Journal of Global Optimization (2020) 76:665–693
https://doi.org/10.1007/s10898-019-00828-4

A hybrid Bregman alternating direction method of multipliers for the linearly constrained difference-of-convex problems

Kai Tu[1] · Haibin Zhang[1] · Huan Gao[2] · Junkai Feng[1]

Received: 6 September 2018 / Accepted: 24 August 2019 / Published online: 29 August 2019

Abstract
In this paper, we propose a hybrid Bregman alternating direction method of multipliers for solving the linearly constrained difference-of-convex problems whose objective can be written as the sum of a smooth convex function with Lipschitz gradient, a proper closed convex function and a continuous concave function. At each iteration, we choose either subgradient step or proximal step to evaluate the concave part. Moreover, the extrapolation technique was utilized to compute the nonsmooth convex part. We prove that the sequence generated by the proposed method converges to a critical point of the considered problem under the assumption that the potential function is a Kurdyka–Łojasiewicz function. One notable advantage of the proposed method is that the convergence can be guaranteed without the Lischitz continuity of the gradient function of concave part. Preliminary numerical experiments show the efficiency of the proposed method.

Keywords Linearly constrained difference-of-convex problems · Bregman distance · Alternating direction method of multipliers · Kurdyka–Łojasiewicz function

1 Introduction

Denote by $\mathbb{R}^d$ the d-dimensional Euclidean space with inner product $\langle \cdot, \cdot \rangle$ and Euclidean norm $\| \cdot \|$. In this paper, we consider the linearly constrained difference-of-convex problem:

✉ Huan Gao
huanhuan135213@163.com

Kai Tu
kaitu_02@163.com

Haibin Zhang
zhanghaibin@bjut.edu.cn

Junkai Feng
fengjunkai1989@126.com

1 College of Applied Sciences, Beijing University of Technology, Beijing 100124, China

2 College of Mathematics and Computational Science, Hunan First Normal University, Changsha 410205, Hunan, China

$$\begin{aligned} \min_{x,y} \quad & f(x) + g(y), \\ \text{s.t.} \quad & Ax + By = b, \end{aligned} \tag{1}$$

where $g : \mathbb{R}^{n_2} \to \mathbb{R}$ is a continuously differentiable convex function, $A \in \mathbb{R}^{m\times n_1}$ and $B \in \mathbb{R}^{m\times n_2}$ are two given matrixes, $b \in \mathbb{R}^m$ is a given vector, and

$$f(x) = f_1(x) - f_2(x), \tag{2}$$

with $f_1, f_2 : \mathbb{R}^{n_1} \to \mathbb{R}$ are two continuous convex functions. Problem (1)–(2) has been encountered in many fields, in particular in machine learning, signal and image processing communities, and hence many numerical methods have been proposed, see [1,9,19,20,31,33–35,44,45,54].

The alternating direction method of multipliers (ADMM) is a powerful tool for solving problem (1), since it reduces the scale of original problem by decomposing it into two subproblems with smaller scales. The success of ADMM is mainly attributed to its subproblems with closed-form solutions. We refer the readers to [12,22,28,41,50,51,53] for the application of ADMM in both convex or nonconvex cases.

The convergence of ADMM and its variants in the convex case are studied intensively in the literature, as well as the results on convergence rate, see e.g. [15–18,24,26]. It is generally very difficult to establish the theoretical convergence of ADMM without the convexity. As we know, a particular condition to guarantee the convergence for nonconvex optimization problems is that the objective function satisfies the Kurdyka–Łojasiewicz (KL) inequality (See Definition 3 below in Sect. 2). For example, if the augmented Lagrangian function is a KL function, then the sequence generated by ADMM converges to a critical point of problem (1), see e.g. [23,27,48,50,51,53].

Nevertheless, in many convex or nonconvex scenarios, there are either without closed-form solutions for the subproblems, or the subproblems are hard to be solved exactly. The Bregman ADMM (BADMM) has thus been proposed for these situations, since an appropriate choice of Bregman distance does simplify the subproblems of ADMM, see [30,45,48,49]. The main difference between the BADMM and the ADMM is that the objection function of subproblems are replaced by the sum of the Bregman distance function and the augmented Lagrangian function.

When BADMM was directly applied to solve problem (1)–(2), at each iteration, one has to solve a difference-of-convex (DC) optimization problem whose objective functions can be written as the difference of two proper closed convex functions. Clearly, these subproblems can be solved by the DC type methods [5,21,29,32,34–36,46,52]. However, when they were applied to solve the related DC subproblem, one may get a critical point rather than the "real" minimizer. Thus, it is difficult to determine the stopping criterion of these nonconvex subproblems.

We now recall some well-known DC type methods for the DC optimization problems. One classical approach is the DC algorithm (DCA) proposed by Pham and An in [46]. By using a specific DC decomposition described in [47], the proximal DCA (pDCA) was thus obtained in [21]. For possibly accelerating pDCA, Wen et al. [52] incorporated the extrapolation technique [7] into proximal DCA and named the resulting method as the proximal DCA with extrapolation (pDCA_e). Later on, based on the Fenchel-Young inequality, Liu et al. [29] defined a new potential function and refined the convergence analysis of pDCA_e to cover the case that the concave part is nonsmooth. One commom palce of DCA, pDCA and pDCA_e is that the subgradient step is used to evaluate the concave part. In convex optimization, from [7] and Theorem 3.2.3 of [38], better convergence rates can be guaranteed for proximal algorithms than for subgradient algorithms. Thus, Banert and Bot [5] utilized the proximal step to evaluate the concave part and proposed a general double-proximal gradient algorithm (DPGA) for a generalized DC problem. Note that [5] also established the convergence theorem of DPGA without the smoothness assumption on the concave part.

Aim to deal with the difficulty of extended BADMM, Sun et al. [45] replaced the concave part of the objective by a linear majorant and proposed an Bregman alternating direction method of multipliers with difference of convex (BADMM-DC) for problem (1)–(2). At each iteration, the subproblems of BADMM-DC are convex, and can be "really" minimized. Suppose that the concave part is a differentiable function with Lipschitz continuous gradient, [45] proved that the whole sequence generated by BADMM-DC converges to a critical point of the augmented Lagrangian function.

However, the concave part is possibly nonsmooth in many practical problems, such as compressed sensing and image deblurring, see [5,29,45]. We also note that BADMM-DC can be regarded as the combination of DCA and BADMM, i.e., one uses the DCA to solve the resulting DC subproblems with one inner iteration at each iteration. Thus, a nature question is: Can we combine other DC type methods, such as pDCA$_e$ and DPGA, with BADMM to accelerate the BADMM-DC witout the Lipschitz continuity assumption on the gradient of the concave part? This paper is devoted to this.

Inspired and motivated by the above results, we combine pDCA$_e$ and DPGA with BADMM to propose a hybrid Bregman alternating direction method of multipliers for solving problem (1)–(2). At each iteration, we choose either subgradient step or proximal step to evaluate the concave part. Moreover, the extrapolation technique was utilized to compute the nonsmooth convex part. Based on the Fenchel-Young inequality, we define a new potential function and prove that the sequence generated by the new method converges to a critical point of problem (1)–(2) under the assumption that the new potential function is a KL function. One notable advantage is that the convergence theorem can be guaranteed without the Lischitz continuity of the gradient of concave part. Moreover, preliminary numerical experiments show that the new proposed method performs well, faster than BADMM-DC.

2 Preliminaries

In this section, we list some basic notations, related definitions and recall some well-known conclusions that will be used in our subsequent analysis.

First, let $C \in \mathbb{R}^{d\times d}$ be a positive definite matrix, and use the operator $\lambda_{\min}(C)$ to denote the smallest eigenvalue of C. Denote $\mathcal{I}$ as the identity matrix. For two arbitrary vectors $v_1 \in \mathbb{R}^n$ and $v_2 \in \mathbb{R}^m$, we simply use $v = (v_1, v_2)$ to denote their adjunction. That is, (v_1, v_2) denotes the vector $(v_1^T, v_2^T)^T$. If $F : \mathbb{R}^d \to 2^{\mathbb{R}^M}$ is a set-valued mapping, its graph is defined by Graph $F = \{(z, u) \in \mathbb{R}^d \times \mathbb{R}^M : u \in F(z)\}$. Given a nonempty closed subset Ω of $\mathbb{R}^d$, let $\mathrm{dist}(u, \Omega)$ denote the distance from u to Ω, i.e., $\mathrm{dist}(u, \Omega) = \inf\{\|u - v\| \mid v \in \Omega\}$. Given a function $h : \mathbb{R}^d \to (-\infty, \infty]$, we denote by $\mathrm{dom}(h)$ the domain of h, namely, $\mathrm{dom}(h) = \{z \in \mathbb{R}^d : h(z) < \infty\}$.

Then, we recall some useful notations and results in convex analysis. Suppose that h is a proper closed convex function. Let h^* denote the convex conjugate of h, i.e., $h^*(\xi) = \sup_{z\in\mathbb{R}^d}\{\langle \xi, z\rangle - h(z)\}$. Then h^* is proper closed convex. Moreover, the Fenchel-Young inequality holds, relating h, h^* and their subgradients: for any z and ξ, it holds that $h(z) + h^*(\xi) \geq \langle z, \xi\rangle$, and the equality holds if and only if $\xi \in \partial h(z)$. Given $v \in \mathbb{R}^d$ and $t > 0$, the proximal operator associated with h, which we denoted by $\mathrm{prox}_{th}(v)$, is defined as the unique minimizer of function $z \to th(z) + \frac{1}{2}\|z - v\|^2$, i.e., $\text{prox}_{th}(v) = \arg\min_{z \in \mathbb{R}^d} \left\{th(z) + \frac{1}{2}\|z - v\|^2\right\}$. The function h is strongly convex, if there exists $\mu > 0$ such that for all $z_1 \in \text{dom}(\partial h)$ and $z_2 \in \mathbb{R}^d$,

$$h(z_2) \geq h(z_1) + \langle \xi, z_2 - z_1 \rangle + \frac{\mu}{2}\|z_2 - z_1\|^2, \quad \forall \xi \in \partial h(z_1).$$

The convex parameter of a function is the largest μ such that the above condition holds. For a convex differentiable function ϕ, the Bregman distance is defined as

$$B_\phi(z_1, z_2) = \phi(z_1) - \phi(z_2) - \langle \nabla\phi(z_2), z_1 - z_2 \rangle.$$

If $\phi(z) = \|z\|^2$, we have that $B_\phi(x, y) = \|x - y\|^2$. If $\phi(z) = z^T C z$, where C is a positive semidefinite matrix, then, $B_\phi(x, y) = (x - y)^T C(x - y)$. We list some useful properties about the Bregman distance as follows.

Proposition 1 *Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be a convex differentiable function and $B_\phi(z_1, z_2)$ be the Bregman distance. Then,*

(i) $B_\phi(z_1, z_2) \geq 0$, $B(z_1, z_1) = 0$, *for any* $z_1, z_2 \in \mathbb{R}^d$;
(ii) *For a fixed* z_2, $B_\phi(\cdot, z_2)$ *is convex;*
(iii) $B_\phi(z_1, z_2) \geq \frac{\upsilon_\phi}{2}\|z_1 - z_2\|^2$, *for any* $z_1, z_2 \in \mathbb{R}^d$, *where* υ_ϕ *is the convex parameter of* ϕ;
(iv) *If* $\nabla\phi$ *is Lipschitz continuous with modulus* L_ϕ, *we have that* $B_\phi(z_1, z_2) \leq \frac{L_\phi}{2}\|z_1 - z_2\|^2$, *for any* $z_1, z_2 \in \mathbb{R}^d$.

Next, we recall a few definitions concerning subdifferentiable calculus for nonsmooth functions, and refer to [39,42] for more details.

Definition 1 Let $h : \mathbb{R}^d \to (-\infty, \infty]$ be a proper lower semi-continuous function. The (limiting) subdifferentiable of h at $z \in \text{dom}(h)$ is given by

$$\partial h(z) = \Bigg\{\xi^* \in \mathbb{R}^d : \exists z_t \to z,\ h(z_t) \to h(z),\ \xi_t \to \xi^* \\ \text{with} \liminf_{y \to z_t} \frac{h(y) - h(z_t) - \langle \xi_t, y - z^t \rangle}{\|y - z^t\|} \geq 0\Bigg\}.$$

We also write $\text{dom}(\partial h) = \{z \in \mathbb{R}^d : \partial h(z) \neq \emptyset\}$. It is well known that if h is convex, the above definition coincides with the usual notion of subdifferentiable in convex analysis, i.e.,

$$\partial h(z) = \{\xi^* \in \mathbb{R}^d : h(v) - h(z) \geq \langle \xi^*, v - z \rangle,\ \text{for all } v \in \mathbb{R}^d\}$$

Lemma 1 *Let $h : \mathbb{R}^d \to (-\infty, \infty]$ be a proper lower semi-continuous function. Then we have the following assertions.*

(a) *Suppose that* $(z_k, \xi_k) \in$ *Graph* (∂h) *be a sequence that converges to* (z, ξ^*), *and that* $h(z_k)$ *converges to* $h(z)$, *then* $(z, \xi^*) \in$ *Graph* (∂h).
(b) *A necessary (but not sufficient) condition for* $z^* \in \mathbb{R}^d$ *to be a local minimizer of* h *is* $\mathbf{0} \in \partial h(z^*)$.
(c) *If* f *is a smooth function, then we have subdifferentiable sum rule*

$$\partial(f + h)(z) = \nabla f(z) + \partial h(z).$$

An important property of subdifferentiable calculus is given as follows.

Lemma 2 *Let $Q_1 : \mathbb{R}^n \times \mathbb{R}^m \to (-\infty, \infty]$ be a function of the form $Q_1(x, y) = f(x) + g(y) + H(x, y)$, where $H(x, y)$ is a continuously differentiable function and $f : \mathbb{R}^n \to (-\infty, \infty]$, $g : \mathbb{R}^m \to (-\infty, \infty]$ are lower semicontinuous functions. Then for all $(x, y) \in dom(Q_1) = dom(f) \times dom(g)$, we have that*

$$\partial Q_1(x, y) = \begin{pmatrix} \partial f(x) + \nabla_x H(x, y) \\ \partial g(y) + \nabla_y H(x, y) \end{pmatrix}.$$

We give the definition of critical points of problem (1)–(2).

