J. Appl. Numer. Optim. 5 (2023), No. 2, pp. 255-269
Available online at http://jano.biemdas.com
https://doi.org/10.23952/jano.5.2023.2.05

THE BOOSTED PROXIMAL DIFFERENCE-OF-CONVEX ALGORITHM

ZIYUE HU[1], QIAO-LI DONG[1], XIAOLONG QIN[2,*]

[1] *Tianjin Key Laboratory for Advanced Signal Processing and College of Science, Civil Aviation University of China, Tianjin 300300, China*
[2] *Center for Advanced Information Technology, Kyung Hee University, Seoul, 02447, South Korea*

Abstract. In this paper, we propose a boosted proximal difference-of-convex algorithm for solving a minimization problem composed of the sum of a smooth convex function and a continuously differentiable convex function minus a continuous and strongly convex function. By adding an additional line search step, the convergence of proximal difference-of-convex algorithm is accelerated. We prove that any limit point of iterative sequence is a critical point of generalized difference-of-convex programming, and the corresponding objective value decreases monotonically and converges. By assuming that the objective function satisfies the strong Kurdyka–Łojasiewicz inequality, we prove the convergence of whole sequence of the proposed algorithm and give the convergence rate. The strong convexity of the convex part of the minimization problem in [F.J. Aragón Artacho, P.T. Vuong, The boosted difference of convex functions algorithm for nonsmooth functions, SIAM J. Optim. 30 (2020), 980-1006] is substituted by the Lipschitz continuity of the gradient of one convex function. Numerical experiments are given to demonstrate the performance of the proposed algorithm compared with the proximal difference-of-convex algorithm.

Keywords. Boosted proximal difference-of-convex algorithm; Difference-of-convex programming; Proximal difference-of-convex algorithm; Strong Kurdyka-Łojasiewicz inequality.

2020 Mathematics Subject Classification. 49J52, 49J53.

1. INTRODUCTION

In this paper, we are concerned with a class of generalized difference-of-convex (DC) programming originally introduced by Tao et al. [1] in 1986

$$\min_{x\in\mathbb{R}^n} F(x) := f(x)+g(x)-h(x), \tag{1.1}$$

where f and g are proper, lower semi-continuous, and convex functions, and h is a convex function. Problem (1.1) has been successfully applied to many problems in science and engineering fields in the last decades, such as compressed sensing [2], machine learning [3, 4, 5, 6], image processing [7], and dimensionality reduction [8]. For the image processing and the compressed sensing, f is a loss function representing data fidelity, while $g-h$ is a regularizer for inducing desirable structures in the solution [9, 10].

The popular methods to solve problem (1.1) are the difference-of-convex algorithm (DCA) and its variants [11, 12, 13]. The core idea of these algorithms is to linearize the concave part of

*Corresponding author.
E-mail address: qxlxajh@163.com (X. Qin).
Received January 22, 2023; Accepted March 14, 2023.

the objective function. So we only need to solve one convex optimization subproblem at each iteration. The iterative format of the DCA for solving problem (1.1) is as follows

$$x^{k+1} \in \arg\min_{x\in\mathbb{R}^n}\{f(x)+g(x)-\langle \xi^k, x\rangle\}, \tag{1.2}$$

where $\xi^k \in \partial h(x^k)$. To solve the subproblem in (1.2) by fully using its separable structure, Gotoh et al. [14] introduced the proximal DC algorithm (pDCA) whose scheme is given below

$$x^{k+1} = \arg\min_{x\in\mathbb{R}^n}\{g(x)+\langle \nabla f(x^k)-\xi^k, x\rangle + \frac{L_{\nabla f}}{2}\|x-x^k\|^2\}, \tag{1.3}$$

where $\xi^k \in \partial h(x^k)$ and $L_{\nabla f}$ is the Lipschitz continuity constant of ∇f. When the nonconvex part of the objective function in (1.1) is zero, pDCA reduces to the classical proximal gradient algorithm for convex programming [15]. Because of this, pDCA may take a lot of iterations in practice as special cases.

Since their inception, how to speed up DC algorithm and the pDCA is a hot issue. In 2018, wen et al. [16] combined the pDCA with inertial extrapolation technology and introduced a proximal difference-of-convex algorithm with extrapolation. The convergence speed of the pDCA is greatly accelerated. Recently, by introducing a new auxiliary function, the authors in [17] relaxed the conditions on h needed for the convergence of the whole iterative sequence.

If $g = 0$, then problem (1.1) reduces to

$$\min_{x\in\mathbb{R}^n} F(x) := f(x) - h(x). \tag{1.4}$$

Under the assumption that f and h are differentiable and strongly convex in (1.4), Aragón Artacho et al. [18] proposed a boosted DC algorithm (BDCA), which can be used to accelerate the convergence of DC algorithm. Recently, Aragón Artacho et al. [19] proposed an improved version of BDCA for the case that h is not differentiable in (1.4). Numerical experiments in [18, 19] demonstrated that the BDCA outperforms the DC algorithm. This advantage has been also confirmed when the BDCA was used to the indefinite kernel support vector machine problem [20].

Inspired by these work, in this paper, we propose a boosted proximal difference-of-convex algorithm (BpDCA) to solve the generalized DC programming (1.1). We prove that any limit point of our iterative sequence is a critical point of the generalized DC programming, and the corresponding objective value decreases monotonically and converges. Under the assumption that F satisfies the strong Kurdyka–Łojasiewicz inequality and ∇g is locally Lipschitz, we prove the convergence of the sequence generated by the BpDCA and give the convergence rate. The strongly convexity of the convex part of the minimization problem in [19] is substituted by the Lipschitz continuity of the gradient of one convex function.

The organization of the paper is as follows. In Section 2, we recall some definitions and known results for further analysis. Section 3 introduces a boosted proximal difference-of-convex algorithm to solve problem (1.1). In this section, the global subsequential convergence, the convergence of the whole sequence, and the convergence rate are established. In Section 4, we present two numerical experiments to demonstrate that the proposed algorithm is superior to the pDCA.

2. Preliminaries

Throughout this paper, the sets of positive integers and real numbers are denoted by $\mathbb{N}$ and $\mathbb{R}$. $\mathbb{R}^n$ is borrowed to denote the n-dimensional Euclidean space with inner product $\langle\cdot,\cdot\rangle$ and Euclidean norm $\|\cdot\|$. The closed ball with center x and radius $r>0$ is presented by $\mathbb{B}(x,r)$.

