ACTA MATHEMATICA VIETNAMICA
Volume 22, Number 1, 1997, pp. 289–355

CONVEX ANALYSIS APPROACH TO D. C. PROGRAMMING: THEORY, ALGORITHMS AND APPLICATIONS

PHAM DINH TAO AND LE THI HOAI AN

Dedicated to Hoang Tuy on the occasion of his seventieth birthday

Abstract. This paper is devoted to a thorough study on convex analysis approach to d.c. (difference of convex functions) programming and gives the State of the Art. Main results about d.c. duality, local and global optimalities in d.c. programming are presented. These materials constitute the basis of the DCA (d.c. algorithms). Its convergence properties have been tackled in detail, especially in d.c. polyhedral programming where it has finite convergence. Exact penalty, Lagrangian duality without gap, and regularization techniques have beeen studied to find appropriate d.c. decompositions and to improve consequently the DCA. Finally we present the application of the DCA to solving a lot of important real-life d.c. programs.

1. Introduction

In recent years there has been a very active research in the following classes of nonconvex and nondifferentiable optimization problems

(1) $\sup\{f(x) : x \in C\},$: where f and C are convex,

(2) $\inf\{g(x) - h(x) : x \in I\!R^n\},$: where g, h are convex,

(3) $\inf\{g(x) - h(x) : x \in C, f_1(x) - f_2(x) \leq 0\},$: where g, h, f_1, f_2 and C are convex.

The main incentive comes from linear algebra, numerical analysis and operations research. Problem (1) is a special case of Problem (2) with $g = \chi_C$, the indicator function of C and $h = f$, while the latter (resp. Problem (3)) can be equivalently transformed into the form of (1) (resp. (2)) by

Received January 22, 1997; in revised form May 7, 1997
Key words. d.c. programming, local and global optimalities, DCA, polyhedral d.c. programming, exact penalty, Lagrangian duality without gap, regularization techniques, escaping procedure, trust region subproblem, nonconvex quadratic programming, multidimensional scaling problem, optimization over the efficient set problem, linear and nonlinear complementarity problems.

introducing an additional scalar variable (resp. via exact penalty relative to the d.c. constraint $f_1(x) - f_2(x) \leq 0$, see [6], [7]). Though the complexity increases from (1) to (3), the solution of one of them implies the solution of the two others. Problem (2) is called a d.c. program whose particular structure has been permitting a good deal of development both in qualitative and quantitative studies (see, e.g., [1]-[8], [41], [44], [70]-[85], [104], [105], [111]-[113]).

There are two different but complementary approaches, we can even say two schools, in d.c. programming.

(i) *Combinatorial approach to global continuous optimization:* it is older and inspired by tools and methods developed in the combinatorial optimization. Nevertheless, new concepts and new methods have been introduced as one works in a continuous approach (see [44] and references therein). People recognizes that it was Hoang Tuy who has incidentally put forward by his pioneering paper in 1964 ([110]) the new global optimization concerning convex maximization over a polyhedral convex set. During the last decade tremendous progress has been made especially in the computational aspect. One can now globally solve larger d.c. programs, especially large-scale low rank nonconvex problems ([47]). However, most robust and efficient global algorithms for solving d.c. programs actually do not meet the expected desire: solving real life d.c. programs in their true dimension.

(ii) *Convex analysis approach to nonconvex programming:* this approach has been less worked out than the preceding one. The explanation can be found in the fact: there is a mine of real world d.c. programs to be solved in the combinatorial optimization. In fact, most real life problems are of nonconvex nature. This approach seemed to take rise in the works of the first author on the computation of bound-norms of matrices (i.e., maximizing a semi-norm on the unit ball of a norm) in 1975 ([70]-[73]). There the convexity of the two d.c. components g and h of the objective function has been used to develop appropriate tools. The main idea in this approach is the use of the d.c. duality, which has been first studied by Toland in 1979 ([108]) who generalized, in a very elegant and natural way, the just mentioned works on convex maximization programming. In contrast with the first approach where many global algorithms have been studied, there have been a very few algorithms for solving d.c. programs in the convex analysis approach. Let us name some people who made contributions to this approach: Pham Dinh Tao, J.F. Toland, J.B. Hiriart-Urruty, Phan Thien Thach, Le Thi Hoai An. Among them J.F. Toland and J.B. Hiriart-Urruty are interested in the theoretical framework only. Here we are interested in local and global optimalities, relationships between local and global solutions to primal and dual d.c. programs and solution algorithms. D.c. algorithms (DCA) based on the duality and local optimality conditions in d.c. optimization has been introduced by Pham Dinh Tao in [76]. Important developments and improvements for DCA from both theoretical and numerical aspects have been completed after the works by the authors of this paper [1], [2], [6], [78]-[83] appeared. These algorithms allow to handle certain classes of large-scale d.c. programs ([1]-[6], [79], [83], [84]). Due to their local character they cannot guarantee the globality of computed solutions for general d.c. programs. In general, DCA converge to a local solution, however we observed from our numerous experiments that DCA converge quite often to a global one ([1]-[6], [79], [83], [84]).

The d.c. objective function (of a d.c. program) has infinitely many d.c. decompositions which may have an important influence on the qualities (robustness, stability, rate of convergence and globability of sought solutions) of DCA. So, it is of particular interest to obtain various equivalent d.c. forms for the primal and dual problems. The Lagrangian duality without gap, exact penalty in d.c. optimization ([1], [7], [78], [82]) and regularization techniques partly answer this concern. In application, regularization techniques using the kernel $\frac{\lambda}{2}\|\cdot\|^2$ and inf-convolution may provide interesting d.c. decompositions of objective functions for DCA ([1], [80], [81]). Furthermore, it is worth noting that by using conjointly suitable d.c. decompositions of convex functions and proximal regularization techniques ([1], [80], [81]) we can obtain the proximal point algorithm ([58], [94]) and the Goldstein-Levitin-Polyak subgradient projection method ([89]) as particular cases of DCA. It would be interesting to find conditions on the choice of the d.c. decomposition and the initial point to ensure the convergence of DCA to a global solution. Polyhedral d.c. optimization occurs when either g or h is polyhedral convex. This class of polyhedral d.c. programs, which plays a key role in nonconvex optimization, possesses worthy properties from both theoretical and computational viewpoints, as necessary and sufficient local optimality conditions, and finite convergence for DCA... In practice, DCA have been successfully applied to many large-scale d.c. optimization problems and proved to be more robust and efficient than related standard methods ([1]-[6], [79], [83], [84]).

The paper is organized as follows. In the next section we study the d.c. optimization framework: duality, local and global optimality. The key point which makes a unified and deep d.c. optimization theory possible relies on the particular structure of the objective function to be minimized on $\mathbb{R}^n$:

$$f(x) = g(x) - h(x)$$

with g and h being convex on $\mathbb{R}^n$. One works then actually with the powerful convex analysis tools applied to the d.c. components g and h of the d.c. function f. The d.c. duality associates a primal d.c. program with a dual d.c. one (with the help of the functional conjugate notion) and states relationships between them. More precisely, the d.c. duality is built on the fundamental feature of proper lower semi-continuous convex functions: such functions $\theta(x)$ are the supremum w.r.t. $y \in \mathbb{R}^n$ of the affine functions

$$\langle x, y\rangle - \theta^*(y).$$

Thanks to a symmetry in the d.c. duality (the bidual d.c. program is exactly the primal one) and the d.c. duality transportation of global minimizers, solving a d.c. program implies solving the dual one and vice-versa. Furthermore, it may be very useful when one is easier to solve than the other. The equality of the optimal value in the primal and dual d.c. programs can be easily translated (with the help of ε-subdifferentials of the d.c. components) in global optimality conditions. These conditions mark the passage from convex optimization to nonconvex optimization, and have been early stated by J.B. Hiriart-Urruty [37] in a more complicated way. They are nice but impractical to use for devising solutions methods to d.c. programs. Local d.c. optimality conditions constitute the basis of DCA. In general, it is not easy to state them as in the global d.c. optimality and at the moment there have been found very few properties which are useful in practice.

