J. Appl. Numer. Optim. 5 (2023), No. 3, pp. 391-398
Available online at http://jano.biemdas.com
https://doi.org/10.23952/jano.5.2023.3.07

PROXIMAL LINEAR METHODS FOR DC COMPOSITE MINIMIZATION PROBLEMS

ABDELLATIF MOUDAFI

Aix-Marseille Université, Laboratoire d'Informatique et Systèmes
(LIS UMR 7020 CNRS / AMU /UTLN) Marseille, France

Abstract. In this paper, we introduce two linearized proximal algorithms for solving DC composite optimization problems. The basic algorithms that we rely are the proximal-linear(ized) methods, which in each iteration solve regularized subproblems formed by linearizing the smooth maps and the concave component, respectively. It is proved that the two proposed algorithms provide descent methods and that if the sequences generated by the algorithms are bounded, every cluster points are critical points of the functions under consideration. Finally, a conclusion is stated and some directions for further research are suggested.

Keywords. Composite functions; DC functions; DCA linearized algorithm; Proximal linearized algorithm; Critical point.

2020 Mathematics Subject Classification. 49J53, 65K10, 49M37, 90C25.

1. Introduction and Preliminaries

In this paper, we are interested in the following class of composite optimization problems

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

where $g : \mathbb{R}^m \to \mathbb{R}\cup\{+\infty\}$ and $h : \mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$ are closed and convex functions and $c : \mathbb{R}^d \to \mathbb{R}^m$ is a smooth map with the composition being on the convex component.

We focus on these optimization problems where the loss function is convex with concave penalty that are common in sparse regression, compressive sensing, sparse approximation and which also brings to mind canonical DC problems and clearly can be used to model reasonably a large class of real-world systems; see, e.g., [1, 2, 3] and the references therein.

A look will also be taken when the composition is on the concave component, namely

$$\min_{x\in\mathbb{R}^d} f(x) := g(x) - h(c(x)), \tag{1.2}$$

where $g : \mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$ and $h : \mathbb{R}^m \to \mathbb{R}\cup\{+\infty\}$ are closed convex functions, and $c : \mathbb{R}^d \to \mathbb{R}^m$ is a smooth map.

Regularized nonlinear least squares and exact penalty formulations of nonlinear programs are classical examples, while notable contemporary instances include robust phase retrieval and matrix

E-mail address: abdellatif.moudafi@univ-amu.fr.
Received September 13, 2022; Accepted August 29, 2023.

factorization problems. The setting where c is the identity function, both (1.1) and (1.2) reduce to

$$\min_{x\in\mathbb{R}^d} f(x) := g(x) - h(x), \tag{1.3}$$

which is nothing else than DC minimization (minimization of a difference of two convex functions, that use convex properties of the two convex functions separately) which is now common place in large-scale optimization. The DCA investigated by Tao and An [4], is a popular algorithm for DC minimization and a proximal point algorithm proposed in [5]. Moudafi and Maingé [6] proposed an alternative proof of the main result of [5] with some extensions. A proximal linearized algorithm for minimizing DC functions was then proposed in [7]. DC optimization algorithms were proved to be particularly successful for analyzing and solving a variety of highly structured and practical problems. As for (1.1), regularized nonlinear least squares and exact penalty formulations of nonlinear programs are classical examples, while notable contemporary instances include robust phase retrieval and matrix factorization problems. In statistical estimation often, one is interested in minimizing an error between a nonlinear/linear process model $G(x)$ and observed data b through amisfit measure h. The resulting problem takes the form

$$\min_x h(b - G(x)) + g(x),$$

where g may be a convex surrogate encouraging prior structural information on x, such as the l_1-norm, the squared l_2-norm classically, but also for example l_1-l_2-norm; see [8, 9] and the references therein. The misfit $h = l_2$-norm appears in particular in nonlinear least squares while $h = l_1$-norm is used in the Least Absolute Deviations technique in regression and for robust phase retrieval. Concave sparse penalty are common in sparse regression, compressive sensing, and sparse approximation [1], Fan and Li [2] proposed the smoothly clipped absolute deviation (SCAD) regularizer that behaves like the l_1-norm near the origin, transitioning (via a concave quadratic) to a constant for large loss values.

