Extending the ergodic convergence rate of the proximal ADMM

Max L.N. Gonçalves * Jefferson G. Melo * Renato D.C. Monteiro †

November 8, 2016

Abstract

Pointwise and ergodic iteration-complexity results for the proximal alternating direction method of multipliers (ADMM) for any stepsize in $(0, (1+\sqrt{5})/2)$ have been recently established in the literature. In addition to giving alternative proofs of these results, this paper also extends the ergodic iteration-complexity result to include the case in which the stepsize is equal to $(1+\sqrt{5})/2$. As far as we know, this is the first ergodic iteration-complexity for the stepsize $(1+\sqrt{5})/2$ obtained in the ADMM literature. These results are obtained by showing that the proximal ADMM is an instance of a non-Euclidean hybrid proximal extragradient framework whose pointwise and ergodic convergence rate are also studied.

2000 Mathematics Subject Classification: 47H05, 47J22, 49M27, 90C25, 90C30, 90C60, 65K10.

Key words: alternating direction method of multipliers, hybrid proximal extragradient method, non-Euclidean Bregman distances, convex program, pointwise iteration-complexity, ergodic iteration-complexity, first-order methods, inexact proximal point method, regular distance generating function.

1 Introduction

This paper considers the following linearly constrained convex problem

$$\inf\{f(y) + g(s) : Cy + Ds = c\} \tag{1}$$

where $\mathcal{S}$, $\mathcal{Y}$ and $\mathcal{X}$ are finite dimensional inner product spaces, $f : \mathcal{Y} \to (-\infty, \infty]$ and $g : \mathcal{S} \to (-\infty, \infty]$ are proper closed convex functions, $C : \mathcal{Y} \to \mathcal{X}$ and $D : \mathcal{S} \to \mathcal{X}$ are linear operators, and $c \in \mathcal{X}$. Convex optimization problems with a separable structure such as (1) appear in many applications areas such as machine learning, compressive sensing and image processing. A well-known method that takes advantage of the special structure of (1) is the alternating direction method of multipliers (ADMM).

Many variants of the ADMM have been considered in the literature; see, for example, [3, 5, 6, 7, 12, 13, 14, 16, 21, 26]. Here, we study the proximal ADMM [6, 8] which, recursively, computes a

*Institute of Mathematics and Statistics, Federal University of Goias, 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 Grant 444134/2014-0, 309370/2014-0 and FAPEG/GO.

†School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205. (email: monteiro@isye.gatech.edu). The work of this author was partially supported by NSF Grant CMMI-1300221.

sequence $\{(s_k, y_k, x_k)\}$ as follows. Given $(s_{k-1}, y_{k-1}, x_{k-1})$, the k-th triple (s_k, y_k, x_k) is determined as

$$\begin{aligned} s_k &= \operatorname{argmin}_s \left\{ g(s) - \langle x_{k-1}, Ds\rangle + \frac{\beta}{2}\|Cy_{k-1} + Ds - c\|^2 + \frac{1}{2}\langle s - s_{k-1}, H(s - s_{k-1})\rangle \right\}, \\ y_k &= \operatorname{argmin}_y \left\{ f(y) - \langle x_{k-1}, Cy\rangle + \frac{\beta}{2}\|Cy + Ds_k - c\|^2 + \frac{1}{2}\langle y - y_{k-1}, G(y - y_{k-1})\rangle \right\}, \\ x_k &= x_{k-1} - \theta\beta\left[Cy_k + Ds_k - c\right] \end{aligned} \tag{2}$$

where $\beta > 0$ is a penalty parameter, $\theta > 0$ is a stepsize parameter, and $H : \mathcal{S} \to \mathcal{S}$ and $G : \mathcal{Y} \to \mathcal{Y}$ are positive semidefinite self-adjoint linear operators. We refer to the subclass obtained from (2) by setting $(H, G) = (0, 0)$ to as the standard ADMM. Also, the proximal ADMM with $(H, G) = (\tau I - \beta D^* D, 0)$ for some $\tau \geq \beta\|D\|^2$ is known as the linearized ADMM or the split inexact Uzawa method (see, e.g., [16, 33, 34]). It has the desirable feature that, for many applications, its subproblems are much easier to solve or even have closed-form solutions (see [5, 16, 31, 32] for more details).

Pointwise and ergodic iteration-complexity results for the proximal ADMM (2) for any $\theta \in (0, (1 + \sqrt{5})/2)$ are established in [4, 13]. Our paper develops alternative pointwise and ergodic iteration-complexity results for the proximal ADMM (2) based on a different but related termination criterion. More specifically, a pointwise iteration-complexity is established for any $\theta \in (0, (1+\sqrt{5})/2)$ and an ergodic one is obtained for any $\theta \in (0, (1 + \sqrt{5})/2]$. Hence, our analysis of the ergodic case includes the case $\theta = (1 + \sqrt{5})/2$ which, as far as we know, has not been established yet. Our approach towards obtaining this extension is based on viewing the proximal ADMM as an instance of a non-Euclidean hybrid proximal extragradient (HPE) framework whose (pointwise and ergodic) complexity is studied and is then used to derive that of the proximal ADMM.

Previous related works. The ADMM was introduced in [9, 11] and is thoroughly discussed in [1, 10]. To discuss complexity results about ADMM, we use the terminology weak pointwise or strong pointwise bounds to refer to complexity bounds relative to the best of the k first iterates or the last iterate, respectively, to satisfy a suitable termination criterion. The first iteration-complexity bound for the ADMM was established only recently in [25] under the assumptions that C is injective. More specifically, the ergodic iteration-complexity for the standard ADMM is derived in [25] for any $\theta \in (0, 1]$ while a weak pointwise iteration-complexity easily follows from the approach in [25] for any $\theta \in (0, 1)$. Subsequently, without assuming that C is injective, [16] established the ergodic iteration-complexity of the proximal ADMM (2) with $G = 0$ and $\theta = 1$ and, as a consequence, of the split inexact Uzawa method [33]. Paper [15] establishes the weak pointwise and ergodic iteration-complexity of another collection of ADMM instances which includes the standard ADMM for any $\theta \in (0, (1+\sqrt{5})/2)$. A strong pointwise iteration-complexity bound for the proximal ADMM (2) with $G = 0$ and $\theta = 1$ is derived in [17]. Finally, a number of papers (see for example [4, 5, 7, 12, 13, 14, 21, 26] and references therein) have extended most of these complexity results to the context of the ADMM class (2) as well as other ADMM classes.

The non-Euclidean HPE framework is a class of inexact proximal point methods for solving the monotone inclusion problem which uses a relative (instead of summable) error criterion. The proximal point method, proposed by Rockafellar [28], is a classical iterative scheme for solving the latter problem. Paper [29] introduces an Euclidean version of the HPE framework. Iteration-complexities of the latter framework are established in [23] (see also [24]). Generalizations of the HPE framework to the non-Euclidean setting are studied in [12, 20, 30]. Applications of the HPE

framework can be found for example in [19, 18, 22, 23, 24, 25].

Organization of the paper. Subsection 1.1 presents our notation and basic results. Section 2 describes the proximal ADMM and present its pointwise and ergodic convergence rate results whose proofs are given in Section 4. Section 3 is devoted to the study of a non-Euclidean HPE framework. This section is divided into two subsections, Subsection 3.1 introduces the framework and presents its convergence rate bounds whose proofs are given in Subsection 3.2.

1.1 Notation and basic results

This subsection presents some definitions, notation and basic results used in this paper.

Let $\mathcal{V}$ be a finite-dimensional real vector space with inner product and associated norm denoted by $\langle\cdot,\cdot\rangle_{\mathcal{V}}$ and $\|\cdot\|_{\mathcal{V}}$, respectively. For a given self-adjoint positive semidefinite linear operator $A:\mathcal{V}\to\mathcal{V}$, the seminorm induced by A on $\mathcal{V}$ is defined by $\|\cdot\|_{\mathcal{V},A}=\langle A(\cdot),\cdot\rangle_{\mathcal{V}}^{1/2}$. For an arbitrary seminorm $\|\cdot\|$ on $\mathcal{V}$, its dual (extended) seminorm, denoted by $\|\cdot\|^*$, is defined as $\|\cdot\|^*:=\sup\{\langle\cdot,v\rangle_{\mathcal{V}}:\|v\|\leq 1\}$.

The following result gives some properties of $\|\cdot\|^*_{\mathcal{V},A}$ whose proof is omitted.

Proposition 1.1. *Let $A:\mathcal{V}\to\mathcal{V}$ be a self-adjoint positive semidefinite linear operator. Then,* $\operatorname{dom}\|\cdot\|^*_{\mathcal{V},A}=Im\,(A)$ *and* $\|A(\cdot)\|^*_{\mathcal{V},A}=\|\cdot\|_{\mathcal{V},A}$.