We shall present therefore in Section 2 the most significative results; most of them are new. In particular, we give there the new elegant property (ii) of Theorem 2 concerning sufficient local d.c. optimality conditions and their consequences, e.g., the d.c. duality transportation of local minimizers. The latter is very helpful to establish relationships between local minimizers of primal and dual d.c. programs. All these results are indispensable to understanding DCA for locally solving primal and dual d.c. programs. The description of DCA and its convergence are presented in Section 3. Section 4 is devoted to polyhedral d.c. programs. Regularization techniques in d.c. programming are studied in Section 5 which emphasizes the role of the proximal regularization in d.c. programming. Some discussions about functions being more convex or less convex and their role in d.c. programming are given in Section 6. Except for the case $h = \frac{\lambda}{2}\|\cdot\|^2 \in \Gamma_o(X)$, one does not know if g^* is less convex than h^* whenever g is more convex than h. A response (positive or negative) to this question or/and, more thoroughtly, a characterization of functions $g, h \in \Gamma_o(X)$ satisfying such a property are particularly important for d.c. programming. They enable us to formulate "false" d.c. programs (i.e., convex programs in reality) whose dual d.c. programs are truly nonconvex. It is surprising enough that the above question remains completely open. We present in Section 7 the relation between DCA and the proximal point algorithm (PPA) (resp. the Goldstein-Levitin-Polyak gradient projection algorithm) in convex programming. Section 8 is related to exact penalty, Lagrangian duality without gap and dimensional reduction in d.c. programming. Last but not least, Section 9 is devoted to the applications of DCA to solving a lot of important real-life d.c. programs, for each of them an appropriate d.c. decomposition and the corresponding DCA are presented. Numerical experiments proved that the DCA applied to these problems converges to a local solution which is quite often a global one. They showed at the same time the robustness and the efficiency of the DCA with respect to related standard methods.

(i) The trust region subproblem ([79], [83]).

The DCA for the trust region subproblem is quite different from related standard algorithms ([19], [63], [77], [91], [96], [100]). It indifferently treats both the normal and hard cases and requires matrix-vectors products only. A very simple numerical procedure has been introduced in [83] to find a new point with smaller value of the quadratic objective function in case of non globality. According to Theorem 4 of [83], the DCA with at most $2m+2$ restartings (m being the number of distinct negative eigenvalue of the matrix being considered) converges to a global solution of the trust region subproblem. In practice the DCA rarely has recourse to the restarting procedure. From the computational viewpoint, a lot of our numerical experiments proved the robustness and efficiency of the DCA with respect to other known algorithms, especially in the large-scale setting ([83]).

(ii) The multidimensional scaling problem (MDS) ([84]).

Recently MDS earned particular attention of researchers by its role in semidefinite programming ([61]), the molecule problem ([34]), the protein structure determination problem ([116], [117]) and the protein folding problem ([30]).

As in the trust region subproblem, the Lagrangian duality relative to MDS has zero gap. That leads to quite appropriate d.c. decompositions and simple DCA (requiring matrix-vector products only). In particular, the reference de Leeuw algorithm has been pointed out as DCA corresponding to some d.c. decomposition.

Except for the case where the dissimilarity matrix represents really the Euclidean distances between n objects (for which the optimal value is zero), the graph of the dual objective function ([84]) can be used to checking the globality of solutions computed by DCA. It is worth noting that MDS can be formulated as a parametrized trust region subproblem and a parametrized DCA applied to this problem is exactly a form of the DCA applied directly to MDS.

(iii) Linearly constrained indefinite quadratic programs ([2]).

Linearly constrained indefinite quadratic problems play an important role in global optimization. We study d.c. theory and its local approach to such problems. The DCA efficiently produces local optima and sometimes produces global optima. We also propose a decomposition branch-and-bound method for globally solving these problems.

(iv) A branch-and-bound method via DCA and ellipsoidal technique for box constrained quadratic programs ([3]).

We propose a new branch-and-bound algorithm using a rectangular partition and ellipsoidal technique for minimizing a nonconvex quadratic function with box constraints. The bounding procedures are investigated by the DCA. This is based upon the fact that the application of the DCA to the problems of minimizing a quadratic form over an ellipsoid and/or over a box is efficient. Some details of the computational aspect of the algorithm are reported.

(v) The DCA with an escaping procedure for globally solving nonconvex quadratic programs ([4]).

(vi) D.c. approach for linearly constrained quadratic zero-one programming problems ([5]).

(vii) Optimization over the efficient set problem ([6]).

We use the DCA for (locally) maximizing a concave, a convex or a quadratic function f over the efficient set of a multiple objective convex program. We also propose a decomposition method for globally solving this problem with f concave. Numerical experiences are discussed.

(viii) Linear and nonlinear complementarity problems.

Difference of subdifferentials (of convex functions) complementarity problem ([85]).

2. Duality and optimality for d.c. program

Let the space $X = I\!R^n$ be equipped with the canonical inner product $\langle \cdot, \cdot \rangle$. Thus, the dual space Y of X can be identified with X itself. Denote by $\Gamma_o(X)$ the set of all proper lower semi-continuous convex functions on X. The conjugate function g^* of $g \in \Gamma_o(X)$ is a function belonging to $\Gamma_o(Y)$ and defined by

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

The Euclidean norm of X is denoted by $\|x\| = \langle x, x\rangle^{1/2}$. For a convex set C in X the indicator function of C is denoted by $\chi_C(x)$ which is equal to 0 if $x \in C$ and $+\infty$ otherwise. We shall use the following usual notations of [93]

$$\text{dom } g = \{x \in X : \ g(x) < +\infty\}.$$

For $\epsilon > 0$ and $x^o \in \text{dom } g$, the symbol $\partial_\epsilon g(x^o)$ denotes ϵ-subdifferential of g at x^o, i.e.,

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

while $\partial g(x^o)$ stands for the usual (or exact) subdifferential of g at x^o. Also, $\text{dom } \partial g = \{x \in X : \ \partial g(x) \neq \emptyset\}$ and $\text{range}\, \partial g = \cup\{\partial g(x) : \ x \in \text{dom } \partial g\}$. We adopt in the sequel the convention $+\infty - (+\infty) = +\infty$. A d.c. program is that of the form

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

where g and h belong to $\Gamma_o(X)$.

Such a function f is called d.c. function on X and g, h are called its d.c. components. If g and h are in addition finite on all of X then one says that $f = g - h$ is finite d.c. function on X. The set of d.c. functions (resp. finite d.c. functions) on X is denoted by $\mathcal{DC}(X)$ (resp. $\mathcal{DC}_f(X)$).

Using the definition of conjugate functions, we have

$$\begin{aligned}\alpha &= \inf\{g(x) - h(x) : x \in X\}\\ &= \inf\{g(x) - \sup\{\langle x, y\rangle - h^*(y) : y \in Y\} : x \in X\}\\ &= \inf\{\beta(y) : \ y \in Y\}\end{aligned}$$

with

$$(P_y) \qquad \beta(y) = \inf\{g(x) - (\langle x, y\rangle - h^*(y)) \ : x \in X\}.$$

It is clear that $\beta(y) = h^*(y) - g^*(y)$ if $y \in \operatorname{dom} h^*$, $+\infty$ otherwise. Finally, we state the dual problem

$$\alpha = \inf\{h^*(y) - g^*(y) : y \in \operatorname{dom} h^*\},$$

that is written, according to the above convention, as

$$\text{(D)} \qquad \alpha = \inf\{h^*(y) - g^*(y) : y \in Y\}.$$

We observe the perfect symmetry between primal and dual programs (P) and (D): the dual program to (D) is exactly (P).

Note that the finiteness of α merely implies that

$$\text{(1)} \qquad \operatorname{dom} g \subset \operatorname{dom} h \quad \text{and} \quad \operatorname{dom} h^* \subset \operatorname{dom} g^*.$$

Such inclusions will be assumed throughout the paper.

This d.c. duality was first studied by J.F. Toland ([108]) in a more general framework. It can be considered as a logical generalization of our earlier works concerning convex maximization (see [71] and references therein).

It is worth noting the richness of the sets $\mathcal{DC}(X)$ and $\mathcal{DC}_f(X)$ ([1], [36], [44], [80]):

(i) $\mathcal{DC}_f(X)$ is a subspace containing the class of lower-$\mathcal{C}^2$ functions (f is said to be lower-$\mathcal{C}^2$ if f is locally a supremum of a family of $\mathcal{C}^2$ functions). In particular, $\mathcal{DC}(X)$ contains the space $\mathcal{C}^{1,1}(X)$ of functions whose gradient is locally Lipschitzian on X.

