Journal of Computational Mathematics, Vol.24, No.4, 2006, 475–480.

ON THE CONVERGENCE OF AN APPROXIMATE PROXIMAL METHOD FOR DC FUNCTIONS *

A. Moudafi   P-E. Maingé
(*Université Antilles Guyane, GRIMAAG, Département Scientifique interfacultaires, 97200 Schelcher, Martinique, France*)

Abstract

In this paper we prove the convergence of the approximate proximal method for DC functions proposed by Sun et al [6]. Our analysis also permits to treat the exact method. We then propose an interesting result in the case where the second component of the DC function is differentiable and provide some computational experiences which proved the efficiency of our method.

Mathematics subject classification: 49J53, 65K10, 49M37, 90C25.
Key words: DC minimization, Critical points, Subdifferentials, Proximal mappings.

1. Introduction

The proximal point algorithm was introduced by Martinet [2] for solving proper lower semicontinuous convex minimization problems and extensively studied by Rockafellar [4] in the context of monotone variational inequalities. It is well-known that if we drop the convexity assumption on the objective function several problems arise. The proximal mapping is not well-defined and in general it is not nonexpansive anymore even in arbitrary small neighbourhoods of minima. Only few research has been proposed concerning the construction of solutions in this nonconvex case, see for instance [3]. Here we focus our attention on the method recently proposed by Sun et al. [6]. To find a critical point of $f := g - h$, this method consists to increasing the function h along the direction of the subgradient and then decreasing the function f thanks to a proximal step. They proved that if the sequence generated by their algorithm is bounded, then every cluster point is a critical point of f. The aim of the paper is to provide a correct proof for the main result in the article [6]. Indeed, we propose a right and elementary proof of the convergence result for the approximate form and we provide conditions ensuring the boundedness of the generated sequences. Aftewards, by means of an epi-convergence argument, we propose an interesting result in the case where the second component of the DC function is differentiable. We then give some numerical experiments which proved the convergence of the algorithm PMDC to local solutions and showed at the same time its robustness and efficiency with respect to the algorithm DCA introduced by Pham Dinh Tao [3].

Let f be a DC function, i.e. $f = g - h$ where f and g are two convex lower semi-continuous and proper functions defined on $\mathbb{R}^n$ satisfying $dom g \cap dom h \neq \emptyset$. We consider the problem:

$$\min_{x\in\mathbb{R}^n} (g(x) - h(x)) \tag{1.1}$$

and the associated dual

$$\min_{y\in\mathbb{R}^n} (h^*(y) - g^*(y)), \tag{1.2}$$

* Received January 4, 2005; Final revised September 19, 2005.

$g^*(y) = \sup_{x\in\mathbb{R}^n}(\langle y,x\rangle - g(x)), h^*$ stand for the conjugate functions of g and h.
It is well known that

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

and that a necessary condition for $x \in dom f$ to be a local minimizer of f is $\partial h(x) \subset \partial g(x)$. As in general this necessary condition is hard to reach, we will focus our attention on finding critical points of f, namely points satisfying the relaxed condition $\partial h(x) \cap \partial g(x) \neq \emptyset$.
Throughout the paper $f := g - h : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a real DC function. We recall that a vector w is called an ε-subgradient (with $\varepsilon \geq 0$) of g at $x \in dom g$, if

$$g(u) \geq g(x) + \langle w, u-x\rangle - \varepsilon \quad \forall u \in \mathbb{R}^n. \tag{1.3}$$

The set of all ε-subgradients of g at x, denoted by $\partial_\varepsilon g(x)$, is called the approximate subdifferential of g at x which reduces to exact subdifferential when $\varepsilon = 0$. We also recall that the Moreau-Yosida approximate and the approximate proximal mapping of g are defined for $c > 0$ by

$$g_c(x) := \inf_{u\in\mathbb{R}^n}\{g(u) + \frac{1}{2c}\|u-x\|^2\} \quad \text{and} \quad prox^{\varepsilon}_{cg}(x) := (I + c\partial_\varepsilon g)^{-1}(x).$$

It is worth mentioning the richness of the class of DC functions which contains the class of lower-$\mathcal{C}^2$ functions and constitutes a minimal realistic extension of the class of convex functions. It has been successfully used in many nonconvex applications such as finance, molecular biology, mutlicommodity network, image restoration processing and seems particularly well suited to model several nonconvex industrial problems (Robotic: computer's vision, fuel mixture ...).

