

# The proximal alternating iterative hard thresholding method for $l_0$ minimization, with complexity $\mathcal{O}(1/\sqrt{k})^{\star}$



Fan Yang<sup>a</sup>, Yi Shen<sup>a,\*</sup>, Zhi Song Liu<sup>b</sup>

<sup>a</sup> Department of Mathematics, Zhejiang Sci-Tech University, Hangzhou 310028, China

<sup>b</sup> School of Mathematics, Physics and Information Science, Zhejiang Ocean University, Zhoushan, 316004, China

## ARTICLE INFO

### Article history:

Received 6 August 2015

Received in revised form 8 July 2016

### Keywords:

Sparse approximation

Alternating minimization

Hard thresholding

Tight wavelet frame

Kurdyka–Łojasiewicz property

## ABSTRACT

Since digital images are usually sparse in the wavelet frame domain, some nonconvex minimization models based on wavelet frame have been proposed and sparse approximations have been widely used in image restoration in recent years. Among them, the proximal alternating iterative hard thresholding method is proposed in this paper to solve the nonconvex model based on wavelet frame. Through combining the proposed algorithm with the iterative hard thresholding algorithm which is well studied in compressed sensing theory, this paper proves that the complexity of the proposed method is  $\mathcal{O}(1/\sqrt{k})$ . On the other hand, a more general nonconvex–nonsmooth model is adopted and the pseudo proximal alternating linearized minimization method is developed to solve the above problem. With the Kurdyka–Łojasiewicz (KL) property, it is proved that the sequence generated by the proposed algorithm converges to some critical points of the corresponding model. Finally, the proposed method is applied to restore the blurred noisy gray images. As the numerical results reveal, the performance of the proposed method is comparable or better than some well-known convex image restoration methods.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

This paper will study how to restore the digital images corrupted by the given blur operators and white Gaussian noise. Mathematically speaking, if the columns of a digital image are stacked one by one, the digital image can be regarded as a column vector  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$ . The generic image restoration problem can be formulated as a linear inverse problem accordingly. Namely, an observed degraded digital image  $\mathbf{b}$  in  $\mathbb{R}^m$  is given by

$$\mathbf{b} = \mathcal{A}\hat{\mathbf{x}} + \mathbf{w},$$

where  $\mathcal{A} \in \mathbb{R}^{m \times n}$  is a known linear blur operator;  $\mathbf{w}$  denotes white Gaussian noise with variance  $\sigma^2$  and  $\hat{\mathbf{x}}$  is the unknown true image. For the given vector  $\mathbf{x}$ ,  $\|\mathbf{x}\|_0$  denotes the number of nonzero entries. Although  $\|\cdot\|_0$  is not a norm, it is still called

<sup>☆</sup> This work is supported by the Zhejiang Provincial Natural Science Foundation of China under Grant No. LY15A010020, the Key project of NSF of China under number 11531013, the Science Foundation of Zhejiang Sci-Tech University under Grant 1113834–Y, the Key Laboratory of Oceanographic Big Data Mining & Application of Zhejiang Province under Grants No. OBDMA201505 and the Open Foundation from Marine Sciences in the Most Important Subjects of Zhejiang under Grant No. 111040602136.

\* Corresponding author.

E-mail addresses: [kiraofsin@yahoo.co.jp](mailto:kiraofsin@yahoo.co.jp) (F. Yang), [yshen@zstu.edu.cn](mailto:yshen@zstu.edu.cn) (Y. Shen), [liuzhisong@zjou.edu.cn](mailto:liuzhisong@zjou.edu.cn) (Z.S. Liu).

as  $l_0$ -norm for convenience. As usual,  $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$  denotes the  $l_1$ -norm and  $\|\mathbf{x}\| = \sqrt{\sum_{i=1}^n x_i^2}$  denotes the  $l_2$ -norm. This paper considers recovering the unknown image by solving the following  $l_0$  minimization model:

$$\min_{\mathbf{x}, \mathbf{y}} \Psi_0(\mathbf{x}, \mathbf{y}) := \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \lambda \|\mathbf{y}\|_0 + \frac{\eta}{2} \|\mathcal{D}^\top \mathbf{x} - \mathbf{y}\|^2, \quad (1.1)$$

where  $\mathcal{D}$  is a given matrix and  $\mathcal{D}^\top$  denotes its transpose, while  $\mathcal{I} \in \mathbb{R}^{n \times n}$  is assumed to be the identity matrix. This paper mainly regards  $\mathcal{D} \in \mathbb{R}^{n \times d}$  as a tight frame, which means  $\mathcal{D}\mathcal{D}^\top = \mathcal{I}$ . Usually, the number of the columns  $d$  is larger than that of the rows  $n$ , indicating  $\mathcal{D}$  is a redundant system in  $\mathbb{R}^n$ . Natural images are often sparse with respect to tight wavelet frames. Hence, the regularization term used for models based on wavelet frame can be the  $l_0$ -norm or  $l_1$ -norm of the wavelet frame coefficients. Other nonconvex image restoration models based on wavelet frame can be found in [1–4] and referred therein. The minimization model (1.1) is solved in [4,5] via the block coordinate descent (BCD) method. Roughly speaking, the first step of the BCD method is deconvoluting in the pixel domain and the second step is denosing in the frame domain. It is proved in [5] that the sequence generated by the BCD method is bounded, which motivates the authors to propose a new method to generate a convergent sequence as below,

$$\begin{cases} \mathbf{x}^{k+1} = \operatorname{argmin}_{\mathbf{x}} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \frac{\eta}{2} \|\mathcal{D}^\top \mathbf{x} - \mathbf{y}^k\|^2, \\ \mathbf{y}^{k+1} \in \operatorname{argmin}_{\mathbf{y}} \lambda \|\mathbf{y}\|_0 + \frac{\eta}{2} \|\mathbf{y} - \mathcal{D}^\top \mathbf{x}^{k+1}\|^2 + \frac{d_k}{2} \|\mathbf{y} - \mathbf{y}^k\|^2, \end{cases} \quad k = 0, 1, 2, \dots, \quad (1.2)$$

where  $d_k$  refer to positive real numbers. The initial value  $\mathbf{y}^0$  is set 0. If all  $d_k$  are set 0, the iterative scheme (1.2) will turn to be the BCD method. The second step of the iterative scheme (1.2) can be connected with the iterative hard thresholding method in [6]. Therefore, the proposed method in this study is called as *Proximal Alternating Iterative Hard Thresholding* (PAIHT) method. To study the convergence properties of the PAIHT method, more general nonconvex–nonsmooth problems of the form will be considered

$$\min_{\mathbf{x}, \mathbf{y}} \Psi(\mathbf{x}, \mathbf{y}) := f(\mathbf{x}) + g(\mathbf{y}) + H(\mathbf{x}, \mathbf{y}). \quad (1.3)$$

If

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2, \quad g(\mathbf{y}) = \lambda \|\mathbf{y}\|_0 \quad \text{and} \quad H(\mathbf{x}, \mathbf{y}) = \frac{\eta}{2} \|\mathcal{D}^\top \mathbf{x} - \mathbf{y}\|^2,$$

then it is easy to observe that the minimization model (1.1) is a special case of the minimization model (1.3). If the functions  $f(\mathbf{x})$  and  $g(\mathbf{y})$  are both semi-continuous, and the function  $H(\mathbf{x}, \mathbf{y})$  is smooth, then the sequence generated by the proximal alternating linearized minimization algorithm [7] is convergent with the Kurdyka–Łojasiewicz (KL) property. Motivated by these works on the KL property, and based on the image restoration model (1.1), this paper explores the nonconvex–nonsmooth minimization problem (1.3) under different situations. For instance, the function  $f(\mathbf{x})$  is convex; the function  $g(\mathbf{y})$  is semi-continuous and the function  $H(\mathbf{x}, \mathbf{y})$  is strongly convex. Subsequently, a new iterative scheme is established for solving the problem (1.3) approximately under the new settings. More specifically, the iteration scheme for solving (1.3) is given by

$$\begin{cases} \mathbf{x}^{k+1} = \operatorname{argmin}_{\mathbf{x}} H(\mathbf{x}, \mathbf{y}^k) + f(\mathbf{x}), \\ \mathbf{y}^{k+1} \in \operatorname{argmin}_{\mathbf{y}} \langle \nabla_{\mathbf{y}} H(\mathbf{x}^{k+1}, \mathbf{y}^k), \mathbf{y} - \mathbf{y}^k \rangle + g(\mathbf{y}) + \frac{d_k}{2} \|\mathbf{y} - \mathbf{y}^k\|^2, \end{cases} \quad (1.4)$$

where  $d_k$  refer to positive real numbers. Since the direct convex method is applied to solve the first subproblem of the iterative scheme (1.4), it is called the *Pseudo Proximal Alternating Linearized Minimization* (PPALM) method.

The rest of this paper is organized as follows. In Section 2, some notations and propositions are introduced and some technical results on nonconvex programming are also included. In Section 3.1, the PAIHT method is proved to be sublinearly convergent. In Section 3.2, the PAIHT method is connected with the PPALM method. Then the convergence analysis of the PPALM method is applied to discuss the convergence properties of the PAIHT method. In Section 4, the comparison results of the proposed iterative algorithms with the BCD method in [5] are reported, and then the image restoration model (1.1) is compared with two convex wavelet frame image restoration models. Some conclusions are given in Section 5.