Definition 2 Let f_2 be a continuous convex function. We say (x^*, y^*, λ^*) is a critical point of problem (1)–(2) if it satisfies

$$\begin{cases} A^T \lambda^* \in \partial f_1(x^*) - \partial f_2(x^*), \\ B^T \lambda^* = \nabla g(y^*), \\ Ax^* + By^* = b. \end{cases} \tag{3}$$

It follows from definition 1.3 in [22] and the Corollary 3.4 in [37] that under some suitable conditions, it can be shown that if (x^*, y^*) is a local minimum of problem (1)–(2), then there exists λ^* such that (x^*, y^*, λ^*) is a critical point of problem (1)–(2). Furthermore, we present a concrete example in Appendix to verify it.

We next recall the Kurdyka–Łojasiewicz (KL) property [2,3,10], which is satisfied by a wide variety of functions such as proper closed semialgebraic functions. KL property plays an important role in convergence analysis of many first-order methods; see, for example, [3,11,13].

For simplicity, for any $\eta > 0$, we use Ξ_η to denote the set of all concave continuous functions $\varphi : [0, \eta) \to [0, \infty)$ that are continuously differentiable on $(0, \eta)$ with positive derivatives and satisfy $\varphi(0) = 0$.

Definition 3 Let $h : \mathbb{R}^d \to (-\infty, \infty]$ be proper and lower semicontinuous. The function h is said to have the KL property at $\bar{z} \in \mathrm{dom}(\partial h)$ if there exist $\eta > 0$, a neighborhood U of $\bar{z}$ and a function $\varphi \in \Xi_\eta$ such that: For any $z \in U$ with $h(\bar{z}) < h(z) < h(\bar{z}) + \eta$, one has

$$\varphi'(h(z) - h(\bar{z}))\, \mathrm{dist}(\mathbf{0}, \partial h(z)) \geq 1.$$

If h satisfy the KL property at all points in $\mathrm{dom}(\partial h)$, then h is called a KL function.

Last, we recall the following result proved in Lemma 6 in [10] concerning the uniformized KL property.

Lemma 3 *Let Ω be a compact set and $F : \mathbb{R}^d \to (-\infty, \infty]$ be a proper and lower semicontinuous function. Suppose that F is constant on Ω and satisfy the KL property at each point of Ω. Then, there exist $\epsilon > 0$, $\eta > 0$ and $\varphi \in \Xi_\eta$ such that for all $\bar{z} \in \Omega$, one has*

$$\varphi'(F(z) - F(\bar{z}))\, dist(\mathbf{0}, \partial F(z)) \geq 1,$$

for all $z \in \{z \in \mathbb{R}^d : dist(z, \Omega) < \epsilon\} \cap \{z \in \mathbb{R}^d \mid F(\bar{z}) < F(z) < F(\bar{z}) + \eta\}$.

3 Hybrid Bregman alternating direction method of multipliers

In this section, we present the algorithm for problem (1)–(2) under the following standard assumption:

Assumption 1 Let υ_ϕ and υ_ψ denote the convex parameter of the convex function ϕ and ψ, respectively. f_1, f_2, g, A, B, ϕ and ψ satisfy

(a) $f_1 : \mathbb{R}^{n_1} \to \mathbb{R}$ and $f_2 : \mathbb{R}^{n_1} \to \mathbb{R}$ are two continuous convex functions;
(b) $g : \mathbb{R}^{n_2} \to \mathbb{R}$ is a continuously differentiable convex function with Lipschitz continuous gradient whose Lipschitz continuity modulus is L_g;
(c) ϕ (resp. ψ) is a convex differentiable function whose gradient function $\nabla\phi$ (resp. $\nabla\psi$) is Lipschitz continuous with modulus L_ϕ (resp. L_ψ);
(d) There is $\eta_0 > 0$ such that $\eta_0\|y\|^2 \leq \|B^T y\|^2$, for any $y \in \mathbb{R}^{n_2}$.

To describe the new proposed method, we introduce the following notations and functions. Define

$$\begin{cases} \eta_1 = \begin{cases} \frac{2(L_g+L_\psi)^2}{\eta_0}, & \text{if } L_\psi \neq 0, \\ \frac{L_g^2}{\eta_0}, & \text{otherwise}, \end{cases} & \eta_2 = \frac{2L_\psi^2}{\eta_0}, \\ b_1 = \frac{\theta_2+\upsilon_\psi}{2} - \frac{\eta_1+\eta_2}{\beta}, & b_2 = \frac{1}{2}(\theta_1 - L_\phi \alpha_{\max}^2), \end{cases} \tag{4}$$

and

$$\tilde{L}_\beta(\xi, x, y, \lambda) = Q(\xi, x) + g(y) - \langle \lambda, Ax + By - b\rangle + \frac{\beta}{2}\|Ax + By - b\|^2, \tag{5}$$

where θ_1 and θ_2 denote the convex parameter of objective function in subproblem (6b) and (6c), respectively, $Q(\xi, x) = f_1(x) + f_2^*(\xi) - \langle \xi, x\rangle$, $\beta > 0$ is the penalty parameter, $\lambda \in \mathbb{R}^m$ is the Lagrange multiplier. Clearly, $\theta_1 \geq \beta\lambda_{\min}(A^T A) + \upsilon_\phi$ and $\theta_2 \geq \beta\lambda_{\min}(B^T B) + \upsilon_\psi$.

We now formally present the hybrid Bregman alternating multipliers of method for solving problem (1)–(2).

Algorithm 1 Hybrid Bregman alternating direction method of multipliers

Input: Take initial point $(x_0, y_0, \lambda_0) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m$, $\xi_0 \in \mathbb{R}^{n_1}$, $x_{-1} = x_0$, and two convex differentiable functions ϕ, ψ, and three parameters $\beta > 0$, $r \geq 0$, $\alpha_k \in [0, \alpha_{\max}]$ such that $b_1 > 0$ and $b_2 > 0$, where b_1 and b_2 were defined in (4).

While a termination criterion is not met, **do**
S1 Set $u_k = x_k + \alpha_k(x_k - x_{k-1})$ and compute the next iteration by

$$\begin{cases} \xi_{k+1} = \arg\min_{\xi\in\mathbb{R}^{n_1}} f_2^*(\xi) - \langle x_k, \xi\rangle + \frac{r}{2}\|\xi - \xi_k\|^2, & \text{(6a)} \\ x_{k+1} = \arg\min_{x\in\mathbb{R}^{n_1}} \tilde{L}_\beta(\xi_{k+1}, x, y_k, \lambda_k) + B_\phi(x, u_k), & \text{(6b)} \\ y_{k+1} = \arg\min_{y\in\mathbb{R}^{n_2}} \tilde{L}_\beta(\xi_{k+1}, x_{k+1}, y, \lambda_k) + B_\psi(y, y_k), & \text{(6c)} \\ \lambda_{k+1} = \lambda_k - \beta(Ax_{k+1} + By_{k+1} - b). & \text{(6d)} \end{cases}$$

S2 Set $k = k + 1$.
end while

Output: (x_k, y_k)

Remark 1 (i) If we take $r \equiv 0$ and $\alpha_k \equiv 0$, Algorithm 1 becomes the BADMM-DC [45].
(ii) Note that Algorithm 1 chooses either subgradient step (i.e., $r = 0$) or proximal step (i.e., $r > 0$) to evaluate the concave part, and that the BADMM-DC just uses subgradient step to evaluate the concave part.

(iii) We emphasize that the convergence theorem of BADMM-DC need the assumption that the gradient of the concave part $-f_2$ is Lipschitz continuous, and that the convergence of Algorithm 1 can be guaranteed under the assumption that the concave part $-f_2$ is nonsmooth, see Theorem 2 in Sect. 4.

Last, we list the differences between pDCA$_e$, DPGA, BADMM-DC and Algorithm 1, when they are applied to a total variation image restoration problem [45]. This problem can be reformulated as

$$\min_{x,y} \ \rho\|Ky\|_1 - \rho\|Ky\|_2 + \frac{1}{2}\|Hy - y^0\|^2, \tag{7}$$

or

$$\min_{x,y} \ \rho\|x\|_1 - \rho\|x\|_2 + \frac{1}{2}\|Hy - y^0\|^2, \quad s.t. \quad x - Ky = \mathbf{0}, \tag{8}$$

where $\rho > 0$ is a regularization parameter, H is a blurred operator, $K : \mathbb{R}^n \to \mathbb{R}^{2n}$ is the discrete gradient operator. See Sect. 5 below for more details.

Let $f_2(x) = \rho\|x\|_2$, $h_1(y) = \rho\|Ky\|_2$ and $g(y) = \frac{1}{2}\|Hy - y^0\|^2$. Given the point y_k, the iterative form of pDCA$_e$ [29,52] for solving problem (7) is

$$\begin{cases} \xi_{k+1} \in \arg\min_{\xi\in\mathbb{R}^d} h_1^*(\xi) - \langle y_k, \xi\rangle, \\ y_{k+1} = \arg\min_y \rho\|Ky\|_1 + \frac{L}{2}\|y - (u_k - \frac{1}{L}(\nabla g(u_k) - \xi_{k+1}))\|^2, \end{cases}$$

where $u_k = y_k + \alpha_k(y_k - y_{k-1})$, $L = \|H\|^2$, with $\alpha_k \in [0, 1)$. By the first order optimal condition, we have that $\xi_{k+1} \in \partial h_1(y_k)$, which means that the concave part $-h_1$ is evaluated by subgradient step. Banert and Bot [5] proposed the DPGA for solving problem (7). Its iterative form is

$$\begin{cases} \xi_{k+1} = \arg\min_\xi f_2^*(\xi) - \langle Ky_k, \xi\rangle + \frac{1}{2\mu_k}\|\xi - \xi_k\|^2, \\ y_{k+1} = \arg\min_y \gamma_k\|Ky\|_1 + \frac{1}{2}\|y - [y_k - \gamma_k(\nabla g(y_k) - K^*\xi_{k+1})]\|^2, \end{cases}$$

where $\gamma_k > 0$ and $\mu_k > 0$ are suitable parameters. By the definition of proximal mapping and $\mu_k > 0$, the update of ξ_{k+1} in DPGA can be reformulated as $\xi_{k+1} = \mathrm{prox}_{\mu_k f_2^*}(\xi_k + \mu_k Ky_k)$. This means that the concave part is evaluated by proximal step.

When BADMM-DC [45] was applied to solve problem (8), it generates the next point $(\xi_{k+1}, x_{k+1}, y_{k+1}, \lambda_{k+1})$ by

$$\begin{cases} \xi_{k+1} = \arg\min_\xi f_2^*(\xi) - \langle x_k, \xi\rangle, \\ x_{k+1} = \arg\min_x \rho\|x\|_1 - \langle \xi_{k+1} + \lambda_k, x\rangle + \frac{\beta}{2}\|x - Ky_k\|^2 + B_\phi(x, x_k), \\ y_{k+1} = \arg\min_y g(y) + \langle K^*\lambda_k, y\rangle + \frac{\beta}{2}\|Ky - x_{k+1}\|^2 + B_\psi(y, y_k), \\ \lambda_{k+1} = \lambda_k - \beta(Ax_{k+1} + By_{k+1} - b). \end{cases}$$

The iterative form of Algorithm 1 for problem (8) is

$$\begin{cases} \xi_{k+1} = \arg\min_\xi f_2^*(\xi) - \langle x_k, \xi\rangle + \frac{r}{2}\|\xi - \xi_k\|^2, \\ x_{k+1} = \arg\min_x \rho\|x\|_1 - \langle \xi_{k+1} + \lambda_k, x\rangle + \frac{\beta}{2}\|x - Ky_k\|^2 + B_\phi(x, u_k), \\ y_{k+1} = \arg\min_y g(y) + \langle K^*\lambda_k, y\rangle + \frac{\beta}{2}\|Ky - x_{k+1}\|^2 + B_\psi(y, y_k), \\ \lambda_{k+1} = \lambda_k - \beta(Ax_{k+1} + By_{k+1} - b), \end{cases}$$

where $r \geq 0$, $u_k = x_k + \alpha_k(x_k - x_{k-1})$ with $\alpha_k \in [0, \alpha_{\max}]$ and $\theta_1 - L_\phi\alpha_{\max}^2 > 0$.

When pDCA$_e$, DPGA, BADMM-DC and Algorithm 1 were applied to solve the above mentioned problem, the main difference of these methods is that pDCA$_e$ and DPGA have to compute the proximal operator of $\|Ky\|_1$, which has no closed form solutions. Hence these methods are computationally expensive. Whereas BADMM-DC and Algorithm 1 just need to solve the proximal operator of l_1 norm, which has the closed form solutions.

Comparing Algorithm 1 with BADMM-DC, Algorithm 1 incooperate the extrapolation technique of pDCA$_e$ and the proximal step stragety of DPGA into BADMM to accelerate BADMM-DC. Moreover, problem (8) does not satisfy the assumptions of convergence theorem of BADMM-DC, since $\|x\|_2$ is nonsmooth. Thus, Sun et al. proved the following convergence theorem of BADMM-DC specially applied to problem (8), which was proved in Theorem 4 of [45].

Theorem 1 *Let K and H denote the discrete gradient operator and a linear operator, respectively. Assume that $H \succ 0$ and $\sigma < \frac{1}{2\|H\|^2}$. Let $c_1 = \lambda_{\min}(KK^*) > 0$ and set $\phi = \frac{r_1}{2}\|\cdot\|^2$, $\psi = \frac{r_2}{2}\|\cdot\|^2$,*

$$r_1 \geq 0,\ r_2 > 0 \text{ and } \max\{\frac{1}{\sigma c_1}, \frac{4r_2^2 + 4(\|H\|^2 + r_2)^2}{r_2 c_1}\} \leq \beta \leq \frac{2}{\sigma c_1}.$$

Let $\{\omega_k = (x_k, y_k, \lambda_k)\}_{k\in\mathbb{N}}$ be generated by BADMM-DC. Then, $\mathbf{0}$ is accumulation point of $\{x_k\}_{k\in\mathbb{N}}$ or $\{\omega_k\}_{k\in\mathbb{N}}$ converges to a critical point of problem (8).

Let us present the convergence theorems of Algorithm 1 particularly applied to problem (8), which will be proved in the Appendix.

Proposition 2 *Let K and H denote the discrete gradient operator and a linear operator, respectively. Assume that $H \succ 0$, $\sigma < \frac{1}{2\|H\|^2}$. Let $c_1 = \lambda_{\min}(KK^*) > 0$, and set $\phi = \frac{r_1}{2}\|\cdot\|^2$, $\psi = \frac{r_2}{2}\|\cdot\|^2$, $\alpha_k \in [0, 1)$, $r \geq 0$,*

$$r_1 \geq 0,\ r_2 \geq 0,\ \text{and } \beta > \max\{\frac{1}{\sigma c_1}, \frac{-c_2 + \sqrt{c_2^2 + 8c_1(\eta_1 + \eta_2)}}{2c_1}\}, \tag{9}$$

where $c_2 = c_1 + 2r_2$. Let $\{d_k = (\xi_k, x_k, x_{k-1}, y_k, y_{k-1}, \lambda_k)\}_{k\in\mathbb{N}}$ be generated by Algorithm 1. Then, the sequence $\{\omega_k = (x_k, y_k, \lambda_k)\}_{k\in\mathbb{N}}$ converges to a critical point of problem (8).