Our goal in this paper centers around prox-linear methods and share the same idea, namely, they linearize some component $h(\cdot)$ or $c(\cdot)$; or both, which we propose to extend to the entire problem classes (1.1) and (1.2). We begin by describing our methods for minimizing composite DC functions which have the property that every cluster points of the sequences are critical points of the composite DC functions. In each iteration of (1.4) (resp. (1.5)), the prox-linear method linearizes the smooth map $c(\cdot)$ and h (resp. only h) solves, respectively the following proximal subproblems:

Given an initial point $x_0 \in dom f$ and a sequence of positive parameters $(\lambda_k)_{k\in\mathbb{N}}$ such that $\liminf_{k\to+\infty}\lambda_k > 0$, calculate $(w_k)_{k\in\mathbb{N}}$ and compute $(x_k)_{k\in\mathbb{N}}$ by

$$\begin{cases} w_k \in \partial h(x_k); \\ x_{k+1} = \arg\min_{x\in\mathbb{R}^d}\left(g(c(x_k) + \nabla c(x_k)(x - x_k)) - \langle w_k, x - x_k\rangle + \frac{1}{2\lambda_k}\|x - x_k\|^2\right). \end{cases} \tag{1.4}$$

If $x_{k+1} = x_k$, stop. Otherwise, set $k := k+1$ and return the first step.

Likewise, calculate $(w_k)_{k\in\mathbb{N}}$ and compute $(x_k)_{k\in\mathbb{N}}$ by

$$\begin{cases} w_k \in \nabla c(x_k)^* \partial h(c(x_k)); \\ x_{k+1} = \arg\min_{x\in\mathbb{R}^d}\left(g(x) - \langle w_k, x - x_k\rangle + \frac{1}{2\lambda_k}\|x - x_k\|^2\right). \end{cases} \tag{1.5}$$

If $x_{k+1} = x_k$, stop. Otherwise, set $k := k+1$ and return the first step.

Note that if c is the identity function, then Algorithm (1.4) and (1.5) become exactly the linearized proximal point algorithm introduced in [7] for DC functions. The setting where h is the null function, Algorithm (1.4) reduces to the proximal-linear algorithm for composite minimization [10] and Algorithm (1.5) is nothing else than the classical proximal point algorithm. The underlying assumption here is that the proximal subproblems can be solved efficiently.

Throughout this paper, we assume that the original function f is bounded below. As it is known that a necessary condition for $x \in dom f$ to be a local minimizer of f is in general hard to be reached, so we focus our attention on finding critical points of f. Before stating the definition of critical points, given a convex function $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, we recall that a vector v is called a subgradient of φ at a point $x \in dom\varphi$ if

$$\varphi(y) \geq \varphi(x) + \langle v, y - x\rangle \; \forall y \in \mathbb{R}^d.$$

The set of all subgradients of φ at x is denoted by $\partial\varphi(x)$, and is called the subdifferential of φ at x. For any point $x \notin dom\varphi$, we define $\partial\varphi(x)$ to be the empty set. For any closed function φ, its Frechét subdifferential at x, $\hat{\partial}\varphi(x)$, is the collection of vectors v such that

$$\varphi(y) \geq \varphi(x) + \langle v, y - x\rangle + o(\|y - x\|) \; \forall y \in \mathbb{R}^d.$$

Unfortunately, $\hat{\partial}\varphi$ can be empty at certain points ever for Lipschitz continuous functions. To avoid this degeneracy, we arrive at the limiting subdifferential, $\partial\varphi$, defined as

$$v \in \partial\varphi(x) \Leftrightarrow \exists x_k \to x, \varphi(x_k) \to \varphi(x), v_k \in \hat{\partial}\varphi(x_k), v_k \to v.$$