## 2. Some preliminaries

This section reviews some notations and propositions. The support of a given vector  $\mathbf{x} = (x_1, \dots, x_n)^\top$  is defined by

$$S(\mathbf{x}) := \{i : x_i \neq 0\}.$$

Let  $\sigma : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper and lower semicontinuous function, then the domain of  $\sigma$  is defined by

$$\operatorname{dom} \sigma := \{\mathbf{x} \in \mathbb{R}^n : \sigma(\mathbf{x}) < +\infty\}.$$

For any  $\mathbf{x} \in \text{dom } \sigma$ , the Fréchet subdifferential  $\hat{\partial}\sigma(\mathbf{x})$  is defined by

$$\hat{\partial}\sigma(\mathbf{x}) = \left\{ \mathbf{v} : \liminf_{\mathbf{y} \neq \mathbf{x}, \mathbf{y} \rightarrow \mathbf{x}} \frac{\sigma(\mathbf{y}) - \sigma(\mathbf{x}) - \langle \mathbf{v}, \mathbf{y} - \mathbf{x} \rangle}{\|\mathbf{y} - \mathbf{x}\|} \geq 0 \right\}$$

and  $\hat{\partial}\sigma(\mathbf{x}) = \emptyset$  if  $\mathbf{x} \notin \text{dom } \sigma$ . The vector  $\mathbf{x} \in \text{dom } \sigma$  is called the critical point if it satisfies  $0 \in \partial\sigma(\mathbf{x})$ , where  $\partial\sigma(\mathbf{x})$  is given by

$$\partial\sigma(\mathbf{x}) := \{ \mathbf{v} : \exists \mathbf{x}^k \rightarrow \mathbf{x}, \sigma(\mathbf{x}^k) \rightarrow \sigma(\mathbf{x}) \text{ and } \mathbf{v}^k \in \hat{\partial}\sigma(\mathbf{x}^k) \rightarrow \mathbf{v} \}.$$

Assume that the coupling function  $H$  in problem (1.3) is continuously differentiable, it then follows from [8, Proposition 10.5] that

$$\partial\Psi(\mathbf{x}, \mathbf{y}) = (\nabla_{\mathbf{x}}H(\mathbf{x}, \mathbf{y}) + \partial f(\mathbf{x}), \nabla_{\mathbf{y}}H(\mathbf{x}, \mathbf{y}) + \partial g(\mathbf{y})) = (\partial_{\mathbf{x}}\Psi(\mathbf{x}, \mathbf{y}), \partial_{\mathbf{y}}\Psi(\mathbf{x}, \mathbf{y})). \quad (2.1)$$

For a proper and lower semicontinuous function  $\sigma : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  and parameter value  $t > 0$ , the proximal mapping is defined by

$$\text{prox}_t^\sigma(\mathbf{x}) := \underset{\mathbf{u}}{\text{argmin}} \left\{ \sigma(\mathbf{u}) + \frac{t}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\}.$$

A mapping  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is said to be  $L$ -Lipschitz continuous on the set  $X \subseteq \mathbb{R}^n$  if there exists  $L > 0$  such that

$$\|f(\mathbf{x}) - f(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\|, \quad \text{for all } \mathbf{x}, \mathbf{y} \in X.$$

A sufficient decrease property of the objective function after a proximal map step for the nonconvex setting is given in [7] as below.

**Lemma 2.1.** Let  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  be a continuously differentiable function with gradient  $\nabla h$  assumed  $L_h$ -Lipschitz continuous. Let  $\sigma : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper and lower semicontinuous function with  $\inf_{\mathbb{R}^n} \sigma > -\infty$ . Fix any  $t > L_h$ . Then for any  $\mathbf{u} \in \text{dom } \sigma$  and any  $\bar{\mathbf{u}} \in \mathbb{R}^n$  defined by

$$\bar{\mathbf{u}} \in \text{prox}_t^\sigma \left( \mathbf{u} - \frac{1}{t} \nabla h(\mathbf{u}) \right),$$

the following inequality

$$h(\bar{\mathbf{u}}) + \sigma(\bar{\mathbf{u}}) \leq h(\mathbf{u}) + \sigma(\mathbf{u}) - \frac{t - L_h}{2} \|\bar{\mathbf{u}} - \mathbf{u}\|^2$$

holds.

The following lemma on the KL inequality is the key to the convergence analysis in Section 3.2.

**Lemma 2.2** (KL Property). Let  $\sigma : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper and lower semicontinuous function. Then it is said to have the KL property at  $\bar{\mathbf{x}} \in \text{dom } \sigma$  if there exist  $\delta \in (0, +\infty]$ , a neighborhood  $U$  of  $\bar{\mathbf{x}}$  and a continuous concave function  $\varphi : [0, \delta) \rightarrow \mathbb{R}^+$  such that:

1.  $\varphi(0) = 0$ ,
2.  $\varphi$  is  $C^1$  on  $(0, \delta)$ ,
3. for any  $s \in (0, \delta)$ ,  $\varphi'(s) > 0$ ,
4. for any  $\mathbf{x} \in U \cap [\sigma(\bar{\mathbf{x}}) < \sigma(\mathbf{x}) < \sigma(\bar{\mathbf{x}}) + \delta]$ , the KL inequality

$$\varphi'(\sigma(\mathbf{x}) - \sigma(\bar{\mathbf{x}})) \text{dist}(\mathbf{x}, \partial\sigma(\mathbf{x})) \geq 1 \quad (2.2)$$

holds. Furthermore, if  $U$  is a compact set,  $\sigma$  is constant on  $U$  and satisfies the KL property at each point of  $U$ , then (2.2) holds for all  $\mathbf{x} \in \{\text{dist}(\mathbf{x}, U) < \epsilon\} \cap [\sigma(\bar{\mathbf{x}}) < \sigma(\mathbf{x}) < \sigma(\bar{\mathbf{x}}) + \delta]$ .

Łojasiewicz proved that the real-analytic functions satisfied the above inequality (2.2) with  $\varphi(s) = s^{1-\theta}$  for  $\theta \in [\frac{1}{2}, 1)$  in [9]. Kurdyka extended the KL property to functions on the o-minimal structure in [10]. Bolte et al. [11,12] extended the KL inequality for nonsmooth subanalytic functions. Bolte in [7] developed it into nonconvex–nonsmooth setting. More discussions on the KL property can be found in [2,7,9–15]. Now the following section will verify that the cost function  $\Psi_0(\mathbf{x}, \mathbf{y})$  given in (1.1) satisfies the KL property.

Since the first term  $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$  and third term  $H(\mathbf{x}, \mathbf{y}) = \frac{\eta}{2} \|\mathcal{D}^\top \mathbf{x} - \mathbf{y}\|^2$  are polynomial functions, they are semi-algebraic functions. It is proved in [7] that the second term  $g(\mathbf{y}) = \lambda \|\mathbf{y}\|_0$  is a semi-algebraic set. Therefore, the cost function  $\Psi_0(\mathbf{x}, \mathbf{y})$  is semi-algebraic which satisfies the KL property.

### 3. Convergence analysis

This section mainly establishes the convergence properties of the PAIHT method. Some general nonconvex optimization algorithms are also studied.

#### 3.1. Convergence analysis of the PAIHT method

For the first subproblem of (1.2), the first order optimality condition gives the fact

$$\mathbf{x}^{k+1} = (\mathcal{A}^\top \mathcal{A} + \eta \mathcal{I})^{-1} (\mathcal{A}^\top \mathbf{b} + \eta \mathcal{D} \mathbf{y}^k). \quad (3.1)$$

If  $\mathcal{A}$  is the convolution operator and the circular boundary condition is used, then the right hand side of (3.1) can be solved by the fast Fourier transform. The second subproblem of (1.2) can be solved by the componentwise hard thresholding operator:

$$[\mathcal{H}_\lambda(\mathbf{c})]_i = \begin{cases} 0, & \text{if } |c_i| < \sqrt{\lambda} \\ \{0, \sqrt{\lambda}\}, & \text{if } |c_i| = \sqrt{\lambda} \\ c_i, & \text{if } |c_i| > \sqrt{\lambda}, \end{cases}$$

where  $\lambda \in \mathbb{R}^+$  is the thresholding value. In summary, the closed form solutions to problem (1.2) are given by

$$\begin{cases} \mathbf{x}^{k+1} = (\mathcal{A}^\top \mathcal{A} + \eta \mathcal{I})^{-1} (\mathcal{A}^\top \mathbf{b} + \eta \mathcal{D} \mathbf{y}^k), \\ \mathbf{y}^{k+1} \in \mathcal{H}_{\lambda_k} \left( \mathbf{y}^k - \frac{\eta}{\eta + d_k} (\mathbf{y}^k - \mathcal{D}^\top \mathbf{x}^{k+1}) \right), \end{cases} \quad k = 0, 1, 2, \dots, \quad (3.2)$$

where  $\lambda_k := \frac{2\lambda}{\eta + d_k}$ . The convergence analysis of the iterative scheme (3.2) is provided by the following.

**Theorem 3.1.** Assume that  $\mathcal{A}^\top \mathcal{A} > 0$ . Then the sequence  $\{(\mathbf{x}^k, \mathbf{y}^k)\}$  generated by the iterative scheme (3.2) converges and there exists a positive integer  $k_0$  such that  $S(\mathbf{y}^k) = S(\mathbf{y}^{k_0})$  for all  $k \geq k_0$ . Let  $(\mathbf{x}^*, \mathbf{y}^*)$  denote the limit of the sequence  $\{(\mathbf{x}^k, \mathbf{y}^k)\}$ . Then,

$$\|\mathbf{y}^k - \mathbf{y}^*\| \leq \mathcal{O}(1/\sqrt{k}), \quad \|\mathbf{x}^k - \mathbf{x}^*\| \leq \mathcal{O}(1/\sqrt{k}), \quad k \rightarrow \infty.$$

**Proof.** Substituting (3.1) into the right hand side of the second subproblem in (1.2),

$$\begin{aligned} & \lambda \|\mathbf{y}\|_0 + \frac{\eta}{2} \|\mathbf{y} - \mathbf{y}^k + \mathbf{y}^k - \mathcal{D}^\top (\mathcal{A}^\top \mathcal{A} + \eta \mathcal{I})^{-1} (\mathcal{A}^\top \mathbf{b} + \eta \mathcal{D} \mathbf{y}^k)\|^2 + \frac{d_k}{2} \|\mathbf{y} - \mathbf{y}^k\|^2 \\ &= \lambda \|\mathbf{y}\|_0 + \frac{\eta}{2} \|\mathbf{y} - \mathbf{y}^k + (\mathcal{I} - \eta \mathcal{D}^\top (\mathcal{A}^\top \mathcal{A} + \eta \mathcal{I})^{-1} \mathcal{D}) \mathbf{y}^k - \mathcal{D}^\top (\mathcal{A}^\top \mathcal{A} + \eta \mathcal{I})^{-1} \mathcal{A}^\top \mathbf{b}\|^2 + \frac{d_k}{2} \|\mathbf{y} - \mathbf{y}^k\|^2. \end{aligned}$$

Since  $\mathcal{D} \mathcal{D}^\top = \mathcal{I}$ , the nonzero eigenvalues of

$$\mathcal{D}^\top (\mathcal{A}^\top \mathcal{A} + \eta \mathcal{I})^{-1} \mathcal{D}$$

coincide with the eigenvalues of the matrix  $(\mathcal{A}^\top \mathcal{A} + \eta \mathcal{I})^{-1}$ . Moreover, the condition  $\mathcal{A}^\top \mathcal{A} > 0$  implies that

$$0 < \mathcal{I} - \eta \mathcal{D}^\top (\mathcal{A}^\top \mathcal{A} + \eta \mathcal{I})^{-1} \mathcal{D} \leq \mathcal{I}.$$

Therefore, the matrix  $\mathcal{I} - \eta \mathcal{D}^\top (\mathcal{A}^\top \mathcal{A} + \eta \mathcal{I})^{-1} \mathcal{D}$  is symmetric positive definite. By Cholesky factorization, see e.g. [16, Theorem 4.2.7], there exists a unique lower triangular matrix  $\mathcal{L}$  such that

$$\mathcal{L} \mathcal{L}^\top = \mathcal{I} - \eta \mathcal{D}^\top (\mathcal{A}^\top \mathcal{A} + \eta \mathcal{I})^{-1} \mathcal{D}.$$