Remark 2 Note that for problem (8), Theorem 1 shows that the sequence generated by BADMM-DC may not converge to a critical point. While Proposition 2 shows that if β satisfy (9), the sequence generated by Algorithm 1 must converge to a critical point of problem (8). From Remark 1 (i), we know that BADMM-DC is a special case of Algorithm 1. This means that if we adjust the choice of β, then the sequence generated by BADMM-DC also must converge to a critical point of problem (8).

4 Convergence analysis

In this section, we prove the convergence of the hybrid Bregman alternating direction multipliers of method under the Assumption 1. The convergence analysis of Algorithm 1 is largely based on the following potential function

$$\Theta(\xi, x, \tilde{x}, y, \tilde{y}, \lambda) = \tilde{L}_\beta(\xi, x, y, \lambda) + \frac{\theta_1}{2}\|x - \tilde{x}\|^2 + \frac{\eta_2}{\beta}\|y - \tilde{y}\|^2, \tag{10}$$

where $\tilde{L}_\beta$ was defined in (5). Moreover, it follows from Fenchel-Young inequality that for any $(\xi, x, \tilde{x}, y, \tilde{y}, \lambda) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_1} \times \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_2} \times \mathbb{R}^m$, $\Theta(x, \xi, \tilde{x}, y, \tilde{y}, \lambda) \geq L_\beta(x, y, \lambda) + \frac{\theta_1}{2}\|x - \tilde{x}\|^2 + \frac{\eta_2}{\beta}\|y - \tilde{y}\|^2$.

We first present the first-order optimal conditions for the subproblems in Algorithm 1, which will be used repeatedly in our convergence analysis below.

$$\begin{cases} \mathbf{0} \in \partial f_2^*(\xi_{k+1}) - x_k + r(\xi_{k+1} - \xi_k), & \text{(11a)} \\ \mathbf{0} \in \partial f_1(x_{k+1}) - \xi_{k+1} - A^T\lambda_k + p_k + \nabla\phi(x_{k+1}) - \nabla\phi(u_k), & \text{(11b)} \\ \mathbf{0} = \nabla g(y_{k+1}) - B^T\lambda_k + q_k + \nabla\psi(y_{k+1}) - \nabla\psi(y_k), & \text{(11c)} \\ Ax_{k+1} + By_{k+1} - b = \dfrac{1}{\beta}(\lambda_k - \lambda_{k+1}), & \text{(11d)} \end{cases}$$

where $p_k = \beta A^T(Ax_{k+1} + By_k - b)$ and $q_k = \beta B^T(Ax_{k+1} + By_{k+1} - b)$.

Now, we begin our analysis with the following lemmas. Note that the proof of Lemma 4 is similar to Lemma 2 in [45]. For the completeness, we rewrite the proof.

Lemma 4 *Suppose that $\{d_k = (\xi_k, x_k, x_{k-1}, y_k, y_{k-1}, \lambda_k)\}_{k\in\mathbb{N}}$ is the sequence generated by Algorithm 1, and that Assumption 1 (b)–(d) hold. Then, for any $k \geq 1$, we have that*

$$\begin{aligned} &\Theta(\xi_{k+1}, x_{k+1}, x_k, y_{k+1}, y_k, \lambda_{k+1}) - \Theta(\xi_{k+1}, x_{k+1}, x_k, y_{k+1}, y_k, \lambda_k) \\ &\leq \frac{\eta_1}{\beta}\|y_{k+1} - y_k\|^2 + \frac{\eta_2}{\beta}\|y_k - y_{k-1}\|^2. \end{aligned}$$

Proof It follows from (6d) and (11c) that

$$\begin{cases} B^T\lambda_{k+1} = \nabla g(y_{k+1}) + \nabla\psi(y_{k+1}) - \nabla\psi(y_k), \\ B^T\lambda_k \quad = \nabla g(y_k) + \nabla\psi(y_k) - \nabla\psi(y_{k-1}), \end{cases} \tag{12}$$

and that

$$\begin{aligned} B^T(\lambda_{k+1} - \lambda_k) = \nabla g(y_{k+1}) - \nabla g(y_k) + \nabla\psi(y_{k+1}) - \nabla\psi(y_k) \\ - (\nabla\psi(y_k) - \nabla\psi(y_{k-1})). \end{aligned} \tag{13}$$

It follows that

$$\begin{aligned} \|\lambda_{k+1} - \lambda_k\| &\leq \frac{\|B^T(\lambda_{k+1} - \lambda_k)\|}{\sqrt{\eta_0}} \\ &\leq \frac{L_g + L_\psi}{\sqrt{\eta_0}}\|y_{k+1} - y_k\| + \frac{L_\psi}{\sqrt{\eta_0}}\|y_k - y_{k-1}\|, \end{aligned} \tag{14}$$

where the first inequality is from Assumption 1 (d), the second inequality follows from (13), the triangle inequality and Assumption 1 (b)–(c). Suppose that $L_\psi \neq 0$. By using that fact that $(h_1 + h_2)^2 \leq 2h_1^2 + 2h_2^2$ for any $h_1, h_2 \in \mathbb{R}$, we get that

$$\|\lambda_{k+1} - \lambda_k\|^2 \leq \eta_1\|y_{k+1} - y_k\|^2 + \eta_2\|y_k - y_{k-1}\|^2. \tag{15}$$

Using the inequality (14) and the definition of η_1 and η_2, the inequality (15) also holds under the case that $L_\psi = 0$. It follows from the definition of Θ, η_1, η_2 and (15) that

$$\begin{aligned} &\Theta(\xi_{k+1}, x_{k+1}, x_k, y_{k+1}, y_k, \lambda_{k+1}) - \Theta(\xi_{k+1}, x_{k+1}, x_k, y_{k+1}, y_k, \lambda_k) \\ &= \frac{1}{\beta}\|\lambda_k - \lambda_{k+1}\|^2 \leq \frac{\eta_1}{\beta}\|y_{k+1} - y_k\|^2 + \frac{\eta_2}{\beta}\|y_k - y_{k-1}\|^2. \end{aligned}$$

This completes the proof. □

Lemma 5 *Suppose that $\{d_k = (\xi_k, x_k, x_{k-1}, y_k, y_{k-1}, \lambda_k)\}_{k\in\mathbb{N}}$ is the sequence generated by Algorithm 1, and that Assumption 1 (b) and (c) holds. Then, for $k \geq 1$, we have that*

$$\begin{aligned}&\Theta(\xi_{k+1}, x_{k+1}, x_k, y_{k+1}, y_k, \lambda_k) - \Theta(\xi_{k+1}, x_{k+1}, x_k, y_k, y_{k-1}, \lambda_k)\\&\leq -\big(\frac{\theta_2 + \upsilon_\psi}{2} - \frac{\eta_2}{\beta}\big)\|y_{k+1} - y_k\|^2 - \frac{\eta_2}{\beta}\|y_k - y_{k-1}\|^2.\end{aligned}$$

Proof For simplify, let $F(y) = \tilde{L}_\beta(\xi_{k+1}, x_{k+1}, y, \lambda_k) + B_\psi(y, y_k)$. By Assumption 1 (b), we have that $\mathrm{dom}(F) = \mathbb{R}^{n_2}$. Since the convex parameter of F is θ_2, it follows from the smoothness of F that for all $v_1, v_2 \in \mathbb{R}^{n_2}$,

$$F(v_2) \geq F(v_1) + \langle \nabla F(v_1), v_2 - v_1\rangle + \frac{\theta_2}{2}\|v_2 - v_1\|^2. \tag{16}$$

Applying (16) with $v_1 = y_{k+1}$ and $v_2 = y_k$, we obtain

$$\begin{aligned}F(y_k) &\geq F(y_{k+1}) + \langle \nabla F(y_{k+1}), y_k - y_{k+1}\rangle + \frac{\theta_2}{2}\|y_{k+1} - y_k\|^2\\&= F(y_{k+1}) + \frac{\theta_2}{2}\|y_{k+1} - y_k\|^2,\end{aligned}\tag{17}$$

where the equality is follows from (11c) and the definition of F. It follows that

$$\begin{aligned}&g(y_{k+1}) - \langle \lambda_k, Ax_{k+1} + By_{k+1} - b\rangle + \frac{\beta}{2}\|Ax_{k+1} + By_{k+1} - b\|^2\\&\leq g(y_k) - \langle \lambda_k, Ax_{k+1} + By_k - b\rangle + \frac{\beta}{2}\|Ax_{k+1} + By_k - b\|^2\\&\quad - B_\psi(y_{k+1}, y_k) - \frac{\theta_2}{2}\|y_{k+1} - y_k\|^2\\&\leq g(y_k) - \langle \lambda_k, Ax_{k+1} + By_k - b\rangle + \frac{\beta}{2}\|Ax_{k+1} + By_k - b\|^2\\&\quad - \frac{\theta_2 + \upsilon_\psi}{2}\|y_{k+1} - y_k\|^2,\end{aligned}\tag{18}$$

where the first inequality follows from (17) and the definition of F and $\tilde{L}_\beta$, the last inequality is from Proposition 1 (iii). By the definition of Θ and (18), we have that

$$\begin{aligned}&\Theta(\xi_{k+1}, x_{k+1}, x_k, y_{k+1}, y_k, \lambda_k) - \Theta(\xi_{k+1}, x_{k+1}, x_k, y_k, y_{k-1}, \lambda_k)\\&\leq -(\frac{\theta_2 + \upsilon_\psi}{2} - \frac{\eta_2}{\beta})\|y_{k+1} - y_k\|^2 - \frac{\eta_2}{\beta}\|y_k - y_{k-1}\|^2,\end{aligned}$$

which completes the proof. □

Lemma 6 *Suppose that Assumption 1 (a) and (c) hold. Let the sequence $\{d_k = (\xi_k, x_k, x_{k-1}, y_k, y_{k-1}, \lambda_k)\}_{k\in\mathbb{N}}$ be generated by Algorithm 1. Then for $k \geq 1$, we have that*

$$\begin{aligned}&\Theta(\xi_{k+1}, x_{k+1}, x_k, y_k, y_{k-1}, \lambda_k) - \Theta(\xi_k, x_k, x_{k-1}, y_k, y_{k-1}, \lambda_k)\\&\leq -\frac{1}{2}\big(\theta_1 - L_\phi\alpha_{\max}^2\big)\|x_k - x_{k-1}\|^2 - r\|\xi_{k+1} - \xi_k\|^2.\end{aligned}$$

Proof For simplify, we define

$$H_k(x) = -\langle \lambda_k, Ax + By_k - b\rangle + \frac{\beta}{2}\|Ax + By_k - b\|^2. \tag{19}$$

By using the similar argument in Lemma 5, we have

$$
\begin{aligned}
& f_1(x_{k+1}) + H_k(x_{k+1}) - \langle \xi_{k+1}, x_{k+1} \rangle + B_\phi(x_{k+1}, u_k) \\
& \quad \leq f_1(x_k) + H_k(x_k) - \langle \xi_{k+1}, x_k \rangle + B_\phi(x_k, u_k) - \frac{\theta_1}{2} \|x_{k+1} - x_k\|^2.
\end{aligned}
$$

Rearranging the above inequality, it yields that

$$
\begin{aligned}
& f_1(x_{k+1}) + H_k(x_{k+1}) \\
& \quad \leq f_1(x_k) + H_k(x_k) - \langle \xi_{k+1}, x_k - x_{k+1} \rangle + B_\phi(x_k, u_k) \\
& \qquad - B_\phi(x_{k+1}, u_k) - \frac{\theta_1}{2} \|x_{k+1} - x_k\|^2 \\
& \quad \leq f_1(x_k) + H_k(x_k) - \langle \xi_{k+1}, x_k - x_{k+1} \rangle + \frac{L_\phi}{2} \|x_k - u_k\|^2 \\
& \qquad - \frac{\theta_1}{2} \|x_{k+1} - x_k\|^2,
\end{aligned}
$$

where the last inequality follows from Proposition 1 (i) and (iv). We also note that from Assumption 1 (a), the function $\xi \to f_2^*(\xi) - \langle x_k, \xi \rangle + \frac{r}{2} \|\xi - \xi_k\|^2$ is convex with convex parameter $r \geq 0$. Using the definition of ξ_{k+1} in (6a), it follows that

$$
f_2^*(\xi_{k+1}) \leq f_2^*(\xi_k) + \langle x_k, \xi_{k+1} - \xi_k \rangle - r \|\xi_{k+1} - \xi_k\|^2.
$$

Adding the above two inequalities, we obtain that

$$
\begin{aligned}
& f_1(x_{k+1}) + f_2^*(\xi_{k+1}) - \langle x_{k+1}, \xi_{k+1} \rangle + H_k(x_{k+1}) \\
& \quad \leq f_1(x_k) + f_2^*(\xi_k) - \langle x_k, \xi_k \rangle + H_k(x_k) + \frac{L_\phi}{2} \|x_k - u_k\|^2 \\
& \qquad - \frac{\theta_1}{2} \|x_{k+1} - x_k\|^2 - r \|\xi_{k+1} - \xi_k\|^2.
\end{aligned} \tag{20}
$$

It follows from (20) and the definition of Θ that

$$
\begin{aligned}
& \Theta(\xi_{k+1}, x_{k+1}, x_k, y_k, y_{k-1}, \lambda_k) - \Theta(\xi_k, x_k, x_{k-1}, y_k, y_{k-1}, \lambda_k) \\
& \quad \leq -\frac{1}{2} \big(\theta_1 - L_\phi \alpha_{\max}^2\big) \|x_k - x_{k-1}\|^2 - r \|\xi_{k+1} - \xi_k\|^2,
\end{aligned}
$$

which implies that the conclusion holds.

By the above three lemmas, we get a descent property for Algorithm 1, i.e., the potential function Θ is decreasing along the sequence generated from Algorithm 1 if the parameters β, $\alpha_{\max}$, and the functions ϕ, ψ are suitably chosen. □

Proposition 3 *Suppose the Assumption 1 holds, and the sequence $\{d_k = (\xi_k, x_k, x_{k-1}, y_k, y_{k-1}, \lambda_k)\}_{k \in \mathbb{N}}$ is generated by Algorithm 1. Then for any $k \geq 1$, we have that*

$$
\Theta(d_{k+1}) - \Theta(d_k) \leq -b_1 \|y_{k+1} - y_k\|^2 - b_2 \|x_k - x_{k-1}\|^2 - r \|\xi_{k+1} - \xi_k\|^2,
$$

where $r \geq 0$, b_1 and b_2 are defined in (4). Moreover, if $b_1 > 0$ and $b_2 > 0$, the sequence $\{\Theta(d_k)\}_{k \in \mathbb{N}}$ is monotonically nonincreasing.