Clearly, $\hat{\partial}\varphi(x) \subset \partial\varphi(x)$ for all x. It is known that the above subdifferential reduces to the classical subdifferential in convex analysis when φ is convex. In addition, if φ is continuously differentiable, then the limiting subdifferential reduces to the gradient, $\nabla\varphi$, of the function φ. From the definition, it follows that if $\bar{x}$ is a local minimizer, then $0 \in \hat{\partial}\varphi(\bar{x})$ and $0 \in \partial\varphi(\bar{x})$, which generalizes the familiar Fermat's rule. The latter condition is in general hard to be reached and we relax it to the following notion of critical point. Throughout this paper, we also assume that the chain rule can be applied when needed which is always the case via qualification conditions as $N_{domg}(c(\bar{x})) \cap Ker(\nabla c(\bar{x})^*) = \{0\}$, N_{domg} being the normal cone to $domg$ or even better like $\mathbb{R}_+\big(domg - c(\bar{x})\big) - \nabla c(\bar{x})(\mathbb{R}^m) = \mathbb{R}^d$; see [11, 12] and the references therein.

Definition 1.1. Recall that $\bar{x} \in \mathbb{R}^d$ is a critical point of f in (1.1) if

$$\nabla c(\bar{x})^* \partial g(c(\bar{x})) \cap \partial h(\bar{x}) \neq \emptyset, \tag{1.6}$$

and that $\bar{x} \in \mathbb{R}^d$ is a critical point of f in (1.2) if

$$\partial g(\bar{x}) \cap \nabla c(\bar{x})^* \partial h(c(\bar{x})) \neq \emptyset. \tag{1.7}$$

2. P-L Method for DC Composite Minimization (1.4)

Throughout this section, we make the following assumptions on the functional components of the problem: $h : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is a proper, closed, convex function, $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is a convex and L-Lipschitz continuous function

$$\|g(x) - g(y)\| \leq L\|x - y\| \; \forall x, y \in \mathbb{R}^m,$$

and $c : \mathbb{R}^m \to \mathbb{R}^d$ is a C^1-smooth mapping with a β-Lipschitz continuous Jacobian map

$$\|\nabla c(x) - \nabla c(y)\| \leq \beta\|x - y\| \;\; \forall x, y \in \mathbb{R}^m.$$

The norm here is the operator norm.

The proposed algorithm seeks critical points of $f : g \circ c - h$ by constructing two sequences $(x_k)_{k\in\mathbb{N}}$ and $(w_k)_{k\in\mathbb{N}}$ by the rules (1.4). To establish its convergence, we begin by showing that Algorithm (1.4) is a descent method.

Theorem 2.1. *The sequence $(x_k)_{k\in\mathbb{N}}$ generated by Algorithm (1.4) satisfies*

(i) either the algorithm stops at a critical point.

(ii) or f decreases strictly, i.e. $f(x_{k+1}) < f(x_k)$ provided that $0 < \lambda_k < \frac{1}{L\beta}$ for all $k \in \mathbb{N}$.

Proof. From the two relations in (1.4) and the subdifferential chain rule, we have

$$w_k \in \partial h(x_k) \text{ and } w_k \in \frac{x_{k+1} - x_k}{\lambda_k} + \nabla c(x_k)^* \partial g\big(c(x_k) + \nabla c(x_k)(x_{k+1} - x_k)\big), \tag{2.1}$$

respectively. If $x_{k+1} = x_k$, then

$$w_k \in \partial h(x_k) \cap \nabla c(x_k)^* \partial g(c(x_k)),$$

which means that x_k is a critical point of f. Now, suppose $x_{k+1} \neq x_k$. By the definition of subdifferential, we can write

$$h(x_{k+1}) \geq h(x_k) + \langle w_k, x_{k+1} - x_k \rangle.$$

On the other hand, since x_{k+1} minimizes

$$g(c(x_k) + \nabla c(x_k)(\cdot - x_k)) - \langle w_k, \cdot - x_k \rangle + \frac{1}{2\lambda_k} \| \cdot - x_k \|^2,$$

we also have

$$g(c(x_k)) \geq g(c(x_k) + \nabla c(x_k)(x_{k+1} - x_k)) - \langle w_k, x_{k+1} - x_k \rangle + \frac{1}{2\lambda_k} \|x_{k+1} - x_k\|^2. \tag{2.2}$$