We define

$$G(\mathbf{y}) := \frac{\eta}{2} \|\mathcal{L}^\top \mathbf{y} - \mathcal{L}^{-1} \mathbf{f}\|^2, \quad \text{where } \mathbf{f} = \mathcal{D}^\top (\mathcal{A}^\top \mathcal{A} + \eta \mathcal{I})^{-1} \mathcal{A}^\top \mathbf{b}.$$

Then

$$\nabla G(\mathbf{y}^k) = \eta (\mathcal{L} \mathcal{L}^\top \mathbf{y}^k - \mathbf{f}).$$

Hence, the second subproblem in (1.2) is equal to

$$\mathbf{y}^{k+1} \in \underset{\mathbf{y}}{\operatorname{argmin}} \lambda \|\mathbf{y}\|_0 + \frac{\eta}{2} \left\| \mathbf{y} - \mathbf{y}^k + \frac{1}{\eta} \nabla G(\mathbf{y}^k) \right\|^2 + \frac{d_k}{2} \|\mathbf{y} - \mathbf{y}^k\|^2. \quad (3.3)$$

Let  $\rho_{\max}(\mathcal{L} \mathcal{L}^\top)$  denote the largest eigenvalue of  $\mathcal{L} \mathcal{L}^\top$ . Then  $\nabla G$  is Lipschitz continuous with constant  $\eta \rho_{\max}(\mathcal{L} \mathcal{L}^\top) (\leq \eta)$  and

$$G(\mathbf{y}^{k+1}) \leq G(\mathbf{y}^k) + \langle \mathbf{y}^{k+1} - \mathbf{y}^k, \nabla G(\mathbf{y}^k) \rangle + \frac{\eta \rho_{\max}(\mathcal{L} \mathcal{L}^\top)}{2} \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2. \quad (3.4)$$

The iterative scheme (3.3) can be regarded as a special type of the proximal iterative hard thresholding algorithm deeply studied in [6,21]. We define

$$F(\mathbf{y}) = \lambda \|\mathbf{y}\|_0 + G(\mathbf{y}).$$

It follows from the inequality (3.3) and the inequality (3.4) that

$$\begin{aligned} F(\mathbf{y}^{k+1}) &= G(\mathbf{y}^{k+1}) + \lambda \|\mathbf{y}^{k+1}\|_0 \\ &\leq G(\mathbf{y}^k) + \langle \mathbf{y}^{k+1} - \mathbf{y}^k, \nabla G(\mathbf{y}^k) \rangle + \frac{\eta \rho_{\max}(\mathcal{L} \mathcal{L}^T)}{2} \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 + \lambda \|\mathbf{y}^{k+1}\|_0 \\ &\leq G(\mathbf{y}^k) + \langle \mathbf{y}^{k+1} - \mathbf{y}^k, \nabla G(\mathbf{y}^k) \rangle + \frac{\eta + d_k}{2} \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 + \lambda \|\mathbf{y}^{k+1}\|_0 \\ &\leq G(\mathbf{y}^k) + \lambda \|\mathbf{y}^k\|_0 = F(\mathbf{y}^k). \end{aligned} \quad (3.5)$$

The inequality (3.5) implies that  $\{F(\mathbf{y}^k)\}$  is nonincreasing and

$$F(\mathbf{y}^k) - F(\mathbf{y}^{k+1}) \geq b - a = \frac{\eta + d_k - \eta \rho_{\max}(\mathcal{L} \mathcal{L}^T)}{2} \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2, \quad (3.6)$$

where

$$\begin{cases} a = G(\mathbf{y}^k) + \langle \mathbf{y}^{k+1} - \mathbf{y}^k, \nabla G(\mathbf{y}^k) \rangle + \frac{\eta \rho_{\max}(\mathcal{L} \mathcal{L}^T)}{2} \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 + \lambda \|\mathbf{y}^{k+1}\|_0, \\ b = G(\mathbf{y}^k) + \langle \mathbf{y}^{k+1} - \mathbf{y}^k, \nabla G(\mathbf{y}^k) \rangle + \frac{\eta + d_k}{2} \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 + \lambda \|\mathbf{y}^{k+1}\|_0. \end{cases}$$

Since  $G(\mathbf{y})$  is bounded below, it is trivial to see that  $\{F(\mathbf{y}^k)\}$  is bounded below. Accordingly,  $\{F(\mathbf{y}^k)\}$  converges to a finite value as  $k \rightarrow \infty$ . This, together with (3.6), implies that

$$\lim_{k \rightarrow \infty} \|\mathbf{y}^{k+1} - \mathbf{y}^k\| = 0. \quad (3.7)$$

From (3.3) we also get

$$\mathbf{y}^{k+1} \in \operatorname{argmin}_{\mathbf{y}} \lambda \|\mathbf{y}\|_0 + \frac{\eta + d_k}{2} \|\mathbf{y}\|^2 - (\eta + d_k) \left\langle \mathbf{y}, \mathbf{y}^k - \frac{1}{\eta + d_k} \nabla G(\mathbf{y}^k) \right\rangle. \quad (3.8)$$

Let  $g_i^k$  denote the  $i$ th entry of the vector  $\nabla G(\mathbf{y}^k)$ . Considering the  $i$ th entrywise minimization problem in (3.8)

$$y_i^{k+1} = \operatorname{argmin}_{y \in \mathbb{R}} \lambda \|y\|_0 + \frac{\eta + d_k}{2} y^2 - (\eta + d_k) \left\langle y, y_i^k - \frac{g_i^k}{\eta + d_k} \right\rangle,$$

we can see that  $y_i^{k+1} = y_i^k - \frac{g_i^k}{\eta + d_k}$  if  $y_i^k - \frac{g_i^k}{\eta + d_k} \geq \sqrt{\frac{2\lambda}{\eta + d_k}}$  and zero otherwise.

Let  $S(\mathbf{y}^k)$  denote the support of the vector  $\mathbf{y}^k$  and  $S^c(\mathbf{y}^k)$  denote the complement set of  $S(\mathbf{y}^k)$ . If  $S(\mathbf{y}^k) \neq S(\mathbf{y}^{k+1})$ , we show that

$$\|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 \geq \min \left\{ \frac{2\lambda}{\eta + d_{k-1}}, \frac{2\lambda}{\eta + d_k} \right\} \quad (3.9)$$

by discussing the following four cases:

1. If  $j \in \overline{S(\mathbf{y}^k)} \cap \overline{S(\mathbf{y}^{k+1})}$ , one obtains that  $y_j^k = y_j^{k+1} = 0$ . Thus,  $|y_j^{k+1} - y_j^k| = 0$ .
2. If  $j \in S(\mathbf{y}^k) \cap \overline{S(\mathbf{y}^{k+1})}$ , one obtains that  $y_j^k \neq 0$  and  $y_j^{k+1} = 0$ . Since only

$$y_j^k = y_j^{k-1} - \frac{g_i^{k-1}}{\eta + d_{k-1}} \geq \sqrt{\frac{2\lambda}{\eta + d_{k-1}}}$$

can make  $y_j^k \neq 0$ , otherwise  $y_j^k = 0$  is a contradiction. Thus we get  $|y_j^{k+1} - y_j^k| \geq \sqrt{\frac{2\lambda}{\eta + d_{k-1}}}$ .

3. If  $j \in \overline{S(\mathbf{y}^k)} \cap S(\mathbf{y}^{k+1})$ , one obtains that  $y_j^k = 0$  and  $y_j^{k+1} \neq 0$ . Thus,

$$y_j^{k+1} = y_j^k - \frac{1}{\eta + d_k} g_i^k \geq \sqrt{\frac{2\lambda}{\eta + d_k}} \quad \text{and} \quad |y_j^{k+1} - y_j^k| \geq \sqrt{\frac{2\lambda}{\eta + d_k}}.$$

4. If  $j \in S(\mathbf{y}^k) \cap S(\mathbf{y}^{k+1})$ , one obtains that  $y_j^k \neq 0$  and  $y_j^{k+1} \neq 0$ . Clearly, we have  $|y_j^{k+1} - y_j^k| \geq 0$ .

It follows from (3.7) and (3.9) that  $S(\mathbf{y}^k)$  does not change when  $k$  is sufficiently large. Hence, there exists a positive integer  $k_0$  such that  $S(\mathbf{y}^k) = S(\mathbf{y}^{k_0})$  holds for all  $k \geq k_0$ . For any  $k \geq k_0$ , the iterative scheme (3.3) becomes the projected gradient scheme

$$\mathbf{y}^{k+1} = \operatorname{argmin}_{\mathbf{y} \in B} \left\{ G(\mathbf{y}^k) + \langle \mathbf{y} - \mathbf{y}^k, \nabla G(\mathbf{y}^k) \rangle + \frac{\eta + d_k}{2} \|\mathbf{y} - \mathbf{y}^k\|^2 \right\}, \quad (3.10)$$

where  $B = \{\mathbf{y} : S(\mathbf{y}) = S(\mathbf{y}^{k_0})\}$ . Let  $\mathbf{y}^*$  denote the solution to minimization problem

$$\min_{\mathbf{y} \in B} G(\mathbf{y}).$$

With the similar proof of [16, Theorem 2.2], we can show that the sequence  $\{\mathbf{y}^k\}$  generated from (3.10) satisfies

$$G(\mathbf{y}^{k+l}) - G(\mathbf{y}^*) \leq \frac{L}{2l} \|\mathbf{y}^k - \mathbf{y}^*\|^2, \quad (3.11)$$

where  $L = \eta + \sup_{k \geq 0} d_k$ . Let  $\rho_{\min}(\mathcal{L}\mathcal{L}^\top)$  denote the smallest eigenvalue of  $\mathcal{L}\mathcal{L}^\top$ . Since  $G(x)$  is strongly convex with modulus  $\eta\rho_{\min}(\mathcal{L}\mathcal{L}^\top)$  and  $\nabla G(\mathbf{y}^*) = 0$ , we have

$$\|\mathbf{y}^{k+l} - \mathbf{y}^*\|^2 \leq \frac{2}{\eta\rho_{\min}(\mathcal{L}\mathcal{L}^\top)} (G(\mathbf{y}^{k+l}) - G(\mathbf{y}^*)), \quad (3.12)$$

where  $l$  is a positive integer. It follows from (3.11) and (3.12) that

$$\|\mathbf{y}^{k_0+l} - \mathbf{y}^*\|^2 \leq \frac{L}{l\eta\rho_{\min}(\mathcal{L}\mathcal{L}^\top)} \|\mathbf{y}^{k_0} - \mathbf{y}^*\|^2.$$

The above inequality implies that

$$\|\mathbf{y}^k - \mathbf{y}^*\| \leq \mathcal{O}(1/\sqrt{k}) \quad \text{as } k \rightarrow \infty.$$

Hence, we have

$$\lim_{k \rightarrow \infty} \mathbf{y}^k = \mathbf{y}^*.$$

This, together with (3.1), implies that  $\mathbf{x}^* = \lim_{k \rightarrow \infty} \mathbf{x}^k$  and

$$\|\mathbf{x}^k - \mathbf{x}^*\| \leq \rho \|\mathbf{y}^k - \mathbf{y}^*\| \leq \mathcal{O}(1/\sqrt{k}) \quad \text{as } k \rightarrow \infty,$$

where  $\rho (< \infty)$  is the spectral norm of  $\eta(\mathcal{A}^\top \mathcal{A} + \eta \mathcal{I})^{-1} \mathcal{D}$ .  $\square$