Now we recall some definitions of the function. The *domain* of the function $f:\mathbb{R}^n\to(-\infty,+\infty]$ is defined by $\operatorname{dom} f := \{x\in\mathbb{R}^n : f(x)<+\infty\}$. We say that f is *proper* if $\operatorname{dom} f\neq\emptyset$. Function f is said to be *coercive* if $f(x)\to+\infty$ whenever $\|x\|\to+\infty$. The *one-sided directional derivative* of f at $x\in\operatorname{dom} f$ for the direction $d\in\mathbb{R}^n$ is defined by

$$f'(x;d) := \lim_{t\downarrow 0}\frac{f(x+td)-f(x)}{t}.$$

Definition 2.1. (i) A function f is called convex if $\operatorname{dom} f$ is a convex set and if, for all $x,y\in\operatorname{dom} f$, $\alpha\in[0,1]$, it holds

$$f(\alpha x+(1-\alpha)y)\le \alpha f(x)+(1-\alpha)f(y).$$

(ii) A function f is said to be θ-strongly convex with $\theta>0$ if $f-\frac{\theta}{2}\|\cdot\|^2$ is convex, i.e.,

$$f(\alpha x+(1-\alpha)y)\le \alpha f(x)+(1-\alpha)f(y)-\frac{\theta}{2}\alpha(1-\alpha)\|x-y\|^2,$$

for all $x,y\in\operatorname{dom} f$ and $\alpha\in[0,1]$.

If f is convex, then

$$f(x)\ge f(y)+\langle\xi,x-y\rangle,\quad \forall x,y\in\operatorname{dom} f, \tag{2.1}$$

where $\xi\in\partial f(y)$ is arbitrary. If function f is convex and differentiable, then

$$f(x)\ge f(y)+\langle\nabla f(y),x-y\rangle,\quad \forall x,y\in\operatorname{dom} f.$$

Moreover, f is θ-strongly convex with $\theta>0$, then

$$f(x)\ge f(y)+\langle\xi,x-y\rangle+\frac{\theta}{2}\|x-y\|^2,\quad \forall x,y\in\operatorname{dom} f,$$

where $\xi\in\partial f(y)$ is arbitrary. If function f is θ-strongly convex and differentiable, then

$$f(x)\ge f(y)+\langle\nabla f(y),x-y\rangle+\frac{\theta}{2}\|x-y\|^2,\quad \forall x,y\in\operatorname{dom} f. \tag{2.2}$$

Definition 2.2. [21] Let $A:\mathbb{R}^n\to 2^{\mathbb{R}^n}$ be a set-valued mapping. One says that

(i) A is monotone if $\langle u-v,x-y\rangle\ge 0$ for any $u\in A(x)$ and $v\in A(y)$.

(ii) A is maximal monotone if there exists no monotone operator $B:\mathbb{R}^n\to 2^{\mathbb{R}^n}$ such that $\operatorname{gra}B$ properly contains $\operatorname{gra}A$, i.e., for every $(x,u)\in\mathbb{R}^n\times\mathbb{R}^n$

$$(x,u)\in\operatorname{gra}A \iff \langle u-v,x-y\rangle\ge 0\ \ \forall (y,v)\in\operatorname{gra}A.$$

(iii) A is ρ-strongly monotone ($\rho>0$) if $\langle u-v,x-y\rangle\ge\rho\|x-y\|^2$ for all $x,y\in\mathbb{R}^n$, $u\in A(x)$ and $v\in A(y)$.

Lemma 2.1. *A function $f:\mathbb{R}^n\to\mathbb{R}$ is strongly convex with modulus ρ if and only if ∂f is strongly monotone with modulus ρ.*

Lemma 2.2. *(Descent lemma [21]) Let $f:\mathbb{R}^n\to\mathbb{R}$ be continuous, convex, and differentiable. Suppose that ∇f is L-Lipschitz continuous. Then it holds that*

$$f(y)\le f(x)+\langle y-x,\nabla f(x)\rangle+\frac{L}{2}\|y-x\|^2,\quad \forall x,y\in\mathbb{R}^n.$$

When dealing with nonconvex and nonsmooth functions, we have to consider Clarke subdifferential, which can be defined in several equivalent ways [22]. For a given locally Lipschitz continuous function $f : \mathbb{R}^n \to (-\infty,+\infty]$, the Clarke subdifferential of f at $\bar{x}$ is given by

$$\partial_C f(\bar{x}) = \text{co}\left\{ \lim_{x\to\bar{x}, x\notin\Omega_f} \nabla f(x) \right\},$$

where co stands for the convex hull and Ω_f denotes the set of Lebesgue measure zero (by Rademacher's theorem), where f fails to be differentiable. If f is also convex on a neighborhood of $\bar{x}$, then $\partial_C f(\bar{x}) = \partial f(\bar{x})$. If f is strictly differentiable at $\bar{x}$, then $\partial_C f(\bar{x}) = \{\nabla f(\bar{x})\}$ [22]. The following lemma is helpful to the calculation of Clarke subdifferential.

Lemma 2.3. [22] *The following assertions hold:*

(i) *For any scalar s, one has $\partial_C(sf)(x) = s\partial_C f(x)$.*
(ii) *$\partial_C(f+g)(x) \subset \partial_C f(x) + \partial_C g(x)$, and equality holds if either f or g is strictly differentiable.*

Definition 2.3. (strong Kurdyka–Łojasiewicz property) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function. One says that f satisfies the strong Kurdyka–Łojasiewicz inequality at $\hat{x} \in \mathbb{R}^n$ if there exist $\sigma \in (0,+\infty]$, a neighborhood U of $\hat{x}$ and a concave function $\phi : [0,\sigma) \to [0,+\infty)$ such that

(i) $\phi(0) = 0$;
(ii) ϕ is of class $\mathscr{C}$ on $(0,\sigma)$;
(iii) $\phi' > 0$ on $(0,\sigma)$;
(iv) for all $x \in U$ with $f(\hat{x}) < f(x) < f(\hat{x}) + \sigma$, one has $\phi'\big(f(x) - f(\hat{x})\big)\,\text{dist}\big(0, \partial_C f(x)\big) \geq 1$.

3. Boosted Proximal Difference-of-Convex Algorithm

In this section, a boosted proximal difference-of-convex algorithm (BpDCA) is introduced and its convergence is established with the aid of strong Kurdyka–Łojasiewicz inequality.

3.1. Algorithm.

Throughout this paper, the following three assumptions are made.

Assumption 1. The function h is strongly convex with modulus $\rho > 0$.

Assumption 2. The function f is smooth convex with a Lipschitz continuous gradient whose Lipschitz continuity modulus is $L_{\nabla f} > 0$. The function g is continuously differentiable on an open set containing $\text{dom}\, h$. The function h is subdifferentiable at every point in $\text{dom}\, h$, i.e., $\partial h(x) \neq \emptyset$ for all $x \in \text{dom}\, h$.

