{1}{2}\|z_{k-1} - \tilde{z}_k\|^2_{\mathcal{Z},M_k} - \|z_k - \tilde{z}_k\|^2_{\mathcal{Z},M_k} &= \frac{1}{2}\|x_{k-1} - x_k\|^2_{\mathcal{X},R_k} + \frac{1}{2}\|y_{k-1} - y_k\|^2_{\mathcal{Y},S_k} \\ &\quad + \frac{1}{2}\|\gamma_{k-1} - \gamma_k\|^2_{\Gamma,H_k^{-1}} + \langle B(y_k - y_{k-1}), \gamma_k - \gamma_{k-1}\rangle_\Gamma \\ &\geq \frac{1}{2}\|y_{k-1} - y_k\|^2_{\mathcal{Y},S_k} + \langle B(y_k - y_{k-1}), \gamma_k - \gamma_{k-1}\rangle_\Gamma. \end{aligned} \tag{61}$$

We will now consider two cases: $k = 1$ and $k > 1$. In the first case, it follows from Lemma 3.6(b), (57), (61) with $k = 1$ and $\eta_0 = 4d_0^2$ that

$$\frac{1}{2}\|z_0 - \tilde{z}_1\|^2_{\mathcal{Z},M_1} - \|z_1 - \tilde{z}_1\|^2_{M_1} - \eta_1 \geq \langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle_\Gamma \geq \frac{1}{2}\|y_1 - y_0\|^2_{\mathcal{Y},S_1} - 4d_0^2 \geq -\eta_0,$$

which gives (58) for $k = 1$. On the other hand, assuming $k > 1$, from Lemma 3.6(c), (57), (61) and some manipulations, we have

$$\begin{aligned}\frac{1}{2}\|z_{k-1} - \tilde{z}_k\|^2_{\mathcal{Z},M_k} - \|z_k - \tilde{z}_k\|^2_{\mathcal{Z},M_k} &\geq \frac{2 - c_{k-1}}{2}\|y_{k-1} - y_k\|^2_{\mathcal{Y},S_k} - \frac{1}{2}\|y_{k-2} - y_{k-1}\|^2_{\mathcal{Y},S_{k-1}} \\ &\geq \eta_k - \eta_{k-1},\end{aligned}$$

where the last inequality is due to the fact that $c_{k-1} \leq 1$ (see Assumption 3.1). Hence, we conclude that (58) holds for all $k \geq 1$. The last statement of the proposition follows directly from (58) and VM-HPE framework's definition. □

We are now ready to prove Theorems 3.2 and 3.3.

Proof of Theorem 3.2: Using Proposition 3.7 and Theorem 2.2, we conclude that, for every $k \geq 1$, (33) holds and there exists $i \leq k$ such that

$$\|M_i(z_{i-1} - z_i)\|^*_{\mathcal{Z},M_i} \leq \frac{d_0}{\sqrt{k}}\sqrt{2(15C_p + 4)}, \tag{62}$$

where $\{M_k\}$ and $\{z_k\}$ are defined in (48) and (57), respectively. Hence, using Proposition 1.1, we obtain

$$\begin{aligned}\|M_i(z_{i-1} - z_i)\|^*_{\mathcal{Z},M_i} &= \|z_{i-1} - z_i\|_{\mathcal{Z},M_i} \\ &= \left(\|x_{i-1} - x_i\|^2_{\mathcal{X},R_i} + \|y_{i-1} - y_i\|^2_{\mathcal{Y},(B^*H_iB+S_i)} + \|\gamma_{i-1} - \gamma_i\|^2_{\Gamma,H_i}\right)^{1/2}.\end{aligned} \tag{63}$$