Combining the two last inequalities, we obtain

$$f(x_k) = g(c(x_k)) - h(x_k) \geq g(c(x_k) + \nabla c(x_k)(x_{k+1} - x_k)) - h(x_{k+1}) + \frac{1}{2\lambda_k} \|x_{k+1} - x_k\|^2. \tag{2.3}$$

This lead to

$$f(x_k) \geq f(x_{k+1}) + g(c(x_k) + \nabla c(x_k)(x_{k+1} - x_k)) - g(c(x_{k+1})) + \frac{1}{2\lambda_k} \|x_{k+1} - x_k\|^2. \tag{2.4}$$

Taking into the fact that

$$\|g(c(x_k) + \nabla c(x_k)(x_{k+1} - x_k)) - g(c(x_{k+1}))\| \leq L \| \big(c(x_k) + \nabla c(x_k)(x_{k+1} - x_k)\big) - c(x_{k+1})\|$$

together with

$$\begin{aligned}
&\| \big(c(x_k) + \nabla c(x_k)(x_{k+1} - x_k)\big) - c(x_{k+1})\| \\
&= \| \int_0^1 \big(\nabla c(x_k + t(x_{k+1} - x_k)) - \nabla c(x_k)\big)(x_{k+1} - x_k) dt \| \\
&\leq \int_0^1 \| \big(\nabla c(x_k + t(x_{k+1} - x_k)) - \nabla c(x_k)\big) \| \|x_{k+1} - x_k\| dt \\
&\leq \beta \|x_{k+1} - x_k\|^2 \int_0^1 t dt = \frac{\beta}{2} \|x_{k+1} - x_k\|^2,
\end{aligned}$$

we finally deduce

$$f(x_k) \geq f(x_{k+1}) + \frac{1}{2}(\lambda_k^{-1} - L\beta) \|x_{k+1} - x_k\|^2. \tag{2.5}$$

Therefore, the algorithm provides a monotonically decreasing sequence $(f(x_k))_{k\in\mathbb{N}}$ provided that $0<\lambda_k\le\frac{1}{L\beta}$, since $x_{k+1}\ne x_k$. □

The following result is a consequence of Theorem 2.1.

Corollary 2.1. *Consider the sequence $(x_k)_{k\in\mathbb{N}}$ generated by Algorithm (1.4). Then sequence $(f(x_k))_{k\in\mathbb{N}}$ is convergent. Furthermore, if f is a continuous function and $(x_k)_{k\in\mathbb{N}}$ is bounded, then $\lim_{k\to+\infty}f(x_k)=f(\bar{x})$ for some cluster point $\bar{x}$ of $(x_k)_{k\in\mathbb{N}}$.*

The following proposition is useful to prove the convergence theorem.

Proposition 2.1. *Let $(x_k)_{k\in\mathbb{N}}$ be generated by Algorithm (1.4), then* $\lim_{k\to+\infty}\|x_{k+1}-x_k\|=0$.