Given a set-valued operator $T:\mathcal{V}\rightrightarrows\mathcal{V}$, its domain and graph are defined as $\operatorname{Dom}T:=\{v\in\mathcal{V}:T(v)\neq\emptyset\}$ and $Gr(T)=\{(v_1,v_2)\in\mathcal{V}\times\mathcal{V}\mid v_2\in T(v_1)\}$, respectively, and its inverse operator $T^{-1}:\mathcal{V}\rightrightarrows\mathcal{V}$ is given by $T^{-1}(v_2):=\{v_1\ :\ v_2\in T(v_1)\}$. The operator T is said to be monotone if

$$\langle u_1-v_1,u_2-v_2\rangle\geq 0\quad\forall\,(u_1,u_2),(v_1,v_2)\,\in\,Gr(T).$$

Moreover, T is maximal monotone if it is monotone and there is no other monotone operator S such that $Gr(T)\subset Gr(S)$. Given a scalar $\varepsilon\geq 0$, the ε-enlargement $T^{[\varepsilon]}:\mathcal{V}\rightrightarrows\mathcal{V}$ of a monotone operator $T:\mathcal{V}\rightrightarrows\mathcal{V}$ is defined as

$$T^{[\varepsilon]}(v):=\{v'\in\mathcal{V}:\langle v'-v_2,v-v_1\rangle\geq-\varepsilon,\ \ \forall(v_1,v_2)\in Gr(T)\}\quad\forall v\in\mathcal{V}. \tag{3}$$

Recall that the ε-subdifferential of a convex function $f:\mathcal{V}\to[-\infty,\infty]$ is defined by $\partial_\varepsilon f(v):=\{u\in\mathcal{V}:f(v')\geq f(v)+\langle u,v'-v\rangle-\varepsilon\ \ \forall v'\in\mathcal{V}\}$ for every $v\in\mathcal{V}$. When $\varepsilon=0$, then $\partial_0 f(x)$ is denoted by $\partial f(x)$ and is called the subdifferential of f at x. The operator ∂f is trivially monotone if f is proper. If f is a proper lower semi-continuous convex function, then ∂f is maximal monotone [27]. The domain of f is denoted by $\operatorname{dom}f$ and the conjugate of f is the function $f^*:\mathcal{V}\to[-\infty,\infty]$ defined as

$$f^*(v)=\sup_{z\in\mathcal{V}}\left(\langle v,z\rangle-f(z)\right)\quad\forall v\in\mathcal{V}.$$

2 Proximal ADMM and its convergence rate

In this section, we recall the proximal ADMM for solving (1) and present pointwise and ergodic convergence rate results. The pointwise convergence rate considers the stepsize parameter in the open interval $(0,(\sqrt{5}+1)/2)$ while the ergodic one includes also the stepsize $(\sqrt{5}+1)/2$.

Throughout this section, we assume that:

A1) the problem (1) has an optimal solution (s^*, y^*) and an associated Lagrange multiplier x^*, or equivalently, the inclusion

$$0 \in T(s, y, x) := \begin{bmatrix} \partial g(s) - D^* x \\ \partial f(y) - C^* x \\ Cy + Ds - c \end{bmatrix} \tag{4}$$

has a solution (s^*, y^*, x^*);

A2) there exists $x \in \mathcal{X}$ such that $(C^*x, D^*x) \in \mathrm{ri}\,(\mathrm{dom}\, f^*) \times \mathrm{ri}\,(\mathrm{dom}\, g^*)$.

Next we state the proximal ADMM for solving the problem (1).

Proximal ADMM

---

(0) Let an initial point $(s_0, y_0, x_0) \in \mathcal{S} \times \mathcal{Y} \times \mathcal{X}$, a penalty parameter $\beta > 0$, a setpsize $\theta > 0$, and self-adjoint positive semidefinite linear operators $H : \mathcal{S} \to \mathcal{S}$ and $G : \mathcal{Y} \to \mathcal{Y}$ be given, and set $k = 1$;

(1) compute an optimal solution $s_k \in \mathcal{S}$ of the subproblem

$$\min_{s \in \mathcal{S}} \left\{ g(s) - \langle D^* x_{k-1}, s \rangle_{\mathcal{S}} + \frac{\beta}{2} \|Cy_{k-1} + Ds - c\|_{\mathcal{X}}^2 + \frac{1}{2} \|s - s_{k-1}\|_{\mathcal{S},H}^2 \right\} \tag{5}$$

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

$$\min_{y \in \mathcal{Y}} \left\{ f(y) - \langle C^* x_{k-1}, y \rangle_{\mathcal{Y}} + \frac{\beta}{2} \|Cy + Ds_k - c\|_{\mathcal{X}}^2 + \frac{1}{2} \|y - y_{k-1}\|_{\mathcal{Y},G}^2 \right\}; \tag{6}$$

(2) set

$$x_k = x_{k-1} - \theta\beta \left[Cy_k + Ds_k - c \right] \tag{7}$$

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

end

The proximal ADMM has different features depending on the choice of the operators H and G. For instance, by taking $(H, G) = (0, 0)$ and $(H, G) = (\tau I - \beta D^* D, 0)$ with $\tau > 0$, it reduces to the standard ADMM and the linearized ADMM, respectively. The latter method is related to the split inexact Uzawa method (see, e.g., [16, 34]) and it basically consists of linearizing the quadratic term $(1/2)\|Cy_{k-1} + Ds - c\|_{\mathcal{X}}^2$ in the standard ADMM and adding a proximal term $(1/2)\|s - s_{k-1}\|_{\mathcal{S},H}^2$. In many applications, the corresponding subproblem (5) for the linearized ADMM is much easier to solve or even has a closed-form solution (see [16, 31, 32] for more details). We also mention that depending on the structure of problem (1), other choices of H and G may be recommended; see, for instance, Section 1.1 of [5]. It is worth pointing out that the condition **A2** is used only to ensure that the subproblems of ADMM as well as some variants have solutions, see for example [25, Proposition 7.2] and [12, comments on page 16]. In particular, under this assumption it is possible to show that the subproblems (5) and (6) have solutions.

The next two results present pointwise and ergodic convergence rate bounds for the proximal ADMM under the assumption that $\theta \in (0, (\sqrt{5}+1)/2)$ and $\theta \in (0, (\sqrt{5}+1)/2]$, respectively. Their statements use the quantities d_0, τ_θ and σ_θ defined as

$$d_0 := d_0(\beta, \theta) = \inf_{(s,y,x)\in T^{-1}(0)} \left\{ \frac{1}{2}\|s_0 - s\|^2_{\mathcal{S},H} + \frac{1}{2}\|y_0 - y\|^2_{\mathcal{Y},(G+\beta C^*C)} + \frac{1}{2\beta\theta}\|x_0 - x\|^2_{\mathcal{X}} \right\}, \quad (8)$$

$$\sigma_\theta := \frac{3\theta^2 - 7\theta + 5 + \sqrt{(3\theta^2 - 7\theta + 5)^2 - 4(2-\theta)(3-\theta)(\theta-1)^2}}{2(3-\theta)}, \quad (9)$$

$$\tau_\theta := 4\max\left\{ \frac{1}{\sqrt{\theta}}, \frac{\sqrt{\theta}}{2-\theta} \right\}. \quad (10)$$

It is easy to verify that $\sigma_\theta \in (0,1)$ whenever $\theta \in (0, (\sqrt{5}+1)/2)$ and $\sigma_\theta = 1$ when $\theta = (\sqrt{5}+1)/2$.

Theorem 2.1. *(**Pointwise convergence of the proximal ADMM**) Consider the sequence $\{(s_k, y_k, x_k)\}$ generated by the proximal ADMM with $\theta \in (0, (\sqrt{5}+1)/2)$, and let $\{\tilde{x}_k\}$ be defined as*

$$\tilde{x}_k = x_{k-1} - \beta(Cy_{k-1} + Ds_k - c). \quad (11)$$

Then, for every $k \in \mathbb{N}$,

$$\begin{pmatrix} H(s_{k-1} - s_k) \\ (G + \beta C^*C)(y_{k-1} - y_k) \\ \frac{1}{\beta\theta}(x_{k-1} - x_k) \end{pmatrix} \in \begin{bmatrix} \partial g(s_k) - D^*\tilde{x}_k \\ \partial f(y_k) - C^*\tilde{x}_k \\ Cy_k + Ds_k - c \end{bmatrix} \quad (12)$$

and there exists $i \le k$ such that

$$\left(\|s_{i-1} - s_i\|^2_{\mathcal{S},H} + \|y_{i-1} - y_i\|^2_{\mathcal{Y},(G+\beta C^*C)} + \frac{1}{\beta\theta}\|x_{i-1} - x_i\|^2_{\mathcal{X}} \right)^{1/2} \le \frac{2\sqrt{d_0}}{\sqrt{k}} \sqrt{\frac{1 + \sigma_\theta + 2\tau_\theta}{1 - \sigma_\theta}}$$

where d_0, σ_θ and τ_θ are as in (8), (9) and (10), respectively.

In contrast to the pointwise convergence rate result stated above, the ergodic convergence rate result stated below holds for the extreme case in which $\theta = (\sqrt{5}+1)/2$.

Theorem 2.2. (Ergodic convergence of the proximal ADMM) *Consider the sequence $\{(s_k, y_k, x_k)\}$ generated by the proximal ADMM with $\theta \in (0, (\sqrt{5}+1)/2]$, and let $\{\tilde{x}_k\}$ be as in (11). Moreover, consider the ergodic sequences $\{(s^a_k, y^a_k, x^a_k, \tilde{x}^a_k)\}$ and $\{\varepsilon^a_k\}$ defined by*

$$(s^a_k, y^a_k, x^a_k, \tilde{x}^a_k) := \frac{1}{k}\sum_{i=1}^{k} (s_i, y_i, x_i, \tilde{x}_i), \quad (\varepsilon^a_{k,s}, \varepsilon^a_{k,y}) = \frac{1}{k}\sum_{i=1}^{k} (\langle r_{i,s}, s_i - s^a_k\rangle, \langle r_{i,y}, y_i - y^a_k\rangle). \quad (13)$$

where

$$(r_{i,s}, r_{i,y}) = (H(s_{i-1} - s_i), (G + \beta C^*C)(y_{i-1} - y_i)). \quad (14)$$

Then, for every $k \in \mathbb{N}$,

$$\begin{pmatrix} H(s^a_{k-1} - s^a_k) \\ (G + \beta C^*C)(y^a_{k-1} - y^a_k) \\ \frac{1}{\beta\theta}(x^a_{k-1} - x^a_k) \end{pmatrix} \in \begin{bmatrix} \partial g_{\varepsilon^a_{k,s}}(s^a_k) - D^*\tilde{x}^a_k \\ \partial f_{\varepsilon^a_{k,y}}(y^a_k) - C^*\tilde{x}^a_k \\ Cy^a_k + Ds^a_k - c \end{bmatrix} \quad (15)$$

$$\left(\|s^a_{k-1}-s^a_k\|^2_{\mathcal{S},H}+\|y^a_{k-1}-y^a_k\|^2_{\mathcal{Y},(G+\beta C^*C)}+\frac{1}{\beta\theta}\|x^a_{k-1}-x^a_k\|^2_{\mathcal{X}}\right)^{1/2}\leq\frac{2\sqrt{2(1+\tau_\theta)d_0}}{k}$$

and

$$\varepsilon^a_{k,s}+\varepsilon^a_{k,y}\leq\frac{3(1+\tau_\theta)[3\theta^2+4\sigma_\theta(\theta^2+\theta+1)]d_0}{\theta^2 k}$$

where d_0, σ_θ *and* τ_θ *are as in* (8), (9) *and* (10), *respectively.*

The proofs of Theorems 2.1 and 2.2 will be presented in Section 4. For this, we first study a non-Euclidean HPE framework from which the proximal ADMM is a special instance.