Assumption 3. The function F is lower bounded, i.e., $\nu := \inf_{x\in\mathbb{R}^n} F(x) > -\infty$.

Definition 3.1. Let F be given in (1.1). One says that $\bar{x}$ is a critical point of F if

$$\nabla f(\bar{x}) + \nabla g(\bar{x}) \in \partial h(\bar{x}).$$

To accelerate the pDCA, we propose the following boosted proximal difference-of-convex algorithm.

Remark 3.1. We review the relation of the BpDCA to previous work.

(i) If one sets $\lambda_k \equiv 0$ for any $k \in \mathbb{N}$, the iterations of the BpDCA and the pDCA coincide. Hence, our convergence results for the BpDCA apply in particular to the pDCA.

(ii) If one sets $f = 0$ and $L_{\nabla f} = 0$, the BpDCA becomes the BDCA in [19].

Algorithm 1 BpDCA

Input: Fix $\alpha > 0, \bar{\lambda} > 0$ and $0 < \beta < 1$. Let x^0 be any initial point.
for $k = 0, 1, 2, \ldots,$ **do**

1. Take any $\xi^k \in \partial h(x^k)$ and solve the strongly convex minimization problem

$$y^k = \operatorname*{arg\,min}_{x \in \mathbb{R}^n} g(x) + \langle \nabla f(x^k) - \xi^k, x\rangle + \frac{L_{\nabla f}}{2}\|x - x^k\|^2. \tag{3.1}$$

2. Set $d_k := y^k - x^k$. If $d^k = 0$, STOP and RETURN x^k. Otherwise, go to Step 3.
3. Set $\lambda_k := \bar{\lambda}$.
   WHILE $F(y^k + \lambda_k d_k) > F(y^k) - \alpha\lambda_k^2\|d_k\|^2$ DO $\lambda_k := \beta\lambda_k$.
4. Set $x^{k+1} := y^k + \lambda_k d_k$. If $x^{k+1} = x^k$ then STOP and RETURN x^k, otherwise set $k := k+1$, and go to Step 1.

end for

3.2. **Global subsequential convergence.** In the following proposition, we prove that $d_k := y_k - x_k$ is a descent direction for F at y_k. Since the value of F is always reduced at y_k with respect to that at x_k, one can achieve a larger decrease by moving along the direction d_k. This simple fact, which is the key idea of the BpDCA, improves the performance of the pDCA in many applications.

Proposition 3.1. *For all $k \in \mathbb{N}$, the following statements hold:*

(i) $F(y^k) \leq F(x^k) - \frac{L_{\nabla f}}{2}\|d_k\|^2$*;*
(ii) $F'(y^k; d_k) \leq -\rho\|d_k\|^2$*;*
(iii) *there exists some $\delta_k > 0$ such that $F(y^k + \lambda_k d_k) \leq F(y^k) - \alpha\lambda^2\|d_k\|^2$ $\forall \lambda \in [0, \delta_k]$, so the backtracking step 3 of BpDCA terminates finitely.*

Proof. (i) In view of the facts that $g(x) + \langle \nabla f(x^k) - \xi^k, x\rangle + \frac{L_{\nabla f}}{2}\|x - x^k\|^2$ in (3.1) is an $L_{\nabla f}$-strongly convex function and y^k is its global minimizer, by (2.2), we have

$$\begin{aligned} &g(y^k) + \langle \nabla f(x^k) - \xi^k, y^k\rangle + \frac{L_{\nabla f}}{2}\|y^k - x^k\|^2 \\ \leq &g(x^k) + \langle \nabla f(x^k) - \xi^k, x^k\rangle + \frac{L_{\nabla f}}{2}\|x^k - x^k\|^2 - \frac{L_{\nabla f}}{2}\|x^k - y^k\|^2, \end{aligned}$$

which implies

$$g(y^k) \leq g(x^k) + \langle \nabla f(x^k) - \xi^k, x^k - y^k\rangle - L_{\nabla f}\|x^k - y^k\|^2. \tag{3.2}$$

From Lemma 2.2 and the fact that ∇f is $L_{\nabla f}$-Lipschitz continuous, it follows

$$f(y^k) \leq f(x^k) + \langle \nabla f(x^k), y^k - x^k\rangle + \frac{L_{\nabla f}}{2}\|y^k - x^k\|^2. \tag{3.3}$$

Due to $\xi^k \in \partial h(x^k)$, (2.1) implies

$$h(y^k) \geq h(x^k) + \langle \xi^k, y^k - x^k\rangle. \tag{3.4}$$

Putting (3.2), (3.3) and (3.4) together yields

$$\begin{aligned} F(y^k) =& f(y^k)+g(y^k)-h(y^k) \\ \leq& f(x^k)+\langle \nabla f(x^k), y^k-x^k\rangle + \frac{L_{\nabla f}}{2}\|y^k-x^k\|^2+g(x^k)+\langle \nabla f(x^k)-\xi^k, x^k-y^k\rangle \\ &-L_{\nabla f}\|x^k-y^k\|^2-\Big(h(x^k)+\langle \xi^k, y^k-x^k\rangle\Big) \\ =& f(x^k)+g(x^k)-h(x^k)-\frac{L_{\nabla f}}{2}\|x^k-y^k\|^2 \\ =& F(x^k)-\frac{L_{\nabla f}}{2}\|d_k\|^2. \end{aligned}$$

The proof is completed.

(ii) It follows from the definition of the one-sided directional derivative $F'(y^k;d_k)$ that

$$\begin{aligned} &F'(y^k;d_k) \\ =&\lim_{t\downarrow 0}\frac{F(y^k+td_k)-F(y^k)}{t} \\ =&\lim_{t\downarrow 0}\frac{f(y^k+td_k)-f(y^k)}{t}+\lim_{t\downarrow 0}\frac{g(y^k+td_k)-g(y^k)}{t}-\lim_{t\downarrow 0}\frac{h(y^k+td_k)-h(y^k)}{t} \\ \leq&\langle \nabla f(y^k), d_k\rangle+\langle \nabla g(y^k), d_k\rangle-\langle v, d_k\rangle \\ =&\langle \nabla f(y^k)+\nabla g(y^k)-v, y^k-x^k\rangle, \end{aligned}$$

where $v\in\partial h(y^k)$. From the first-order optimality condition of (3.1), we obtain

$$\nabla g(y^k)+\nabla f(x^k)+L_{\nabla f}(y^k-x^k)=\xi^k\in\partial h(x^k). \tag{3.5}$$

By Lemma 2.1 and the fact that h is ρ-strongly convex, we have that ∂h is ρ-strongly monotone. Combining this and (3.5), it holds that

$$\langle \nabla g(y^k)+\nabla f(x^k)+L_{\nabla f}(y^k-x^k)-v, x^k-y^k\rangle \geq \rho\|x^k-y^k\|^2.$$