On the other hand, using Proposition 1.1 and the definition in (33), we find

$$\begin{aligned}&\|x_{i-1} - x_i\|_{\mathcal{X},R_i} = \|R_i(x_{i-1} - x_i)\|^*_{\mathcal{X},R_i} = \|r_{i,x}\|^*_{\mathcal{X},R_i}, \\ &\|y_{i-1} - y_i\|_{\mathcal{Y},(B^*H_iB+S_i)} = \|(B^*H_iB + S_i)(y_{i-1} - y_i)\|^*_{\mathcal{Y},(B^*H_iB+S_i)} = \|r_{i,y}\|^*_{\mathcal{Y},(B^*H_iB+S_i)}, \\ &\|\gamma_{i-1} - \gamma_i\|_{\Gamma,H_i^{-1}} = \|H_i^{-1}(\gamma_{i-1} - \gamma_i)\|^*_{\Gamma,H_i^{-1}} = \|r_{i,\gamma}\|^*_{\Gamma,H_i^{-1}},\end{aligned}$$

which, combined with (62) and (63), proves (34). □

Proof of Theorem 3.3: Combining Proposition 3.7 and Theorem 2.3, 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 $k \geq 1$,

$$\max\left\{\|r^a_{k,x}\|^*_{\mathcal{X},R_k}, \|r^a_{k,y}\|^*_{\mathcal{Y},(B^*H_kB^*+S_k)}, \|r^a_{k,\gamma}\|^*_{\Gamma,H_k^{-1}}\right\} \leq \|(r^a_{k,x}, r^a_{k,y}, r^a_{k,\gamma})\|^*_{\mathcal{Z},M_k} \leq \frac{\sqrt{5}\mathcal{E}\, d_0}{k}, \tag{64}$$

$$\varepsilon^a_k = \frac{1}{k}\left(\sum_{i=1}^k \langle r_{i,x}, x_i - x^a_k\rangle_{\mathcal{X}} + \sum_{i=1}^k \langle r_{i,y}, y_i - y^a_k\rangle_{\mathcal{Y}} + \sum_{i=1}^k \langle r_{i,\gamma}, \tilde{\gamma}_i - \tilde{\gamma}^a_k\rangle_{\Gamma}\right) \leq \frac{\widetilde{\mathcal{E}} d_0^2}{k}, \tag{65}$$

where $\mathcal{E}$ and $\widetilde{\mathcal{E}}$ are defined as in Theorem 2.3 and Theorem 3.3, respectively. On the other hand, (33), (37) and (38) yield

$$Ax_k + By_k = r_{k,\gamma} + b, \quad Ax^a_k + By^a_k = r^a_{k,\gamma} + b.$$

Additionally, (37), (38) and some algebraic manipulations give

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

Hence, combining the identity in (65) with the last two displayed equations, we also find

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

where the last equality is due to the definitions of $\varepsilon_{k,x}^a$ and $\varepsilon_{k,y}^a$ in (39). Therefore, the inequalities in (41) and (42) now follows from (64) and (65), respectively.

To finish the proof of the theorem, note that direct use of Theorem 1.3(b) (for f and g), (33) and (37)–(39) give $\varepsilon_{k,x}^a, \varepsilon_{k,y}^a \geq 0$ and (40). □

A Proof of Theorems 2.2 and 2.3

We start by presenting the following two Lemmas.

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

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

Proof. Direct calculations yield

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

□

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

(a) *we have*

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

(b) *we have*

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

where C_P and M_0 are as in (13) *and* Assumption 2.1*, respectively.*

Proof. (a) From Lemma A.1 with $(z, z_+, \tilde{z}) = (z_{k-1}, z_k, \tilde{z}_k)$ and $M = M_k$, (10) and (11), we obtain

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

Hence, (a) follows from the above inequality, the fact that $0 \in T(z^*)$ and $r_k \in T(\tilde{z}_k)$ (see (10)), and the monotonicity of T.

(b) Using (a), (8) and Assumption 2.1, we find

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

Thus, the result follows by applying the above inequality recursively and by using (13). □

We are now ready to prove Theorem 2.2.

