Critical point concepts: a fixed-point approach

Abdellatif Moudafi

Critical Points Concept: A Fixed Point approach

Abdellatif Moudafi
Aix Marseille Université, CNRS, LSIS UMR 7296,13397, Marseille, France
abdellatif.moudafi@univ-amu.fr

Abstract

Inspired by the notion of critical points for DC functions and given two mappings P and Q, we introduce *the concept of critical points in the fixed-point context* and design and algorithm for finding such points. Connections are then made with the DC optimization case. We show that the proposed Algorithmic approach coincides with the celebrate DCA introduced by Pham Dinh Tao. The case of maximal monotone operators is also stated and investigated via a characterization of their associated resolvents.

AMS Subject Classification: Primary, 49J53, 65K10; Secondary, 49M37, 90C25.

Key words: Critical point, fixed-point problem, DC optimization, DCA, proximity, resolvent, convex function, conjugate function, monotone operators, duality.

1 Introduction and preliminaries

DC Programming and DCA were introduced by Pham Dinh Tao in 1985 in their preliminary form and extensively developed by Le Thi Hoai An and Pham Dinh Tao since 1994 to become now classical, see for example [1] and references therein. Their original key idea relies on the DC structure of the objective function in nonconvex programs which are explored and exploited in a deep and suitable way.

A DC program is of the form :

$$(P) \quad \inf\{f(x) := g(x) - h(x) : x \in X\},$$

with g, h two proper, convex and lower semicontinuous functions.
Such a function f is called a DC function. DC programming provided an extension of convex programming, not too large to still allow using convex analysis and convex optimization tools but sufficiently wide to cover most real world nonconvex problems. The convexity of the two DC components g and h of the objective function has been used to develop appropriate tools from both theoretical and algorithmic viewpoints. The other support of this approach is the DC duality, which has been first studied by J.F. Toland [9]. Many authors were interested in solution algorithms and in local and global optimality conditions, relationships between local and global solutions to primal DC programs and their dual, namely

$$(D) \quad \inf\{h^*(y) - g^*(y) : y \in Y\},$$

where Y is the dual space of X, which can be identified with X itself, and $g^*(\cdot) := \sup_{z\in X}\{\langle z, \cdot\rangle - g(z)\}, h^*(\cdot) := \sup_{z\in X}\{\langle z, \cdot\rangle - h(z)\}$ denote the conjugate functions of g and h, respectively.

The complexity of DC programs resides on the lack of practical global optimality conditions. So the following necessary local optimality conditions for (P) were developed:

$$\partial g(\bar{x}) \cap \partial h(\bar{y}) \neq \emptyset, \tag{1}$$

such a point $\bar{x}$ is called a critical point of $g - h$ or for (P), and

$$\partial g(\bar{x}) \subset \partial h(\bar{y}). \tag{2}$$

The condition (2) is also sufficient for many important classes of DC programs, [1] and references therein.

Based on local optimality conditions and duality in DC programming, the DCA introduced by Pham Dinh Tao generates two sequences $\{x_k\}$ and $\{y_k\}$, candidates to be optimal solutions of the primal and dual programs respectively and $\{x_k\}$ (resp. $\{y_k\}$) converges to a solution $\bar{x}$ of (P) (resp. a solution $\bar{y}$ of (D)) verifying local optimality conditions and such that $\bar{x} \in \partial g^*(\bar{y}); \bar{y} \in \partial h(\bar{x})$.
These two sequences $\{x_k\}$ and $\{y_k\}$ are determined by the next scheme (DCA):

$$y_k \in \partial h(x_k); x_{k+1} \in \partial g^*(y_k).$$

We are interested, in this short paper, in developing similar ideas in the fixed-point case.

2 Critical point notion and algorithmic aspect

To begin with, let us define critical points of two mappings.

Definition 2.1 *Let P and Q be two given mappings defined on a Hilbert space H, a point $\bar{x}$ is a critical point of (P, Q) if and only if*

$$\exists \bar{y} \text{ such that } P(\bar{x} + \bar{y}) = \bar{x} = Q(\bar{x} + \bar{y}). \tag{3}$$

This is equivalent, using the corresponding dual mappings, to

$$\exists \bar{y} \text{ such that } (I - P)(\bar{x} + \bar{y}) = \bar{y} = (I - Q)(\bar{x} + \bar{y}). \tag{4}$$

Observe that when one of the two mappings is equal to the identity, for example $Q = I$, relation (3) reduces to the following classical fixed point problem:

$$\text{find } \bar{x} \text{ such that } \bar{x} = P(\bar{x}).$$

Now, let us made the connection with DC programming.