We now ready to present the following proposition, which characterizes the distance of a subgradient of potential function from zero.

Proposition 4 *Let* $\{d_k = (\xi_k, x_k, x_{k-1}, y_k, y_{k-1}, \lambda_k)\}_{k\in\mathbb{N}}$ *be the sequence generated by Algorithm 1. Define* $\xi_k^* = x_{k-1} - x_k - r(\xi_k - \xi_{k-1})$, $x_k^* = 2A^T(\lambda_{k-1} - \lambda_k) - (\nabla\phi(x_k) - \nabla\phi(u_{k-1})) + (\theta_1\mathcal{I} - \beta A^T A)(x_k - x_{k-1}) + A^T(\lambda_{k-1} - \lambda_{k-2})$, $x_{k-1}^* = \theta_1(x_{k-1} - x_k)$, $y_k^* = B^T(\lambda_{k-1} - \lambda_k) + \frac{2\eta_2}{\beta}(y_k - y_{k-1}) - (\nabla\psi(y_k) - \nabla\psi(y_{k-1}))$, $y_{k-1}^* = \frac{2\eta_2}{\beta}(y_{k-1} - y_k)$ *and* $\lambda_k^* = \frac{1}{\beta}(\lambda_k - \lambda_{k-1})$. *Then we have that* $(\xi_k^*, x_k^*, x_{k-1}^*, y_k^*, y_{k-1}^*, \lambda_k^*) \in \partial\Theta(d_k)$. *Moreover, there exists* $\zeta > 0$ *such that*

$$
\begin{aligned}
dist(0, \partial\Theta(d_k)) \le \zeta(&\|x_k - x_{k-1}\| + \|x_{k-1} - x_{k-2}\| + \|\lambda_{k-1} - \lambda_{k-2}\| \\
&+ \|\lambda_k - \lambda_{k-1}\| + \|y_k - y_{k-1}\|) + r\|\xi_k - \xi_{k-1}\|.
\end{aligned}
$$

Proof It follows from the definition of Θ, Lemma 2 and (11d) that $\partial\Theta(d_k) = (\partial_x\Theta(d_k), \partial_\xi\Theta(d_k), \partial_{\tilde{x}}\Theta(d_k), \partial_y\Theta(d_k), \partial_{\tilde{y}}\Theta(d_k), \partial_\lambda\Theta(d_k))$ with

$$
\begin{cases}
\partial_\xi\Theta(d_k) = \partial f_2^*(\xi_k) - x_k, \\
\partial_x\Theta(d_k) = \partial f_1(x_k) - \xi_k + \theta_1(x_k - x_{k-1}) + A^T(\lambda_{k-1} - 2\lambda_k), \\
\partial_{\tilde{x}}\Theta(d_k) = \theta_1(x_{k-1} - x_k), \\
\partial_y\Theta(d_k) = \nabla g(y_k) - B^T\lambda_k + B^T(\lambda_{k-1} - \lambda_k) + \frac{2\eta_2}{\beta}(y_k - y_{k-1}), \\
\partial_{\tilde{y}}\Theta(d_k) = \frac{2\eta_2}{\beta}(y_{k-1} - y_k), \\
\partial_\lambda\Theta(d_k) = -(Ax_k + By_k - b).
\end{cases}
\tag{21}
$$

By rewriting the first-order condition of the subproblems in Algorithm 1, we have

$$
\begin{cases}
x_{k-1} - r(\xi_k - \xi_{k-1}) \in \partial f_2^*(\xi_k), \\
A^T[\lambda_{k-1} - \beta(Ax_k + By_{k-1} - b)] + \xi_k - (\nabla\phi(x_k) - \nabla\phi(u_{k-1})) \in \partial f_1(x_k), \\
\nabla g(y_k) = B^T\lambda_k - (\nabla\psi(y_k) - \nabla\psi(y_{k-1})), \\
Ax_k + By_k - b = -\frac{1}{\beta}(\lambda_k - \lambda_{k-1}).
\end{cases}
$$

Substituting the above inclusions and equalities into (21) and using the definition of ξ_k^*, x_k^*, y_k^* and λ_k^*, it yields that

$$
\xi_k^* \in \partial_\xi\Theta(d_k), \quad x_k^* \in \partial_x\Theta(d_k), \quad y_k^* \in \partial_y\Theta(d_k), \quad \lambda_k^* \in \partial_\lambda\Theta(d_k).
$$

It follows that $(\xi_k^*, x_k^*, x_{k-1}^*, y_k^*, y_{k-1}^*, \lambda_k^*) \in \partial\Theta(d_k)$. Clearly, there exists $\zeta > 0$ such that

$$
\begin{aligned}
\mathrm{dist}(0, \partial\Theta(d_k)) \le \zeta(&\|x_k - x_{k-1}\| + \|x_{k-1} - x_{k-2}\| + \|\lambda_{k-1} - \lambda_{k-2}\| \\
&+ \|\lambda_k - \lambda_{k-1}\| + \|y_k - y_{k-1}\|) + r\|\xi_k - \xi_{k-1}\|,
\end{aligned}
$$

which completes the proof. □

Denote the set of all limit points of $\{d_k\}_{k\in\mathbb{N}}$ by $\Omega(d_0)$. Before proving several properties of the set $\Omega(d_0)$, we give the following important result.

Proposition 5 *Suppose that the Assumption 1 holds, and that* $b_1 > 0$ *and* $b_2 > 0$. *Assume that the sequence* $\{d_k = (\xi_k, x_k, x_{k-1}, y_k, y_{k-1}, \lambda_k)\}_{k\in\mathbb{N}}$ *generated by Algorithm 1, and that the sequence* $\{\omega_k = (x_k, y_k, \lambda_k)\}_{k\in\mathbb{N}}$ *is bounded. Then*

$$
\sum_{k=0}^{\infty} \|\omega_{k+1} - \omega_k\|^2 < \infty \quad and \quad \sum_{k=0}^{\infty} r\|\xi_{k+1} - \xi_k\|^2 < \infty.
$$

Proof It follows from Proposition 3 that

$$b_3(\|x_k - x_{k-1}\|^2 + \|y_{k+1} - y_k\|^2) + r\|\xi_{k+1} - \xi_k\|^2 \leq \Theta(d_k) - \Theta(d_{k+1}),$$

where $b_3 = \min\{b_1, b_2\} > 0$. Moreover, the sequence $\{\Theta(d_k)\}_{k\in\mathbb{N}}$ is monotonically nonincreasing. Summing up the above inequality for $k = 0, \cdots, N$, it follows that

$$\sum_{k=0}^{N} b_3(\|x_k - x_{k-1}\|^2 + \|y_{k+1} - y_k\|^2) + r\|\xi_{k+1} - \xi_k\|^2 \\ \leq \Theta(d_0) - \Theta(d_{N+1}). \tag{22}$$

From the inclusion (11b), we obtain

$$\xi_{k+1} \in \partial f_1(x_{k+1}) - A^T\lambda_k + \beta A^T(Ax_{k+1} + By_k - b) + (\nabla\phi(x_{k+1}) - \nabla\phi(u_k)).$$

Since the sequence $\{\omega_k\}_{k\in\mathbb{N}}$ is bounded, it follows that Proposition 16.20 in [6] and the boundedness of $\{u_k\}_{k\in\mathbb{N}}$ and $\{x_{k+1}\}_{k\in\mathbb{N}}$ that $\{\partial f_1(x_{k+1})\}_{k\in\mathbb{N}}$ is also bounded. Thus, the sequences $\{\xi_k\}_{k\in\mathbb{N}}$ and $\{d_k\}_{k\in\mathbb{N}}$ are also bounded, which means that $\{d_k\}_{k\in\mathbb{N}}$ has at least one cluster point. Let d^* be a cluster point of $\{d_k\}_{k\in\mathbb{N}}$ and let $\{d_{k_j}\}_{j\in\mathbb{N}}$ be the subsequence converging to it, i.e., $d_{k_j} \to d^*$. By Assumption 1 (a)–(b), we obtain that the function Θ is lower semicontinuous on $\mathbb{R}^{n_1} \times \mathbb{R}^{n_1} \times \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_2} \times \mathbb{R}^m$, and hence $\Theta(d^*) \leq \liminf_{j\to\infty} \Theta(d_{k_j})$. Consequently, $\{\Theta(d_{k_j})\}_{j\in\mathbb{N}}$ is bounded from below, which together with the fact that $\{\Theta(d_k)\}_{k\in\mathbb{N}}$ is nonincreasing, means that the sequence $\{\Theta(d_k)\}_{k\in\mathbb{N}}$ is convergent and

$$\lim_{k\to\infty} \Theta(d_k) \geq \Theta(d^*). \tag{23}$$

Letting $N \to \infty$ in (22) and using (23), we have

$$\sum_{k=0}^{\infty} \left\{ b_3(\|x_k - x_{k-1}\|^2 + \|y_{k+1} - y_k\|^2) + r\|\xi_{k+1} - \xi_k\|^2 \right\} \leq \Theta(d_0) - \Theta(d^*).$$

Thus, we have $\sum_{k=0}^{\infty} \|x_{k+1} - x_k\|^2 < \infty$, $\sum_{k=0}^{\infty} \|y_{k+1} - y_k\|^2 < \infty$, and $\sum_{k=0}^{\infty} r\|\xi_{k+1} - \xi_k\|^2 < \infty$. By (15), it follows that $\sum_{k=0}^{\infty} \|\lambda_{k+1} - \lambda_k\|^2 < \infty$, which completes the proof. □

We are now ready to summarize several properties of limit point set $\Omega(d_0)$ in the following proposition.

Proposition 6 *Suppose that the Assumption 1 holds, and that $b_1 > 0$ and $b_2 > 0$. Let $\{d_k = (\xi_k, x_k, x_{k-1}, y_k, y_{k-1}, \lambda_k)\}_{k\in\mathbb{N}}$ be the sequence generated by Algorithm 1. The sequence $\{\omega_k = (x_k, y_k, \lambda_k)\}_{k\in\mathbb{N}}$ is assumed to be bounded. The following assertions hold.*

(i) *$\Omega(d_0)$ is a nonempty compact set, and $\lim_{k\to\infty} dist(d_k, \Omega(d_0)) = 0$;*
(ii) *Any accumulation point of (x_k, y_k, λ_k) is a critical point of L_β;*
(iii) *The limit $\lim_{k\to\infty} \Theta(d_k) = \Lambda$ exists and $\Theta(\bar{d}) = \Lambda$ for all $\bar{d} \in \Omega(d_0)$.*

Proof We proof the assertions item by item.

(i) From the proof of Proposition 5, we see that the sequences $\{\xi_k\}_{k\in\mathbb{N}}$ and $\{d_k\}_{k\in\mathbb{N}}$ are bounded, provided that the sequence $\{\omega_k\}_{k\in\mathbb{N}}$ is bounded. By the definition of $\Omega(d_0)$, the proof of this assertion is trivial.

(ii) Let $\omega^* = (x^*, y^*, \lambda^*)$ be an accumulation point of $\{\omega_k\}_{k\in\mathbb{N}}$, then there exists an index set I_1 such that $\lim_{k(\in I_1)\to\infty} \omega_k = \omega^*$. For simplify, we define $d^* = (\xi^*, x^*, x^*, y^*, y^*, \lambda^*)$. By Proposition 5, we obtain that $\lim_{k\to\infty} \|\omega_{k+1} - \omega_k\|^2 = 0$ and $\lim_{k\to\infty} r\|\xi_{k+1} - \xi_k\|^2 = 0$. From the proof of Proposition 5, we see that the sequence $\{\xi_k\}_{k\in\mathbb{N}}$ is bounded. Thus, there exit $\xi^* \in \mathbb{R}^{n_1}$ and a subindex set I_2 of I_1 such that

$$\begin{cases} \lim_{k(\in I_2)\to\infty} x_{k-1} = x^*, & \lim_{k(\in I_2)\to\infty} x_{k+1} = x^*, \\ \lim_{k(\in I_2)\to\infty} \xi_{k+1} = \xi^*, & \lim_{k(\in I_2)\to\infty} y_{k-1} = y^*, \\ \lim_{k(\in I_2)\to\infty} d_{k+1} = d^*. \end{cases} \tag{24}$$

From the lower semicontinuity of Θ, we have

$$\lim \inf_{k(\in I_2)\to\infty} \Theta(\xi_{k+1}, x_{k+1}, x_k, y_k, y_{k-1}, \lambda_k) \geq \Theta(d^*). \tag{25}$$

By the definition of ξ_{k+1} and x_{k+1} as the minimizer in (6a) and (6b), we have that

$$\begin{cases} f_2^*(\xi_{k+1}) \leq f_2^*(\xi^*) - \langle x_k, \xi^* - \xi_{k+1}\rangle + \frac{r}{2}(\|\xi^* - \xi_k\|^2 - \|\xi_{k+1} - \xi_k\|^2), \\ f_1(x_{k+1}) - \langle \xi_{k+1}, x_{k+1}\rangle + H_k(x_{k+1}) \leq f_1(x^*) - \langle \xi_{k+1}, x^*\rangle + H_k(x^*) \\ \quad + B_\phi(x^*, u_k) - B_\phi(x_{k+1}, x_k), \end{cases}$$

where the function H_k was defined in (19). Adding the above two inequalities, we obtain

$$\begin{aligned} &f_2^*(\xi_{k+1}) + f_1(x_{k+1}) - \langle \xi_{k+1}, x_{k+1}\rangle + H_k(x_{k+1}) \\ &\quad \leq f_2^*(\xi^*) + f_1(x^*) - \langle \xi_{k+1}, x^*\rangle - \langle x_k, \xi^* - \xi_{k+1}\rangle + H_k(x^*) \\ &\quad\quad + \frac{r}{2}(\|\xi^* - \xi_k\|^2 - \|\xi_{k+1} - \xi_k\|^2) + B_\phi(x^*, u_k) - B_\phi(x_{k+1}, x_k). \end{aligned} \tag{26}$$

