Annals of Operations Research 133, 23–46, 2005

The DC (Difference of Convex Functions) Programming and DCA Revisited with DC Models of Real World Nonconvex Optimization Problems

LE THI HOAI AN lethi@sciences.univ-metz.fr
Laboratoire de l'Informatique Théorique et Appliquée, UFR MIM, Université de Metz, Ile du Saulcy 57045 Metz, France

PHAM DINH TAO* pham@insa-rouen.fr
Laboratory of Modelling, Optimization and Operations Research, National Institute for Applied Sciences-Rouen, BP 8, F 76 131 Mont Saint Aignan Cedex, France

Abstract. The DC programming and its DC algorithm (DCA) address the problem of minimizing a function $f = g - h$ (with g, h being lower semicontinuous proper convex functions on $\mathbb{R}^n$) on the whole space. Based on local optimality conditions and DC duality, DCA was successfully applied to a lot of different and various nondifferentiable nonconvex optimization problems to which it quite often gave global solutions and proved to be more robust and more efficient than related standard methods, especially in the large scale setting. The computational efficiency of DCA suggests to us a deeper and more complete study on DC programming, using the special class of DC programs (when either g or h is polyhedral convex) called polyhedral DC programs. The DC duality is investigated in an easier way, which is more convenient to the study of optimality conditions. New practical results on local optimality are presented. We emphasize regularization techniques in DC programming in order to construct suitable equivalent DC programs to nondifferentiable nonconvex optimization problems and new significant questions which have to be answered. A deeper insight into DCA is introduced which really sheds new light on DCA and could partly explain its efficiency. Finally DC models of real world nonconvex optimization are reported.

Keywords: DC programming, DC algorithms (DCA), DC duality, local optimality conditions, global optimality conditions, polyhedral DC programming, regularization techniques

1. Introduction

In recent years there has been a very active research in nonconvex programming. A great deal of work involves *global optimization* (which is concerned with finding *global solutions* to nonconvex programs) whose main tools and solution methods are developed according to the spirit of the combinatorial optimization, but with the difference that one works in *the continuous framework* (see Horst and Tuy, 1996; Horst, Pardalos, and Thoai, 1995; Horst and Thoai, 1999). People recognize that it was Hoang Tuy who has incidentally put forward, by his pioneering paper in 1964 (Tuy, 1964), the new global

* Corresponding author. Address: LMI – INSA de Rouen, Place Emile Blondel, BP 8, 76131 Mont Saint Aignan, France.

optimization concerning convex maximization over a polyhedral convex set. It is worth noting that the works by Hoang Tuy, R. Horst, H. Benson, P.M. Pardalos, H. Konno, Le Dung Muu, Le Thi Hoai An, Nguyen Van Thoai, Phan Thien Thach, and Pham Dinh Tao are among the most important contributions to this approach. During the last decade tremendous progress has been made, especially in the computational aspect. One can now globally solve specially structured nonconvex problems arising from applications with larger dimensions, especially large scale but low rank nonconvex problems (Konno, Thach, and Tuy, 1999). However, most robust and efficient global algorithms actually do not meet the expected desire: solving real life programs in their true dimension.

Beside this combinatorial approach to global continuous optimization, the *convex analysis approach* to nonconvex programming has been much less worked. It seemed to take rise in the works of Pham Dinh Tao on the computation of bound-norms of matrices (i.e. maximizing a seminorm over the unit ball of a norm) in 1974 (Tao, 1975, 1976, 1981, 1984, 1986; and references therein). These works are extended in a natural and logical way to the DC (difference of convex functions) program:

$$\alpha = \inf\{f(x) := g(x) - h(x) \colon x \in X\}$$

where $X = \mathbb{R}^n$ is the usual Euclidean space and g, h are lower semicontinuous proper convex functions on X (i.e. $g, h \in \Gamma_0(X)$, see section 2). Indeed we would like to make an extension of convex programming, not too large to still allow using the arsenal of powerful tools in convex analysis and convex optimization but sufficiently wide to cover most real world nonconvex optimization problems. The set of DC functions $f = g - h$, being the vector subspace spanned by the convex cone $\Gamma_0(X)$, is then closed under all the operations usually considered in optimization (see, e.g., Tao and An, 1997; Tuy, 1998; and references therein). There 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 in 1978 (Toland, 1978, 1979) who generalized, in a very elegant and natural way, the just mentioned works by Pham Dinh Tao on convex maximization programming (g then is the indicator function of a nonempty closed convex set in X). In contrast with the combinatorial approach where many global algorithms have been studied, there have been a very few algorithms for solving DC programs in the convex analysis approach. Here we are interested in local and global optimality conditions, relationships between local and global solutions to primal DC programs and their dual

$$\alpha = \inf\{h^*(y) - g^*(y) \colon y \in Y\}$$

(where Y is the dual space of X, which can be identified with X itself, and g^*, h^* denote the conjugate functions of g and h, respectively) and solution algorithms.

DC algorithms (DCA), based on local optimality conditions and the duality in DC programming, have been introduced by Pham Dinh Tao in 1986 (Tao, 1986, 1988) as an extension of the aforementioned subgradient algorithms (for convex maximization programming) to DC programming. Important developments and improvements for

DCA from both theoretical and computational aspects have been completed since 1993 throughout the joint works of Le Thi Hoai An and Pham Dinh Tao (see References). Due to its local character it cannot guarantee the globality of computed solutions for general DC programs. However, we observe that with a suitable starting point it converges quite often to a global one. In practice, DCA was successfully applied to a lot of different and various nonconvex optimization problems to which it quite often gave global solutions and proved to be more robust and more efficient than related standard methods, especially in the large scale setting (see section 3 and References). To our knowledge, DCA is actually one of the rare algorithms for nonsmooth nonconvex programming which allow to solve large-scale DC programs. It is worth noting that (see, e.g., (An and Tao, 2002b)) for suitable DC decompositions, DCA generates almost standard algorithms in convex and nonconvex programming.

The computational efficiency of DCA for solving nonsmooth nonconvex optimization problems suggested to us *a deeper and more complete study on DC programming in the convex analysis approach.* That is the purpose of the paper.

In section 2 we present the backbone of the paper which is dealing with:

(i) New results on local optimality in DC programming and a special class of DC programs (when either g or h are polyhedral convex) called polyhedral DC programs. Polyhedral DC programs, which play a key role in nonconvex programming, constitute here the main stream of the exposition.

(ii) The DC duality is formulated, using the main feature of lower semicontinuous proper convex functions on X (namely such a function is characterized as the supremum of a collection of affine minorizations), in an easier way which is more convenient to the study of optimality conditions in DC programming and DCA.

(iii) We emphasize the still open problem whose answer is crucial to a deeper understanding of the DC structure: does it exist false DC program whose dual program is really DC program? This problem is intimately related to the notion *more convex and less convex* introduced by J.J. Moreau (Moreau, 1965). We discuss also regularization techniques in DC programming in order to construct suitable equivalent DC programs to nonconvex programs.

(iv) Using polyhedral DC programming, a deeper insight into DCA is introduced which *really sheds new light on DCA and could partly explain its efficiency.*

We briefly indicate in section 3 the succinct list of DC programs which have been successfully treated by DCA. Finally concluding remarks are reported in the last section.

2. DC programming and DCA

In this section we summarize the material needed for an easy understanding of DC programming and DCA. We are working with the space $X = \mathbb{R}^n$ which is equipped with the canonical inner product $\langle \cdot, \cdot \rangle$ and the corresponding Euclidean norm $\| \cdot \|$, thus the dual space Y of X can be identified with X itself. We follow (Rockafellar, 1970) for definitions of usual tools of convex analysis where functions could take the infinite values $\pm\infty$. A function $\theta : X \to \mathbb{R} \cup \{\pm\infty\}$ is said to be *proper* if it takes nowhere the value $-\infty$ and is not identically equal to $+\infty$. The effective domain of θ, denoted by $\operatorname{dom}\theta$, is

$$\operatorname{dom}\theta = \{x \in X:\ \theta(x) < \infty\}.$$

The set of all lower semicontinuous proper convex functions on X is denoted by $\Gamma_0(X)$. For $g \in \Gamma_0(X)$, the *conjugate function* g^* of g is a function belonging to $\Gamma_0(Y)$ and defined by

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

Note that $g^{**} = g$.

Let $g \in \Gamma_0(X)$ and let $x^0 \in \operatorname{dom} g$ and $\epsilon > 0$. Then $\partial_\epsilon g(x^0)$ stands for the *ϵ-subdifferential* of g at x^0 and is given by

$$\partial_\epsilon g(x^0) = \{y^0 \in Y:\ g(x) \geqslant g(x^0) + \langle x - x^0, y^0\rangle - \epsilon,\ \forall x \in X\}$$

while $\partial g(x^0)$ corresponding to $\epsilon = 0$ stands for the *usual* (*or exact*) *subdifferential* of g at x^0. Recall that

$$y^0 \in \partial g(x^0) \iff x^0 \in \partial g^*(y^0) \iff \langle x^0, y^0\rangle = g(x^0) + g^*(y^0).$$

One says that g is *subdifferentiable* at x^0 if $\partial g(x^0)$ is nonempty. It has been proved that (Rockafellar, 1970)

$$\operatorname{ri}(\operatorname{dom} g) \subset \operatorname{dom}\partial g \subset \operatorname{dom} g$$

where $\operatorname{ri}(\operatorname{dom} g)$ stands for the relative interior of $\operatorname{dom} g$ and $\operatorname{dom}\partial g := \{x \in X:\ \partial g(x) \neq 0\}$. The interior of a set S in X is denoted $\operatorname{int} S$. Also, the indicator function χ_C of a closed convex set C is defined by $\chi_C(x) = 0$ if $x \in C$, and $+\infty$ otherwise.