2. Approximate Proximal Point Algorithms

2.1 ε-Proximal Method for DC functions

The method we will study is an approximate form of the scheme by Sun et al which is based on the following equivalence:

$$x \text{ is a critical point of } g-h \Leftrightarrow x = prox_{cg}(x+cw), \forall c > 0 \text{ and } w \in \partial h(x).$$

Thanks to this fixed-point formulation, Sun et al. [6] proposed an algorithm for finding a critical point of a DC function. This method combines the proximal point algorithm with the subgradient method. Here we consider the approximate version obtained by replacing the exact subdifferential by the approximate one, since the function h (respectively g) is assumed to be convex, proper and lower semicontinuous, $\partial_0 h(x) = \partial h(x)$, for any x. Furthermore, directly from the definition it follows that $0 \leq \varepsilon_1 \leq \varepsilon_2 \Rightarrow \partial_{\varepsilon_1} h(x) \subseteq \partial_{\varepsilon_2} h(x)$. Thus $\partial_\varepsilon h(x)$ is an enlargement of $\partial h(x)$. The use of elements in $\partial_\varepsilon h$ instead of ∂h allows an extra degree of freedom which is very useful in various applications. On the other hand, setting $\varepsilon = 0$ one retrieves the exact subdifferential, so that the exact method can be also treated. For all these reasons, we consider the following scheme:

Algorithm. Proximal Method for DC Functions (PMDC)

Setp 1: Given $x_0, c_0 \geq c$. Set $k = 0$.
Step 2: Compute $w_k \in \partial_{\varepsilon_k} h(x_k)$ and set $y_k = x_k + c_k w_k$.
Step 3: Compute $x_{k+1} \in prox^{\varepsilon_k}_{c_k g}(y_k)$ (Proximal step).
If $x_{k+1} = x_k$ stop. Otherwise increase k by 1 and loop to step 2.

The following proposition contains the convergence results of PMDC.

Theorem 2.1. *Assume that $f := g - h$ is bounded from below, $c_k \geq c > 0$ for any $k \in \mathbb{N}$ and suppose that $\sum_{k=0}^{+\infty} \varepsilon_k < +\infty$, then the sequence $(f(x_k))_{k\in\mathbb{N}}$ is convergent and the sequence (x_k) is asymptotically regular in the following sense: $\lim_{k\to+\infty} c_k^{-1}\|x_k - x_{k+1}\| = 0$. Moreover, if the sequences (x_k) and (w_k) are bounded, then every cluster-point x_∞ and w_∞ of the sequences (x_k) and (w_k) are critical points of the functions $g-h$ and $h^* - g^*$, respectively.*

Proof. From the equality in step 3, we have

$$c_k^{-1}(x_k - x_{k+1}) + w_k \in \partial_{\varepsilon_k} g(x_{k+1}),$$

which implies that

$$g(x_k) \geq g(x_{k+1}) + \langle w_k + c_k^{-1}(x_k - x_{k+1}), x_k - x_{k+1}\rangle - \varepsilon_k.$$