3 A non-Euclidean HPE framework

This section describes and derives convergence rate bounds for a non-Euclidean HPE framework for solving monotone inclusion problems. Subsection 3.1 describes the non-Euclidean HPE framework and its corresponding pointwise and ergodic convergence rate bounds. Subsection 3.2 gives the proofs for the two convergence rate results stated in Subsection 3.1.

3.1 A non-Euclidean HPE framework and its convergence rate

Let $\mathcal{Z}$ be finite-dimensional inner product real vector space. We start by introducing the definition of a distance generating function and its corresponding Bregman distance adopted in this paper.

Definition 3.1. *A proper lower semi-continuous convex function* $w:\mathcal{Z}\to(-\infty,\infty]$ *is called a distance generating function if* $\mathrm{int}(\mathrm{dom}\, w)=\mathrm{Dom}\,\partial w\neq\emptyset$ *and* w *is continuously differentiable on this interior. Moreover,* w *induces the Bregman distance* $dw:\mathcal{Z}\times\mathrm{int}(\mathrm{dom}\, w)\to\mathbb{R}$ *defined as*

$$(dw)(z';z):=w(z')-w(z)-\langle\nabla w(z),z'-z\rangle_{\mathcal{Z}}\quad\forall(z',z)\in\mathcal{Z}\times\mathrm{int}(\mathrm{dom}\, w).\tag{16}$$

For simplicity, for every $z\in\mathrm{int}(\mathrm{dom}\, w)$, the function $(dw)(\,\cdot\,;z)$ will be denoted by $(dw)_z$ so that

$$(dw)_z(z')=(dw)(z';z)\quad\forall(z',z)\in\mathcal{Z}\times\mathrm{int}(\mathrm{dom}\, w).$$

The following useful identities follow straightforwardly from (16):

$$\nabla(dw)_z(z')=-\nabla(dw)_{z'}(z)=\nabla w(z')-\nabla w(z)\quad\forall z,z'\in\mathrm{int}(\mathrm{dom}\, w),\tag{17}$$

$$(dw)_v(z')-(dw)_v(z)=\langle\nabla(dw)_v(z),z'-z\rangle_{\mathcal{Z}}+(dw)_z(z')\quad\forall z'\in\mathcal{Z},\ \forall v,z\in\mathrm{int}(\mathrm{dom}\, w).\tag{18}$$

Our analysis of the non-Euclidean HPE framework requires the distance generating function to be regular with respect to a seminorm according to the following definition.

Definition 3.2. *Let distance generating function* $w:\mathcal{Z}\to[-\infty,\infty]$, *seminorm* $\|\cdot\|$ *in* $\mathcal{Z}$ *and convex set* $Z\subset\mathrm{int}(\mathrm{dom}\, w)$ *be given. For given positive constants* m *and* M, w *is said to be* (m,M)*-regular with respect to* $(Z,\|\cdot\|)$ *if*

$$(dw)_z(z')\geq\frac{m}{2}\|z-z'\|^2\quad\forall z,z'\in Z,\tag{19}$$

$$\|\nabla w(z)-\nabla w(z')\|^*\leq M\|z-z'\|\quad\forall z,z'\in Z.\tag{20}$$

We now make some remarks about the class of regular distance generating functions as in Definition 3.2, which was first introduced in [12]. First, if the seminorm in Definition 3.2 is a norm, then (19) implies that w is strongly convex, in which case the corresponding dw is said to be nondegenerate. However, since $\|\cdot\|$ is not assumed to be a norm, a regular distance generating function w does not need to be strongly convex, or equivalently, dw can be degenerate. Second, some examples of (m, M)-regular distance generating functions can be found in [12, Example 2.3]. For the purpose of analyzing the proximal ADMM, we make use of the distance generating function given by $w(\cdot) = (1/2)\|\cdot\|^2_{\mathcal{Z},Q}$ where Q is a self-adjoint positive semidefinite linear operator. This w can be easily shown to be $(1, 1)$-regular with respect to $(\mathcal{Z}, \|\cdot\|_{\mathcal{Z},Q})$. Third, if $w : \mathcal{Z} \to [-\infty, \infty]$ is (m, M)-regular with respect to $(Z, \|\cdot\|)$, then

$$\frac{m}{2}\|z - z'\|^2 \le (dw)_z(z') \le \frac{M}{2}\|z - z'\|^2 \quad \forall z, z' \in Z. \tag{21}$$

Throughout this section, we assume that $w : \mathcal{Z} \to [-\infty, \infty]$ is an (m, M)-regular distance generating function with respect to $(Z, \|\cdot\|)$ where $Z \subset \mathrm{int}(\mathrm{dom}\, w)$ is a closed convex set and $\|\cdot\|$ is a seminorm in $\mathcal{Z}$. Our problem of interest in this section is the monotone inclusion problem (MIP)

$$0 \in T(z) \tag{22}$$

where $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ is a maximal monotone operator and the following conditions hold:

B1) $\mathrm{Dom}\, T \subset Z$;

B2) the solution set $T^{-1}(0)$ of (22) is nonempty.

We now state a non-Euclidean HPE (NE-HPE) framework for solving (22).

NE-HPE framework for solving (22).

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

(1) choose $\lambda_k > 0$ and find $(\tilde{z}_k, z_k, \varepsilon_k, \eta_k) \in Z \times Z \times \mathbb{R}_+ \times \mathbb{R}_+$ such that

$$r_k := \frac{1}{\lambda_k}\nabla(dw)_{z_k}(z_{k-1}) \in T^{[\varepsilon_k]}(\tilde{z}_k), \tag{23}$$

$$(dw)_{z_k}(\tilde{z}_k) + \lambda_k\varepsilon_k + \eta_k \le \sigma(dw)_{z_{k-1}}(\tilde{z}_k) + \eta_{k-1}; \tag{24}$$

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

end

We now make some remarks about the NE-HPE framework. First, [12] studies an NE-HPE framework based on a regular distance generating function w for solving a monotone inclusion problem consisting of the sum of T and a μ-monotone operator S with respect to w. The latter notion implies strong monotonicity of S when dw is nondegenerate (see [12, Assumption (A1)]). Second, the NE-HPE does not specify how to find λ_k and $(\tilde{z}_k, z_k, \varepsilon_k)$ satisfying (23) and (24). The particular scheme for computing λ_k and $(\tilde{z}_k, z_k, \varepsilon_k)$ will depend on the instance of the framework under consideration and the properties of the operator T. Third, if w is strongly convex on Z and $\sigma = 0$,

then (24) implies that $\varepsilon_k = 0$ and $z_k = \tilde{z}_k$ for every k, and hence that $r_k \in T(z_k)$ in view of (23). Therefore, the HPE error conditions (23)-(24) can be viewed as a relaxation of an iteration of the exact non-Euclidean proximal point method, namely,

$$0 \in \frac{1}{\lambda_k}\nabla(dw)_{z_{k-1}}(z_k) + T(z_k).$$

Fourth, if w is strongly convex on Z, then it can be shown that the above inclusion has a unique solution z_k, and hence that, for any given $\lambda_k > 0$, there exists a triple $(\tilde{z}_k, z_k, \varepsilon_k)$ of the form $(z_k, z_k, 0)$ satisfying (23)-(24) with $\sigma = 0$. Clearly, computing the triple in this (exact) manner is expensive, and hence computation of (inexact) quadruples satisfying the HPE (relative) error conditions with $\sigma > 0$ is more computationally appealing.

We end this subsection by presenting pointwise and ergodic convergence rate results for the NE-HPE framework whose proofs are given in the next subsection. Their statements use the quantity $(dw)_0$ defined as

$$(dw)_0 = \inf\{(dw)_{z_0}(z^*) : z^* \in T^{-1}(0)\}. \tag{25}$$

Theorem 3.3. *(**Pointwise convergence of the NE-HPE**) Consider the sequence $\{(r_k, \varepsilon_k, \lambda_k)\}$ generated by the NE-HPE framework with $\sigma < 1$. Then, for every $k \geq 1$, $r_k \in T^{[\varepsilon_k]}(\tilde{z}_k)$ and the following statements hold:*

(a) if $\underline{\lambda} := \inf_{j\geq 1}\lambda_j > 0$, then there exists $i \leq k$ such that

$$\|r_i\|^* \leq \frac{2M}{\sqrt{m}}\sqrt{\frac{(1+\sigma)(dw)_0 + 2\eta_0}{1-\sigma}\left(\frac{\lambda_i^{-1}}{\sum_{j=1}^k \lambda_j}\right)} \leq \frac{2M}{\underline{\lambda}\sqrt{mk}}\sqrt{\frac{(1+\sigma)(dw)_0 + 2\eta_0}{1-\sigma}}$$

$$\varepsilon_i \leq \frac{(1+\sigma)(dw)_0 + 2\eta_0}{(1-\sigma)\sum_{i=1}^k \lambda_i} \leq \frac{(1+\sigma)(dw)_0 + 2\eta_0}{(1-\sigma)\underline{\lambda}k};$$

(b) there exists an index $i \leq k$ such that

$$\|r_i\|^* \leq \frac{2M}{\sqrt{m}}\sqrt{\frac{(1+\sigma)(dw)_0 + 2\eta_0}{1-\sigma}\left(\frac{1}{\sum_{j=1}^k \lambda_j^2}\right)}, \qquad \varepsilon_i \leq \frac{[(1+\sigma)(dw)_0 + 2\eta_0]\lambda_i}{(1-\sigma)\sum_{j=1}^k \lambda_j^2},$$

where $(dw)_0$ is as defined in (25).