A function $\theta \in \Gamma_0(X)$ is said to be *polyhedral convex* if (Rockafellar, 1970)

$$\theta(x) = \max\{\langle a^i, x\rangle - \beta_i:\ i = 1, \ldots, m\} + \chi_S(x), \quad \forall x \in X, \tag{1}$$

where $a^i \in Y$, $\alpha_i \in \mathbb{R}$ for $i = 1, \ldots, m$ and S is a nonempty polyhedral convex set in X. Recall that (Rockafellar, 1970) the conjugate of a polyhedral convex function is polyhedral convex and the sum of polyhedral convex functions is polyhedral convex too.

Let $\rho \geqslant 0$ and C be a convex subset of X. One says that a function $\theta : C \to \mathbb{R} \cup \{+\infty\}$ is *ρ-convex* if

$$\theta[\lambda x + (1-\lambda)x'] \leqslant \lambda\theta(x) + (1-\lambda)(x') - \frac{\lambda(1-\lambda)}{2}\rho\|x - x'\|^2, \quad \forall\lambda \in]0, 1[,\ \forall x, x' \in C.$$

It amounts to saying that $\theta - (\rho/2)\|\cdot\|^2$ is convex on C. The *modulus of strong convexity* of θ on C, denoted by $\rho(\theta, C)$ or $\rho(\theta)$ if $C = X$, is given by

$$\rho(\theta, C) = \sup\{\rho \geqslant 0:\ \theta - (\rho/2)\|\cdot\|^2 \text{ is convex on } C\}. \tag{2}$$

Clearly, θ is convex on C if and only if $\rho(\theta, C) \geqslant 0$. One says that θ is *strongly convex* on C if $\rho(\theta, C) > 0$.

A function $f \in \Gamma_0(X)$ is said to be *essentially differentiable* if it satisfies the following three conditions (Rockafellar, 1970):

(i) $\text{int}(\text{dom } f) \neq \emptyset$,

(ii) f is differentiable on $\text{int}(\text{dom } f)$,

(iii) $\lim_{k\to+\infty} \|\nabla f(x^k)\| = +\infty$ for every sequence $\{x^k\} \subset \text{int}(\text{dom } f)$ converging to a boundary point x of $\text{int}(\text{dom } f)$.

A function $f \in \Gamma_0(X)$ is said to be strictly convex on a convex set C of dom f if

$$f\big(\lambda x + (1-\lambda)x'\big) < \lambda f(x) + (1-\lambda) f\big(x'\big), \quad \forall \lambda \in]0,1[,\ \forall x, x' \in C,\ x \neq x'.$$

For $g, h \in \Gamma_0(X)$, a general *DC program* is that of the form

$$(\text{P}_{\text{dc}}) \quad \alpha = \inf\big\{f(x) = g(x) - h(x)\colon\ x \in X\big\},$$

where, to avoid ambiguity, we adopt the convention (which is justified in section 2.2)

$$+\infty - (+\infty) = +\infty. \tag{3}$$

One says that $g - h$ is a *DC decomposition* (or DC representation) of f, and g, h are its *convex DC components*. If g and h are finite on X, then $f = g - h$ is said to be a finite DC function on X. The set of DC functions (resp. finite DC functions) on X is denoted by $\mathcal{DC}(X)$ (resp. $\mathcal{DC}_f(X)$).

Note that the finiteness of α merely implies that

$$\text{dom}\, g \subset \text{dom}\, h \quad \text{and} \quad \text{dom}\, h^* \subset \text{dom}\, g^*. \tag{4}$$

Such inclusions will be assumed throughout the paper.

A point x^* is said to be *a local minimizer* of $g - h$ if $g(x^*) - h(x^*)$ is finite (i.e., $x^* \in \text{dom}\, g \cap \text{dom}\, h$) and there exists a neighbourhood $\mathcal{U}$ of x^* such that

$$g(x^*) - h(x^*) \leqslant g(x) - h(x), \quad \forall x \in \mathcal{U}. \tag{5}$$

Under the convention $+\infty - (+\infty) = +\infty$, the property (5) is equivalent to $g(x^*) - h(x^*) \leqslant g(x) - h(x), \forall x \in \mathcal{U} \cap \text{dom}\, g$.

A point x^* is said to be *a critical point* of $g - h$ if $\partial g(x^*) \cap \partial h(x^*) \neq \emptyset$.

It is worth noting the richness of $\mathcal{DC}(X)$ and $\mathcal{DC}_f(X)$ (An, 1997; Tao and An, 1997, 1998; and references therein): they contain almost realistic objective functions and are closed under all the operations usually considered in optimization.

DC programming is a natural extension of convex maximization in which the function g is the indicator function of a nonempty closed convex set C. In the convex analysis approach to nonsmooth nonconvex optimization, convex maximization has been extensively studied since 1974 by Pham Dinh Tao (Tao, 1981, 1984, 1985, 1986, and references therein) who has introduced subgradient algorithms for solving convex maximization problems.

Let us mention, at the very beginning, the main features of DC programming which have been developed and exploited in our works:

(i) a DC function has infinitely many DC decompositions;

(ii) the DC duality and optimality in DC programming are built uniquely from DC components and their conjugates (but not from the DC function f itself);

(iii) DC algorithms (DCA) are also constructed from DC components and their conjugates.

In (P_{dc}) the nonconvexity comes from the concavity of the function $-h$ (except the case h is affine since (P_{dc}) then is a convex program). Let us outline our way of introducing the duality in DC programming.

2.1. Polyhedral DC programs and their dual DC programs

Polyhedral DC programs are DC programs in which one of DC components is polyhedral convex. Consider first the case where h is a polyhedral convex function given by (1). According to (4), (P_{dc}) can be written as

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

where

$$\widetilde{h}(x) = \max\{\langle a^i, x\rangle - \beta_i:\ i = 1, \ldots, m\}, \quad \forall x \in X.$$

Hence (P_{dc}) takes the form ($\mathcal{M} := \{1, \ldots, m\}$)

$$\begin{aligned} \alpha &= \inf_{x\in X} \inf_{i\in\mathcal{M}} \{f(x) = g(x) - [\langle a^i, x\rangle - \beta_i]\} \\ &= \inf_{i\in\mathcal{M}} \inf_{x\in X} \{f(x) = g(x) - [\langle a^i, x\rangle - \beta_i]\}. \end{aligned} \tag{6}$$

That gives rise to convex programs

$$(\widetilde{P}_i) \quad \alpha(i) = \inf_{x\in X}\{f(x) = g(x) - [\langle a^i, x\rangle - \beta_i]\}$$

whose the solution set $\widetilde{\mathcal{P}}_i$ is $\partial g^*(a^i)$. Let

$$\mathcal{M}(\alpha) := \{i \in \mathcal{M}:\ \alpha(i) = \alpha\} \quad \text{and} \quad \mathcal{M}(x) := \{i \in \mathcal{M}:\ \langle a^i, x\rangle - \beta_i = \widetilde{h}(x)\}$$

then the solution set $\widetilde{\mathcal{P}}$ of $(\widetilde{P})$ is given by

$$\begin{aligned} \alpha &= \min\{\alpha(i):\ i \in \mathcal{M}\}, \\ \widetilde{\mathcal{P}} &= \bigcup\{\widetilde{\mathcal{P}}_i:\ i \in \mathcal{M}(\alpha)\} = \bigcup\{\partial g^*(a^i):\ i \in \mathcal{M}(\alpha)\}. \end{aligned} \tag{7}$$

Solving the DC program $(\widetilde{P})$ amounts then to solving convex programs $(\widetilde{P}_i)$ with $i \in \mathcal{M}$. These results hide a certain duality we are now displaying. From (6) we have

$$\alpha = \inf_{i\in\mathcal{M}}\{\beta_i - g^*(a^i)\}. \tag{8}$$

The next result, stated for polyhedral DC programming (Tao and An, 1997), is crucial to a good understanding of the general DC programming from both theoretical and algorithmic viewpoints.

Proposition 1. There holds

$$(\widetilde{\mathrm{D}}) \quad \alpha = \inf\left\{\widetilde{h}^*(y) - g^*(y)\colon\ y \in co\left\{a^i\colon\ i \in \mathcal{M}\right\}\right\}$$

with the optimal value α satisfying

$$\alpha = \inf\left\{\widetilde{h}^*\left(a^i\right) - g^*\left(a^i\right)\colon\ i \in \mathcal{M}\right\}.$$

According to (3) and (4), $(\widetilde{\mathrm{D}})$ can be written as

$$\alpha = \inf\left\{\widetilde{h}^*(y) - g^*(y)\colon\ y \in Y\right\} \tag{9}$$

and is said to be the *dual program* of $(\widetilde{\mathrm{P}})$. It is *polyhedral DC program* too.