### 3.2. Global convergence of the PPALM method

Note that the iterative scheme (1.2) is equal to

$$\begin{cases} \mathbf{x}^{k+1} = \operatorname{argmin}_{\mathbf{x}} \frac{1}{2} \|\mathcal{A}\mathbf{x} - \mathbf{b}\|^2 + \frac{\eta}{2} \|\mathcal{D}^\top \mathbf{x} - \mathbf{y}^k\|^2, \\ \mathbf{y}^{k+1} \in \operatorname{argmin}_{\mathbf{y}} \lambda \|\mathbf{y}\|_0 + \eta \langle \mathbf{y}^k - \mathcal{D}^\top \mathbf{x}^{k+1}, \mathbf{y} - \mathbf{y}^k \rangle + \frac{\eta + d_k}{2} \|\mathbf{y} - \mathbf{y}^k\|^2, \end{cases} \quad k = 0, 1, 2, \dots$$

If we set  $\hat{d}_k = \eta + d_k$ , the above scheme is obviously a special case of (1.4). In this subsection, we consider the more general iterative scheme (1.4), i.e.,

$$\begin{cases} \mathbf{x}^{k+1} = \operatorname{argmin}_{\mathbf{x}} f(\mathbf{x}) + H(\mathbf{x}, \mathbf{y}^k), \\ \mathbf{y}^{k+1} \in \operatorname{argmin}_{\mathbf{y}} \langle \nabla_{\mathbf{y}} H(\mathbf{x}^{k+1}, \mathbf{y}^k), \mathbf{y} - \mathbf{y}^k \rangle + g(\mathbf{y}) + \frac{d_k}{2} \|\mathbf{y} - \mathbf{y}^k\|^2. \end{cases} \quad (3.13)$$

For the sake of convenience, we define

$$\mathbf{z}^k := (\mathbf{x}^k, \mathbf{y}^k), \quad k = 1, 2, \dots,$$

and make the following assumptions for (3.13):

**Assumption 3.1.** (i) The function  $f$  is convex, the function  $g$  is proper and lower semicontinuous and the function  $H$  is strongly convex.

(ii)  $\inf_{\mathbb{R}^n \times \mathbb{R}^d} \Psi > -\infty$ ,  $\inf_{\mathbb{R}^n} f > -\infty$  and  $\inf_{\mathbb{R}^d} g > -\infty$ .

(iii) For any fixed  $\mathbf{x}^k$ , the partial gradient  $\nabla_{\mathbf{x}}H(\mathbf{x}^k, \mathbf{y})$  is globally Lipschitz with moduli  $L_1(\mathbf{x}^k)$ , that is

$$\|\nabla_{\mathbf{x}}H(\mathbf{x}^k, \mathbf{y}^k) - \nabla_{\mathbf{x}}H(\mathbf{x}^k, \mathbf{y}^{k-1})\| \leq L_1(\mathbf{x}^k)\|\mathbf{y}^k - \mathbf{y}^{k-1}\|.$$

Likewise, the partial gradient  $\nabla_{\mathbf{y}}H(\mathbf{x}^k, \mathbf{y})$  is globally Lipschitz with moduli  $L_2(\mathbf{x}^k)$ .

(iv) For  $i = 1, 2$ , there exist  $\lambda_i^+, \lambda_i^-$  such that

$$\lambda_1^- \leq L_1(\mathbf{x}^k) \leq \lambda_1^+ \quad \text{and} \quad \lambda_2^- \leq L_2(\mathbf{x}^k) \leq \lambda_2^+,$$

for any  $k \in \mathbb{N}$ .

(v) For any  $k \in \mathbb{N}$ ,

$$c^- \leq c_k \leq c^+ \quad \text{and} \quad d^- \leq d_k \leq d^+.$$

**Lemma 3.1** (Sufficient Decrease Property). Let  $\{\mathbf{z}^k\}_{k \in \mathbb{N}}$  be a sequence generated by (3.13). If Assumption 3.1 holds, then there exists a positive constant  $\rho_1$  such that

$$\frac{\rho_1}{2} \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 \leq \Psi(\mathbf{z}^k) - \Psi(\mathbf{z}^{k+1}), \quad k = 0, 1, 2, \dots, \quad (3.14)$$

where  $\rho_1 = \min\{c^-, d^- - \lambda_2^+\}$ . Moreover,  $\|\mathbf{z}^{k+1} - \mathbf{z}^k\| \rightarrow 0$  as  $k \rightarrow +\infty$ .

**Proof.** Since  $\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} H(\mathbf{x}, \mathbf{y}^k) + f(\mathbf{x})$ , the first-order optimality condition implies that

$$\nabla H(\mathbf{x}^{k+1}, \mathbf{y}^k) + \nabla f(\mathbf{x}^{k+1}) = 0. \quad (3.15)$$

Note that the function  $H(\mathbf{x}, \mathbf{y})$  is strongly convex, then there exists a constant  $c_k > 0$  such that

$$H(\mathbf{x}^k, \mathbf{y}^k) \geq H(\mathbf{x}^{k+1}, \mathbf{y}^k) + \langle \nabla H(\mathbf{x}^{k+1}, \mathbf{y}^k), \mathbf{x}^k - \mathbf{x}^{k+1} \rangle + \frac{c_k}{2} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2.$$

Since  $f(\mathbf{x})$  is convex, we also have

$$f(\mathbf{x}^k) \geq f(\mathbf{x}^{k+1}) + \langle \nabla f(\mathbf{x}^{k+1}), \mathbf{x}^k - \mathbf{x}^{k+1} \rangle.$$

It follows that

$$\begin{aligned} H(\mathbf{x}^k, \mathbf{y}^k) + f(\mathbf{x}^k) &\geq H(\mathbf{x}^{k+1}, \mathbf{y}^k) + \langle \nabla H(\mathbf{x}^{k+1}, \mathbf{y}^k), \mathbf{x}^k - \mathbf{x}^{k+1} \rangle + \frac{c_k}{2} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 + f(\mathbf{x}^k) \\ &= H(\mathbf{x}^{k+1}, \mathbf{y}^k) + \langle \nabla f(\mathbf{x}^{k+1}), \mathbf{x}^{k+1} - \mathbf{x}^k \rangle + f(\mathbf{x}^k) + \frac{c_k}{2} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 \\ &\geq H(\mathbf{x}^{k+1}, \mathbf{y}^k) + f(\mathbf{x}^{k+1}) + \frac{c_k}{2} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2, \end{aligned} \quad (3.16)$$

where the second equality uses (3.15). We note that

$$\mathbf{y}^{k+1} \in \arg\min_{\mathbf{y}} \langle \nabla_{\mathbf{y}}H(\mathbf{x}^{k+1}, \mathbf{y}^k), \mathbf{y} - \mathbf{y}^k \rangle + g(\mathbf{y}) + \frac{d_k}{2} \|\mathbf{y} - \mathbf{y}^k\|^2,$$

and the partial gradient  $\nabla_{\mathbf{y}}H(\mathbf{x}^{k+1}, \mathbf{y})$  is globally Lipschitz with moduli  $L_2(\mathbf{x}^{k+1})$ . Applying Lemma 2.1 with

$$h(\cdot) := H(\mathbf{x}^{k+1}, \cdot), \quad \sigma = g, \quad t = d_k > L_2(\mathbf{x}^{k+1}),$$

we have

$$H(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}) + g(\mathbf{y}^{k+1}) \leq H(\mathbf{x}^{k+1}, \mathbf{y}^k) + g(\mathbf{y}^k) - \left( \frac{d_k - L_2(\mathbf{x}^{k+1})}{2} \right) \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2. \quad (3.17)$$

It follows from (3.16) and (3.17) that

$$\begin{aligned} \Psi(\mathbf{x}^k, \mathbf{y}^k) - \Psi(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}) &= H(\mathbf{x}^k, \mathbf{y}^k) + f(\mathbf{x}^k) + g(\mathbf{y}^k) - H(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}) - f(\mathbf{x}^{k+1}) - g(\mathbf{y}^{k+1}) \\ &\geq \frac{c_k}{2} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + \frac{d_k - L_2(\mathbf{x}^{k+1})}{2} \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 \\ &\geq \frac{\rho_1}{2} \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2, \end{aligned} \quad (3.18)$$

where  $\rho_1 = \min\{c^-, d^- - \lambda_2^+\}$ . This shows that the sequence  $\{\Psi(\mathbf{z}^k)\}_{k \in \mathbb{N}}$  is nonincreasing. Since  $\Psi$  is assumed to be bounded below, it converges to some real number  $\underline{\Psi}$ . Summing up the above inequality (3.18) from  $k = 0$  to  $k = n - 1$ , we have

$$\sum_{k=0}^{n-1} \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 \leq \frac{2}{\rho_1} (\Psi(\mathbf{z}^0) - \Psi(\mathbf{z}^n)) \leq \frac{2}{\rho_1} (\Psi(\mathbf{z}^0) - \underline{\Psi}).$$

This implies that  $\lim_{k \rightarrow +\infty} \|\mathbf{z}^{k+1} - \mathbf{z}^k\| = 0$ .  $\square$

**Lemma 3.2** (Subgradient Lower Bound for Iterates Gap). Suppose that [Assumption 3.1](#) holds. Let  $\{\mathbf{z}^k\}_{k \in \mathbb{N}}$  be a sequence generated by [\(3.13\)](#). For all  $k \in \mathbb{N}$ , we define

$$\begin{cases} \mathcal{Q}_{\mathbf{x}}^k := \nabla_{\mathbf{x}} H(\mathbf{x}^k, \mathbf{y}^k) + \nabla f(\mathbf{x}^k), \\ \mathcal{Q}_{\mathbf{y}}^k := d_{k-1}(\mathbf{y}^{k-1} - \mathbf{y}^k) + \nabla_{\mathbf{y}} H(\mathbf{x}^k, \mathbf{y}^k) - \nabla_{\mathbf{y}} H(\mathbf{x}^k, \mathbf{y}^{k-1}). \end{cases} \quad (3.19)$$

If  $\{\mathbf{z}^k\}_{k \in \mathbb{N}}$  is bounded, then  $(\mathcal{Q}_{\mathbf{x}}^k, \mathcal{Q}_{\mathbf{y}}^k) \in \partial \Psi(\mathbf{x}^k, \mathbf{y}^k)$  and  $\text{dist}(0, \partial \Psi(\mathbf{z}^k)) \leq (d^+ + \lambda_1^+ + \lambda_2^+) \|\mathbf{z}^k - \mathbf{z}^{k-1}\|$ .