By the definition of Θ and H_k, we get

$$\begin{aligned} &\Theta(\xi_{k+1}, x_{k+1}, x_k, y_k, y_{k-1}, \lambda_k) \\ &\quad = f_1(x_{k+1}) + f_2^*(\xi_{k+1}) - \langle x_{k+1}, \xi_{k+1}\rangle + H_k(x_{k+1}) \\ &\quad\quad + g(y_k) + \frac{\theta_1}{2}\|x_{k+1} - x_k\|^2 + \frac{\eta_2}{\beta}\|y_k - y_{k-1}\|^2. \end{aligned} \tag{27}$$

It follows from (26) and (27) that

$$\begin{aligned} &\Theta(\xi_{k+1}, x_{k+1}, x_k, y_k, y_{k-1}, \lambda_k) \\ &\quad \leq f_2^*(\xi^*) + f_1(x^*) - \langle \xi_{k+1}, x^*\rangle - \langle x_k, \xi^* - \xi_{k+1}\rangle + H_k(x^*) \\ &\quad\quad + \frac{r}{2}(\|\xi^* - \xi_k\|^2 - \|\xi_{k+1} - \xi_k\|^2) + B_\phi(x^*, u_k) - B_\phi(x_{k+1}, x_k) \\ &\quad\quad + g(y_k) + \frac{\theta_1}{2}\|x_{k+1} - x_k\|^2 + \frac{\eta_2}{\beta}\|y_k - y_{k-1}\|^2. \end{aligned} \tag{28}$$

Taking the limit in (28) and combining with (24) and Proposition 5, we have

$$\begin{aligned} &\lim \sup_{k(\in I_2)\to\infty} \Theta(\xi_{k+1}, x_{k+1}, x_k, y_k, y_{k-1}, \lambda_k) \\ &\quad \leq \Theta(d^*) + \lim \sup_{k(\in I_2)\to\infty} \frac{r}{2}\|\xi^* - \xi_k\|^2 + \lim \sup_{k(\in I_2)\to\infty} \frac{L_\phi}{2}\|x^* - u_k\|^2 \\ &\quad = \Theta(d^*) + \lim \sup_{k(\in I_2)\to\infty} \frac{r}{2}\|\xi^* - \xi_k\|^2. \end{aligned} \tag{29}$$

If $r = 0$, then $\limsup_{k(\in I_2)\to\infty} \frac{r}{2}\|\xi^* - \xi_k\|^2 = 0$ holds. Otherwise, by Proposition 5 and (24), we have that $\lim_{k\to\infty} \|\xi_{k+1} - \xi_k\| = 0$ and

$$\lim_{k(\in I_2)\to\infty} \|\xi_k - \xi^*\|^2 = 0,$$

which together with (29) implies that

$$\lim \sup_{k(\in I_2)\to\infty} \Theta(\xi_{k+1}, x_{k+1}, x_k, y_k, y_{k-1}, \lambda_k) \le \Theta(d^*). \tag{30}$$

Since $\Theta(\cdot)$ is lower semicontinuous, it follows that

$$\lim \inf_{k(\in I_2)\to+\infty} \Theta(\xi_{k+1}, x_{k+1}, x_k, y_k, y_{k-1}, \lambda_k) \ge \Theta(d^*). \tag{31}$$

From (30) and (31), we obtain that

$$\begin{aligned}\Theta(d^*) &= \lim_{k(\in I_2)\to+\infty} \Theta(\xi_{k+1}, x_{k+1}, x_k, y_k, y_{k-1}, \lambda_k)\\ &= \lim_{k(\in I_2)\to+\infty} \{Q(\xi_{k+1}, x_{k+1}) + S(x_{k+1}, x_k, y_k, y_{k-1}, \lambda_k)\}\end{aligned}$$

where $S(x, \tilde{x}, y, \tilde{y}, \lambda) = g(y) - \langle\lambda, Ax + By - b\rangle + \frac{\beta}{2}\|Ax + By - b\|^2 + \frac{\theta_1}{2}\|x - \tilde{x}\|^2 + \frac{\eta_2}{\beta}\|y - \tilde{y}\|^2$. It follows from the triangle inequality that

$$\begin{aligned}&|Q(\xi_{k+1}, x_{k+1}) - Q(\xi^*, x^*)|\\ &\quad\le |\Theta(\xi_{k+1}, x_{k+1}, x_k, y_k, y_{k-1}, \lambda_k) - \Theta(\xi^*, x^*, x^*, y^*, y^*, \lambda^*)|\\ &\qquad + |S(x_{k+1}, x_k, y_k, y_{k-1}, \lambda_k) - S(x^*, x^*, y^*, y^*, \lambda^*)|.\end{aligned} \tag{32}$$

Taking the limit in (32) and using the continuity of S, we get that

$$\lim_{k(\in I_2)\to+\infty} Q(\xi_{k+1}, x_{k+1}) = Q(\xi^*, x^*).$$

Recall that $Q(\xi, x) = f_1(x) + f_2^*(\xi) - \langle x, \xi\rangle$. It follows from (11a) and (11b) that for any $k(\in I_2)$,

$$\begin{cases} v_{k+1} = x_k - x_{k+1} - r(\xi_{k+1} - \xi_k) \in \partial f_2^*(\xi_{k+1}) - x_{k+1},\\ e_{k+1} = A^T\lambda_k - p_k - (\nabla\phi(x_{k+1}) - \nabla\phi(u_k)) \in \partial f_1(x_{k+1}) - \xi_{k+1}.\end{cases}$$

Combining Lemma 2 with the definition of v_{k+1} and e_{k+1}, it yields that $(v_{k+1}, e_{k+1}) \in \partial Q(\xi_{k+1}, x_{k+1})$. By Proposition 5 and (24), we obtain that

$$\lim_{k(\in I_2)\to\infty} (v_{k+1}, e_{k+1}) = (0, A^T\lambda^*).$$

It follows from Lemma 1 (a) that $(0, A^T\lambda^*) \in \partial Q(\xi^*, x^*)$, i.e.,

$$0 = \partial f_2^*(\xi^*) - x^* \quad\text{and}\quad A^T\lambda^* \in \partial f_1(x^*) - \xi^*. \tag{33}$$

By the continuity of f_2 and (33), we get that $\xi^* \in \partial f_2(x^*)$ and

$$A^T\lambda^* \in \partial f_1(x^*) - \partial f_2(x^*). \tag{34}$$

Furthermore, because of the continuity of ∇g, letting the index k in (11c) and (11d) belong to I_2 and trend to ∞, we have that

$$\nabla g(y^*) = B^T\lambda^* \quad\text{and}\quad Ax^* + By^* = b.$$

This together with (34) and (3) means that (x^*, y^*, λ^*) is a critical point of L_β.

(iii) For any point $d^* = (\xi^*, x^*, \tilde{x}^*, y^*, \tilde{y}^*, \lambda^*) \in \Omega(d_0)$, then there exists an index set I_3 such that $\lim_{k(\in I_3)\to\infty} d_k = d^*$. From Proposition 5, we see that $\tilde{x}^* = x^*$, $\tilde{y}^* = y^*$ and $d^* = (\xi^*, x^*, x^*, y^*, y^*, \lambda^*)$. By the lower semicontinuity of Θ, we have

$$\lim_{k(\in I_3)\to\infty} \inf \Theta(\xi_k, x_k, x_{k-1}, y_k, y_{k-1}, \lambda_{k-1}) \geq \Theta(d^*). \tag{35}$$

It follows from the continuity of Θ with respect to y and λ that

$$\lim_{k(\in I_3)\to\infty} \sup \Theta(d_k) = \lim_{k(\in I_3)\to\infty} \sup \Theta(\xi_k, x_k, x_{k-1}, y_{k-1}, y_{k-2}, \lambda_{k-1}). \tag{36}$$

By the similar way in item (ii), we have obtain

$$\lim_{k(\in I_3)\to\infty} \Theta(\xi_k, x_k, x_{k-1}, y_{k-1}, y_{k-2}, \lambda_{k-1}) = \Theta(d^*). \tag{37}$$

It follows from (35)–(37) that

$$\lim_{k(\in I_3)\to\infty} \Theta(d_k) = \Theta(d^*),$$

which together the fact that $\{\Theta(d_k)\}_{k\in\mathbb{N}}$ is decreasing implies that

$$\lim_{k\to\infty} \Theta(d_k) = \Theta(d^*).$$

Therefore, Θ is constant on $\Omega(d_0)$. Moreover, $\inf_{k\in\mathbb{N}} \Theta(d_k) = \lim_{k\to\infty} \Theta(d_k)$. □

By combining with Proposition 3 with Propositions 6, we are now ready to present the global convergence analysis of Algorithm 1.

Theorem 2 *Suppose that the Assumption 1 holds, and that $b_1 > 0$ and $b_2 > 0$. Let $\{d_k = (\xi_k, x_k, x_{k-1}, y_k, y_{k-1}, \lambda_k)\}_{k\in\mathbb{N}}$ be the sequence generated by Algorithm 1. Suppose that the sequence $\{\omega_k = (x_k, y_k, \lambda_k)\}_{k\in\mathbb{N}}$ is assumed to be bounded. Suppose that Θ is a KL function. Then, $\{\omega_k\}_{k\in\mathbb{N}}$ has finite length, that is*

$$\sum_{k=0}^{\infty} \|\omega_{k+1} - \omega_k\| < \infty,$$

and as a consequence, $\{\omega_k\}_{k\in\mathbb{N}}$ converges to a critical point of L_β.

Proof In the view of Proposition 6, it suffices to prove that $\{\omega_k\}_{k\in\mathbb{N}}$ is convergent and $\sum_{k=1}^{\infty} \|\omega_{k+1} - \omega_k\| < \infty$. To this end, we first recall from Propositions 3 and 6 (iii) that the sequence $\{\Theta(d_k)\}_{k\in\mathbb{N}}$ is nonincreasing and $\zeta = \lim_{k\to\infty} \Theta(d_k)$ exists. In the following, we consider two cases.

Case 1 Suppose that $\Theta(d_N) = \Lambda$ for some $N \geq 1$. Since $\{\Theta(d_k)\}_{k\in\mathbb{N}}$ is decreasing, we must have $\Theta(d_k) = \Lambda$ for all $k \geq N$. It follows from Proposition 3 that $r(\xi_{N+t} - \xi_N) = 0$, $x_{N+t} = x_N = x_{N-1}$ and $y_{N+t} = y_N$ for all $t \geq 0$. Hence, $\{x_k\}_{k\in\mathbb{N}}$ and $\{y_k\}_{k\in\mathbb{N}}$ converges finitely. Moreover, it follows from (15) that for $k \geq N + 1$,

$$\|\lambda_{k+1} - \lambda_k\|^2 \leq \eta_1 \|y_{k+1} - y_k\|^2 + \eta_2 \|y_k - y_{k-1}\|^2 = 0.$$

which implies that $\lambda_{N+1+t} = \lambda_{N+1}$ for any $t \geq 0$, and the sequence $\{\omega_{N+1+t}\}_{t\in\mathbb{N}}$ is constant. Consequently, we obtain that $\{\omega_k\}$ is a convergent sequence in this case.

Case 2 Suppose that $\Theta(d_k) > \Lambda$, for any $k \geq 1$. Note that

$$\lim_{k\to\infty} \text{dist}(d_k, \Omega(d_0)) = 0.$$

It follows that for all $\epsilon > 0$, there exists k_1 such that for any $k > k_1$, $\operatorname{dist}(d_k, \Omega(d_0)) < \epsilon$. Again since $\lim_{k\to\infty} \Theta(d_k) = \Lambda$, it follows that for all $\eta > 0$, there exists $k_2 > 0$, such that for any $k > k_2$, $\Theta(d_k) < \Lambda + \eta$. Consequently, for all $\epsilon, \eta > 0$, when $k \geq M = \max\{k_1, k_2\} + 1$,

$$\operatorname{dist}(d_k, \Omega(d_0)) < \epsilon, \quad \Lambda < \Theta(d_k) < \Lambda + \eta.$$

Since $\Omega(d_0)$ is a nonempty compact set and Θ is constant on $\Omega(d_0)$, it follows from Lemma 3 that for any $k \geq M$,

$$\operatorname{dist}(0, \partial\Theta(d_k))\varphi'(\Theta(d_k) - \Lambda) \geq 1. \tag{38}$$

For notation simplicity, we define

$$\Delta_k = \varphi(\Theta(d_k) - \Lambda) - \varphi(\Theta(d_{k+1}) - \Lambda).$$

Since $\Theta(d_k) - \Theta(d_{k+1}) = (\Theta(d_k) - \Lambda) - (\Theta(d_{k+1}) - \Lambda)$, from the concavity of φ, we get that

$$\Delta_k \geq \varphi'(\Theta_1(d_k) - \Lambda)(\Theta_1(d_k) - \Theta_1(d_{k+1})). \tag{39}$$

It follows that

$$\begin{aligned}
& b_3\big(\|x_k - x_{k-1}\|^2 + \|y_{k+1} - y_k\|^2\big) + r\|\xi_{k+1} - \xi_k\|^2 \\
&\quad \leq b_1\|y_{k+1} - y_k\|^2 + b_2\|x_k - x_{k-1}\|^2 + r\|\xi_{k+1} - \xi_k\|^2 \\
&\quad \leq \Theta(d_k) - \Theta(d_{k+1}) \leq \operatorname{dist}(0, \partial\Theta(d_k))\varphi'(\Theta(d_k) - \Lambda)(\Theta(d_k) - \Theta(d_{k+1})) \\
&\quad \leq \Delta_k \operatorname{dist}(0, \partial\Theta(d_k)) \\
&\quad \leq \zeta\Delta_k(\|x_k - x_{k-1}\| + \|x_{k-1} - x_{k-2}\| + \|\lambda_{k-1} - \lambda_{k-2}\| + \|\lambda_k - \lambda_{k-1}\| \\
&\qquad + \|y_k - y_{k-1}\|) + \Delta_k r\|\xi_k - \xi_{k-1}\|,
\end{aligned} \tag{40}$$

where the first inequality is from the definition of b_1, b_2 and b_3, the second inequality follows from Proposition 3, the third inequality is obtained by (38), the fourth inequality is from (39) and the last inequality follows from Proposition 4. Let $\delta_1 = \sqrt{\eta_2}$ and $\delta_2 = \sqrt{\eta_2}$ and take $\gamma > \max\{1, \frac{1+2(\delta_1+\delta_2)}{2}\}$. Now, we divide the proof of this case into two parts.