(ii) Under some caution we can say that $\mathcal{DC}(X)$ is the subspace generated by the convex cone $\Gamma_o(X)$: $\mathcal{DC}(X) = \Gamma_o(X) - \Gamma_o(X)$. This relation marks the passage from convex optimization to nonconvex optimization and also indicates that $\mathcal{DC}(X)$ constitutes a minimal realistic extension of $\Gamma_o(X)$.

(iii) $\mathcal{DC}_f(X)$ is closed under all the operations usually considered in optimization. In particular, a linear combination of $f_i \in \mathcal{DC}(X)$ belongs to $\mathcal{DC}_f(X)$, a finite supremum of d.c. functions is d.c..

Let us give below some useful formulations relative to these results. If $f_i \in \mathcal{DC}(X)$, $f_i = g_i - h_i$ for $i = 1, \dots, m$, then

$$\min_i f_i = \sum_{i=1}^{m} g_i - \max_i \Big[h_i + \sum_{j=1, j\neq i}^{m} g_j\Big].$$

If $f = g - h$, then

$$f^+ = \max(g,h) - h, \quad f^- = \max(g,h) - g, \quad |f| = 2\max(g,h) - (g+h).$$

Proposition 1 ([36], [80]). *Every nonnegative d.c. function $f = g - h$ (g, $h \in \Gamma_o(X)$) admits a nonnegative d.c. decomposition, i.e., $f = g_1 - h_1$ with g_1, h_1 being in $\Gamma_o(X)$ and nonnegative.*

Proof. The functions g_1 and h_1 are defined in [36] by

$$g_1 = g - (\langle b, \cdot\rangle - h^*(b)), \quad h_1 = h - (\langle b, \cdot\rangle - h^*(b))$$

with $b \in \operatorname{dom} h^*$.

The following nonnegative decomposition of f, with λ being a positive number given in [80], is intimately related to the proximal regularization technique (Section 5)

$$g_1 = g + \frac{\lambda}{2}\|\cdot\|^2 + (h + \frac{\lambda}{2}\|\cdot\|^2)^*(0),$$
$$h_1 = h + \frac{\lambda}{2}\|\cdot\|^2 + (h + \frac{\lambda}{2}\|\cdot\|^2)^*(0). \quad \square$$

Remark that Proposition 1 remains true in $\mathcal{DC}(X)$.

More generally, every $f \in \mathcal{DC}(X)$ admits (by using f^+ and f^-) a nonnegative d.c. decomposition. It follows that if f_1, $f_2 \in \mathcal{DC}(X)$, then by taking their nonnegative d.c. decomposition $f_i = g_i - h_i$, $i = 1, 2$, the product $f_1 \cdot f_2$ is a d.c. function ([36])

$$f_1.f_2 = \frac{1}{2}\left[(g_1+g_2)^2 + (h_1+h_2)^2\right] - \frac{1}{2}\left[(g_1+h_1)^2 + (g_2+h_2)^2\right].$$

These result have been extended to d.c. functions [80] as follows: Let $f_i \in \mathcal{DC}(X)$, $f_i = g_i - h_i$ be such that

$$\operatorname{dom} g_i = C \subset \operatorname{dom} h_i \quad \text{for } i = 1, \ldots, m$$

then the effective domain of $f = \min f_i$ is C and $f = g - h$ with $g = \sum_{i=1}^{n} g_i$ and $h = \max_i (h_i + \sum_{j=1, j \neq i}^{n} g_j)$.

If $f = g - h$ with $\operatorname{dom} g = \operatorname{dom} h = C$, then • C is the effective domain of the next d.c. functions

$$f^+ = \max(g, h) - h, \quad f^- = \max(g, h) - g,$$

$$|f| = 2\max(g, h) - (g + h).$$

• f admits a nonnegative d.c. decomposition $f = g_1 - h_1$ with $\operatorname{dom} g_1 = \operatorname{dom} h_1 = C$.

If $f_i = g_i - h_i$ with $\operatorname{dom} g_i = \operatorname{dom} h_i = C$ for $i = 1, 2$, then

∘ they admit a nonnegative d.c. decomposition $f_i = \overline{g}_i - \overline{h}_i$ with $\operatorname{dom} \overline{g}_i = \operatorname{dom} \overline{h}_i = C$ for $i = 1, 2$;

∘ the product $f_1 \cdot f_2$ is a d.c. function

$$f_1 \cdot f_2 = \frac{1}{2}\Big[(\overline{g}_1 + \overline{g}_2)^2 + (\overline{h}_1 + \overline{h}_2)^2\Big] - \frac{1}{2}\Big[(\overline{g}_1 + \overline{h}_1)^2 + (\overline{g}_2 + \overline{h}_2)^2\Big]$$

with $\operatorname{dom} f_1 \cdot f_2 = C$.

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

$$g(x^*) - h(x^*) \leq g(x) - h(x), \quad \forall x \in U. \tag{2}$$

Under the convention $+\infty - (+\infty) = +\infty$, the property (2) is equivalent to $g(x^*) - h(x^*) \leq g(x) - h(x)$, $\forall x \in U \cap \operatorname{dom} g$.

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

$\operatorname{int} S$ denotes the interior of the set S in X. Moreover, if S is convex, then $\operatorname{ri} S$ stands for the relative interior of S.

A convex function f on X is said to be *essentially differentiable* if it satisfies the following three conditions ([93]) :

(i) $C = \operatorname{int}\,(\operatorname{dom}\, f) \neq \emptyset$,

(ii) f is differentiable on C,

(iii) $\lim_{k\to\infty} \|\nabla f(x^k)\| = +\infty$ for every sequence $\{x^k\}$ which converges to a point at the boundary of C.

Let $\rho \geq 0$ and C be a convex subset of X. One says that a function $\theta : C \longrightarrow I\!R \cup \{+\infty\}$ is ρ-convex if

$$\theta[\lambda x + (1-\lambda)x'] \leq \lambda\theta(x) + (1-\lambda)\theta(x') - \frac{\lambda(1-\lambda)}{2}\rho\|x - x'\|^2,$$

$$\forall \lambda \in]0,1[, \ \forall x, x' \in C.$$

It amounts to say 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:

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

One say that θ is *strongly convex* on C if $\rho(\theta, C) > 0$.

For $f \in \Gamma_o(X)$ and $\lambda > 0$ the Moreau-Yosida regularization of f with parameter λ, denoted by f_λ, is the inf-convolution of f and $\frac{1}{2\lambda}\|\cdot\|^2$. The function f_λ is continuously differentiable, underapproximates f without changing the set of minimizers and $(f_\lambda)_\mu = f_{\lambda+\mu}$. More precisely, $\nabla f_\lambda = \frac{1}{\lambda}[I - (I + \lambda\partial f)^{-1}]$ is Lipschitzian with ratio $\frac{1}{\lambda}$. The operator $(I + \lambda\partial f)^{-1}$ is called the proximal mapping associated to λf ([94]).

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

$$\mathcal{P}_\ell = \{x^* \in X : \partial h(x^*) \subset \partial g(x^*)\},$$
$$\mathcal{D}_\ell = \{y^* \in Y : \partial g^*(y^*) \subset \partial h^*(y^*)\}.$$

We present below some fundamental results of d.c. optimization which constitute the basis of DCA presented in Subsection 3.3.

2.1 Duality and global optimality for d.c. optimization

Theorem 1. *Let $\mathcal{P}$ and $\mathcal{D}$ be the solution sets of problems (P) and (D), respectively. Then*

(i) *$x \in \mathcal{P}$ if and only if $\partial_\epsilon h(x) \subset \partial_\epsilon g(x) \ \ \forall \epsilon > 0$.*

(ii) *Dually, $y \in \mathcal{D}$ if and only if $\partial_\epsilon g^*(y) \subset \partial_\epsilon h^*(y) \ \ \forall \epsilon > 0$.*

(iii) *$\cup\{\partial h(x) : x \in \mathcal{P}\} \subset \mathcal{D} \subset \operatorname{dom} h^*$.*

The first inclusion becomes equality if g^* is subdifferentiable in $\mathcal{D}$ (in particular if $\mathcal{D} \subset \mathrm{ri}\,(\mathrm{dom}\, g^*)$ or if g^* is subdifferentiable in $\mathrm{dom}\, h^*$).

