Alternative DC Algorithm for Partial DC programming

Le Thi Hoai An,[*] Huynh Van Ngai[†] and Pham Dinh Tao[‡]

Abstract

DC Programming and DCA have been introduced by Pham Dinh Tao in 1985 and extensively developed by Le Thi Hoai An and Pham Dinh Tao since 1994. These approaches have been successfully applied to solving real life problems in their large scale setting. In this paper, we introduce an alternative DC algorithm for solving partial DC programs of the form

$$\min f(x, y) \quad \text{s.t.} \quad (x, y) \in \mathbb{R}^n \times R^m,$$

where, $f : \mathbb{R}^n \times R^m \to \mathbb{R} \cup \{+\infty\}$ is a partially DC function, that is, for each $x \in \mathbb{R}^n$ and each $y \in \mathbb{R}^m$, $f(x, \cdot)$ and $f(\cdot, y)$ are DC functions in the second variable and the first variable, respectively. This proposed algorithm is an natural extension of the standard DC algorithm. Furthermore, we also consider an inexact version of this alternative DC algorithm. The convergence of these proposed algorithms (both the exact and inexact versions) are investigated. The applications to nonconvex feasibility problems and to matrix factorization problems are reported.

Keywords: DC program, DC algorithm, subanalytic, subdifferential, Lojasiewicz exponent

Mathematics Subject Classification. 90C30, 90C26, 90C52, 65K05, 47J25, 47N10

1 Introduction

DC programming and DCA (DC Algorithms) have been introduced by Pham Dinh Tao in 1985 and extensively developed by Le Thi Hoai An and Pham Dinh Tao since 1994. DC programming plays a key role in nonconvex programming because almost nonconvex programs encountered in practice are DC programs. Based on local optimality conditions and DC duality, DCA is one of efficient algorithms for nonconvex programs, especially for nonsmooth ones. Due to its local character it cannot guarantee the globality of computed solutions for general DC programs. However, we observe that, with a suitable starting point, it converges quite often to a global one. In practice, DCA was successfully applied to a lot of different and various nonconvex programs to which it gave almost always global solutions and proved to be more robust and more efficient than related standard methods, especially in the large-scale setting. On the other hand, it is worth noting that, with appropriate DC decompositions, DCA permits to find again standard optimization algorithms for convex and nonconvex programming (see [14], [15], [22], [23] and references therein). Let $\Gamma_0(\mathbb{R}^n)$ denote the convex cone of all lower semicontinuous proper convex functions on $\mathbb{R}^n$. The vector space of DC functions, $DC(\mathbb{R}^n) = \Gamma_0(\mathbb{R}^n) - \Gamma_0(\mathbb{R}^n)$, is

[*]Laboratoiry of Theoretical and Applied Computer Science, UFR MIM, Université de Lorraine, Site de Metz Ile du Saulcy 57045 Metz, France, lethi@univ-metz.fr

[†]Department of Mathematics, University of Quynhon, 170 An Duong Vuong, Qui Nhon, Vietnam, ngaivn@yahoo.com

[‡]Laboratory of Modelling, Optimization and Operations Research, LMI-National Institute for Applied Sciences-Rouen, BP 8, F 76 131 Mon Saint Aignan Cedex, France pham@insa-rouen.fr

quite large to contain almost real life objective functions and is closed under all the operations usually considered in Optimization.

In this paper, we are interested to the *partial DC programs* of the form

$$(P) \qquad \min f(x,y) \quad \text{s.t.} \quad (x,y) \in \mathbb{R}^n \times R^m. \tag{1}$$

Where, $f : \mathbb{R}^n \times R^m \to \mathbb{R} \cup \{+\infty\}$ is a partially DC function in the sense: for each $x \in \mathbb{R}^n$ and each $y \in \mathbb{R}^m$, $f(x,\cdot)$ and $f(\cdot,y)$ are DC functions. An important subclass of the class of partially DC functions is the one of functions f such that there are two extended valued partial convex functions $g(x,y)$ and $h(x,y)$ such that

$$f(x,y) = g(x,y) - h(x,y), \ \ (x,y) \in \mathbb{R}^n \times \mathbb{R}^m.$$

Here, and in throughout the paper, a convention $+\infty - (+\infty) = +\infty$ is used. By definition, a function in the two variable $(x,y) \in \mathbb{R}^n \times \mathbb{R}^m$ is said to be *partially convex* if it is convex in each variable when the other variable is fixed.

Since a partially convex function is not necessarily convex, this class of partial DC optimization problems which is broader than usual DC programs, includes many practical optimization models. Especially, under suitable assumptions, it covers the class of nonconvex and nonsmooth problems of the form

$$\min f_1(x) + f_2(y) + H(x,y), \ \text{ s. t. } \ (x,y) \in \mathbb{R}^n \times \mathbb{R}^m.$$

This latter class of problems is studied recenly in [3], [4], [5]. In these works, the approach is based on the proximal regularization of the Gauss-Seidel scheme, and based on the Kurdyka-Lojasiewicz inequality, the authors have established the convergent results for the considered methods. However, an inconvenience of that approach is that, one has to solve a nonconvex subproblem at each iterative step.

In this work, we propose an altenative DC algorithm for solving partial DC programs (P) . The terminology "alternative DCA" means that at each step, we apply the DCA alternatively for solving DC problems when the other variable is fixed. We also consider an inexact version of this algorithm by using approximate subdifferentials. We show that these algorithms (both the exact version and the inexact one) establish the interesting convergence.

The rest of the paper is organized as follows. In the next preliminary section, we recall the basis notions and results from Nonsmooth Analysis and Convex Analysis as well. Some new results on the relationship between subdifferentials and partial subdifferentials are given; a brief presentation on DCA and DC programming is reported, and some basis facts on subanalytic sets and functions are recalled. In Section 3, we introduce an alternative DC algorithm and its inexact version, and establish the convergence results as the convergence of subsequences; the convergence of the whole sequence under the Lojasiewicz inequality and under the Kurdyka-Lojasiewicz inequality. In Section 4, we consider generalized partial DC programs and propose a generalized version of the alternative DC algorithm for solving this program. We show that the convergence results in Section 3 remain valid for this generalized version. We present in the final section the applications to nonconvex feasibility and matrix factorization problems.

2 Preliminaries

2.1 Fréchet and limiting subdifferentials of lower semicontinuous functions

Let us recall some notions from Convex Analysis and Nonsmooth Analysis, which will be needed thereafter (see, e.g., [20], [24], [25]). In the sequel, the space $\mathbb{R}^n$ is equipped with the canonical inner product $\langle\cdot\rangle$. Its dual space is identified with $\mathbb{R}^n$ itself. $\mathcal{S}(\mathbb{R}^n)$ denotes the set of lower semicontinuous functions $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. The open ball with the center $x \in \mathbb{R}^n$ and radius $\varepsilon > 0$ is denoted by $B(x, \varepsilon)$; while the unit ball (i.e., the ball with the center at the origin and unit radius) is denoted by B. A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called $\rho-$convex for some $\rho(f) \geq 0$, if for all $x, y \in \mathbb{R}^n$, $\lambda \in [0, 1]$ one has

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\rho}{2}\lambda(1-\lambda)\|x-y\|^2.$$

The supremum of all $\rho \geq 0$ such that the above inequality is verified is called the convex modulus of f, which is denoted by $\rho(f)$.

The subdifferential of a convex function $f \in \mathcal{S}(\mathbb{R}^n)$ at $x \in \text{Dom } f$ is defined by

$$\partial f(x) = \{x^* \in \mathbb{R}^n : \quad \langle x^*, y - x\rangle \leq f(y) - f(x) \quad \forall y \in \mathbb{R}^n\}.$$

We set $\partial f(x) = \emptyset$ if $x \notin \text{Dom } f$. The $\varepsilon-$subdifferential with respect to a given $\varepsilon > 0$ is defined by

$$\partial_\varepsilon f(x) = \{x^* \in \mathbb{R}^n : \quad \langle x^*, y - x\rangle \leq f(y) - f(x) + \varepsilon \quad \forall y \in \mathbb{R}^n\},$$

and $\partial_\varepsilon f(x) = \emptyset$ if $x \notin \text{Dom } f$.

Note that for any $\varepsilon \geq 0$, one has the following relation between the approximative subdifferentials of a convex function f and its conjugate f^* :

$$x^* \in \partial_\varepsilon f(x) \quad \Longleftrightarrow \quad x \in \partial_\varepsilon f^*(x^*). \tag{2}$$

For a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous real extended valued function. The *Fréchet subdifferential* of f at $x \in \text{Dom } f$ is defined by

$$\partial^F f(x) = \left\{x^* \in \mathbb{R}^n : \ \liminf_{h\to 0} \frac{f(x+h) - f(x) - \langle x^*, h\rangle}{\|h\|} \geq 0\right\}.$$

For $x \notin \text{Dom } f$, we set $\partial^F f(x) = \emptyset$. The *limiting subdifferential* of f at $x \in \mathbb{R}^n$ is

$$\hat{\partial} f(x) = \{x^* \in \mathbb{R}^n : \ \exists (x_k, f(x_k)) \to (x, f(x)),\ x_k^* \in \partial^F f(x_k)) \quad \text{with } (x_k^*) \to x^* \text{ as } k \to \infty\}.$$

It is worth to mention that $\partial^F f(x)$ is not necessarily closed, althought $\hat{\partial} f(x)$ is closed, for any $x \in \mathbb{R}^n$. A point $x_0 \in \mathbb{R}^n$ is called a *Fréchet (limiting) critical* point for the function f, if $0 \in \partial^F f(x)$ ($0 \in \hat{\partial} f(x_0)$, respectively).

When f is a convex function , then the Fréchet subdifferential and the limiting subdifferential coincide with the subdifferential in the sense of Convex Analysis. Moreover, if f is a DC function, i.e., $f := g - h$, where g, h is convex functions, then

$$\partial^F f(x) \subseteq \partial f(x) \subseteq \partial g(x) - \partial h(x)$$

wherever h is continuous at x. Especially, if h is differentiable at x, then one has the equality:

$$\partial^F f(x) = \hat{\partial} f(x) = \partial g(x) - \nabla h(x).$$

The limiting subdifferential enjoys the following sum rule:

For two lower semicontnuous functions $f, g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ such that either f or g is locally Lipchitz at $\bar{x} \in \text{Dom } f \cap \text{Dom } g$, one has

$$\hat{\partial}(f+g)(x) \subseteq \hat{\partial} f(\bar{x}) + \hat{\partial} g(\bar{x}).$$

Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be a function in the two variables $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. Denote by $\partial_x^F f(x,y)$, $\partial_y^F f(x,y)$, the partial Fréchet subdifferential of f with respect to the variables x, y, respectively. It is obvious that

$$\partial^F f(x,y) \subseteq \partial_x^F f(x,y) \times \partial_y^F f(x,y).$$

The following lemma gives a sufficient condition to guarantee the inverse inclusion.

Lemma 2.1 *Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be a function in the two variables $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ and let $(\bar{x}, \bar{y}) \in \text{Dom } f$ be given. If for $x^* \in \partial_x^F f(\bar{x}, \bar{y})$, the following condition is satisfied: for any $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\langle x^*, x - \bar{x} \rangle \leq f(x,y) - f(\bar{x}, y) + \varepsilon \|x - \bar{x}\|, \ \forall (x,y) \in B(x,\delta) \times B(\bar{y}, \delta), \tag{3}$$

then one has

$$\{x^*\} \times \partial_y^F f(x,y) \subseteq \partial^F f(x,y).$$

As a result, if $f(x,y)$ is of the form

$$f(x,y) = f_1(x) + f_2(y) + Q(x,y), \ \ (x,y) \in \mathbb{R}^n \times \mathbb{R}^m,$$

where, Q is a continuously differentiable at $(\bar{x}, \bar{y}) \in \text{Dom } f$, then

$$\partial^F f(\bar{x}, \bar{y}) = \partial_x^F f(\bar{x}, \bar{y}) \times \partial_y^F f(\bar{x}, \bar{y}).$$

Proof. It is straightforward from the definition of the Fréchet subdifferential. □

A function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is said to be *partially convex* if for each $x \in \mathbb{R}^n$ and each $y \in \mathbb{R}^m$, the functions $f(x, \cdot)$ and $f(\cdot, y)$ are convex.

Corollary 2.1 *Suppose that $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be a partially convex function in the two variables $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ and that, for $(\bar{x}, \bar{y}) \in \text{Dom } f$, the partial subdifferential $\partial_x f(\bar{x}, y)$ is lower semicontinuous with respect to the variable y at $\bar{y}$, in the sense: for any $\varepsilon > 0$, there is $\delta > 0$ such that*

$$\partial_x f(\bar{x}, \bar{y}) \subseteq \partial_x f(\bar{x}, y) + \varepsilon B_{\mathbb{R}^n} \quad \forall y \in B(\bar{y}, \delta).$$

Then one has

$$\partial^F f(\bar{x}, \bar{y}) = \partial_x f(\bar{x}, \bar{y}) \times \partial_y f(\bar{x}, \bar{y}).$$

Proof. Let $\varepsilon > 0$, and let $\delta > 0$ such that

$$\partial_x f(\bar{x}, \bar{y}) \subseteq \partial_x f(\bar{x}, y) + \varepsilon B_{\mathbb{R}^n} \quad \forall y \in B(\bar{y}, \delta).$$

Then, for any $x^* \in \partial_x^F f(\bar{x}, \bar{y}) = \partial_x^F f(\bar{x}, \bar{y})$,

$$x^* \in \partial_x f(\bar{x}, y) + \varepsilon B_{\mathbb{R}^n} \quad \forall y \in B(\bar{y}, \delta).$$

Hence,

$$\langle x^*, x - \bar{x} \rangle \leq f(x, y) - f(\bar{x}, y) + +\varepsilon \|x - \bar{x}\|, \ \forall (x, y) \in B(x, \delta) \times B(\bar{y}, \delta),$$

and the conclusion follows from Lemma 2.1. □

The next lemma gives the relationship between the Fréchet/limiting subdifferential and the partial subdifferentials of a partially convex function, by using directional derivatives.

Lemma 2.2 *Let $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous partially convex function. Let $(\bar{x}, \bar{y}) \in \mathrm{Dom}\ g$ be given. Then the followings hold.*

(i) $\partial^F g(\bar{x}, \bar{y}) \subseteq \partial_x g(\bar{x}, \bar{y}) \times \partial_y g(\bar{x}, \bar{y})$.

(ii) If for each $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, the functions $g(x, \cdot)$, $g(\cdot, y)$ are continuous at $\bar{x}$, $\bar{y}$, respectively, then

$$\hat{\partial} g(\bar{x}, \bar{y}) \subseteq \partial_x g(\bar{x}, \bar{y}) \times \partial_y g(\bar{x}, \bar{y}).$$

Moreover, the equality is valid if in addition, either $g'_x((\bar{x}, y), u)$ is uniformly lower semicontinuous in the variable y for $u \in B_{\mathbb{R}^n}$ or $g'_y((x, \bar{y}), v)$ is uniformly lower semicontinuous in the variable x for $v \in B_{\mathbb{R}^m}$, and in this case, one has

$$\partial^F g(\bar{x}, \bar{y}) = \hat{\partial} g(\bar{x}, \bar{y}) = \partial_x g(\bar{x}, \bar{y}) \times \partial_y g(\bar{x}, \bar{y}).$$

Proof. The inclusion (i) is obvious from definition. For (ii), let $(x^*, y^*) \in \hat{\partial} g(\bar{x}, \bar{y})$. By the definition, there exist sequences $(x_k, y_k) \to (\bar{x}, \bar{y})$ with $g(x_k, y_k) \to (\bar{x}, \bar{y})$; $(x_k^*, y_k^*) \in \partial^F g(x_k, y_k)$ such that $(x_k^*, y_k^*) \to (x^*, y^*)$. Thanks to (i), $x_k^* \in \partial_x g(x_k, y_k)$ and $y_k^* \in \partial_y g(x_k, y_k)$. Hence,

$$\langle x_k^*, x - x_k \rangle \leq g(x, y_k) - g(x_k, y_k) \quad \text{for all} \quad x \in \mathbb{R}^n;$$

$$\langle y_k^*, y - y_k \rangle \leq g(x_k, y) - g(x_k, y_k) \quad \text{for all} \quad y \in \mathbb{R}^m.$$

Since $g(x, \cdot)$, $g(\cdot, y)$ are continuous at $\bar{x}$, $\bar{y}$, respectively, by letting $k \to \infty$ in the two inequalities, one obtains

$$\langle x^*, x - \bar{x} \rangle \leq g(x, \bar{y}) - g(\bar{x}, \bar{y}) \quad \text{for all} \quad x \in \mathbb{R}^n;$$

$$\langle y^*, y - \bar{y} \rangle \leq g(\bar{x}, y) - g(\bar{x}, \bar{y}) \quad \text{for all} \quad y \in \mathbb{R}^m.$$

That is $(x^*, y^*) \in \partial_x g(\bar{x}, \bar{y}), \times \partial_y g(\bar{x}, \bar{y})$.

