On the iteration-complexity of a non-Euclidean hybrid proximal extragradient framework and of a proximal ADMM

Max L. N. Gonçalves, Jefferson G. Melo & Renato D. C. Monteiro

To cite this article: Max L. N. Gonçalves, Jefferson G. Melo & Renato D. C. Monteiro (2019): On the iteration-complexity of a non-Euclidean hybrid proximal extragradient framework and of a proximal ADMM, Optimization, DOI: 10.1080/02331934.2019.1652297

To link to this article: https://doi.org/10.1080/02331934.2019.1652297

Published online: 09 Aug 2019.
On the iteration-complexity of a non-Euclidean hybrid proximal extragradient framework and of a proximal ADMM

Max L. N. Gonçalves a, Jefferson G. Melo a and Renato D. C. Monteiro b

aIME, Universidade Federal de Goias, Goiânia, Brazil; bSchool of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA

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.

1. Introduction
This paper considers the following linearly constrained convex problem:

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

where $S$, $Y$ and $X$ are finite-dimensional inner product spaces, $f : Y \to (-\infty, \infty]$ and $g : S \to (-\infty, \infty]$ are proper closed convex functions, $C : Y \to X$ and $D : S \to X$ are linear operators, and $c \in X$. 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, [1–11]. Here, we study the proximal ADMM [4,12] which, recursively, computes a 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

\[
\begin{align*}
s_k &= \arg\min_s \left\{ g(s) - \langle x_{k-1}, Ds \rangle + \beta \frac{1}{2} \| Cy_{k-1} + Ds - c \|_2^2 \\
&\quad + \frac{1}{2} \langle s - s_{k-1}, H(s - s_{k-1}) \rangle \right\}, \\
y_k &= \arg\min_y \left\{ f(y) - \langle x_{k-1}, Cy \rangle + \beta \frac{1}{2} \| Cy + Ds_k - c \|_2^2 \\
&\quad + \frac{1}{2} \langle y - y_{k-1}, G(y - y_{k-1}) \rangle \right\}, \\
x_k &= x_{k-1} - \theta \beta \left[ Cy_k + Ds_k - c \right],
\end{align*}
\]

where \(\beta > 0\) is a penalty parameter, \(\theta > 0\) is a stepsize parameter, and \(H : S \to S\) and \(G : X \to X\) 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 \(\tau \geq \beta \| D \|_2^2\) is known as the linearized ADMM or the split inexact Uzawa method (see, e.g. [1,9,13,14]). It has the desirable feature that, for many applications, its subproblems are much easier to solve or even have closed-form solutions (see [3,9,15,16] for more details).

Pointwise and ergodic iteration-complexity results for the proximal ADMM (2) for any \(\theta \in (0, (1 + \sqrt{5})/2)\) were established in [17]. 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 [18,19] and is thoroughly discussed in [20,21]. 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 [22] under the assumptions that \(C\) is injective. More specifically, the ergodic iteration-complexity for the standard ADMM is derived in [22] for any \(\theta \in (0, 1)\) while a weak pointwise iteration-complexity easily follows from the approach in [22] for any \(\theta \in (0, 1)\). Paper [1] analysed a primal–dual scheme for solving a saddle-point problem associated
to (1) with $C = I$ and $c = 0$, and established, as a by-product, an ergodic iteration-
complexity bound for the linearized ADMM with $\theta = 1$. An ergodic iteration-
complexity result for the latter algorithm applied to solve (1) was obtained in [9].
It should be noted however that [1, 9, 23] do not provide any details on how to
obtain an easily verifiable ergodic termination criterion with a well-established
iteration-complexity bound. A strong pointwise iteration-complexity bound for
the proximal ADMM (2) with $G = 0$ and $\theta = 1$ was derived in [24]. A num-
ber of papers (see for example [3, 5, 6, 8, 10, 11, 17, 25] and references therein) have
extended most of these complexity results to the context of the ADMM class (2)
as well as other ADMM classes. Finally, during the review process of this paper,
we became aware of paper [26], which derived ergodic complexity results for the
proximal ADMM with stepsize parameter $\theta \in (0, (1 + \sqrt{5})/2)$. It was shown that
the latter algorithm is an instance of an HPE scheme which, on the other hand, is
a special instance of the non-Euclidean HPE framework analysed here. It should
be mentioned that the analysis presented in [26] does not include the extreme
case $\theta = (1 + \sqrt{5})/2$.

The non-Euclidean HPE framework is a class of inexact proximal point meth-
ods for solving the monotone inclusion problem which uses a relative (instead of
summable) error criterion. The proximal point method, proposed by Rockafel-
lar [27], is a classical iterative scheme for solving the latter problem. Paper [28]
introduces a Euclidean version of the HPE framework. Iteration-complexities of
the latter framework are established in [29] (see also [30]). Generalizations of the
HPE framework to the non-Euclidean setting are studied in [6, 31, 32]. However,
none of the aforementioned HPE frameworks includes the proximal ADMM con-
sidered here as a special case. Applications of the HPE framework can be found
for example in [22, 29, 30, 33–35].

Organization of the paper. Section 1.1 presents our notation and basic results.
Section 2 describes the proximal ADMM and presents 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, Section 3.1 introduces the framework and presents its convergence
rate bounds whose proofs are given in Section 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 asso-
ciated 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 $\| \cdot \|_{V,A}^*$ whose proof is omitted.

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

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

The operator $T$ is said to be monotone if

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

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

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

Recall that the $\varepsilon$-subdifferential of a convex function $f : V \to [-\infty, \infty]$ is defined by $\partial_{\varepsilon} f(v) := \{ u \in V : f(v') \geq f(v) + (u, v' - v) - \varepsilon \forall v' \in V \}$ for every $v \in V$. 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 $\partial f$ is trivially monotone if $f$ is proper. If $f$ is a proper lower semi-continuous convex function, then $\partial f$ is maximal monotone [36]. The domain of $f$ is denoted by $\text{dom } f$ and the conjugate of $f$ is the function $f^* : V \to [-\infty, \infty]$ defined as $f^*(v) = \sup_{z \in V} \{ (v, z) - f(z) \} \quad \forall v \in V$.

**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 $(0, (√5 + 1)/2)$ while the ergodic one includes also the stepsize $(√5 + 1)/2$.