**Proof.** Since  $f(\mathbf{x})$  is convex, it follows from [\(2.1\)](#) that

$$\partial \Psi(\mathbf{x}, \mathbf{y}) = (\partial_{\mathbf{x}} \Psi(\mathbf{x}, \mathbf{y}), \partial_{\mathbf{y}} \Psi(\mathbf{x}, \mathbf{y})) = (\nabla_{\mathbf{x}} H(\mathbf{x}, \mathbf{y}) + \nabla f(\mathbf{x}), \nabla_{\mathbf{y}} H(\mathbf{x}, \mathbf{y}) + \partial g(\mathbf{y})).$$

By the second iteration step of [\(3.13\)](#), we have

$$\mathbf{y}^k \in \underset{\mathbf{y} \in \mathbb{R}^d}{\text{argmin}} \langle \mathbf{y} - \mathbf{y}^{k-1}, \nabla_{\mathbf{y}} H(\mathbf{x}^k, \mathbf{y}^{k-1}) \rangle + \frac{d_{k-1}}{2} \|\mathbf{y} - \mathbf{y}^{k-1}\|^2 + g(\mathbf{y}).$$

The optimality condition is

$$\nabla_{\mathbf{y}} H(\mathbf{x}^k, \mathbf{y}^{k-1}) + d_{k-1}(\mathbf{y}^k - \mathbf{y}^{k-1}) + \partial g(\mathbf{y}^k) = 0.$$

Hence,

$$\mathcal{Q}_{\mathbf{y}}^k = \nabla_{\mathbf{y}} H(\mathbf{x}^k, \mathbf{y}^k) + \partial g(\mathbf{y}^k).$$

Together with [\(3.19\)](#), we have  $(\mathcal{Q}_{\mathbf{x}}^k, \mathcal{Q}_{\mathbf{y}}^k) \in \partial \Psi(\mathbf{x}^k, \mathbf{y}^k)$ . Since  $\nabla H$  is Lipschitz continuous and  $\{\mathbf{z}^k\}_{k \in \mathbb{N}}$  is bounded, we have

$$\begin{aligned} \|\mathcal{Q}_{\mathbf{y}}^k\| &\leq d_{k-1} \|\mathbf{y}^{k-1} - \mathbf{y}^k\| + \|\nabla_{\mathbf{y}} H(\mathbf{x}^k, \mathbf{y}^k) - \nabla_{\mathbf{y}} H(\mathbf{x}^k, \mathbf{y}^{k-1})\| \\ &\leq (d_{k-1} + \lambda_2^+) \|\mathbf{y}^k - \mathbf{y}^{k-1}\| \\ &\leq (d^+ + \lambda_2^+) \|\mathbf{z}^k - \mathbf{z}^{k-1}\|. \end{aligned} \quad (3.20)$$

Similarly, we can get

$$\begin{aligned} \|\mathcal{Q}_{\mathbf{x}}^k\| &= \|\nabla_{\mathbf{x}} H(\mathbf{x}^k, \mathbf{y}^k) + \nabla f(\mathbf{x}^k)\| \\ &= \|\nabla_{\mathbf{x}} H(\mathbf{x}^k, \mathbf{y}^k) - \nabla_{\mathbf{x}} H(\mathbf{x}^k, \mathbf{y}^{k-1})\| \\ &\leq \lambda_1^+ \|\mathbf{y}^k - \mathbf{y}^{k-1}\| \\ &\leq \lambda_1^+ \|\mathbf{z}^k - \mathbf{z}^{k-1}\|, \end{aligned} \quad (3.21)$$

where the second equality is due to the first order optimality condition  $\nabla_{\mathbf{x}} H(\mathbf{x}^k, \mathbf{y}^{k-1}) + \nabla f(\mathbf{x}^k) = 0$ . Then it follows from [\(3.20\)](#) and [\(3.21\)](#) that

$$\text{dist}(0, \partial \Psi(\mathbf{z}^k)) \leq \|\mathcal{Q}_{\mathbf{x}}^k\| + \|\mathcal{Q}_{\mathbf{y}}^k\| \leq (d^+ + \lambda_1^+ + \lambda_2^+) \|\mathbf{z}^k - \mathbf{z}^{k-1}\|. \quad \square \quad (3.22)$$

Next we summarize some properties for the limiting point set. Let  $\mathbf{z}^0$  be a starting point of the PPALM method. We define the crit  $\Psi$  as the set of critical points of  $\Psi$  and denote the set of limit points by

$$\omega(\mathbf{z}^0) = \{\mathbf{z}^* : \text{there exists a subsequence } \{\mathbf{z}^{k_q}\} \text{ such that } \mathbf{z}^{k_q} \rightarrow \mathbf{z}^* \text{ as } q \rightarrow +\infty\}.$$

**Lemma 3.3.** Suppose that [Assumption 3.1](#) holds. Let  $\{\mathbf{z}^k\}$  be a sequence generated by [\(3.13\)](#). Then the following statements hold,

1.  $\emptyset \neq \omega(\mathbf{z}^0) \subset \text{crit} \Psi$ ;
2.  $\lim_{k \rightarrow \infty} \text{dist}(\mathbf{z}^k, \omega(\mathbf{z}^0)) = 0$ ;
3.  $\omega(\mathbf{z}^0)$  is a nonempty, compact and connected set;
4.  $\Psi$  is finite and constant on  $\omega(\mathbf{z}^0)$ .

**Proof.** The proof follows from [\[7, Lemma 3.5\]](#). Let  $\mathbf{z}^* = (\mathbf{x}^*, \mathbf{y}^*)$  be a limit point of  $\{\mathbf{z}^k\}_{k \in \mathbb{N}} = \{(\mathbf{x}^k, \mathbf{y}^k)\}_{k \in \mathbb{N}}$ . Consequently, there exists a subsequence  $\{(\mathbf{x}^{k_q}, \mathbf{y}^{k_q})\}_{q \in \mathbb{N}}$  such that  $(\mathbf{x}^{k_q}, \mathbf{y}^{k_q}) \rightarrow (\mathbf{x}^*, \mathbf{y}^*)$  as  $q \rightarrow +\infty$ . Hence, for any  $\delta > 0$ ,  $\|\mathbf{x}^{k_q} - \mathbf{x}^*\| \leq \delta$  as  $q \rightarrow +\infty$ . Furthermore, the convexity of  $f(\mathbf{x})$  implies that there exists a constant  $\zeta$  such that

$$\|f(\mathbf{x}^{k_q}) - f(\mathbf{x}^*)\| \leq \zeta \|\mathbf{x}^{k_q} - \mathbf{x}^*\| \leq \zeta \delta$$

holds when  $q \rightarrow +\infty$ . It concludes that  $f(\mathbf{x}^{k_q})$  tends to  $f(\mathbf{x}^*)$  as  $q \rightarrow +\infty$ .

Since  $g(\mathbf{y})$  is lower semicontinuous, we obtain that

$$\liminf_{q \rightarrow +\infty} g(\mathbf{y}^{k_q}) \geq g(\mathbf{y}^*). \quad (3.23)$$



From the second iteration step of (3.13), we have

$$\mathbf{y}^{k+1} \in \operatorname{argmin}_{\mathbf{y}} \langle \nabla_{\mathbf{y}} H(\mathbf{x}^{k+1}, \mathbf{y}^k), \mathbf{y} - \mathbf{y}^k \rangle + g(\mathbf{y}) + \frac{d_k}{2} \|\mathbf{y} - \mathbf{y}^k\|^2.$$

Therefore, we obtain that

$$\begin{aligned} & \langle \nabla_{\mathbf{y}} H(\mathbf{x}^{k+1}, \mathbf{y}^k), \mathbf{y}^{k+1} - \mathbf{y}^k \rangle + g(\mathbf{y}^{k+1}) + \frac{d_k}{2} \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 \\ & \leq \langle \nabla_{\mathbf{y}} H(\mathbf{x}^{k+1}, \mathbf{y}^k), \mathbf{y}^* - \mathbf{y}^k \rangle + g(\mathbf{y}^*) + \frac{d_k}{2} \|\mathbf{y}^* - \mathbf{y}^k\|^2. \end{aligned} \quad (3.24)$$

Setting  $k = k_q - 1$  in the above inequality and letting  $q \rightarrow +\infty$ , we get

$$\begin{aligned} \limsup_{q \rightarrow +\infty} g(\mathbf{y}^{k_q}) & \leq \limsup_{q \rightarrow +\infty} \langle \nabla_{\mathbf{y}} H(\mathbf{x}^{k_q}, \mathbf{y}^{k_q-1}), \mathbf{y}^* - \mathbf{y}^{k_q-1} \rangle + g(\mathbf{y}^*) + \frac{d_k}{2} \|\mathbf{y}^* - \mathbf{y}^{k_q-1}\|^2 \\ & = g(\mathbf{y}^*), \end{aligned}$$

where the last inequality is implied by the fact that  $\mathbf{y}^{k_q-1} \rightarrow \mathbf{y}^*$  as  $q \rightarrow +\infty$ . Together with (3.23), we conclude that  $g(\mathbf{y}^{k_q})$  tends to  $g(\mathbf{y}^*)$  as  $q \rightarrow +\infty$ . Accordingly, we have

$$\begin{aligned} \lim_{q \rightarrow +\infty} \Psi(\mathbf{x}^{k_q}, \mathbf{y}^{k_q}) & = \lim_{q \rightarrow +\infty} \{H(\mathbf{x}^{k_q}, \mathbf{y}^{k_q}) + f(\mathbf{x}^{k_q}) + g(\mathbf{y}^{k_q})\} \\ & = H(\mathbf{x}^*, \mathbf{y}^*) + f(\mathbf{x}^*) + g(\mathbf{y}^*) \\ & = \Psi(\mathbf{x}^*, \mathbf{y}^*). \end{aligned}$$

From Lemmas 3.1 and 3.2 we also have  $(\mathcal{Q}_{\mathbf{x}}^k, \mathcal{Q}_{\mathbf{y}}^k) \in \partial \Psi(\mathbf{x}^k, \mathbf{y}^k)$  and  $(\mathcal{Q}_{\mathbf{x}}^k, \mathcal{Q}_{\mathbf{y}}^k) \rightarrow (0, 0)$  as  $k \rightarrow +\infty$ . The closedness property of  $\partial \Psi$  implies that  $(0, 0) \in \partial \Psi(\mathbf{x}^*, \mathbf{y}^*)$ . Therefore,  $(\mathbf{x}^*, \mathbf{y}^*)$  is a critical point of  $\Psi$ . This completes the proof of item 1. The items 2, 3, 4 follow easily with [7, Lemma 3.5].  $\square$

Now we use the KL property to prove the global convergence properties for the sequence  $\{\mathbf{z}^k\}$  generated by the PPALM method.

**Theorem 3.2.** Suppose that  $\Psi$  is a KL function such that Assumption 3.1 holds. Let  $\{\mathbf{z}^k\}$  be a sequence generated by (3.13) which is assumed to be bounded. Then the sequence  $\{\mathbf{z}^k\}_{k \in \mathbb{N}}$  converges to a critical point of  $\Psi$ .