Assume now that, for instance, $g'_x((\bar{x}, y), u)$ is uniformly lower semicontinuous in the variable y for $u \in B_{\mathbb{R}^n}$. Then, for any $\varepsilon > 0$, there is $\delta > 0$ such that

$$g'_x((\bar{x}, y), u) \geq g'_x((\bar{x}, \bar{y}), u) - \varepsilon \quad \forall y \in B(\bar{y}, \delta), \ \forall u \in B_{\mathbb{R}^n}.$$

By the positive homogeneousness of $g'_x((\bar{x},y),u)$, in the variable u, one has

$$g'_x((\bar{x},y),x-\bar{x}) \geq g'_x((\bar{x},\bar{y}),x-\bar{x}) - \varepsilon\|x-\bar{x}\| \quad \forall y \in B(\bar{y},\delta),\ \forall x \in \mathbb{R}^n.$$

Let $(x^*,y^*) \in \partial_x g(\bar{x},\bar{y}) \times \partial_y g(\bar{x},\bar{y})$ be given. For $y \in B(\bar{y},\delta)$ and $x \in \mathbb{R}^n$, one has

$$\begin{aligned} g(x,y) - g(\bar{x},\bar{y}) &= [g(x,y) - g(\bar{x},y)] + [g(\bar{x},y) - g(\bar{x},\bar{y})] \\ &\geq g'_x((\bar{x},y),x-\bar{x}) + \langle y^*, y-\bar{y}\rangle \\ &\geq g'_x((\bar{x},\bar{y}),x-\bar{x}) - \varepsilon\|x-\bar{x}\| + \langle y^*, y-\bar{y}\rangle \\ &\geq \langle x^*, x-\bar{x}\rangle + \langle y^*, y-\bar{y}\rangle - \varepsilon\|x-\bar{x}\|. \end{aligned}$$

Therefore,

$$(x^*,y^*) \in \partial^F g(\bar{x},\bar{y}) \subseteq \partial g(\bar{x},\bar{y}),$$

which completes the proof of the lemma. □

A relationship between the sudifferentials and the partial subdifferentials of partially DC functions is reported in the following lemma.

Lemma 2.3 *Let $g,h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be lower semicontinuous partially convex functions. Let*

$$f(x,y) := g(x,y) - h(x,y), \quad (x,y) \in \mathbb{R}^n \times \mathbb{R}^m.$$

For given $(\bar{x},\bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$, suppose that $g(\cdot,y)$, $g(x,\cdot)$ are continuous at $\bar{x}$, $\bar{y}$, respectively, and h is locally Lipschitz around $(\bar{x},\bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$. Then one has

$$\partial^F f(\bar{x},\bar{y}) \subseteq \hat{\partial} f(\bar{x},\bar{y}) \subseteq (\partial_x g(\bar{x},\bar{y}) - \partial_x h(\bar{x},\bar{y})) \times (\partial_y g(\bar{x},\bar{y}) - \partial_y h(\bar{x},\bar{y})).$$

Moreover, the equalities hold when h is continuously differentiable at $(\bar{x},\bar{y})$ and the either $g'_x((\bar{x},y),u)$ is uniformly lower semicontinuous in the variable y for $u \in B_{\mathbb{R}^n}$ or $g'_y((x,\bar{y}),v)$ is uniformly lower semicontinuous in the variable x for $v \in B_{\mathbb{R}^m}$.

Proof. Since h is locally Lipschitz, by the sum rule for the limiting subdifferential, one has

$$\hat{\partial} f(\bar{x},\bar{y}) \subseteq \hat{\partial} g(\bar{x},\bar{y}) + \hat{\partial}(-h)(\bar{x},\bar{y})$$

By Lemma 2.2,

$$\hat{\partial} g(\bar{x},\bar{y}) \subseteq \partial_x g(\bar{x},\bar{y}) \times \partial_y g(\bar{x},\bar{y}).$$

Therefore, it suffices to show that

$$\hat{\partial}(-h)(\bar{x},\bar{y}) \subseteq -\partial_x h(\bar{x},\bar{y}) \times \partial_y h(\bar{x},\bar{y}).$$

Indeed, let $(x^*,y_k^*) \in \hat{\partial}(-h)(\bar{x},\bar{y})$. Then there are sequences $(x_k,y_k) \to (\bar{x},\bar{y})$; $(x_k^*,y_k^*) \in \partial^F(-h)(x_k,y_k)$ such that $(x_k^*,y_k^*) \to (x^*,y^*)$. By the definition of the Fréchet subdifferential, for each index k, for any $\varepsilon > 0$, there is $\delta > 0$ such that

$$\langle x_k^*, u\rangle + \langle y_k^*, v\rangle \leq -h(x_k+u, y_k+v) + h(x_k,y_k) + \varepsilon(\|u\| + \|v\|) \quad \forall (u,v) \in \delta B_{\mathbb{R}^n \times \mathbb{R}^m}.$$

It implies that

$$\langle x_k^*, u\rangle \leq -h(x_k+u, y_k) + h(x_k,y_k) + \varepsilon\|u\| \quad \forall u \in \delta B_{\mathbb{R}^n};$$

$$\langle y_k^*, v\rangle \le -h(x_k, y_k+v) + h(x_k, y_k) + \varepsilon\|v\| \quad \forall v \in \delta\mathbb{R}^m.$$

For any $d \in \mathbb{R}^n$, when t is sufficiently small such as $t\|d\| \le \delta$, one has

$$\langle x_k^*, -td\rangle \le -h(x_k - td, y_k) + h(x_k, y_k) + \varepsilon t\|d\|.$$

Thanks to the convexity of h, one has

$$h(x_k, y_k) - h(x_k - td, y_k) \le h(x_k + td) - h(x_k, y_k).$$

By taking this into account of the preceding relation, one derives that for for $d \in \mathbb{R}^n$, when $t > 0$ is sufficiently small,

$$\langle x_k^*, -d\rangle \le \frac{h(x_k + td, y_k) - h(x_k, y_k)}{t} + \varepsilon\|d\|,$$

which implies

$$\langle -x_k^*, d\rangle \le h'_x((x_k, y_k), d) + \varepsilon\|d\| \le h(x_k + d, y_k) - h(x_k, y_k) + \varepsilon\|d\|.$$

Here, the second inequality follows from the convexity of $h(\cdot, y_k)$, which guarantees

$$h'_x((x_k, y_k), d) \le h(x_k + d, y_k) - h(x_k, y_k).$$

By letting $k \to \infty$, and as $\varepsilon > 0$ is arbitrary, one obtains

$$\langle -x^*, d\rangle \le h(\bar{x} + d, \bar{y}) - h(\bar{x}, \bar{y}) \quad \text{for all } d \in \mathbb{R}^n.$$

That is, $-x^* \in \partial_x h(\bar{x}, \bar{y})$. Similarly, one also obtains $-y^* \in \partial_y h(\bar{x}, \bar{y})$.

When h is continuously differentiable at $(\bar{x}, \bar{y})$, then

$$\hat{\partial} f(\bar{x}, \bar{y}) = \partial g(\bar{x}, \bar{y}) - \nabla h(\bar{x}, \bar{y}).$$

The conclusion follows directly from Lemma 2.2 (ii). □

Recall also that the Clarke subdifferential of a locally Lipschitz $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ at x is defined by

$$\partial^\circ f(x) = \{x^* \in \mathbb{R}^n : \ \langle x^*, d\rangle \le f^\circ(x, d)\ \forall d \in \mathbb{R}^n\}.$$

Where, $f^\circ(x, d)$ is the Clarke generalized directional derivative defined by

$$f^\circ(x, d) = \limsup_{z \to x, t \downarrow 0} \frac{f(z + td) - f(z)}{t}.$$

The following well-known properties are useful in the sequel.

(i) If f is locally Lipschitz around x, then $\partial^0 f(x) = \text{co}\hat{\partial} f(x)$, which coincides with the subdifferential in Convex Analysis when f is a convex function.

(ii) If $f := g - h$; g is a continuous convex function and h is a differentiable convex function around $x \in \mathbb{R}^n$,, then

$$\partial^\circ f(x) = \hat{\partial} f(x) = \partial^F f(x) = \partial g(x) - \nabla h(x).$$

(ii) If $f := g - h$; g is a differentiable convex function and h is a continuous convex function around $x \in \mathbb{R}^n$, then

$$\partial^\circ f(x) = \nabla g(x) - \partial h(x).$$

If $0 \in \partial^\circ f(x)$, then we say that a point $x \in \text{Dom } f$ is a Clarke critical point of f.

2.2 Subanalytic functions and Lojasiewicz inequality

We briefly recall the notion of subanalytic functions (see [18], [19]).

Definition 2.1 *(i) A subset C of $\mathbb{R}^n$ is said to be semianalytic if each point of $\mathbb{R}^n$, there exists a neighborhood V such that $C \cap V$ is of the following form:*

$$C \cap V = \bigcup_{i=1}^{p} \bigcap_{j=1}^{q} \{x \in V : \; f_{ij}(x) = 0, \, g_{ij}(x) > 0\},$$

where f_{ij}, $g_{ij} : V \to \mathbb{R}$ $(1 \leq i \leq p, \, 1 \leq j \leq q$ $)$ are real-analytic functions.

(ii) A subset C of $\mathbb{R}^n$ is called subanalytic if each point of $\mathbb{R}^n$, there exists a neighborhood V such that

$$C \cap V = \{x \in \mathbb{R}^n : \; \exists y \in R^m, \, (x, y) \in D\},$$

where D is a bounded seminalytic subset of $\mathbb{R}^n \times \mathbb{R}^m$ with $m \geq 1$.

(iii) A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be subanalytic if its graph gph f *is a subanalytic subset of $R^n \times \mathbb{R}$.*

It is obvious that the class of subanalytic sets (resp. functions) contains all analytic sets (resp. functions). Let us list some of the elementary properties of subanalytic sets and subanalytic functions (see, e.g., [10], [18], [26]):

- (i) Subanalytic sets are closed under locally finite union and intersection. The complement of a subanalytic set is subanalytic.
- (ii) The closure, the interior, the boundary of a subanalytic set are subanalytic.
- (iii) A closed $C \subseteq \mathbb{R}^n$ is subanalytic *iff* its indicator function χ_C, defined by $\chi_C(x) = 0$ if $x \in C$ and $+\infty$ otherwise, is subanalytic.
- (iv) Given a subanalytic set C, the distance function $d_C(x) := \inf_{z \in C} \|x - z\|$ is a subanalytic function.
- (v) Let $f, g : X \to \mathbb{R}$ be continuous subanalytic functions, where $X \subseteq \mathbb{R}^n$ is a subanalytic set. Then the sum $f + g$ is subanalytic if f maps bounded sets on bounded sets, or if both two functions are bounded from below.
- (vi) Let $X \subseteq \mathbb{R}^n$, $T \subseteq \mathbb{R}^m$ be subanalytic sets, where T is compact. If $f : \; X \times T \to \mathbb{R}$ is a continuous subanalytic function, then $g(x) := \min_{t \in T} f(x, t)$ is continuous subanalytic.
- (vii) Let $X \subseteq \mathbb{R}^n$ be subanalytic set and let $f, g : X \to \mathbb{R}$ be subanalytic functions. Then, $f + g$ is subanalytic if f maps bounded sets on bounded sets, or both two functions are bounded from below.

The following proposition gives the subanalyticity of the conjugate of a convex function.

Proposition 2.1 *If $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a lower semicontinuous subanalytic strongly convex function then its conjugate f^* is a $C^{1,1}$ (the class of functions whose derivative is Lipschitz) subanalytic convex function.*

Proof. It is well-known that the conjugate of a lower semicontinuous strongly convex function is of the class $C^{1,1}$. Let us prove that f^* is subanalytic. Let $\rho > 0$ is the modulus of the strong convexity of f. Take $\rho_1 \in (0, \rho)$ and set $h(x) := f(x) - \frac{\rho_1}{2}\|x\|^2$, $x \in \mathbb{R}^n$. Then h is a strongly convex function. One has, by the definition of the conjugate

$$\begin{aligned} f^*(x) = -\inf_{y\in\mathbb{R}^n}\{f(y) - \langle x, y\rangle\} &= -\inf_{y\in\mathbb{R}^n}\{h(y) + \tfrac{\rho_1}{2}\|y\|^2 - \langle x, y\rangle\} \\ &= -\inf_{y\in\mathbb{R}^n}\{h(y) + \tfrac{\rho_1}{2}\|x/\rho_1 - y\|^2\} + \tfrac{1}{2}\|x\|^2 = -\varphi(x) + \tfrac{\rho_1^{-1/2}}{2}\|x\|^2, \end{aligned}$$

where,

$$\varphi(x) = \inf_{y\in\mathbb{R}^n}\{h(y) + \frac{\rho_1}{2}\|x/\rho_1 - y\|^2\},\ x \in \mathbb{R}^n.$$

According to Proposition 2.9 in ([7]), the function φ is subanalytic. Thus f^* is subanalytic. □

The class of subanalytic functions possesses a nice property called the Lojasiewicz (gradient) inequality. This inequality has been established by Lojasiewicz ([18] for differentiable subanalytic functions, and then it is generalized by Bolte-Daniliidis-Lewis ([7]), to nonsmooth analytic functions. Let us recall the nonsmooth version of the important inequality, which is useful in our convergence analysis of algorithm.

Theorem 2.1 *(Theorem 3.1, [7]) Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a subanalytic function such that its domain* Dom f *is closed and $f|_{\text{Dom } f}$ is continuous and let x_0 is a critical point of f. Then there exist $\theta \in [0, 1)$, $L > 0$ and a neighborhood V of x_0 such that the following inequality holds.*

$$|f(x) - f(x_0)|^\theta \le L\|x^*\| \quad \text{for all } x \in V,\ x^* \in \partial^F f(x),$$

where a convention $0^0 = 1$ is used.

The number θ in the theorem is called a *Lojasiewicz exponent* of the critical point x_0.

A general version of the Lojasiewicz inequality established by Kyrdyka [**?**] that is called "Kurdyka-Lojasiewicz inequality" is satisfies for definable functions on o-minimal structures. This general inequality was also developped for nonsmooth functions by Bolte *et al* [9]. We will recall it in Section 3.

2.3 Brief presentation on DC programming and DC algorithm

We consider now a standard d.c. program, that is, an optimization problem of the form:

$$(\mathcal{P}) \quad \alpha = \inf\{f(x) := g(x) - h(x) : \ \ x \in \mathbb{R}^n\}$$

where g, h belong to $\Gamma_0(\mathbb{R}^n)$, the class of lower semicontinuous proper convex functions on $\mathbb{R}^n$. The function f is called a DC function on $\mathbb{R}^n$. Remark that the closed convex constraint set C is incorporated in the first convex DC component g with the help of its indicator function χ_C ($\chi_C\ (x) := 0$ if $x \in C, +\infty$ otherwise). Recall the natural convention $+\infty - (+\infty) = +\infty$ and that $\alpha \in \mathbb{R}$ implies *dom* $g := \{x \in \mathbb{R}^n : g(x) < +\infty\} \subset$ *dom* h. The dual problem of ($\mathcal{P}$ is defined by

$$(\mathcal{D}) \quad \alpha = \inf\{h^*(y) - g^*(y) : \ \ y \in \mathbb{R}^n\}.$$

where g^*, h^* are the conjugate functions of g, h, respectively, i.e.,

$$g^*(y) := \sup\{\langle x, y\rangle - g(x) : x \in \mathbb{R}^n\}.$$

A point $x^* \in \mathbb{R}^n$ (resp. $y^* \in \mathbb{R}^n$) is called a *weak critical* point of d.c problem $(\mathcal{P})$ if $0 \in \partial g(x^*) \cap \partial h(x^*)$ (resp. $0 \in \partial g^*(y^*) \cap \partial h^*(y^*)$.