Case 2a Suppose that $r = 0$. From Eq. (40), we have

$$\begin{aligned}
& \frac{b_3}{2}\big(\|x_k - x_{k-1}\| + \|y_{k+1} - y_k\|\big)^2 \\
&\quad \leq \zeta\Delta_k(\|x_k - x_{k-1}\| + \|x_{k-1} - x_{k-2}\| + \|\lambda_{k-1} - \lambda_{k-2}\| + \|\lambda_k - \lambda_{k-1}\|).
\end{aligned}$$

By using the fact that $2\sqrt{h_1 h_2} \leq \gamma h_1 + \frac{h_2}{\gamma}$ for all $h_1, h_2 > 0$, we obtain that

$$\begin{aligned}
& \|x_k - x_{k-1}\| + \|y_{k+1} - y_k\| \\
&\quad \leq \frac{1}{2\gamma}(\|x_k - x_{k-1}\| + \|x_{k-1} - x_{k-2}\| + \|y_k - y_{k-1}\|) \\
&\qquad + \frac{1}{2\gamma}(\|\lambda_{k-1} - \lambda_{k-2}\| + \|\lambda_k - \lambda_{k-1}\|) + \frac{\gamma\zeta}{2b_3}\Delta_k.
\end{aligned} \tag{41}$$

On the other hand, it follows from (14) that

$$\begin{cases}
\|\lambda_k - \lambda_{k-1}\| \leq \delta_1\|y_k - y_{k-1}\| + \delta_2\|y_{k-1} - y_{k-2}\| \\
\|\lambda_{k-1} - \lambda_{k-2}\| \leq \delta_1\|y_{k-1} - y_{k-2}\| + \delta_2\|y_{k-2} - y_{k-3}\|.
\end{cases} \tag{42}$$

Then substituting (42) into (41), and rearranging the terms, we have that

$$\begin{aligned}
&(1-\frac{1}{\gamma})\|x_k-x_{k-1}\|+(1-\delta_3)\|y_{k+1}-y_k\| \\
&\quad\le \delta_3(\|y_k-y_{k-1}\|-\|y_{k+1}-y_k\|)+\delta_4(\|y_{k-1}-y_{k-2}\|-\|y_k-y_{k-1}\|) \\
&\qquad+\frac{1}{2\gamma}(\|x_{k-1}-x_{k-2}\|-\|x_k-x_{k-1}\|)+\frac{\gamma\zeta}{2b_3}\Delta_k \\
&\qquad+\frac{\delta_2}{2\gamma}(\|y_{k-2}-y_{k-3}\|-\|y_{k-1}-y_{k-2}\|), \qquad (43)
\end{aligned}$$

where $\delta_3=\frac{1+2\delta_1+2\delta_2}{2\gamma}$ and $\delta_4=\frac{\delta_1+2\delta_2}{2\gamma}$. Thus, summing (43) from $k=M$ to ∞, it yields that

$$\begin{aligned}
&\sum_{k=M}^{\infty}\left\{\left(1-\frac{1}{\gamma}\right)\|x_k-x_{k-1}\|+(1-\delta_3)\|y_{k+1}-y_k\|\right\} \\
&\quad\le \delta_3(\|y_M-y_{M-1}\|+\|y_{M-1}-y_{M-2}\|+\|y_{M-2}-y_{M-3}\|) \\
&\qquad+\frac{1}{2\gamma}\|x_{M-1}-x_{M-2}\|+\frac{\zeta\gamma}{2b_3}\varphi(\Theta(d_M)-\Theta(d^*)).
\end{aligned}$$

By the choice of γ, we get that $\sum_{k=M}^{\infty}\|x_k-x_{k-1}\|<\infty$ and $\sum_{k=M}^{\infty}\|y_{k+1}-y_k\|<\infty$, which implies that $\{x_k\}$ and $\{y_k\}$ are convergent. Moreover, from (42), we see that $\sum_{k=M}^{\infty}\|\lambda_k-\lambda_{k-1}\|<\infty$, which shows that $\{\lambda_k\}$ is convergent. Consequently, we conclude that $\{\omega_k\}$ is a convergent sequence.

Case 2b Suppose that $r>0$. Let $b_4=\min\{b_3,r\}$ and $b_5=\max\{\zeta,r\}$. It follows from (40) that

$$\begin{aligned}
&b_4(\|x_k-x_{k-1}\|^2+\|y_{k+1}-y_k\|^2+\|\xi_{k+1}-\xi_k\|^2) \\
&\quad\le b_5\Delta_k(\|x_k-x_{k-1}\|+\|x_{k-1}-x_{k-2}\| \\
&\qquad+\|\lambda_{k-1}-\lambda_{k-2}\|+\|\lambda_k-\lambda_{k-1}\|+\|\xi_k-\xi_{k-1}\|). \qquad (44)
\end{aligned}$$

By Cauchy-Schwarz inequality, we get

$$\begin{aligned}
&(\|x_k-x_{k-1}\|+\|y_{k+1}-y_k\|+\|\xi_{k+1}-\xi_k\|)^2 \\
&\quad\le 3(\|x_k-x_{k-1}\|^2+\|y_{k+1}-y_k\|^2+\|\xi_{k+1}-\xi_k\|^2). \qquad (45)
\end{aligned}$$

It follows from (44)–(45) that

$$\begin{aligned}
&\|x_k-x_{k-1}\|+\|y_{k+1}-y_k\|+\|\xi_{k+1}-\xi_k\| \\
&\quad\le \frac{1}{2\gamma}(\|x_k-x_{k-1}\|+\|x_{k-1}-x_{k-2}\|+\|\lambda_{k-1}-\lambda_{k-2}\| \\
&\qquad+\|\lambda_k-\lambda_{k-1}\|+\|y_k-y_{k-1}\|+\|\xi_k-\xi_{k-1}\|)+\frac{3\gamma b_5\Delta_k}{2b_4}. \qquad (46)
\end{aligned}$$

Similar, substituting (42) into (46), and rearranging the terms, we have

$$\begin{aligned}
&\left(1-\frac{1}{\gamma}\right)\|x_k-x_{k-1}\|+(1-\delta_3)\|y_{k+1}-y_k\|+\left(1-\frac{1}{2\gamma}\right)\|\xi_{k+1}-\xi_k\| \\
&\quad\le \frac{3\gamma b_5\Delta_k}{2b_4}+\frac{1}{2\gamma}(\|x_{k-1}-x_{k-2}\|-\|x_k-x_{k-1}\|)
\end{aligned}$$

$$+\delta_3(\|y_k - y_{k-1}\| - \|y_{k+1} - y_k\|) + \delta_4(\|y_{k-1} - y_{k-2}\| - \|y_k - y_{k-1}\|)$$
$$+\frac{\delta_2}{2\gamma}(\|y_{k-2} - y_{k-3}\| - \|y_{k-1} - y_{k-2}\|) + \frac{1}{2\gamma}(\|\xi_k - \xi_{k-1}\| - \|\xi_{k+1} - \xi_k\|).$$

Using the similar way in Case 2a, we can obtain that the sequence $\{\omega_k\}$ is convergent.

From Theorem 3 of [45], we get that if $\frac{1}{\sigma\eta_0} \leq \beta \leq \frac{2}{\sigma\eta_0}$, the sequence $\{\omega_k = (x_k, y_k, \lambda_k)\}_{k\in\mathbb{N}}$ generated by BADMM-DC is bounded. Here, we relax the upper bound of β and give a similar sufficient condition to guarantee the sequence $\{\omega_k = (x_k, y_k, \lambda_k)\}_{k\in\mathbb{N}}$ generated by Algorithm 1 is bounded. □

Lemma 7 *Let $\{\omega_k = (x_k, y_k, \lambda_k)\}_{k\in\mathbb{N}}$ be the sequence generated by Algorithm 1. Suppose that there exists $\sigma > 0$ such that*

$$\hat{g} = \inf_y\{g(y) - \sigma\|\nabla g(y)\|^2\} > -\infty,$$

and that $g(y) - \sigma\|\nabla g(y)\|^2$ is coercive. Suppose that $f_1(x) - f_2(x)$ is coercive, and that $\beta \geq \frac{1}{\sigma\eta_0}$. Then $\{\omega_k\}_{k\in\mathbb{N}}$ is bounded.

Proof From (12), it yields that $B^T\lambda_k = \nabla g(y_k) + (\nabla\psi(y_k) - \nabla\psi(y_{k-1}))$. By simple computations, we obtain that

$$\|\lambda_k\|^2 \leq \frac{2}{\eta_0}\|\nabla g(y_k)\|^2 + \eta_2\|y_k - y_{k-1}\|^2, \tag{47}$$

It follows that for all $k \geq 1$,

$$\begin{aligned}
&\Theta(\xi_1, x_1, x_0, y_1, y_0, \lambda_1)\\
&\quad \geq \Theta(\xi_k, x_k, x_{k-1}, y_k, y_{k-1}, \lambda_k)\\
&\quad \geq f_1(x_k) - f_2(x_k) + g(y_k) - \frac{\|\lambda_k\|^2}{2\beta} + \frac{\eta_2}{\beta}\|y_k - y_{k-1}\|^2\\
&\qquad + \frac{\beta}{2}\left\|Ax_k + By_k - b - \frac{\lambda_k}{\beta}\right\|^2\\
&\quad \geq f_1(x_k) - f_2(x_k) + g(y_k) - \sigma\|\nabla g(y_k)\|^2 + \frac{\eta_2}{2\beta}\|y_k - y_{k-1}\|^2\\
&\qquad + \left(\sigma - \frac{1}{\beta\eta_0}\right)\|\nabla g(y_k)\|^2 + \frac{\beta}{2}\left\|Ax_k + By_k - b - \frac{\lambda_k}{\beta}\right\|^2,
\end{aligned} \tag{48}$$

where the first inequality is from Proposition 3, the second inequality holds due to the Fenchel-Young inequality, the last inequality follows from (47). By (48) and the assumption, it follows that the sequences $\{x_k\}_{k\in\mathbb{N}}$ and $\{y_k\}_{k\in\mathbb{N}}$ are bounded, which together with (47) implies the sequence $\{\lambda_k\}_{k\in\mathbb{N}}$ is bounded. This completes the proof. □

5 Numerical experiments

In this section, we apply Algorithm 1 to solve the total variation image restoration problem [45] and the least squares problems with l_{1-2} regularization problem [52]. We also compare the performance of our method with DPGA [5], pDCA$_e$ [29,52] and BADMM-DC [45]. The MATLAB codes are run on a PC (with Intel (R) Core (TM) i7-6700HQ CPU 2.60 GHZ) under MATLAB Version 8.6.0.267246 (R2015b) Service Pack 1 which contains optimization Toolbox version 7.3.

In Tables 1, 2 and 3, "iter." denotes the number of iteration, "cpu" denotes the CPU time in seconds. In Table 1, "SNR" denotes the signal-to-noise ratio and "SNR-de" denotes the SNR of the degraded images. In Tables 2 and 3, "obj." denotes the objective function value of optimization problem, "$t_{\lambda_{\max}}$" denotes the time for computing $\lambda_{\max}(A^T A)$, where $\lambda_{\max}(A^T A)$ is the Lipschitz constant of the smooth part of the least squares problems with l_{1-2} regularization. In Table 3, "max" means the number of iterations hits 6000. We report the above mentioned values, averaged over the 10 random instances.

5.1 Total variation image restoration problem

In this subsection, we report the numerical performance of Algorithm 1 on solving the total variation image restoration problem. We first briefly review its background, see [25,33,43] for more details.

Given an image of the size $m_1 \times m_2$ pixels, where m_1 and m_2 denote the numbers of pixels in horizontal and vertical directions, respectively, it can be presented by a vector $y \in \mathbb{R}^n$ with $n = m_1 m_2$. Suppose that the original image $\bar{y}$ is blurred by a linear operator $H : \mathbb{R}^n \to R^n$ (e.g. a convolution by a blurring kernel). Furthermore, it is corrupted with a noise υ. Hence, the main goal of image processing is to find $\bar{y} \in \mathbb{R}^n$ such that

$$y^0 = H\bar{y} + \upsilon, \tag{49}$$

where y^0 is the degraded image. Since model (49) is usually ill-posed, certain regularization techniques are required. One of the most popular regularization technique is the total variation (TV) regularization proposed in [4,43], mainly because of its capability of preserving the edges of images. Thus, one can reconstruct the original image y by considering the variational regularization model

$$\min_y \; R(y) + \frac{1}{2}\|Hy - y^0\|^2, \tag{50}$$

where $\frac{1}{2}\|Hy - y^0\|^2$ is the data fitting term and $R(y)$ is the regularization term, related to the a priori information of the target image.

Note that a natural image is (approximated) sparse after certain transformation W, e.g., gradient operator [43] or wavelet transform [14]. By taking such properties into account, it is natural to take $R(y) = \|Wy\|_0$ in problem (50). However, finding a global minimizer of L_0 norm regularized problem is NP hard, one popular technology is to replace L_0 norm by L_1 norm, see e.g., [14]. If $W = K$, where K is the gradient operator, then we have that $R(y) = \|Ku\|_1$, which is the famous ROF model [43]. It has been shown that the nonconvex sparse metrics perform better that the convex l_1 in terms of promoting sparsity. A class of nonconvex functions enjoy the following form

$$R(y) = f_1(Wy) - f_2(Wy),$$

where f_1 and f_2 are both proper closed convex functions. We refer the readers to [5,19,33,45,54,55] for the choice of f_1 and f_2. Thus, problem (50) becomes the following Total Variational regularization problem

$$\min_y \; f_1(Wy) - f_2(Wy) + \frac{1}{2}\|Hy - y^0\|^2,$$

Table 1 Numerical results for the first example

Original images	SNR-de	DPGA in [5]			pDCA$_e$ in [29,52]			BADMM-DC in [45]			Algorithm 1		
		iter.	cpu	SNR	iter.	cpu	SNR	iter.	cpu	SNR	iter.	cpu	SNR
Cameraman.png	17.36	43	0.6	18.73	51	0.6	22.02	16	0.1	28.46	16	0.1	29.51
Lena.png	23.58	39	4.8	24.91	40	4.1	27.85	9	0.4	33.90	10	0.5	34.47
couple.bmp	19.79	41	5.1	21.33	51	5.1	25.20	12	0.5	31.93	12	0.6	32.48
lighthouse.bmp	16.20	44	5.4	17.72	51	5.0	21.44	16	0.7	28.94	16	0.8	29.35