2.2. *Duality in DC programming*

Proposition 1 can be extended to the general case where $h \in \Gamma_0(X)$ is arbitrary by using the fundamental characterization of a convex function $\theta \in \Gamma_0(X)$ as *the pointwise supremum of a collection of affine minorizations*:

$$\theta(x) = \sup\left\{\langle x, y\rangle - \theta^*(y)\colon\ y \in Y\right\}, \quad \forall x \in X \tag{10}$$

(the polyhedral convex function h corresponds to *a finite collection of affine minorizations).* Indeed, replacing the function h by its expression of the form (10) in $(\mathrm{P}_{\mathrm{dc}})$, we have

$$\alpha = \inf\left\{g(x) - \sup\left\{\langle x, y\rangle - h^*(y)\colon\ y \in Y\right\}\colon\ x \in X\right\} = \inf\left\{\alpha(y)\colon\ y \in Y\right\}$$

with

$$(\mathrm{P}_y) \quad \alpha(y) := \inf\left\{g(x) - \left[\langle x, y\rangle - h^*(y)\right]\colon\ x \in X\right\}.$$

It is clear that (P_y) is a convex program and

$$\alpha(y) = \begin{cases} h^*(y) - g^*(y) & \text{if } y \in \operatorname{dom} h^*, \\ +\infty & \text{otherwise.} \end{cases} \tag{11}$$

Finally we state the dual program of $(\mathrm{P}_{\mathrm{dc}})$

$$\alpha = \inf\left\{h^*(y) - g^*(y)\colon\ y \in \operatorname{dom} h^*\right\}$$

that is written, according to (3), as

$$(\mathrm{D}_{\mathrm{dc}}) \quad \alpha = \inf\left\{h^*(y) - g^*(y)\colon\ y \in Y\right\}.$$

We observe the perfect symmetry between primal and dual DC programs: the dual to $(\mathrm{D}_{\mathrm{dc}})$ is exactly $(\mathrm{P}_{\mathrm{dc}})$.

These results generalize, in a very elegant and natural way, the early works of Pham Dinh Tao on convex maximization programming, just mentioned above. The formulation of dual DC programs can also be deduced from the general framework on duality in nonconvex programming by J.F. Toland (1978, 1979).

2.3. *Optimality conditions in DC programming*

Let $\mathcal{P}$ and $\mathcal{D}$ denote the solution sets of problems (P_{dc}) and (D_{dc}), respectively, and let

$$\mathcal{P}_l = \left\{x^* \in X\colon \partial h(x^*) \subset \partial g(x^*)\right\}, \qquad \mathcal{D}_l = \left\{y^* \in Y\colon \partial g^*(y^*) \subset \partial h^*(y^*)\right\}.$$

We first present the well-known fundamental result on DC programming on which the DCA relies (An, 1997; Tao and An, 1998).

Theorem 1. (i) Transportation of global minimizers: $\bigcup\{\partial h(x)\colon x \in \mathcal{P}\} \subset \mathcal{D} \subset \operatorname{dom} h^*$. The first inclusion becomes identity if g^* is subdifferentiable in $\mathcal{D}$ (in particular if $\mathcal{D} \subset \operatorname{ri}(\operatorname{dom} g^*)$ or if g^* is subdifferentiable in $\operatorname{dom} h^*$). In this case $\mathcal{D} \subset (\operatorname{dom} \partial g^* \cap \operatorname{dom} \partial h^*)$.

(ii) If x^* is a local minimizer of $g - h$, then $x^* \in \mathcal{P}_l$. The converse statement holds if h is a polyhedral convex function.

(iii) Let x^* be a critical point of $g - h$ and $y^* \in \partial g(x^*) \cap \partial h(x^*)$. Let $\mathcal{U}$ be a neighbourhood of x^* such that $(\mathcal{U} \cap \operatorname{dom} g) \subset \operatorname{dom} \partial h$. If for any $x \in \mathcal{U} \cap \operatorname{dom} g$ there is $y \in \partial h(x)$ such that $h^*(y) - g^*(y) \geqslant h^*(y^*) - g^*(y^*)$, then x^* is a local minimizer of $g - h$. More precisely,

$$g(x) - h(x) \geqslant g(x^*) - h(x^*), \quad \forall x \in \mathcal{U} \cap \operatorname{dom} g.$$

(iv) Transportation of local minimizers: Let $x^* \in \operatorname{dom} \partial h$ be a local minimizer of $g - h$ and let $y^* \in \partial h(x^*)$ (i.e., $\partial h(x^*)$ is nonempty and x^* admits a neighbourhood $\mathcal{U}$ such that $g(x) - h(x) \geqslant g(x^*) - h(x^*), \forall x \in \mathcal{U} \cap \operatorname{dom} g$). Under the assumption

$$y^* \in \operatorname{int}(\operatorname{dom} g^*) \quad \text{and} \quad \partial g^*(y^*) \subset \mathcal{U}, \tag{12}$$

for example if g^* is differentiable at y^*, we have y^* is a local minimizer of $h^* - g^*$.

Remark 1. (i) Naturally, by the symmetry of the DC duality, theorem 1 has its dual counterpart.

(ii) All these results concern only the DC components g and h (but not f).

Problem (P_{dc}) is a "*false*" *DC program* if the function $f = g - h$ is actually convex on X. In this case we have

Proposition 2. Let $f = g - h$, with $g, h \in \Gamma_0(X)$ verifying $\operatorname{dom} g \subset \operatorname{dom} h$ and $\operatorname{ri}(\operatorname{dom} g) \cap \operatorname{ri}(\operatorname{dom} h) \neq \emptyset$. Let $x_0 \in \operatorname{dom} g$ (resp. $\operatorname{dom} h$) be a point where g (resp. h) is continuous. If f is convex, then

(i) h is continuous at x_0,

(ii) $\partial f(x_0) = \partial g(x_0) \overset{*}{-} \partial h(x_0)$,

(iii) $0 \in \partial f(x_0) \Leftrightarrow \partial h(x_0) \subset \partial g(x_0)$.

Proof. We have $\operatorname{dom} f = \operatorname{dom} g$ and $g = f + h$. Hence $\partial g(x_0) = \partial f(x_0) + \partial h(x_0)$ in virtue of $\operatorname{ri}(\operatorname{dom} g) \cap \operatorname{ri}(\operatorname{dom} h) \neq \emptyset$. If g is continuous at $x_0 \in \operatorname{dom} g$, we have

$$x_0 \in \operatorname{ri}(\operatorname{dom} g) = \operatorname{int}(\operatorname{dom} g) \subset \operatorname{int}(\operatorname{dom} h).$$

So the closed convex set $\partial h(x_0)$ is bounded and it follows from (An, 1994; Pschenichny, 1971)

$$\partial f(x_0) = \partial g(x_0) \overset{*}{-} \partial h(x_0) := \bigcap \{\partial g(x_0) - z\colon z \in \partial h(x_0)\}$$

(the symbol $C \overset{*}{-} D$ stands for the $*$ *difference* of the two sets C and D of X: $C \overset{*}{-} D = \{x \in X\colon x + D \subset C\}$ (Pschenichny, 1971)). □

The equivalence (iii) then is immediate.

Remark 2. Convex programs can then be recasted (and solved) in the DC programming as follows: Let $f, \theta \in \Gamma_0(X)$ be such that $\operatorname{dom} f \subset \operatorname{dom} \theta$ and $\operatorname{ri}(\operatorname{dom} f) \cap \operatorname{ri}(\operatorname{dom} \theta) \neq \emptyset$, then the convex program

$$\inf\{f(x)\colon x \in X\}$$

is equivalent to the false DC program

$$\inf\{(f + \theta)(x) - \theta(x)\colon x \in X\} \tag{13}$$

in the following sense:

(i) they have the same optimal value and the same solution set,

(ii) we have

$$0 \in \partial f(x^*) \Longrightarrow \partial\theta(x^*) \subset \partial(f + \theta)(x^*) = \partial f(x^*) + \partial\theta(x^*) \tag{14}$$

and the other implication holds if, in addition, θ is continuous at x^*.

The following result on local optimality in DC programming is very useful for a large class of real world DC programs (An, 1997, 2003; An and Tao, 1999, 2000a, 2000b, 2001a, 2001b, 2001c, 2002a, 2003a, 2003b; An, Tao, and Hao, 2003, 2004b).

Proposition 3. Let $f = g - h$, with $g, h \in \Gamma_0(X)$ verifying $\operatorname{dom} g \subset \operatorname{dom} h$. If there exists a convex neighbourhood $\mathcal{U}$ of x^* such that f is finite and convex on $\mathcal{U}$, then the following two assertions are equivalent:

(i) $0 \in \partial(f + \chi_{\mathcal{U}})(x^*)$,

(ii) $\partial h(x^*) \subset \partial g(x^*)$.