For a d.c. optimization problem with a set constraint:

$$\inf\{f(x) : \quad x \in C\},$$

where f is a d.c. function and $C \subseteq \mathbb{R}^n$ is a nonempty convex set, we can equivalently transform it into a standard d.c. program by using the indicator function of C as follow.

$$\inf\{f(x) + \chi_C(x) : \quad x \in \mathbb{R}^n\}.$$

Note that if f is a subanalytic function and C is a subanalytic set then so is the function $f + \chi_C$.

In the convex approach to DC programming, the DCA, based on local optimality and DC duality. The DCA consists in the construction of the two sequences $\{x^k\}$ and $\{y^k\}$ (candidates for being primal and dual solutions, respectively) that we improve at each iteration (thus, the sequences $\{g(x^k) - h(x^k)\}$ and $\{h^*(y^k) - g^*(y^k)\}$ are decreasing) in an appropriate way such that their corresponding limits x^∞ and y^∞ satisfy local optimality conditions (see, e.g., [14], [15], [22], [23]).

DC Algorithm (DCA) (simplified form) consists of constructing the two sequences $\{x^k\}$ and $\{y^k\}$, starting a given $x^0 \in \text{Dom } g$ by setting

$$(DCA) \qquad y^k \in \partial h(x^k); \quad x^{k+1} \in \partial g^*(y^k).$$

3 Partial DC programming and alternative DC algorithm

Consider a partial DC program of the form:

$$(PDCP) \qquad \inf\{f(x, y) = g(x, y) - h(x, y) : \quad (x, y) \in \mathbb{R}^n \times \mathbb{R}^m\}.$$

Where, $g, h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ are lower semicontinuous partial convex functions. Such the function f is called a *partially DC* function. A subset $C \in \mathbb{R}^n \times \mathbb{R}^m$ is said to be *partially convex* if for each $x \in \mathbb{R}^n$ and each $y \in \mathbb{R}^m$, the following sets are convex:

$$C_x = \{y \in \mathbb{R}^m : \quad (x, y) \in C\}; \quad C_y = \{x \in \mathbb{R}^n : \quad (x, y) \in C\}.$$

Obviously, if C is convex then it is also partially convex, and C is a partially convex set *iff* its indicator function χ_C is a partially convex function. By the help of indicator functions, we can equivalently transform a partial DC program with a set constraint:

$$\inf\{f(x, y) : \quad (x, y) \in C\},$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is a partial DC function and $C \subseteq \mathbb{R}^n \times \mathbb{R}^m$ is a partial convex set, into a partial DC optimization problem (PDCP) as follows.

$$\inf\{f(x, y) + \chi_C(x, y) : \quad (x, y) \in \mathbb{R}^n \times \mathbb{R}^m\}.$$

Definition 3.1 *A point $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ is said to be a weak critical point of the partial DC program (PDCP) if*

$$0_{\mathbb{R}^n} \in \partial_x g(x, y) \cap \partial_x h(x, y) \quad \textit{and} \quad 0_{\mathbb{R}^m} \in \partial_y g(x, y) \cap \partial_y h(x, y)$$

According to Lemma 2.3, a Fréchet critical point is a weak critical point. When h is of C^1 class around x and g satisfies the assumptions of Lemma 2.2, then a weak critical point is a (Fréchet) critical point.

The alternative DC algorithm we propose here for solving (PDCP) consists of alternately constructing the two sequences $\{(x_k, y_k)\}$ $\{(x_k^*, y_k^*)\}$ starting $(x_0, y_0) \in \text{Dom } g$. The terminology "alternative DCA" means that at each step, we apply DCA alternatively for solving a DC program in variables x and y. The basis of the algorithm is described as follows.

ADCA: Alternative DC Algorithm

1. Initialization: Given $(x_0, y_0) \in \text{Dom } g$.
2. For $k = 0, 1, ...$, generating the sequences $\{(x_k, y_k)\}$ and $\{(x_k^*, y_k^*)\}$ by
   - Compute $x_k^* \in \partial_x h(x_k, y_k)$.
   - Compute $x_{k+1} \in \partial_x g^*(\cdot, y_k)(x_k^*)$; i.e., x_{k+1} is a solution of the convex program

   $$\min\{g(x, y_k) - \langle x, x_k^* \rangle : \; x \in \mathbb{R}^n\}. \tag{4}$$

   - Compute $y_k^* \in \partial_y h(x_{k+1}, y_k)$.
   - Compute $y_{k+1} \in \partial_y g^*(x_{k+1}, \cdot)(y_k^*)$ by solving the convex program

   $$\min\{g(x_{k+1}, y) - \langle y_k^*, y \rangle : \; y \in \mathbb{R}^m\}. \tag{5}$$

3. If stopping criterion is met, then stop and we have (x_{k+1}, y_{k+1}) is the computed solution, otherwise, set $k = k + 1$ and go to Step 2.

Output (x_k, y_k) and $f(x_k, y_k)$ as the best known solution and objective function value.

Due to the inexactness of practical computations, more general, we consider an inexact version of (ADCA) as follows. Recall that $x_\varepsilon \in \mathbb{R}^n$ is an $\varepsilon-$solution of the problem $\inf\{f(x) : \quad x \in \mathbb{R}^n\}$, for some $\varepsilon > 0$ if

$$f(x_\varepsilon) \leq f(x) + \varepsilon \quad \text{for all } x \in \mathbb{R}^n.$$

IADCA: Inexact Alternative DC Algorithm

1. Initialization: Given $(x_0, y_0) \in \text{Dom } g$ and $\varepsilon_0 \geq 0$.
2. For $k = 0, 1, ...$, update $\varepsilon_k \geq 0$ and generating the sequences $\{(x_k, y_k)\}$ and $\{(x_k^*, y_k^*)\}$ by
   - Compute $x_k^* \in \partial_x^{\varepsilon_k} h(x_k, y_k)$.
   - Compute $x_{k+1} \in \partial_x^{\varepsilon_k} g^*(\cdot, y^k)(x_k^*)$; i.e., x_{k+1} is an ε_k-solution of the convex program

   $$\min\{g(x, y_k) - \langle x, x_k^* \rangle : \; x \in \mathbb{R}^n\}. \tag{6}$$

- Compute $y_k^* \in \partial_y^{\varepsilon_k} h(x_{k+1}, y_k)$.
- Compute $y_{k+1} \in \partial_y^{\varepsilon_k} g^*(x_{k+1}, \cdot)(y_k^*)$ by searching an ε_k-solution the convex program

$$\min\{g(x_{k+1}, y) - \langle y_k^*, y\rangle : \ y \in \mathbb{R}^m\}. \tag{7}$$

3. If stopping criterion is met, then stop and we have (x_{k+1}, y_{k+1}) is the computed solution, otherwise, set $k = k+1$ and go to Step 2.

3.1 Convergence Analysis of PDCA

Let $g, h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be lower semicontinuous partially convex functions. For each$x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, denote by $\rho(g, y)$, $\rho(g, x)$ and $\rho(h, x)$, $\rho(h, y)$ the convex modulus of $g(\cdot, y)$, $g(\cdot, y)$, $g(\cdot, y)$, $g(\cdot, y)$, respectively. For $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, set

$$\rho(y) := \rho(g, y) + \rho(h, y) \quad \text{and} \quad \rho(x) := \rho(g, x) + \rho(h, x).$$

Let us Consider the partial DC program (PDCP). Firstly, we establish some lemmata which are needed in the proof of the convergence results.

Lemma 3.1 *Let $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous convex function and $\varepsilon > 0$. For $\in \text{Dom } \varphi$ and $x^* \in \partial^\varepsilon \varphi(x)$, there exists $z \in \bar{B}(x, \sqrt{\varepsilon})$ such that*

$$x^* \in \partial\varphi(z) + \sqrt{\varepsilon} B_{\mathbb{R}^n}.$$

Proof. By the definition,

$$\langle x^*, y - x\rangle \le \varphi(y) - \varphi(x) + \varepsilon, \ \forall y \in \mathbb{R}^n,$$

equivalently,

$$g(x) - \langle x^*, x\rangle \le g(y) - \langle x^*, y\rangle + \varepsilon, \ \forall y \in \mathbb{R}^n.$$

By applying the Ekeland variational principle ([11]) to the function $y \mapsto g(y) - \langle x^*, y\rangle$, we find $z \in \bar{B}(x, \sqrt{\varepsilon})$ such that

$$g(z) - \langle x^*, z\rangle \le g(y) - \langle x^*, y\rangle + \sqrt{\varepsilon}\|y - z\|, \ \forall y \in \mathbb{R}^n.$$

Therefore,

$$x^* \in \partial(g + \sqrt{\varepsilon}\| \cdot - z\|)(z) = \partial\varphi(z) + \sqrt{\varepsilon} B_{\mathbb{R}^n}.$$

□

Lemma 3.2 *Let $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a $\rho-$convex function. For any $\varepsilon \ge 0$, $t \in (0, 1)$ and any $x^* \in \partial^\varepsilon \varphi(x)$, $x \in \text{Dom } \varphi$, one has*

$$\langle x^*, z - x\rangle \le \varphi(z) - \varphi(x) - \frac{\rho(1-t)}{2}\|z - x\|^2 + \varepsilon/t \ \ \forall z \in \mathbb{R}^n.$$

Proof. As $x^* \in \partial^\varepsilon \varphi(x)$, one has

$$\langle x^*, z - x\rangle \le \varphi(z) - \varphi(x) + \varepsilon \ \ \forall z \in \mathbb{R}^n.$$

For $z \in \mathbb{R}^n$ and $t \in (0,1)$,

$$\varphi(x + t(z-x)) \leq t\varphi(z) + (1-t)\varphi(x) - \frac{\rho t(1-t)}{2}\|z-x\|^2.$$

By taking $x + t(z-x)$ into account of z in the first relation, and in using this iequality, one obtains

$$\langle x^*, z-x\rangle \leq \varphi(z) - \varphi(x) - \frac{\rho(1-t)}{2}\|z-x\|^2 + \varepsilon/t.$$

. □

The following lemma is crucial to establish the convergence of (ADCA).

Lemma 3.3 *Let* $\{w_k\} := \{(x_k, y_k)\}$ *be a sequence generated by (IADCA), with respect to a updated sequence* $\{\varepsilon_k\}$ *with* $\varepsilon_k \geq 0$ *for all* $k \geq 0$*..*

(i) For every $k = 0, 1, ...,$ *one has*

$$\frac{\rho(y_k)}{4}\|x_{k+1} - x_k\|^2 + \frac{\rho(x_{k+1})}{4}\|y_{k+1} - y_k\|^2 \leq f(x_{k+1}, y_{k+1}) - f(x_k, y_k) + 8\varepsilon_k.$$

As a result, for the exact algorithm ADCA, i.e., $\varepsilon_k = 0$ *for all* $k \geq 0$*, then the sequence* $\{f(x_k, y_k)\}$ *is nonincreasing.*

(ii) If $\sum_{k=1}^{\infty} \varepsilon_k < \infty$ *and*

$$\liminf_{k\to\infty} \min\{\rho(y_k), \rho(x_k)\} > 0 \quad \textit{and} \quad \liminf_{k\to\infty} f(x_k, y_k) > -\infty,$$

then

$$\sum_{k=1}^{\infty} \|w_k - w_{k+1}\|^2 < +\infty, \textit{ and therefore } \lim_{k\to\infty} \|w_k - w_{k+1}\| = 0.$$

Proof. (i). As $x_k^* \in \partial_x^{\varepsilon_k} h(x_k, y_k)$, by using Lemma 3.2 with $t = 1/2$, one has

$$\langle x_k^*, x_{k+1} - x_k\rangle \leq h(x_{k+1}, y_k) - h(x_k, y_k) - \frac{\rho(h, y_k)}{4}\|x_{k+1} - x_k\|^2 + 2\varepsilon_k. \tag{8}$$

Since $x_{k+1} \in \partial^{\varepsilon_k} g^*(\cdot, y_k)(x_k^*)$, equivalently, $x_k^* \in \partial_x^{\varepsilon_k} g(x_{k+1}, y_k)$, one also has

$$\langle x_k^*, x_k - x_{k+1}\rangle \leq g(x_k, y_k) - g(x_{k+1}, y_k) - \frac{\rho(g, y_k)}{4}\|x_{k+1} - x_k\|^2 + 2\varepsilon_k. \tag{9}$$

By adding inequalities (8) and (9), one obtains

$$\frac{\rho(y_k)}{4}\|x_{k+1} - x_k\|^2 \leq f(x_k, y_k) - f(x_{k+1}, y_k) + 4\varepsilon_k. \tag{10}$$

Similarly, one has

$$\langle y_k^*, y_{k+1} - y_k\rangle \leq h(x_{k+1}, y_{k+1}) - h(x_{k+1}, y_k) - \frac{\rho(h, x_{k+1})}{4}\|y_{k+1} - y_k\|^2 + 2\varepsilon_k,$$

and

$$\langle y_k^*, y_k - y_{k+1}\rangle \le g(x_{k+1}, y_k) - g(x_{k+1}, y_{k+1}) - \frac{\rho(g, x_{k+1})}{4}\|y_{k+1} - y_k\|^2 + 2\varepsilon_k,$$

which imply

$$\frac{\rho(x_{k+1})}{4}\|y_{k+1} - y_k\|^2 \le f(x_{k+1}, y_k) - f(x_{k+1}, y_{k+1}) + 4\varepsilon_k. \tag{11}$$

By adding (10) and (11), one obtains (i).

(ii). By the assumption, there is an index $k_0 \ge 0$ such that for some $\rho > 0$,

$$\min\{\rho(x_k), \rho(y_k)\} \ge \rho > 0 \quad \text{for all} \quad k \ge k_0.$$

Then, according to part (i), one has

$$\begin{aligned} \tfrac{\rho}{4}\textstyle\sum_{k=k_0}^{\infty}\|w_{k+1} - w_k\|^2 &\le \textstyle\sum_{k=k_0}^{\infty}\left(\frac{\rho(y_k)}{4}\|x_{k+1} - x_k\|^2 + \frac{\rho(x_{k+1})}{4}\|y_{k+1} - y_k\|^2\right) \\ &\le f(x_{k_0}, y_{k_0}) - \liminf_{k\to\infty} f(x_k, y_k) + 8\textstyle\sum_{k=k_0}^{\infty}\varepsilon_k < +\infty. \end{aligned}$$

Thus (ii) is proved. □

By making use of this lemma, we establish the following convergence result.

Theorem 3.1 *Consider problem (PDCP). Assume that $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous and $g, h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ are lower semicontinuous partially convex functions and that for each $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, the function $g(\cdot, y)$; $h(\cdot, y)$ and $g(x, \cdot)$; $h(x, \cdot)$ are continuous. Let $\{w_k\} := \{(x_k, y_k)\}$ be a sequence generated by (IADCA), with respect to a updated sequence $\{\varepsilon_k\}$ with $\varepsilon_k \ge 0$ for all $k \ge 0$ such that*

$$\sum_{k=0}^{\infty}\varepsilon_k < +\infty \quad \textit{and} \quad \liminf_{k\to\infty}\min\{\rho(y_k), \rho(x_k)\} > 0.$$

Then for any subsequence $\{w_{k_l}\} := \{(x_{k_l}, y_{k_l})\}$ of $\{w_k\}$, converging to $\bar{w} = (\bar{x}, \bar{y})$ such that $\{(x_{k_l}^*, x_{k_l}^*)\}$ is bounded, the point $\bar{w}$ is a weak critical point of problem (PDCP).