Proof of Theorem 2.2: First, note that the desired inclusion holds due to (10). Now, using (7) and (11), we obtain, respectively,

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

Combining the above inequalities, we find

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

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

$$\sum_{i=1}^{k} \|z_{i-1} - z_i\|^2_{\mathcal{Z},M_i} \leq \frac{2(1+\sigma)C_P(\|z^* - z_0\|^2_{\mathcal{Z},M_0} + \eta_0) + 2(1-\sigma)\eta_0}{(1-\sigma)}, \tag{66}$$

for all $z^* \in T^{-1}(0)$. Hence, (15) follows from Proposition 1.1, (10), (14), (66) and the fact that $\sum_{i=1}^{k} t_i \geq k \min_{i=1,\ldots,k}\{t_i\}$. □

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

Proposition A.3. *Let $\{z_k\}$, $\{M_k\}$ and $\{\eta_k\}$ be generated by the* VM-HPE framework *and consider $\{\tilde{z}^a_k\}$ and $\{\varepsilon^a_k\}$ as in* (18)*. Then, for every $k \geq 1$,*

$$\varepsilon^a_k \leq \frac{1}{2k}\left(\eta_0 + \|\tilde{z}^a_k - z_0\|^2_{\mathcal{Z},M_0} + \sum_{i=1}^{k} c_{i-1}\|\tilde{z}^a_k - z_{i-1}\|^2_{\mathcal{Z},M_{i-1}}\right), \tag{67}$$

where M_0 and $\{c_k\}$ are given in Assumption 3.1*.*

Proof. Using Lemma A.1 with $(z^*, z, z_+, \tilde{z}) = (\tilde{z}^a_k, z_{i-1}, z_i, \tilde{z}_i)$ and $M = M_i$, (10) and (11), we find, for every $i = 1, \ldots, k$,

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

where the second inequality is due to the fact that $1-\sigma \geq 0$. Hence, using Assumption 2.1 and simple calculations, we obtain

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

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

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

which clearly gives (67). □

Proof of Theorem 2.3: Note first that the desired inclusion and the first inequality in (20) follow from (10), (18) and Theorem 1.3(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 (67). Note that, using the convexity of $\|\cdot\|^2_{M_{i-1}}$, inequality (7) and (18), we find

$$\|\tilde{z}_k^a - z_{i-1}\|^2_{\mathcal{Z},M_{i-1}} \leq \frac{1}{k}\sum_{j=1}^{k}\|\tilde{z}_j - z_{i-1}\|^2_{\mathcal{Z},M_{i-1}} \leq \frac{2}{k}\sum_{j=1}^{k}\Big(\|\tilde{z}_j - z_j\|^2_{\mathcal{Z},M_{i-1}} + \|z_j - z_{i-1}\|^2_{\mathcal{Z},M_{i-1}}\Big). \quad (68)$$

From (13), we have $M_{i-1} \preceq C_P M_j$ for all $j=1,\dots,k$. Hence, using Proposition 1.2, inequality (11), Lemma A.2(b) and (14), we find

$$\begin{aligned}
\sum_{j=1}^{k}\|\tilde{z}_j - z_j\|^2_{\mathcal{Z},M_{i-1}} &\leq C_P\sum_{j=1}^{k}\|\tilde{z}_j - z_j\|^2_{\mathcal{Z},M_j}\\
&\leq C_P\sum_{j=1}^{k}\Big(\sigma\|\tilde{z}_j - z_{j-1}\|^2_{\mathcal{Z},M_j} + \eta_{j-1} - \eta_j\Big)\\
&\leq \frac{\sigma}{1-\sigma}C_p^2(d_0^2+\eta_0) + C_P\eta_0. \qquad (69)
\end{aligned}$$

On the other hand, using (7), $M_{i-1} \preceq C_P M_j$ for all $j=1,\dots,k$, Proposition 1.2, Lemma A.2(b) and (14), we obtain

$$\begin{aligned}
\sum_{j=1}^{k}\|z_j - z_{i-1}\|^2_{\mathcal{Z},M_{i-1}} &\leq 2\sum_{j=1}^{k}\Big(\|z_j - z^*\|^2_{\mathcal{Z},M_{i-1}} + \|z^* - z_{i-1}\|^2_{\mathcal{Z},M_{i-1}}\Big)\\
&\leq 2\sum_{j=1}^{k}\Big(C_P\|z_j - z^*\|^2_{\mathcal{Z},M_j} + \|z^* - z_{i-1}\|^2_{\mathcal{Z},M_{i-1}}\Big)\\
&\leq 2(1+C_P)C_P(d_0^2+\eta_0)k. \qquad (70)
\end{aligned}$$