Therefore, we deduce that

$$\langle \nabla f(y^k)+\nabla g(y^k)-v, x^k-y^k\rangle+\langle \nabla f(x^k)-\nabla f(y^k), x^k-y^k\rangle-L_{\nabla f}\|x^k-y^k\|^2\geq\rho\|x^k-y^k\|^2,$$

which implies that

$$\begin{aligned} &\langle \nabla f(y^k)+\nabla g(y^k)-v, y^k-x^k\rangle \\ \leq&\langle \nabla f(x^k)-\nabla f(y^k), x^k-y^k\rangle-(L_{\nabla f}+\rho)\|x^k-y^k\|^2 \\ \leq&-\rho\|x^k-y^k\|^2, \end{aligned}$$

where the last inequality comes from Cauchy-Schwarz inequality and the $L_{\nabla f}$-Lipschitz continuity of ∇f. Hence, $F'(y^k;d_k)\leq-\rho\|d_k\|^2$ holds.

(iii) If $d_k=0$, the conclusion is obviously valid. So let us consider the case of $d_k\neq 0$. From the definition of the one-sided directional derivative $F'(y^k;d_k)$, we have

$$\lim_{\lambda\downarrow 0}\frac{F(y^k+\lambda d_k)-F(y^k)}{\lambda}=F'(y^k;d_k)\leq-\rho\|d_k\|^2<-\frac{\rho}{2}\|d_k\|^2<0.$$

Thus there exists $\tilde{\lambda}_k > 0$ such that $\frac{F(y^k+\lambda d_k)-F(y^k)}{\lambda} \le -\frac{\rho}{2}\|d_k\|^2$ for all $\lambda \in (0,\tilde{\lambda}_k]$, which yields

$$F(y^k+\lambda d_k) \le F(y^k) - \frac{\rho\lambda}{2}\|d_k\|^2, \quad \forall \lambda \in (0,\tilde{\lambda}_k].$$

Setting $\delta_k := \min\{\tilde{\lambda}_k, \frac{\rho}{2\alpha}\}$, we obtain

$$F(y^k+\lambda d_k) \le F(y^k) - \alpha\lambda^2\|d_k\|^2, \quad \forall \lambda \in (0,\delta_k].$$

The proof is completed. □

Theorem 3.1. *For any $x_0 \in \mathbb{R}^n$, either BpDCA returns a critical point of problem (1.1) or it generates an infinite sequence such that the following holds:*

(i) *$\{F(x^k)\}_{k\in\mathbb{N}}$ is monotonically decreasing and convergent to some F^*.*

(ii) *Any limit point of $\{x^k\}_{k\in\mathbb{N}}$ is a critical point to problem (1.1). In addition, if F is coercive, then there exists a subsequence of $\{x^k\}_{k\in\mathbb{N}}$ which converges to a critical point of problem (1.1).*

(iii) *$\sum_{k=0}^{+\infty}\|d_k\|^2 < +\infty$. Further, if there is some $\bar{\lambda}$ such that $\lambda_k \le \bar{\lambda}$ for all $k \in \mathbb{N}$, then $\sum_{k=0}^{+\infty}\|x^{k+1}-x^k\|^2 < +\infty$.*

Proof. If BpDCA stops at step 2 and returns x^k, then $x^k = y^k$. Combining this and (3.5) yields

$$\nabla g(x^k) + \nabla f(x^k) = \xi^k \in \partial h(x^k).$$

Thus x^k is a critical point of the problem (1.1). Otherwise, BpDCA will produce an infinite sequence.

(i) By Proposition 3.1 and step 3 of BpDCA, we obtain

$$F(x^{k+1}) \le F(y^k) - \alpha\lambda_k^2\|d_k\|^2 \le F(x^k) - \left(\alpha\lambda_k^2 + \frac{L_{\nabla f}}{2}\right)\|d_k\|^2. \tag{3.6}$$

Therefore, we see that $\{F(x^k)\}_{k\in\mathbb{N}}$ is monotonically decreasing. Since function F is lower bounded, it holds that $\{F(x^k)\}_{k\in\mathbb{N}}$ converges to some F^*.

(ii) From (3.6), we deduce

$$\frac{L_{\nabla f}}{2}\|d_k\|^2 \le \left(\alpha\lambda_k^2 + \frac{L_{\nabla f}}{2}\right)\|d_k\|^2 \le F(x^k) - F(x^{k+1}). \tag{3.7}$$

Because of the convergence of $\{F(x^k)\}_{k\in\mathbb{N}}$, we have

$$\lim_{k\to+\infty}\|y^k - x^k\|^2 = \lim_{k\to+\infty}\|d_k\|^2 = 0. \tag{3.8}$$

If $\bar{x}$ is a limit point of $\{x^k\}_{k\in\mathbb{N}}$, then there exists a subsequence $\{x^{k_i}\}_{i\in\mathbb{N}} \subset \{x^k\}_{k\in\mathbb{N}}$ such that $\lim_{i\to+\infty} x^{k_i} \to \bar{x}$. Furthermore, by (3.8), we have $\lim_{i\to+\infty} y^{k_i} \to \bar{x}$. Substituting k_i for k in (3.5) yields

$$\nabla g(y^{k_i}) + \nabla f(x^{k_i}) + L_{\nabla f}(y^{k_i} - x^{k_i}) \in \partial h(x^{k_i}). \tag{3.9}$$

Taking the limit in (3.9) and combining the closedness of ∂h and the continuity of ∇f and ∇g, we deduce $\nabla g(\bar{x}) + \nabla f(\bar{x}) \in \partial h(\bar{x})$. Thus $\bar{x}$ is a critical point of problem (1.1). When F is coercive, it follows from (i) that $\{x^k\}_{k\in\mathbb{N}}$ is bounded, which implies that the limit point set of $\{x^k\}_{k\in\mathbb{N}}$ is not empty. Therefore, there exists a subsequence of $\{x^k\}_{k\in\mathbb{N}}$ which converges to a critical point of problem (1.1).

(iii) Summing (3.7) from $k=0$ to K, we have

$$\frac{L_{\nabla f}}{2}\sum_{k=0}^{K}\|d_k\|^2 \le F(x^0)-F(x^{K+1}) \le F(x^0)-\nu < +\infty,$$

thanks to the lower boundedness of F. Let $K\to+\infty$. It holds that $\sum_{k=0}^{+\infty}\|d_k\|^2<+\infty$. It follows from the Step 4 of BpDCA that $x^{k+1}-x^k = y^k - x^k + \lambda_k d_k = (1+\lambda_k)d_k$. So, we have

$$\sum_{k=0}^{+\infty}\|x^{k+1}-x^k\|^2 = \sum_{k=0}^{+\infty}(1+\lambda_k)^2\|d_k\|^2 \le (1+\bar{\lambda})^2\sum_{k=0}^{+\infty}\|d_k\|^2 < +\infty.$$

□

3.3. **Global convergence and convergence rate.** Firstly, we prove the convergence of the sequence generated by BpDCA as long as the sequence has a cluster point at which F satisfies the strong Kurdyka-Łojasiewicz inequality and ∇g is locally Lipschitz.