Proof. From (2.5), we have that $\sum_{k=0}^{n-1}\frac{1-\lambda_kL\beta}{2\lambda_k}\|x_{k+1}-x_k\|^2\le f(x_0)-f(x_k)$. Since f is bounded from below and $\liminf_{k\to+\infty}\lambda_k>0$, we obtain $\sum_{k=0}^{+\infty}\frac{1-\lambda_kL\beta}{2\lambda_k}\|x_{k+1}-x_k\|^2<+\infty$ and thus the sequence $(x_k)_{k\in\mathbb{N}}$ is asymptotically regular, namely $\lim_{k\to+\infty}\|x_{k+1}-x_k\|=0$. □

Theorem 2.2. *Suppose that $(x_k)_{k\in\mathbb{N}}$ is bounded. Then every cluster-point $\bar{x}$ of $(x_k)_{k\in\mathbb{N}}$ is a critical point of the function f, namely*

$$\partial g(\bar{x})\cap\nabla c(\bar{x})^*\partial h(c(\bar{x}))\ne\emptyset.$$

Proof. Note that from the first part of (2.1) and the convexity of g, if $(x_k)_{k\in\mathbb{N}}$ is bounded, then $(w_k)_{k\in\mathbb{N}}$ is also bounded. Let $\bar{x}$ and $\bar{w}$ be cluster points of the sequences $(x_k)_{k\in\mathbb{N}}$ and $(w_k)_{k\in\mathbb{N}}$, respectively. Then, there exist two subsequences (x_{k_ν}) and (w_{k_ν}) converging respectively to $\bar{x}$ and $\bar{w}$. Since g is convex and lower semicontinuous, it follows from the first part of (2.1) and maximal monotonicity of its subdifferential that $\bar{w}\in\partial h(\bar{x})$ when $\nu\to+\infty$.

Now, we claim that $\bar{w}\in\nabla c(\bar{x})\partial g(c(\bar{x}))$. The optimality condition of the minimization problem in algorithm (1.4) (i.e., the second part of (2.1)) reads as

$$w_{k_\nu}=\frac{x_{k_\nu+1}-x_{k_\nu}}{\lambda_{k_\nu}}+\nabla c(x_{k_\nu})^*z_{k_\nu}\ \text{ for some }\ z_{k_\nu}\in\partial g\big(c(x_{k_\nu})+\nabla c(x_{k_\nu})(x_{k_\nu+1}-x_{k_\nu})\big).$$

Taking into account the C^1-smoothness of c, convex and lower semicontinuity of g together with asymptotical regularity of $(x_k)_{k\in\mathbb{N}}$, we obtain at the limit $\bar{w}=\nabla c(\bar{x})^*\bar{z}$ with $\bar{z}\in\partial g(c(\bar{x}))$. These lead finally to

$$\partial g(\bar{x})\cap\nabla c(\bar{x})^*\partial g(c(\bar{x}))\ne\emptyset,$$

which means that $\bar{x}$ is a critical point of f. □

3. P-L Method for DC Composite Minimization (1.5)

Throughout this section, we make the following assumptions on the functional components of the problem: $g:\mathbb{R}^d\to\mathbb{R}\cup\{+\infty\}$ is a proper, closed, and convex function, $h:\mathbb{R}^m\to\mathbb{R}\cup\{+\infty\}$ is a convex and L-Lipschitz continuous function

$$\|h(x)-h(y)\|\le L\|x-y\|,\ \ \forall x,y\in\mathbb{R}^m,$$

and $c:\mathbb{R}^d\to\mathbb{R}^m$ is a C^1-smooth mapping with a β-Lipschitz continuous Jacobian map

$$\|\nabla c(x)-\nabla c(y)\|\le\beta\|x-y\|,\ \ \forall x,y\in\mathbb{R}^d.$$

The proposed algorithm look for critical points of $f:=g-h\circ c$ by constructing two sequences $(x_k)_{k\in\mathbb{N}}$ and $(w_k)_{k\in\mathbb{N}}$ by the following rules

$$\begin{cases} w_k \in \nabla c(x_k)^* \partial h(c(x_k)); \\ x_{k+1} = \arg\min_{x \in \mathbb{R}^d} \left(g(x) - \langle w_k, x - x_k \rangle + \frac{1}{2\lambda_k} \|x - x_k\|^2 \right). \end{cases} \tag{3.1}$$

We begin by showing that Algorithm (1.5) is also a descent method.