Proposition 2.1 *Let h and g be two given proper, convex and lower semicontinuous functions on a Hilbert space H, then $\bar{x}$ is a critical point of their proximal mappings if, and only if, $\bar{x}$ is a critical point of $g - h$.*

Proof: Indeed, it is clear that

$$prox_h(\bar{x}+\bar{y}) = \bar{x} \Leftrightarrow \bar{y} \in \partial h(\bar{x})$$

and

$$prox_g(\bar{x}+\bar{y}) = \bar{x} \Leftrightarrow \bar{y} \in \partial g(\bar{x}).$$

From which the desired result follows.

Recall that the proximity of a function f is defined by

$$prox_f(x) = (I+\partial f)^{-1}(x) = arg\min_z\{f(z) + 1/2\|x-z\|^2\},$$

$\partial f(x) := \{x^* \in Y; f(z) \geq f(x) + \langle x^*, z-x\rangle \ \forall z \in X\}$ denotes the subdifferential of f at x. □

In view of the definition of critical points, we are now ready to introduce our Algorithm:

Algorithm: Given x_k, compute y_k by

$$y_k = (I-P)(y_k + x_k); \tag{5}$$

then, compute x_{k+1} by

$$x_{k+1} = Q(y_k + x_{k+1}). \tag{6}$$

Remark 2.1 *In the special case of the DC Optimization, the proposed Algorithm coincides with the celebrate DCA introduced by Pham Dinh Tao. Indeed, by taking $P = prox_h$ and $Q = prox_g$.*

$$y_k = (I - prox_h)(y_k + x_k) \Leftrightarrow y_k = prox_{h^*}(y_k + x_k) \Leftrightarrow x_k \in \partial h^*(y_k) \Leftrightarrow y_k \in \partial h(x_k);$$

and

$$x_{k+1} = prox_g(y_k + x_{k+1}) \Leftrightarrow y_k \in \partial g(x_{k+1}) \Leftrightarrow x_{k+1} \in \partial g^*(y_k).$$

The DC decomposition of a function is not unique. Indeed, we can write

$$g - h = (g+\varphi) - (h+\varphi),$$

φ being any convex finite function. So, we without loss of generality, we can assume that g, h to be strongly convex which amounts to saying that their subdifferential operators are strongly monotone which, in turn, ensures the fact that their proximal mappings are Banach contractions, see for example [2]-Lemma 2.3.

We propose, the following convergence result in the case where the mappings Q and $I - P$ are assumed to be Banach contractions.

Theorem 2.2 *Assume that Q and $I-P$ are Banach contractions with constants $\sigma, \tau \in]0,1[$, respectively and such that $\sigma+\tau < 1$. Then the sequence $\{x_k\}$ strongly converges to the critical point $\bar{x}$ of (P,Q) and $\{y_k\}$ strongly converges to $\bar{y}$ which solves the dual problem (4).*

Proof: Indeed, assumption on Q yields

$$\|x_{k+1} - \bar{x}\| \leq \sigma \|x_{k+1} + y_k - (\bar{x} + \bar{y})\|,$$

from which, we deduce that

$$\|x_{k+1} - \bar{x}\| \leq \frac{\sigma}{1-\sigma} \|y_k - \bar{y}\|. \tag{7}$$

Likewise, assumption on $I - P$ leads to

$$\|y_k - \bar{y}\| \leq \tau \|x_k + y_k - (\bar{x} + \bar{y})\|,$$

from which, we infer that

$$\|y_k - \bar{y}\| \leq \frac{\tau}{1-\tau} \|x_k - \bar{x}\|. \tag{8}$$

Combining inequalities (7) and (8), we obtain

$$\|x_{k+1} - \bar{x}\| \leq \frac{\sigma\tau}{(1-\sigma)(1-\tau)} \|x_k - \bar{x}\|.$$

Since, by hypothesis $\kappa := \frac{\sigma\tau}{(1-\sigma)(1-\tau)} \in]0,1[$, we obtain the strong convergence of the sequence $\{x_k\}$ to the critical point $\bar{x}$ of (P,Q) and by passing to the limit in (8), we obtain that $\{y_k\}$ strongly converges to $\bar{y}$ which solves the dual problem (4). □

Remark 2.2 *i) The hypothesis on the dual operator $I - P$ of P can be expressed in terms of a condition on P. Indeed, it is easily seen that $I - P$ is a τ-contraction if, and only if, P satisfies the following relation*

$$\langle Px - Py, x - y \rangle \geq \frac{1-\beta}{2} \|x - y\|^2 + \frac{1}{2} \|Px - Py\|^2. \tag{9}$$