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

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

has a solution $(s^*, y^*, x^*)$;
there exists \( x \in \mathcal{X} \) such that \((C^*x, D^*x) \in \text{ri} (\text{dom } f^*) \times \text{ri} (\text{dom } g^*)\).

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 stepsize \(\theta > 0\), 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

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

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

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

(2) set

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

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. [9,14]) 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 [9,15,16] 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 [3]. 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 [22, Proposition 7.2] and [6, 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 \(d_0, \tau_0\) and \(\sigma_0\).
defined as

\[
\begin{align*}
d_0 :=& d_0(\beta, \theta) = \inf_{(s, y, x) \in \mathcal{T}^{-1}(0)} \left\{ \frac{1}{2} \| s_0 - s \|^2_{S, H} + \frac{1}{2} \| y_0 - y \|^2_{Y, (G + \beta C^* C)} \\
&+ \frac{1}{2 \beta \theta} \| x_0 - x \|^2_X \right\}, \\
\sigma_\theta :=& \frac{3 \theta^2 - 7 \theta + 5 + \sqrt{(3 \theta^2 - 7 \theta + 5)^2 - 4(2 - \theta)(3 - \theta)(\theta - 1)^2}}{2(3 - \theta)}, \\
\tau_\theta :=& 4 \max \left\{ \frac{1}{\sqrt{\theta}}, \frac{\sqrt{\theta}}{2 - \theta} \right\}.
\end{align*}
\]

(8), (9), (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

\[
\tilde{x}_k = x_{k-1} - \beta (C y_{k-1} + D s_k - c).
\]

Then, for every \( k \in \mathbb{N} \),

\[
\begin{pmatrix}
\frac{H(s_{k-1} - s_k)}{1/\beta \theta (x_{k-1} - x_k)} \\
\frac{(G + \beta C^* C)(y_{k-1} - y_k)}{1/\beta \theta (x_{k-1} - x_k)}
\end{pmatrix} \in \begin{pmatrix}
\partial g(s_k) - D^* \tilde{x}_k \\
\partial f(y_k) - C^* \tilde{x}_k \\
C y_k + D s_k - c
\end{pmatrix}
\]

(12)

and there exists \( i \leq k \) such that

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

where \( d_0, \sigma_\theta \) and \( \tau_\theta \) 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 \( \theta \in (0, (\sqrt{5} + \sqrt{5}) + 1)/2 \), and let \( \tilde{x}_k \) be defined as

\[
\tilde{x}_k = x_{k-1} - \beta (C y_{k-1} + D s_k - c).
\]

Then, for every \( k \in \mathbb{N} \),

\[
\begin{pmatrix}
\frac{H(s_{k-1} - s_k)}{1/\beta \theta (x_{k-1} - x_k)} \\
\frac{(G + \beta C^* C)(y_{k-1} - y_k)}{1/\beta \theta (x_{k-1} - x_k)}
\end{pmatrix} \in \begin{pmatrix}
\partial g(s_k) - D^* \tilde{x}_k \\
\partial f(y_k) - C^* \tilde{x}_k \\
C y_k + D s_k - c
\end{pmatrix}
\]

(12)

and there exists \( i \leq k \) such that

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

where \( d_0, \sigma_\theta \) and \( \tau_\theta \) are as in (8), (9) and (10), respectively.
1)/2], and let \{\tilde{x}_k\} be as in (11). Moreover, consider the ergodic sequences \{(s_k^a, y_k^a, x_k^a, \tilde{x}_k^a)\} and \{\varepsilon_k^a\} defined by

\[
(s_k^a, y_k^a, x_k^a, \tilde{x}_k^a) := \frac{1}{k} \sum_{i=1}^{k} (s_i, y_i, x_i, \tilde{x}_i), \quad (\varepsilon_{k,s}^a, \varepsilon_{k,y}^a)
\]

\[
= \frac{1}{k} \sum_{i=1}^{k} ((r_{i,s}, s_i - s_k^a), (r_{i,y}, y_i - y_k^a)), \quad (13)
\]

where

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

Then, for every \(k \in \mathbb{N}\),

\[
\begin{pmatrix}
H(s_{k-1}^a - s_k^a) \\
(G + \beta C^* C)(y_{k-1}^a - y_k^a) \\
\frac{1}{\beta \theta} (x_{k-1}^a - x_k^a)
\end{pmatrix}
\begin{bmatrix}
\partial g_{s_{k,s}}(s_k^a) - D^* \tilde{x}_k^a \\
\partial f_{s_{k,y}}(y_k^a) - C^* \tilde{x}_k^a \\
C y_k^a + D s_k^a - c
\end{bmatrix}, \quad (15)
\]

\[
\leq 2\sqrt{\frac{2}{k}} \frac{(1 + \tau_\theta) d_0}{\theta}
\]

and

\[
\varepsilon_{k,s}^a + \varepsilon_{k,y}^a \leq \frac{3(1 + \tau_\theta)[3\theta^2 + 4\sigma_\theta(\theta^2 + \theta + 1)]d_0}{\theta^2 k} \quad (17)
\]

where \(d_0, \sigma_\theta\) and \(\tau_\theta\) 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. Section 3.1 describes the non-Euclidean HPE framework and its corresponding pointwise and ergodic convergence rate bounds. Section 3.2 gives the proofs for the two convergence rate results stated in Section 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 \( w : Z \to (-\infty, \infty] \) is called a distance-generating function if \( \text{int}(\text{dom} \ w) = \text{Dom} \ \partial w \neq \emptyset \) and \( w \) is continuously differentiable on this interior. Moreover, \( w \) induces the Bregman distance \( d_w : Z \times \text{int}(\text{dom} \ w) \to \mathbb{R} \) defined as

\[
(d_w)(z'; z) := w(z') - w(z) - \langle \nabla w(z), z' - z \rangle_Z \quad \forall (z', z) \in Z \times \text{int}(\text{dom} \ w).
\]

(18)

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

\[
(d_w)_z(z') = (d_w)(z'; z) \quad \forall (z', z) \in Z \times \text{int}(\text{dom} \ w).
\]