Theorem 3.2. *For any $x_0\in\mathbb{R}^n$, consider the sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by the BpDCA. Suppose that $\{x^k\}_{k\in\mathbb{N}}$ has a cluster point x^*, that ∇g is locally Lipschitz continuous around x^* and that F satisfies the strong Kurdyka–Łojasiewicz inequality at x^*. Then $\{x^k\}_{k\in\mathbb{N}}$ converges to x^*, which is a critical point to problem (1.1).*

Proof. Since x^* is a cluster point of $\{x^k\}_{k\in\mathbb{N}}$, there exists a subsequence $\{x^{k_m}\}_{m\in\mathbb{N}}$ of $\{x^k\}_{k\in\mathbb{N}}$ such that $\lim_{m\to+\infty}x^{k_m}\to x^*$. From Theorem 3.1 (i), we have $\lim_{k\to+\infty}F(x^k)\to F^*$. Therefore, it follows from the continuity of F that

$$F(x^*) = \lim_{m\to\infty}F(x^{k_m}) = \lim_{k\to\infty}F(x^k) = F^*.$$

Thus F is finite and has the same value F^* at every cluster point of $\{x^k\}_{k\in\mathbb{N}}$.

Assume $F(x^k)=F^*$ for some $k>1$. Since $\{F(x^k)\}_{k\in\mathbb{N}}$ is decreasing and convergent to F^*, we have $F(x^k)=F(x^{k+1})$. From (3.7), we deduce that $d_k=0$, which implies that BpDCA terminates after a finite number of steps. Next, we assume that $F(x^k)>F^*$ for all $k\in\mathbb{N}$. Since ∇g is locally Lipschitz around x^*, there exist some constants $L_{\nabla g}\ge 0$ and $\delta_1>0$ such that

$$\|\nabla g(x)-\nabla g(y)\| \le L_{\nabla g}\|x-y\|, \quad \forall x,y\in\mathbb{B}(x^*,\delta_1). \tag{3.10}$$

Further, since F satisfies the strong Kurdyka-Łojasiewicz inequality at x^*, there exist $\sigma\in(0,+\infty)$, a neighborhood U of x^* and a continuous and concave function $\phi:[0,\sigma]\to[0,+\infty)$ such that

$$\phi'\big(F(x)-F^*\big)\,\mathrm{dist}\big(0,\partial_C F(x)\big)\ge 1 \tag{3.11}$$

for all $x\in\Lambda$, where

$$\Lambda = \{x\in\mathbb{R}^n : x\in U\}\cap\{x\in\mathbb{R}^n : F^*<F(x)<F^*+\sigma\}.$$

Take δ_2 small enough that $\mathbb{B}(x^*,\delta_2)\subset U$ and set $\delta := \frac{1}{2}\min\{\delta_1,\delta_2\}$. Let

$$M := \max_{\lambda\ge 0}\frac{2(L_{\nabla g}+L_{\nabla f})(1+\lambda)}{2\alpha\lambda^2+L_{\nabla f}}, \tag{3.12}$$

which is attained at $\hat{\lambda} = -1+\sqrt{1+\frac{L_{\nabla f}}{2\alpha}}$. Since $\lim_{k\to+\infty}F(x^k)=F^*$ and $F(x^k)>F^*$ for all $k\in\mathbb{N}$, there exists an index N_1 such that

$$F^*<F(x^k)<F^*+\sigma, \quad \forall k\ge N_1.$$

From Theorem 3.1 (iii), we obtain $d_k = y^k - x^k \to 0$. Thus there exists an index N_2 such that

$$\|y^k - x^k\| \leq \delta, \quad \forall k \geq N_2.$$

Because $\lim_{m\to+\infty} x^{k_m} = x^*$ and ϕ is continuous, one sees that there exists an index $N \geq \max\{N_1, N_2\}$ such that

$$x^N \in \mathbb{B}(x^*, \delta) \tag{3.13}$$

and

$$\|x^N - x^*\| + M\phi\big(F(x^N) - F^*\big) < \delta. \tag{3.14}$$

For all $k \geq N$ such that $x^k \in \mathbb{B}(x^*, \delta)$, we deduce

$$\|y^k - x^*\| \leq \|y^k - x^k\| + \|x^k - x^*\| \leq 2\delta \leq \delta_1.$$

which implies that $y^k \in \mathbb{B}(x^*, \delta_1)$. Therefore, by (3.10), it holds

$$\|\nabla g(y^k) - \nabla g(x^k)\| \leq L_{\nabla g}\|y^k - x^k\|. \tag{3.15}$$

On the other hand, for all $k \geq N$ such that $x^k \in \mathbb{B}(x^*, \delta)$, we have $x^k \in \Lambda$. Therefore, from (3.11), for all $k \geq N$ such that $x^k \in \mathbb{B}(x^*, \delta)$, we have

$$\phi'\big(F(x^k) - F^*\big)\,\mathrm{dist}\big(0, \partial_C F(x^k)\big) \geq 1. \tag{3.16}$$

Using the definition of x^{k+1}, we obtain

$$\begin{aligned} x^{k+1} &= y^k + \lambda_k d_k \\ &= y^k + \lambda_k(y^k - x^k) \\ &= (1+\lambda_k)y^k - (1+\lambda_k)x^k + x^k \\ &= (1+\lambda_k)(y^k - x^k) + x^k \end{aligned} \tag{3.17}$$

which implies that $y^k - x^k = \frac{1}{1+\lambda_k}(x^{k+1} - x^k)$. From (3.5) and Lemma 2.3, we obtain

$$\begin{aligned} \nabla g(y^k) - \nabla g(x^k) + L_{\nabla f}(y^k - x^k) &\in \partial h(x^k) - \nabla f(x^k) - \nabla g(x^k) \\ &= \partial_C\big(-F(x^k)\big) = -\partial_C F(x^k). \end{aligned} \tag{3.18}$$