Clearly, P is both $\frac{1-\beta}{2}$-strongly monotone and $\frac{1}{2}$cocoercive.
ii) We would like to emphasize that this type of hypothesis can be relaxed by assuming, for example, $I - P$ satisfying, for some $\tau \in]0,1[$ and $\kappa \in [0,1[$, the following assumption

$$\|(I-P)(x) - (I-P)(y)\| \leq \tau \|x - y\| + \kappa \|P(x) - P(y)\| \ \forall x, y, \tag{10}$$

the condition on the parameters, in Theorem 2.2 should be replaced by $(1+\tau)\kappa + \sigma < 1$. Obviously we can do so for both operators Q and $I - P$ with an appropriate condition on the parameters.

In the DC optimization case, we have the following result:

Proposition 2.3 *Let h and g be two given proper, convex and lower semicontinuous functions on a Hilbert space H. Assume that h is κ-strongly convex and g differentiable with L-Lipschitz gradient and $0 < L < \kappa$. Then the sequence $\{x_k\}$ strongly converges to the critical point $\bar{x}$ of $g - h$ and $\{y_k\}$ strongly converges to $\bar{y}$ which is a critical point of the dual problem.*

Proof: The function h is κ-strongly convex is equivalent to the fact that its subdifferential operator ∂h is κ-strongly monotone which is equivalent, in turn, to the fact that the associated proximal mapping, $prox_h$, is a Banach contraction with constant $\frac{1}{1+\kappa}$, see for example [4]-xi-Theorem 2.1. On the other hand, it is well known that the assumptions on g are equivalent to the $\frac{1}{L}$-strong convexity of its conjugate function g^* and thus to the property of Banach contraction with constant $\frac{L}{1+L}$ of its associated proximity mapping, see for instance [4]-xii-Theorem 2.1. The result follows by applying Theorem 2.2, since the condition on the parameters is fulfilled. Indeed the assumption on the parameters in Theorem 2.2, in this case, reduces to $0 < L < \kappa$. □

Remark 2.3 *We would like to emphasize that by particularizing Theorem 2.2 to DC programming, we obtain in fact a more general result than proposition 2.3. We will develop it below in the more general setting of monotone operators. We will see that assumptions on h and g can be replaced respectively by the following weaker conditions:*

$$\langle u - v, x - y \rangle \geq \frac{1-\tau^2}{2\tau^2}\|x - y\|^2 - \frac{1}{2}\|u - v\|^2 \quad \forall x, y \; \forall u \in \partial h(x), v \in \partial h(y) \tag{11}$$

and

$$\langle u - v, x - y \rangle \geq \frac{1-\sigma^2}{2\sigma^2}\|u - v\|^2 - \frac{1}{2}\|x - y\|^2 \quad \forall x, y \; \forall u \in \partial g(x), v \in \partial g(y). \tag{12}$$

3 Maximal Monotone Operators

It is well known that firmly nonexpansive mappings are important in fixed-point theory, because of the correspondance with maximal monotone operators. Let A be a multivalued maximal monotone operator, we denote by $J_A := (I + A)^{-1}$ its resolvent. Observe that when $A = \partial f$ with f a proper, convex and lower semicontinuous function, $J_{\partial f}$ is nothing but $prox_f$ the proximal mapping of f. Recall also the following resolvent identity

$$I = J_A + J_{A^{-1}},$$

A^{-1} being the inverse operator of A. Note also that if $A = \partial f$, then its inverse $A^{-1} = (\partial f)^{-1} = \partial f^*$ and the resolvent equation takes the following form to $I = prox_f + prox_{f^*}$.

Now, let us state a characterization, proposed in [4]-xiii-Theorem2.1, of the Banach contraction of the resolvent of a maximal monotone operator B in terms of a property of B:

Fact 1: The resolvent operator J_B is a Banach contraction with constant $\tau \in]0, 1[$ if, and only, if the operator B satisfies the following relation

$$\forall x, y \; \forall u \in B(x), v \in B(y) \;\; \langle u - v, x - y \rangle \geq \frac{1-\tau^2}{2\tau^2}\|x - y\|^2 - \frac{1}{2}\|u - v\|^2. \tag{13}$$

Observe that this class of operators is more large than that of strongly monotone ones. Furthermore, given a maximal monotone operator A, we deduce the following dual fact:

Fact 2: $I - J_A = J_{A^{-1}}$ is a Banach contraction with constant $\sigma \in]0,1[$ if, and only, if the operator A satisfies the following relation

$$\forall x, y\ \forall u \in A(x), v \in A(y)\ \ \langle u - v, x - y\rangle \geq \frac{1-\sigma^2}{2\sigma^2}\|u - v\|^2 - \frac{1}{2}\|x - y\|^2, \qquad (14)$$

clearly this class of operators is more large than that of cocoercive ones.