The following useful identities follow straightforwardly from (18):

\[
\nabla (d_w)_z(z') = -\nabla (d_w)_{z'}(z) = \nabla w(z') - \nabla w(z) \quad \forall z, z' \in \text{int}(\text{dom} \ w),
\]

\[
(d_w)_v(z') - (d_w)_v(z) = \langle \nabla (d_w)_v(z), z' - z \rangle_Z + (d_w)_z(z') \quad \forall z' \in Z,
\]

\[
\times \forall v, z \in \text{int}(\text{dom} \ w).
\]

(19)

(20)

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 \( w : Z \to [-\infty, \infty] \), seminorm \( \| \cdot \| \) in \( Z \) and convex set \( Z \subset \text{int}(\text{dom} \ w) \) be given. For given positive constants \( m \) and \( M \), \( w \) is said to be \((m, M)\)-regular with respect to \((Z, \| \cdot \|)\) if

\[
(d_w)_z(z') \geq \frac{m}{2} \| z - z' \|^2 \quad \forall z, z' \in Z,
\]

(21)

\[
\| \nabla w(z) - \nabla w(z') \|^* \leq M \| z - z' \| \quad \forall z, z' \in Z.
\]

(22)

We now make some remarks about the class of regular distance-generating functions as in Definition 3.2, which was first introduced in [6]. First, if the seminorm in Definition 3.2 is a norm, then (21) 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 [6, Example 2.3]. For the purpose of analysing the proximal ADMM, we make use of the distance-generating function given by \( w(:) = (1/2) \| \cdot \|^2_{Z,Q} \) where \( Q \) is a self-adjoint positive semidefinite linear operator. This \( w \) can be easily shown to be \((1, 1)\)-regular with respect to \((Z, \| \cdot \|_{Z,Q})\). Third, if \( w : Z \to [-\infty, \infty] \) is
(m, M)-regular with respect to \((Z, \| \cdot \|)\), then
\[
\frac{m}{2} \| z - z' \|^2 \leq (dw)z(z') \leq \frac{M}{2} \| z - z' \|^2 \quad \forall \, z, z' \in Z.
\] (23)

Throughout this section, we assume that \( w : Z \to [-\infty, \infty] \) is an \((m, M)\)-regular distance-generating function with respect to \((Z, \| \cdot \|)\) where \( Z \subset \text{int}(\text{dom} \, w) \) is a closed convex set and \( \| \cdot \| \) is a seminorm in \( Z \). Our problem of interest in this section is the monotone inclusion problem (MIP)
\[
0 \in T(z) \tag{24}
\]
where \( T : Z \rightrightarrows Z \) is a maximal monotone operator and the following conditions hold:

(B1) \( \text{Dom} \, T \subset Z \);
(B2) the solution set \( T^{-1}(0) \) of (24) is nonempty.

We now state a non-Euclidean HPE (NE-HPE) framework for solving (24).

**NE-HPE framework for solving (24).**

(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
\[
\begin{align*}
\varepsilon_k &:= \frac{1}{\lambda_k} \nabla (dw)z_k (z_{k-1}) \in T[\varepsilon_k](\tilde{z}_k), \tag{25} \\
(dw)z_k(\tilde{z}_k) + \lambda_k \varepsilon_k &+ \eta_k \leq \sigma (dw)z_{k-1}(\tilde{z}_k) + \eta_{k-1}; \tag{26}
\end{align*}
\]
(2) set \( k \leftarrow k + 1 \) and go to step 1.

We now make some remarks about the NE-HPE framework. First, [6] 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 \( \mu \)-monotone operator \( S \) with respect to \( w \). The latter notion implies strong monotonicity of \( S \) when \( dw \) is nondegenerate (see [6, Assumption (A1)]). Second, the NE-HPE does not specify how to find \( \lambda_k \) and \( (\tilde{z}_k, z_k, \varepsilon_k) \) satisfying (25) and (26). The particular scheme for computing \( \lambda_k \) 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 (26) 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 (25). Therefore, the HPE error conditions (25)–(26) can be viewed as a relaxation of an iteration of
the exact non-Euclidean proximal point method, namely,

\[ 0 \in \frac{1}{\lambda_k} \nabla (d w)_{z_{k-1}}(z_k) + T(z_k). \]

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 (25)–(26) 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 \((d w)_0\) defined as

\[ (d w)_0 = \inf \{(d w)_{z_0}(z^*) : z^* \in T^{-1}(0)\}. \tag{27} \]

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

\[
\|r_i\| \leq \frac{2M}{\sqrt{m}} \sqrt{\frac{(1 + \sigma)(d w)_0 + 2\eta_0}{1 - \sigma} \left( \frac{\lambda_i^{-1}}{\sum_{j=1}^{k} \lambda_j} \right)}
\]

\[
\leq \frac{2M}{\lambda \sqrt{mk}} \sqrt{\frac{(1 + \sigma)(d w)_0 + 2\eta_0}{1 - \sigma}}
\]

\[
\varepsilon_i \leq \frac{(1 + \sigma)(d w)_0 + 2\eta_0}{(1 - \sigma) \sum_{i=1}^{k} \lambda_i} \leq \frac{(1 + \sigma)(d w)_0 + 2\eta_0}{(1 - \sigma) \lambda_k};
\]

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

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

\[
\varepsilon_i \leq \frac{[(1 + \sigma)(d w)_0 + 2\eta_0] \lambda_i}{(1 - \sigma) \sum_{j=1}^{k} \lambda_j^2},
\]

where \((d w)_0\) is as defined in (27).
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

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

(28)

The following result provides convergence rate bounds for $\|r_k^a\|_*$ and $\varepsilon_k^a$. The pair $(r_k^a, \varepsilon_k^a)$ 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

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

where

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

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

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

$$\rho_k \leq \frac{(dw)_0 + \eta_0}{1 - \sigma}; \quad (29)$$

(b) Dom $T$ is bounded, in which case

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

where $D := \sup \{\min\{(dw)_y(y'), (dw)_{y'}(y)\} : y, y' \in \text{Dom } T\}$ is the diameter of Dom $T$ with respect to $dw$, and $(dw)_0$ is as defined in (27).