From now on, we focus on the ergodic convergence of the NE-HPE framework. For $k \geq 1$, define $\Lambda_k := \sum_{i=1}^k \lambda_i$ and the ergodic iterate $(\tilde{z}_k^a, r_k^a, \varepsilon_k^a)$ as

$$\tilde{z}_k^a = \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i \tilde{z}_i, \quad r_k^a := \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i r_i, \quad \varepsilon_k^a := \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i \left(\varepsilon_i + \langle r_i, \tilde{z}_i - \tilde{z}_k^a\rangle\right). \tag{26}$$

The following result provides convergence rate bounds for $\|r_k^a\|^*$ and ε_k^a. The pair (r_k^a, ε_k^a) plays the role of a residual for $\tilde{z}_k^a$.

Theorem 3.4. (Ergodic convergence of the NE-HPE) *For every $k \geq 1$, $r_k^a \in T^{[\varepsilon_k^a]}(\tilde{z}_k^a)$ and*

$$\|r_k^a\|^* \leq \frac{2\sqrt{2}M((dw)_0 + \eta_0)^{1/2}}{\sqrt{m}\Lambda_k}, \quad \varepsilon_k^a \leq \left(\frac{3M}{m}\right)\left[\frac{3((dw)_0 + \eta_0) + \sigma\rho_k}{\Lambda_k}\right].$$

where

$$\rho_k := \max_{i=1,\ldots,k}(dw)_{z_{i-1}}(\tilde{z}_i).$$

Moreover, the sequence $\{\rho_k\}$ is bounded under either one of the following situations:

(a) $\sigma < 1$, in which case

$$\rho_k \leq \frac{(dw)_0 + \eta_0}{1-\sigma}; \tag{27}$$

(b) $\operatorname{Dom} T$ is bounded, in which case

$$\rho_k \leq \frac{2M}{m}[(dw)_0 + \eta_0 + D]$$

where $D := \sup\{\min\{(dw)_y(y'), (dw)_{y'}(y)\} : y, y' \in \operatorname{Dom} T\}$ is the diameter of $\operatorname{Dom} T$ with respect to dw, and $(dw)_0$ is as defined in (25).

The bound on ε_k^a presented in Theorem 3.4 depends on the quantity ρ_k which is bounded under the assumption $\sigma < 1$ or boundedness of $\operatorname{Dom} T$. As we will show in Section 4, proximal ADMM is an instance of the NE-HPE in which the stepsize $\theta = (\sqrt{5}+1)/2$ corresponds to the parameter $\sigma = 1$. Even in this case, the sequence $\{\rho_k\}$ is bounded regardless the boundedness of $\operatorname{Dom} T$.

3.2 Convergence rate analysis of the NE-HPE framework

The main goal of this subsection is to present the proofs of Theorems 3.3 and 3.4. Toward this goal, we first establish some technical lemmas which provide useful properties of regular Bregman distances and of the NE-HPE framework.

Lemma 3.5. *Let $w : \mathcal{Z} \to [-\infty, \infty]$ be an (m, M)-regular distance generating function with respect to $(Z, \|\cdot\|)$ as in Definition 3.2. Then, the following statements hold:*

(a) for every $z, z' \in Z$, we have

$$(\|\nabla(dw)_{z'}(z)\|^*)^2 \leq \frac{2M^2}{m}\min\{(dw)_z(z'), (dw)_{z'}(z)\}; \tag{28}$$

(b) for every $l \geq 1$ and $u_0, u_1, \ldots, u_l \in Z$, we have

$$(dw)_{u_0}(u_l) \leq \frac{lM}{m}\sum_{i=1}^{l}\min\{(dw)_{u_{i-1}}(u_i), (dw)_{u_i}(u_{i-1})\}. \tag{29}$$

Proof. (a) It is easy to see that (28) immediately follows from (17), (19) and (20).

(b) It follows from the second inequality in (21) that

$$(dw)_{u_0}(u_l) \leq \frac{M}{2}\|u_l - u_0\|^2 \leq \frac{M}{2}\left(\sum_{i=1}^{l}\|u_i - u_{i-1}\|\right)^2 \leq \frac{lM}{2}\sum_{i=1}^{l}\|u_i - u_{i-1}\|^2$$

which, in view of (19), immediately implies (29). □

The next result presents some useful estimates related to the sequence generated by the NE-HPE framework.

Lemma 3.6. *For every $k \geq 1$, the following statements hold:*

(a) for every $z \in \operatorname{dom} w$, we have

$$(dw)_{z_{k-1}}(z) - (dw)_{z_k}(z) = (dw)_{z_{k-1}}(\tilde{z}_k) - (dw)_{z_k}(\tilde{z}_k) + \lambda_k \langle r_k, \tilde{z}_k - z \rangle_{\mathcal{Z}};$$

(b) for every $z \in \operatorname{dom} w$, we have

$$(dw)_{z_{k-1}}(z) - (dw)_{z_k}(z) + \eta_{k-1} \geq (1-\sigma)(dw)_{z_{k-1}}(\tilde{z}_k) + \lambda_k(\langle r_k, \tilde{z}_k - z \rangle_{\mathcal{Z}} + \varepsilon_k) + \eta_k;$$

(c) for every $z^* \in T^{-1}(0)$, we have

$$(dw)_{z_{k-1}}(z^*) - (dw)_{z_k}(z^*) + \eta_{k-1} \geq (1-\sigma)(dw)_{z_{k-1}}(\tilde{z}_k) + \eta_k;$$

(d) for every $z^* \in T^{-1}(0)$, we have

$$(dw)_{z_k}(z^*) + \eta_k \leq (dw)_{z_{k-1}}(z^*) + \eta_{k-1}.$$

Proof. (a) Using (18) twice and using the definition of r_k given by (23), we obtain

$$\begin{aligned}
(dw)_{z_{k-1}}(z) - (dw)_{z_k}(z) &= (dw)_{z_{k-1}}(z_k) + \langle \nabla (dw)_{z_{k-1}}(z_k), z - z_k \rangle_{\mathcal{Z}} \\
&= (dw)_{z_{k-1}}(z_k) + \langle \nabla (dw)_{z_{k-1}}(z_k), \tilde{z}_k - z_k \rangle_{\mathcal{Z}} + \langle \nabla (dw)_{z_{k-1}}(z_k), z - \tilde{z}_k \rangle_{\mathcal{Z}} \\
&= (dw)_{z_{k-1}}(\tilde{z}_k) - (dw)_{z_k}(\tilde{z}_k) + \langle \nabla (dw)_{z_{k-1}}(z_k), z - \tilde{z}_k \rangle_{\mathcal{Z}} \\
&= (dw)_{z_{k-1}}(\tilde{z}_k) - (dw)_{z_k}(\tilde{z}_k) + \lambda_k \langle r_k, \tilde{z}_k - z \rangle_{\mathcal{Z}}.
\end{aligned}$$

(b) This statement follows as an immediate consequence of (a) and (24).
(c) This statement follows from (b), the fact that $0 \in T(z^*)$ and $r_k \in T^{[\varepsilon_k]}(\tilde{z}_k)$, and (3).
(d) This statement follows as an immediate consequence of (c) and $\sigma \leq 1$.

□

The pointwise convergence rate bounds for the NE-HPE framework will follow directly from the next result which estimates the residual pair (r_i, ε_i).

Lemma 3.7. *Let $\{(r_k, \varepsilon_k, \lambda_k)\}$ and (η_0, σ) be given by the NE-HPE framework and assume that $\sigma < 1$. Then, for every $t \in \mathbb{R}$ and every $k \geq 1$, there exists an $i \leq k$ such that*

$$\|r_i\|^* \leq \frac{2M}{\sqrt{m}} \sqrt{\frac{(1+\sigma)(dw)_0 + 2\eta_0}{1-\sigma} \left(\frac{\lambda_i^{t-2}}{\sum_{j=1}^k \lambda_j^t} \right)}, \qquad \varepsilon_i \leq \frac{(1+\sigma)(dw)_0 + 2\eta_0}{1-\sigma} \left(\frac{\lambda_i^{t-1}}{\sum_{j=1}^k \lambda_j^t} \right)$$

where $(dw)_0$ is as defined in (25).