Combining (3.15), (3.17) and (3.18), we deduce

$$\begin{aligned} \mathrm{dist}\big(0, \partial_C F(x^k)\big) &\leq \|\nabla g(x^k) - \nabla g(y^k) + L_{\nabla f}(x^k - y^k)\| \\ &\leq \|\nabla g(x^k) - \nabla g(y^k)\| + L_{\nabla f}\|x^k - y^k\| \\ &\leq (L_{\nabla g} + L_{\nabla f})\|y^k - x^k\| \\ &= \frac{L_{\nabla g} + L_{\nabla f}}{1+\lambda_k}\|x^{k+1} - x^k\|. \end{aligned} \tag{3.19}$$

For all $k \geq N$ such that $x^k \in \mathbb{B}(x^*,\delta)$, it follows from (3.19) that

$$\begin{aligned}
&\frac{L_{\nabla g}+L_{\nabla f}}{1+\lambda_k}\|x^{k+1}-x^k\|\big(\phi(F(x^k)-F^*)-\phi(F(x^{k+1})-F^*)\big)\\
\geq&\operatorname{dist}\big(0,\partial_C F(x^k)\big)\big(\phi(F(x^k)-F^*)-\phi(F(x^{k+1})-F^*)\big)\\
\geq&\operatorname{dist}\big(0,\partial_C F(x^k)\big)\phi'\big(F(x^k)-F^*\big)\big(F(x^k)-F(x^{k+1})\big)\\
\geq&F(x^k)-F(x^{k+1})\\
\geq&\left(\alpha\lambda_k^2+\frac{L_{\nabla f}}{2}\right)\|y^k-x^k\|^2\\
=&\frac{2\alpha\lambda_k^2+L_{\nabla f}}{2(1+\lambda_k)^2}\|x^{k+1}-x^k\|^2,
\end{aligned} \tag{3.20}$$

where the second inequality comes the fact that ϕ is a concave function, the third and fourth inequalities hold due to (3.16) and (3.7), and the last equality originates from (3.17). Using (3.20) and (3.12), we have

$$\begin{aligned}
\|x^{k+1}-x^k\| \leq&\frac{2(L_{\nabla g}+L_{\nabla f})(1+\lambda_k)}{2\alpha\lambda_k^2+L_{\nabla f}}\big(\phi(F(x^k)-F^*)-\phi(F(x^{k+1})-F^*)\big)\\
\leq&M\big(\phi(F(x^k)-F^*)-\phi(F(x^{k+1})-F^*)\big).
\end{aligned} \tag{3.21}$$

Next, we prove that $x^k \in \mathbb{B}(x^*,\delta)$ for all $k \geq N$. From (3.13), the claim holds for $k=N$. we suppose that it holds for $k=N,N+1,\dots,N+p-1$ with $p\geq 1$. Then, we know that (3.21) is valid for $k=N,N+1,\dots,N+p-1$. Hence,

$$\begin{aligned}
\|x^{N+p}-x^*\| \leq&\|x^N-x^*\|+\sum_{i=1}^{p}\|x^{N+i}-x^{N+i-1}\|\\
\leq&\|x^N-x^*\|+M\sum_{i=1}^{p}\left[\phi\left(F(x^{N+i-1})-F^*\right)-\phi\left(F(x^{N+i})-F^*\right)\right]\\
\leq&\|x^N-x^*\|+M\phi\left(F(x^N)-F^*\right)<\delta,
\end{aligned} \tag{3.22}$$

where the last inequality follows from (3.14). We can conclude from (3.22) that $x^{N+p} \in \mathbb{B}(x^*,\delta)$. Therefore, it holds that $x^k \in \mathbb{B}(x^*,\delta)$ for all $k \geq N$. Summing (3.21) from $k=N$ to J, we have

$$\sum_{k=N}^{J}\|x^{k+1}-x^k\| \leq M\phi\left(F(x^N)-F^*\right)<+\infty.$$

Taking the limit as $J \to +\infty$, we deduce $\sum_{k=1}^{+\infty}\|x^{k+1}-x^k\|<+\infty$, which implies that $\{x^k\}_{k\in\mathbb{N}}$ is a Cauchy sequence. Further, the sequence $\{x^k\}_{k\in\mathbb{N}}$ converges to x^*. From Theorem 3.1 (ii), x^* is a critical point to problem (1.1). □

Employing arguments which are similar to those used in the proof of [19, Theorem 4.9], we can easily obtain the following rate of convergence of BpDCA.

Theorem 3.3. (Rate of Convergence) *Let the sequence $\{x^k\}_{k\in\mathbb{N}}$ be generated by* BpDCA *and suppose further that $\{x^k\}_{k\in\mathbb{N}}$ converges to some x^*. Suppose that ∇g is locally Lipschitz continuous around x^* and that F satisfies the strong Kurdyka-Łojasiewicz inequality at x^* with $\phi(s)=cs^{1-\theta}$ for some $c>0$ and $\theta\in[0,1)$. Then*

(i) *If* $\theta = 0$, *the sequence* $\{x^k\}_{k\in\mathbb{N}}$ *converges to* x^* *in a finite number of steps.*

(ii) *If* $\theta \in (0, \frac{1}{2}]$, *the sequence* $\{x^k\}_{k\in\mathbb{N}}$ *converges linearly to* x^*.

(iii) *If* $\theta > (\frac{1}{2}, 1)$, *then there exists* $\eta > 0$ *such that for all sufficiently large* k, $\|x^k - x^*\| < \eta k^{-\frac{1-\theta}{2\theta-1}}$.

4. Numerical Experiments

In this section, we illustrate performance of our algorithm BpDCA and compare it with the pDCA (1.3) through two numerical examples. All the codes are written in MATLAB (version R2017a) and run on a personal ASUS computer with Intel(R) Core(TM) m3-7Y30 CPU @ 1.00GHz 1.61GHz and RAM 8.00GB.

In our numerical tests, we focus on the following DC regularized least squares problem:

$$\min_{x\in\mathbb{R}^n} \Psi(x) = \frac{1}{2}\|Ax-b\|^2 + g(x) - h(x), \tag{4.1}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, g is a continuously differentiable convex function, and h is a continuous and strongly convex function. Observe that $L = \lambda_{\max}(A^T A)$.

We present numerical experiments for solving the problem (4.1) on random instances generated as follows. We first generate an $m \times n$ matrix A with i.i.d. standard Gaussian entries, and then normalize this matrix so that the columns of A have unit norms. A subset D of the size K is then chosen uniformly at random from $\{1, 2, 3, \ldots, n\}$ and a K-sparse vector y having i.i.d. standard Gaussian entries on D is generated. Finally, we set $b = Ay$.