Proof. Let $\{w_{k_l}\} := \{(x_{k_l}, y_{k_l})\}$ of $\{w_k\}$ such that it converges to $\bar{w} = (\bar{x}, \bar{y})$ and $\{w_{k_l}^*\} := \{(x_{k_l}^*, x_{k_l}^*)\}$ is bounded. Without loss of generality, assume that the sequence $\{w_k^*\}$ converges to w^*. By the lower semicontinuity of f, there is an index l_0 such that

$$f(x_{k_l}, y_{k_l}) > f(\bar{x}, \bar{y}) - 1 \quad \forall l \ge l_0.$$

For any $k \ge 0$, pick an index $l \ge l_0$ such that $k \le k_l$. Thanks to Lemma 3.3 (i), one has

$$f(x_k, y_k) \ge f(x_{k_l}, y_{k_l}) - 8\sum_{i=k}^{k_l-1}\varepsilon_i \ge f(\bar{x}, \bar{y}) - 8\sum_{i=0}^{\infty}\varepsilon_i - 1.$$

Thus $\liminf_{k\to\infty} f(x_k, y_k) > -\infty$. By virtue of Lemma 3.3 (ii), one obtains $\lim_{k\to\infty}\|w_{k+1} - w_k\| = 0$. Therefore, $\lim_{l\to\infty} w_{k_l} = \lim_{l\to\infty} w_{k_l+1} = \bar{w}$.

As $x_{k_l}^* \in \partial_x^{\varepsilon_k} h(x_{k_l}, y_{k_l})$,

$$\langle x_{k_l}^*, x - x_{k_l}\rangle \le h(x, y_{k_l}) - h(x_{k_l}, y_{k_l}) + \varepsilon_{k_l} \quad \text{for all} \quad x \in \mathbb{R}^n.$$

By letting $l \to \infty$, since $\lim_{k\to\infty} \varepsilon_k = 0$; $\liminf_{l\to\infty} h(x_{k_l}, y_{k_l}) \geq h(\bar{x}, \bar{y})$ as well as $\lim_{l\to\infty} h(x, y_{k_l}) = h(x, \bar{y})$, for each $x \in \mathbb{R}^n$, one obtains

$$\langle x^*, x - \bar{x}\rangle \leq h(x, \bar{y}) - h(\bar{x}, \bar{y}) \quad \forall x \in \mathbb{R}^n,$$

that is, $x^* \in \partial_x h(\bar{x}, \bar{y})$. Next, since $x_{k_l+1} \in \partial_x^{\varepsilon_{k_l}} g^*(\cdot, y_{k_l})(x^*_{k_l})$, equivalently,

$$\langle x^*_{k_l}, x - x_{k_l+1}\rangle \leq g(x, y_{k_l}) - g(x_{k_l+1}, y_{k_l}) + \varepsilon_{k_l} \quad \text{for all} \quad x \in \mathbb{R}^n.$$

As before, let $l \to \infty$ to obtain $x^* \in \partial_x g(\bar{x}, \bar{y})$, that shows

$$x^* \in \partial_x g(\bar{x}, \bar{y}) \cap \partial_x h(\bar{x}, \bar{y})$$

Similarly, we also derive

$$y^* \in \partial_y g(\bar{x}, \bar{y}) \cap \partial_y h(\bar{x}, \bar{y}).$$

That is, $\bar{w}$ is a weak critical point of (PDCP). □

Next, we establish the convergence results of the whole sequence $\{x_k\}$ generated by ADCA and by IADCA when the cost function is subanalytic. The proof of the following convergence theorems are based on the nonsmooth version of the Lojasiewicz inequality (Theorem 3.1, [7]), inspired by Lojasiewicz [18] and Attouch- Bolte [2]. We make use of the following assumptions:

(A1) The function g is differentiable with locally Lipschitz drivative, and for each $(x, y) \in \text{Dom } f$, the partial subdifferentials $\partial_x h(u, \cdot)$, $\partial_y h(\cdot, v)$ are locally Lipschitz uniformly in u, v, respectively around (x, y) on Dom h, that is, for each $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$, there exist $\kappa, \delta > 0$ such that

$$\partial_x h(u, v_1) \subseteq \partial_x h(u, v_2) + \kappa\|v_1 - v_2\| B_{\mathbb{R}^n} \quad \forall (u, v_1), (u, v_2) \in \text{Dom } h \cap B((x, y), \delta); \tag{12}$$

$$\partial_y h(u_1, v) \subseteq \partial_y h(u_2, v) + \kappa\|u_1 - u_2\| B_{\mathbb{R}^m} \quad \forall (u_1, v), (u_2, v) \in \text{Dom } h \cap B((x, y), \delta). \tag{13}$$

(A2) The function h is differentiable with locally Lipschitz deriative, and for each $(x, y) \in \text{Dom } f$, $\partial_x g(u, \cdot)$, $\partial_y g(\cdot, v)$ are locally Lipschitz uniformly in u, v, respectively around (x, y) on Dom g.

Obviously, (A1) is satisfied if g is differentiable with locally Lipschitz drivative and h is of the form

$$h(x, y) := h_1(x) + h_2(y) + Q(x, y), \;\; (x, y) \in \mathbb{R}^m \times \mathbb{R}^n,$$

where, Q is continuously differentiable with locally Lipschitz derivative. And so (A2) is satisfied when h is differentiable with locally Lipschitz drivative; g is

$$g(x, y) := g_1(x) + g_2(y) + Q(x, y), \;\; (x, y) \in \mathbb{R}^m \times \mathbb{R}^n.$$

Note that if (A1) is satisfied , then the assumption of Corollary 2.1 is fulfilled for the function h at all $(x, y) \in \text{Dom } f$, therefore, one has, for $(x, y) \in \text{Dom } f$,

$$\partial^F(-f)(x, y) = \partial_x h(x, y) \times \partial_y h(x, y) - \nabla g(x, y). \tag{14}$$

Similarly, if (A2) is satisfied , then the assumption of Corollary 2.1 is fulfilled for the function g, and one has, for $(x, y) \in \text{Dom } f$,

$$\partial^F f(x,y) = \partial_x g(x,y) \times \partial_y g(x,y) - \nabla h(x,y). \tag{15}$$

For $w_0 \in \mathbb{R}^n \times \mathbb{R}^m$ and a sequence of positives $\{\varepsilon_k\}$, set

$$C(w_0, \{\varepsilon_k\}) = \left\{ \begin{array}{ll} \bar{w} \in \mathbb{R}^m \times \mathbb{R}^n : & \bar{w} \text{ is a limit point of a sequence generated by IDCA} \\ & \text{from the starting point } w_0, \text{ with respect to } \{\varepsilon_k\} \end{array} \right\}.$$

Theorem 3.2 *Let us consider partial DC problem (PDCP). Let $(x_0, y_0) \in$ Dom f be given. Let the sequences $\{w_k\} = \{(x_k, y_k)\}$ and $\{w_k^*\} = \{(x_k^*, y_k^*)\}$ are generated by IADCA . Suppose that the following assumptions are satisfied:*

(i) The function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is subanalytic, and Dom f *is closed; $f|_{\text{Dom } f}$ is locally Lipschitz.*

(ii) The function $g, h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ are lower semicontinuous partially convex; and either (A1) or (A2) is satisfied.

(iii) There exists a sequence $\{r_k\}$ of positive reals such that $C(w_0, \{r_k\})$ is bounded.

Then there exists a sequence $\{\beta_k\}$ of positive reals such that for all sequences $\{x_k\}$ generated by IADCA from the starting point w_0 with respect to an updated sequence $\{\varepsilon_k\}$ satisfying

$$\liminf_{k \to \infty} \min\{\rho(x_k), \rho(y_k)\} > \rho > 0; \tag{16}$$

$$0 \le \varepsilon_k \le \min\left\{r_k, \beta_k \|w_k - w_{k-1}\|^2\right\}, \quad k = 1, 2, ..., \tag{17}$$

Then, the sequence $\{w_k\}$ converges to a (weak) critical point (PDCP). Moreover, under (A1) this weak limit point is a Clarke critical point, and under (A2), it is a Fréchet critical point.

Proof. Consider the closure hull $\bar{C}(w_0, \{r_k\})$ of $C(w_0, \{r_k\})$. Then, $\bar{C}(w_0, \{r_k\})$ is compact. According to the Lojasiewicz inequality, applied for the subanalytic functions f and $-f$, by shrinking $\varepsilon(x)$ if necessary, for each $w \in \bar{C}(w_0, \{r_k\})$, we can find $L(w) > 0$; $\theta(w) \in [0, 1)$ such that

$$|f(u) - f(w)|^{\theta(w)} \le L(w)\|u^*\| \quad \text{for all } u \in B(w, \varepsilon(w));\ u^* \in \partial^F f(u) \cup \partial^F(-f)(u). \tag{18}$$

By the compactness of $\bar{C}(w_0, \{r_k\})$, there exist $z_1, z_2, ..., z_p \in \bar{C}(w_0, \{r_k\})$ such that $\bar{C}(w_0, \{r_k\}) \subseteq \bigcup_{i=1}^p B(z_i, \varepsilon(z_i)/2)$. Set

$$\varepsilon := \min\{\varepsilon(z_i)/2 :\ i = 1, ..., p\}, \quad \theta = \max\{\theta(z_i) :\ i = 1, ..., p\},$$

and $L = \max\{L(z_i) :\ i = 1, ..., p\}$. Let $w \in \bar{C}(w_0, \{r_k\})$. There is some z_i, such that $w \in B(z_i, \varepsilon(z_i)/2)$. Moreover, we can find $\delta_w \in (0, \varepsilon)$ such that $|f(u) - f(w)| < 1$ for $u \in \text{Dom } f \cap B(w, \delta_w)$ and

$$|f(u) - f(w)| \le |f(u) - f(z_i)| \quad \text{for all } u \in B(w, \delta_w).$$

Since $B(w,\delta_w) \subseteq B(z_i, \varepsilon(z_i))$, from (18), one derives

$$\begin{aligned} |f(u)-f(w)|^{\theta} &\le |f(u)-f(w)|^{\theta(z_i)} \le |f(u)-f(z_i)|^{\theta(z_i)} \\ &\le L\|u^*\| \quad \text{for all } u \in B(w,\delta_w);\ u^* \in \partial^F f(u) \cup \partial^F(-f)(u). \end{aligned} \tag{19}$$

Let $\{\beta_k\}$ be a sequence of positives such that $\sum_{k=0}^{\infty} \beta_k^{\theta/2} < \infty$. Let $\{\varepsilon_k\}$ and $\{x_k\}$ generated by IADCA such that (16) and (17) are satisfied. According to Lemma 3.3, when k is sufficiently large, one has

$$\frac{\rho}{4}\|w_k - w_{k+1}\|^2 \le f(w_k) - f(w_{k+1}) + 8\varepsilon_k. \tag{20}$$

Without loss of generality, assume that $\beta_k < \rho_k/64$ and (20) holds for all $k = 0, 1, \ldots$. By the way, $\sum_{k=0}^{\infty} \epsilon_k < \infty$, and $\{w_k\}$ is bounded. Furthermore, under (A1) or (A2), according to Lemma 3.1, the sequence $\{w_k^*\}$ is also bounded. By virtue of Lemma 3.3, $\sum_{k=0}^{\infty} \|w_k - w_{k+1}\|^2 < \infty$, and each limit point of $\{w_k\}$ is a weak critical point of (PDCP). Pick a subsequence $\{w_{k_l}\}$ of $\{w_k\}$ such that $\{f(w_{k_l})\}$ is convergent. For k sufficiently large, and for indices l_1, l_2 such that $k_{l_1} < k < k_{l_2}$, by (20), one has

$$f(w_{k_{l_1}}) - f(x_k) = \sum_{j=k_{l_1}}^{k-1} [f(w_j) - f(w_{j+1})] \ge \frac{\rho}{4} \sum_{j=k_{l_1}}^{k-1} \|w_j - w_{j+1}\|^2 - 8 \sum_{j=k_{l_1}}^{k-1} \varepsilon_j;$$

$$f(w_k) - f(w_{k_{l_2}}) = \sum_{j=k}^{k_{l_2}-1} [f(w_j) - f(w_{j+1})] \ge \frac{\rho}{4} \sum_{j=k}^{k_{l_2}-1} \|w_j - w_{j+1}\|^2 - 8 \sum_{j=k}^{k_{l_2}-1} \varepsilon_j.$$

These inequality together the convergence of $\sum_{k=0}^{\infty} \|w_k - w_{k+1}\|^2$ and $\sum_{k=0}^{\infty} \epsilon_k$ yield the convergence of the whole sequence $\{f(w_k)\}$. Without loss of generality, assume that $\lim_{k\to\infty} f(w_k) = 0$. Moreover,for $k = 0, 1, \ldots$, one has

$$f(w_{k+1}) + 8\varepsilon_{k+1} \le f(w_{k+1}) + \frac{\rho}{4}\|w_k - w_{k+1}\|^2 \le f(w_k) + 8\varepsilon_k.$$

That is, the sequence $\{f(w_k) + 8\varepsilon_k\}$ is a decreasing sequence, and therefore,

$$f(w_k) + 8\varepsilon_k \ge 0, \quad \text{for all } k = 0, 1, \ldots.$$

Denote by $\mathcal{C}$ the set of all limit points of $\{w^k\}$. By assumption, $\{w^k\}$ is bounded, then obviously $\mathcal{C}$ is a compact set, and moreover $\mathcal{C} \subseteq \bar{C}(w_0, \{r_k\})$. Let us consider the following cases.

Case 1. (A1) is verified. By assumptions, for each $w \in \mathcal{C}$, pick $\kappa_w, \delta_w > 0$ such that

$$\|\nabla g(w') - \nabla g(w'')\| \le \kappa_w \|w' - w''\| \quad \forall w', w'' \in B(w, \delta_w);$$

$$|f(w') - f(w'')| \le \kappa_w \|w' - w''\| \quad \forall w', w'' \in B(w, \delta_w) \cap \text{Dom } f;$$

$$\partial_x h(u, v_1) \subseteq \partial_x h(u, v_2) + \kappa_w \|v_1 - v_2\| B_{\mathbb{R}^n} \quad \forall (u, v_1), (u, v_2) \in B(w, \delta_w) \cap \text{Dom } f;$$

$$\partial_y h(u_1, v) \subseteq \partial_y h(u_2, v) + \kappa_w \|u_1 - u_2\| B_{\mathbb{R}^m} \quad \forall (u_1, v), (u_2, v) \in B(w, \delta_w) \cap \text{Dom } f.$$

By virtue of (19), assume that for the same δ_w as above, one has

$$|f(w')|^{\theta} \le L\|w^*\| \quad \text{for all } w' \in B(w, \delta_w);\ u^* \in \partial^F f(w') \cup \partial^F(-f)(w').$$

By the compactness of $\mathcal{C}$, we derive easily that there are some $\delta > 0$ and $\kappa > 0$ such that

$$|f(w)|^\theta \le L\|w^*\| \quad \text{for all } w \in \mathbb{R}^n \times \mathbb{R}^m \text{ with } d_{\mathcal{C}}(w) < \delta,\ w^* \in \partial^F f(u) \cup \partial^F(-f)(u), \tag{21}$$

and

$$\|\nabla g(w') - \nabla g(w'')\| \le \kappa\|w' - w''\| \quad \forall w', w'' \in B(w, \delta_w) \text{ with } d_{\mathcal{C}}(w') < \delta,\ \|w' - w''\| < \delta; \tag{22}$$

$$|f(w') - f(w'')| \le \kappa\|w' - w''\| \quad \forall w', w'' \in \text{Dom } f \text{ with } d_{\mathcal{C}}(w') < \delta,\ \|w' - w''\| < \delta; \tag{23}$$

$$\partial_x h(u, v_1) \subseteq \partial_x h(u, v_2) + \kappa\|v_1 - v_2\| B_{\mathbb{R}^n} \quad \forall (u, v_1), (u, v_2) \in \text{Dom } f \text{ with } d_{\mathcal{C}}((u, v_1)) < \delta,\ \|v_1 - v_2\| < \delta; \tag{24}$$

$$\partial_y h(u_1, v) \subseteq \partial_y h(u_2, v) + \kappa\|u_1 - u_2\| B_{\mathbb{R}^m} \quad \forall (u_1, v), (u_2, v) \in \text{Dom } f \text{ with } d_{\mathcal{C}}((u_1, v)) < \delta,\ \|u_1 - u_2\| < \delta. \tag{25}$$

Let an index k_0 such that

$$d_{\mathcal{C}}(w_k) < \delta/4;\ \|w_k - w_{k+1}\| < \delta/4 \ \text{ and } \varepsilon_k^{1/2} < \delta/4 \ \ \forall k \ge k_0.$$