It amounts to say that, under the above assumptions, the necessary condition of local DC optimality (theorem 1)

$$\partial h(x^*) \subset \partial g(x^*)$$

is also sufficient: $f(x) \geqslant f(x^*)$ for all $x \in \mathcal{U}$.

Proof. By hypothesis, we have

$$\mathcal{U} \subset \operatorname{dom} f = \operatorname{dom} g. \tag{15}$$

Hence

$$x^* \in \operatorname{int}\mathcal{U} \subset \operatorname{int}(\operatorname{dom} g) \subset \operatorname{int}(\operatorname{dom} h), \tag{16}$$

and

$$g + \chi_{\mathcal{U}} = (f + \chi_{\mathcal{U}}) + h.$$

On the other hand, we can write (Laurent, 1972)

$$\partial g(x^*) = \partial(g + \chi_{\mathcal{U}})(x^*) = \partial(f + \chi_{\mathcal{U}})(x^*) + \partial h(x^*).$$

Since the closed convex set $\partial h(x^*)$ is bounded, we deduce that (An, 1994; Pschenichny, 1971)

$$\partial(f + \chi_{\mathcal{U}})(x^*) = \partial g(x^*) \stackrel{*}{-} \partial h(x^*).$$

The proof is then complete by using the same reasonings as proposition 2. □

Consider now a DC program of the form

$$\inf\{\varphi(x) = k(x) - h(x) \colon x \in C\} \tag{17}$$

with $k, h \in \Gamma_0(X)$ both being finite on X and C being a closed convex set of X. This DC program can be rewritten in the canonical form

$$\inf\{(\varphi + \chi_C)(x) = (k + \chi_C)(x) - h(x) \colon x \in X\}$$

or

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

where

$$g = k + \chi_C.$$

A local optimality condition for problem (17), similar to proposition 3, is given below.

Proposition 4. If there is a convex neighbourhood $\mathcal{U}$ of x^* such that the function $\varphi = k - h$ is convex on $C \cap \mathcal{U}$, then the following two assertions are equivalent:

(i) $0 \in \partial(\varphi + \chi_{C\cap\mathcal{U}})(x^*)$,

(ii) $\partial h(x^*) \subset \partial(k + \chi_C)(x^*)$.

It amounts to say that, under the above assumptions, the necessary condition of local DC optimality (theorem 1 with $g = k + \chi_C$)

$$\partial h(x^*) \subset \partial g(x^*)$$

is also sufficient: $\varphi(x) \geqslant \varphi(x^*)$ for all $x \in C \cap \mathcal{U}$.

Proof. By assumption, the function $\varphi + \chi_{C\cap\mathcal{U}}$ is convex on X and

$$k + \chi_{C\cap\mathcal{U}} = h + (\varphi + \chi_{C\cap\mathcal{U}}).$$

Hence

$$\partial(k + \chi_{C\cap\mathcal{U}})(x^*) = \partial h(x^*) + \partial(\varphi + \chi_{C\cap\mathcal{U}})(x^*).$$

It follows that (An, 1994; Pschenichny, 1971)

$$\partial(\varphi + \chi_{C\cap\mathcal{U}})(x^*) = \partial(k + \chi_{C\cap\mathcal{U}})(x^*) \overset{*}{-} \partial h(x^*). \tag{18}$$

But from

$$k + \chi_{C\cap\mathcal{U}} = k + \chi_C + \chi_{\mathcal{U}} \quad \text{and} \quad \partial\chi_{\mathcal{U}}(x^*) = \{0\}$$

we deduce that

$$\partial(\varphi + \chi_{C\cap\mathcal{U}})(x^*) = \partial(k + \chi_C)(x^*) + \partial\chi_{\mathcal{U}}(x^*) = \partial(k + \chi_C)(x^*).$$

Finally, as in the preceding demonstrations, we conclude the assertions in proposition 4. □

The equality of the optimal value in the primal and dual DC programs can be easily translated (with the help of *ϵ-subdifferential* of the DC components) in global optimality conditions, namely x^* is a global solution to (P_{dc}) if and only if (see (An, 1997; Hiriat Urruty, 1989; Tao and An, 1997) and references therein)

$$\partial_\epsilon h(x^*) \subset \partial_\epsilon g(x^*), \quad \forall\epsilon \geqslant 0.$$

Unfortunately, as we foresee, these conditions are rather difficult to use for devising solution methods to DC programs.

2.4. False DC program whose dual is a real DC program?

Thanks to a *symmetry* in the DC duality (the bidual DC program is exactly the primal one) and the *DC duality transportation of global minimizers* (see theorem 1), solving a DC program implies solving the dual one and *vice versa*. It may be useful if one of them is easier to solve than the other. It leads to the open problem whose answer is crucial to a deep understanding of the DC structure: *does it exist false DC program whose dual program is really DC program*? This problem is intimately related to the notion *more convex and less convex* introduced by J.J. Moreau (1965): Let $g, h \in \Gamma_0(X)$, the function

g is said to be more convex than h (or h is said to be less convex than g), and we write $g \succ h$ (or $h \prec g$) if

$$g = h + \theta, \quad \text{with } \theta \in \Gamma_0(X).$$

According to (Moreau, 1965) we have, for $h(x) := \frac{1}{2}\langle x, Ax\rangle + \langle x, b\rangle$ with A being a symmetric positive definite $n \times n$ matrix

$$g \succ h \iff g^* \prec h^*.$$

It is very important to study the set of functions $g, h \in \Gamma_0(X)$ verifying one of the following properties:

(i) $g \succ h$ and $g^* \prec h^*$.
In this case both the corresponding primal and dual DC programs are convex.

(ii) $g \succ h$ but not $g^* \prec h^*$.
The primal DC program is then convex but not its dual. Regarding problem (13), it amounts to find $\theta \in \Gamma_0(X)$ such that $(f + \theta)^* \prec \theta^*$ does not hold.

Another important feature of the DC structure, which must be taken into account while studying solution algorithms, is regularization techniques in DC programming which have been first studied by Pham Dinh Tao in 1986 (Tao, 1988) and they were used to improve DCA in numerical solutions of many real world nonconvex programs (Tao and An, 1997, and references therein).

Remark 3. Different ways to formulate equivalent DC programs to nonconvex optimization problems can be used affording Lagrangian duality without gap or exact penalty techniques (An, Tao, and Muu, 1999; Tao and An, 1994, 1995; see also Thach, 1993b, 1994, 1996).

Using the theoretical framework of DC programming just above displayed, we are now in a position to outline the DCA for general DC programs.

2.5. *The DCA for general DC programs*

The DCA consists in the construction of the two sequences $\{x^k\}$ and $\{y^k\}$ (candidates for being primal and dual solutions, respectively) that we improve at each iteration (thus, the sequences $\{g(x^k) - h(x^k)\}$ and $\{h^*(y^k) - g^*(y^k)\}$ are decreasing) in an appropriate way such that their corresponding limits x^∞ and y^∞ satisfy the local optimality condition

$$\partial h(x^\infty) \subset \partial g(x^\infty) \quad \text{and} \quad \partial g^*(y^\infty) \subset \partial h^*(y^\infty), \quad \text{i.e.} \quad (x^\infty, y^\infty) \in \mathcal{P}_l \times \mathcal{D}_l$$

(where $\mathcal{P}_l = \{x^* \in X\colon \partial h(x^*) \subset \partial g(x^*)\}$ and $\mathcal{D}_l = \{y^* \in Y\colon \partial g^*(y^*) \subset \partial h^*(y^*)\}$) or are critical points of $g - h$ and $h^* - g^*$, respectively.

These sequences are generated as follows: x^{k+1} (resp. y^k) is a solution to the *convex program* (P_k) (resp. (D_k)) defined by

$$(P_k) \quad \begin{cases} \alpha_k = \inf\{g(x) - [h(x^k) + \langle x - x^k, y^k\rangle]\} \\ \text{s.t.} \quad x \in X, \end{cases}$$

$$(D_k) \quad \begin{cases} \inf\{h^*(y) - [g^*(y^{k-1}) + \langle x^k, y - y^{k-1}\rangle]\} \\ \text{s.t.} \quad y \in Y. \end{cases}$$

In view of the relation: (P_k) (resp. (D_k)) is obtained from (P_{dc}) (resp. (D_{dc})) by replacing h (resp. g^*) with its affine minorization defined by $y^k \in \partial h(x^k)$ (resp. $x^k \in \partial g^*(y^{k-1})$), the DCA yields the next scheme

$$y^k \in \partial h(x^k); \quad x^{k+1} \in \partial g^*(y^k). \tag{19}$$

It corresponds actually to the *simplified* DCA (which will be shortly called DCA through the paper for simplicity) where x^{k+1} (resp. y^k) is arbitrarily chosen in $\partial g^*(y^k)$ (resp. $\partial h(x^k)$). Remark that

$$h(x^k) + \langle x - x^k, y^k\rangle = \langle x, y^k\rangle - h^*(y^k) \quad \text{and}$$
$$g^*(y^{k-1}) + \langle x^k, y - y^{k-1}\rangle = \langle x^k, y\rangle - g(x^k)$$

according to (19) and so problem (P_k) is exactly (P_y) with $y = y^k$. We have the analogous result for problem (D_k) via the DC duality.