Theorem 3.1. *The sequence $(x_k)_{k\in\mathbb{N}}$ generated by Algorithm (1.5) satisfies*

(i) either the algorithm stops at a critical point.

(ii) or f decreases strictly, i.e. $f(x_{k+1}) < f(x_k)$ provided that $0 < \lambda_k < \frac{1}{L\beta}$ for all $k \in \mathbb{N}$.

Proof. In view of (1.5), the subdifferential chain rule yields

$$w_k \in \nabla c(x_k)^* \partial h(c(x_k)) \text{ and } w_k + \frac{x_k - x_{k+1}}{\lambda_k} \in \partial g(x_{k+1}),$$

respectively. If $x_{k+1} = x_k$, then $w_k \in \partial g(x_k) \cap \nabla c(x_k)^* \partial h(c(x_k))$, which means that x_k is a critical point of f.

Now, we suppose $x_{k+1} \neq x_k$. Since $w_k = \nabla c(x_k)^* v_k$ for some $v_k \in \partial h(c(x_k))$, by the definition of subdifferential, we can write

$$h(c(x_{k+1})) \geq h(c(x_k)) + \langle v_k, c(x_{k+1}) - c(x_k) \rangle. \tag{3.2}$$

On the other hand, as x_{k+1} minimizes $g(\cdot) - \langle w_k, \cdot - x_k \rangle + \frac{1}{2\lambda_k} \| \cdot - x\|^2$, we also have

$$g(x_k) \geq g(x_{k+1}) - \langle w_k, x_{k+1} - x_k \rangle + \frac{1}{2\lambda_k} \|x_{k+1} - x_k\|^2. \tag{3.3}$$

Combining the last inequalities, we obtain

$$\begin{aligned} f(x_k) &= g(x_k) - h(c(x_k)) \\ &\geq f(x_{k+1}) + \langle v_k, c(x_{k+1}) - c(x_k) \rangle - \langle w_k, x_{k+1} - x_k \rangle + \frac{1}{2\lambda_k} \|x_{k+1} - x_k\|^2. \end{aligned} \tag{3.4}$$

In view of

$$\begin{aligned} \langle v_k, c(x_{k+1}) - c(x_k) \rangle - \langle w_k, x_{k+1} - x_k \rangle &= \langle v_k, c(x_{k+1}) - \big(c(x_k) + \nabla c(x_k)(x_{k+1} - x_k)\big) \rangle \\ &\geq -\tfrac{L\beta}{2} \|x_{k+1} - x_k\|^2, \end{aligned}$$

this lead again to

$$f(x_k) \geq f(x_{k+1}) + \frac{1}{2}(\lambda_k^{-1} - L\beta) \|x_{k+1} - x_k\|^2. \tag{3.5}$$

Therefore, this algorithm provides also a monotonically decreasing sequence $(f(x_k))_{k\in\mathbb{N}}$ since $0 < \lambda_k < \frac{1}{L\beta}$ and $x_{k+1} \neq x_k$. □

The following results are consequences of Theorem 3.1.

Remark 3.1. Consider the sequence $(x_k)_{k\in\mathbb{N}}$ generated by Algorithm (1.5), the sequence $(f(x_k))_{k\in\mathbb{N}}$ is convergent. Moreover, taking into account (3.5) and summing along the indices $i = 0, \cdots, n$, we obtain that the sequence $(x_k)_{k\in\mathbb{N}}$ is again asymptotically regular, namely $\lim_{k\to+\infty} \|x_{k+1} - x_k\| = 0$. Furthermore, if f is a continuous function and $(x_k)_{k\in\mathbb{N}}$ is bounded, then $\lim_{k\to+\infty} f(x_k) = f(\bar{x})$ for some cluster point $\bar{x}$ of $(x_k)_{k\in\mathbb{N}}$.