The bound on $\varepsilon_k^a$ presented in Theorem 3.4 depends on the quantity $\rho_k$ which is bounded under the assumption $\sigma < 1$ or boundedness of 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 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 \( w : Z \rightarrow [-\infty, \infty] \) 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
\[
(\|\nabla (dw)_{z'}(z)\|^*)^2 \leq \frac{2M^2}{m} \min\{(dw)_z(z'), (dw)_{z'}(z)\}; \tag{30}
\]

(b) for every \( l \geq 1 \) and \( u_0, u_1, \ldots, u_l \in Z \), we have
\[
(dw)_{u_0}(u_l) \leq \frac{lM}{m} \sum_{i=1}^{l} \min\{(dw)_{u_i-1}(u_i), (dw)_{u_i}(u_{i-1})\}. \tag{31}
\]

Proof: (a) It is easy to see that (30) immediately follows from (19), (21) and (22).

(b) It follows from the second inequality in (23) 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 (21), immediately implies (31).

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 \( z \in \text{dom } w \), we have
\[
(dw)_{z_{k-1}}(z) - (dw)_{z_k}(z) = (dw)_{z_{k-1}}(\tilde{z}_k) - (dw)_{z_k}(\tilde{z}_k) + \lambda_k \langle r_k, \tilde{z}_k - z \rangle Z;
\]

(b) for every \( z \in \text{dom } w \), we have
\[
(dw)_{z_{k-1}}(z) - (dw)_{z_k}(z) + \eta_{k-1} \geq (1 - \sigma)(dw)_{z_{k-1}}(\tilde{z}_k) + \lambda_k \langle r_k, \tilde{z}_k - z \rangle Z + \varepsilon_k + \eta_k;
\]

(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 (20) twice and using the definition of $r_k$ given by (25), we obtain

\[
(dw)_{z_{k-1}}(z) - (dw)_{z_k}(z) = (dw)_{z_{k-1}}(z_k) + \langle \nabla (dw)_{z_{k-1}}(z_k), z - z_k \rangle_Z
\]

\[
= (dw)_{z_{k-1}}(z_k) + \langle \nabla (dw)_{z_{k-1}}(z_k), \tilde{z}_k - z_k \rangle_Z
\]

\[
+ \langle \nabla (dw)_{z_{k-1}}(z_k), z - \tilde{z}_k \rangle_Z
\]

\[
= (dw)_{z_{k-1}}(\tilde{z}_k) - (dw)_{z_k}(\tilde{z}_k)
\]

\[
+ \langle \nabla (dw)_{z_{k-1}}(z_k), z - \tilde{z}_k \rangle_Z
\]

\[
= (dw)_{z_{k-1}}(\tilde{z}_k) - (dw)_{z_k}(\tilde{z}_k) + \lambda_k \langle r_k, \tilde{z}_k - z \rangle_Z.
\]

(b) This statement follows as an immediate consequence of (a) and (26).

(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 $(r_i, \varepsilon_i)$.

**Lemma 3.7:** Let $\{(r_k, \varepsilon_k, \lambda_k)\}$ and $(\eta_0, \sigma)$ 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

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

\[
\varepsilon_i \leq \frac{(1 + \sigma)(dw)_0 + 2\eta_0}{1 - \sigma} \left( \frac{\lambda_i^{t-1}}{\sum_{j=1}^k \lambda_j^t} \right)
\]

where $(dw)_0$ is as defined in (27).

**Proof:** For every $i \geq 1$, define

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

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 holds, we have, from (19) and (25), for every $i \geq 1$

\[
\lambda_i \|r_i\|_* = \|\nabla (dw)_{z_{i-1}}(\tilde{z}_i) - \nabla (dw)_{z_i}(\tilde{z}_i)\|_* \leq \|\nabla (dw)_{z_{i-1}}(\tilde{z}_i)\|_* + \|\nabla (dw)_{z_i}(\tilde{z}_i)\|_*
\]

\[
\leq \frac{\sqrt{2}M}{\sqrt{m}} \left[ (dw)_{z_{i-1}}(\tilde{z}_i)^{1/2} + (dw)_{z_i}(\tilde{z}_i)^{1/2} \right] \leq \frac{\sqrt{2}M}{\sqrt{m}} \left[ (dw)_{z_{i-1}}(\tilde{z}_i)^{1/2} + (\sigma (dw)_{z_{i-1}}(\tilde{z}_i) + \eta_{i-1} - \eta_i)^{1/2} \right],
\]

where the second and third inequalities are due to (30) and (26), respectively. Hence,

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

The previous estimative together with (26) and definition of $\theta_i$ imply that

\[
\theta_i \leq (1 + \sigma) (dw)_{z_{i-1}}(\tilde{z}_i) + (\eta_{i-1} - \eta_i), \quad \forall \ i \geq 1.
\]

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 (27). As a consequence, we obtain the conclusion of the lemma.

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

**Proof of Theorem 3.3:** The inclusion $r_k \in T^{[\epsilon_k]}(\tilde{z}_k)$ holds due to (25). Statements (a) and (b) follow directly from Lemma 3.7 with $t = 1$ and $t = 2$, respectively.

**Proof of Theorem 3.4:** The inclusion $r^a_k \in T^{[\epsilon_k]}(\tilde{z}^a_k)$ follows from the transportation formula (see [37, Theorem 2.3]). Now, let $z^* \in T^{-1}(0)$. Using (19), (25)
and (28), we easily see that

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

which, together with (30) and Lemma 3.6(d), imply that

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

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

\[ (dw)_{z_0}(z) + \eta_0 \geq \sum_{i=1}^{k} \lambda_i (\langle r_i, z_i - z \rangle) + \varepsilon_i. \]

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

\[ \max_{i=1,\ldots,k} (dw)_{z_0}(\tilde{z}_i) + \eta_0 \geq (dw)_{z_0}(\tilde{z}_k^d) + \eta_0 \geq \sum_{i=1}^{k} \lambda_i (\langle r_i, z_i - \tilde{z}_k^d \rangle) + \varepsilon_i = \Lambda_k \varepsilon_k^d, \]

where the equality is due to (28). On the other hand, (31) implies that, for every \(i \geq 1\) and \(z^* \in T^{-1}(0)\),

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

where the second inequality is due to (31) and Lemma 3.6(d), and the last inequality is due to Lemma 3.6(d). Combining the above relations with (32) and using the definitions of \(\rho_k\) and \((dw)_0\) and the fact that \(M/m \geq 1\), we conclude that the bound on \(\varepsilon_k^d\) holds.