**Proof.** Since  $\{\mathbf{z}^k\}$  is bounded, there exists a subsequence  $\{\mathbf{z}^{k_q}\}_{q \in \mathbb{N}}$  such that  $\mathbf{z}^{k_q} \rightarrow \mathbf{z}^*$  as  $q \rightarrow \infty$ . Moreover, the conclusion that  $\mathbf{z}^*$  is a critical point of  $\Psi$  follows from the item 1 of Lemma 3.3. Since  $\lim_{k \rightarrow +\infty} \|\mathbf{z}^{k+1} - \mathbf{z}^k\| = 0$  in Lemma 3.1, we obtain that

$$\mathbf{z}^{k_q+1} \rightarrow \mathbf{z}^* \quad \text{as } q \rightarrow \infty. \quad (3.25)$$

In a similar way as in Lemma 3.3 we get that

$$f(\mathbf{x}^{k_q+1}) \rightarrow f(\mathbf{x}^*) \quad \text{as } q \rightarrow +\infty$$

along with (3.25). Setting  $k = k_q$  in (3.24) and letting  $q \rightarrow +\infty$ , it is easy to check that

$$\limsup_{q \rightarrow +\infty} g(\mathbf{y}^{k_q+1}) \leq g(\mathbf{y}^*).$$

Also note that  $g(\mathbf{y})$  is lower semicontinuous, we have

$$\liminf_{q \rightarrow +\infty} g(\mathbf{y}^{k_q+1}) \geq g(\mathbf{y}^*).$$

Then

$$g(\mathbf{y}^{k_q+1}) \rightarrow g(\mathbf{y}^*) \quad \text{as } q \rightarrow +\infty.$$

Therefore, we obtain that

$$\lim_{q \rightarrow +\infty} \Psi(\mathbf{x}^{k_q+1}, \mathbf{y}^{k_q+1}) = \Psi(\mathbf{x}^*, \mathbf{y}^*). \quad (3.26)$$

Combining (3.26) and the existence of  $\lim_{k \rightarrow +\infty} \Psi(\mathbf{z}^k)$ , we conclude that

$$\lim_{k \rightarrow +\infty} \Psi(\mathbf{z}^k) = \Psi(\mathbf{z}^*). \quad (3.27)$$

In fact, if there exists a sufficiently large  $\bar{k}$  such that  $\Psi(\mathbf{z}^{\bar{k}}) = \Psi(\mathbf{z}^*)$ , then Lemma 3.1 shows that  $\mathbf{z}^{\bar{k}} = \mathbf{z}^k$  for all  $k > \bar{k}$ . Accordingly, the sequence  $\{\mathbf{z}^k\}_{k \in \mathbb{N}}$  is stationary and the conclusion of this theorem holds trivially. Otherwise, we have

$\Psi(\mathbf{z}^k) \neq \Psi(\mathbf{z}^*)$  for all  $k > 0$ . Since the inequality (3.14) implies that  $\{\Psi(\mathbf{z}^k)\}$  is a nonincreasing sequence, it concludes that  $\Psi(\mathbf{z}^*) < \Psi(\mathbf{z}^k)$  for all  $k > 0$ . On the other hand, (3.27) implies that for any  $\delta > 0$  there exists a nonnegative integer  $k_0$  such that  $\Psi(\mathbf{z}^k) < \Psi(\mathbf{z}^*) + \delta$  for all  $k > k_0$ . Since  $\lim_{k \rightarrow \infty} \text{dist}(\mathbf{z}^k, \omega(\mathbf{z}^0)) = 0$  in Lemma 3.3, for any  $\epsilon > 0$  there exists a positive integer  $k_1$  such that  $\text{dist}(\mathbf{z}^k, \omega(\mathbf{z}^0)) < \epsilon$  for all  $k > k_1$ . We thus define  $L = \max\{k_0, k_1\}$ . Note that  $\omega(\mathbf{z}^0)$  is nonempty and compact;  $\Psi$  is a constant on  $\omega(\mathbf{z}^0)$ . By Lemma 2.2, there exists a continuous concave function  $\varphi$  such that

$$\varphi'(\Psi(\mathbf{z}^k) - \Psi(\mathbf{z}^*)) \text{dist}(0, \partial\Psi(\mathbf{z}^k)) \geq 1 \quad \text{for all } k > L.$$

This, together with (3.22), leads to

$$\varphi'(\Psi(\mathbf{z}^k) - \Psi(\mathbf{z}^*)) \geq \frac{1}{d^+ + \lambda_1^+ + \lambda_2^+} \|\mathbf{z}^k - \mathbf{z}^{k-1}\|^{-1}. \quad (3.28)$$

The concavity of  $\varphi$  implies that

$$\varphi(\Psi(\mathbf{z}^k) - \Psi(\mathbf{z}^*)) - \varphi(\Psi(\mathbf{z}^{k+1}) - \Psi(\mathbf{z}^*)) \geq \varphi'(\Psi(\mathbf{z}^k) - \Psi(\mathbf{z}^*))(\Psi(\mathbf{z}^k) - \Psi(\mathbf{z}^{k+1})).$$

For the sake of convenience, we define  $\Delta_{p,q} := \varphi(\Psi(\mathbf{z}^p) - \Psi(\mathbf{z}^*)) - \varphi(\Psi(\mathbf{z}^q) - \Psi(\mathbf{z}^*))$  for all  $p, q \in \mathbb{N}$ . By (3.14), we have

$$\Delta_{k,k+1} \geq \frac{\rho_1}{2(d^+ + \lambda_1^+ + \lambda_2^+)} \frac{\|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2}{\|\mathbf{z}^k - \mathbf{z}^{k-1}\|}.$$

Setting  $C := 2(d^+ + \lambda_1^+ + \lambda_2^+)/\rho_1$ , we have

$$\|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 \leq C \Delta_{k,k+1} \|\mathbf{z}^k - \mathbf{z}^{k-1}\| \leq \left( \frac{\|\mathbf{z}^k - \mathbf{z}^{k-1}\| + C \Delta_{k,k+1}}{2} \right)^2.$$

Hence,

$$2\|\mathbf{z}^{k+1} - \mathbf{z}^k\| \leq \|\mathbf{z}^k - \mathbf{z}^{k-1}\| + C \Delta_{k,k+1}. \quad (3.29)$$

Summing up (3.29) from  $i = L + 1$  to  $i = k$ , we have

$$\begin{aligned} 2 \sum_{i=L+1}^k \|\mathbf{z}^{i+1} - \mathbf{z}^i\| &\leq \sum_{i=L+1}^k \|\mathbf{z}^i - \mathbf{z}^{i-1}\| + C \sum_{i=L+1}^k \Delta_{i,i+1} \\ &\leq \sum_{i=L+1}^k \|\mathbf{z}^{i+1} - \mathbf{z}^i\| + \|\mathbf{z}^{L+1} - \mathbf{z}^L\| + C \Delta_{L+1,k+1}. \end{aligned}$$

Thus, for any  $k > L$ ,

$$\sum_{i=L+1}^k \|\mathbf{z}^{i+1} - \mathbf{z}^i\| \leq \|\mathbf{z}^{L+1} - \mathbf{z}^L\| + C \varphi(\Psi(\mathbf{z}^{L+1}) - \Psi(\mathbf{z}^*)). \quad (3.30)$$

This implies that

$$\sum_{k=1}^{\infty} \|\mathbf{z}^{k+1} - \mathbf{z}^k\| < +\infty. \quad (3.31)$$

For  $q > p > L$ , we have

$$\|\mathbf{z}^q - \mathbf{z}^p\| = \left\| \sum_{k=p}^{q-1} (\mathbf{z}^{k+1} - \mathbf{z}^k) \right\| \leq \sum_{k=p}^{q-1} \|\mathbf{z}^{k+1} - \mathbf{z}^k\|.$$

It follows from (3.31) that  $\sum_{k=L+1}^{\infty} \|\mathbf{z}^{k+1} - \mathbf{z}^k\| \rightarrow 0$  as  $L \rightarrow +\infty$ . Accordingly, the sequence  $\{\mathbf{z}^k\}_{k \in \mathbb{N}}$  is a Cauchy sequence. Therefore, we conclude that the sequence  $\{\mathbf{z}^k\}_{k \in \mathbb{N}}$  converges to a critical point of  $\Psi$ .  $\square$

Following [13, Theorem 5], we furthermore establish the convergence rate of the PPALM method.

**Lemma 3.4 (Convergence Rate).** Suppose that Assumption 3.1 holds.  $\Psi$  satisfies the KL inequality with  $\varphi(s) = \omega s^{1-\theta}$  for  $\omega > 0$  and  $\theta \in [0, 1)$ . Let  $\{\mathbf{z}^k\}$  be a bounded sequence generated by (3.13) and converges to a critical point  $\mathbf{z}^*$ . Then the following statements hold:

1. If  $\theta = 0$ , then sequence  $\{\mathbf{z}^k\}$  converges to  $\mathbf{z}^*$  after a finite number of steps,
2. If  $\theta \in (0, \frac{1}{2}]$ , then there exist  $Q > 0$  and  $\gamma \in (0, 1)$  such that  $\|\mathbf{z}^k - \mathbf{z}^*\| \leq Q\gamma^k$ ,
3. If  $\theta \in (\frac{1}{2}, 1)$ , then there exists a constant  $Q > 0$  such that  $\|\mathbf{z}^k - \mathbf{z}^*\| \leq Q(k - L)^{\frac{1-\theta}{1-2\theta}}$  for all  $k > L$ .

**Proof.** We define  $\Delta_k = \sum_{i=k}^{+\infty} \|\mathbf{z}^{i+1} - \mathbf{z}^i\|$ , then for any  $k > 0$  the value of  $\Delta_k$  is finite by Theorem 3.2. The triangle inequality implies that  $\Delta_k \geq \|\mathbf{z}^{+\infty} - \mathbf{z}^k\|$ . Assume that  $\Delta_k > 0$  for all  $k > 0$ . Note that  $\varphi(s) = \omega s^{1-\theta}$  decreases to zero. By (3.30), we obtain

$$\Delta_{L+1} \leq (\Delta_L - \Delta_{L+1}) + C\varphi(\Psi(\mathbf{z}^{L+1}) - \Psi(\mathbf{z}^*)) = (\Delta_L - \Delta_{L+1}) + \omega C(\Psi(\mathbf{z}^{L+1}) - \Psi(\mathbf{z}^*))^{1-\theta}. \quad (3.32)$$

Setting  $M = d^+ + \lambda_1^+ + \lambda_2^+$  and using (3.28), we obtain

$$\omega(1-\theta)(\Psi(\mathbf{z}^{L+1}) - \Psi(\mathbf{z}^*))^{-\theta} \geq \frac{1}{M} \|\mathbf{z}^{L+1} - \mathbf{z}^L\|^{-1}.$$

This, together with (3.32), implies that

$$\Delta_{L+1} \leq (\Delta_L - \Delta_{L+1}) + \omega^{\frac{1}{\theta}} C(M(1-\theta))^{\frac{1-\theta}{\theta}} \|\mathbf{z}^{L+1} - \mathbf{z}^L\|^{\frac{1-\theta}{\theta}} = (\Delta_L - \Delta_{L+1}) + C_1(\Delta_L - \Delta_{L+1})^{\frac{1-\theta}{\theta}},$$

where  $C_1 = \omega^{\frac{1}{\theta}} C(M(1-\theta))^{\frac{1-\theta}{\theta}}$ .