In this case $\mathcal{D} \subset (\mathrm{dom}\, \partial g^* \cap \mathrm{dom}\, \partial h^*)$.

(iv) $\cup\{\partial g^*(y) : y \in \mathcal{D}\} \subset \mathcal{P} \subset \mathrm{dom}\, g$.

The first inclusion becomes equality if h is subdifferentiable in $\mathcal{P}$ (in particular if $\mathcal{P} \subset \mathrm{ri}\,(\mathrm{dom}\, h)$ or if h is subdifferentiable in $\mathrm{dom}\, g$). In this case $\mathcal{P} \subset (\mathrm{dom}\, \partial g \cap \mathrm{dom}\, \partial h)$.

The relationship between primal and dual solutions: $\cup\{\partial h(x) : x \in \mathcal{P}\} \subset \mathcal{D}$ and $\cup\{\partial g^*(y) : y \in \mathcal{D}\} \subset \mathcal{P}$ is due to J.F. Toland ([108])

in the general context of duality principle dealing with linear vector spaces in separating duality. A direct proof of the results (except for Properties (i) and (ii)) of Theorem 1, based on the theory of subdifferential for convex functions is given in [1], [76], [80]. The properties (i) and (ii) have been first established by J.B. Hiriart-Urruty ([37]). His proof (based on his earlier work concerning the behaviour of the ϵ-directional derivative of a convex function as a function of the parameters ϵ) is rather complicated. The following proof of these properties is very simple and well suited to our d.c. duality framework ([1], [76], [80]). In fact, they are nothing but a geometrical translation of the equality of the optimal value in the primal and dual d.c. programs (P) and (D).

Indeed, in virtue of the d.c. duality, $x^* \in \mathcal{P}$ if and only if $x^* \in \mathrm{dom}\, g$ and

$$g(x^*) - h(x^*) \leq h^*(y) - g^*(y), \quad \forall y \in \mathrm{dom}\, h^*, \tag{4}$$

i.e.,

$$g(x^*) + g^*(y) \leq h(x^*) + h^*(y), \quad \forall y \in \mathrm{dom}\, h^*. \tag{5}$$

On the other hand, for $x^* \in \mathrm{dom}\, h$, the property $\partial_\epsilon h(x^*) \subset \partial_\epsilon g(x^*)\ \forall \epsilon > 0$ is, by definition, equivalent to

$$\forall \epsilon > 0,\ \langle x^*, y\rangle + \epsilon \geq h(x^*) + h^*(y) \Rightarrow \langle x^*, y\rangle + \epsilon \geq g(x^*) + g^*(y). \tag{6}$$

It is easy to see the equivalence between (5) and (6), and property (i) thus is proved.

The global optimality condition in (i) is difficult to use for deriving solution methods to problem (P). The algorithms DCA which will be described in Subsection 3.1. are based on local optimality conditions. The relations (ii) and (iv) indicate that solving the primal d.c. program (P) implies solving the dual d.c. program (D) and vice-versa. It may be useful if one of them is easier to solve than the other.

2.2. Duality and local optimality conditions for d.c. optimization

Theorem 2. (i) *If x^* is a local minimizer of $g-h$, then $x^* \in \mathcal{P}_\ell$.*

(ii) *Let x^* be a critical point of $g-h$ and $y^* \in \partial g(x^*) \cap \partial h(x^*)$. Let U be a neighbourhood of x^* such that $U \cap \operatorname{dom} g \subset \operatorname{dom} \partial h$.*

If for any $x \in U \cap \operatorname{dom} g$ there is $y \in \partial h(x)$ such that $h^*(y) - g^*(y) \geq h^*(y^*) - g^*(y^*)$, then x^* is a local minimizer of $g-h$. More precisely,

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

Property (i) is well known ([1], [37], [76], [108]). To facilitate the reading we give below a short proof for it.

Property (ii) is new, it establishes an interesting sufficient condition (dealing with the d.c. duality) for the local d.c. optimality.

Proof. (i) If x^* is a local minimizer of $g-h$, then there exists a neighbourhood of x^* such that

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

Hence, for $y^* \in \partial h(x^*)$ we have $g(x) - g(x^*) \geq \langle x - x^*, y^* \rangle$, $\forall x \in U \cap \operatorname{dom} g$. The convexity of g then implies that $y^* \in \partial g(x^*)$.

(ii) The condition $y^* \in \partial g(x^*) \cap \partial h(x^*)$ implies $g(x^*) + g^*(y^*) = \langle x^*, y^* \rangle = h(x^*) + h^*(y^*)$. Hence,

$$g(x^*) - h(x^*) = h^*(y^*) - g^*(y^*). \tag{7}$$

For any $x \in U \cap \operatorname{dom} g$, by assumption, there is $y \in \partial h(x)$ such that

$$h^*(y) - g^*(y) \geq h^*(y^*) - g^*(y^*). \tag{8}$$

On the other hand, we have $h(x) + h^*(y) = \langle x, y \rangle \leq g(x) + g^*(y)$. Hence,

$$g(x) - h(x) \geq h^*(y) - g^*(y). \tag{9}$$

Combining (7), (8), (9), we get $g(x) - h(x) \geq g(x^*) - h(x^*)$, $\forall x \in U \cap \operatorname{dom} g$.

□

Corollary 1 (sufficient local optimality). *Let x^* be a point that admits a neighbourhood U such that $\partial h(x) \cap \partial g(x^*) \neq \emptyset$, $\forall x \in U \cap \operatorname{dom} g$. Then x^* is a local minimizer of $g-h$. More precisely, $g(x) - h(x) \geq g(x^*) - h(x^*)$, $\forall x \in U \cap \operatorname{dom} g$.*

Proof. Let $x \in U \cap \operatorname{dom} g$ and let $y \in \partial h(x) \cap \partial g(x^*)$. Since $y \in \partial h(x)$ we have $h(x) + h^*(y) = \langle x, y\rangle \leq g(x) + g^*(y)$. So $g(x) - h(x) \geq h^*(y) - g^*(y)$. Similarly, $y \in \partial g(x^*)$ implies that $g(x^*) + g^*(y) = \langle x^*, y\rangle \leq h(x^*) + h^*(y)$. Then, $h^*(y) - g^*(y) \geq g(x^*) - h(x^*)$.

If $y^* \in \partial h(x^*) \cap \partial g(x^*)$, then $g(x^*) + g^*(y^*) = \langle x^*, y^*\rangle = h(x^*) + h^*(y^*)$.

Hence, $g(x^*) - h(x^*) = h^*(y^*) - g^*(y^*)$. The assumptions of (ii) in Theorem 2 are fulfilled. Thus, the proof is complete. □

A function $\theta \in \Gamma_o(X)$ is said to be polyhedral convex if ([93])

$$\theta(x) = \max\{\langle a_i, x\rangle - \alpha_i : \ i = 1, \dots, m\} + \chi_C(x) \quad \forall x \in X,$$

where C is a nonempty polyhedral convex set in X.

Polyhedral d.c. optimization occurs when either g or h is polyhedral convex. This class of d.c. optimization problems, which is frequently encountered in practice, enjoys interesting properties (from both theoretical and practical viewpoints) concerning local optimality and the convergence of DCA, and has been extensively developed in [1], [81].

Corollary 1 can be used to prove the following well known result due to C. Michelot (in the case where g, h belonging to $\Gamma_o(X)$ are finite on the whole X and generalized by us ([1], [81]) to the case of arbitrary g and h belonging to $\Gamma_o(X)$): the converse of property (i) of Theorem 2 in case h is polyhedral convex.

Corollary 2 (sufficient strict local optimality). *If* $x^* \in int(\operatorname{dom} h)$ *verifies* $\partial h(x^*) \subset \operatorname{int}(\partial g(x^*))$, *then* x^* *is a strict local minimizer of* $g - h$.