We now establish the bounds on \(\rho_k\) under either one of the conditions (a) or (b). First, if \(\sigma < 1\), then it follows from Lemma 3.6(c,d) that
\[(1 - \sigma)(dw)_{z_{i-1}}(\tilde{z}_i) \leq (dw)_{z_{i-1}}(z^*) + \eta_{k-1} \leq (dw)_{z_0}(z^*) + \eta_0 \quad \text{for every } i \geq 1 \text{ and } z^* \in T^{-1}(0), \text{ and hence that } (29) \text{ holds. Assume now that } \text{Dom } T \text{ is bounded. Then, it follows from inequality (31) and Lemma 3.6(d) that, for every } i \geq 1 \text{ and } z^* \in T^{-1}(0),
\]
\[
(dw)_{z_{i-1}}(\tilde{z}_i) \leq \frac{2M}{m} \left[ (dw)_{z_{i-1}}(z^*) + \min\{(dw)_{\tilde{z}_i}(z^*), (dw)_{z^*}(\tilde{z}_i)\} \right]
\]
\[
\leq \frac{2M}{m} \left[ (dw)_{z_0}(z^*) + \eta_0 + D \right]
\]
which, in view of definitions of \(\rho_k\) and \((dw)_0\), proves (b). \(\blacksquare\)

### 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 byproduct, 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,

\[
0 \in H(s_k - s_{k-1}) + \left[ \partial g(s_k) - D^*\tilde{x}_k \right], \quad (33)
\]
\[
0 \in (G + \beta C^*C)(y_k - y_{k-1}) + \left[ \partial f(y_k) - C^*\tilde{x}_k \right], \quad (34)
\]
\[
0 = \frac{1}{\theta \beta} (x_k - x_{k-1}) + \left[ Cy_k + Ds_k - c \right], \quad (35)
\]
\[
\tilde{x}_k - x_{k-1} = \beta C(y_k - y_{k-1}) + \frac{x_k - x_{k-1}}{\theta}. \quad (36)
\]

**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 (33). Now, from the optimality condition of (6) and definition of \(\tilde{x}_k\) in (11), we obtain

\[
0 \in \partial f(y_k) - C^*x_{k-1} + \beta C^*(Cy_{k-1} + Ds_k - c) + G(y_k - y_{k-1}) \nonumber \\
= \partial f(y_k) - C^*[x_{k-1} + \beta(Cy_{k-1} + Ds_k - c)] \\
+ \beta C^*C(y_k - y_{k-1}) + G(y_k - y_{k-1}) \nonumber \\
= \partial f(y_k) - C^*\tilde{x}_k + \beta C^*C(y_k - y_{k-1}) + G(y_k - y_{k-1}),
\]
which proves (34). Moreover, (35) follows immediately from (7). On the other hand, it follows from definition of $x_k$ in (7) that

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

which, combined with definition of $\tilde{x}_k$ in (11), yields (36).

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 $Z$, the seminorm $\| \cdot \|$ on $Z$, the distance-generating function $w : Z \to [-\infty, \infty]$ and the convex set $Z \subset \text{int}(\text{dom} w)$. We consider $Z := \mathcal{S} \times \mathcal{Y} \times \mathcal{X}$ 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

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

$$w(z) := \frac{1}{2} \| (s, y, x) \|^2, \quad Z := Z \quad (37)$$

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

$$(dw)_z(z') = \frac{1}{2} \| s' - s \|_{\mathcal{S}, H}^2 + \frac{1}{2} \| y' - y \|_{\mathcal{Y}, (G + \beta C^* C)}^2 + \frac{1}{2 \beta \theta} \| x' - x \|_{\mathcal{X}}^2 \quad (38)$$

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

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

$$Q(s, y, x) = (Hs, (G + \beta C^* C)y, x/(\beta \theta)) \quad \forall \ (s, y, x) \in Z, \quad (39)$$

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

$$\Delta s_k = s_k - s_{k-1}, \quad \Delta y_k = y_k - y_{k-1}, \quad \Delta x_k = x_k - x_{k-1}. \quad (40)$$

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

\[
\frac{1}{\sqrt{\theta}} \left( \frac{1}{2} \| \Delta y_1 \|_{\mathcal{Y},G}^2 - \frac{1}{\sqrt{\theta}} (C \Delta y_1, \Delta x_1)_\mathcal{X} \right) \leq \tau_0 d_0,
\]

where \( d_0 \) and \( \tau_0 \) are as in (8) and (10), respectively.

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

\[
\frac{1}{\theta} (C \Delta y_k, \Delta x_k)_\mathcal{X} \geq \frac{1 - \theta}{\theta} (C \Delta y_k, \Delta x_{k-1})_\mathcal{X} + \frac{1}{2} \| \Delta y_k \|_{\mathcal{Y},G}^2
\]

\[
- \frac{1}{2} \| \Delta x_{k-1} \|_{\mathcal{Y},G}^2, \quad \forall k \geq 2.
\]

Proof: (a) Let a point \( z^* := (s^*, y^*, x^*) \) be such that \( 0 \in T(s^*, y^*, x^*) \) (see Assumption A1). Since \( \langle x, x'_\rangle_\mathcal{X} \leq (1/2)(\|x\|_\mathcal{X}^2 + \|x'_\|_\mathcal{X}^2) \) for every \( x, x' \in \mathcal{X} \), using (40) we obtain

\[
\frac{1}{2} \| \Delta y_1 \|_{\mathcal{Y},G}^2 - \frac{1}{\sqrt{\theta}} (C \Delta y_1, \Delta x_1)_\mathcal{X} \leq \frac{1}{2\beta \theta} \| \Delta x_1 \|_{\mathcal{X}}^2 + \frac{\beta}{2} \| C \Delta y_1 \|_{\mathcal{X}}^2 + \frac{1}{2} \| \Delta y_1 \|_{\mathcal{Y},G}^2
\]

\[
\leq \frac{1}{\beta \theta} \| x_1 - x^* \|_{\mathcal{X}}^2 + \beta \| C(y_1 - y^*) \|_{\mathcal{X}}^2
\]

\[
+ \| y_1 - y^* \|_{\mathcal{Y},G}^2
\]

\[
+ \frac{1}{\beta \theta} \| x_0 - x^* \|_{\mathcal{X}}^2 + \beta \| C(y_0 - y^*) \|_{\mathcal{X}}^2
\]

\[
+ \| y_0 - y^* \|_{\mathcal{Y},G}^2
\]

which, combined with (38) and simple calculus, yields

\[
\frac{1}{\sqrt{\theta}} \left( \frac{1}{2} \| \Delta y_1 \|_{\mathcal{Y},G}^2 - \frac{1}{\sqrt{\theta}} (C \Delta y_1, \Delta x_1)_\mathcal{X} \right) \leq \frac{2}{\sqrt{\theta}} \left( (dw)_{z_1}(z^*) + (dw)_{z_0}(z^*) \right).
\]