Proof. For every $i \geq 1$, define

$$\theta_i = \max\left\{ \frac{m\lambda_i^2 \left(\|r_i\|^*\right)^2}{4M^2}, \lambda_i \varepsilon_i \right\}.$$

It is easy to see that the conclusion of the lemma will follow if we show that, for every $i \geq 1$, we have

$$(1-\sigma)\sum_{i=1}^{k} \theta_i \leq (1+\sigma)(dw)_0 + 2\eta_0.$$

In order to show that the last inequality hold, we have, from (17) and (23), for every $i \geq 1$

$$\begin{aligned}
\lambda_i \|r_i\|^* &= \|\nabla(dw)_{z_{i-1}}(\tilde{z}_i) - \nabla(dw)_{z_i}(\tilde{z}_i)\|^* \leq \|\nabla(dw)_{z_{i-1}}(\tilde{z}_i)\|^* + \|\nabla(dw)_{z_i}(\tilde{z}_i)\|^* \\
&\leq \frac{\sqrt{2}M}{\sqrt{m}} \left[(dw)_{z_{i-1}}(\tilde{z}_i)^{1/2} + (dw)_{z_i}(\tilde{z}_i)^{1/2} \right] \\
&\leq \frac{\sqrt{2}M}{\sqrt{m}} \left[(dw)_{z_{i-1}}(\tilde{z}_i)^{1/2} + (\sigma(dw)_{z_{i-1}}(\tilde{z}_i) + \eta_{i-1} - \eta_i)^{1/2} \right],
\end{aligned}$$

where the second and third inequalities are due to (28) and (24), respectively. Hence,

$$\frac{m\lambda_i^2 \left(\|r_i\|^*\right)^2}{2M^2} \leq 2(1+\sigma)(dw)_{z_{i-1}}(\tilde{z}_i) + 2(\eta_{i-1} - \eta_i).$$

The previous estimative together with (24) and definition of θ_i imply that $\theta_i \leq (1+\sigma)(dw)_{z_{i-1}}(\tilde{z}_i) + (\eta_{i-1} - \eta_i)$ for every $i \geq 1$. Thus, if $z^* \in T^{-1}(0)$, it follows from Lemma 3.6(c) that

$$(1-\sigma)\sum_{i=1}^{k} \theta_i \leq (1+\sigma)[(dw)_{z_0}(z^*) - (dw)_{z_k}(z^*) + \eta_0 - \eta_k] + (1-\sigma)(\eta_0 - \eta_k) \leq (1+\sigma)(dw)_{z_0}(z^*) + 2\eta_0.$$

The desired inequality follows from the latter inequality and the definition of $(dw)_0$ in (25). As a consequence, we obtain the conclusion of the lemma. □

Now we are ready to prove Theorems 3.3 and 3.4 stated in Subsection 3.1.

Proof of Theorem 3.3: The inclusion $r_k \in T^{[\varepsilon_k]}(\tilde{z}_k)$ holds due to (23). Statements (a) and (b) follow directly from Lemma 3.7 with t = 1 and t = 2, respectively. □

Proof of Theorem 3.4: The inclusion $r_k^a \in T^{[\varepsilon_k^a]}(\tilde{z}_k^a)$ follows from the transportation formula (see [2, Theorem 2.3]). Now, let $z^* \in T^{-1}(0)$. Using (17), (23) and (26), we easily see that

$$\Lambda_k r_k^a = \nabla(dw)_{z_0}(z_k) = \nabla(dw)_{z_0}(z^*) - \nabla(dw)_{z_k}(z^*)$$

which, together with (28) and Lemma 3.6(d), imply that

$$\begin{aligned}
\Lambda_k \|r_k^a\|^* &\leq \|\nabla(dw)_{z_0}(z^*)\|^* + \|\nabla(dw)_{z_k}(z^*)\|^* \\
&\leq \frac{\sqrt{2}M}{\sqrt{m}} [(dw)_{z_0}(z^*)^{1/2} + (dw)_{z_k}(z^*)^{1/2}] \leq \frac{2\sqrt{2}M}{\sqrt{m}} ((dw)_{z_0}(z^*) + \eta_0)^{1/2}.
\end{aligned}$$

This inequality together with definition of $(dw)_0$ clearly imply the bound on $\|r_k^a\|^*$. To show the bound on ε_k^a, first note that Lemma 3.6(b) implies that for every $z \in W$,

$$(dw)_{z_0}(z) + \eta_0 \geq \sum_{i=1}^{k} \lambda_i(\langle r_i, \tilde{z}_i - z\rangle_{\mathcal{Z}} + \varepsilon_i).$$

Letting $z = \tilde{z}_k^a$ in the last inequality and using the fact that $(dw)_{z_0}(\cdot)$ is convex, we obtain

$$\max_{i=1,\ldots,k}(dw)_{z_0}(\tilde{z}_i) + \eta_0 \geq (dw)_{z_0}(\tilde{z}_k^a) + \eta_0 \geq \sum_{i=1}^{k} \lambda_i(\langle r_i, \tilde{z}_i - \tilde{z}_k^a\rangle_{\mathcal{Z}} + \varepsilon_i) = \Lambda_k \varepsilon_k^a \tag{30}$$

where the equality is due to (26). On the other hand, (29) implies that, for every $i \geq 1$ and $z^* \in T^{-1}(0)$,

$$\begin{aligned}
(dw)_{z_0}(\tilde{z}_i) &\leq \frac{3M}{m}\left[(dw)_{z_i}(\tilde{z}_i) + (dw)_{z_i}(z^*) + (dw)_{z_0}(z^*)\right] \\
&\leq \frac{3M}{m}\left[\sigma(dw)_{z_{i-1}}(\tilde{z}_i) + \eta_{i-1} + (dw)_{z_{i-1}}(z^*) + \eta_{i-1} + (dw)_{z_0}(z^*)\right] \\
&\leq \frac{3M}{m}\left[\sigma(dw)_{z_{i-1}}(\tilde{z}_i) + 2((dw)_{z_{i-1}}(z^*) + \eta_{i-1}) + (dw)_{z_0}(z^*)\right] \\
&\leq \frac{3M}{m}\left[\sigma(dw)_{z_{i-1}}(\tilde{z}_i) + 3(dw)_{z_0}(z^*) + 2\eta_0\right]
\end{aligned}$$

where the second inequality is due to (29) and Lemma 3.6(d), and the last inequality is due to Lemma 3.6(d). Combining the above relations with (30) and using the definitions of ρ_k and $(dw)_0$ and the fact that $M/m \geq 1$, we conclude that the bound on ε_k^a holds.

We now establish the bounds on ρ_k under either one of the conditions (a) or (b). First, if $\sigma < 1$, then it follows from Lemma 3.6(c)-(d) that $(1-\sigma)(dw)_{z_{i-1}}(\tilde{z}_i) \leq (dw)_{z_{i-1}}(z^*) + \eta_{k-1} \leq (dw)_{z_0}(z^*) + \eta_0$ for every $i \geq 1$ and $z^* \in T^{-1}(0)$, and hence that (27) holds. Assume now that $\operatorname{Dom} T$ is bounded. Then, it follows from inequality (29) and Lemma 3.6(d) that, for every $i \geq 1$ and $z^* \in T^{-1}(0)$,

$$(dw)_{z_{i-1}}(\tilde{z}_i) \leq \frac{2M}{m}\left[(dw)_{z_{i-1}}(z^*) + \min\{(dw)_{\tilde{z}_i}(z^*), (dw)_{z^*}(\tilde{z}_i)\}\right] \leq \frac{2M}{m}\left[(dw)_{z_0}(z^*) + \eta_0 + D\right]$$

which, in view of definitions of ρ_k and $(dw)_0$, proves (b). □

4 Convergence rate analysis of the proximal ADMM

Our goal in this section is to show that the proximal ADMM is an instance of the NE-HPE framework for solving the inclusion problem (4) and, as a by-product, establish its pointwise and ergodic convergence rate bounds presented in Section 2.

We start by presenting a preliminary technical result about the proximal ADMM.

Lemma 4.1. *Consider the triple (s_k, y_k, x_k) generated at the k-iteration of the proximal ADMM and the point $\tilde{x}_k$ defined in (11). Then,*

$$0 \in H(s_k - s_{k-1}) + [\partial g(s_k) - D^*\tilde{x}_k], \tag{31}$$

$$0 \in (G + \beta C^*C)(y_k - y_{k-1}) + [\partial f(y_k) - C^*\tilde{x}_k], \tag{32}$$

$$0 = \frac{1}{\theta\beta}(x_k - x_{k-1}) + [Cy_k + Ds_k - c], \tag{33}$$

$$\tilde{x}_k - x_{k-1} = \beta C(y_k - y_{k-1}) + \frac{x_k - x_{k-1}}{\theta}. \tag{34}$$

Proof. From the optimality condition of (5), we have

$$0 \in \partial g(s_k) - D^*(x_{k-1} - \beta(Cy_{k-1} + Ds_k - c)) + H(s_k - s_{k-1}),$$

which, combined with definition of $\tilde{x}_k$ in (11), yields (31). Now, from the optimality condition of (6) and definition of $\tilde{x}_k$ in (11), we obtain

$$\begin{aligned} 0 &\in \partial f(y_k) - C^*x_{k-1} + \beta C^*(Cy_k + Ds_k - c) + G(y_k - y_{k-1}) \\ &= \partial f(y_k) - C^*[x_{k-1} + \beta(Cy_{k-1} + Ds_k - c)] + \beta C^*C(y_k - y_{k-1}) + G(y_k - y_{k-1}) \\ &= \partial f(y_k) - C^*\tilde{x}_k + \beta C^*C(y_k - y_{k-1}) + G(y_k - y_{k-1}). \end{aligned}$$

which proves (32). Moreover, (33) follows immediately from (7). On the other hand, it follows from definition of x_k in (7) that

$$\frac{x_k - x_{k-1}}{\theta} + \beta C(y_k - y_{k-1}) = -\beta(Cy_{k-1} + Ds_k - c)$$

which, combined with definition of $\tilde{x}_k$ in (11), yields (34).