Let $k \ge k_0$ be arbitrary. Since

$$x_k^* \in \partial_x^{\varepsilon_k} h(x_k, y_k) \cap \partial_x^{\varepsilon_k} g(x_{k+1}, y_k),$$

Lemma 3.1 yields the existence of $u_k \in \bar{B}(x_k, \sqrt{\varepsilon_k})$ and of $u_k' \in \bar{B}(x_{k+1}, \sqrt{\varepsilon_k})$ such that

$$x_k^* \in \partial_x h(u_k, y_k) + \sqrt{\varepsilon_k} B_{\mathbb{R}^n}; \tag{26}$$

$$\begin{aligned} x_k^* \in \ & \nabla_x g(u_k', y_k) + \sqrt{\varepsilon_k} B_{\mathbb{R}^n} \subseteq \nabla_x g(u_k, y_k) + (\kappa\|u_k - u_k'\| + \sqrt{\varepsilon_k}) B_{\mathbb{R}^n} \\ & \subseteq \nabla_x g(u_k, y_k) + (\kappa\|x_k - x_{k+1}\| + (2\kappa + 1)\sqrt{\varepsilon_k}) B_{\mathbb{R}^n}. \end{aligned} \tag{27}$$

Similarly, as

$$y_k^* \in \partial_y^{\varepsilon_k} h(x_{k+1}, y_k) \cap \partial_x^{\varepsilon_k} g(x_{k+1}, y_{k+1}),$$

there exist $v_k \in \bar{B}(y_k, \sqrt{\varepsilon_k})$ and of $v_k' \in \bar{B}(y_{k+1}, \sqrt{\varepsilon_k})$ such that

$$y_k^* \in \partial_y h(x_{k+1}, v_k) + \sqrt{\varepsilon_k} B_{\mathbb{R}^m}; \tag{28}$$

$$\begin{aligned} y_k^* \in \ & \nabla_y g(x_{k+1}, v_k') + \sqrt{\varepsilon_k} B_{\mathbb{R}^m} \subseteq \nabla_y g(x_{k+1}, v_k) + (\kappa\|v_k - v_k'\| + \sqrt{\varepsilon_k}) B_{\mathbb{R}^m} \\ & \subseteq \nabla_y g(x_{k+1}, v_k) + (\kappa\|y_k - y_{k+1}\| + (2\kappa + 1)\sqrt{\varepsilon_k}) B_{\mathbb{R}^n}. \end{aligned} \tag{29}$$

Note that (A1) guarantees that the assumption of Corollary 2.1 is satisfied for the functtion h (at points under consideration). Hence,

$$\partial^F(-f)(u_k, v_k) = (\partial_x h(u_k, v_k) - \nabla_x g(u_k, v_k)) \times (\partial_y h(u_k, v_k) - \nabla_y g(u_k, v_k)) = \partial_x f(u_k, v_k) \times \partial_y f(u_k, v_k). \tag{30}$$

From (26) and (28), by (A2), one has

$$\begin{aligned} (x_k^*, y_k^*) \in \ & \partial h(u_k, v_k) + (\kappa(\|y_k - v_k\| + \|u_k - x_{k+1}\|) + 2\sqrt{\varepsilon_k}) B_{\mathbb{R}^n \times \mathbb{R}^m} \\ & \subseteq \partial h(u_k, v_k) + (\kappa\|w_k - w_{k+1}\| + (2\kappa + 2)\sqrt{\varepsilon_k}) B_{\mathbb{R}^n \times \mathbb{R}^m}. \end{aligned}$$

From (27) and (29),

$$\begin{aligned} (x_k^*, y_k^*) \in \ & \nabla g(u_k, v_k) + (\kappa(\|y_k - v_k\| + \|u_k - x_{k+1}\| + \|w_k - w_{k+1}\|) + 2\sqrt{\varepsilon_k}) B_{\mathbb{R}^n \times \mathbb{R}^m} \\ & \subseteq \nabla g(u_k, v_k) + (2\kappa\|w_k - w_{k+1}\| + (4\kappa + 2)\sqrt{\varepsilon_k}) B_{\mathbb{R}^n \times \mathbb{R}^m}. \end{aligned}$$

The latter inclusions together relation (30) imply that there exists $(u_k^*, v_k^*) \in \partial^F(-f)(u_k, v_k)$ such that

$$\|(u_k^*, v_k^*)\| \le 3\kappa\|w_k - w_{k+1}\| + (6\kappa+3)\sqrt{\varepsilon_k}. \tag{31}$$

Hence, by (21),

$$|f(u_k, v_k)|^\theta \le L[3\kappa\|w_k - w_{k+1}\| + (6\kappa+3)\sqrt{\varepsilon_k}]. \tag{32}$$

Then, (23) yields

$$\begin{aligned} 0 \le (f(x_k, y_k) + 8\varepsilon_k)^\theta & \le (|f(u_k, v_k)| + 2\kappa\sqrt{\varepsilon_k} + 8\varepsilon_k)^\theta \\ & \le |f(u_k, v_k)|^\theta + (2\kappa)^\theta \varepsilon_k^{\theta/2} + (8\varepsilon_k)^\theta \\ & \le 3L\kappa\|w_k - w_{k+1}\| + \alpha_k. \end{aligned} \tag{33}$$

Where, $\alpha_k := L(6\kappa+3)\varepsilon_k^{1/2} + (2\kappa)^\theta\varepsilon_k^{\theta/2} + (8\varepsilon_k)^\theta$. By the definition of $\{\varepsilon_k\}$ and the boundedness of $\{x_k\}$, then $\sum_{k=0}^\infty \alpha_k < \infty$.

Set $\gamma_k = f(w_k) + 8\varepsilon_k$, $k = 0, 1, ...$ By using the concavity of the function $t \in \mathbb{R} \mapsto t^{1-\theta}$ on $(0, +\infty)$, from relation (20) and the preceding inequality, one derives that

$$\gamma_k^{1-\theta} - \gamma_{k+1}^{1-\theta} \ge (1-\theta)\gamma_k^{-\theta}(\gamma_k - \gamma_{k+1}) \ge \frac{(1-\theta)[\rho\|w_k - w_{k+1}\|^2/4 - 8\varepsilon_{k+1}]}{3L\|w_k - w_{k+1}\| + \alpha_k}. \tag{34}$$

As

$$0 \le \varepsilon_{k+1} \le \beta_k\|w_k - w_{k+1}\|^2 \le \rho\|w_k - w_{k+1}\|^2/64,$$

it follows that

$$\frac{\|w_k - w_{k+1}\|^2}{\|w_k - w_{k+1}\| + \frac{\alpha_k}{3L\kappa}} \le \frac{24L\kappa}{\rho(1-\theta)}(\gamma_k^{1-\theta} - \gamma_{k+1}^{1-\theta}).$$

By using the inequality

$$b \le \frac{b^2}{a} + a/4, \quad a > 0,\ b \ge 0,$$

one obtains

$$\begin{aligned} \|w_k - w_{k+1}\| & \le \tfrac{\|w_k - w_{k+1}\|^2}{\|w_k - w_{k+1}\| + \frac{\alpha_k}{3L\kappa}} + \tfrac{1}{4}\|w_k - w_{k+1}\| + \tfrac{\alpha_k}{12L\kappa} \\ & \le \tfrac{24L\kappa}{\rho(1-\theta)}(\gamma_k^{1-\theta} - \gamma_{k+1}^{1-\theta}) + \tfrac{1}{4}\|w_k - w_{k+1}\| + \tfrac{\alpha_k}{12L\kappa}, \end{aligned}$$

which follows that

$$\|w_k - w_{k+1}\| \le \frac{32L\kappa}{\rho(1-\theta)}(\gamma_k^{1-\theta} - \gamma_{k+1}^{1-\theta}) + \frac{\alpha_k}{9L\kappa}.$$

Therefore,

$$\sum_{k=k_0}^{\infty} \|w_k - w_{k+1}\| \le \frac{32L\kappa}{\rho(1-\theta)}\gamma_{k_0}^{1-\theta} + \frac{1}{9L\kappa}\sum_{k=k_0}^{\infty} \alpha_k < \infty.$$

Thus, the sequence $\{w_k\}$ converges to an unique limit point.

Case 2. (A2) is satisfied. Similarly to Case 1, we can find $\delta > 0$ and $\kappa > 0$ such that

$$|f(w)|^\theta \le L\|w^*\| \quad \text{for all } w \in \mathbb{R}^n \times \mathbb{R}^m \text{ with } d_{\mathcal{C}}(w) < \delta,\ w^* \in \partial^F f(u), \tag{35}$$

and

$$\|\nabla h(w') - \nabla h(w'')\| \le \kappa\|w' - w''\| \quad \forall w', w'' \in B(w, \delta_w) \text{ with } d_{\mathcal{C}}(w') < \delta,\ \|w' - w''\| < \delta; \tag{36}$$

$$|f(w') - f(w'') \leq \kappa \|w' - w''\| \quad \forall w', w'' \in \text{Dom } f \text{ with } d_{\mathcal{C}}(w') < \delta, \ \|w' - w''\| < \delta; \tag{37}$$

$$\partial_x g(u, v_1) \subseteq \partial_x g(u, v_2) + \kappa \|v_1 - v_2\| B_{\mathbb{R}^n} \quad \forall (u, v_1), (u, v_2) \in \text{Dom } f \text{ with } d_{\mathcal{C}}((u, v_1)) < \delta, \ \|v_1 - v_2\| < \delta; \tag{38}$$

$$\partial_y g(u_1, v) \subseteq \partial_g h(u_2, v) + \kappa \|u_1 - u_2\| B_{\mathbb{R}^m} \quad \forall (u_1, v), (u_2, v) \in \text{Dom } f \text{ with } d_{\mathcal{C}}((u_1, v)) < \delta, \ \|u_1 - u_2\| < \delta. \tag{39}$$

Pick k_0 such that

$$d_{\mathcal{C}}(w_{k-1}) < \delta/4; \ \|w_k - w_{k-1}\| < \delta/4 \ \text{ and } \varepsilon_{k-1}^{1/2} < \delta/4 \ \ \forall k \geq k_0.$$

Consider indices $k \geq k_0$. As

$$x_{k-1}^* \in \partial_x^{\varepsilon_k} h(x_{k-1}, y_{k-1}) \cap \partial_x^{\varepsilon_k} g(x_k, y_{k-1});$$

$$y_{k-1}^* \in \partial_y^{\varepsilon_k} h(x_k, y_{k-1}) \cap \partial_x^{\varepsilon_k} g(x_k, y_k),$$

By following the same arguments as in Case 1, but exchanging the roles of g and h, we can find $(u_k, v_k) \in \bar{B}(x_k, \sqrt{\varepsilon_{k-1}}) \times \bar{B}(y_k, \sqrt{\varepsilon_{k-1}})$, $(u_k^*, v_k^*) \in \partial^F f(u_k, v_k)$ such that

$$\|(u_k^*, v_k^*)\| \leq 3\kappa \|w_k - w_{k-1}\| + (6\kappa + 3)\sqrt{\varepsilon_{k-1}}. \tag{40}$$

And therefore,

$$|f(u_k, v_k)|^\theta \leq L[3\kappa \|w_k - w_{k-1}\| + (6\kappa + 3)\sqrt{\varepsilon_{k-1}}]. \tag{41}$$

Then, (37) yields

$$\begin{aligned} 0 \leq (f(x_k, y_k) + 8\varepsilon_k)^\theta & \leq (|f(u_k, v_k)| + 2\kappa\sqrt{\varepsilon_{k-1}} + 8\varepsilon_k)^\theta \\ & \leq |f(u_k, v_k)|^\theta + (2\kappa)^\theta \varepsilon_{k-1}^{\theta/2} + (8\varepsilon_k)^\theta \\ & \leq 3L\kappa \|w_k - w_{k-1}\| + \alpha_k'. \end{aligned} \tag{42}$$

Here, $\alpha_k' := L(6\kappa + 3)\varepsilon_{k-1}^{1/2} + (2\kappa)^\theta \varepsilon_{k-1}^{\theta/2} + (8\varepsilon_k)^\theta$, and so $\sum_{k=1}^\infty \alpha_k' < \infty$. With $\gamma_k = f(w_k) + 8\varepsilon_k$, $k = 0, 1, ...$ as in Case 1, one has

$$\gamma_k^{1-\theta} - \gamma_{k+1}^{1-\theta} \geq (1-\theta)\gamma_k^{-\theta}(\gamma_k - \gamma_{k+1}) \geq \frac{(1-\theta)[\rho\|w_k - w_{k+1}\|^2/4 - 8\varepsilon_{k+1}]}{3L\kappa \|w_k - w_{k-1}\| + \alpha_k'}, \tag{43}$$

$$\frac{\|w_k - w_{k+1}\|^2}{\|w_k - w_{k-1}\| + \frac{\alpha_k'}{3L\kappa}} \leq \frac{24L\kappa}{\rho(1-\theta)}(\gamma_k^{1-\theta} - \gamma_{k+1}^{1-\theta}).$$

By using again the inequality $b \leq \frac{b^2}{a} + a/4$, for $a > 0$, $b \geq 0$, one derives

$$\begin{aligned} \|w_k - w_{k+1}\| & \leq \tfrac{\|w_k - w_{k+1}\|^2}{\|w_k - w_{k-1}\| + \frac{\alpha_k}{3L\kappa}} + \tfrac{1}{4}\|w_k - w_{k+1}\| + \tfrac{\alpha_k}{12L\kappa} \\ & \leq \tfrac{12L\kappa}{\rho(1-\theta)}(\gamma_k^{1-\theta} - \gamma_{k+1}^{1-\theta}) + \tfrac{1}{4}\|w_k - w_{k-1}\| + \tfrac{\alpha_k'}{12L\kappa}, \end{aligned}$$

which follows that

$$\|w_k - w_{k+1}\| \leq \frac{32L\kappa}{1-\theta}(\gamma_k^{\rho(1-\theta)} - \gamma_{k+1}^{1-\theta}) + \frac{\alpha_k}{9L\kappa}.$$

Therefore, by summing these inequalities for $k = k_0$ to arbitrary N, and then letting $N \to \infty$, on obtains

$$\sum_{k=k_0}^{\infty} \|w_k - w_{k+1}\| \leq \frac{1}{3}\|w_{k_0} - w_{k_0-1}\| + \frac{32L\kappa}{\rho(1-\theta)}\gamma_{k_0}^{1-\theta} + \frac{1}{9L\kappa}\sum_{k=k_0}^{\infty} \alpha_k' < \infty.$$

This implies that the sequence $\{x^k\}$ is convergent. The proof is completed. □

For the exact alternative DC algorithm ADCA, to obtain the convergence result, we need only a weaker assumption that guarantees the Lojasiewicz inequality for f: $f\ |_{\mathrm{Dom}\ f}$ is continuous, instead of $f\ |_{\mathrm{Dom}\ f}$ is locally Lipschitz. The proof of the convergence theorem for ADCA is completely similar although so much simpler to the one of the preceding theorem. Furtheremore, we establish the convergent rates of the sequence $\{w_k\}$ generated by ADCA.

Theorem 3.3 *Let us consider partial DC problem (PDCP). Let $(x_0, y_0) \in \mathrm{Dom}\ f$ be given. Let the sequences $\{w_k\} = \{(x_k, y_k)\}$ and $\{w_k^*\} = \{(x_k^*, y_k^*)\}$ are generated by ADCA . Suppose that the following assumptions are satisfied:*

- *(i) The function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is subanalytic, and $\mathrm{Dom}\ f$ is closed; $f\ |_{\mathrm{Dom}\ f}$ is continuous.*
- *(ii) The function $g, h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ are lower semicontinuous partially convex; and either (A1) or (A2) is satisfied.*
- *(iii) The sequence $\{w_k\}$ is bounded.*

If

$$\liminf_{k\to\infty} \min\{\rho(x_k), \rho(y_k)\} > \rho > 0, \tag{44}$$

then the sequence $\{w_k\}$ converges to a (Clarke) critical point (PDCP). Moreover, under (A2), this critical point is a Fréchet critical point. Moreover, if w^∞ is the limit point of $\{w^k\}$ with a Lojasiewicz exponent $\theta \in [0, 1)$. Then there exists constants $\tau_1, \tau_2 > 0$ such that when k is sufficiently large, one has

$$\|w_k - w_\infty\| \le \sum_{j=k}^{\infty} \|w_j - wj + 1\| \le \tau_1 \|w_k - w_{k-1}\| + \tau_2 \|w_k - w_{k-1}\|^{\frac{1-\theta}{\theta}}. \tag{45}$$

As a result, one has

- *If $\theta \in (1/2, 1)$ then $\|w_k - w_\infty\| \le ck^{\frac{1-\theta}{1-2\theta}}$ for some $c > 0$.*
- *If $\theta \in (0, 1/2]$ then $\|w_k - w_\infty\| \le cq^k$ for some $c > 0; q \in (0, 1)$.*
- *If $\theta = 0$ then $\{w_k\}$ is convergent in a finite number of steps.*

Proof. For $\kappa, L, \theta, \delta, k_0$ with respect to each cases 1 and 2 as in the preceding theorem, we follow the same arguments, but here we work with the subdifferential instead of $\varepsilon-$subdifferential. Let $k \ge k_0$ be arbitrary. Set $\gamma_k = f(x_k)$; $r_k = \sum_{j=k}^{\infty} \|w_j - w_{j+1}\|$, $k \ge k_0$.