In the *complete* form of DCA, we impose the following natural choice

$$x^{k+1} \in \arg\min\{g(x) - h(x) \colon x \in \partial g^*(y^k)\} \tag{20}$$

and

$$y^k \in \arg\min\{h^*(y) - g^*(y) \colon y \in \partial h(x^k)\}. \tag{21}$$

Problems (20) and (21) are equivalent to convex maximization problems (22) and (23), respectively

$$x^{k+1} \in \arg\min\{\langle x, y^k\rangle - h(x) \colon x \in \partial g^*(y^k)\}, \tag{22}$$

$$y^k \in \arg\min\{\langle x^k, y\rangle - g^*(y) \colon y \in \partial h(x^k)\}. \tag{23}$$

The complete DCA ensures that $(x^\infty, y^\infty) \in \mathcal{P}_l \times \mathcal{D}_l$. It can be viewed as a sort of the decomposition approach of the primal and dual problems (P_{dc}), (D_{dc}). From a practical point of view, although problems (20) and (21) are simpler than (P_{dc}), (D_{dc}) (we work in $\partial h(x^{k+1})$ and $\partial g^*(y^k)$ with convex maximization problems), they remain nonconvex programs and thus are still hard to solve. In practice, except for the cases where the convex maximization problems (22) and (23) are easy to treat, one generally uses the simplified DCA to solve DC programs.

The DCA was introduced by Pham Dinh Tao in 1986 as an extension of the aforementioned subgradient algorithms (for convex maximization programming) to DC programming. However, this field has been really developed since 1994 by joint works of Le Thi Hoai An and Pham Dinh Tao (An, 1994, 1997, 2000, 2003; An and Tao, 1997, 1998, 1999, 2000a, 2000b, 2001a, 2001b, 2001c, 2002a, 2002b, 2003a, 2003b; An, Tao, and Muu, 1996, 1998, 1999, 2003a, 2003b; An, Tao, and Thoai, 2002; An, Tao, and Hao, 2002, 2003, 2004a, 2004b; Tao and An, 1994, 1995, 1998, 1997) for solving nonsmooth nonconvex optimization problems. To our knowledge, DCA is actually one of a few algorithms (in the convex analysis approach to DC programming) which allow to solve large-scale DC programs.

It had been proved in Pham Dinh Tao and Le Thi Hoai An (see, e.g., An, 1997; Tao and An, 1998, 1997) that for the simplified DCA we have

(i) The sequences $\{g(x^k) - h(x^k)\}$ and $\{h^*(y^k) - g^*(y^k)\}$ are decreasing and

- $g(x^{k+1}) - h(x^{k+1}) = g(x^k) - h(x^k)$ if and only if $y^k \in \partial g(x^k) \cap \partial h(x^k)$, $y^k \in \partial g(x^{k+1}) \cap \partial h(x^{k+1})$ and $[\rho(g) + \rho(h)]\|x^{k+1} - x^k\| = 0$.
- $h^*(y^{k+1}) - g^*(y^{k+1}) = h^*(y^k) - g^*(y^k)$ if and only if $x^{k+1} \in \partial g^*(y^k) \cap \partial h^*(y^k)$, $x^{k+1} \in \partial g^*(y^{k+1}) \cap \partial h^*(y^{k+1})$ and $[\rho(g^*) + \rho(h^*)]\|y^{k+1} - y^k\| = 0$.

DCA terminates at the kth iteration if either of the above equalities holds.

(ii) If $\rho(g) + \rho(h) > 0$ (resp. $\rho(g^*) + \rho(h^*) > 0$), then the series $\{\|x^{k+1} - x^k\|^2\}$ (resp. $\{\|y^{k+1} - y^k\|^2\}$) converges.

(iii) If the optimal value α of problem (P_{dc}) is finite and the sequences $\{x^k\}$ and $\{y^k\}$ are bounded, then every limit point x^∞ (resp. y^∞) of the sequence $\{x^k\}$ (resp. $\{y^k\}$) is a critical point of $g - h$ (resp. $h^* - g^*$).

(iv) DCA has a linear convergence for general DC programs.

(v) In polyhedral DC programs, the sequences DCA $\{x^k\}$ and $\{y^k\}$ contain finitely many elements and DCA has a finite convergence.
We have the same results for the complete DCA, except that in (i) (resp. (iii)) we must add the following property:
$\partial h(x^k) \subset \partial g(x^k)$ and $\partial g^*(y^k) \subset \partial h^*(y^k)$ (resp. $\partial h(x^\infty) \subset \partial g(x^\infty)$ and $\partial g^*(y^\infty) \subset \partial h^*(y^\infty)$).

(vi) If DCA converges to a point x^* that admits a convex neighbourhood in which the objective function f is finite and convex (i.e. the function f is locally convex at x^*) and if the second DC component h is differentiable at x^*, then x^* is a local minimizer for (P_{dc}). For the complete DCA the second condition is not needed. Property (vi) has of course its dual part.

For more details, see (An, 1997; Tao and An, 1998, 1997) and references therein.

Remark 4. (i) There are as many DCA as there are DC decompositions. It is of particular interest to study various equivalent DC forms for the primal and dual DC problems. It is worth mentioning, for instance, that by using suitable DC decompositions of convex functions we can obtain almost standard algorithms for convex and nonconvex programming (An and Tao, 2002b).

(ii) The choice of the DC decomposition of the objective function in a DC program and the initial point for DCA are open questions to be studied. Of course, this depends strongly on the very specific structure of the problem being considered. In practice, for solving a given DC program, we try to choose g and h such that sequences $\{x^k\}$ and $\{y^k\}$ can be easily calculated, i.e. either they are in explicit form or their computations are inexpensive.

2.6. Deeper insight into DCA

The special class of polyhedral DC programs, which plays a key role in nonconvex optimization, possesses worthy properties, from both theoretical and computational viewpoints, *necessary and sufficient local optimality conditions, and finite convergence for DCA* (see theorem 1, sections 2.1, 2.3 and 2.5, (An, 1997; Tao and An, 1998, 1997)). Here also they are used to give a deeper interpretation of DCA.

Denote by h_k (resp. h^k) the affine (resp. polyhedral convex) minorization of the convex function h:

$$h_k(x) = h\big(x^k\big) + \big\langle x - x^k, y^k\big\rangle = \big\langle x, y^k\big\rangle - h^*\big(y^k\big), \quad \forall x \in X, \tag{24}$$

$$h^k(x) := \sup\big\{h_i(x):\ i = 0, \dots, k\big\} = \sup\big\{\big\langle x, y^i\big\rangle - h^*\big(y^i\big):\ i = 0, \dots, k\big\}, \quad \forall x \in X, \tag{25}$$

where the sequences $\{x^k\}$ and $\{y^k\}$ are generated as above, i.e. as corresponding solutions of the relaxed convex program (P_k) and (D_k) which are obtained from the original program $(\mathrm{P}_{\mathrm{dc}})$ and $(\mathrm{D}_{\mathrm{dc}})$ by replacing the convex function h with h_k and g^* by its affine minorization, $y \to g^*(y^k) + \langle x^{k+1}, y - y^k\rangle$.

Likewise, for the dual part, let $(g^*)_k$ (resp. $(g^*)^k$) be the affine (resp. polyhedral convex) minorization of the convex function g^*:

$$(g^*)_k(y) := g^*\big(y^{k-1}\big) + \big\langle y - y^{k-1}, x^k\big\rangle = \big\langle y, x^k\big\rangle - g\big(x^k\big) \quad \forall y \in Y, \tag{26}$$

and

$$(g^*)^k(y) := \sup\big\{(g^*)_i(y):\ i = 1, \dots, k\big\} = \sup\big\{\big\langle y, x^i\big\rangle - g\big(x^i\big):\ i = 1, \dots, k\big\} \quad \forall y \in Y. \tag{27}$$

Since a proper lower semicontinuous convex function θ is characterized as the supremum of a collection of its affine minorizations, namely

$$\theta(x) = \sup\big\{\langle x, y\rangle - \theta^*(y):\ y \in Y\big\}, \quad \forall x \in X,$$

it is better, instead of the affine minorizations h_k and $(g^*)_k$, *to use the polyhedral convex functions h^k and $(g^*)^k$ to under approximate the convex functions h and g^*, respectively* (we then tighten up these approximations at each iteration)

$$h_k \leqslant h^k \leqslant h, \quad \alpha_k \geqslant \beta_k \geqslant \alpha, \quad \forall k \geqslant 0. \tag{28}$$

In other words we are dealing with the following relaxed programs (which still are (non-convex) DC programs as opposed to the relaxed convex programs (P_k))

$$\beta_k = \inf\{g(x) - h^k(x):\ x \in X\}, \tag{29}$$

and

$$\inf\{h^*(y) - (g^*)^k(y):\ y \in Y\} \tag{30}$$

which actually are polyhedral DC programs.