(41)

On the other hand, consider

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

(42)

Lemma 4.1 implies that inclusion (25) is satisfied for \( (z_0, z_1, \tilde{z}_1, \lambda_1, \varepsilon_1) \) with \( T \) and \( dw \) as in (4) and (38), respectively. Hence, it follows from Lemma 3.6(a) with \( z = z^*, \lambda_1 = 1 \) and the fact that \( \langle r_1, \tilde{z}_1 - z^* \rangle \geq 0 \) (because \( 0 \in T(z^*) \) and \( r_1 \in T(\tilde{z}_1) \)) that

\[
(dw)_{z_1}(z^*) \leq (dw)_{z_0}(z^*) + (dw)_{z_1}(\tilde{z}_1) - (dw)_{z_0}(\tilde{z}_1).
\]

(43)
Using the definitions in (38) and (42) and equation in (36), we obtain

\[(d \omega)_{z_1}(\tilde{z}_1) - (d \omega)_{z_0}(\tilde{z}_1) \leq \frac{1}{2\beta \theta} \|\tilde{x}_1 - x_1\|_\chi^2 - \frac{\beta}{2} \|C(y_1 - y_0)\|_\chi^2
- \frac{1}{2\beta \theta} \|\tilde{x}_1 - x_0\|_\chi^2
= \frac{(\theta - 1)}{2\beta \theta^2} \|x_1 - x_0\|_\chi^2 - \frac{1}{2} \|x_1 - x_0\|_\chi^2 \frac{1}{\theta \sqrt{\beta}}
+ \sqrt{\beta} C(y_1 - y_0) \|_\chi^2
\leq \frac{(\theta - 1)}{2\beta \theta^2} \|x_1 - x_0\|_\chi^2.\]

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

\[(d \omega)_{z_1}(\tilde{z}_1) \leq (d \omega)_{z_0}(\tilde{z}_1).\]  

(44)

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

\[(d \omega)_{z_1}(\tilde{z}_1) - (d \omega)_{z_0}(\tilde{z}_1) \leq \frac{(\theta - 1)}{2\beta \theta^2} \|x_1 - x_0\|_\chi^2
\leq \frac{2(\theta - 1)}{\theta} \left( \frac{\|x_1 - x^*\|_\chi^2}{2\beta \theta} + \frac{\|x_0 - x^*\|_\chi^2}{2\beta \theta} \right)
\leq \frac{2(\theta - 1)}{\theta} \left((d \omega)_{z_1}(z^*) + (d \omega)_{z_0}(z^*)\right),\]

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 (38) and definitions of \(z_0, z_1\) and \(z^*\). Hence, combining the last estimative with (43), we obtain

\[(d \omega)_{z_1}(z^*) \leq \frac{\theta}{2 - \theta} \left(1 + \frac{2(\theta - 1)}{\theta}\right) (d \omega)_{z_0}(z^*) = \frac{3\theta - 2}{2 - \theta} (d \omega)_{z_0}(z^*),\]

which, combined with (44), yields

\[(d \omega)_{z_1}(z^*) \leq \max \left\{1, \frac{3\theta - 2}{2 - \theta}\right\} (d \omega)_{z_0}(z^*).\]

Therefore, statement (a) follows from (41), the last inequality, definition of \(\tau_\theta\) in (10) and the fact that \(d_0\) (as defined in (8)) satisfies \(d_0 = \inf_{z \in T^{-1}(0)} (d \omega)_{z_0}(z)\).
(b) From the inclusion (34) and relation (36), 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

\[ 0 \leq \left\{ \frac{1}{\theta} C^*(x_k - x_{k-1}) - \frac{(1 - \theta)}{\theta} C^*(x_{k-1} - x_{k-2}) - G(y_k - y_{k-1}) \right. \]
\[ + G(y_{k-1} - y_{k-2}), y_k - y_{k-1} \right\}_{\mathcal{Y}}, \]

which, combined with (40), yields

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

Hence, item (b) follows from the last inequality and the fact that

\[ \langle Gy, y' \rangle_{\mathcal{Y}} \leq (1/2)(\|y\|^2_{\mathcal{Y},G} + \|y'\|^2_{\mathcal{Y},G}), \quad \forall y, y' \in \mathcal{Y}. \]

Therefore, the proof of the lemma is concluded. ■

We now present some properties of the parameter $\sigma_\theta$ defined in (9).

**Lemma 4.3:** Let $\theta \in (0, (\sqrt{5} + 1)/2]$ be given and consider the parameter $\sigma_\theta$ 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

\[ M_\theta(\sigma) := \begin{bmatrix} \sigma (1 + \theta) - 1 & (\sigma + \theta - 1)(1 - \theta) \\ (\sigma + \theta - 1)(1 - \theta) & \sigma - (1 - \theta)^2 \end{bmatrix}; \quad (45) \]

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

(c) the matrix $M_\theta(\sigma)$ in (45) 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 $\det(M_\theta(\sigma)) \leq 0$ for $\sigma$ 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 $1/3 \leq 1/(1 + \theta)$.

(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 $d_w$ as in (4) and (38), 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

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

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

$$\eta_0 = \tau_0 d_0, \quad \eta_k = [\sigma_0 - (\theta - 1)^2] \lambda_k = 1, \quad \varepsilon_k = 0 \quad \forall \; k \geq 1,$$

where $d_0$, $\sigma_0$, $\tau_0$ and $(\Delta x_k, \Delta y_k)$ are as in (8), (9), (10) and (40), 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 $\sigma = \sigma_0$.