Proof. From the upper semicontinuity of the operator ∂h at $x^* \in \operatorname{int}(\operatorname{dom} h)$ ([93]) it follows that for any open set O containing $\partial h(x^*)$ there is a neighbourhood U of x^* such that $\partial h(x) \subset O$, $\forall x \in U$. Hence, by letting $O = \operatorname{int}(\partial g(x^*))$ and taking Corollary 1 into account, we have x^* is a local minimizer of $g - h$. But x^* is actually a strict local minimizer of $g - h$. Indeed, since $\partial h(x)$ is compact for $x \in V = U \cap \operatorname{int}(\operatorname{dom} h)$, we have $\forall x \in V \ \ \exists \epsilon(x) > 0$ such that $\partial h(x) + \epsilon(x)B \subset O$ (B being the closed unit ball of the Euclidean norm).

Now let $x \in V \backslash \{x^*\}$ and $y \in \partial h(x)$. Then

$$\begin{aligned} g(x) - g(x^*) &\geq \langle x - x^*, y + \frac{\epsilon(x)}{\|x - x^*\|}(x - x^*)\rangle \\ &= \epsilon(x)\|x - x^*\| + \langle x - x^*, y\rangle \\ &\geq \epsilon(x)\|x - x^*\| + h(x) - h(x^*). \end{aligned}$$

The proof is complete. □

It may happen that the dual d.c. program (D) is easier to locally solve than the primal d.c. program (P). So it is useful to state results relative to the d.c. duality transportation of local minimizers. Paradoxically, such a result is more complicated than the d.c. duality transportation of global minimizers in Theorem 1.

Corollary 3 (d.c. duality transportation of a local minimizer). *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 U such that $g(x) - h(x) \geq g(x^*) - h(x^*)$, $\forall x \in U \cap \operatorname{dom} g$). If*

$$(10) \qquad y^* \in \operatorname{int}(\operatorname{dom} g^*) \text{ and } \quad \partial g^*(y^*) \subset U$$

((10) holds if g^* is differentiable at y^*), then y^* is a local minimizer of $h^* - g^*$.

Proof. According to (i) of Theorem 2 we have $y^* \in \partial h(x^*) \subset \partial g(x^*)$. So, $x^* \in \partial g^*(y^*) \cap \partial h^*(y^*)$. Under the assumption (10) and the upper semicontinuity of ∂g^*, y^* admits a neighbourhood $V \subset \operatorname{int}(\operatorname{dom} g^*)$ such that ([93]) $\partial g^*(V) \subset U$. More precisely, $\partial g^*(V) \subset (U \cap \operatorname{dom} g)$, since we have ([93]) $\operatorname{range} \partial g^* = \operatorname{dom} \partial g$ and $\operatorname{dom} \partial g \subset \operatorname{dom} g$. Using the dual property (in the d.c. duality) in (ii) of Theorem 2, we deduce that y^* is a local minimizer of $h^* - g^*$.

If g^* is differentiable at y^*, then $x^* = \partial g^*(y^*)$ and we have (10) ([93]). □

By the symmetry of the d.c. duality, Corollary 3 has its corresponding dual part.

Remark 1. This result improves an earlier result of J.F. Toland ([108]) where he assumed that g^* is differentiable on the whole dual space Y. In [1], [81] we have proved that this result remains true if g^* is only essentially differentiable.

3. D.c. algorithm (DCA) for general d.c. programs

3.1. Description of DCA for general d.c. programs

For each fixed $x^* \in X$ we consider the problem

$$(S(x^*)) \qquad \inf\{h^*(y) - g^*(y) : y \in \partial h(x^*)\},$$

which is equivalent to the convex maximization one

$$\inf\{\langle x^*, y\rangle - g^*(y) : y \in \partial h(x^*)\}.$$

Similarly, for each fixed $y^* \in Y$, for duality, we define the problem

$$(T(y^*)) \qquad \inf\{g(x) - h(x) : x \in \partial g^*(y^*)\}.$$

This problem is equivalent to

$$\inf\{\langle x, y^*\rangle - h(x) : x \in \partial g^*(y^*)\}.$$

Let $\mathcal{S}(x^*)$, $\mathcal{T}(y^*)$ denote the solution sets of Problems $(S(x^*))$ and $(T(y^*))$, respectively.

The complete form of DCA is based upon duality of d.c. optimization defined by (P) and (D). It allows approximating a point $(x^*, y^*) \in \mathcal{P}_\ell \times \mathcal{D}_\ell$. From a point $x^o \in \operatorname{dom} g$ given in advance, the algorithm consists of constructing two sequences $\{x^k\}$ and $\{y^k\}$ defined by

$$y^k \in \mathcal{S}(x^k); \quad x^{k+1} \in \mathcal{T}(y^k). \tag{11}$$

The complete DCA can be viewed as a sort of decomposition approach of the primal and dual problems (P), (D). From a practical point of view, although the problems $(S(x^k))$ and $(T(x^k))$ are simpler than (P), (D) (we work in $\partial h(x^k)$ and $\partial g^*(y^k)$ with convex maximization problems), they remain nonconvex programs and thus are still of a difficult task (see Subsection 3.3). In practice the following simplified form of DCA is used:

• **Simplified form of DCA:**

The idea of the simplified DCA is to construct two sequences $\{x^k\}$ and $\{y^k\}$ (candidates to primal and dual solutions) which are easy to calculate and satisfy the following conditions:

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

(ii) Every limit point x^* (resp. y^*) of the sequence $\{x^k\}$ (resp. $\{y^k\}$) is a critical point of $g - h$ (resp. $h^* - g^*$).

These conditions suggest constructing two sequences $\{x^k\}$ and $\{y^k\}$, starting from a given point $x^o \in \operatorname{dom} g$ by setting

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

Interpretation of the simplified DCA:

At each iteration k we do the following:

$$x^k \in \partial g^*(y^{k-1}) \to y^k \in \partial h(x^k)$$
$$= \operatorname{argmin}\{h^*(y) - [g^*(y^{k-1}) + \langle x^k, y - y^{k-1}\rangle] : y \in Y\}, \quad (D_k)$$
$$y^k \in \partial h(x^k) \to x^{k+1} \in \partial g^*(y^k)$$
$$= \operatorname{argmin}\{g(x) - [h(x^k) + \langle x - x^k, y^k\rangle] : x \in X\}. \quad (P_k)$$

Problem (P_k) is a convex program obtained from (P) by replacing h with its affine minorization defined by $y^k \in \partial h(x^k)$. Similarly, the convex problem (D_k) is obtained from (D) by using the affine minorization of g^* defined by $x^k \in \partial g^*(y^{k-1})$. Here we can see the complete symmetry between problems (P_k) and (D_k), and between the sequences $\{x^k\}$ and $\{y^k\}$ relative to the duality of d.c. optimization. The two forms of DCA are identical if g^* and h are essentially differentiable.

• **Well definiteness of DCA**:

DCA is well defined if one can construct two sequences $\{x^k\}$ and $\{y^k\}$ as above from an arbitrary initial point $x^o \in \operatorname{dom} g$. We have $x^{k+1} \in \partial g^*(y^k)$ and $y^k \in \partial h(x^k)$, $\forall k \geq 0$. So $\{x^k\} \subset \operatorname{range} \partial g^* = \operatorname{dom} \partial g$ and $\{y^k\} \subset \operatorname{range} \partial h = \operatorname{dom} \partial h^*$. Then it is clear that

Lemma 1. *Sequences $\{x^k\}$, $\{y^k\}$ in DCA are well defined if and only if*

$$\operatorname{dom} \partial g \subset \operatorname{dom} \partial h \text{ and } \operatorname{dom} \partial h^* \subset \operatorname{dom} \partial g^*.$$

Since for $\varphi \in \Gamma_o(X)$ we have $\operatorname{ri}(\operatorname{dom} \varphi) \subset \operatorname{dom} \partial\varphi \subset \operatorname{dom} \varphi$ ([93]) (ri (dom φ) stands for the relative interior of dom φ) we can say, under the essential assumption (1), that DCA is in general well defined.

Remark 2. A d.c. function f has infinitely many d.c. decompositions. For example if $f = g - h$, then $f = (g + \theta) - (h + \theta)$ for every $\theta \in \Gamma_o(X)$ finite on the whole X. It is clear that the primal d.c. programs (P) corresponding to the two d.c. decompositions of the objective function f are identical. But their dual programs are quite different and so is DCA relative to these d.c. decompositions. In other words, there are as many DCA as there are d.c. decompositions of the objective function f. It is so useful to find a suitable d.c. decomposition of f since it may have an important influence on the efficiency of DCA for its solution. This question is intimately related to the regularization techniques in d.c. programming ([1], [76], [80]).