It naturally leads us to the following crucial questions:

- how to solve the polyhedral DC programs (29) and (30)?
- what exactly are the relationships between the sequences $\{x^k\}$, $\{y^k\}$ and solutions to problems (29) and (30)?
- what additional assumptions (on the tightening of the approximations h^k and $(g^*)^k$ for the convex functions h and g^*, respectively) to ensure the convergence of the sequences $\{x^k\}$, $\{y^k\}$ to optimal solutions to problems (P_{dc}) and (D_{dc}), respectively?

Answers to these questions are presented below. First, by writing problem (29) in the form

$$\inf\{g(x) - \sup\{h_i(x):\ i = 0, \ldots, k\}:\ x \in X\},$$

we see that it is equivalent to the problem

$$\inf_{x\in X}\inf_{i=0,\ldots,k}\{g(x) - h_i(x)\},$$

and is so to

$$\inf_{i=0,\ldots,k}\inf_{x\in X}\{g(x) - h_i(x)\}.$$

Since x^{i+1} is an optimal solution to (P_i), we easily deduce that x^l, with $l \in \arg\min\{g(x^{i+1}) - h_i(x^{i+1}):\ i = 0, \ldots, k\}$, is an optimal solution to problem (29).

On the other hand, we have

$$g(x^{i+1}) - h(x^{i+1}) \leqslant g(x^{i+1}) - h_i(x^{i+1}), \quad \text{for } i = 0, \ldots, k.$$

So

$$g(x^{k+1}) - h(x^{k+1}) \leqslant \inf\{g(x^{i+1}) - h_i(x^{i+1}):\ i = 0, \ldots, k\} \tag{31}$$

since the sequence $\{g(x^i) - h(x^i)\}$ is decreasing.

It is clear that if

$$h_k(x^{k+1}) = h(x^{k+1}), \tag{32}$$

then $h^k(x^{k+1}) = h(x^{k+1})$ and the equality holds in (31)

$$g(x^{k+1}) - h(x^{k+1}) = g(x^{k+1}) - h_k(x^{k+1}) = \inf\{g(x^{i+1}) - h_i(x^{i+1}):\ i = 0, \ldots, k\}.$$

Hence x^{k+1} is an optimal solution to the polyhedral DC program (29).

Similarly, for the dual part, problem (30) can be written as

$$\inf_{i=1,\dots,k} \inf_{y\in Y}\big\{h^*(y) - (g^*)_i(y)\big\}.$$

Since y^i is an optimal solution to problem (D_i), it follows that y^l, with $l \in \arg\min\{h^*(y^i) - (g^*)_i(y^i)\colon\ i = 1,\dots,k\}$, is an optimal solution to problem (30).

Like the primal part, we have

$$h^*\big(y^i\big) - g^*\big(y^i\big) \leqslant h^*\big(y^i\big) - (g^*)_i(y) \quad \text{for } i = 1,\dots,k$$

and

$$h^*\big(y^k\big) - g^*\big(y^k\big) \leqslant \inf\big\{h^*\big(y^i\big) - (g^*)_i(y),\ i = 1,\dots,k\big\} \tag{33}$$

since the sequence $\{h^*(y^i) - g^*(y^i)\}$ is decreasing. Moreover, if

$$g^*\big(y^k\big) = (g^*)_k\big(y^k\big), \tag{34}$$

then $g^*(y^k) = (g^*)^k(y^k)$ and the equality holds in (33)

$$h^*\big(y^k\big) - g^*\big(y^k\big) = h^*\big(y^k\big) - (g^*)_k\big(y^k\big) = \inf\big\{h^*\big(y^i\big) - (g^*)_i\big(y^i\big)\colon\ i = 1,\dots,k\big\}$$

and y^k is a solution to problem (30).

To point out relationships between the sequences $\{x^k\}$, $\{y^k\}$ and solutions to problems (29) and (30) and also additional assumptions for ensuring the convergence of DCA to solutions of (P_{dc}) and (D_{dc}), we shall now find conditions ensuring the equalities (32) and (34). For this we will distinguish two cases:

(A) There is some k such that $h_k(x^{k+1}) = h(x^{k+1})$ or $g(y^k) = (g^*)_k(y^k)$.

In this case, the following result can be deduced from the general convergence theorem of DCA (An, 1997; Tao and An, 1998, 1997)

Theorem 2. (i) $g(x^{k+1}) - h(x^{k+1}) = h^*(y^k) - g^*(y^k)$ if and only if $h_k(x^{k+1}) = h(x^{k+1})$,

(ii) $h^*(y^k) - g^*(y^k) = g(x^k) - h(x^k)$ if and only if $g^*(y^k) = (g^*)_k(y^k)$,

(iii) $g(x^{k+1}) - h(x^{k+1}) = g(x^k) - h(x^k)$ if and only if $h_k(x^{k+1}) = h(x^{k+1})$ and $g^*(y^k) = (g^*)_k(y^k)$. Consequently, if $g(x^{k+1}) - h(x^{k+1}) = g(x^k) - h(x^k)$ (and the common value thus equals $h^*(y^k) - g^*(y^k)$), then the following statements hold:

- x^{k+1} (resp. y^k) is a solution to problem (29) (resp. (30)).
- Moreover, if h and h^k coincide at some solution to (P_{dc}) or g^* and $(g^*)^k$ coincide at some solution to (D_{dc}), then x^{k+1} (resp. y^k) is also a solution to (P_{dc}) (resp. (D_{dc})) and $h^k(x^{k+1}) = h(x^{k+1})$.

(B) $h_k(x^{k+1}) < h(x^{k+1})$ for every k.

This case corresponds to making k equal to infinity in (A) and the *asymptotic behaviour* of DCA can be described as below.

Assume the optimal value of problem (P_{dc}) be finite and the sequences $\{x^k\}$ and $\{y^k\}$ generated by DCA be bounded. Then

$$g(x^\infty) - h(x^\infty) = \inf\{g(x^{i+1}) - h_i(x^{i+1}): \ i = 0, \dots, \infty\}$$

for every limit point x^∞ of the sequence $\{x^k\}$, where h_∞ denotes the affine minorization of h:

$$h_\infty(x) := h(x^\infty) + \langle x - x^\infty, y^\infty \rangle = \langle x, y^\infty \rangle - h^*(y^\infty), \quad \forall x \in X$$

with $y^\infty \in \partial h(x^\infty)$ being a limit point of $\{y^k\}$.

Consequently the point x^∞ is a solution to the DC program

$$\inf\{g(x) - h^\infty(x): \ x \in X\}, \tag{35}$$

where the convex function h^∞ is defined by

$$h^\infty(x) := \sup\{\langle x, y^i \rangle - h^*(y^i): \ i = 0, \dots, \infty\}, \quad \forall x \in X.$$

Similarly, for the dual DC program (D_{dc}), the related affine minorizations of g^* are defined by

$$(g^*)_\infty(y) := g^*(y^\infty) + \langle y - y^\infty, x^\infty \rangle = \langle y, x^\infty \rangle - g(x^\infty), \quad \forall y \in Y$$

and

$$\begin{aligned}(g^*)^\infty(y) &:= \sup\{(g^*)_i(y): \ i = 1, \dots, \infty\} \\ &= \sup\{\langle y, x^i \rangle - g(x^i): \ i = 1, \dots, \infty\}, \quad \forall y \in Y.\end{aligned}$$

As above we have

$$h^*(y^\infty) - (g^*)^\infty(y^\infty) = \inf\{h^*(y^i) - (g^*)_i(y^i): \ i = 1, \dots, \infty\}$$

and the point y^∞ is a solution to the DC program

$$\inf\{h^*(y) - (g^*)^\infty(y): \ y \in Y\}. \tag{36}$$

Moreover, the optimal values of problems (35) and (36) are equal

$$g(x^\infty) - h^\infty(x^\infty) = h^*(y^\infty) - (g^*)^\infty(y^\infty).$$

Summing up, DCA generates two sequences $\{x^k\}$ and $\{y^k\}$ (candidates to primal and dual solutions, respectively) which are improved at each iteration (the sequences $\{g(x^k) - h(x^k)\}$ and $\{h^*(y^k) - g^*(y^k)\}$ are decreasing) and serve, in turn, to construct affine minorizations of h (the affine functions h_i), and affine minorizations of g^* (the affine functions $(g^*)_i$). These functions will be used to construct convex programs (P_i) and (D_i) – whose solutions are the next iterations of the sequences $\{x^k\}$ and $\{y^k\}$ – and to built up tighter and tighter polyhedral convex minorizations of the functions h and g^* so that DCA have the following asymptotic behaviour.

Theorem 3. If the optimal value α of the DC program (P_{dc}) is finite and the sequences $\{x^k\}$, $\{y^k\}$ are bounded, then every limit point x^∞ (resp. y^∞) of $\{x^k\}$ (resp. $\{y^k\}$) is a solution to the approximated DC program

$$\inf\{g(x) - h^\infty(x):\ x \in X\}$$

and

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

respectively. Moreover, the optimal values of these problems are equal

$$g(x^\infty) - h^\infty(x^\infty) = h^*(y^\infty) - (g^*)^\infty(y^\infty).$$

Consequently, the nearer the function h^∞ (resp. $(g^*)^\infty$) approaches to the function h (resp. g^*), the better x^∞ (resp. y^∞) approximates an optimal solution to the DC program (P_{dc}) (resp. (D_{dc})). Furthermore, if either of the following conditions holds

(i) the functions h^∞ and h coincide at some optimal solution to (P_{dc}),

(ii) the functions $(g^*)^\infty$ and g^* coincide at some optimal solution to (D_{dc}),

then x^∞ and y^∞ are also optimal solutions to (P_{dc}) and (D_{dc}), respectively.