Case 1. The relation

$$x_k^* = \nabla_x g(x_{k+1}, y_k) \in \partial_x h(x_k, y_k)$$

implies that

$$x_k^* - \nabla_x g(x_k, y_k) \in \partial_x^F (-f)(x_k, y_k); \ \ \|x_k^* - \nabla_x g(x_k, y_k)\| \le \kappa \|x_k - x_{k+1}\|.$$

The relation

$$y_k^* = \nabla_x g(x_{k+1}, y_{k+1}) \in \partial_y h(x_{k+1}, y_k),$$

together with (A1) imply

$$\begin{aligned} y_k^* - \nabla_y g(x_k, y_k) &\in \partial_y h(x_k, y_k) - \nabla_y g(x_k, y_k) + \kappa\|x_k - x_{k+1}\| \\ &= \partial_y^F(-f)(x_k, y_k) + \kappa\|x_k - x_{k+1}\|; \end{aligned}$$

and

$$\|y_k^* - \nabla_y g(x_k, y_k)\| = \|\nabla_y g(x_{k+1}, y_{k+1}) - \nabla_y g(x_k, y_k)\| \le \kappa\|w_k - w_{k+1}\|.$$

Hence, there is $w_k^* \in \partial^F(-f)(w_k) = \partial_x^F(-f)(w_k) \times \partial_y^F(-f)(w_k)$ such that

$$\|w_k^*\| \le 3\kappa\|w_k - w_{k+1}\|.$$

Therefore,

$$f(w_k)^\theta \le 3L\kappa\|w_k - w_{k+1}\|.$$

As in the proof of Theorem 3.2, one has

$$\|w_k - w_{k+1}\| \le \frac{32L\kappa}{\rho(1-\theta)}(\gamma_k^{1-\theta} - \gamma_{k+1}^{1-\theta}),$$

which follows that

$$r_k = \sum_{j=k}^{\infty} \|w_j - w_{j+1}\| \le \frac{32L\kappa}{\rho(1-\theta)}\gamma_k^{1-\theta} \le \frac{32.3^{(1-\theta)/\theta}(L\kappa)^{1/\theta}}{\rho(1-\theta)}\|w_k - w_{k+1}\|^{(1-\theta)/\theta}.$$

Case 2. Similarly, one derives that there exists $w_k^* \in \partial^F f(w_k)$ such that

$$\|w_k^*\| \le 3\kappa\|w_k - w_{k-1}\|,$$

and therefore,

$$f(w_k)^\theta \le 3L\kappa\|w_k - w_{k-1}\|;$$

$$\|w_k - w_{k+1}\| \le\le \frac{32L\kappa}{\rho(1-\theta)}(\gamma_k^{1-\theta} - \gamma_{k+1}^{1-\theta}) + \frac{1}{4}\|w_k - w_{k-1}\|.$$

Hence,

$$\begin{aligned} r_k &\le \tfrac{r_k + \|w_k - w_{k-1}\|}{4} + \tfrac{32L\kappa}{\rho(1-\theta)}\gamma_k^{1-\theta} \\ &\le \tfrac{r_k + \|w_k - w_{k-1}\|}{4} + \tfrac{32.3^{(1-\theta)/\theta}(L\kappa)^{1/\theta}}{\rho(1-\theta)}\|w_k - w_{k-1}\|^{(1-\theta)/\theta}. \end{aligned}$$

It follows that

$$r_k \le \frac{\|x^k - x^{k-1}\|}{3} + \frac{128.3^{(1-\theta)/\theta}(L\kappa)^{1/\theta}}{3\rho(1-\theta)}\|w_k - w_{k-1}\|^{(1-\theta)/\theta}.$$

Thus (45) is proved, and the sequence $\{w_k\}$ converges to some w_∞.

Since $\lim_{k\to\infty} \|w_k - w_{k+1}\| = 0$, then without loss of generality, we can assume that $\|w_k - w_{k+1}\| < 1$ for all $k = 0, 1, ...$ By setting $\tau = \tau_1 + \tau_2$ in (45), one has

- If $\theta \in (1/2, 1]$, then $r_k \le \tau(r_{k-1} - r_k)^{\frac{1-\theta}{\theta}}$. Hence, by the concavity of the function $t \mapsto t^{\frac{1-2\theta}{1-\theta}}$ on $(0, +\infty)$,

$$r_k^{\frac{1-2\theta}{1-\theta}} - r_{k-1}^{\frac{1-2\theta}{1-\theta}} \ge \frac{1-2\theta}{1-\theta} r_k^{\frac{-\theta}{1-\theta}}(r_k - r_{k-1}) \ge \frac{2\theta - 1}{1-\theta)\tau}.$$

Thus,

$$r_k^{\frac{1-2\theta}{1-\theta}} = r_0^{\frac{1-2\theta}{1-\theta}} + \sum_{j=1}^{k} \left(r_k^{\frac{1-2\theta}{1-\theta}} - r_{k-1}^{\frac{1-2\theta}{1-\theta}} \right) \geq r_0^{\frac{1-2\theta}{1-\theta}} + \frac{2\theta - 1}{1-\theta)\tau} k.$$

Consequently, $r_k \leq ck^{\frac{1-\theta}{1-2\theta}}$, for some $c > 0$.

- If $\theta \in (0, 1/2]$, then $r_k \leq \tau(r_{k-1} - r_k)$, $k = 1.2, ...$ Therefore, $r_k \leq \frac{\tau}{\tau+1} r_{k-1}$. It follows that $r_k \leq cq^k$, where, $q = \tau/(\tau + 1)$; $c = r_0$.
- If $\theta = 0$ then $\|w_k - w_{k+1}\| \geq 1/L$ for $w_k \neq w_{k+1}$ with k sufficiently large. From the inequality

$$f(w_k) - f(w_{k+1}) \geq \frac{\rho}{4} \|x^k - x^{k+1}\|^2 \geq \frac{\rho}{4L^2}$$

for $w_k \neq w_{k+1}$, we derive that (ADCA) terminates after a finite number of iteration steps.

□

With very similar arguments, next we will show that the convergence result for ADCA remains valid for a more general class of functions for which the Kurdyka-Lojasiewicz inequality is satisfied. First we recall this important inequality. Denote by $\mathcal{M}$ the class of continuous concave functions $\varphi : [0, +\infty) \to [0, +\infty)$ verifying:

- $\varphi(0) = 0$ and φ is continuously differentiable on $(0, +\infty)$.
- $\varphi'(t) > 0$ for all $t > 0$.

Recall that a lower semicontinuous function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to verify the *Kurdyka-Lojasiewicz inequality* around $\bar{x} \in \text{Dom } f$ if there exist a function $\varphi \in \mathcal{M}$ and $\delta > 0$ such that

$$(KLI) \quad \varphi'(f(x) - f(\bar{x}))\|x^*\| \geq 1, \;\; \forall x \in B(\bar{x}, \delta) \text{ with } f(x) - f(\bar{x}) \in (0, \delta), \; x^* \in \partial f(x).$$

This remarkable inequality is a nonsmooth version of the Kurdyka-Lojasiewicz inequality ([13]). Due to the original works by Lojasiewicz [17], [19], then Kurdyka [13], and the recent extentions to nonsmooth functions by Bolts *et al* [7], [9], the class of functions verifying the Kurdyka-Lojasiewicz inequality is very ample, for example, it consists of semi-algebraic, subanalytic, and log-exp (see, e.g., [13], [7], [**?**] and references therein).

Theorem 3.4 *Consider partial DC problem (PDCP). Let* $(x_0, y_0) \in \text{Dom } f$ *be given. Let the sequences* $\{w_k\} = \{(x_k, y_k)\}$ *and* $\{w_k^*\} = \{(x_k^*, y_k^*)\}$ *are generated by ADCA . Suppose that The sequence* $\{w_k\}$ *is bounded. Then the sequence* $\{f(w_k)\}$ *is decreasing and therefore it converges to some* $c \in \mathbb{R}$. *Suppose further that the following assumptions are satisfied:*

(i) $\text{Dom } f$ *is closed;* $f |_{\text{Dom } f}$ *is continuous, and the function* $|f - c|$ *verifies (KLI) at all limit points of* $\{w_k\}$.

(ii) The function $g, h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ *are lower semicontinuous partially convex; and either (A1) or (A2) is satisfied.*

If

$$\liminf_{k\to\infty} \min\{\rho(x_k), \rho(y_k)\} > \rho > 0, \tag{46}$$

then the sequence $\{w_k\}$ *converges to a (Clarke) critical point (PDCP). Moreover, under (A2), this critical point is a Fréchet critical point.*

Sketch of the proof. The decreasing property of the sequence $\{f(w_k)\}$ follows immediately from Lemma 3.3. Without loss of generality, assume $c = 0$, and also, according to that lemma, we can assume that

$$\frac{\rho}{2}\|w_k - w_{k+1}\|^2 \le f(w_k) - f(w_{k+1}), \quad k = 0, 1, \ldots.$$

Recall that $\mathcal{C}$ denotes the set of limit points of $\{f(w_k)\}$. Then $\mathcal{C}$ is compact. For the case 1 where (A1) is satisfied, by the compactness of $\mathcal{C}$, we can derives $\kappa, \delta > 0$ and a function $\varphi \in \mathcal{M}$ such that

$$\|\nabla g(w') - \nabla g(w'')\| \le \kappa\|w' - w''\| \quad \forall w', w'' \in B(w, \delta_w) \text{ with } d_{\mathcal{C}}(w') < \delta,\ \|w' - w''\| < \delta;$$

$$\varphi'(|f(w)|)\|w^*\| \ge 1 \quad \forall w', w'' \in \text{Dom } f \text{ with } d_{\mathcal{C}}(w) < \delta,\ f(w) \in (-\delta, \delta),\ w^* \in \partial|f|(w);$$

$$\partial_x h(u, v_1) \subseteq \partial_x h(u, v_2) + \kappa\|v_1 - v_2\|B_{\mathbb{R}^n} \quad \forall (u, v_1), (u, v_2) \in \text{Dom } f \text{ with } d_{\mathcal{C}}((u, v_1)) < \delta,\ \|v_1 - v_2\| < \delta;$$

$$\partial_y h(u_1, v) \subseteq \partial_y h(u_2, v) + \kappa\|u_1 - u_2\|B_{\mathbb{R}^m} \quad \forall (u_1, v), (u_2, v) \in \text{Dom } f \text{ with } d_{\mathcal{C}}((u_1, v)) < \delta,\ \|u_1 - u_2\| < \delta.$$

As in the proof of Theorem 3.3, by noting that $\partial|f|(w) = \partial(-f)(w)$ for $f(w) < 0$, when k is sufficiently large, there is some $w^* \in \partial(-f)(w_k)$ such that

$$\|w^*\| \le 3\kappa\|w_k - w_{k+1}\|,$$

following

$$\varphi'(f(w_k)) \ge \frac{1}{3\kappa\|w_k - w_{k+1}\|}.$$

Hence, by making use of the concavity of φ, one derives that

$$\|w_k - w_{k+1}\| \le \frac{2}{3\rho\kappa}[\varphi(f(w_k)) - \varphi(f(w_{k+1})],$$

and the convergence of $\{w_k\}$ is followed. Similarly to the case (A2), when k is sufficiently large, one has

$$\varphi'(f(w_k)) \ge \frac{1}{3\kappa\|w_k - w_{k-1}\|},$$

which implies that

$$\frac{\|w_k - w_{k+1}\|^2}{\|w_k - w_{k-1}\|} \le \frac{2}{3\rho\kappa}[\varphi(f(w_k)) - \varphi(f(w_{k+1}))].$$

This relation yields the convergence of the sequence $\{w_k\}$ as in the proof of the preceding theorems.□

Observe from the proof, In the case of (A2), instead of the assumption: $|f - c|$ satisfies (KLI), we need only (KLI) for f at limit points.

4 Generalized partial DC programming

Consider generalized partial DC programs of the following form:

$$(GPDCP) \qquad \min\{f(x,y): \ (x,y) \in \mathbb{R}^n \times \mathbb{R}^m\}.$$

Where, $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is a proper lower semicontinuous functions such that for each $y \in \mathbb{R}^n$, $f(\cdot, y)$ is a DC function on $\mathbb{R}^m$, and for each $y \in \mathbb{R}^n$, $f(\cdot, y)$ is a DC function on $\mathbb{R}^m$, and their decompositions are available:

$$f(x,y) := g_y(x) - h_y(x), \ \ x \in \mathbb{R}^n, \quad g_y, h_y \in \Gamma_0(\mathbb{R}^n); \tag{47}$$

$$f(x,y) := g_x(y) - h_x(y), \ \ x \in \mathbb{R}^n, \quad g_x, h_x \in \Gamma_0(\mathbb{R}^m). \tag{48}$$

We make use of the following hypothesis:

(H) For each $(\bar{x}, \bar{y}) \in \text{Dom } f$, $g_y(\bar{x})$, $h_y(\bar{x})$ as continuous functions in the variable y on $\bar{y}$, and $g_x(\bar{y})$, $h_x(\bar{y})$ as continuous functions in the variable x at $\bar{x}$. Let us introduce the notion of weak critical points for (GPDCP), which is an extension of the one for (PDCP).

Definition 4.1 *A point $(x,y) \in \mathbb{R}^n \times \mathbb{R}^m$ is said to be a weak critical point of (GPDCP) if*

$$\partial g_y(x) \cap \partial h_y(x) \neq \emptyset \quad \textit{and} \quad \partial g_x(y) \cap \partial h_x(y) \neq \emptyset.$$

Obviously, a Fréchet/limiting/Clarke critical point is a weak critical point. The following lemmata gives some sufficient conditions under which a weak critical point is a Fréchet/Clarke critical point.

Lemma 4.1 *Let f be a generalized partial DC function defined by (47), (48), and let $(\bar{x}, \bar{y}) \in \text{Dom } f$. Suppose the following two conditions are satisfied*

(i) The subdifferential $\partial g_y(\bar{x})$ regarded as a multifunction in y is lower semicontinuous at $\bar{y}$ in the sense: for any $\varepsilon > 0$, there is $\delta > 0$ such that

$$\partial g_{\bar{y}}(x) \subseteq \partial g_y(\bar{x}) + \varepsilon, \ \ \forall y \in B(\bar{y}, \delta).$$

(ii) There is $\delta > 0$ such that the function h_y is differentiable on $B(\bar{x}, \delta)$ for y near $\bar{y}$, and $\nabla h_y(x)$, regarded as a function in the two variable x, y is continuous at $(\bar{x}, \bar{y})$.

Then one has

$$\partial^F f(\bar{x}, \bar{y}) = \partial^F_x f(\bar{x}, \bar{y}) \times \partial^F_x f(\bar{x}, \bar{y}).$$

As a result, if $(\bar{x}, \bar{y})$ is a weak critical point of (GPDCP) and $h_{\bar{x}}$ is differentiable at $\bar{y}$., then it is a Fréchet critical point of (GPDCP).