3.2. Convergence of DCA for general d.c. programs

Let ρ_i and ρ_i^*, $(i = 1, 2)$ be real nonnegative numbers such that $0 \le \rho_i < \rho(f_i)$ (resp. $0 \le \rho_i^* < \rho(f_i^*)$) where $\rho_i = 0$ (resp. $\rho_i^* = 0$) if $\rho(f_i) = 0$ (resp. $\rho(f_i^*) = 0$) and ρ_i (resp. ρ_i^*) may take the value $\rho(f_i)$ (resp. $\rho(f_i^*)$) if it is attained. We next set $f_1 = g$ and $f_2 = h$. Also let $dx^k := x^{k+1} - x^k$ and $dy^k := y^{k+1} - y^k$.

The basic convergence theorem of DCA for general d.c. programming will be stated below.

Theorem 3. *Suppose that the sequences $\{x^k\}$ and $\{y^k\}$ are defined by the simplified DCA. Then we have*

(i) $$(g-h)(x^{k+1}) \le (h^*-g^*)(y^k) - \max\left\{\frac{\rho_2}{2}\|dx^k\|^2, \frac{\rho_2^*}{2}\|dy^k\|^2\right\}$$
$$\le (g-h)(x^k) - \max\left\{\frac{\rho_1+\rho_2}{2}\|dx^k\|^2, \frac{\rho_1^*}{2}\|dy^{k-1}\|^2 + \frac{\rho_2}{2}\|dx^k\|^2, \frac{\rho_1^*}{2}\|dy^{k-1}\|^2 + \frac{\rho_2^*}{2}\|dy^k\|^2\right\}.$$

The equality $(g-h)(x^{k+1}) = (g-h)(x^k)$ holds if and only if $x^k \in \partial g^*(y^k)$, $y^k \in \partial h(x^{k+1})$ and $(\rho_1+\rho_2)dx^k = \rho_1^* dy^{k-1} = \rho_2^* dy^k = 0$. In this case

- *$(g-h)(x^{k+1}) = (h^*-g^*)(y^k)$ and x^k, x^{k+1} are the critical points of $g-h$ satisfying $y^k \in (\partial g(x^k) \cap \partial h(x^k))$ and $y^k \in (\partial g(x^{k+1}) \cap \partial h(x^{k+1}))$,*
- *y^k is a critical point of $h^* - g^*$ satisfying $[x^k, x^{k+1}] \subset ((\partial g^*(y^k) \cap \partial h^*(y^k))$,*
- *$x^{k+1} = x^k$ if $\rho(g) + \rho(h) > 0, y^k = y^{k-1}$ if $\rho(g^*) > 0$ and $y^k = y^{k+1}$ if $\rho(h^*) > 0$.*

(ii) *Similarly, for the dual problem we have*

$$(h^*-g^*)(y^{k+1}) \le (g-h)(x^{k+1}) - \max\left\{\frac{\rho_1}{2}\|dx^{k+1}\|^2, \frac{\rho_1^*}{2}\|dy^k\|^2\right\}$$
$$\le (h^*-g^*)(y^k) - \max\left\{\frac{\rho_1}{2}\|dx^{k+1}\|^2 + \frac{\rho_2}{2}\|dx^k\|^2, \frac{\rho_1^*}{2}\|dy^k\|^2 + \frac{\rho_2}{2}\|dx^k\|^2, \frac{\rho_1^*+\rho_2^*}{2}\|dy^k\|^2\right\}.$$

The equality $(h^*-g^*)(y^{k+1}) = (h^*-g^*)(y^k)$ holds if and only if $x^{k+1} \in \partial g^*(y^{k+1})$, $y^k \in \partial h(x^{k+1})$ and $(\rho_1^*+\rho_2^*)dy^k = \rho_2 dx^k = \rho_1 dx^{k+1} = 0$. In this case

- $(h^* - g^*)(y^{k+1}) = (g - h)(x^{k+1})$ *and* y^k, y^{k+1} *are the critical points of* $h^* - g^*$ *satisfying* $x^{k+1} \in (\partial g^*(y^k) \cap \partial h^*(y^k))$ *and* $x^{k+1} \in (\partial g^*(y^{k+1}) \cap \partial h^*(y^{k+1}))$,
- x^{k+1} *is a critical point of* $g - h$ *satisfying* $[y^k, y^{k+1}] \subset ((\partial g(x^{k+1}) \cap \partial h(x^{k+1}))$,
- $y^{k+1} = y^k$ *if* $\rho(g^*) + \rho(h^*) > 0, x^{k+1} = x^k$ *if* $\rho(h) > 0$ *and* $x^{k+1} = x^{k+2}$ *if* $\rho(g) > 0$.

(iii) *If* α *is finite then the decreasing sequences* $\{(g-h)(x^k)\}$ *and* $\{(h^* - g^*)(y^k)\}$ *converge to the same limit* $\beta \geq \alpha$, *i.e.,* $\lim_{k\to+\infty}(g - h)(x^k) = \lim_{k\to+\infty}(h^* - g^*)(y^k) = \beta$. *If* $\rho(g) + \rho(h) > 0$ *(resp.* $\rho(g^*) + \rho(h^*) > 0$*), then* $\lim_{k\to+\infty}\{x^{k+1} - x^k\} = 0$ *(resp.* $\lim_{k\to+\infty}\{y^{k+1} - y^k\} = 0$*).*

Moreover, $\lim_{k\to+\infty}\{g(x^k) + g^*(y^k) - \langle x^k, y^k\rangle\} = 0 = \lim_{k\to+\infty}\{h(x^{k+1}) + h^*(y^k) - \langle x^{k+1}, y^k\rangle\}$.

(iv) *If* α *is finite and the sequences* $\{x^k\}$ *and* $\{y^k\}$ *are bounded, then for every limit* x^* *of* $\{x^k\}$ *(resp.* y^* *of* $\{y^k\}$*) there exists a cluster point* y^* *of* $\{y^k\}$ *(resp.* x^* *of* $\{x^k\}$*) such that*

- $(x^*, y^*) \in [\partial g^*(y^*) \cap \partial h^*(y^*)] \times [\partial g(x^*) \cap \partial h(x^*)]$ *and* $(g - h)(x^*) = (h^* - g^*)(y^*) = \beta$,
- $\lim_{k\to+\infty}\{g(x^k) + g^*(y^k)\} = \lim_{k\to+\infty}\langle x^k, y^k\rangle$.

Proof. First, we need the following results

Proposition 2. *Suppose that the sequences* $\{x^k\}$ *and* $\{y^k\}$ *are generated by the simplified DCA. Then we have*

(i) $(g-h)(x^{k+1}) \leq (h^*-g^*)(y^k) - \frac{\rho_2}{2}\|dx^k\|^2 \leq (g-h)(x^k) - \frac{\rho_1+\rho_2}{2}\|dx^k\|^2$.