Proof. It is immediate by using the argument in the proof of theorem 2 (just setting $k = +\infty$ in (A)) and the general convergence theorem of DCA (see An, 1997; Tao and An, 1998, 1997). □

Remark 5. (i) Theorem 2 holds especially in polyhedral DC programs where DCA has a finite convergence. The hidden features of DCA reside in (iv) and (v):

- the limit point x^{k+1} (resp. y^k) is not only the solution of (P_k) (resp. (D_k)) but also the solution to the more tightly approximate problem (29) (resp. (30)) of (P_{dc}) (resp. (D_{dc})),
- $\beta_k + \epsilon_k \leqslant \alpha \leqslant \beta_k$ where $\epsilon_k := \inf\{h^k(x) - h(x):\ x \in \mathcal{P}\} \leqslant 0$ and the more ϵ_k is near zero (i.e. the more the polyhedral convex minorization h^k is close to h over $\mathcal{P}$) the more x^{k+1} is near $\mathcal{P}$.

(ii) Theorem 3 corresponds to the infinite convergence of DCA ($k = +\infty$).

In practice, in order to globally solve (P_{dc}) and (D_{dc}), it is of particular interest to have the function h^k (resp. $(g^*)^k$), with k being finite or $+\infty$, tight under-approximate the function h (resp. g^*) especially in a neighborhood of a solution to problem (P_{dc}) (resp. (D_{dc})). To this end, we have to find suitable DC decompositions and initial points for DCA according to the specific structure of the DC program being considered.

3. DC models of real world nonsmooth nonconvex optimization problems

The major difficulty in nonconvex programming resides in the fact that there is, in general, no practicable global optimality conditions. Thus, checking globalilty of solutions computed by local algorithms is only possible in the cases where optimal values are known a priori or by comparison with global algorithms which, unfortunately, cannot be applied to large scale problems. A pertinent comparison of local algorithms should be based on the following aspects:

- mathematical foundations of the algorithms,
- rate of convergence and running-time,
- ability to treating large-scale problems,
- quality of computed solutions: the lower the corresponding value of the objective is, the better the local algorithm will be,
- the degree of dependence on initial points: the larger the set (made up of starting points which ensure convergence of the algorithm to a global solution) is, the better the algorithm will be.

DCA seems to meet these features since it was successfully applied to a lot of different and various nonconvex optimization problems to which it quite often gave global solutions and proved to be more robust and more efficient than related standard methods, especially in the large-scale setting. We present below the list of these problems:

1. The Trust Region subproblems (An, 1997, 2000; Conn, Gould, and Toint, 2000; Tao and An, 1998);

2. Nonconvex Quadratic Programs and Quadratic Zero-One Programming problems (An, 1997; An and Tao, 1997, 1998, 2001a; An, Tao, and Muu, 1998);

3. Multiple Criteria Optimization: Optimization over the Efficient and Weakly Efficient Sets (An, 1997; An, Tao, and Muu, 1996, 2003a; An, Tao, and Thoai, 2002);

4. Optimization in Mathematical Programming problems with Equilibrium Constraints (An, Tao, and Muu, 2003b);

5. Optimization in Mathematics Finance;

6. Molecular Optimization via the Distance Geometry Problems (An, 1997, 2003; An and Tao, 2000a, 2000b, 2003a, 2003b);

7. Multidimensional Scaling Problems (MDS) (An, 1997; An and Tao, 2001c);

8. Optimization for Restoration of Signals and Images (An, 1997; An and Tao, 2001b);

9. Tikhonov Regularization for Nonlinear Ill-Posed problems (An, Tao, and Hao, 2002, 2003, 2004b);

10. Nonconvex Multicommodity Network Optimization problems (An and Tao, 2002a);

11. Data Mining.

4. Concluding remarks

Using the polyhedral DC programming as the main stream of the exposition, we have revisited the DC programming and DCA in an easy way which is more suitable to applications. New results on optimality conditions and a deeper insight into DCA (again with the support of the polyhedral DC programming) allow a good understanding of the DC programming and partly explain the computational efficiency of DCA.

To guarantee globality of sought solutions or to improve their quality, it is advised to combine DCA with the following techniques in a deeper and efficient way:

1. Branch-and-Bound Techniques (see, e.g., 2, 3, 4, 8 in the preceding list): the most popular approach in global optimization;
2. Continuation methods (see, e.g., An, 2003; More and Wu, 1997, 1996; Wu, 1996). In the continuation approach, the original objective function is gradually transformed into a smoother with fewer local minimizers. An optimization algorithm is then applied to the transformed function, tracing the minimizers back to the original function.
3. Monotonic optimization, recently developed by H. Tuy (2000).

To our knowledge, these approaches are not much evolved in the literature. We are convinced that they will constitute new trends in nonconvex programming.

References

An, Le Thi Hoai. (1994). "Analyse numérique des algorithmes de l'optimisation DC. Approches locale et globale. Codes et simulations numériques en grande dimension. Applications." Thèse de Doctorat de l'Université de Rouen.

An, Le Thi Hoai. (1997). "Contribution à l'optimisation non convexe et l'optimisation globale: Théorie, algorithmes et applications." Habilitation à Diriger des Recherches, Université de Rouen.

An, Le Thi Hoai. (2000). "An Efficient Algorithm for Globally Minimizing a Quadratic Function under Convex Quadratic Constraints." *Math. Programming Ser. A* 87(3), 401–426.

An, Le Thi Hoai. (2003). "Solving Large Scale Molecular Distance Geometry Problem by a Smoothing Technique via the Gaussian Transform an DC Programming." *J. Global Optim.* 27, 375–397.

An, Le Thi Hoai and Pham Dinh Tao. (1997). "Solving a Class of Linearly Constrained Indefinite Quadratic Problems by DC Algorithms." *J. Global Optim.* 11, 253–285.

An, Le Thi Hoai, Pham Dinh Tao, and Le Dung Muu. (1996). "Numerical Solution for Optimization over the Efficient Set by DC Optimization Algorithm." *Oper. Res. Lett.* 19, 117–128.

An, Le Thi Hoai and Pham Dinh Tao. (1998). "A Branch-and-Bound Method via DC Optimization Algorithm and Ellipsoidal Techniques for Box Constrained Nonconvex Quadratic Programming Problems." *J. Global Optim.* 13, 171–206.

An, Le Thi Hoai, Pham Dinh Tao, and Le Dung Muu. (1998). "A Combined DC Optimization – Ellipsoidal Branch-and-Bound Algorithm for Solving Nonconvex Quadratic Programming Problems." *J. Combin. Optim.* 2(1), 9–28.

An, Le Thi Hoai, Pham Dinh Tao, and Le Dung Muu. (1999). "Exact Penalty in DC Programming." *Vietnam J. Math.* 27(2), 169–178.

An, Le Thi Hoai and Pham Dinh Tao. (1999). "DCA Revisited and DC Models of Real World Nonconvex Optimization Problems." Technical Report, LMI, Insa-Rouen.

An, Le Thi Hoai and Pham Dinh Tao. (2000a). "DC Programming Approach for Large-Scale Molecular Optimization via the General Distance Geometry Problem." In *Optimization in Computational Chemistry and Molecular Biology: Local and Global Approaches*. Nonconvex Optimization and Its Applications, Vol. 40. Kluwer Academic, pp. 301–339.

An, Le Thi Hoai and Pham Dinh Tao. (2000b). "Large Scale Molecular Conformation via the Exact Distance Geometry Problem." In *Optimization*. Lecture Notes in Economics and Mathematical Systems, Vol. 481. Heidelberg: Springer, pp. 260–277.

An, Le Thi Hoai and Pham Dinh Tao. (2001a). "A Continuous Approach for Large-Scale Linearly Constrained Quadratic Zero-One Programming." *Optim.* 50(1,2), 93–120.

An, Le Thi Hoai and Pham Dinh Tao. (2001b). "DC Optimization Approaches via Markov Models for Restoration of Signals (1-D) and (2-D)." In *Advances in Convex Analysis and Global Optimization*. Nonconvex Optimization and Its Applications, Vol. 54. Kluwer Academic, pp. 300–317.

An, Le Thi Hoai and Pham Dinh Tao. (2001c). "DC Programming Approach and Solution Algorithm to the Multidimensional Scaling Problem." In *From Local to Global Optimization*. Nonconvex Optimization and Its Applications, Vol. 53. Kluwer Academic, pp. 231–276.

An, Le Thi Hoai and Pham Dinh Tao. (2002a). "DC Programming Approach for Multicommodity Network Optimization Problems with Step Increasing Cost Functions." *J. Global Optim.* 22(1–4), 204–233. Special Issue, dedicated to Professor R. Horst on the occasion of his 60th birthday.