We next consider two different classes of regularizers: the ℓ_{1-2} regularizer and the logarithmic regularizer. In the numerical results listed in the following tables, we consider $(m, n, K) = (120i, 512i, 20i)$ for $i = 1, 2, \ldots, 10$. We run 10 instances randomly for each (m, n, K) and report the number of iterations (Iter), CPU times in seconds (CPU time) and the function values at termination (fval), averaged the 10 random instances. We terminate the algorithms in the experiments when

$$\frac{\|x^k - x^{k-1}\|}{\max\{1, \|x^k\|\}} < 10^{-2}.$$

4.1. **The least squares problem with the ℓ_{1-2} regularizer.** In this subsection, we consider the ℓ_{1-2} regularized least squares problem:

$$\min_{x\in\mathbb{R}^n} \Psi_{\ell_{1-2}}(x) = \frac{1}{2}\|Ax-b\|^2 + \mu\|x\|_1 - \mu\|x\|_2, \tag{4.2}$$

where $\mu > 0$ is the regularization parameter. This problem takes the form of (1.1) with $f(x) = \frac{1}{2}\|Ax-b\|^2$, $g(x) = \mu\|x\|_1$, and $h(x) = \mu\|x\|_2$. In addition, we suppose that A in (4.2) does not have zero columns. Using this hypothesis, we see that if we choose $\mu < \frac{1}{2}\|A^T b\|_\infty$, then the assumptions in Theorem 3.2 are satisfied (see Subsection 5.1 in [16]). We set $\mu = 0.5$ in (4.2).

We now make comparisons of two algorithms with a randomly initial point belonging to the product space $(0,1)^n$. We take $\bar{\lambda} = 50$ for BpDCA and the parameters of different problem sizes are selected as shown in Table 1. In the process of numerical experiments, we find that the different parameters have an influence on the search time of the step 3 in Algorithm 1. Table 2 illustrates that BpDCA behaves better than pDCA in Iter, CPU time and fval. As the problem size becomes larger, the advantages become more obvious. In addition, the number of iterations of pDCA is about 6 times that of BpDCA, and the CPU times is about 1.5 times.

TABLE 1. Selection of parameters α and β of different problem sizes in BpDCA.

i	α	β
1, 2, 3, 4	0.6	0.6
5, 6, 7	0.8	0.5
8, 9, 10	0.5	0.6

TABLE 2. Comparison of BpDCA and pDCA for solving (4.2).

Problem size			Iter		CPU time		fval	
m	n	K	BpDCA	pDCA	BpDCA	pDCA	BpDCA	pDCA
120	512	20	1056	6043	1.45	3.4984	2.8895e-2	3.1042e-2
240	1024	40	2188	12463	4.3875	10.6422	3.8662e-2	4.0173e-2
360	1536	60	3243	18835	29.1016	41.7219	4.4610e-2	4.5929e-2
480	2048	80	4395	25461	64.2281	87.0922	4.9500e-2	5.0210e-2
600	2560	100	5329	31499	108.475	181.7094	5.4036e-2	5.4408e-2
720	3072	120	6275	36903	214.6484	367.4141	5.3719e-2	5.4202e-2
840	3584	140	7880	46682	340.1516	575.8125	5.6342e-2	5.6830e-2
960	4096	160	8929	52438	575.2766	797.3859	5.5155e-2	5.6359e-2
1080	4608	180	10387	60342	855.2875	1160.1922	5.7326e-2	5.8441e-2
1200	5120	200	11714	68040	1146.0766	1565.7125	5.7869e-2	5.8674e-2

To illustrate the ability of recovering the original sparse solution by BpDCA and pDCA, we plot in Figure 1 the true solution and the solutions obtained on a random instance with $(m,n,K) = (480,2048,80)$. The true solution is represented by asterisks, while circles are the estimates obtained by BpDCA and pDCA. We see that the estimates obtained by BpDCA and pDCA are quite close to the true values.

4.2. **The least squares problem with the logarithmic regularizer.** The least squares problem with the logarithmic regularization function is defined as:

$$\min_{x\in\mathbb{R}^n} \Psi_{\log}(x) = \frac{1}{2}\|Ax-b\|^2 + \sum_{i=1}^{n} \left[\mu \log(|x_i|+\varepsilon) - \mu \log \varepsilon\right], \tag{4.3}$$

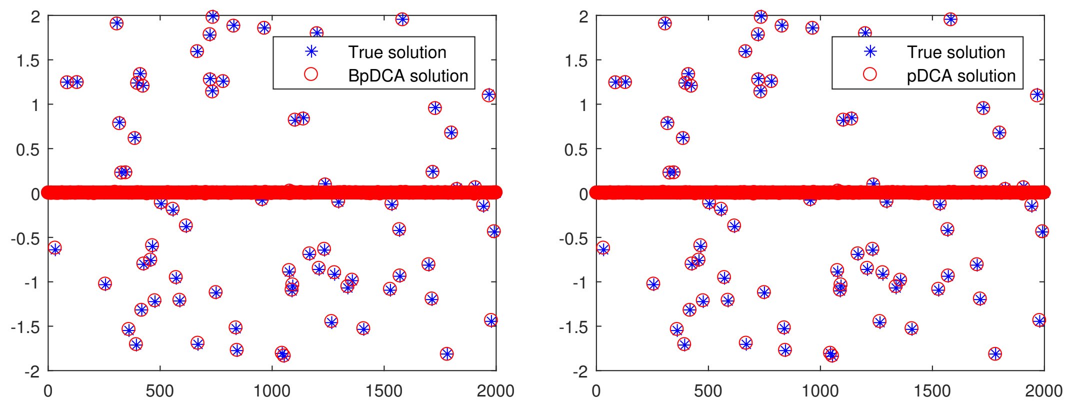

FIGURE 1. The true solution and the solution by BpDCA (left) and pDCA (right)

where $\varepsilon > 0$ is a constant, and $\mu > 0$ is the regularization parameter. Take $f(x) = \frac{1}{2}\|Ax-b\|^2$, $g(x) = \frac{\mu}{\varepsilon}\|x\|_1$ and $h(x) = \sum_{i=1}^{n} \mu \left[\frac{|x_i|}{\varepsilon} - \log(|x_i|+\varepsilon) + \log\varepsilon\right]$. It is not hard to demonstrate that assumptions in Theorem 3.2 are satisfied (see Subsection 5.2 in [16]). We set $\mu = 0.5$ and $\varepsilon = 3$ in (4.3).

TABLE 3. Comparison of BpDCA and pDCA for solving (4.3).