□

In order to show that the proximal ADMM is an instance of the NE-HPE framework, we need to introduce the elements required by the setting of Section 3, namely, the space $\mathcal{Z}$, the seminorm $\|\cdot\|$ on $\mathcal{Z}$, the distance generating function $w : \mathcal{Z} \to [-\infty, \infty]$ and the convex set $Z \subset \operatorname{int}(\operatorname{dom} w)$. We consider $\mathcal{Z} := \mathcal{S} \times \mathcal{Y} \times \mathcal{X}$ and endow it with the inner product given by

$$\langle z, z' \rangle := \langle s, s' \rangle_{\mathcal{S}} + \langle y, y' \rangle_{\mathcal{Y}} + \langle x, x' \rangle_{\mathcal{X}} \quad \forall\, z = (s, y, x),\ z' = (s, y, x).$$

The seminorm $\|\cdot\|$, the function w and the set Z are defined as

$$\|z\| := \left(\|s\|^2_{\mathcal{S},H} + \|y\|^2_{\mathcal{Y},(G+\beta C^*C)} + \frac{1}{\beta\theta}\|x\|^2_{\mathcal{X}} \right)^{1/2} \quad w(z) := \frac{1}{2}\|(s, y, x)\|^2, \quad Z := \mathcal{Z} \tag{35}$$

for every $z = (s, y, x) \in \mathcal{Z}$. Clearly, the Bregman distance associated with w is given by

$$(dw)_z(z') = \frac{1}{2}\|s' - s\|^2_{\mathcal{S},H} + \frac{1}{2}\|y' - y\|^2_{\mathcal{Y},(G+\beta C^*C)} + \frac{1}{2\beta\theta}\|x' - x\|^2_{\mathcal{X}} \tag{36}$$

for every $z = (s, y, x) \in \mathcal{Z}$ and $z' = (s', y', x') \in \mathcal{Z}$.

Using Proposition 1.1 and the fact that $\|\cdot\| = \|\cdot\|_{\mathcal{Z},Q}$ where Q is the self-adjoint positive semidefinite linear operator given by

$$Q(s, y, x) = (Hs, (G + \beta C^*C)y, x/(\beta\theta)) \quad \forall (s, y, x) \in \mathcal{Z},$$

it is easy to see that the function w is a $(1, 1)$-regular distance generating function with respect to $(Z, \|\cdot\|)$.

To simplify some relations in the proofs given below, define

$$\Delta s_k = s_k - s_{k-1}, \quad \Delta y_k = y_k - y_{k-1}, \quad \Delta x_k = x_k - x_{k-1}. \tag{37}$$

The following technical result will be used to prove that the proximal ADMM is an instance of the NE-HPE framework.

Lemma 4.2. *Let $\{(s_k, y_k, x_k)\}$ be the sequence generated by the proximal ADMM. Then, the following statements hold:*

(a) if $\theta < 2$, then

$$\frac{1}{\sqrt{\theta}}\left(\frac{1}{2}\|\Delta y_1\|^2_{\mathcal{Y},G} - \frac{1}{\sqrt{\theta}}\langle C\Delta y_1, \Delta x_1\rangle_{\mathcal{X}}\right) \le \tau_\theta d_0$$

where d_0 and τ_θ are as in (8) *and* (10)*, respectively.*

(b) for any $\theta > 0$, we have

$$\frac{1}{\theta}\langle C\Delta y_k, \Delta x_k\rangle_{\mathcal{X}} \ge \frac{1-\theta}{\theta}\langle C\Delta y_k, \Delta x_{k-1}\rangle_{\mathcal{X}} + \frac{1}{2}\|\Delta y_k\|^2_{\mathcal{Y},G} - \frac{1}{2}\|\Delta y_{k-1}\|^2_{\mathcal{Y},G}, \quad \forall k \ge 2.$$

Proof. (a) Let a point $z^* := (s^*, y^*, x^*)$ be such that $0 \in T(s^*, y^*, x^*)$ (see assumption **A1**). Since $\langle x, x'\rangle_{\mathcal{X}} \le (1/2)(\|x\|^2_{\mathcal{X}} + \|x'\|^2_{\mathcal{X}})$ for every $x, x' \in \mathcal{X}$, using (37) we obtain

$$\begin{aligned}
\frac{1}{2}\|\Delta y_1\|^2_{\mathcal{Y},G} - \frac{1}{\sqrt{\theta}}\langle C\Delta y_1, \Delta x_1\rangle_{\mathcal{X}} &\le \frac{1}{2\beta\theta}\|\Delta x_1\|^2_{\mathcal{X}} + \frac{\beta}{2}\|C\Delta y_1\|^2_{\mathcal{X}} + \frac{1}{2}\|\Delta y_1\|^2_{\mathcal{Y},G} \\
&\le \frac{1}{\beta\theta}\|x_1 - x^*\|^2_{\mathcal{X}} + \beta\|C(y_1 - y^*)\|^2_{\mathcal{X}} + \|y_1 - y^*\|^2_{\mathcal{Y},G} \\
&\quad + \frac{1}{\beta\theta}\|x_0 - x^*\|^2_{\mathcal{X}} + \beta\|C(y_0 - y^*)\|^2_{\mathcal{X}} + \|y_0 - y^*\|^2_{\mathcal{Y},G}
\end{aligned}$$

which, combined with (36) and simple calculus, yields

$$\frac{1}{\sqrt{\theta}}\left(\frac{1}{2}\|\Delta y_1\|^2_{\mathcal{Y},G} - \frac{1}{\sqrt{\theta}}\langle C\Delta y_1, \Delta x_1\rangle_{\mathcal{X}}\right) \le \frac{2}{\sqrt{\theta}}\left((dw)_{z_1}(z^*) + (dw)_{z_0}(z^*)\right). \tag{38}$$

On the other hand, consider

$$z_0 = (s_0, y_0, x_0), \quad z_1 = (s_1, y_1, x_1), \quad \tilde{z}_1 = (s_1, y_1, \tilde{x}_1), \quad \lambda_1 = 1, \quad \varepsilon_1 = 0. \tag{39}$$

Lemma 4.1 implies that inclusion (23) is satisfied for $(z_0, z_1, \tilde{z}_1, \lambda_1, \varepsilon_1)$ with T and dw as in (4) and (36), respectively. Hence, it follows from Lemma 3.6(a) with $z = z^*$, $\lambda_1 = 1$ and the fact that $\langle r_1, \tilde{z}_1 - z^*\rangle \ge 0$ (because $0 \in T(z^*)$ and $r_1 \in T(\tilde{z}_1)$) that

$$(dw)_{z_1}(z^*) \le (dw)_{z_0}(z^*) + (dw)_{z_1}(\tilde{z}_1) - (dw)_{z_0}(\tilde{z}_1). \tag{40}$$

Using the definitions in (36) and (39) and equation in (34), we obtain

$$\begin{aligned}(dw)_{z_1}(\tilde{z}_1) - (dw)_{z_0}(\tilde{z}_1) &\leq \frac{1}{2\beta\theta}\|\tilde{x}_1 - x_1\|_{\mathcal{X}}^2 - \frac{\beta}{2}\|C(y_1 - y_0)\|_{\mathcal{X}}^2 - \frac{1}{2\beta\theta}\|\tilde{x}_1 - x_0\|_{\mathcal{X}}^2 \\ &= \frac{(\theta-1)}{2\beta\theta^2}\|x_1 - x_0\|_{\mathcal{X}}^2 - \frac{1}{2}\left\|\frac{x_1 - x_0}{\theta\sqrt{\beta}} + \sqrt{\beta}C(y_1 - y_0)\right\|_{\mathcal{X}}^2 \\ &\leq \frac{(\theta-1)}{2\beta\theta^2}\|x_1 - x_0\|_{\mathcal{X}}^2.\end{aligned}$$

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

$$(dw)_{z_1}(\tilde{z}_1) \leq (dw)_{z_0}(\tilde{z}_1). \tag{41}$$

Now, if $\theta \in (1, 2)$, we have

$$\begin{aligned}(dw)_{z_1}(\tilde{z}_1) - (dw)_{z_0}(\tilde{z}_1) &\leq \frac{(\theta-1)}{2\beta\theta^2}\|x_1 - x_0\|_{\mathcal{X}}^2 \leq \frac{2(\theta-1)}{\theta}\left(\frac{\|x_1 - x^*\|_{\mathcal{X}}^2}{2\beta\theta} + \frac{\|x_0 - x^*\|_{\mathcal{X}}^2}{2\beta\theta}\right) \\ &\leq \frac{2(\theta-1)}{\theta}\left[(dw)_{z_1}(z^*) + (dw)_{z_0}(z^*)\right]\end{aligned}$$

where the second inequality is due to the fact that $2ab \leq a^2 + b^2$ for all $a, b \geq 0$, and the last inequality is due to (36) and definitions of z_0, z_1 and z^*. Hence, combining the last estimative with (40), we obtain

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

which, combined with (41), yields

$$(dw)_{z_1}(z^*) \leq \max\left\{1, \frac{3\theta-2}{2-\theta}\right\}(dw)_{z_0}(z^*).$$

Therefore, statement (a) follows from (38), the last inequality, definition of τ_θ in (10) and the fact that d_0 (as defined in (8)) satisfies $d_0 = \inf_{z \in T^{-1}(0)}(dw)_{z_0}(z)$.

(b) From the inclusion (32) and relation (34), we see that, for every $j \geq 1$,

$$\partial f(y_j) \ni C^*(\tilde{x}_j - \beta C(y_j - y_{j-1})) - G(y_j - y_{j-1}) = \frac{1}{\theta}C^*(x_j - (1-\theta)x_{j-1}) - G(y_j - y_{j-1}).$$

For every $k \geq 2$, using the previous inclusion for $j = k-1$ and $j = k$, it follows from the monotonicity of the subdifferential of f that

$$0 \leq \left\langle \frac{1}{\theta}C^*(x_k - x_{k-1}) - \frac{(1-\theta)}{\theta}C^*(x_{k-1} - x_{k-2}) - G(y_k - y_{k-1}) + G(y_{k-1} - y_{k-2}), y_k - y_{k-1}\right\rangle_{\mathcal{Y}},$$

which, combined with (37), yields

$$\frac{1}{\theta}\langle C\Delta y_k, \Delta x_k\rangle_{\mathcal{X}} \geq \frac{(1-\theta)}{\theta}\langle C\Delta y_k, \Delta x_{k-1}\rangle_{\mathcal{X}} + \|\Delta y_k\|_{\mathcal{Y},G}^2 - \langle G\Delta y_{k-1}, \Delta y_k\rangle_{\mathcal{Y}}.$$

Hence, item (b) follows from the last inequality and the fact that $\langle Gy, y'\rangle_{\mathcal{Y}} \leq (1/2)(\|y\|_{\mathcal{Y},G}^2 + \|y'\|_{\mathcal{Y},G}^2)$ for every $y, y' \in \mathcal{Y}$. Therefore, the proof of the lemma is concluded. □

We now present some properties of the parameter σ_θ defined in (9).