**Case 1:**  $\theta \in (0, \frac{1}{2}]$ . It is easy to see that  $\frac{1-\theta}{\theta} \geq 1$ . Since there exists an integer  $K$  such that  $0 \leq \Delta_{L-1} - \Delta_L < 1$  for all  $L > K$  holds, we obtain

$$\Delta_{L+1} \leq (1 + C_1)(\Delta_L - \Delta_{L+1}).$$

This implies that  $\Delta_{L+1} \leq \frac{1+C_1}{2+C_1} \Delta_L$ . Choosing  $\gamma = \frac{1+C_1}{2+C_1}$ , the item 2 immediately follows.

**Case 2:**  $\theta \in (\frac{1}{2}, 1)$ , thus  $\frac{1-\theta}{\theta} \in (0, 1)$ . Similarly, we have  $(\Delta_{L-1} - \Delta_L)^{\frac{1-\theta}{\theta}} \geq \Delta_{L-1} - \Delta_L$ , which implies that

$$\Delta_{L+1} \leq (1 + C_1)(\Delta_L - \Delta_{L+1})^{\frac{1-\theta}{\theta}}, \quad \text{for all } L > K.$$

We define  $J(s) = s^{-\frac{\theta}{1-\theta}}$ , then

$$\begin{aligned} 1 &\leq (1 + C_1)^{\frac{\theta}{1-\theta}} (\Delta_L - \Delta_{L+1}) \Delta_{L+1}^{-\frac{\theta}{1-\theta}} \\ &= (1 + C_1)^{\frac{\theta}{1-\theta}} \left( \frac{\Delta_L}{\Delta_{L+1}} \right)^{\frac{\theta}{1-\theta}} (\Delta_L - \Delta_{L+1}) J(\Delta_L) \\ &\leq (1 + C_1)^{\frac{\theta}{1-\theta}} \left( \frac{\Delta_L}{\Delta_{L+1}} \right)^{\frac{\theta}{1-\theta}} \int_{\Delta_{L+1}}^{\Delta_L} J(s) ds \\ &= \frac{1-\theta}{1-2\theta} (1 + C_1)^{\frac{\theta}{1-\theta}} \left( \frac{\Delta_L}{\Delta_{L+1}} \right)^{\frac{\theta}{1-\theta}} \left( \Delta_L^{\frac{1-2\theta}{1-\theta}} - \Delta_{L+1}^{\frac{1-2\theta}{1-\theta}} \right). \end{aligned}$$

Therefore,

$$\Delta_{L+1}^{\frac{1-2\theta}{1-\theta}} - \Delta_L^{\frac{1-2\theta}{1-\theta}} \geq \frac{2\theta-1}{1-\theta} (1 + C_1)^{\frac{1-\theta}{\theta}} \left( \frac{\Delta_{L+1}}{\Delta_L} \right)^{\frac{\theta}{1-\theta}}.$$

Setting  $\nu = \frac{1-2\theta}{1-\theta} < 0$ , we have

$$\Delta_{L+1}^{\nu} - \Delta_L^{\nu} \geq \frac{-\nu}{(1 + C_1)^{1-\nu}} \left( \frac{\Delta_{L+1}}{\Delta_L} \right)^{1-\nu}.$$

Let  $\mu$  be the positive constant such that

$$\frac{-\nu}{(1 + C_1)^{1-\nu}} \mu = \mu^{\frac{\nu}{1-\nu}} - 1.$$

Then the above equation has a unique solution  $0 < \mu < 1$ . Assume that  $\left( \frac{\Delta_{L+1}}{\Delta_L} \right)^{1-\nu} > \mu$ , we thus have

$$\Delta_{L+1}^{\nu} - \Delta_L^{\nu} > \mu^{\frac{\nu}{1-\nu}} - 1.$$

Moreover, we have the following estimation for  $\left( \frac{\Delta_{L+1}}{\Delta_L} \right)^{1-\nu} \leq \mu$ ,

$$\begin{aligned} \left( \frac{\Delta_{L+1}}{\Delta_L} \right)^{1-\nu} \leq \mu &\Rightarrow \Delta_{L+1} \leq \mu^{\frac{1}{1-\nu}} \Delta_L \Rightarrow \Delta_{L+1}^{\nu} \geq \mu^{\frac{\nu}{1-\nu}} \Delta_L^{\nu} \\ &\Rightarrow \Delta_{L+1}^{\nu} - \Delta_L^{\nu} \geq (\mu^{\frac{\nu}{1-\nu}} - 1) \Delta_L^{\nu} \\ &\Rightarrow \Delta_{L+1}^{\nu} - \Delta_L^{\nu} \geq \mu^{\frac{\nu}{1-\nu}} - 1, \end{aligned}$$

where the last inequality follows from  $\Delta_L \leq 1$  (for  $L$  sufficiently large). Therefore,

$$\Delta_{i+1}^v - \Delta_i^v \geq \mu^{\frac{v}{1-v}} - 1, \quad \text{for all } i > K. \quad (3.33)$$

Summing up (3.33) from  $i = L$  to  $i = k - 1$  leads to

$$\Delta_k^v \geq \Delta_k^v - \Delta_L^v \geq \left( \mu^{\frac{v}{1-v}} - 1 \right) (k - L).$$

The item 3 now immediately follows from

$$\Delta_k \leq \left[ \left( \mu^{\frac{v}{1-v}} - 1 \right) (k - L) \right]^{\frac{1}{v}} \leq Q (k - L)^{\frac{1-\theta}{1-2\theta}},$$

where  $Q = (\mu^{\frac{v}{1-v}} - 1)^{\frac{1-\theta}{1-2\theta}}$ .

**Case 3:** If  $\theta = 0$ , we have  $\varphi(s) = cs$  and  $\varphi'(s) = c$ . Using (3.28), we obtain

$$\|\mathbf{z}^{k+1} - \mathbf{z}^k\| \geq \frac{1}{cM}. \quad (3.34)$$

It follows from (3.14) that

$$\|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 \leq \frac{2}{\rho_1} (\Psi(\mathbf{z}^k) - \Psi(\mathbf{z}^{k+1})). \quad (3.35)$$

The above inequalities (3.34) and (3.35), together with the fact that  $\Psi(\mathbf{z}^k)$  decreases to zero, show that sequence  $\{\mathbf{z}^k\}$  converges after a finite number of steps.  $\square$

Since the PAIHT method can be regarded as a special application case of the PPALM method and the nonconvex model (1.1) satisfies the KL property, the sequence generated by the PAIHT method also satisfies Lemma 3.4. Furthermore, we have the following corollary:

**Corollary 3.1.** Let  $\{(\mathbf{x}^k, \mathbf{y}^k)\}_{k=1}^\infty$  be the sequence generated by the proximal alternating iterative hard thresholding method (1.2). Then the sequence  $\{(\mathbf{x}^k, \mathbf{y}^k)\}_{k=1}^\infty$  converges to a critical point of (1.1).

#### 4. Numerical results

In this section, some numerical experiments are carried out to illustrate the proposed models and algorithms for image restoration. The experiments are performed under Windows 7 and MATLAB R2015a running on a notebook computer equipped with an Intel Core i7-3610QM CPU (2.30 GHz) and 8 GB RAM memory. In Section 3.1, it is found that the iterative scheme (3.3) becomes the projected gradient scheme (3.10) after some steps, which also motivates us to apply the accelerated proximal gradient method in [17] to solve (3.3). Algorithm 1 presents the summary of the accelerated iteration scheme. The piecewise spline tight wavelet frame constructed in [18] is chosen in numerical experiments for all the models based on wavelet frame. The level of the wavelet frame decomposition is set to be four.

To verify the benefit of the accelerated step, the Algorithm 1 is firstly compared with the iteration scheme (3.2) and the BCD method by solving the model (1.1) for image restoration problem. The test image size of Cameraman is  $256 \times 256$ , and the image is scaled into the range between 0 and 1. The clean image is degraded by going through a motion kernel of  $15 \times 30$  (applied with the MATLAB functions `imfilter` and `fspecial`) and following an additive zero-mean white Gaussian noise with a standard deviation  $3/255$ . The regularization parameters of the model (1.1) are set as  $\lambda = 10^{-5}$  and  $\eta = 0.2$ . The regularization parameter  $d_k$  of the Algorithm 1 and iteration scheme (3.2) are set to be  $0.01\eta$ . The energy values of the three methods with iterations from 50 to 400 are described in Fig. 1. It is observed that the accelerated step helps to accelerate the convergence speed of the method and the parameters  $d_k$  keep the energy curve more stable.

To illustrate the performance of the image restoration model (1.1), this paper compares it with the analysis model in [19] and the  $l_1$ -based model in [3]:

$$\min_{\mathbf{x}, \mathbf{y}} \Psi_1(\mathbf{x}, \mathbf{y}) := \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \lambda \|\mathbf{y}\|_1 + \frac{\eta}{2} \|\mathcal{D}^T \mathbf{x} - \mathbf{y}\|^2. \quad (4.1)$$

The split Bregman algorithm is used to solve the analysis based approach (see e.g. [19]). By applying alternating minimization algorithm to solve the  $l_1$ -based model (4.1), we note that all the algorithms have similar deconvolution steps such as (3.1). For fair comparison, it is assumed that all algorithms are under circular boundary condition and solved by the fast Fourier transform. Therefore, the complexity of all the methods will be the same and the number of iterations can represent the speed of methods. It is revealed in [1,5] that initializations are important for  $l_0$ -based iterative schemes in some cases. For the Gaussian blur kernel, the penalty decomposition method in [5,20] is adopted to search a better local minimizer. In other words, the parameter  $\eta$  in Algorithm 1 may be updated if necessary. For other blur kernels, the Algorithm 1 is applied with

**Algorithm 1** Accelerated PAIHT scheme for (1.1)

**Input:** Let  $\eta > 0, \lambda > 0$ . Set  $k = 1, t_1 = 1$  and  $d_k \in [d_{\min}, d_{\max}], \mathbf{y}^1 = \hat{\mathbf{y}}^1 = 0$ .