The equality $(g - h)(x^{k+1}) = (g - h)(x^k)$ *holds if and only if*

$$x^k \in \partial g^*(y^k),\ y^k \in \partial h(x^{k+1}) \text{ and } (\rho_1 + \rho_2)\|dx^k\| = 0.$$

(ii) *Similarly, by duality, we have*

$$\begin{aligned}(h^* - g^*)(y^{k+1}) &\leq (g - h)(x^{k+1}) - \frac{\rho_1^*}{2}\|dy^k\|^2 \\ &\leq (h^* - g^*)(y^k) - \frac{\rho_1^* + \rho_2^*}{2}\|dy^k\|^2.\end{aligned}$$

The equality $(h^* - g^*)(y^{k+1}) = (h^* - g^*)(y^k)$ *holds if and only if*

$$x^{k+1} \in \partial g^*(y^{k+1}),\ y^k \in \partial h(x^{k+1}) \text{ and } (\rho_1^* + \rho_2^*)\|dy^k\| = 0.$$

Proof of Proposition 2. (i) The inclusion $y^k \in \partial h(x^k)$ follows that $h(x^{k+1}) \geq h(x^k) + \langle x^{k+1} - x^k, y^k \rangle + \frac{\rho_2}{2}\|dx^k\|^2$. Hence,

$$(g-h)(x^{k+1}) \leq g(x^{k+1}) - \langle x^{k+1} - x^k, y^k \rangle - h(x^k) - \frac{\rho_2}{2}\|dx^k\|^2. \tag{12}$$

Likewise, $x^{k+1} \in \partial g^*(y^k)$ implies

$$g(x^k) \geq g(x^{k+1}) + \langle x^k - x^{k+1}, y^k \rangle + \frac{\rho_1}{2}\|dx^k\|^2.$$

So,

$$g(x^{k+1}) - \langle x^{k+1} - x^k, y^k \rangle - h(x^k) \leq (g-h)(x^k) - \frac{\rho_1}{2}\|dx^k\|^2. \tag{13}$$

On the other hand,

$$x^{k+1} \in \partial g^*(y^k) \Leftrightarrow \langle x^{k+1}, y^k \rangle = g(x^{k+1}) + g^*(y^k), \tag{14}$$

$$y^k \in \partial h(x^k) \Leftrightarrow \langle x^k, y^k \rangle = h(x^k) + h^*(y^k). \tag{15}$$

Thus,

$$g(x^{k+1}) - \langle x^{k+1} - x^k, y^k \rangle - h(x^k) = h^*(y^k) - g^*(y^k). \tag{16}$$

Finally, combining (12), (13) and (16), we get

$$(g-h)(x^{k+1}) \leq (h^* - g^*)(y^k) - \frac{\rho_2}{2}\|dx^k\|^2 \leq (g-h)(x^k) - \frac{\rho_1 + \rho_2}{2}\|dx^k\|^2. \tag{17}$$

If $\rho_1 + \rho_2 > 0$, the last statement of (i) is an immediate consequence of the equality $dx^k = 0$ and the construction of the sequences $\{x^k\}$ and $\{y^k\}$.

It is clear that there exist ρ_1 and ρ_2 such that $\rho_1 + \rho_2 > 0$ if and only if $\rho(h) + \rho(g) > 0$. In the case $\rho(h) = \rho(g) = 0$, (15) implies the equivalence between $(g-h)(x^{k+1}) = (g-h)(x^k)$ and the combination (18) and (19):

$$(g-h)(x^{k+1}) = (h^* - g^*)(y^k), \tag{18}$$

$$(h^* - g^*)(y^k) = (g-h)(x^k). \tag{19}$$

We then deduce from (14) and (18) that $h(x^{k+1}) + h^*(y^k) = \langle x^{k+1}, y^k\rangle$, i.e., $y^k \in \partial h(x^{k+1})$.

Similarly, (15) and (19) give $g(x^k) + g^*(y^k) = \langle x^k, y^k\rangle$, i.e., $x^k \in \partial g^*(y^k)$.

Property (ii) is analogously proved. □

The following result is an important consequence of Proposition 2

Corollary 4.

$$\begin{aligned}\text{(i) } (g-h)(x^{k+1}) &\le (h^*-g^*)(y^k) - \frac{\rho_2}{2}\|dx^k\|^2\\ &\le (g-h)(x^k) - \Big[\frac{\rho_1^*}{2}\|dy^{k-1}\|^2 + \frac{\rho_2}{2}\|dx^k\|^2\Big].\\ \text{(ii) } (g-h)(x^{k+1}) &\le (h^*-g^*)(y^k) - \frac{\rho_2^*}{2}\|dy^k\|^2\\ &\le (g-h)(x^k) - \Big[\frac{\rho_1^*}{2}\|dy^{k-1}\|^2 + \frac{\rho_2^*}{2}\|dy^k\|^2\Big].\end{aligned}$$

The equality $(g-h)(x^{k+1}) = (g-h)(x^k)$ *holds if and only if*

$$x^k \in \partial g^*(y^k),\ y^k \in \partial h(x^{k+1}) \text{ and } (\rho_1+\rho_2)dx^k = \rho_1^* dy^{k-1} = \rho_2^* dy^k = 0.$$

Similarly, by duality, we have

$$\begin{aligned}\text{iii) } (h^*-g^*)(y^{k+1}) &\le (g-h)(x^{k+1})\frac{\rho_1^*}{2}\|dy^k\|^2\\ &\le (h^*-g^*)(y^k) - \Big[\frac{\rho_1^*}{2}\|dy^k\|^2 + \frac{\rho_2}{2}\|dx^k\|^2\Big].\\ \text{(iv) } (h^*-g^*)(y^{k+1}) &\le (g-h)(x^{k+1}) - \frac{\rho_1}{2}\|dx^{k+1}\|^2\\ &\le (h^*-g^*)(y^k) - \Big[\frac{\rho_1}{2}\|dx^{k+1}\|^2 + \frac{\rho_2}{2}\|dx^k\|^2\Big].\end{aligned}$$

The equality $(h^*-g^*)(y^{k+1}) = (h^*-g^*)(y^k)$ *holds if and only if*

$$\begin{aligned}&x^{k+1} \in \partial g^*(y^{k+1}),\ y^k \in \partial h(x^{k+1}) \text{ and}\\ &(\rho_1^*+\rho_2^*)dy^k = \rho_2 dx^k = \rho_1 dx^{k+1} = 0.\end{aligned}$$

Proof. The inequalities in (i) and (ii) are easily deduced from Properties (i) and (ii) of Proposition 2. The inequalities in (iii) and (iv) can be shown by the same arguments as in the proof of Proposition 2. □

We are now in a position to demonstrate Theorem 3.

Proof of Theorem 3. Properties (i) and (ii) are proved analogously, therefore we give here the proof for (i) only.

The first inequality of (i) is an immediate consequence of (i) and (ii) of Corollary 4.

- If $\rho_2\|dx^k\|^2 \le \rho_2^*\|dy^k\|^2$, then (ii) of Corollary 4 follows that

$$(h^* - g^*)(y^k) - \max\left\{\frac{\rho_2}{2}\|dx^k\|^2, \frac{\rho_2^*}{2}\|dy^k\|^2\right\} = (h^* - g^*)(y^k) - \frac{\rho_2^*}{2}\|dy^k\|^2$$
$$\le (g-h)(x^k) - \left\{\frac{\rho_1^*}{2}\|dy^{k-1}\|^2 + \frac{\rho_2^*}{2}\|dy^k\|^2\right\}.$$

Parallelly, Property (i) of Corollary 4 implies

$$(h^* - g^*)(y^k) - \frac{\rho_2^*}{2}\|dy^k\|^2 \le (h^* - g^*)(y^k) - \frac{\rho_2}{2}\|dx^k\|^2$$
$$\le (g-h)(x^k) - \left[\frac{\rho_1^*}{2}\|dy^{k-1}\|^2 + \frac{\rho_2}{2}\|dx^k\|^2\right].$$

On the other hand, by (i) of Proposition 2

$$(h^* - g^*)(y^k) - \frac{\rho_2^*}{2}\|dy^k\|^2 \le (h^* - g^*)(y^k) - \frac{\rho_2}{2}\|dx^k\|^2$$
$$\le (g-h)(x^k) - \frac{\rho_1 + \rho_2}{2}\|dx^k\|^2.$$

Combining these inequalities, we get the second inequality of (i).

- If $\rho_2^*\|dy^k\|^2 \le \rho_2\|dx^k\|^2$, then by using the same arguments we can easily show the second inequality of (i). The first property of (iii) is evident. We will prove the last one. Taking (16) and (i) into account, we have

$$\lim_{k\to+\infty}(g-h)(x^{k+1}) = \lim_{k\to+\infty}\{g(x^{k+1}) - \langle x^{k+1} - x^k, y^k\rangle - h(x^k)\}$$
$$= \lim_{k\to+\infty}(g-h)(x^k).$$

The second equality implies $\lim_{k\to+\infty}\{g(x^{k+1}) - \langle x^{k+1} - x^k, y^k\rangle - g(x^k)\} = 0$, i.e., $\lim_{k\to+\infty}\{g(x^k) + g^*(y^k) - \langle x^k, y^k\rangle\} = 0$, since $x^{k+1} \in \partial g^*(y^k)$.