Lemma 4.3. *Let $\theta \in (0, (\sqrt{5}+1)/2]$ be given and consider the parameter σ_θ as defined in (9). Then, the following statements hold:*

(a) $\sigma = \sigma_\theta$ is the largest root of the equation $\det(M_\theta(\sigma)) = 0$ and $\det(M_\theta(\sigma)) > 0$ for every $\sigma > \sigma_\theta$ where $\det(\cdot)$ denotes the determinant function and

$$M_\theta(\sigma) := \begin{bmatrix} \sigma(1+\theta) - 1 & (\sigma+\theta-1)(1-\theta) \\ (\sigma+\theta-1)(1-\theta) & \sigma - (1-\theta)^2 \end{bmatrix}; \tag{42}$$

(b) $1/3 < \max\{(1-\theta)^2, 1-\theta, 1/(1+\theta)\} \le \sigma_\theta \le 1$;

(c) the matrix $M_\theta(\sigma)$ in (42) is positive semidefinite for $\sigma = \sigma_\theta$.

Proof. (a) It is a simple algebraic computation to see that $\sigma = \sigma_\theta$ is the largest root of the second-order equation $\det(M_\theta(\sigma)) = 0$.

(b) The second inequality follows by (a) and the fact that $\det(M_\theta(\sigma)) \le 0$ for σ equal to $(1-\theta)^2$, $1-\theta$ and $1/(1+\theta)$. Now, the first and third inequalities are due to the fact that $\theta \in (0, (\sqrt{5}+1)/2]$ and $1/3 \le 1/(1+\theta)$.

(c) Statements (a) and (b) imply that $\det(M_\theta(\sigma_\theta)) = 0$ and the main diagonal entries of $M_\theta(\sigma_\theta)$ are nonnegative. Since $M_\theta(\sigma)$ is symmetric, we then conclude that (c) holds. □

The next result shows that the proximal ADMM can be seen as an instance of the NE-HPE framework.

Theorem 4.4. *Consider the operator T and Bregman distance dw as in (4) and (36), respectively. Let $\{(s_k, y_k, x_k)\}$ be the sequence generated by the proximal ADMM with $\theta \in (0, (\sqrt{5}+1)/2]$ and consider $\{\tilde{x}_k\}$ as in (11). Define*

$$z_{k-1} = (s_{k-1}, y_{k-1}, x_{k-1}), \quad \tilde{z}_k = (s_k, y_k, \tilde{x}_k), \quad \lambda_k = 1, \quad \varepsilon_k = 0 \quad \forall k \ge 1, \tag{43}$$

and the sequence $\{\eta_k\}$ as

$$\eta_0 = \tau_\theta d_0, \quad \eta_k = [\sigma_\theta - (\theta-1)^2]\frac{\|\Delta x_k\|_{\mathcal{X}}^2}{2\beta\theta^3} + \frac{\sigma_\theta + \theta - 1}{2\theta}\|\Delta y_k\|_{\mathcal{Y},G}^2, \quad \forall k \ge 1 \tag{44}$$

where d_0, σ_θ, τ_θ and $(\Delta x_k, \Delta y_k)$ are as in (8), (9), (10) and (37), respectively. Then, the sequence $\{(z_k, \tilde{z}_k, \lambda_k, \varepsilon_k, \eta_k)\}$ is an instance of the NE-HPE framework with input $z_0 = (s_0, y_0, x_0), \eta_0$ and $\sigma = \sigma_\theta$.

Proof. The inclusion (23) follows from (31)-(33), (43) and definitions of T and dw. Now it remains to show that the error condition (24) holds. First of all, it follows from (34), (36), (37) and (43) that

$$\begin{aligned}(dw)_{z_k}(\tilde{z}_k) + \lambda_k\varepsilon_k &= \frac{1}{2\beta\theta}\|\tilde{x}_k - x_k\|_{\mathcal{X}}^2 = \frac{1}{2\beta\theta}\left\|\beta C\Delta y_k + \frac{1-\theta}{\theta}\Delta x_k\right\|_{\mathcal{X}}^2 \\ &= \frac{\beta}{2\theta}\|C\Delta y_k\|_{\mathcal{X}}^2 + \frac{(1-\theta)}{\theta^2}\langle C\Delta y_k, \Delta x_k\rangle_{\mathcal{X}} + \frac{(1-\theta)^2}{2\beta\theta^3}\|\Delta x_k\|_{\mathcal{X}}^2. \end{aligned} \tag{45}$$

Also, (36), (37) and (43) imply that

$$(dw)_{z_{k-1}}(\tilde{z}_k) = \frac{1}{2}\|\Delta s_k\|^2_{\mathcal{S},H} + \frac{1}{2}\|\Delta y_k\|^2_{\mathcal{Y},G} + \frac{\beta}{2}\|C\Delta y_k\|^2_{\mathcal{X}} + \frac{1}{2\beta\theta}\|x_{k-1} - \tilde{x}_k\|^2_{\mathcal{X}}. \tag{46}$$

It follows from (34) and (37) that

$$\|x_{k-1} - \tilde{x}_k\|^2_{\mathcal{X}} = \left\|\beta C\Delta y_k + \frac{1}{\theta}\Delta x_k\right\|^2_{\mathcal{X}} = \beta^2\|C\Delta y_k\|^2_{\mathcal{X}} + \frac{2\beta}{\theta}\langle C\Delta y_k, \Delta x_k\rangle_{\mathcal{X}} + \frac{1}{\theta^2}\|\Delta x_k\|^2_{\mathcal{X}}$$

which, combined with (46), yields

$$(dw)_{z_{k-1}}(\tilde{z}_k) = \frac{\|\Delta s_k\|^2_{\mathcal{S},H}}{2} + \frac{\|\Delta y_k\|^2_{\mathcal{Y},G}}{2} + \frac{\beta(\theta+1)\|C\Delta y_k\|^2_{\mathcal{X}}}{2\theta} + \frac{1}{\theta^2}\langle C\Delta y_k, \Delta x_k\rangle_{\mathcal{X}} + \frac{\|\Delta x_k\|^2_{\mathcal{X}}}{2\beta\theta^3}. \tag{47}$$

Therefore, combining (45) and (47), we see, after simple algebraic manipulations, that the error condition (24) is satisfied if and only if

$$\begin{aligned}[\sigma(1+\theta)-1]\frac{\beta\|C\Delta y_k\|^2_{\mathcal{X}}}{2\theta} + \left[\sigma - (\theta-1)^2\right]\frac{\|\Delta x_k\|^2_{\mathcal{X}}}{2\beta\theta^3} + \sigma\frac{\|\Delta y_k\|^2_{\mathcal{Y},G}}{2} + \frac{(\sigma+\theta-1)}{\theta^2}\langle C\Delta y_k, \Delta x_k\rangle_{\mathcal{X}} \\ \geq \eta_k - \eta_{k-1} - \sigma\frac{\|\Delta s_k\|^2_{\mathcal{S},H}}{2}.\end{aligned} \tag{48}$$

We now show that inequality (48) with $\sigma = \sigma_\theta$ holds for $k = 1$. Indeed, it follows from definition of η_1 and $\sigma_\theta \geq 1/(1+\theta)$ (see Lemma 4.3 (b)) that

$$\begin{aligned}&[\sigma_\theta(1+\theta)-1]\frac{\beta\|C\Delta y_1\|^2_{\mathcal{X}}}{2\theta} + \left[\sigma_\theta - (1-\theta)^2\right]\frac{\|\Delta x_1\|^2_{\mathcal{X}}}{2\beta\theta^3} + \sigma_\theta\frac{\|\Delta y_1\|^2_{\mathcal{Y},G}}{2} + \frac{(\sigma_\theta+\theta-1)}{\theta^2}\langle C\Delta y_1, \Delta x_1\rangle_{\mathcal{X}} \\ &\geq \left[\sigma_\theta - \frac{\sigma_\theta+\theta-1}{\theta} + \frac{\sigma_\theta+\theta-1}{\theta^{3/2}}\right]\frac{\|\Delta y_1\|^2_{\mathcal{Y},G}}{2} + \eta_1 + \frac{(\sigma_\theta+\theta-1)}{\theta^{3/2}}\left(\frac{1}{\sqrt{\theta}}\langle C\Delta y_1, \Delta x_1\rangle_{\mathcal{X}} - \frac{1}{2}\|\Delta y_1\|^2_{\mathcal{Y},G}\right) \\ &\geq \left[\sigma_\theta - \frac{\sigma_\theta+\theta-1}{3\theta}\right]\frac{\|\Delta y_1\|^2_{\mathcal{Y},G}}{2} + \eta_1 - \frac{(\sigma_\theta+\theta-1)}{\theta}\tau_\theta d_0 \\ &\geq \eta_1 - \eta_0\end{aligned}$$

where the second inequality follows from the fact that $\sqrt{\theta} \leq 3/2$ and Lemmas 4.2(a) and 4.3(b), and the third inequality is due to the fact that $1/3 \leq \sigma_\theta \leq 1$ (see Lemma 4.3(b)) and definition of η_0. Therefore, inequality (48) holds with $k = 1$ and $\sigma = \sigma_\theta$.