1: **for**  $k = 1, 2, \dots$ , compute **do**

2:  $\lambda_k = \frac{2\lambda}{\eta + d_k}$

3:

$$\mathbf{x}^{k+1} = (\mathcal{A}^T \mathcal{A} + \eta \mathcal{L})^{-1} (\mathcal{A}^T \mathbf{b} + \eta \mathcal{D} \hat{\mathbf{y}}^k)$$

4:

$$\mathbf{y}^{k+1} \in \mathcal{H}_{\lambda_k} \left( \hat{\mathbf{y}}^k - \frac{\eta}{\eta + d_k} (\hat{\mathbf{y}}^k - \mathcal{D}^T \mathbf{x}^{k+1}) \right)$$

5:  $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$

6:

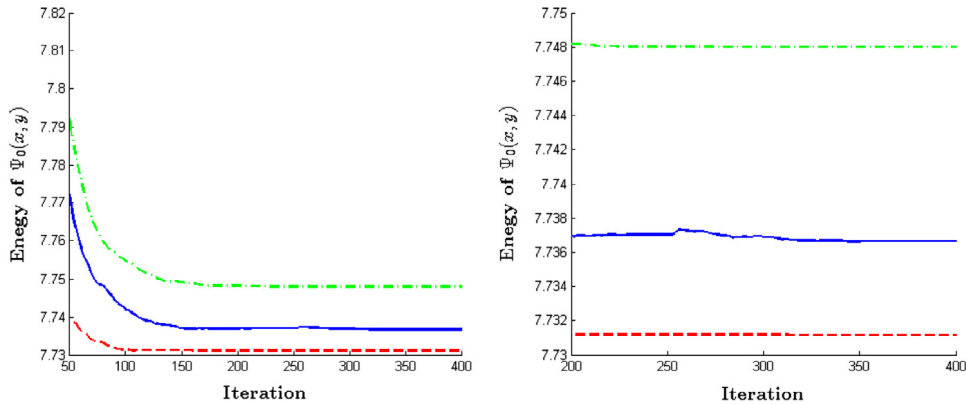
$$\hat{\mathbf{y}}^{k+1} = \mathbf{y}^{k+1} + \left( \frac{t_k - 1}{t_{k+1}} \right) (\mathbf{y}^{k+1} - \mathbf{y}^k)$$

7: Update  $d_{k+1}$  and  $\eta$  if necessary.

8: If stopping criterion satisfies go to Output.

9: **end for**

**Output:**  $\mathbf{x}^{k+1}, \mathbf{y}^{k+1}$ .



**Fig. 1.** Energy decay of the model (1.1). Green dash dot: by the iterative scheme (3.2) with  $d_k = 0.01\eta$ . Blue Line: by the BCD method in [5]. Red dash: by Algorithm 1 with  $d_k = 0.01\eta$ . Left: Iteration from 50 to 400; Right: A zoom-in view. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the fixed  $\eta$ . This paper sets  $d_k = 0.01\eta$  for Algorithm 1. The stopping criteria for the Algorithm 1 and the model (4.1) are as below,

$$\frac{|\Psi_i(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}) - \Psi_i(\mathbf{x}^k, \mathbf{y}^k)|}{\max(|\Psi_i(\mathbf{x}^{k+1}, \mathbf{y}^{k+1})|, 1)} \leq 10^{-3}, \quad i = 0, 1.$$

The stopping criterion for split Bregman algorithm is set default, while all the other parameters are manually tuned for different blur kernels to obtain the best restoration results. Once the blur kernel is fixed, the parameters will be fixed. To measure the quality of restored images, the Peak Signal-to-Noise Ratio (PSNR) value is utilized, which is defined by

$$\text{PSNR} = 10 \log_{10} \left( \frac{mn}{\|\hat{\mathbf{x}} - \mathbf{x}\|^2} \right),$$

where  $\mathbf{x} \in \mathbb{R}^{m \times n}$  denotes the restored image and  $\hat{\mathbf{x}} \in \mathbb{R}^{m \times n}$  denotes the clean image. All three methods are tested on six different images. The results are reported in Table 1. It is observed that the Algorithm 1 can mostly achieve the best reconstruction qualities measured by PSNR values and the speed of the proposed algorithm is the fastest. From the zoom-in views of comparison results in Fig. 2, it is observed that the restoration results of the proposed algorithm products present fewer artifacts compared with other methods.

**Table 1**Numerical results for image deblurring with noise level  $\sigma = 5/255$ .

Image	Kernel	Alg. In [19]	Iter	Model (4.1)	Iter	Alg. 1	Iter
Cameraman256	('average', 9)	24.33	80	24.29	67	<b>24.65</b>	15
	('disk', 3)	25.81	71	25.73	65	<b>25.92</b>	21
	('gaussian', 25, 2)	24.58	80	24.54	48	<b>24.62</b>	37
	('motion', 15, 30)	25.21	74	25.10	78	<b>25.53</b>	26
House256	('average', 9)	29.81	69	29.61	52	<b>30.07</b>	10
	('disk', 3)	31.06	62	30.81	52	<b>31.07</b>	9
	('gaussian', 25, 2)	30.00	69	29.96	36	<b>30.60</b>	28
	('motion', 15, 30)	29.65	69	29.20	63	<b>30.45</b>	16
Peppers256	('average', 9)	26.51	84	26.26	81	<b>26.79</b>	14
	('disk', 3)	26.74	72	26.56	71	<b>26.82</b>	16
	('gaussian', 25, 2)	25.12	79	25.08	44	<b>25.39</b>	37
	('motion', 15, 30)	25.63	75	25.65	82	<b>26.24</b>	23
Boat512	('average', 9)	26.83	75	26.78	52	<b>27.06</b>	12
	('disk', 3)	<b>28.51</b>	67	28.39	50	28.42	13
	('gaussian', 25, 2)	<b>27.57</b>	78	27.52	37	27.49	34
	('motion', 15, 30)	26.99	64	26.80	61	<b>27.40</b>	18
Flintstones512	('average', 9)	22.25	86	22.14	83	<b>22.77</b>	21
	('disk', 3)	<b>25.48</b>	77	25.31	78	24.92	29
	('gaussian', 25, 2)	24.08	93	24.03	69	<b>24.26</b>	51
	('motion', 15, 30)	23.88	78	23.76	92	<b>24.24</b>	36
Lena512	('average', 9)	29.22	75	29.11	42	<b>29.30</b>	10
	('disk', 3)	31.21	66	30.89	42	<b>31.29</b>	10
	('gaussian', 25, 2)	30.49	77	30.44	30	<b>30.62</b>	29
	('motion', 15, 30)	29.38	63	29.11	53	<b>29.88</b>	16

**Fig. 2.** Zoom-in views of the restoration results. Top: Image (Lena) is corrupted by  $25 \times 2$  Gaussian kernel and zero mean white Gaussian noise with  $\sigma = 5/255$ . Bottom: Image (House) is corrupted by  $15 \times 30$  motion kernel and zero mean white Gaussian noise with  $\sigma = 5/255$ .

## 5. Conclusions

This paper has proposed the PAIHT method for solving the nonconvex frame based image restoration model (1.1). By considering the nonconvex model (1.1) as a special case of the more general nonconvex and nonsmooth model (1.3), this paper furthermore develops the PPALM method for solving (1.3) and proves its convergence. Since the PAIHT method can be regarded as a special case of the PPALM method, some convergence properties of the PAIHT method can be obtained. Numerical experiments show that the proposed image restoration algorithm often performs better than some of the other convex image restoration algorithms. However, the accelerated step in Algorithm 1 is still lack of theoretical analysis, which can be our future research directions.

## Acknowledgments

The authors are grateful to the editor and anonymous reviewers for their constructive suggestions and comments to improve the paper, especially the proof of [Theorem 3.1](#). The authors also thank J.-F. Cai and B. Dong for providing the source codes for the image-restoration algorithms in [\[5,19\]](#) respectively.

## References

- [1] C. Bao, B. Dong, L. Hou, Z. Shen, X. Zhang, X. Zhang, Extrapolated proximal iterative hard thresholding methods for wavelet frame based image restoration, manuscript, 2015.
- [2] C. Bao, H. Ji, Z. Shen, Convergence analysis for iterative data-driven tight frame construction scheme, *Appl. Comput. Harmon. Anal.* 38 (3) (2015) 510–523.
- [3] A. Danielyan, V. Katkovnik, K. Egiazarian, BM3D frames and variational image deblurring, *IEEE Trans. Image Process.* 21 (4) (2012) 1715–1728.
- [4] J. Portilla, Image restoration through  $l_0$  analysis-based sparse optimization in tight frames, in: 16th IEEE International Conference on Image Processing, 2009.
- [5] Y. Zhang, B. Dong, Z. Lu,  $l_0$  minimization of wavelet frame based image restoration, *Math. Comp.* 82 (282) (2013) 995–1015.
- [6] Z. Lu, Iterative hard thresholding methods for  $l_0$  regularized convex cone programming, *Math. Program.* 147 (1–2) (2014) 125–154.
- [7] J. Bolte, S. Sabach, M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, *Math. Program.* 146 (1–2) (2014) 459–494.
- [8] R.T. Rockafellar, R.J.B. Wets, *Variational Analysis*, Springer Science & Business Media, 2009.
- [9] S. Łojasiewicz, Une propriété topologique des sous ensembles analytiques réels, in: *Les Équations aux Dérivées Partielles*, Éditions du centre National de la Recherche Scientifique, Paris, 1963, pp. 87–89.
- [10] K. Kurdyka, On gradients of functions definable in o-minimal structures, *Ann. Inst. Fourier* 48 (3) (1998) 769–783.
- [11] J. Bolte, A. Daniilidis, A. Lewis, The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems, *SIAM J. Optim.* 17 (4) (2007) 1205–1223.
- [12] J. Bolte, A. Daniilidis, O. Ley, L. Mazet, Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity, *Trans. Amer. Math. Soc.* 362 (6) (2010) 3319–3363.
- [13] H. Attouch, J. Bolte, On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, *Math. Program.* 116 (1–2) (2009) 5–16.
- [14] H. Attouch, J. Bolte, P. Redont, A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka–Łojasiewicz inequality, *Math. Oper. Res.* 35 (2) (2010) 438–457.
- [15] Y. Xu, W. Yin, A block coordinate descent method for multi-convex optimization with applications to nonnegative tensor factorization and completion, *SIAM J. Imag. Sci.* 6 (3) (2012) 1758–1789.
- [16] G.H. Golub, C.F. Van Loan, *Matrix Computations*, JHU Press, 2012.
- [17] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imag. Sci.* 2 (1) (2009) 183–202.
- [18] A. Ron, Z. Shen, Affine systems in  $L_2(\mathbb{R}^d)$ : the analysis of the analysis operator, *J. Funct. Anal.* 148 (2) (1997) 408–447.
- [19] J.F. Cai, S. Osher, Z. Shen, Split Bregman methods and frame based image restoration, *Multiscale Model. Simul.* 8 (2) (2009) 337–369.
- [20] Z. Lu, Y. Zhang, Sparse approximation via penalty decomposition methods, *SIAM J. Optim.* 23 (4) (2013) 2448–2478.
- [21] T. Blumensath, M.E. Davies, Iterative thresholding for sparse approximations, *J. Fourier Anal. Appl.* 14 (5–6) (2008) 629–654.