Also, since $w_k \in \partial_{\varepsilon_k} h(x_k)$, we have

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

Adding the last inequalities, we obtain

$$f(x_{k+1}) \leq f(x_k) - c_k^{-1}\|x_k - x_{k+1}\|^2 + 2\varepsilon_k, \tag{2.1}$$

which in turn implies that

$$f(x_{k+1}) \leq f(x_k) + 2\varepsilon_k, \tag{2.2}$$

hence

$$f(x_k) \leq f(x_0) + 2\sum_{i=0}^{k-1} \varepsilon_i$$

which together with the assumption $\sum_{k=0}^{+\infty} \varepsilon_k < +\infty$ implies that the sequence $(f(x_k))$ is bounded from above. On the other hand, f is assumed bounded from below by assumption, hence $(f(x_k))$ is bounded and thus has at least one cluster point. In fact the whole sequence converges. Indeed, let $(f(x_k))$ admits two cluster points f_1^∞ and f_2^∞ such that $f_1^\infty < f_2^\infty$ and let $(f(x_{\nu_1}))$ and $(f(x_{\nu_2}))$ be two subsequences converging to f_1^∞ and f_2^∞, respectively. Set $\tilde{\varepsilon} = \frac{f_2^\infty - f_1^\infty}{4}$. Then, there exist two integers $\tilde{\nu}_1, \tilde{\nu}_2$ such that

$$f(x_{\nu_1}) < f_1^\infty + \tilde{\varepsilon} \quad \forall \nu_1 > \tilde{\nu}_1 \quad \text{and} \quad f(x_{\nu_2}) > f_2^\infty - \tilde{\varepsilon} \quad \forall \nu_2 > \tilde{\nu}_2.$$

Moreover, since $\sum_{k=0}^{+\infty} \varepsilon_k < +\infty$, there exists $\tilde{\nu}_3 \in I\!N$ such that $\sum_{k=\tilde{\nu}_3}^{+\infty} \varepsilon_k < \tilde{\varepsilon}$. Now, choose an integer $\bar{\nu}_1$ such that $\bar{\nu}_1 \geq \max\{\tilde{\nu}_1, \tilde{\nu}_2, \tilde{\nu}_3\}$. Then, by virtue of (2.5), we can write

$$f(x_{\nu_1}) < f(x_{\bar{\nu}_1}) + 2\sum_{i=\bar{\nu}_1}^{\nu_1 - 1} \varepsilon_i < f_1^\infty + 3\tilde{\varepsilon} = f_2^\infty - \tilde{\varepsilon} \;\forall \nu_1 > \bar{\nu}_1.$$

This is a contradiction, and hence $(f(x_k))$ has at the most one cluster point.
With this result in hand, we infer from (2.4) that

$$\lim_{k \to +\infty} c_k^{-1}\|x_k - x_{k+1}\|^2 = 0,$$

which in turn, since $c_k \geq c > 0$, implies that $\lim\limits_{k \to +\infty} c_k^{-1}\|x_k - x_{k+1}\| = 0$.
Now let's consider two subsequences (x_{k_ν}) and (w_{k_ν}) of (x_k) and (w_k) (we will use the same notation for the index even if it needs extracting other subsequences) converging respectively to x_∞ and w_∞. By passing to the limit in the following relations

$$c_{k_\nu}^{-1}(x_{k_\nu} - x_{k_\nu + 1}) + w_{k_\nu} \in \partial_{\varepsilon_{k_\nu}} g(x_{k_\nu + 1}) \text{ and } w_{k_\nu} \in \partial_{\varepsilon_{k_\nu}} h(x_{k_\nu})$$

and taking into account the fact that the multi-valued maps $\partial_{(\cdot)} f(\cdot)$ and $\partial_{(\cdot)} h(\cdot)$ are closed on $I\!R_+ \times I\!R^n$, we obtain

$$w_\infty \in \partial g(x_\infty) \quad \text{and} \quad w_\infty \in \partial h(x_\infty),$$

from which we infer that $\partial g(x_\infty) \cap \partial h(x_\infty) \neq \emptyset$, in other words that x_∞ is a critical point of $g - h$ and by duality that w_∞ is a critical point of $h^* - g^*$.
Remark 2.1. It should be noticed that the assumption that (x_k) is bounded holds true if, for example, the function f is inf-compact, namely for each λ the $\{x \in I\!R^n, f(x) \leq \lambda\}$ is a compact set. Indeed, $(x_k) \subset \{x;\; f(x) \leq f(x_0) + \sum_{k=0}^{+\infty} \varepsilon_k\}$. Furthermore, it is a well known fact that the sequence (w_k) is bounded if, for instance, $x_k \in int(domh)$.