It follows from inequalities (68)–(70) and the fact that $k \geq 1$ that

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

which, combined with Proposition A.3 and the first condition in (12), yields

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

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

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

$$k\, r_k^a = \sum_{i=1}^{k} r_i = M_1(z_0 - z^*) - M_k(z_k - z^*) + \sum_{i=1}^{k-1}(M_{i+1} - M_i)(z_i - z^*). \tag{71}$$

From (13), we obtain $M_1 \preceq C_P M_k$ and $M_1 \preceq C_P M_0$. Hence, it follows from Propositions 1.1 and 1.2 that

$$\begin{aligned}\|M_1(z_0 - z^*)\|^*_{\mathcal{Z},M_k} &\leq \sqrt{C_P}\|M_1(z_0 - z^*)\|^*_{\mathcal{Z},M_1}\\ &= \sqrt{C_P}\|z_0 - z^*\|_{\mathcal{Z},M_1}\\ &\leq C_p\|z_0 - z^*\|_{\mathcal{Z},M_0}.\end{aligned} \tag{72}$$

Direct use of Proposition 1.1 yields

$$\|M_k(z_k - z^*)\|^*_{\mathcal{Z},M_k} = \|z_k - z^*\|_{\mathcal{Z},M_k}. \tag{73}$$

Next step is to estimate the general term in the summation in (71). To do this, first note that using Assumption 2.1, we find

$$0 \preceq L_i := M_{i+1} - M_i + c_i M_{i+1} \preceq c_i(2+c_i)M_i\,, \qquad \forall\, i = 1,\ldots,k-1, \tag{74}$$

and so

$$\begin{aligned}\|(M_{i+1} - M_i)(z_i - z^*)\|^*_{\mathcal{Z},M_k} &= \|(L_i - c_i M_{i+1})(z_i - z^*)\|^*_{\mathcal{Z},M_k}\\ &\leq \|L_i(z_i - z^*)\|^*_{\mathcal{Z},M_k} + c_i\|M_{i+1}(z_i - z^*)\|^*_{\mathcal{Z},M_k}.\end{aligned} \tag{75}$$

From (13) and the last inequality in (74), we obtain, respectively, $M_i \preceq C_p M_k$ and $L_i \preceq c_i(2+c_i)M_i$. Hence, using Propositions 1.1 and 1.2, we have

$$\begin{aligned}\|L_i(z_i - z^*)\|^*_{\mathcal{Z},M_k} &\leq \sqrt{C_P}\|L_i(z_i - z^*)\|^*_{\mathcal{Z},M_i}\\ &\leq \sqrt{C_P}\sqrt{c_i(2+c_i)}\|L_i(z_i - z^*)\|^*_{\mathcal{Z},L_i}\\ &= \sqrt{C_P}\sqrt{c_i(2+c_i)}\|z_i - z^*\|_{\mathcal{Z},L_i}\\ &\leq \sqrt{C_P}\, c_i(2+c_i)\|z_i - z^*\|_{\mathcal{Z},M_i}.\end{aligned} \tag{76}$$

Again, from (13), we obtain $M_{i+1} \preceq C_P M_k$ and $M_{i+1} \preceq (1+c_i)M_i$, and consequently

$$\begin{aligned}\|M_{i+1}(z_i - z^*)\|^*_{\mathcal{Z},M_k} &\leq \sqrt{C_p}\|M_{i+1}(z_i - z^*)\|^*_{\mathcal{Z},M_{i+1}}\\ &= \sqrt{C_p}\|z_i - z^*\|_{\mathcal{Z},M_{i+1}}\\ &\leq \sqrt{C_P(1+c_i)}\|z_i - z^*\|_{\mathcal{Z},M_i}.\end{aligned} \tag{77}$$