Proof. It suffices to show that

$$\partial^F_x f(\bar{x}, \bar{y}) \times \partial^F_x f(\bar{x}, \bar{y}) \subseteq \partial^F f(\bar{x}, \bar{y}).$$

Let $(x^*, y^*) \in \partial^F_x f(\bar{x}, \bar{y}) \times \partial^F_x f(\bar{x}, \bar{y})$. By the asumptions, for $\varepsilon > 0$, there is $\delta > 0$ such that

$$\partial g_{\bar{y}}(x) \subseteq \partial g_y(\bar{x}) + \varepsilon, \ \ \forall y \in B(\bar{y}, \delta);$$

$$\|\nabla h_y(x) - \nabla h_{\bar{y}}(\bar{x})\| < \varepsilon, \ \forall (x,y) \in B((\bar{x},\bar{y}),\delta),$$

As

$$x^* \in \partial_x^F(\bar{x},\bar{y}) = \partial g_{\bar{y}}(\bar{x}) - \nabla h_{\bar{y}}(\bar{x}),$$

then, for $y \in B(\bar{y},\delta)$, one has

$$\begin{aligned} x^* + \nabla h_y(\bar{x}) \quad &= [x^* + \nabla h_{\bar{y}}(\bar{x})] + [\nabla h_y(\bar{x}) - \nabla h_{\bar{y}}(\bar{x})] \\ &\subseteq \partial g_{\bar{y}}(\bar{x}) + \varepsilon B_{R^n} \subseteq \partial g_y(\bar{x}) + 2\varepsilon B_{R^n}. \end{aligned}$$

Hence,

$$\langle x^*, x - \bar{x}\rangle \le g_y(x) - g_y(\bar{x}) - \langle \nabla h_y(\bar{x}), x - \bar{x}\rangle + 2\varepsilon\|x - \bar{x}\|, \ \forall x \in \mathbb{R}^n, \ y \in B(\bar{y},\delta). \tag{49}$$

Since $y^* \in \partial_y^F f(\bar{x},\bar{y})$ we can assume that for the same $\delta > 0$ as before, one has

$$\langle y^*, y - \bar{y}\rangle \le f(\bar{x},y) - f(\bar{x},\bar{y}) + \varepsilon\|y - \bar{y}\|, \ \forall y \in B(\bar{y},\delta). \tag{50}$$

Let $(x,y) \in B((\bar{x},\bar{y}),\delta)$. By the mean value theorem, applied to the function h_y, there exists $z \in [x,\bar{x}]$ such that

$$h_y(x) - h_y(\bar{x}) = \langle \nabla h_y(z), x - \bar{x}\rangle.$$

It implies that

$$|h_y(x) - h_y(\bar{x}) - \langle \nabla h_y(\bar{x}), x - \bar{x}\rangle| \le \varepsilon\|x - \bar{x}\|.$$

By combinning this relation and (49), one obtains

$$\begin{aligned} \langle x^*, x - \bar{x}\rangle \quad &\le g_y(x) - g_y(\bar{x}) - (h_y(x) - h_y(\bar{x})) + 3\varepsilon\|x - \bar{x}\| \\ &= f(x,y) - f(\bar{x},y) + 3\varepsilon\|x - \bar{x}\|. \end{aligned}$$

By adding this inequality and (50),

$$\langle x^*, x - \bar{x}\rangle + \langle y^*, y - \bar{y}\rangle \le f(x,y) - f(\bar{x},\bar{y}) + 3\varepsilon\|x - \bar{x}\| + \varepsilon\|y - \bar{y}\|.$$

Thus $(x^*, y^*) \in \partial^F f(\bar{x},\bar{y})$.

If $h_{\bar{x}}$ is diffentiable at $\bar{y}$ and $(\bar{x},\bar{y})$ is a weak critical point of (GPDCP) , then

$$(0,0) \in (\partial g_{\bar{y}}(\bar{x}) - \nabla h_{\bar{y}}(\bar{x})) \times (\partial g_{\bar{x}}(\bar{y}) - \nabla h_{\bar{x}}(\bar{y})) = \partial_x^F f(\bar{x},\bar{y}) \times \partial_y^F f(\bar{x},\bar{y}) = \partial^F f(\bar{x},\bar{y}).$$

That is, $(\bar{x},\bar{y})$ is a Fréchet critical point of f. □

Lemma 4.2 *Let f be a generalized partially DC function defined by (47), (48), and let $(\bar{x},\bar{y}) \in \operatorname{Dom} f$. Suppose that f is locally Lipschitz around $(\bar{x},\bar{y})$ and that the following two conditions are satisfied*

(i) $\partial h_y(\bar{x})$ regarded as a multifunction in y is lower semicontinuous at $\bar{y}$.

(ii) There is $\delta > 0$ such that the function g_y is differentiable on $B(\bar{x},\delta)$ for y near $\bar{y}$, and $\nabla g_y(x)$, regarded as a function in the two variable x, y is continuous at $(\bar{x},\bar{y})$.

If $(\bar{x},\bar{y})$ is a weak critical point of (GPDCP) and $g_{\bar{x}}$ is differentiable at $\bar{y}$., then this point is a Clarke critical point of (GPDCP).

Proof. We see that the assumptions (i), (ii) guarantees that the function $(-f)$ satisfies all assumptions of Lemma 4.1 where the roles of g_y, h_y and of g_x, h_x are exchanged, respectively. Hence,

$$(\partial h_{\bar{y}}(\bar{x}) - \nabla g_{\bar{y}}(\bar{x})) \times (\partial h_{\bar{x}}(\bar{y}) - \nabla g_{\bar{x}}(\bar{y})) = \partial_x(-f)(\bar{x},\bar{y}) \times \partial_y^F(-f)(\bar{x},\bar{y}) = \partial^F(-f)(\bar{x},\bar{y}).$$

Since

$$\partial^F(-f)(\bar{x},\bar{y}) \subseteq \partial^0(-f)(\bar{x},\bar{y}) = -\partial f(\bar{x},\bar{y}),$$

one obtains

$$(\partial g_{\bar{y}}(\bar{x}) - \nabla h_{\bar{y}}(\bar{x})) \times (\partial g_{\bar{x}}(\bar{y}) - \nabla h_{\bar{x}}(\bar{y})) \subseteq \partial^0 f(\bar{x},\bar{y}),$$

and the conclusion follows. □

We propose the generalized versions of ADCA and of IADCA for solving (GPDCP) as follows.

GADCA: Generalized Alternative DC Algorithm

1. Initialization: Given $(x_0, y_0) \in \text{Dom } f$.
2. For $k = 0, 1, \ldots$, generating the sequences $\{(x_k, y_k)\}$ and $\{(x_k^*, y_k^*)\}$ by
   - Compute $x_k^* \in \partial h_{y_k}(x_k)$.
   - Compute $x_{k+1} \in \partial g_{y_k}^*(x_k^*)$; i.e., x_{k+1} is a solution of the convex program
   $$\min\{g_{y_k}(x) - \langle x, x_k^*\rangle : \ x \in \mathbb{R}^n\}. \tag{51}$$
   - Compute $y_k^* \in \partial h_{x_{k+1}}(y_k)$.
   - Compute $y_{k+1} \in \partial_y g_{x_{k+1}}^*(y_k^*)$ by solving the convex program
   $$\min\{g_{x_{k+1}}(y) - \langle y_k^*, y\rangle : \ y \in \mathbb{R}^m\}. \tag{52}$$
3. If stopping criterion is met, then stop and we have (x_{k+1}, y_{k+1}) is the computed solution, otherwise, set $k = k + 1$ and go to Step 2.

Output (x_k, y_k) and $f(x_k, y_k)$ as the best known solution and objective function value.

GIADCA: Generalized Inexact Alternative DC Algorithm

1. Initialization: Given $(x_0, y_0) \in \text{Dom } f$ and $\varepsilon_0 \geq 0$.
2. For $k = 0, 1, \ldots$, update $\varepsilon_k \geq 0$ and generating the sequences $\{(x_k, y_k)\}$ and $\{(x_k^*, y_k^*)\}$ by
   - Compute $x_k^* \in \partial^{\varepsilon_k} h_{x_k}(y_k)$.
   - Compute $x_{k+1} \in \partial^{\varepsilon_k} g_{y^k}^*(x_k^*)$; i.e., x^{k+1} is an ε_k-solution of the convex program
   $$\min\{g_{y_k}(x) - \langle x, x_k^*\rangle : \ x \in \mathbb{R}^n\}. \tag{53}$$
   - Compute $y_k^* \in \partial^{\varepsilon_k} h_{x^{k+1}}(y_k)$.
   - Compute $y_{k+1} \in \partial^{\varepsilon_k} g_{x_{k+1}}^*(y_k^*)$ by searching an ε_k-solution the convex program
   $$\min\{g(x_{k+1}, y) - \langle y_k^*, y\rangle : \ y \in \mathbb{R}^m\}. \tag{54}$$

3. If stopping criterion is met, then stop and we have (x_{k+1}, y_{k+1}) is the computed solution, otherwise, set $k = k + 1$ and go to Step 2.

For $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$, denote by

$$\rho_x = \rho(g_x) + \rho(h_x), \quad \rho_y = \rho(g_y) + \rho(h_y).$$

Note that if $g_y(x)$ (or g_x, h_x, h_y) is continuous in variable y, for each $x \in \mathbb{R}^n$, then for any sequences $\{(x_k, y_k)\}$ converging to (x_∞, y_∞), and $\{x_k^*\} \subseteq \mathbb{R}^n$ with $x_k^* \in \partial g_{y_k}(x_k)$; $x_k^* \to x^*$ as $k \to \infty$, one has $x^* \in \partial g_{y_\infty}(x_\infty)$.

By this remark, under hypothesis (H), the convergence result of Theorem 34 stated for (GIADCA) also holds.

Theorem 4.1 *Consider problem (GPDCP). Assume that hypothesis (H) is satisfied. Let* $\{w_k\} := \{(x_k, y_k)\}$ *be a sequence generated by (GIADCA), with respect to a updated sequence* $\{\varepsilon_k\}$ *with* $\varepsilon_k \geq 0$ *for all* $k \geq 0$ *such that*

$$\sum_{k=0}^{\infty} \varepsilon_k < +\infty \quad \textit{and} \quad \liminf_{k\to\infty} \min\{\rho(y_k), \rho(x_k)\} > 0.$$

Then for any subsequence $\{w_{k_l}\} := \{(x_{k_l}, y_{k_l})\}$ *of* $\{w_k\}$*, converging to* $\bar{w} = (\bar{x}, \bar{y})$ *such that* $\{(x_{k_l}^*, x_{k_l}^*)\}$ *is bounded, the point* $\bar{w}$ *is a weak critical point of problem (GPDCP).*

Still, instead of assumptions (A1), (A2), we make use of the following general versions (GA1) and (GA2), under which the convergences of (GIACDA) and (GADCA) when the objective function f is assumed to be subnanalytic (or more general, to verify (KLI)) remain valid.

(GA1) The functions g_x, g_y are differentiable such that $\nabla g_y(x)$, $\nabla g_x(y)$ are locally Lipschitz in the joint variable (x, y), and for each $\bar{x} \in \mathbb{R}^n$, $\bar{y} \in \mathbb{R}^m$, the multifunction $\partial h_y(\bar{x})$, $\partial h_x(\bar{y})$ are locally Lipschitz in variables y, x, respectively.

(GA2) The functions h_x, h_y are differentiable such that $\nabla h_y(x)$, $\nabla h_x(y)$ are locally Lipschitz in the joint variable (x, y), and for each $\bar{x} \in \mathbb{R}^n$, $\bar{y} \in \mathbb{R}^m$, the multifunction $\partial g_y(\bar{x})$, $\partial g_x(\bar{y})$ are locally Lipschitz in variables y, x, respectively.

Remark 4.1 In Theorems 3.2, 3.3, and 3.4, we make use of (H) and (GA1), (GA2), instead of (A1), (A2), respectively, then these theorems remain valid for (GPDCP).

5 Applications

5.1 Nonconvex Feasibility Problems

We consider the problem of finding a point in the intersection of sets which is not necessarily convex. This common problem plays an important role in many diverse areas of Mathamatics and Physis. Let $C_1, C_2 \subseteq \mathbb{R}^n$ be nonempty closed subsets of $\mathbb{R}^n$. The feasible problem for two sets is stated by

$$(FP) \quad \text{Finding a point } x \in C_1 \cap C_2.$$

For convex feasibility problems, i.e., the sets under consideration are convex, there are many theoretical results as well as algorithms aspects in the literature. For details, the reader is referred to the exellent review paper [6] by Bauschke-Borwein, and the references given therein. The best celebrated algorithm for solving this convex feasibility problem is the *alternating projection algorithm* due to Von Neumann. Recently, some extended variants of this method of alternating projection to nonconvex problems have been investigated in [3], [4].

We will show that the (nonconvex) feasibility problem (FP) can be reformulated equivalently as a partial DC program (PDCP), and therefore the algorithm ADCA can be worked. For given $\alpha_1, \alpha_2 > 0$, regarded as penalized parameters, problem (FP) is transformed to the following optimization problem:

$$\min\{f(x,y) := \alpha_1 d^2_{C_1}(x) + \alpha_2 d^2_{C_2}(y) + \|x-y\|^2 : \quad (x,y) \in \mathbb{R}^n \times \mathbb{R}^n\}. \tag{55}$$

Note that

$$\begin{aligned} d^2_{C_1}(x) &= \min\nolimits_{u\in C_1} \|x-u\|^2 = \|x\|^2 + \min\nolimits_{u\in C_1}[-2\langle x,u\rangle + \|u\|^2 \\ &= \|x\|^2 - \max\nolimits_{u\in C_1}[2\langle x,u\rangle - \|u\|^2], \end{aligned}$$

is a DC function, and so is the function

$$d^2_{C_2}(y) = \|y\|^2 - \max_{v\in C_2}[2\langle y,v\rangle - \|v\|^2].$$

Hence, $f(x,y)$ is a DC function (therefore, a partial DC function) by the decomposition:

$$f(x,y) = g(x,y) - h(x,y), \quad (x,y) \in \mathbb{R}^n,$$

$$g(x,y) := \alpha_1\|x\|^2 + \alpha_2\|y\|^2 + \|x-y\|^2;$$

$$h(x,y) := \alpha_1 \max_{u\in C_1}[2\langle x,u\rangle - \|u\|^2] + \alpha_2 \max_{v\in C_2}[2\langle y,v\rangle - \|v\|^2].$$

Denote by P_{C_1}, P_{C_2}, the (multi-valued) projection operators on C_1, C_2, rspectively. That is,

$$P_{C_1}(x) = \text{argmin}\{\|x-u\|^2 : \ u \in C_1\}; \ P_{C_2}(y) = \text{argmin}\{\|y-v\|^2 : \ v \in C_2\}, \ (x,y) \in \mathbb{R}^n \times \mathbb{R}^n.$$

One has, from the subdifferential formula for supremum functions in Convex Analysis,

$$\partial_x h(x,y) = 2\alpha_1 \text{co} P_{C_1}(x), \ \ \partial_y h(x,y) = 2\alpha_2 \text{co} P_{C_2}(y), \ (x,y) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Where, "co" stands for the convex hull (of a set). By checking directly, It is easy to see that for $y \in \mathbb{R}^n$, $x^* \in \mathbb{R}^n$,

$$\partial_x g^*(\cdot,y)(x^*) = \left\{ \frac{2y + x^*}{2(\alpha_1+1)} \right\},$$

and similarly, for $x \in \mathbb{R}^n$, $y^* \in \mathbb{R}^n$,

$$\partial_y g^*(x,\cdot)(y^*) = \left\{ \frac{2x + y^*}{2(\alpha_2+1)} \right\}.$$

Algorithm ADCA applied to problem (55) is the following.

ACDA-FP

1. Initialization: Select a starting point $(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^n$.

2. For $k = 0, 1, ...$, a sequence $\{(x_k, y_k)\}$ is generated by

   2.1. Compute
$$u_k \in P_{C_1}(x_k), \quad x_{k+1} = \frac{y_k + \alpha_1 u_k}{\alpha_1 + 1}.$$

   2.2. Compute
$$v_k \in P_{C_2}(y_k), \quad y_{k+1} = \frac{x_{k+1} + \alpha_2 v_k}{\alpha_2 + 1}.$$

Remark 5.1 The algorithm above invokes the projections on (noncovex) sets which is generally difficult in practical computations. If there are residual functions $r_1(x)$, $r_2(y)$ with respect to C_1, C_2, respectively, which are asummed to be DC functions such that

$$d^2_{C_1}(x) \le r_1(x), \quad d^2_{C_2}(y) \le r_2(y), \text{ for all } x, y \in \mathbb{R}^n,$$

$$x \in C_1 \text{ iff } r_1(x) = 0; \quad y \in C_2 \text{ iff } r_2(y) = 0,$$

then problem (FP) can be transformed to the following

$$\min\{f(x,y) := \alpha_1 r_1(x) + \alpha_2 r_2(y) + \|x - y\|^2 : \quad (x,y) \in \mathbb{R}^n \times \mathbb{R}^n\}. \tag{56}$$

This problem is a DC program (therefore, a partial DC program). By making use of suitable DC decompositions of r_1, r_2, the partial DC algorithm can be applied.