Table 2 Numerical results for problem (52) with $\lambda = 1 \times 10^{-3}$

Problem size				DPGA in [5]			pDCA$_e$ in [29,52]			BADMM-DC in [45]			Algorithm 1		
n	m	s	$t_{\lambda_{\max}}$	iter.	cpu	obj.	iter.	cpu	obj.	iter.	cpu	obj.	iter.	cpu	obj.
2560	720	80	0.1	5339	6.3	5.9167e−02	599	0.7	5.9153e−02	5150	6.1	5.9166e−02	466	0.6	5.9153e−02
5120	1440	160	0.6	5291	31.7	1.2419e−01	601	3.7	1.2416e−01	5121	30.6	1.2418e−01	479	2.9	1.2415e−01
7680	2160	240	0.8	5333	68.9	1.8850e−01	601	7.8	1.8846e−01	5145	66.8	1.8850e−01	464	6.1	1.8846e−01
10240	2880	320	1.4	5368	123.6	2.5149e−01	601	13.9	2.5144e−01	5176	119.2	2.5149e−01	466	13.8	2.5144e−01
12800	3600	400	2.5	5441	195.9	3.1568e−01	601	21.4	3.1562e−01	5249	185.7	3.1568e−01	469	16.7	3.1561e−01
15360	4320	480	4.1	5485	273.3	3.8212e−01	601	33.2	3.8204e−01	5292	270.2	3.8212e−01	470	23.6	3.8204e−01
17920	5040	560	6.3	5432	371.6	4.4877e−01	601	41.3	4.4867e−01	5238	360.9	4.4876e−01	469	32.0	4.4867e−01
20480	5760	640	8.7	5451	495.5	5.1179e−01	601	54.3	5.1169e−01	5257	463.3	5.1178e−01	466	40.8	5.1168e−01
23040	6480	720	11.5	5557	619.6	5.8116e−01	601	69.4	5.8104e−01	5358	621.5	5.8115e−01	469	52.1	5.8104e−01
25600	7200	800	15.7	5498	788.0	6.4690e−01	601	98.4	6.4677e−01	5303	749.9	6.4689e−01	472	66.0	6.4676e−01

Table 3 Numerical results for problem (52) with $\lambda = 5 \times 10^{-4}$

Problem size				DPGA in [5]			pDCA$_e$ in [29,52]			BADMM-DC in [45]			Algorithm 1		
n	m	s	$t_{\lambda_{\max}}$	iter.	cpu	obj.	iter.	cpu	obj.	iter.	cpu	obj.	iter.	cpu	obj.
2560	720	80	0.1	5999	6.9	3.3580e−02	901	1.1	2.9499e−02	Max	6.9	3.2920e−02	651	0.8	2.9497e−02
5120	1440	160	0.7	5999	35.2	6.8352e−02	901	5.4	6.0430e−02	Max	35.7	6.7152e−02	638	3.8	6.0425e−02
7680	2160	240	0.7	5999	77.3	1.0568e−01	941	12.3	9.2908e−02	Max	78.7	1.0386e−01	679	8.9	9.2902e−02
10240	2880	320	1.3	5999	142.4	1.4499e−01	941	21.2	1.2755e−01	Max	134.4	1.4253e−01	681	15.3	1.2755e−01
12800	3600	400	2.4	5999	231.9	1.8197e−01	981	34.8	1.5954e−01	Max	214.3	1.7896e−01	746	26.5	1.5954e−01
15360	4320	480	4.1	5999	307.1	2.1244e−01	961	48.6	1.8897e−01	Max	305.0	2.0892e−01	645	32.7	1.8897e−01
17920	5040	560	6.5	5999	411.0	2.5467e−01	1001	72.3	2.2479e−01	Max	413.3	2.5048e−01	726	49.7	2.2479e−01
20480	5760	640	8.7	5999	534.9	2.9117e−01	981	86.9	2.5788e−01	Max	534.4	2.8643e−01	675	59.5	2.5787e−01
23040	6480	720	11.7	5999	673.1	3.2420e−01	981	109.2	2.8754e−01	Max	670.0	3.1886e−01	657	77.3	2.8753e−01
25600	7200	800	15.7	5999	849.9	3.6715e−01	1001	144.8	3.2387e−01	Max	847.4	3.6117e−01	708	101.2	3.2386e−01

or

$$\min_{x,y} \ f_1(x) - f_2(x) + \frac{1}{2}\|Hy - y^0\|^2, \quad s.t. \quad x - Wy = \mathbf{0}.$$

In this experiment, we test four grayscale images: the cameraman.png (256 × 256), the lenna.png (512 × 512), the couple.bmp (512 × 512) and the lighthouse.bmp (512 × 512). The blurring operator H is generated by the Matlab order H = fspecial('average', 5) and the noise υ is generated by the Gaussian noise $\mathcal{N}(0, 0.1)$ and $y^0 = H\bar{y} + \upsilon$. Here, we use the following linearly constrained difference-of-convex problem [45] to deblurr the above degraded images:

$$\min_{x,y} \ \rho\|x\|_1 - \rho\|x\|_2 + \frac{1}{2}\|Hy - y^0\|^2, \quad s.t. \quad x - Ky = \mathbf{0}, \tag{51}$$

where $K : \mathbb{R}^n \to \mathbb{R}^{2n}$ is the discrete gradient operator.

In the implementation of these methods, we set $\rho = 0.1$ and choose $\gamma_k = \frac{1}{8\mu}$ and $\mu_k = \frac{1}{8\mu}$ for DPGA with $\mu = 10$, $\beta = 1$, $\phi = 0.5\|\cdot\|^2$, $\psi = 0.5\|\cdot\|^2$ for BADMM-DC, $\beta = 1$, $r = 500$, $\phi = 0.5\|\cdot\|^2$, $\psi = 0.5\|\cdot\|^2$ for Algorithm 1. For pDCA$_e$ and Algorithm 1, we update the α_k by the following scheme $\alpha_k = \frac{\theta_{k-1}-1}{\theta_k}$ with $\theta_{k+1} = \frac{1+\sqrt{1+4\theta_k^2}}{2}$, where $\theta_{-1} = \theta_0 = 1$. Note that both pDCA$_e$ and DPGA have to compute the proximal operator of $f_1 \circ K$, which has no closed form solutions. Thus, we apply the method in [8] to compute the approximate solution of the proximal operator of $\|\cdot\|_1 \circ K$ at the given points. Here, we adpot the following stop criterion for these methods:

$$\frac{\|y_k - y_{k-1}\|_F}{\|y_k\|_F} < 10^{-3}.$$

To be fair, we set the initial points of these methods be the zero vector.

From [40], the SNR in the unit of dB is usually used to measure the quality of the restored images, and it is defined by SNR $= 20\log_{10}\frac{\|\bar{y}\|}{\|\bar{y}-y_1\|}$, where y_1 is the restored image and $\bar{y}$ is the original one. Note that the capability of achieving a higher SNR value reflects a better quality of the restored image for a method. The numerical results are presented in Tables 1. Moreover, In Fig. 1, we present the original image (cameraman.png), degraded image and the reconstructed images, which were recoveried by setting $\alpha = 0.1$ in problem (51). Form Table 1, we can see that after the almost same iterations and CPU time, Algorithm 1 can recovery the degraded images with higher SNR value than BADMM-DC.

5.2 The least squares problems with l_{1-2} regularization

In this subsection, we test the efficiency of Algorithm 1, DPGA, pDCA$_e$ and BADMM-DC on the least squares problems with l_{1-2} regularization, which has been encountered in compressed sensing, see [32,52] for more details.

The least squares problems with l_{1-2} regularization is to solve the following minimization problem:

$$\min_{x\in\mathbb{R}^n} \ \lambda\|x\|_1 - \lambda\|x\|_2 + \frac{1}{2}\|Ax - b\|^2, \tag{52}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $\lambda > 0$ is the regularization parameter. Clearly, problem (52) can be recast into problem (1)–(2) with $n_1 = n$, $n_2 = m$,

$$f_1(x) = \lambda\|x\|_1, \ f_2(x) = \lambda\|x\|_2, \ g(y) = \frac{1}{2}\|y\|^2, \ B = -\mathcal{I}.$$

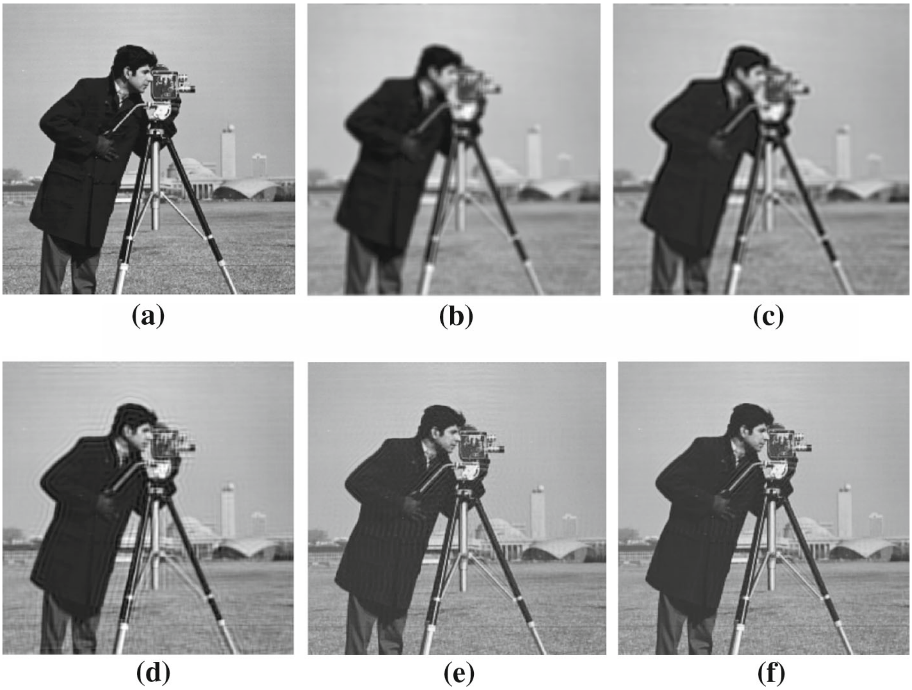

Fig. 1 Original image (**a**), degraded image (**b**) and recovered images (**c**–**f**). Images **c**–**f** is recovered by DPGA, pDCA$_e$, BADMM-DC, Algorithm 1. Here, the SNR values of images **b**–**f** is 17.37 dB, 18.73 dB, 22.02 dB, 28.46 dB and 29.51 dB, respectively

In this experiments, for given (m, n, s), we generate the matrix $A \in \mathbb{R}^{m\times n}$ and the vector $b \in \mathbb{R}^m$ as follow: we first generate A with i.i.d. standard Gaussian entries, and then normalize each column of A. Second, we set $b = A\bar{x} + 0.01 \cdot \epsilon$, where $\epsilon \in \mathbb{R}^m$ is a random vector with i.i.d. standard Gaussian entries, $\bar{x}$ is a random vector with i.i.d. standard Gaussian entries on the index set T and other entries are set to be zeros. Note that T is chosen uniformly at random from $\{1, 2, \cdots, n\}$ such that the cardinality of T is s.

In the implementation of these methods, we take $\gamma_k = \frac{49}{100L}$ and $\mu_k = 40$ for DPGA, where $L = \lambda_{\max}(A^T A)$, and set $\beta = 0.5$, $\psi \equiv 0$, $\phi(x) = \frac{1}{2}\|x\|_Q^2$ for BADMM-DC, where $Q = t\mathcal{I} - \beta A^T A$ with $t = 1.01 * \beta * L$. The choice of α_k in pDCA$_e$ and Algorithm 1 is what the reference [52] proposed. Moreover, we let $r = 30$, and set the other parameters of Algorithm 1 be same as that of BADMM-DC. To be fair, we set the initial points of these methods be the zero vector. We adpot the stop criterion of [52] for these methods:

$$\frac{\|x_k - x_{k-1}\|}{\max\{\|x_k\|, 1\}} < 10^{-5}.$$

The numerical results are presented in Tables 2 and 3, which correspond with $\lambda = 1 \times 10^{-3}$ and $\lambda = 5 \times 10^{-4}$, respectively. We can see that Algorithm 1 outperforms DPGA, pDCA$_e$ and BADMM-DC.

Acknowledgements This research was supported by the National Natural Science Foundation of China Grants 11801161, 61179033 and 11771003, and Natural Science Foundation of Hunan Province of China Grant

2018JJ3093. The authors are very grateful to Beijing Innovation Center for Engineering Science and Advanced Technology, Peking University and Beijing University of Technology for their joint project support. The first author Kai Tu would like to thank Prof. Penghua Yin from University of California, Los Angeles for providing the codes of [45], and Dr. Wenxing Zhang from University of Electronic Science and Technology of China and Dr. Benxing Zhang from Guilin University of Electronic Technology for the advices and discussions on the codes of the total variation image restoration problem. The authors also take this opportunity to thank the anonymous referees for their patient and valuable comments, which improved the quality of this paper greatly.

Appendix

Proof of Proposition 2

Proof Clearly, problem (51) is the special case of problem (1)–(2) with $f_1(x) = \rho\|x\|_1$, $f_2(x) = \rho\|x\|_2$ and $g(y) = \frac{1}{2}\|Hy - y^0\|^2$, $A = \mathcal{I}$, $B = -K$ and $b = \mathbf{0}$. Moreover, f_1, f_2, g, A and B satisfy Assumption 1 (a)–(d). It follows from the choice of ϕ and ψ that $L_\phi = r_1$, $L_\psi = r_2$, $\upsilon_\phi = r_1$ and $\upsilon_\psi = r_2$. By simple computations, we have that $b_1 > 0$ and $b_2 > 0$. It follows from $\sigma < \frac{1}{2\|H\|^2}$ that

$$
\begin{aligned}
g(y) - \sigma\|\nabla g(y)\|^2 &= \frac{1}{2}\|Hy - y^0\|^2 - \sigma\|H^*(Hy - y^0)\|^2 \\
&\geq (\frac{1}{2} - \sigma\|H\|^2)\|Hy - y^0\|^2 \geq 0. \qquad (53)
\end{aligned}
$$