Hence, using (13) and (75)–(77), we find

$$\begin{aligned}\|(M_{i+1}-M_i)(z_i-z^*)\|^*_{\mathcal{Z},M_k} &\le c_i\sqrt{C_p}\left(1+(1+c_i)+\sqrt{1+c_i}\,\right)\|z_i-z^*\|_{\mathcal{Z},M_i}\\ &\le c_i\sqrt{C_P}\left(1+C_P+\sqrt{C_P}\right)\|z_i-z^*\|_{\mathcal{Z},M_i}. \end{aligned}\tag{78}$$

Finally, using the definition of d_0 in (14), (71)–(73), (78) and Lemma A.2(b), we conclude that

$$\begin{aligned} k\|r_k^a\|^*_{\mathcal{Z},M_k} &\le \|M_1(z_0-z^*)\|^*_{\mathcal{Z},M_k}+\|M_k(z_k-z^*)\|^*_{\mathcal{Z},M_k}+\sum_{i=1}^{k-1}\|(M_{i+1}-M_i)(z_i-z^*)\|^*_{\mathcal{Z},M_k}\\ &\le \left(C_P+1+C_S\sqrt{C_P}\left(1+C_P+\sqrt{C_P}\right)\right)\max_{i=0,\dots,k}\|z_i-z^*\|_{\mathcal{Z},M_i}\\ &\le \sqrt{C_P}\left(C_P+1+C_S\sqrt{C_P}\left(1+C_P+\sqrt{C_P}\right)\right)\sqrt{d_0^2+\eta_0}\\ &= \left((1+C_P)\left(\sqrt{C_P}+C_SC_P\right)+C_SC_P^{3/2}\right)\sqrt{d_0^2+\eta_0}, \end{aligned}$$

which gives (19) for the case $k\ge 2$. Note now that by (13), we have $M_1\preceq C_PM_0$ and so using Propositions 1.1 and 1.2, Lemma A.2(b), (14) and the second identity in (18) with $k=1$, we find

$$\begin{aligned} \|r_1^a\|^*_{\mathcal{Z},M_1}=\|r_1\|^*_{\mathcal{Z},M_1} &= \|M_1(z_0-z_1)\|^*_{\mathcal{Z},M_1}\\ &= \|z_0-z_1\|_{\mathcal{Z},M_1}\\ &\le \|z_0-z^*\|_{\mathcal{Z},M_1}+\|z_1-z^*\|_{\mathcal{Z},M_1}\\ &\le \sqrt{C_P}\|z_0-z^*\|_{\mathcal{Z},M_0}+\|z_1-z^*\|_{\mathcal{Z},M_1}\\ &\le (1+\sqrt{C_P})\sqrt{C_P}\sqrt{d_0^2+\eta_0}\,, \end{aligned}$$

which in turn, combined with the fact that $C_P\ge 1$, gives (19) for $k=1$. □

References

[1] H. Attouch and M. Soueycatt. 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.

[2] S. Banert, R. I. Bot, and E. R. Csetnek. Fixing and extending some recent results on the ADMM algorithm. *Avaliable on http://www.arxiv.org.*

[3] B.He, H. Liu, Z. Wang, and X. Yuan. A strictly contractive peaceman–rachford splitting method for convex programming. *SIAM J. Optim.*, 24(3):1011–1040, 2014.

[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(1):1–122, 2011.

[5] R. S. Burachik, A. N. Iusem, and B. F. Svaiter. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Anal.*, 5(2):159–180, 1997.

[6] R. S. Burachik, C. A. Sagastizábal, and B. F. Svaiter. ϵ-enlargements of maximal monotone operators: theory and applications. In *Reformulation: nonsmooth, piecewise smooth, semismooth and smoothing methods (Lausanne, 1997)*, volume 22 of *Appl. Optim.*, pages 25–43. Kluwer Acad. Publ., Dordrecht, 1999.

[7] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.*, 40(1):120–145, 2011.

[8] Y. Cui, X. Li, D. Sun, and K. C. Toh. 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.

[9] D. Davis and W. Yin. Convergence rate analysis of several splitting schemes. *Avaliable on http://www.arxiv.org.*

[10] W. Deng and W. Yin. On the global and linear convergence of the generalized alternating direction method of multipliers. *J. Sci. Comput.*, pages 1–28, 2015.

[11] J. Eckstein. Some saddle-function splitting methods for convex programming. *Optim. Method Softw.*, 4(1):75–83, 1994.