These classes were used to proving the convergence of some algorithms for variational and systems of inequalities, see for instance [10] and references therein.

Now by taking $P = J_B, Q = J_A$ with A, B two maximal monotone operator, a simple computation shows that:

Proposition 3.1

$$\bar{x} \textit{ is a critical point of } (J_A, J_B) \Leftrightarrow A(\bar{x}) \cap B(\bar{x}) \neq \emptyset,$$

which is nothing but the definition of $\bar{x}$ a critical point of $A - B$, see for example [7].

We have the following convergence result:

Proposition 3.2 *Let B and A be two given maximal monotone operators on a Hilbert space H verifying (13) and (14), respectively and assume that $\tau + \sigma < 1$. Then the sequence $\{x_k\}$ strongly converges to the critical point $\bar{x}$ of $A - B$ and $\{y_k\}$ strongly converges to $\bar{y}$ which is a critical point of the dual problem.*

Proof: The proof follows by using Fact 1, Fact 2 and by applying Theorem 2.2 with $P = J_A$ and $Q = J_B$. □

Remark 3.1 *We can obtain the definition of critical points for two equilibrium bifunction (F, G) by means of the well-known associated monotone operators or resolvents. A simple computation show that a critical point $\bar{x}$ for (F, G) can be defined as*

$$\exists \bar{y} \textit{ such that } \ \forall \xi\ \ \langle \bar{y}, \xi - \bar{x}\rangle \leq \min\big(F(\bar{x}, \xi), G(\bar{x}, \xi)\big).$$

But it is more challenging to finding other classes of mappings R bigger than that (respectively its dual) of operators satisfying for example, for some $\tau \in]0,1[$ and $\kappa \in [0,1[$, the following relation

$$\|R(x) - R(y)\| \leq \sigma\|x - y\| + \kappa\|(I - R)(x) - (I - R)(y)\|\ \forall x, y \textit{ with } 2\sigma + \kappa < 1,$$

assuring the convergence of the proposed algorithm in the context of fixed-points and also in the DC Optimization setting. Since we have at our disposal nice mathematical concepts and properties like essential strict convexity, Legendre functions, duality between strong convexity and strong smoothness with their local versions, see for instance [3] and [5]. Duality between strong convexity and strong smoothness plays a key role, for example, in applications to logarithmic regret in strongly convex repeated games and regret bound in machine learning, see for example [6] and [8] and references therein.

References

[1] Le Thi Hoai An, Pham Dinh Tao, The DC (Difference of Convex Functions) Programming and DCA Revisited with DC Models of Real World Nonconvex Optimization Problems. *Annals OR* 133(1-4) (2005), pp. 23-46.

[2] H. Attouch, A. Moudafi and H. Riahi, Quantitative stability analysis for maximal monotone operators and semi-groups of contractions, *Journal Nonlinear Analysis, Theory, Methods & Applications*, vol. 21, (1993), pp. 697-723

[3] H.H. Bauschke, J.M. Borwein, and P.L. Combettes: Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces, *Communications in Contemporary Mathematics* 3(4) (2001), pp. 615-647.

[4] H.H. Bauschke, S.M. Moffat, and X. Wang, Firmly nonexpansive mappings and maximally monotone operators: correspondence and duality, *Set-Valued and Variational Analysis* 20 (2012), pp. 131-153.

[5] R. Goebel and R.-T. Rockafellar, Local strong convexity and local Lipschitz continuity of the gradients of convex functions, *J. Convex Analysis* 15 (2008), pp. 263–270.

[6] S. Kakade, S. Shalev-Shwartz, and A. Tewari. On the duality of strong convexity and strong smoothness: Learning applications and matrix regularization. Manuscript, (2009).

[7] A. Moudafi, On critical points of the difference of two maximal monotone operators, *Afrika Matematika*, vol. 26 (3) (2013), pp. 457-463.

[8] Sh. Shalev-Shwartz and Y. Singer, Logarithmic Regret Algorithms for Strongly Convex Repeated Games,Technical Report, The Hebrew University, (2007).

[9] J. F. Toland, Duality in Nonconvex Optimization, *J. Math. Anal. Appl.*, vol. 66, Issue 2 (1978,) pp. 399-415.

[10] R. U. Verma, Generalized system for relaxed cocoercive variational inequalities and projection methods, *J. Optim. Theory Appl.*, 121(2004), pp. 203-210.