Example. We consider the problem of maximizing a convex lower semi-continuous function h on a closed convex set C of $I\!R^n$. This problem has received a great attention and can be rewritten as a DC programming problem, namely

$$-\min_{x\in I\!R^n}\{\delta_C(x)-h(x)\}, \tag{2.3}$$

where δ_C stands for the indicator function of C.
In this context Algorithm PMDC takes the following form:

$$x_{k+1}=proj_C^{\varepsilon_k}(x_k+c_k w_k), \quad \text{where } w_k\in\partial h(x_k),$$

more precisely x_{k+1} solves the approximate variational inequality:

$$\langle c_k^{-1}(x_k-x_{k+1})+w_k, u-x_{k+1}\rangle \le \varepsilon_k \quad \forall u\in C.$$

2.2 An Interesting Special Case

Let us consider now the case where the function h is differentiable. In this context the exact algorithm (i.e with $\varepsilon_k=0$ for all $k\in I\!N$) reduces to

$$x_{k+1}=prox_{c_k g}(x_k+c_k\nabla h(x_k)),$$

which looks like the prox-gradient algorithm and we have the following result:
Proposition 2.1. *If the exact algorithm converges to a point x^∞ that admits a neighborhood, $\mathcal{V}(x_\infty)$, in which the function f is convex and the function h is differentiable with a Lipschitz gradient and if we assume that the sequence (c_k) is bounded, then x_∞ is a local minimizer of f.*

Proof. From theorem 2.1, we have that $\lim_{k\to+\infty}\|x_k-x_{k+1}\|=0$ which together with the fact that ∇h is Lipschitz continuous on $\mathcal{V}(x_\infty)$ yield, for k sufficiently large, that

$$\lim_{k\to+\infty}\|\nabla h(x_k)-\nabla h(x_{k+1})\|=0.$$

We also have

$$c_k^{-1}(x_k-x_{k+1})+\nabla h(x_k)\in\partial g(x_{k+1}),$$

which can be rewritten as

$$c_k^{-1}(x_k-x_{k+1})+\nabla h(x_k)-\nabla h(x_{k+1})\in\partial g((x_{k+1}))-\nabla h(x_{k+1})\subset\partial(g-h)(x_{k+1}),$$

in other words, for all $x\in\mathcal{V}(x_\infty)$

$$f(x)\ge f(x_{k+1})+\langle c_k^{-1}(x_k-x_{k+1})+\nabla h(x_k)-\nabla h(x_{k+1}), x-x_{k+1}\rangle. \tag{2.4}$$

According to what we saw before, we have

$$\lim_{k\to+\infty}(c_k^{-1}(x_k-x_{k+1})+\nabla h(x_k)-\nabla h(x_{k+1}))=0,$$

and taking into account the fact that f is lower semi-continuous which ensures that

$$\liminf_{k\to+\infty}f(x_k)\ge f(x_\infty),$$

we obtain the required result by passing to the lower limit in relation (2.7).
Remark 2.2. We would like to point out that the assumption on h is automatically satisfied if we replace h by its Moreau-Yosida approximate which is $\mathcal{C}^{1,1}$ and since $h_\lambda\uparrow h$ as $\lambda\downarrow 0$, the function $g-h_\lambda\downarrow cl(g-h)$ as $\lambda\downarrow 0$ and thus approximates the function $g-h$ in the sense of epi-convergence if h is continuous.