[12] E. X. Fang, B. He, H. Liu, and X. Yuan. Generalized alternating direction method of multipliers: new theoretical insights and applications. *Math. Prog. Comp.*, 7(2):149–187, 2015.

[13] M. Fazel, T. K. Pong, D. Sun, and P. Tseng. Hankel matrix rank minimization with applications to system identification and realization. *SIAM J. Matrix Anal. Appl.*, 34(3):946–977, 2013.

[14] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.*, 2:17–40, 1976.

[15] R. Glowinski and A. Marroco. 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.

[16] M. L. N. Gonçalves, J. G. Melo, and R. D. C. Monteiro. Extending the ergodic convergence rate of the proximal ADMM. *Avaliable on http://www.arxiv.org.*

[17] M. L. N. Gonçalves, J. G. Melo, and R. D. C. Monteiro. Improved pointwise iteration-complexity of a regularized ADMM and of a regularized non-euclidean HPE framework. *To appear in SIAM J. Optim.*

[18] W. W. Hager, M. Yashtini, and H. Zhang. An $O(1/k)$ convergence rate for the variable stepsize Bregman operator splitting algorithm. *SIAM J. Numer. Anal.*, 54(3):1535–1556, 2016.

[19] B. He, L.-Z. Liao, D. Han, and H. Yang. A new inexact alternating directions method for monotone variational inequalities. *Math. Program.*, 92(1, Ser. A):103–118, 2002.

[20] B. He and H. Yang. Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities. *Oper. Res. Lett.*, 23(35):151 – 161, 1998.

[21] B. He and X. Yuan. On the $\mathcal{O}(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. *SIAM Journal on Numer. Anal.*, 50(2):700–709, 2012.

[22] B. He and X. Yuan. On non-ergodic convergence rate of Douglas-Rachford alternating direction method of multipliers. *Numer. Math.*, 130(3):567–577, 2015.

[23] B. S. He, H. Yang, and L. Wang. Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities. *J. Optim. Theory. Appl.*, 106(2):337–356, 2000.

[24] Y. He and R. D. C. Monteiro. An accelerated HPE-type algorithm for a class of composite convex-concave saddle-point problems. *SIAM J. Optim.*, 26(1):29–56, 2016.

[25] T. Lin, S. Ma, and S. Zhang. An extragradient-based alternating direction method for convex minimization. *Found. Comput. Math.*, pages 1–25, 2015.

[26] P. A. Lotito, L. A. Parente, and M. V. Solodov. A class of variable metric decomposition methods for monotone variational inclusions. *J. Convex Anal.*, 16(3&4):857–880, 2009.

[27] M. Marques Alves, R. D. C. Monteiro, and B. F. Svaiter. Regularized HPE-type methods for solving monotone inclusions with improved pointwise iteration-complexity bounds. *SIAM J. Optim.*, 26(4):2730–2743, 2016.

[28] R. D. C. Monteiro and B. F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM J. Optim.*, 20(6):2755–2787, 2010.

[29] R. D. C. Monteiro and B. F. Svaiter. 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.

[30] R. D. C. Monteiro and B. F Svaiter. Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. *SIAM J. Optim.*, 23(1):475–507, 2013.

[31] Y. Ouyang, Y. Chen, G. Lan, and E. Pasiliao, Jr. An accelerated linearized alternating direction method of multipliers. *SIAM J. Imaging Sci.*, 8(1):644–681, 2015.

[32] L. A. Parente, P. A. Lotito, and M. V. Solodov. A class of inexact variable metric proximal point algorithms. *SIAM J. Optim.*, 19(1):240–260, 2008.

[33] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.

[34] R. T. Rockafellar. On the maximal monotonicity of subdifferential mappings. *Pacific J. Math.*, 33:209–216, 1970.

[35] M. V. Solodov. 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.

[36] M. V. Solodov and B. F. Svaiter. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal.*, 7(4):323–345, 1999.

[37] X. Zhang, M. Burger, X. Bresson, and S. Osher. Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *SIAM J. Imaging Sci.*, 3(3):253–276, 2010.