We next show that inequality (48) with $\sigma = \sigma_\theta$ holds for $k \geq 2$. Using Lemma 4.2(b) and the definition of $\{\eta_k\}$ in (44), we see, after simple calculus, that a sufficient condition for (48) to hold with $\sigma = \sigma_\theta$ and $k \geq 2$ is that

$$(\sigma_\theta(1+\theta)-1)\beta\frac{\|C\Delta y_k\|^2_{\mathcal{X}}}{2} + [\sigma_\theta - (1-\theta)^2]\frac{\|\Delta x_{k-1}\|^2_{\mathcal{X}}}{2\beta\theta^2} + \frac{(\sigma_\theta+\theta-1)(1-\theta)}{\theta}\langle C\Delta y_k, \Delta x_{k-1}\rangle_{\mathcal{X}} \geq 0.$$

Hence, since the last inequality holds due to Lemma 4.3(c), we conclude the proof of the theorem. □

Now we are ready to present the proof of the pointwise convergence rate of the proximal ADMM.

Proof of Theorem 2.1: Since $\sigma_\theta \in [0,1)$ for any $\theta \in (0,(\sqrt{5}+1)/2)$ and w as defined in (35) is a $(1,1)$-regular distance generating function and $\underline{\lambda} := \inf \lambda_k = 1$, we obtain by combining Theorems 4.4 and 3.3(a) that inclusion (12) holds and there exists $i \leq k$ such that

$$\left(\|s_{i-1}-s_i\|^2_{\mathcal{S},H} + \|y_{i-1}-y_i\|^2_{\mathcal{Y},(G+\beta C^*C)} + \frac{1}{\beta\theta}\|x_{i-1}-x_i\|^2_{\mathcal{X}}\right)^{1/2} \leq \frac{2}{\sqrt{k}}\sqrt{\frac{(1+\sigma_\theta)d_0+2\eta_0}{1-\sigma_\theta}},$$

where we also used the definition of the norm $\|\cdot\|$ in (35) and Proposition 1.1. The result now follows from the last inequality and definition of η_0 in (44). □

In order to establish the ergodic convergence rate of the proximal ADMM, we need the next auxiliary result.

Lemma 4.5. *Let $\{(s_k,y_k,x_k)\}$ be the sequence generated by the proximal ADMM and $\{\tilde{x}_k\}$ be given by (11). Then, the pair $(z_{k-1},\tilde{z}_k)$ as defined in (43) satisfies*

$$(dw)_{z_{k-1}}(\tilde{z}_k) \leq \frac{4(1+\tau_\theta)(\theta^2+\theta+1)}{\theta^2}d_0 \quad k \geq 1,$$

where dw is the Bregman distance given in (36), d_0 and τ_θ are as in (8) and (10), respectively.

Proof. It follows from (36) and (43) that

$$(dw)_{z_{k-1}}(\tilde{z}_k) = \frac{1}{2}\|\Delta s_k\|^2_{\mathcal{S},H} + \frac{1}{2}\|\Delta y_k\|^2_{\mathcal{Y},G} + \frac{\beta}{2}\|C\Delta y_k\|^2_{\mathcal{X}} + \frac{1}{2\beta\theta}\|x_{k-1}-\tilde{x}_k\|^2_{\mathcal{X}}. \tag{49}$$

On the other hand, using (34) we have

$$\begin{aligned}\frac{1}{2\beta\theta}\|x_{k-1}-\tilde{x}_k\|^2_{\mathcal{X}} &= \frac{1}{2\beta\theta}\|\beta C\Delta y_k + \frac{\Delta x_k}{\theta}\|^2_{\mathcal{X}} \\ &= \frac{\beta}{2\theta}\|C\Delta y_k\|^2_{\mathcal{X}} + \frac{1}{\beta\theta^2}\langle \beta C\Delta y_k, \Delta x_k\rangle_{\mathcal{X}} + \frac{1}{2\beta\theta^3}\|\Delta x_k\|^2_{\mathcal{X}} \\ &\leq \frac{\beta(\theta+1)}{2\theta^2}\|C\Delta y_k\|^2_{\mathcal{X}} + \frac{\theta+1}{2\beta\theta^3}\|\Delta x_k\|^2_{\mathcal{X}}\end{aligned}$$

where the inequality is due to Cauchy-Schwarz inequality and the fact that $2ab \leq a^2+b^2$ for all $a,b \geq 0$. Combining the last inequality and (49), we have

$$\begin{aligned}(dw)_{z_{k-1}}(\tilde{z}_k) &\leq \frac{(\theta^2+\theta+1)}{\theta^2}\Big[\frac{1}{2}\|\Delta s_k\|^2_{\mathcal{S},H} + \frac{1}{2}\|\Delta y_k\|^2_{\mathcal{Y},G} + \frac{\beta}{2}\|C\Delta y_k\|^2_{\mathcal{X}} + \frac{1}{2\beta\theta}\|\Delta x_k\|^2_{\mathcal{X}}\Big] \\ &\leq \frac{2(\theta^2+\theta+1)}{\theta^2}\left[(dw)_{z_{k-1}}(z^*) + (dw)_{z_k}(z^*)\right]\end{aligned}$$

where the last inequality is due to (36) and the fact that $2ab \leq a^2+b^2$ for all $a,b \geq 0$. Hence, since by Theorem 4.4 the proximal ADMM is an instance of the NE-HPE framework, it follows from the last estimative and Lemma 3.6(d) that

$$(dw)_{z_{k-1}}(\tilde{z}_k) \leq \frac{4(\theta^2+\theta+1)}{\theta^2}((dw)_{z_0}(z^*)+\eta_0)$$

which, combined with the definition of η_0 in (44) and the fact that d_0 (as defined in (8)) satisfies $d_0 = \inf_{z\in T^{-1}(0)}(dw)_{z_0}(z)$, proves the result. □

Next we present the proof of the ergodic iteration-complexity bound for the proximal ADMM.

Proof of Theorem 2.2: First, it follows from Theorem 4.4 that the proximal ADMM with $\theta \in (0, (\sqrt{5}+1)/2]$ is an instance of the NE-HPE applied to problem (4) in which $\sigma := \sigma_\theta$, $\{(z_k, \tilde{z}_k, \lambda_k, \varepsilon_k)\}$ and $\{\eta_k\}$ are as defined in (9), (43) and (44), respectively. Moreover, inclusions (31)-(33) is equivalent to

$$H(s_{k-1}-s_k)+D^*\tilde{x}_k \in \partial g(s_k), \quad (G+\beta C^*C)(y_{k-1}-y_k)+C^*\tilde{x}_k \in \partial f(y_k), \quad \frac{1}{\beta\theta}(x_{k-1}-x_k) = Cy_k+Ds_k-c.$$

Hence, using (13) we obtain trivially the third inclusion of (15) while the first and second inclusions of (15) follow from (13) and the transportation formula (see [2, Theorem 2.3]). Now, since w as defined in (35) is a $(1,1)$-regular distance generating function and $1 = \underline{\lambda} = \inf \lambda_k$, we obtain from Theorem 3.4 and Lemma 4.5 that

$$\|(r^a_{k,s}, r^a_{k,y}, r^a_{k,x})\|^* \le \frac{2\sqrt{2(d_0+\eta_0)}}{k}, \; \varepsilon^a_k \le \frac{3[3\theta^2(d_0+\eta_0) + 4\sigma_\theta(1+\tau_\theta)(\theta^2+\theta+1)d_0]}{\theta^2 k}, \forall k \ge 1. \quad (50)$$

Hence, since $(r^a_{k,s}, r^a_{k,y}, r^a_{k,x}) = (H(s^a_{k-1} - s^a_k), (G+\beta C^*C)(y^a_{k-1} - y^a_k), (x^a_{k-1} - x^a_k)/(\beta\theta)$, it follows from the definition of $\|\cdot\|$ in (35) and Proposition 1.1 that

$$\|(r^a_{k,s}, r^a_{k,y}, r^a_{k,x})\|^* = \left(\|s^a_{k-1} - s^a_k\|^2_{\mathcal{S},H} + \|y^a_{k-1} - y^a_k\|^2_{\mathcal{Y},(G+\beta C^*C)} + \frac{1}{\beta\theta}\|x^a_{k-1} - x^a_k\|^2_{\mathcal{X}}\right)^{1/2}.$$

Therefore, the result follows from (50) and the definition of η_0 in (44). □

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

[10] R. Glowinski. *Numerical Methods for Nonlinear Variational Problems*. Springer Series in Computational Physics. Springer-Verlag, 1984.

[11] 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.

[12] 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. *Avaliable on http://www.arxiv.org*.

[13] Y. Gu, B. Jiang, and H. Deren. A semi-proximal-based strictly contractive Peaceman-Rachford splitting method. *Avaliable on http://www.arxiv.org*.

[14] 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.

[15] B. He, F. Ma, and X. Yuan. On the step size of symmetric alternating directions method of multipliers. *Avaliable on http://www.optimization-online.org*.

[16] 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.

[17] 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.

[18] Y. He and R. D. C. Monteiro. Accelerating block-decomposition first-order methods for solving composite saddle-point and two-player Nash equilibrium problems. *SIAM J. Optim.*, 25(4):2182–2211, 2015.

[19] 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.

[20] O. Kolossoski and R. D. C. Monteiro. An accelerated non-euclidean hybrid proximal extragradient-type algorithm for convex-concave saddle-point problems. *Avaliable on http://www.optimization-online.org*.

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

[22] 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. *Avaliable on http://www.arxiv.org*.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

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

[28] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.*, 14(5):877–898, 1976.

[29] 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.

[30] M. V. Solodov and B. F. Svaiter. An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. *Math. Oper. Res.*, 25(2):214–230, 2000.

[31] X. Wang and X Yuan. The linearized alternating direction method of multipliers for dantzig selector. *SIAM J. Sci. Comput.*, 34(5):2792–2811, 2012.

[32] J. Yang and X. Yuan. Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization. *Math. Comput.*, 82(281):301–329, 2013.

[33] 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.

[34] X. Zhang, M. Burger, and S. Osher. A unified primal-dual algorithm framework based on Bregman iteration. *J. Sci. Comput.*, 46(1):20–46, 2010.