---

[1] Indeed, since $g_\lambda\uparrow g$ as $\lambda\downarrow 0$, by an epi-convergence argument, see for example [5], we have $\partial g-\nabla h\subset \liminf_{\lambda\to 0}(\nabla g_\lambda-\nabla h)=\liminf_{\lambda\to 0}\nabla(g_\lambda-h)\subset\limsup_{\lambda\to 0}\nabla(g_\lambda-h)\subset\partial(g-h)$

2.3 Numerical Experiments

This part is intended to illustrate, by means of some examples, the convergence of the algorithm PMDC to local solutions and to show at the same time its robustness and efficiency with respect to DCA. It is worth mentioning that the source codes of all numerical experiences are written in Fortran.

First, we focus our attention on problem (2.6) with a polyhedral set C, it is worth mentioning that in this case the convergence of the algorithm is finite.

Example 1. We are concerned here with the following problem

$$\max_{(x,y)\in C} h(x,y),$$

where $h(x,y) = |x| + |y|$ and $C = \{(u,v) \in \mathbb{R}^2; \quad |u| \leq 1, |v| \leq 1\}$. Let us note that h admits over C four local maxima at $(1,1)$, $(-1,1)$, $(-1,-1)$ and $(1,-1)$.

Table 1. Solutions found by PMDC with respect to different initial points

	$c_k = 1$		$c_k = 0.5$	
(x_0, y_0)	iter	(x^*, y^*)	iter	(x^*, y^*)
$(-7,7)$	1	$(-1,1)$	1	(-1,1)
$(-1,0.5)$	1	$(-1,1)$	1	(-1,1)
$(7,7)$	1	$(1,1)$	1	(1,1)
$(0.5,-0.5)$	2	$(1,1)$	3	(1,1)
$(-0.5,0.5)$	1	$(-1,1)$	1	(-1,1)
$(0.1,0.1)$	1	$(1,1)$	2	(1,1)
$(0.1,-0.1)$	2	$(1,1)$	3	(1,1)
$(-0.1,0.1)$	1	$(-1,1)$	2	(-1,1)
$(-0.1,-0.1)$	1	$(-1,-1)$	2	(-1,-1)

Now, let us take up an example solved in [9] by the algorithm DCA. Table 2.3 shows clearly the robustness and efficiency of PMDC with respect to DCA.

Example 2. The function f is defined by

$$f(x,y) = \frac{1}{2}\begin{pmatrix} -83.75 & 28.34 \\ 28.34 & -48.24 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + (17.72 \quad 15.22)\begin{pmatrix} x \\ y \end{pmatrix},$$

and the problem under consideration is

$$\min_{(x,y)\in C} f(x,y) \quad \text{with} \quad C = \{(u,v) \in \mathbb{R}^2; \quad 0 \leq u \leq 1, \quad 0 \leq v \leq 1\},$$

which amounts to

$$\min_{(x,y)\in\mathbb{R}^2} \{\delta_C(x,y) - (-f(x,y))\}.$$

Over the square C, f admits four local maxima at $(0,0)$, $(0,1)$, $(1,1)$ and $(1,0)$.

Table 2. Solutions found by PMDC and DCA / to the same initial points

	PMDC/$c_k = 1$		PMDC/$c_k = 0.1$		DCA	
(x_0, y_0)	iter	(x^*, y^*)	iter	(x^*, y^*)	iter	(x^*, y^*)
$(\mathbf{0.24}, \mathbf{0.32})$	**1**	$(\mathbf{0}, \mathbf{0})$	**1**	$(\mathbf{0}, \mathbf{0})$	**9**	$(\mathbf{0}, \mathbf{1})$
$(\mathbf{0.24}, \mathbf{0.31})$	**1**	$(\mathbf{0}, \mathbf{0})$	**1**	$(\mathbf{0}, \mathbf{0})$	**8**	$(\mathbf{0}, \mathbf{0})$
$(0.32, 0.25)$	1	$(1,0)$	2	$(1,0)$	3	$(1,0)$
$(0.25, 0.75)$	1	$(0,1)$	1	$(0,1)$	2	$(0,1)$

Last but not least, we give a Maximum Eigenvalue numerical application.