5.2 Matrix Factorization Problems

We consider in this subsection an application of the partial DC programming to matrix factorization problems. This class of matrix factorization problems plays a crucial role in data mining as well as in diverse applications. Let m, n, p be given positive integers with $p \le \min\{m, n\}$, and let $A \in \mathbb{R}^{m\times n}$ be a given matrix. Denote by

$$\mathbb{R}^{m\times p}_+ := \{X \in \mathbb{R}^{m\times p} : \quad X \ge 0\}; \quad \mathbb{R}^{p\times n}_+ := \{Y \in \mathbb{R}^{p\times n} : \quad Y \ge 0\}.$$

Let $C_1 \subseteq \mathbb{R}^{m\times p}_+$ and $C_2 \subseteq \mathbb{R}^{p\times n}_+$ be given two nonempty closed sets. By the optimization approach, a nonnegative matrix factorization problem is formulated as a optimization problem of the form

$$(MF) \quad \min\{d(A, XY) : \quad X \in C_1, \ Y \in C_2\}.$$

Where, $d : \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n} \to \mathbb{R}_+$ define a *gap* of the approximation A by XY, verifying $d(U, V) = 0$ *iff* $U = V$.

Let $R_1 : \mathbb{R}^{m\times p} \to \mathbb{R}_+$ and $R_2 : \mathbb{R}^{p\times n} \to \mathbb{R}_+$ be *residual* functions of C_1, C_2, respectively. That is,

$$X \in \mathbb{R}^{m\times p}, \ R_1(X) = 0 \quad \text{iff} \quad X \in C_1;$$

$$Y \in \mathbb{R}^{p\times n}, \ R_2(Y) = 0 \quad \text{iff} \quad Y \in C_2.$$

By using the penalization approach, we consider the following penalized problem of (MF):

$$(PMF) \quad \min\{f(X,Y) := \alpha_1 R_1(X) + \alpha_2 R_2(Y) + d(A, XY) : \quad X \in \mathbb{R}^{m\times p}, \ Y \in \mathbb{R}^{p\times n}\}.$$

Where, $\alpha_1, \alpha_2 > 0$ are penalty parameters. Under suitable conditions on functions R_1, R_2 and d, it is possible to formulate (PMF) as partial DC programs, or generalized partial DC programs. For instance, when R_1, R_2 and $d(A, \cdot)$ are DC functions, then $f(X,Y)$ is a partial DC function. Indeed, consider the DC decompositions of $R_1, R_2,\ d(A,\cdot)$:

$$R_1(X) = g_1(X) - h_1(X),\ X \in \mathbb{R}^{m\times p},\ g_1,\ h_1 \text{ are convex on } \mathbb{R}^{m\times p};$$

$$R_2(Y) = g_2(Y) - h_2(Y),\ Y \in \mathbb{R}^{p\times n},\ g_2,\ h_2 \text{ are convex on } \mathbb{R}^{p\times n}$$

and

$$d(A,U) = d_1(U) - d_2(U),\ U \in \mathbb{R}^{m\times n},\ d_1,\ d_2 \text{ are convex on } \mathbb{R}^{m\times n}.$$

By setting

$$g(X,Y) := g_1(X) + g_2(Y) + d_1(XY), \quad X \in \mathbb{R}^{m\times p},\ Y \in \mathbb{R}^{p\times n};$$

$$h(X,Y) := g_1(X) + g_2(Y) + d_2(XY), \quad X \in \mathbb{R}^{m\times p},\ Y \in \mathbb{R}^{p\times n},$$

then obviously, g, h is partial convex functions, and therefore

$$f(X,Y) = g(X,Y) - h(X,Y), \quad X \in \mathbb{R}^{m\times p},\ Y \in \mathbb{R}^{p\times n}$$

is a partial DC function.

We illustrate this by considering the case where the gap function d is defined by

$$d(U,V) := \|U - V\|_F^2, \quad U, V \in \mathbb{R}^{m\times n},$$

where, the Frobenius norm $\|\cdot\|_F$ is defined for any matrix M, by

$$\|M\|_F := \sqrt{\mathrm{Trace}(M^T M)}.$$

Taking the residual fuctions: $R_1(X) = d^2_{C_1}(X)$, $R_2(Y) = d^2_{C_2}(Y)$, consider the following problem:

$$\min\{f(X,Y) := \alpha_1 d^2_{C_1}(X) + \alpha_2 d^2_{C_2}(Y) + \|A - XY\|_F^2 : \ \ X \in \mathbb{R}^{m\times p},\ Y \in \mathbb{R}^{p\times n}\}. \tag{57}$$

By setting

$$g(X,Y) := \alpha_1 \|X\|_F^2 + \alpha_2 \|Y\|_F^2 + \|A - XY\|_F^2;$$

$$h(X,Y) := \alpha_1 \max_{U\in C_1} [2\langle X, U\rangle - \|U\|_F^2] + \alpha_2 \max_{V\in C_2} [2\langle Y, V\rangle - \|V\|_F^2],$$

problem (57) is rewritten as the partial DC program:

$$\min\{f(X,Y) := g(X,Y) - h(X,Y) : \ \ X \in \mathbb{R}^{m\times p},\ Y \in \mathbb{R}^{p\times n}\}. \tag{58}$$

For $X \in \mathbb{R}^{m\times p}$, $Y \in \mathbb{R}^{p\times n}$, one has

$$\partial_X h(X,Y) = 2\alpha_1 \mathrm{co} P_{C_1}(X), \ \ \partial_Y h(X,Y) = 2\alpha_2 \mathrm{co} P_{C_2}(Y). \tag{59}$$

For $U \in \mathbb{R}^{m\times p}$, $Y \in \mathbb{R}^{p\times n}$, the solution of the convex optimization problem

$$\min\{G(X,Y) - \langle X, U\rangle : \ \ X \in \mathbb{R}^{m\times p}\}$$

is determined by by $\nabla_X g(X,Y) = U$, equivalently,

$$2\alpha_1 X + 2(XY - A)Y^T = U,$$

which gives

$$X = (2AY^T + U)(2\alpha_1 I_p + 2YY^T)^{-1}, \tag{60}$$

here, I_p stands for the identity matrix of p order. Similarly, for $V \in \mathbb{R}^{p\times n}$, $X \in \mathbb{R}^{m\times p}$, the solution of the convex program

$$\min\{G(X,Y) - \langle Y, V\rangle : \quad Y \in \mathbb{R}^{p\times n}\}$$

is

$$Y = (2\alpha_2 I_p + 2X^T X)^{-1}(2X^T A + V). \tag{61}$$

By making use of (59), (60) and (61), one derives the following algorithm ADCA for solving (58):

ACDA1-FM

1. Initialization: Select a starting point $(X_0, Y_0) \in \mathbb{R}^{m\times p} \times \mathbb{R}^{p\times n}$.
2. For $k = 0, 1, ...$, a sequence $\{(X_k, Y_k)\}$ is generated by

   2.1. Compute
   $$U_k \in P_{C_1}(X_k), \quad X_{k+1} = (AY_k^T + \alpha_1 U_k)(\alpha_1 I_p + Y_k Y_k^T)^{-1}.$$

   2.2. Compute
   $$V_k \in P_{C_2}(Y_k), \quad Y_{k+1} = (\alpha_2 I_p + X_{k+1}^T X_{k+1})^{-1}(X_{k+1}^T A + V_k).$$

We consider an other partial DC decomposition for problem (57). By setting $Q(X,Y) = \|A - XY\|_F^2$, the partial derivatives of Q :

$$X \mapsto \nabla_X Q(X,Y) = 2(XY - A)Y^T, \quad Y \mapsto \nabla_Y Q(X,Y) = 2X^T(XY - A),$$

are Lipschitz with modulus $L_Y = 2\|YY^T\|_F$, $L_X = 2\|X^T X\|_F$, respectively. Therefore, the functions

$$X \mapsto L_Y\|X\|_F^2 - Q(X,Y) \quad \text{and} \quad Y \mapsto L_X\|Y\|_F^2 - Q(X,Y)$$

are convex functions, for each $Y \in \mathbb{R}^{p\times n}$, each $X \in \mathbb{R}^{m\times p}$, respectively. By this, for each $Y \in \mathbb{R}^{p\times n}$, the function $f(\cdot, Y)$ admets the following DC decomposition:

$$f(X,Y) = G_Y(X) - H_Y(X), \quad X \in \mathbb{R}^{m\times p}, \tag{62}$$

here,

$$G_Y(X) = (\alpha_1 + L_Y)\|X\|_F^2 + \alpha_2 d_{C_2}^2(Y);$$
$$H_Y(X) = \alpha_1 \max_{U\in C_1}[2\langle X, U\rangle - \|U\|_F^2] + L_Y\|X\|_F^2 - Q(X,Y)$$

are convex functions on $\mathbb{R}^{m\times p}$. And for each $X \in \mathbb{R}^{m\times p}$, the function $f(X, \cdot)$ has the DC decomposition:

$$f(X,Y) = G_X(Y) - H_X(Y), \quad X \in \mathbb{R}^{m\times p}, \tag{63}$$

here,

$$G_X(Y) = (\alpha_2 + L_X)\|Y\|_F^2 + \alpha_1 d_{C_1}^2(X);$$
$$H_X(Y) = \alpha_2 \max_{V \in C_2}[2\langle Y, V\rangle - \|V\|_F^2] + L_X\|Y\|_F^2 - Q(X,Y)$$

are convex functions on $\mathbb{R}^{p\times n}$.

For $X \in \mathbb{R}^{m\times p}$, $Y \in \mathbb{R}^{p\times n}$, one has

$$\partial H_Y(X) = 2\alpha_1 \text{co} P_{C_1}(X) + 2L_Y X - 2(XY - A)Y^T;$$

$$\partial H_X(Y) = 2\alpha_2 \text{co} P_{C_2}(Y) + 2L_X Y - X^T(XY - A).$$

For $U \in \mathbb{R}^{m\times p}$, $Y \in \mathbb{R}^{p\times n}$, the solution of the convex optimization problem

$$\min\{G_Y(X) - \langle X, U\rangle : \quad X \in \mathbb{R}^{m\times p}\}$$

is given by

$$X = \frac{U}{2(\alpha_1 + L_Y)}.$$

And for $V \in \mathbb{R}^{p\times n}$, $X \in \mathbb{R}^{m\times p}$, the solution of the convex program

$$\min\{G(X,Y) - \langle Y, V\rangle : \quad Y \in \mathbb{R}^{p\times n}\}$$

is

$$Y = \frac{V}{2(\alpha_2 + L_X)}.$$

Hence, the algorithm GADCA for solving (58) is given as follows.

ACDA2-FM

1. Initialization: Select a starting point $(X_0, Y_0) \in \mathbb{R}^{m\times p} \times \mathbb{R}^{p\times n}$.
2. For $k = 0, 1, ...$, a sequence $\{(X_k, Y_k)\}$ is generated by

   2.1. Compute
   $$U_k \in P_{C_1}(X_k), \quad X_{k+1} = \frac{\alpha_1 U_k + L_{Y_k} X_k - (X_k Y_k - A)Y_k^T}{\alpha_1 + L_{Y_k}}.$$

   2.2. Compute
   $$V_k \in P_{C_2}(Y_k), \quad Y_{k+1} = \frac{\alpha_2 V_k + L_{X_{k+1}} Y_k - X_{k+1}^T(X_{k+1}Y_k - A)}{\alpha_2 + L_{X_{k+1}}}.$$

References

[1] P.-A. Absil, R. Mahony, B. Andrews, Convergence of the iterates of descent methods for analytic cost functions, *SIAM J. Optim.,* **16**, 531-547 (2005).

[2] H. Attouch, J. Bolte, The convergence of the proximal algorithm for nonsmooth functions involving analytic features, *Math. Program.* **116**, 5-16 (2009).

[3] H. Attouch, J. Bolte, P. Redont and A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kyrdyka-Lojasiewicz inequality, *Mathematics of Operations research,* **35**, 438-457 (2010).

[4] H. Attouch, J. Bolte and B. F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward spliting, and regularized Gauss-Seidel methods, *Mathematical Programming,* Ser. A **137**, 91-129 (2013).

[5] J. Bolte, S. Sabach and M. Teboule, Proximal alternating linearized minimization for nonconvex and nonsmooth problems (*Preprint*).

[6] H. H. Bauschke, J. M. Borwein, On projection algorithms for solving convex feasibility problems, *SIAM review*, **38** (3), 367-426 (1996).

[7] J. Bolte, A. Daniliidis, A. Lewis, Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamic systems, *SIAM Optim.*, **17** (4), 1205-1223 (2007).

[8] J. Bolte, A. Daniliidis, O. Ley, L. Mazet, Characterizations of Lojasiewicz inequalities: Subgradient flows, talweg, convexity, *Trans. Amer. Math. Soc.*, **362**, 3319-3363 (2010).

[9] J. Bolte, A. Daniilidis, A. Lewis and M. Shiota, Clarke subgradients of stratifiable functions, *SIAM Optim.*, **18**, 556-572 (2007).

[10] E. Bierstone, P. Milman, Semianalytic and subanalytic sets, *IHES Publ. Math.*, **67**, 5-42 (1988).

[11] Ekeland I., On the variational principle, *J. Math. Anal. and Appl.,* **47**, 324–353, (1974).

[12] J.B. Hiriart-Urruty, C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Parts I&II, Springer Verlag (1991).

[13] K. Kurdyka, On gradients of functions definable in o-minimal structures, *Annales de l'Institut Fourier,* **48**, 769-783 (1998).

[14] Le Thi Hoai An, Pham Dinh Tao, Large scale global molecular optimization from distance matrices by a DC optimization appoach, *SIAM J. Optim,* **14**(1), 77-116, (2003).

[15] Le Thi Hoai An and Pham Dinh Tao, The DC (difference of convex functions) Programming and DCA revisited with DC models of real world nonconvex optimization problems. *Annals of Operations Research*, **133**, pp. 23-48, (2005).

[16] Le Thi Hoai An, Huynh Van Ngai and Pham Dinh Tao, Convergence Analysis of DC Algorithm for DC programming with subanalytic data, (*preprint*) (2010), available at http://www.optimization-online.org/DBFILE/2013/08/3996.pdf

[17] S. Lojasiewicz, Sur le problème de la division, *Studia Mathematica,* **18**, 87-136(1959).

[18] S. Lojasiewicz, Une propriété topologique des sous-ensembles analytiques réels, *in les Equations aux dérivées Partielles, Editions du Centre National de la Recherche Scientifique, Paris,* 87-89 (1963).

[19] S. Lojasiewicz, Sur la géométrie semi- et sous-analytique, *Anal de l'institute Fourier,* **43**, 1575-1595 (1993).

[20] B. S. Mordukhovich, *Variational analysis and generalized differentiation. I. Basic theory.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 330. Springer-Verlag, Berlin, 2006.

[21] H. V. Ngai, M. Théra, Error bounds for systems of lower semicontinuous functions in Asplund spaces, *Math. Program. ,* **116**, 397-427 (2009).

[22] Pham Dinh Tao, Le Thi Hoai An, Convex analysis approach to D.C. Programming: Theory, algorithms and applications, *Acta Mathematica Vietnamica,* **22**, 289-355(1997).

[23] Pham Dinh Tao, Le Thi Hoai An, A DC Optimization algorithm for solving the trust region subproblem, *SIAM J. Optim.,* **8**, 2, 476-505(1998).

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

[25] R.T. Rockafellar, R. J-B. Wets, *Variational Analysis*, Springer, New York (1998).

[26] M. Shiota, *Geometry of subanalytic and semialgebraic sets,* Progress in Math., 150. Birkhauser Boston, Inc., Boston, MA (1997).