**Proof:** The inclusion (25) follows from (33)–(35), (46) and definitions of $T$ and $d_w$. Now it remains to show that the error condition (26) holds. First of all, it follows from (36), (38), (40) and (46) that

$$(d_w)_k(\tilde{z}_k) + \lambda_k \varepsilon_k = \frac{1}{2\beta \theta} \lambda_k \varepsilon_k = \frac{1}{2\beta \theta} \lambda_k \varepsilon_k$$

$$= \frac{\beta}{2\theta} \lambda_k \varepsilon_k = \frac{\beta}{2\theta} \lambda_k \varepsilon_k$$

and

$$(d_w)_{k-1}(\tilde{z}_k) = \frac{1}{2} \lambda_k \varepsilon_k = \frac{1}{2} \lambda_k \varepsilon_k$$

Also, (38), (40) and (46) imply that

$$(d_w)_{k-1}(\tilde{z}_k) = \frac{1}{2} \lambda_k \varepsilon_k = \frac{1}{2} \lambda_k \varepsilon_k$$

It follows from (36) and (40) that

$$\|x_{k-1} - \tilde{x}_k\|^2 = \beta C \lambda_k \varepsilon_k + \frac{\beta}{\theta} \lambda_k \varepsilon_k$$

which, combined with (49), yields

$$(d_w)_{k-1}(\tilde{z}_k) = \frac{\lambda_k \varepsilon_k}{2} = \frac{\lambda_k \varepsilon_k}{2}$$

Finally, the next result shows that...
Therefore, combining (48) and (50), we see, after simple algebraic manipulations, that the error condition (26) is satisfied if and only if

\[
[\sigma(1 + \theta) - 1] \frac{\beta}{2\theta} \| C\Delta y_k \|_\chi^2 + [\sigma - (\theta - 1)^2] \frac{\| \Delta x_k \|_\chi^2}{2\beta\theta^3} + \sigma \frac{\| \Delta y_k \|_{2,G}^2}{2} \\
+ \frac{(\sigma + \theta - 1)}{\theta^2} (C\Delta y_k, \Delta x_k)_\chi \\
\geq \eta_k - \eta_{k-1} - \sigma \frac{\| \Delta s_k \|_{2,H}^2}{2}.
\]  

(51)

We now show that inequality (51) with \( \sigma = \sigma_\theta \) holds for \( k = 1 \). Indeed, it follows from definition of \( \eta_1 \) and \( \sigma_\theta \geq 1/(1 + \theta) \) (see Lemma 4.3(b)) that

\[
[\sigma_\theta(1 + \theta) - 1] \frac{\beta}{2\theta} \| C\Delta y_1 \|_\chi^2 + [\sigma_\theta - (1 - \theta)^2] \frac{\| \Delta x_1 \|_\chi^2}{2\beta\theta^3} + \sigma_\theta \frac{\| \Delta y_1 \|_{2,G}^2}{2} \\
+ \frac{(\sigma_\theta + \theta - 1)}{\theta^2} (C\Delta y_1, \Delta x_1)_\chi \\
\geq \left[ \sigma_\theta - \frac{\sigma_\theta + \theta - 1}{\theta} + \frac{\sigma_\theta + \theta - 1}{\theta^3/2} \right] \frac{\| \Delta y_1 \|_{2,G}^2}{2} + \eta_1 \\
+ \frac{(\sigma_\theta + \theta - 1)}{\theta^3/2} \left( \frac{1}{\sqrt{\theta}} (C\Delta y_1, \Delta x_1)_\chi - \frac{1}{2} \| \Delta y_1 \|_{2,G}^2 \right) \\
\geq \left[ \sigma_\theta - \frac{\sigma_\theta + \theta - 1}{3\theta} \right] \frac{\| \Delta y_1 \|_{2,G}^2}{2} + \eta_1 - \frac{(\sigma_\theta + \theta - 1)}{\theta} \tau_0 d_0 \\
\geq \eta_1 - \eta_0,
\]

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 \( \eta_0 \). Therefore, inequality (51) holds with \( k = 1 \) and \( \sigma = \sigma_\theta \).

We next show that inequality (51) with \( \sigma = \sigma_\theta \) holds for \( k \geq 2 \). Using Lemma 4.2(b) and the definition of \( \{ \eta_k \} \) in (47), we see, after simple calculus, that a sufficient condition for (51) to hold with \( \sigma = \sigma_\theta \) and \( k \geq 2 \) is that

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

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 (37) is a $(1, 1)$-regular distance-generating function and $\lambda := \inf \lambda_k = 1$, we obtain by combining Theorems 4.4 and 3.3(a) that inclusion (12) holds and there exists $i \leq k$ such that

\[
\left( \|s_{i-1} - s_i\|_{\Delta^\theta}^2 + \|y_{i-1} - y_i\|_{\Delta^\theta_{G+\beta C^* C}}^2 + \frac{1}{\beta \theta} \|x_{i-1} - x_i\|_{\Delta^\theta}^2 \right)^{1/2} \leq \frac{2}{\sqrt{k}} \sqrt{\frac{(1 + \sigma_\theta)d_0 + 2\eta_0}{1 - \sigma_\theta}},
\]

where we also used the definition of the norm $\| \cdot \|$ in (37) and Proposition 1.1. The result now follows from the last inequality and definition of $\eta_0$ in (47).

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 (46) 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 (38), $d_0$ and $\tau_\theta$ are as in (8) and (10), respectively.

Proof: It follows from (38) and (46) that

\[
(dw)_{z_{k-1}}(\tilde{z}_k) = \frac{1}{2} \|\Delta s_k\|_{\Delta^\theta}^2 + \frac{1}{2} \|\Delta y_k\|_{\Delta^\theta_{G}}^2 + \frac{\beta}{2} \|C\Delta y_k\|_{\Delta^\theta}^2 + \frac{1}{2\beta \theta} \|x_{k-1} - \tilde{x}_k\|_{\Delta^\theta}^2.
\]