An, Le Thi Hoai and Pham Dinh Tao. (2002b). "DC Programming. Theory, Algorithms, Applications: The State of the Art." In *First International Workshop on Global Constrained Optimization and Constraint Satisfaction*. Valbonne-Sophia Antipolis, France, October 2–4, 2002. Research Report, 28 pages. Laboratory of Modeling, Optimization & Operations Research, Insa-Rouen, France (2002).

An, Le Thi Hoai and Pham Dinh Tao. (2003a). "Large Scale Global Molecular Optimization from Exact Distance Matrices by a DC Optimization Approach." *SIAM J. Optim.* 14(1), 77–114.

An, Le Thi Hoai and Pham Dinh Tao. (2003b). "A New Algorithm for Solving Large Scale Molecular Distance Geometry Problems, Applied Optimization." In *Hight Performance Algorithms and Software for Nonlinear Optimization*. Kluwer Academic, pp. 276–296.

An, Le Thi Hoai, Pham Dinh Tao, and Nguyen Van Thoai. (2002). "Combination between Local and Global Methods for Solving an Optimization Problem over the Efficient Set." *European J. Oper. Res.* 142, 257–270.

An, Le Thi Hoai, Pham Dinh Tao, and Le Dung Muu. (2003a). "Simplicially Constrained DC Optimization over the Efficient Set and Weakly Efficient Sets." *J. Optim. Theory Appl.* 117(3), 503–531.

An, Le Thi Hoai, Pham Dinh Tao, and Le Dung Muu. (2003b). "A DC Optimization Approach to Mathematical Programming Problems with Linear Equilibrium Constraints." Submitted.

An, Le Thi Hoai, Pham Dinh Tao, and Dinh Nho Hao. (2002). "Towards Tikhonov Regularization of Nonlinear Ill-Posed Problems: A DC Programming Approach." *C.R. Acad. Sci. Paris Ser. I* 335, 1073–1078.

An, Le Thi Hoai, Pham Dinh Tao, and Dinh Nho Hao. (2003). "Solving Inverse Problems for an Elliptic Equations by D.C. (Difference of Convex Functions) Programming." *J. Global Optim.* 25(4), 407–423.

An, Le Thi Hoai, Pham Dinh Tao, and Dinh Nho Hao. (2004a). "On the Ill-Posedness of the Trust Region Subproblem." *J. Inverse Ill-Posed Problems*, to appear.

An, Le Thi Hoai, Pham Dinh Tao, and Dinh Nho Hao. (2004b). "DC Programming Approach to Tikhonov Regularization for Nonlinear Ill-Posed Problems." Submitted.

Conn, A.R., N.I.M. Gould, and P.L. Toint. (2000). *Trust Region Methods*. MPS–SIAM Series on Optimization.

Hiriat Urruty, J.B. (1989). "Conditions nécessaires et suffisantes d'optimalité globule en optimisation de différences de deux fonctions convexes." *C.R. Acad. Sci. Paris Ser. I* 309, 459–462.

Hiriat Urruty, J.B. and C. Lemarechal. (1993). *Convex Analysis and Minimization Algorithms*. Berlin: Springer.

Horst, R., P.M. Pardalos, and Nguyen Van Thoai. (1995). *Introduction to Global Optimization*. Kluwer Academic.

Horst, R. and Nguyen Van Thoai. (1999). “DC Programming: Overview.” *J. Optim. Theory Appl.* 2103(1), 1–43.

Horst, R. and Hoang Tuy. (1996). *Global Optimization (Deterministic Approaches)*, 3rd ed. Springer.

Konno, H., Phan Thien Thach and Hoang Tuy. (1999). *Optimization on Low Rank Nonconvex Structures*. Kluwer Academic.

Laurent, P.J. (1972). *Approximation et optimisation*. Paris: Hermann.

Mahey, P. and Pham Dinh Tao. (1993). “Partial Regularization of the Sum of Two Maximal Monotone Operators.” *Math. Modell. Numer. Anal. (M^2AN)* 27, 375–395.

Mahey, P. and Pham Dinh Tao. (1995). “Proximal Decomposition of the Graph of Maximal Monotone Operator.” *SIAM J. Optim.* 5, 454–468.

More, J.J. and Z. Wu. (1997). “Global Continuation for Distance Geometry Problems.” *SIAM J. Optim.* 8, 814–836.

More, J.J. and Z. Wu. (1996). “Distance Geometry Optimization for Protein Structures.” Preprint MCS-P628-1296, Argonne National Laboratory, Argonne, IL.

Moreau, J.J. (1965). “Proximité et dualité dans un espace Hilbertien.” *Bull. Soc. Math. France* 93, 273–299.

Pardalos, P.M. and J.B. Rosen. (1987). *Constrained Global Optimization: Algorithms and Applications*, Lecture Notes in Computer Sciences, Vol. 268. Springer.

Pschenichny, B.N. (1971). *Contrôle optimal et jeux differentiels*. Rocquencourt: Cahiers de l’IRIA.

Rockafellar, R.T. (1970). *Convex Analysis*. Princeton: Princeton University.

Rockafellar, R.T. (1976). “Monotone Operators and the Proximal Point Algorithm.” *SIAM J. Control Optim.* 14, 877–898.

Tao, Pham Dinh. (1975). “Elements homoduaux relatifs à un couple de normes (φ, ψ). Applications au calcul de $S_{\varphi\psi}(A)$.” Technical Report, Grenoble.

Tao, Pham Dinh. (1976). “Calcul du maximum d’une forme quadratique définie positive sur la boule unité de la norme du max.” Technical Report, Grenoble.

Tao, Pham Dinh. (1981). “Contribution à la théorie de normes et ses applications à l’analyse numérique.” Thèse de Doctorat d’Etat Es Science, Université Joseph Fourier, Grenoble.

Tao, Pham Dinh. (1984). “Convergence of Subgradient Method for Computing the Bound Norm of Matrices.” *Linear Algebra Appl.* 62, 163–182.

Tao, Pham Dinh. (1985). “Algorithmes de calcul d’une forme quadratique sur la boule unité de la norme maximum.” *Numer. Math.* 45, 377–440.

Tao, Pham Dinh. (1986). *Algorithms for Solving a Class of Non Convex Optimization Problems. Methods of Subgradients*. Fermat Days 85. Mathematics for Optimization, Elsevier Science Publishers, B.V. North-Holland.

Tao, Pham Dinh. (1988). “Duality in D.C. (Difference of Convex Functions) Optimization. Subgradient Methods.” In *Trends in Mathematical Optimization*, International Series of Numerical Mathematics, Vol. 84. Birkhäuser, pp. 277–293.

Tao, Pham Dinh and Le Thi Hoai An. (1994). “Stabilité de la dualité lagrangienne en optimisation DC (différence de deux fonctions convexes).” *C.R. Acad. Sci. Paris Sér. I* 318, 379–384.

Tao, Pham Dinh and Le Thi Hoai An. (1995). “Lagrangian Stability and Global Optimality in Nonconvex Quadratic Minimization over Euclidean Balls and Spheres.” *J. Convex Anal.* 2, 263–276.

Tao, Pham Dinh and Le Thi Hoai An. (1998). “DC Optimization Algorithms for Solving the Trust Region Subproblem.” *SIAM J. Optim.* 8, 476–505.

Tao, Pham Dinh and Le Thi Hoai An. (1997). “Convex Analysis Approach to DC Programming: Theory, Algorithms and Applications.” *Acta Math. Vietnamica* 22(1), 287–355.

Thach, Phan Thien. (1993a). “DC Sets, D.C. Functions and Nonlinear Equations.” *Math. Programming* 58, 415–428.

Thach, Phan Thien. (1993b). “Global Optimality Criteria and Duality with Zero Gap in Nonconvex Optimization Problems.” *SIAM J. Math. Anal.* 24, 1537–1556.

Thach, Phan Thien. (1994). “A Nonconvex Duality with Zero Gap and Applications.” *SIAM J. Optim.* 4, 44–64.

Thach, Phan Thien. (1996). “On the Degree and Separability of Nonconvexity and Applications to Optimization Problems.” *Math. Programming* 77, 23–47.

Toland, J.F. (1978). “Duality in Nonconvex Optimization.” *J. Math. Anal. Appl.* 66, 399–415.

Toland, J.F. (1979). “On Subdifferential Calculus and Duality in Nonconvex Optimization.” *Bull. Soc. Math. France*, Mémoire 60, 177–183.

Tuy, Hoang. (1964). “Concave Programming under Linear Constraints.” *Translated Soviet Mathematics* 5, 1437–1440.

Tuy, Hoang. (1998). *Convex Analysis and Global Optimization*. Kluwer Academic.

Tuy, Hoang. (2000). “Monotonic Optimization: Problems and Solution Approaches.” *SIAM J. Optim.* 2, 464–494.

Wu, Z. (1996). “The Effective Energy Transformation Scheme as a Special Continuation Approach to Global Optimization with Application to Molecular Conformation.” *SIAM J. Optim.* 6, 748–768.