Theorem 3.2. *Suppose that $(x_k)_{k\in\mathbb{N}}$ is bounded. Then every cluster-point $\bar{x}$ of $(x_k)_{k\in\mathbb{N}}$ is a critical point of the function f, namely*

$$\partial g(\bar{x}) \cap \nabla c(\bar{x})^* \partial h(c(\bar{x})) \neq \emptyset.$$

Proof. From (1.5), convexity of h together with C^1 smoothness of c, if $(x_k)_{k\in\mathbb{N}}$ is bounded, then $(w_k)_{k\in\mathbb{N}}$ is bounded too. Let $\bar{x}$ and $\bar{w}$ be cluster points of the sequences $(x_k)_{k\in\mathbb{N}}$ and $(w_k)_{k\in\mathbb{N}}$, respectively. Then, there exist two subsequences (x_{k_ν}) and (w_{k_ν}) converging respectively to $\bar{x}$ and $\bar{w}$. Since h is convex and lower semicontinuous and c is C^1 smooth, it follows from (1.5) and maximal monotonicity of subdifferential of h that $\bar{w}\in\nabla c(\bar{x})^*\partial h(c(\bar{x}))$ when $\nu\to+\infty$. Now, we clam that $\bar{w}\in\partial g(\bar{x})$. The optimality condition of the minimization problem in algorithm (1.5) reads as

$$w_{k_\nu}+\frac{x_{k_\nu}-x_{k_\nu+1}}{\lambda_{k_\nu}}\in\partial g(x_{k_\nu+1}).$$

Taking into account the convex and lower semicontinuity of g and thus maximal monotonicity of its subdifferential, together with asymptotical regularity of $(x_k)_{k\in\mathbb{N}}$, we obtain at the limit $\nu\to+\infty$ that $\bar{w}\in\partial g(\bar{x})$. These lead finally to

$$\partial g(\bar{x})\cap\nabla c(\bar{x})^*\partial h(c(\bar{x}))\neq\emptyset,$$

which means again that $\bar{x}$ is a critical point of f. □

Remark 3.2. It is worth mentioning that the results presented are still true if we consider approximate versions obtained by replacing the exact subdifferential by the approximate one such as in [6] and inexact versions handling approximate solutions of subproblems as in [7]. A generalized version of (1.4) may be proposed by using quasi distance as regularization term which preserves the nice properties of convexity, continuity and coercivity of the Euclidean norm [13]. This kind of generalized method is more appropriate and is a nice tool to model the dynamics of human behaviors in the context of the variational rationality approach [14]. It is also well known that inertial versions of prox-(linear) methods automatically accelerate the convergence. We hope that this paper may stimulate further research involving proximal linearized algorithms for composite DC functions and these concepts. Also, one can replace g in (1.1) (resp. h in (1.2)) by a smooth approximation (for instance, its Moreau envelope which is convex, Lipschitz and enjoys the key properties to be continuously differentiable, with Lipschitz gradient and which is a good approximation of the associated function) and then minimize the resulting composite function by prox-linear methods suggested in this paper.

To conclude, we foresee further progress in this topics in the near future beginning with global and Linear convergences of the sequences (1.4) and (1.5) under suitable additional assumptions such as Kurdyka-Lojasiewicz property, which is satisfied by a wide variety of functions such as proper closed semi algebraic functions, and plays an important role in the convergence analysis of many first-order methods. Another focus of research regarding the composite problem $\min_x h(c(x))+g(x)$, introduced in [15], is to investigate the following general composite envelope

$$\varphi_\lambda(x):=\min_{u\in\mathbb{R}^d}\left(h(c(x)+\nabla c(x)(u-x))+g(u)+\frac{1}{2\lambda}\|u-x\|^2\right)$$

in line with the composite Moreau envelope studied in the excellent paper [16]. Relying on

$$S_\lambda(x):=\arg\min_{u\in\mathbb{R}^d}\left(h(c(x)+\nabla c(x)(u-x))+g(u)+\frac{1}{2\lambda}\|u-x\|^2\right)$$

and

$$R_\lambda(x)=\frac{x-S_\lambda(x)}{\lambda}$$

one can devise Newton-like algorithms with fast asymptotic convergence rates by using the properties of this more general envelope that also reads as

$$\varphi_\lambda(x) = h\big(c(x) - \lambda\langle \nabla c(x), R_\lambda(x)\rangle\big) + g(S_\lambda(x)) + \frac{\lambda}{2}\|R_\lambda(x)\|^2.$$