It follows that for any $k \geq 1$,

$$
\begin{aligned}
&\Theta(\xi_1, x_1, x_0, y_1, y_0, \lambda_1) \\
&\quad \geq \rho(\|x_k\|_1 - \|x_k\|_2) + g(y_k) - \sigma\|\nabla g(y_k)\|^2 + \frac{\eta_2}{2\beta}\|y_k - y_{k-1}\|^2 \\
&\qquad + (\sigma - \frac{1}{\beta\eta_0})\|\nabla g(y_k)\|^2 + \frac{\beta}{2}\|x_k - Ky_k - \frac{\lambda_k}{\beta}\|^2, \\
&\quad \geq a_1 t_0\|y_k - z\|^2 + \frac{\beta}{2}\|x_k - Ky_k - \frac{\lambda_k}{\beta}\|^2 + t_1, \qquad (54)
\end{aligned}
$$

where $a_1 = (\frac{1}{2} - \sigma\|H\|^2)$, $z = \frac{1}{t_0}H^*y^0$ and $t_1 = a_1(\|y_0\|^2 - t_0\|z\|^2)$ with $t_0 = \lambda_{\min}(H^*H)$, the first inequality follows from (48), the second inequality is from $\inf_x\{\|x\|_1 - \|x\|_2\} \geq 0$ and (53). Since H has full column rank, it follows from (54) that the sequences $\{y_k\}_{k\in\mathbb{N}}$ and $\{x_k - Ky_k - \frac{\lambda_k}{\beta}\}_{k\in\mathbb{N}}$ are bounded, which together with (47) implies the sequences $\{\lambda_k\}_{k\in\mathbb{N}}$ and $\{x_k\}_{k\in\mathbb{N}}$ are bounded. Thus, the sequence $\{\omega_k\}_{k\in\mathbb{N}}$ is bounded. Now, we point that for this problem, the potential function $\Theta(\xi, x, \tilde{x}, y, \tilde{y}, \lambda)$, defined in (10), is a KL function. In fact, it follows from the definition of f_1, f_2 and Θ that

$$
\begin{aligned}
&\Theta(\xi, x, \tilde{x}, y, \tilde{y}, \lambda) \\
&\quad = \rho\|x\|_1 + I_\Omega(\xi) - \langle \xi, x\rangle + \frac{1}{2}\|Hy - y^0\|^2 - \langle \lambda, x - Ky - b\rangle \\
&\qquad + \frac{\beta}{2}\|x - Ky - b\|^2 + \frac{\theta_1}{2}\|x - \tilde{x}\|^2 + \frac{\eta_2}{\beta}\|y - \tilde{y}\|^2,
\end{aligned}
$$

where $I_\Omega(\xi)$ is the indicator function of closed convex set $\Omega = \{\xi \in \mathbb{R}^{n_1} \mid \|\xi\|^2 \leq \rho^2\}$. Clearly, Ω is a semi-algebraic set. By [10], we know that indicator function of semi-algebraic set is semi-algebraic, and that $\|\cdot\|_p$ is semi-algebraic whenever p is rational, i.e., $p = \frac{p_1}{p_2}$ where p_1 and p_2 are positive integers. Using the fact that finite sums of semi-algebraic functions is a semi-algebraic function, it yields that Θ is a semi-algebraic function, and hence it is a KL function. Note that all assumptions in Theorem 2 hold, which means the conclusion holds. □

Proposition 7 *Consider the total variation image restoration problem* [45]*:*

$$\min_{x,y} \ \rho\|x\|_1 - \rho\|x\|_2 + \frac{1}{2}\|Hy - y^0\|^2, \quad s.t. \quad x - Ky = \mathbf{0}, \tag{55}$$

where $\rho > 0$ is a regularization parameter, H is a blurred operator, $K : \mathbb{R}^n \to \mathbb{R}^{2n}$ is the discrete gradient operator. Suppose that $(\bar{x}, \bar{y})$ is a local minimum of problem (55), then there exists $\bar{\lambda}$ such that $(\bar{x}, \bar{y}, \bar{\lambda})$ is a critical point of problem (55), i.e., $(\bar{x}, \bar{y}, \bar{\lambda})$ satisfy the inclusion (3).

Proof We note that problem (55) is equivalent to the following problem

$$\min_{x,y} \ \rho\|Ky\|_1 - \rho\|Ky\|_2 + \frac{1}{2}\|Hy - y^0\|^2. \tag{56}$$

Since $(\bar{x}, \bar{y})$ is a local minimum of problem (55), it yields that $\bar{x} = K\bar{y}$, and that $\bar{y}$ is a local minimum of problem (56). It follows from Lemma 1 (b) that

$$\mathbf{0} \in \partial((\rho\|\cdot\|_1 - \rho\|\cdot\|_2) \circ K)(\bar{y}) + H^*(H\bar{y} - y^0), \tag{57}$$

where H^* is the adjoint operator of H. Since $\|\cdot\|_1$ is a proper convex function and $\|\cdot\|_2$ is a continuous convex function, it follows from the Corollary 3.4 in [37] and (57) that

$$\mathbf{0} \in \partial(\rho\|\cdot\|_1 \circ K)(\bar{y}) - \partial(\rho\|\cdot\|_2 \circ K)(\bar{y}) + H^*(H\bar{y} - y^0).$$

Note that $\partial(\rho\|\cdot\|_1 \circ K)(\bar{y}) = \rho K^*\partial\|K\bar{y}\|_1$ and $\partial(\rho\|\cdot\|_2 \circ K)(\bar{y}) = \rho K^*\partial\|K\bar{y}\|_2$, where K^* is the adjoint operator of K. Take $\bar{\xi}_1 \in \rho\partial\|\bar{x}\|_1$ and $\bar{\xi}_2 \in \rho\partial\|\bar{x}\|_2$, such that $K^*(\bar{\xi}_1 - \bar{\xi}_2) + H^*(H\bar{y} - y^0) = \mathbf{0}$. Letting $\bar{\lambda} = \bar{\xi}_1 - \bar{\xi}_2$, it yields that

$$\begin{cases} \bar{\lambda} \in \partial f_1(\bar{x}) - \partial f_2(\bar{x}), \\ -K^*\bar{\lambda} = \nabla g(\bar{y}), \\ \bar{x} - K\bar{y} = \mathbf{0}. \end{cases}$$

This completes the proof. □

References

1. An, L.T.H., Belghiti, M.T., Tao, P.D.: A new efficient algorithm based on DC programming and DCA for clustering. J. Global Optim. **37**(4), 593–608 (2007)
2. Attouch, H., Redont, P., Soubeyran, A.: A new class of alternating proximal minimization algorithms with costs-to-move. SIAM J. Optim. **18**(3), 1061–1081 (2007)
3. Attouch, H., Bolte, J.: On the convergence of the proximal algorithms for nonsmooth functions involving analytic features. Math. Program. **116**(1–2), 5–16 (2009)
4. Bai, M.R., Zhang, X.J., Shao, Q.Q.: Adaptive correction procedure for TVL1 image deblurring under impulse noise. Inverse Probl. **32**(8), 085004 (2016)
5. Banert, S., Bot, R.I.: A general double-proximal gradient algorithm for d.c. programming. Math. Program. (2018). https://doi.org/10.1007/s10107-018-1292-2
6. Bauschke, H.H., Combettes, P.L.: Convex analysis and monotone operator theory in Hilbert spaces. Springer, New York (2011)
7. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)

8. Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problem. IEEE Trans. Image Process. **18**(11), 2419–2434 (2009)
9. Becker, S., Bobin, J., Candès, E.: NESTA: a fast and accurate first-order method for sparse recovery. SIAM J. Imaging Sci. **4**(1), 1–39 (2009)
10. Bolte, J., Sabach, S., Teboule, M.: Proximal alternating linerized minimization for nonconvex and nonsmooth problems. Math. Program. **146**(1–2), 459–494 (2014)
11. Bot, R.I., Csetnek, E.R.: An inertial Tseng's type proximal algorithm for nonsmooth and nonconvex optimization problems. J. Optim. Theory Appl. **171**(2), 600–616 (2016)
12. Boyd, S., Parikh, N., Chu, E., Peleato, B., Echstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**(1), 1–122 (2011)
13. Bredies, K., Lorenz, D.A., Reiterer, S.: Minimization of nonsmooth, nonconvex functionals by iterative thresholding. J Optim. Theory Appl. **165**(1), 78–112 (2015)
14. Cai, J., Chan, R.H., Shen, L., Shen, Z.: Convergence analysis of tight framelet approach for missing data recovery. Adv. Comput. Math. **31**(1), 87–113 (2009)
15. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**(1–3), 293–318 (1992)
16. Eckstein, J., Yao, W.: Relative-error approximate versions of Douglas Rachford splitting and special cases of the ADMM. Math. Program. **170**(2), 417–444 (2018)
17. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. **2**, 17–40 (1976)
18. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R. (eds.) Augmented Lagrangian Methods Applications to the Numerical Solution of Boundary-Value Problems, pp. 299–331. North-Holland, Amsterdam (1983)
19. Gasso, G., Rakotomamonjy, A., Canu, S.: Recovering sparse signals with a certain family of nonconvex penalties and DC programming. IEEE Trans. Signal Process. **57**(12), 4686–4698 (2009)
20. Geremew, W., Nam, N.M., Semenova, A., Boginski, V., Pasiliao, E.: A DC programming approach for solving multicast network design problems via the Nesterov smoothing technique. J. Global Optim. **72**(4), 705–729 (2018)
21. Gotoh, J., Takeda, A., Tono, K.: DC formulations and algorithms for sparse optimization problems. Math. Program. **169**(1), 141–176 (2018)
22. Gonçalves, M.L.N., Melo, J.G., Monteiro, R.D.C.: Convergence rate bounds for a proximal ADMM with over-relaxation stepsize parameter for solving nonconvex linearly constrained problems (2017). arXiv preprint arXiv:1702.01850v2
23. Guo, K., Han, D.R., Wu, T.T.: Convergence of ADMM for optimization problem nonseparable nonconvex objective and linear constraints. Int. J. Comput. Math. **94**(8), 1653–1669 (2017)
24. Han, D.R., Yuan, X.M.: Local linear convergence of the alternating direction method of multipliers for quadratic programs. SIAM J. Numer. Anal. **51**(6), 3446–3457 (2013)
25. Hansen, P.C., Nagy, J.G., OĹeary, D.P.: Deblurring Images: Matrices, Spectra, and Filtering. SIAM, Philadelphia (2006)
26. He, B.S., Yuan, X.M.: On the $O(1/n)$ convergence rate of the Douglas–Rachford alternating direction method. SIAM J. Numer. Anal. **50**(2), 700–709 (2012)
27. Li, G.Y., Pong, T.K.: Global convergence of splitting methods for nonconvex composite optimization. SIAM J. Optim. **25**(4), 2434–2460 (2015)
28. Liavas, A.P., Sidiropoulos, N.D.: Parallel algorithms for constrained tensor factorization via alternating direction method of multipliers. IEEE Trans. Signal. Process. **63**(20), 5450–5463 (2015)
29. Liu, T.X., Pong, T.K., Takeda, A.: A refined convergence analysis of pDCA$_e$ with applications to simultaneous sparse recovery and outlier detection. Comput. Optim. Appl. **73**(1), 69–100 (2019)
30. Liu, Q.H., Shen, X.Y., Gu, Y.T.: Lineralized ADMM for non-convex non-smooth optimization with convergence analysis (2017). arXiv preprint arXiv:1705.02502
31. Lou, Y.F., Yin, P.H., Xin, J.: Point source super-resolution via non-convex l_1 based methods. J. Sci. Comput. **68**(3), 1082–1100 (2016)
32. Lou, Y.F., Yan, M.: Fast l_1-l_2 minimization via a proximal operator. J. Sci. Comput. **74**(2), 767–785 (2018)
33. Lou, Y.F., Zeng, T.Y., Osher, S., Xin, J.: A weighted difference of anisotropic and isotropic total variation model for image processing. SIAM J. Imaging Sci. **8**(3), 1798–1823 (2015)
34. Lu, Z.S., Li, X.R.: Sparse recovery via partial regularization: models, theory, and algorithms. Math. Oper. Res. **43**(4), 1290–1316 (2018)
35. Lu, Z.S., Zhou, Z.R., Sun, Z.: Enhanced proximal DC algorithms with extrapolation for a class of structured nonsmooth DC minimization. Math. Program. **176**(1–2), 369–401 (2019)

36. Maingé, P.E., Moudafi, A.: Convergence of new inertial proximal methods for DC programming. SIAM J. Optim. **19**(1), 397–413 (2008)
37. Mordukhovich, B.S., Nam, N.M., Yen, N.D.: Fréchet subdifferential calculus and optimality conditions in nondifferentiable programming. Optimization **55**(5–6), 685–708 (2006)
38. Nesterov, Y.: Introductory Lectures on Convex Optimization. A Basic Course. Kluwer, Boston (2004)
39. Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation. I: Basic Theory, II: Applications. Springer, Berlin (2006)
40. Pratt, W.K.: Digital Image Processing: PIKS Scientific Inside. Wiley, Hoboken (2001)
41. Parihk, N., Boyd, S.: Proximal algorithms. Found. Trends Optim. **1**(3), 127–239 (2013)
42. Rockafellar, R.T., Wets, R.: Variational Analysis. Grundlehren Math., Wiss., vol. 317. Springer, Berlin (1998)
43. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D. **60**(1–4), 259–268 (1992)
44. Souza, J.C.O., Oliveira, P.R.: A proximal point algorithm for DC fuctions on Hadamard manifolds. J. Global Optim. **63**(4), 797–810 (2015)
45. Sun, T., Yin, P.H., Cheng, L.Z., Jiang, H.: Alternating direction method of multipliers with difference of convex functions. Adv. Comput. Math. **44**, 723–744 (2018)
46. Tao, P.D., An, L.T.H.: Convex analysis approach to DC programming: theory, algorithms and applications. Acta Math. Vietnam. **22**(1), 289–355 (1997)
47. Tao, P.D., An, L.T.H.: A DC optimization algorithm for solving the trust-region subproblem. SIAM J. Optim. **8**(2), 476–505 (1998)
48. Wang, F.H., Xu, Z.B., Xu, H.K.: Convergence of Bregman alternating direction method with multipliers for nonconvex composite problems (2014). arXiv preprint arXiv:1410.8625
49. Wang, H.F., Kong, L.C., Tao, J.Y.: The linearized alternating direction method of multipliers for sparse group LAD model. Optim. Lett. **13**, 505–525 (2019)
50. Wu, Z.M., Li, M., Wang, D.Z.W., Han, D.R.: A symmetric alternating direction method of multipliers for separable nonconvex minimization problems. Asia Pac. J. Oper. Res. **34**, 1750030 (2017)
51. Wang, Y., Yao, W.T., Zeng, J.S.: Global convergence of ADMM in nonconvex nonsmooth optimization. J. Sci. Comput. **78**, 29–63 (2019)
52. Wen, B., Chen, X.J., Pong, T.K.: A proximal difference-of-convex algorithm with extrapolation. Comput. Optim. Appl. **69**(2), 297–324 (2018)
53. Yang, L., Pong, T.K., Chen, X.J.: Alternating direction method of multipliers for a class of nonconvex and nonsmooth problems with applications to background/foreground extraction. SIAM J. Imaging Sci. **10**(1), 74–110 (2017)
54. Yin, P.H., Liu, Y.F., He, Q., Xin, J.: Minimization of l_{1-2} for compressed sensing. SIAM J. Sci. Comput. **37**(1), 536–563 (2015)
55. Zhang, T.: Some sharp performance bounds for the least squares regression with l_1 regularization. Ann. Stat. **37**(5A), 2109–2144 (2009)