On the other hand, using (36) we have

\[
\frac{1}{2\beta \theta} \|x_{k-1} - \tilde{x}_k\|_{\Delta^\theta}^2 = \frac{1}{2\beta \theta} \|\beta C\Delta y_k + \frac{\Delta x_k}{\theta} \|_{\Delta^\theta}^2 = \frac{\beta}{2\theta} \|\Delta y_k\|_{\Delta^\theta}^2 + \frac{1}{\beta \theta^2} \langle \beta C\Delta y_k, \Delta x_k \rangle_{\Delta^\theta} + \frac{1}{2\beta \theta^3} \|\Delta x_k\|_{\Delta^\theta}^2 \leq \frac{\beta(\theta + 1)}{2\theta^2} \|\Delta y_k\|_{\Delta^\theta}^2 + \frac{\theta + 1}{2\beta \theta^3} \|\Delta x_k\|_{\Delta^\theta}^2.
\]
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 (52), we have

$$
(dw)_{z_{k-1}}(\tilde{z}_k) \leq \frac{(\theta^2 + \theta + 1)}{\theta^2} \times \left[ \frac{1}{2} \|s_k\|^2_{S:H} + \frac{1}{2} \|\Delta y_k\|^2_{\gamma,G} + \frac{\beta}{2} \|C\Delta y_k\|^2_{\chi} + \frac{1}{2\beta \theta} \|\Delta x_k\|^2_{\chi} \right]
$$

$$
\leq \frac{2(\theta^2 + \theta + 1)}{\theta^2} \left[ (dw)_{z_{k-1}}(z^*) + (dw)_{z_k}(z^*) \right],
$$

where the last inequality is due to (38) 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

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

which, combined with the definition of $\eta_0$ in (47) 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), (46) and (47), respectively. Hence, since $w$ as defined in (37) is a $(1, 1)$-regular distance-generating function and $1 = \lambda = \inf \lambda_k$, we obtain from Proposition 1.1 and Theorem 3.4, and Lemma 4.5 that, for every $k \geq 1$,

$$
\left( \|s_k^a - s_k^a\|^2_{S:H} + \|y_k^a - y_k^a\|^2_{\gamma(G + \beta C^*C)} + \frac{1}{\beta \theta} \|x_k^a - x_k^a\|^2_{\chi} \right)^{1/2}
\leq \frac{2\sqrt{2(d_0 + \eta_0)}}{k}
$$

(53)

and

$$
\varepsilon_k^a = \frac{1}{k} \sum_{i=1}^k \left( (r_{i,s}^2, s_i - s_i^a) + (r_{i,y}^2, y_i - y_i^a) + \langle r_{i,x}, \tilde{x}_i - \tilde{x}_k^a \rangle \right)
\leq \frac{3[3\theta^2(d_0 + \eta_0) + 4\sigma \theta (1 + \tau \theta) (\theta^2 + \theta + 1)d_0]}{\theta^2 k},
$$

(54)

where $(r_{i,s}, r_{i,y}, r_{i,x}) = (H(s_{i-1} - s_i), (G + \beta C^*C)(y_{i-1} - y_i), (x_{i-1} - x_i))/\beta \theta)$. Moreover, (13) and (35) yield

$$
Ds_k + Cy_k = r_{k,x} + c,
$$

$$
r_k^a := \frac{1}{k} \sum_{i=1}^k r_{i,x} = Ds_k^a + Cy_k^a - c.
$$
Additionally, (13) and some algebraic manipulations imply that
\[
\sum_{i=1}^{k} \langle r_{i,x}, x_i - x_k^a \rangle = \sum_{i=1}^{k} \langle r_{i,x} - r_{k,x}^a, x_i - x_k^a \rangle = \sum_{i=1}^{k} \langle r_{i,x} - r_{k,x}^a, x_i \rangle = \sum_{i=1}^{k} \langle Ds_i - Ds_k^a + Cy_i - Cy_k^a, x_i \rangle.
\]

Hence, the inequalities in (16) and (17) now follow from (53) and (54), and simple calculus.

To finish the proof of the theorem, note that direct use of the transportation formula (see [37, Theorem 2.3]) and (33)–(35) give \( \varepsilon_{k,x}, \varepsilon_{k,y} \geq 0 \) and (15).

\[\square\]

5. Concluding remark

This paper developed alternative pointwise and ergodic iteration-complexity results for the proximal ADMM for solving convex linearly constrained optimization problems. A pointwise iteration-complexity was established for any stepsize \( \theta \in (0, (1 + \sqrt{5})/2) \) and an ergodic one was obtained for any \( \theta \in (0, (1 + \sqrt{5})/2) \). Hence, our analysis of the ergodic case included the stepsize \( \theta = (1 + \sqrt{5})/2 \) which, as far as we know, has not been established yet.

Regarding the asymptotic convergence in the ergodic case of the proximal ADMM with \( \theta = (\sqrt{5} + 1)/2 \), if the linear operator \( Q \) as in (39) is positive definite, then the ergodic sequence \( \{ (s_k^a, y_k^a, x_k^a) \} \) (see (13)) is bounded and its accumulation points are solutions of the Lagrangian system (4). Indeed, it follows from Theorem 4.4 that the proximal ADMM can be seen as an instance of the NE-HPE framework whose nondegenerate Bregman distance \( d_w \) is as in (38). Hence, Lemma 3.6(d) with \( \{ z_k \} \) as in (46) implies that \( \{ (s_k, y_k, x_k) \} \) is bounded which in turn implies that \( \{ z_k^a := (s_k^a, y_k^a, x_k^a) \} \) is bounded. Thus, from Lemma 4.5, we also obtain the boundedness of the sequence \( \{ z_k^a := (s_k^a, y_k^a, x_k^a) \} \). Hence, it follows from Theorem 2.2 that any accumulation point of \( \{ (s_k^a, y_k^a, x_k^a) \} \) is a solution of (4). Now, if \( \theta \in (0, (\sqrt{5} + 1)/2) \) and \( Q \) is positive definite, then it can be shown, by using a similar argument, that \( \{ (s_k, y_k, x_k) \} \) and \( \{ (s_k^a, y_k^a, x_k^a) \} \) converge to a solution of (4).

Acknowledgments

The authors would like to thank an anonymous reviewer and the associate editor for their insightful comments on earlier drafts of this paper. We also thank the reviewer for bringing [26] to our attention.

Disclosure statement

No potential conflict of interest was reported by the authors.
Funding

The work of these authors was supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) Grants 302666/2017-6 and 406975/2016-7. The work of this author was partially supported by NSF Grant CMMI-1300221.

ORCID

Max L. N. Gonçalves http://orcid.org/0000-0001-9563-1101

References