Problem size			Iter		CPU time		fval	
m	n	K	BpDCA	pDCA	BpDCA	pDCA	BpDCA	pDCA
120	512	20	3269	17126	5.5688	11.6859	4.8309e-3	5.1096e-3
240	1024	40	6730	34357	29.8844	40.7406	5.9484e-3	6.1199e-3
360	1536	60	10545	53789	93.9156	162.3656	6.3557e-3	6.4732e-3
480	2048	80	14669	74857	219.6922	404.0094	6.4069e-3	6.7847e-3
600	2560	100	17995	90230	369.5703	677.7266	6.6728e-3	6.9916e-3
720	3072	120	21519	108831	627.7609	1058.9016	6.3252e-3	7.0984e-3
840	3584	140	24387	122795	889.2719	1519.7547	7.1042e-3	7.4170e-3
960	4096	160	28089	141383	1294.7156	2267.0359	6.5396e-3	7.3894e-3
1080	4608	180	32225	163572	1745.0188	3215.6313	6.8676e-3	7.5324e-3
1200	5120	200	35586	178893	2355.1281	4203.2953	7.0109e-3	7.6203e-3

We now make comparisons of two algorithms with a randomly initial point belonging to the product space $(0,1)^n$. We take $\bar{\lambda} = 50$, $\alpha = 0.5$ and $\beta = 0.2$ for BpDCA. From Table 3, we draw the conclusions similar to that in Section 4.1. Figure 2 demonstrates the ability of recovering the original sparse solution by BpDCA and pDCA for a random instance with $(m,n,K) = (480,2048,80)$.

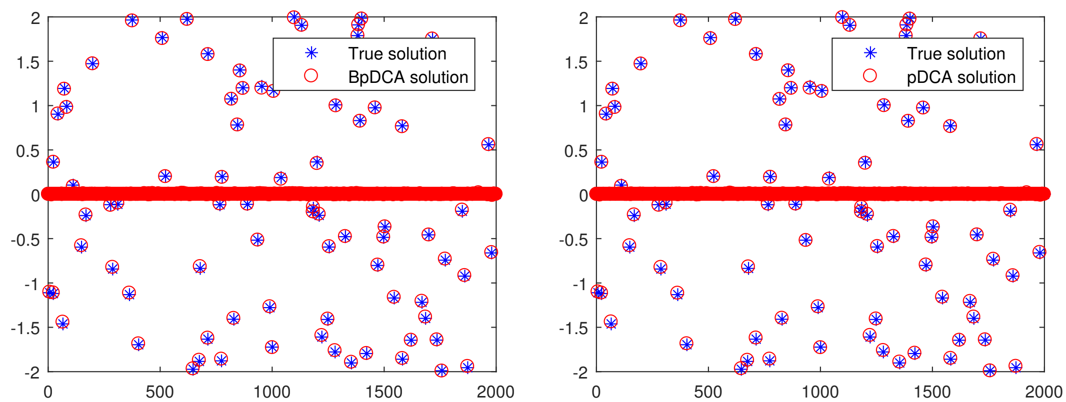

FIGURE 2. The true solution and the solution by BpDCA (left) and pDCA (right)

Acknowledgements

The first and second authors were supported by Scientific Research Project of Aeronautical Science Foundation of China (20200008067001).

REFERENCES

[1] P.D. Tao, E.B. Souad, Algorithms for solving a class of nonconvex optimization problems. Methods of subgradients, In: Hiriart-Urruty J.B. (Ed.), Fermat days 85: Mathematics for Optimization, North-Holland Mathematics Studies, vol. 129, pp. 249-271, 1986.

[2] Y. Lou, P. Yin, Q. He, J. Xin, Minimization of ℓ_{1-2} for compressed sensing, J. Sci. Comput. 37 (2015), A536–A563.

[3] M. Nouiehed, J.S. Pang, M. Razaviyayn, On the pervasiveness of difference-convexity in optimization and statistics, Math. Program. 174 (2019), 195-222.

[4] H.A. Le Thi, T. Pham Dinh, H.M. Le, X.T. Vo, DC approximation approaches for sparse optimization, Eur. J. Oper. Res. 244 (2015), 26-46.

[5] G. Li, L. Yang, Z. Wu, G. Wu, D.C. programming for sparse proximal support vector machines, Info. Sc. 547 (2021), 187-201.

[6] M. Gaudioso, E. Gorgone, J.B. Hiriart-Urruty, Feature selection in SVM via polyhedral k-norm, Optim. Lett. 14 (2020), 19-36.

[7] K. Tu, H. Zhang, H. Gao, et al., A hybrid Bregman alternating direction method of multipliers for the linearly constrained difference-of-convex problems, J. Glob. Optim. 76 (2020), 665-693.

[8] M. Ding, X. Song, B. Yu, An inexact proximal DC algorithm with sieving strategy for rank constrained least squares semidefinite programming, J. Sci. Comput. 91 (2022), 75.

[9] T. Liu, T.K. Pong, A. Takeda, A refined convergence analysis of pDCAe with applications to simultaneous sparse recovery and outlier detection, Comput. Optim. Appl. 73 (2019), 69-100.

[10] Y. Lou, M. Yan, Fast L1-L2 Minimization via a proximal operator, J. Sci. Comput. 74 (2918), 767-785.
[11] T. Liu, T.K. Pong, A. Takeda, A successive difference-of-convex approximation method for a class of non-convex nonsmooth optimization problems, Math. Program. 176 (2019), 339-367.
[12] H.A. Le Thi, P.D. Tao, DC programming and DCA: thirty years of developments, Math. Program. 169 (2018), 5-68.
[13] S. He, Q.L. Dong, T.M. Rassias, Contractive difference-of-convex algorithms, submitted to J. Global Optim.
[14] J. Gotoh, A. Takeda, K. Tono, DC formulations and algorithms for sparse optimization problems, Math. Program. 169 (2018), 141-176.
[15] B. O′Donoghue, E. Candès, Adaptive restart for accelerated gradient schemes, Found. Comput. Math. 15 (2015), 715-732.
[16] B. Wen, X. Chen, T.K. Pong, A proximal difference-of-convex algorithm with extrapolation, Comput. Optim. Appl. 69 (2018), 297-324.
[17] L. Gao, B. Wen, Convergence rate analysis of an extrapolated proximal difference-of-convex algorithm, J. Appl. Math. Comput. (2022) doi: 10.1007/s12190-022-01797-w.
[18] F.J. Aragón Artacho, R.M.T. Fleming, P.T. Vuong, Accelerating the DC algorithm for smooth functions, Math. Program. 169 (2018), 95-118.
[19] F.J. Aragón Artacho, P.T. Vuong, The boosted difference of convex functions algorithm for nonsmooth functions, SIAM J. Optim. 30 (2020), 980-1006.
[20] H.M. Xu, H. Xue, X.H. Chen, Y.Y. Wang, Solving indefinite kernel support vector machine with difference of convex functions programming, AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, pp. 2782–2788, 2017.
[21] H.H. Bauschke, P.L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, 2nd ed. Springer, New York, 2017.
[22] F.H. Clarke, Optimization and Nonsmooth Analysis, 2nd ed., Classics in Appl. Math. 5, SIAM, Philadelphia, 1990.