Example 3. The maximum eigenvalue of a symmetric positive definite matrix Q can be reached by solving problem (2.6) with

$$h(x) = \langle x, Qx \rangle \quad \text{and} \quad C = \{x \in \mathbb{R}^n; \|x\| \leq 1\}.$$

Let us define the matrix Q by $Q = \begin{pmatrix} 3/2 & 1/2 \\ 1/2 & 3/2 \end{pmatrix}$ and note that its eigenvalues, which are local critical points of (2.6), are given by $sol_1 = 1$ and $sol_2 = 2$.

Table 3. Results given by PMDC with respect to different initial points

iter	$\mathrm{sol}_{(0.1,0.5)}$	$\mathrm{sol}_{(-0,5,0.1)}$	$\mathrm{sol}_{(0.5,-0.5)}$
1	0.439	1.552	1.000
2	1.862	1.774	1.000
3	1.979	1.904	1.000
4	1.992	1.963	1.000
5	1.997	1.986	1.000
6	1.999	1.995	1.000

Conclusion. This paper is aimed at studying the convergence properties of an approximate proximal point algorithm for minimizing DC functions. It is worth mentioning that if $\varepsilon_k = 0\ \forall k \in I\!N$, the approximate algorithm reduces to the exact one by Sun et al, and in this case the presented analysis provides an alternative and elementary proof of their convergence result whose proof is not self-contained and based on the composition of multi-valued maps. A full numerical evaluation of our results exceeds the scope of this paper. However, we gave some numerical experiments which proved the convergence of the algorithm PMDC to local solutions and clearly showed its robustness and efficiency.

Our further developments will be consecrated to analyzing the bundle method as suggested by one of the two referees, to developing a quasi Newton-PPA, to comparing its robustness and performance to the bundle method and DCA, and also to globalizing the PMDC by combining this method with global techniques (branch-and-bound ...) in a deeper way.

Acknowledgment. The authors thank the two anonymous referees for their careful reading of the paper and especially the one who suggested to provide numerical experiments. This paves the way for interesting theoretical questions.

References

[1] J.B. Hiriart-Urruty and C. Lemaréchal, Convex analysis and minimization algorithms. Springer-Verlag, Berlin, 1993.

[2] B. Martinet, Algorithmes pour la résolution des problèmes d'optimisation et de minmax, Thèse d'état Université de Grenoble, France (1972).

[3] D.T. Pham and E.B. Souad, Algorithms for solving a class of nonconvex optimization problems: Methods of subgradient, Fermat days 85, Mathematics for optimization, Elsevier, North Holland, (1986), 249-270.

[4] R.T. Rockafellar, Monotone operator and the proximal point algorithm, *SIAM J. Control. Opt.*, **14**:5 (1976), 877-898.

[5] R.T. Rockafellar and R. Wets, Variational Analysis, Springer, Berlin, 1988.

[6] Wen-yu Sun, Raimundo J.B. Sampaio, M.A.B. Candido, Proximal point algorithm for minimization of DC function, *Journal of computational Mathematics*, **21**:4 (2003), 451-462.

[7] J.F. Toland, Duality in non convex optimization, *J. of Mathematical Analysis and Applications*, **66** (1978), 399-415.

[8] K.C. Kiwiel, Proximity control in bundle methods for convex nondifferentiable minimization, *Mathematical Programming*, **46** (1990), 105-122.

[9] F.C. Akoa, Approches de points intérieurs et de la programmation DC en optimisation non convexes, Thèse Institut National des Sciences Appliquées de Rouen, France (2005).

[10] B. Lemaire, The proximal algorithm, New methods in optimization and their industrial uses, *Int. Ser. Numer. Math.*, **87** (1989), 73-87.