If h is the identity function, then φ_λ become exactly the composite Moreau envelope introduced in [16] and if h is the null function, then φ_λ reduces to the classical Moreau envelope. It would be also interesting to develop the nice case $g = \delta_C$ and $h = \delta_Q$, i.e., the split feasibility problems governed by closed sets C and Q with a nonlinear smooth model c and related extended CQ-algorithm versions.

REFERENCES

[1] Y. Chen, D. Ge, M. Wang, Z. Wang, Y. Ye, H. Yin, Strong NP-hardness for sparse optimization with concave penalty functions. Proceedings of the 34th International Conference on Machine Learning, PMLR 70 2017, 740-74.

[2] J. Fan, R. Li, Variable selection via nonconvex penalized likelihood and its oracle properties, J. Amer. Stat. Assoc. 96 (2001), 1348–1361.

[3] Z. Jin, D.Y. Gao, On DC optimization problems, In D. Y. Go et al. (eds) vol. 37, pp. 203-221, Springer, 2017.

[4] P.D. Tao, L.T.H. An, Convex analysis approach to dc programming: Theory, algorithms and applications, Acta Math. Vietnam. 22 (1997), 289-355.

[5] W. Sun, R.J.B. Sampaio, M.A.B. Candido, Proximal point algorithm for minimization of DC functions, J. Comput. Math. 21 (2003), 451-462.

[6] A. Moudafi, P.-E. Maingé, On the convergence of an approximate proximal method for d.c. functions. J. Comput. Math. 24 (2006), 475-480.

[7] J. C. O. Souza, P. R. Oliveira, A. Soubeyran, Global convergence of a proximal linearized algorithm for difference of convex functions, Optim. Lett. 10 (2016), 1529-1539.

[8] P. Yin, Y. Lou, Q. He, J. Xin, Minimization of l_{1-2} for compressed sensing, SIAM J. Sci. Comput. 37 (2015), 536-563.

[9] A. Moudafi, H.K. Xu, A DC regularization of split minimization problems, Appl. Anal. Optim. 2 (2018), 285-297.

[10] A.S. Lewis, S.J. Wright, A proximal method for composite minimization, Math. Program. 158 (2016), 501–546.

[11] J.V. Burke, T. Hoheisel, Q.V. Nguyen, A study of convex convex-composite functions via infimal convolution with applications, Math. Oper. Res. 46 (2021), 1324-1348.

[12] B. S. Mordukhovich, Variational Analysis and Generalized Differentiation. I, Volume 330 of Grundlehren der MathematischenWissenschaften, Fundamental Principles of Mathematical Sciences, Springer, Berlin 2006.

[13] J.X. Cruz Neto, P.R. Oliveira, A. Soubeyran, J.C. O. Souza, A generalized proximal linearized algorithm for DC functions with application to the optimal size of the firm problem, Ann. Oper. Res. 289 (2002), 313-339.

[14] G.C. Bento, A. Soubeyran, Generalized inexact proximal algorithms: Routines formation with resistance to change, following worthwhile changes, J. Optim. Theory Appl. 166 (2015), 172-187.

[15] D. Drusvyatskiy, C. Paquette, C, Efficiency of minimizing compositions of convex functions and smooth maps, Math. Program. 178 (2019), 503–558.

[16] P. Patrinos, A. Bemporad, Proximal Newton Methods for Convex Composite Optimization, 52nd IEEE Conference on Decision and Control, Florence, 2013.