Likewise, it results from the first equality that $\lim_{k\to+\infty}\{h(x^{k+1}) - \langle x^{k+1} - x^k, y^k\rangle - h(x^k)\} = 0$, i.e., $\lim_{k\to+\infty}\{h(x^{k+1}) + h^*(y^k) - \langle x^{k+1}, y^k\rangle\} = 0$ since $y^k \in \partial h(y^k)$.

(iv) We assume α is finite and the sequences $\{x^k\}$ and $\{y^k\}$ are bounded. Let x^* be a limit point of $\{x^k\}$. For the sake of simplicity we write $\lim_{k\to+\infty} x^k = x^*$. We can suppose (by extracting a subsequence if necessary) that the sequence $\{y^k\}$ converges to $y^* \in \partial h(x^*)$. Property (iii) then implies

$$\lim_{k\to+\infty}\{g(x^k) + g^*(y^k)\} = \lim_{k\to+\infty}\langle x^k, y^k\rangle = \langle x^*, y^*\rangle.$$

Let $\theta(x,y) = g(x) + g^*(y)$ for $(x,y) \in X \times Y$. It is clear that $\theta \in \Gamma_o(X \times Y)$. Then the lower semicontinuity of θ implies

$$\theta(x^*, y^*) \leq \lim_{k\to+\infty}\inf \theta(x^k, y^k) = \lim_{k\to+\infty} \theta\langle x^k, y^k\rangle = \langle x^*, y^*\rangle,$$

i.e., $\theta(x^*, y^*) = g(x^*) + g^*(y^*) = \langle x^*, y^*\rangle$. In other words, $y^* \in \partial g(x^*)$. According to Lemma 2 stated below we have

$$\lim_{k\to+\infty} h(x^k) = h(x^*) \text{ and } \lim_{k\to+\infty} h^*(y^k) = h^*(y^*),$$

since $y^k \in \partial h(x^k)$, $x^k \to x^*$ and $y^k \to y^*$.

Hence, in virtue of (iii),

$$\begin{aligned}
\lim_{k\to+\infty}(g-h)(x^k) &= \lim_{k\to+\infty} g(x^k) - \lim_{k\to+\infty} h(x^k) \\
&= \lim_{k\to+\infty} g(x^k) - h(x^*) = \beta, \\
\lim_{k\to+\infty}(h^*-g^*)(y^k) &= \lim_{k\to+\infty} h^*(y^k) - \lim_{k\to+\infty} g^*(y^k) \\
&= h^*(y^*) - \lim_{k\to+\infty} g^*(y^k) = \beta.
\end{aligned}$$

It then suffices to show that

$$\lim_{k\to+\infty} g(x^k) = g(x^*); \quad \lim_{k\to+\infty} g^*(y^k) = g^*(y^*).$$

Since $\lim_{k\to+\infty} g(x^k)$ and $\lim_{k\to+\infty} g^*(y^k)$ exist, (iii) implies

$$g(x^*) + g^*(y^*) = \lim_{k\to+\infty} \{g(x^k) + g^*(y^k)\} = \lim_{k\to+\infty} g(x^k) + \lim_{k\to+\infty} g^*(y^k).$$

Further, because of the lower semicontinuity of g and g^*,

$$\lim_{k\to+\infty} g(x^k) = \lim_{k\to+\infty} \inf g(x^k) \geq g(x^*),$$
$$\lim_{k\to+\infty} g^*(y^k) = \lim_{k\to+\infty} \inf g^*(y^k) \geq g^*(y^*).$$

The former equalities imply that these last inequalities are in fact equalities. The proof of Theorem 3 is complete. □

Lemma 2. *Let $h \in \Gamma_o(X)$ and $\{x^k\}$ be a sequence of elements in X such that*

(i) $x^k \to x^*$,

(ii) *There exists a bounded sequence $\{y^k\}$ with $y^k \in \partial h(x^k)$,*

(iii) *$\partial h(x^*)$ is nonempty.*

Then $\lim_{k\to+\infty} h(x^k) = h(x^*)$.

Proof. Indeed, let $y^* \in \partial h(x^*)$. Then $h(x^k) \geq h(x^*) + \langle x^k - x^*, y^*\rangle$, $\forall k$. Since $y^k \in \partial h(x^k)$, we have $h(x^*) \geq h(x^k) + \langle x^* - x^k, y^k\rangle$, $\forall k$. Hence, $h(x^k) \leq h(x^*) + \langle x^k - x^*, y^k\rangle$, $\forall k$. As $x^k \to x^*$, we have $\lim_{k\to+\infty} \langle x^k - x^*, y^*\rangle = 0$. Moreover, $\lim_{k\to+\infty} \langle x^k - x^*, y^k\rangle = 0$, since the sequence $\{y^k\}$ is bounded. Consequently, $\lim_{k\to+\infty} h(x^k) = h(x^*)$. □

Comments on Theorem 3.

(i) Properties (i) and (ii) prove that DCA is a descent method for both primal and dual programs. DCA provides critical points for (P) and (D) after finitely many operations if there is no strict decrease of the primal (or dual) objective function.

(ii) If C and D are convex sets such that $\{x^k\} \subset C$ and $\{y^k\} \subset D$, then Theorem 3 remains valid if we replace $\rho(f_i)$ by $\rho(f_i, C)$ and $\rho(f_i^*)$ by $\rho(f_i^*, D)$ for $i = 1, 2$. By this way we may improve the results in the theorem.

(iii) In (ii) of Theorem 3, the convergence of the whole sequence $\{x^k\}$ (resp. $\{y^k\}$) can be ensured under the following conditions ([65], [71]):

- $\{x^k\}$ is bounded;
- The set of limit points of $\{x^k\}$ is finite;
- $\lim_{k\to+\infty} \|x^{k+1} - x^k\| = 0.$

(iv) In general, the qualities (robustness, stability, rate of convergence and globality of sought solutions) of DCA, in both complete and simplified forms, depend upon the d.c. decomposition of the function f. Theorem 3 shows how strong convexity of d.c. components in primal and dual problems can influence on DCA. To make the d.c. components (of the primal objective function $f = g - h$) strongly convex we usually apply the following process

$$f = g - h = \Big(g + \frac{\lambda}{2}\|\cdot\|^2\Big) - \Big(h + \frac{\lambda}{2}\|\cdot\|^2\Big).$$

In this case the d.c. components in the dual problem will be continuously differentiable. Parallelly inf-convolution of g and h with $\frac{\lambda}{2}\|\cdot\|^2$ will make the d.c. components (in dual problem) strongly convex and the d.c. components of the primal objective function continuously differentiable. For a detailed study of regularization techniques in d.c. optimization, see Section 5 and [1], [76], [80].

3.3. How to restart simplified DCA for obtaining x^* such that $\partial h(x^*) \subset \partial g(x^*)$

As mentioned above, the complete DCA theoretically provides a x^* such that $\partial h(x^*) \subset \partial g(x^*)$. In practice, except for the cases where the convex maximization problems ($S(x^k)$ and $(T(y^k))$ are easy to solve, one generally uses the simplified DCA. It is worth noting that if the simplified DCA terminates at some point x^* for which $\partial h(x^*)$ is not contained in $\partial g(x^*)$, then one can reduce the objective function value by restarting it from a new initial point $x^o = x^*$ with $y^o \in \partial h(x^o)$ such that $y^o \notin \partial g(x^o)$. In fact, since

$$g(x^1) + g^*(y^o) = \langle x^1, y^o\rangle \le h(x^1) - h(x^o) + \langle x^o, y^o\rangle,$$

and $\langle x^o, y^o\rangle < g(x^o) + g^*(y^o)$ because $y^o \notin \partial g(x^o)$, we have

$$g(x^1) + g^*(y^o) < h(x^1) - h(x^o) + g(x^o) + g^*(y^o).$$

Hence,

$$g(x^1) - h(x^1) < g(x^o) - h(x^o).$$

4. Polyhedral d.c. optimization problems and finite convergence of DCA with fixed choices of subgradients

4.1. Polyhedral d.c. program

We suppose that in Problem (P) either g or h is polyhedral convex. We may assume that h is a polyhedral convex function given by $h(x) = \max\{\langle a^i, x\rangle - \alpha^i : i = 1,\dots,m\} + \chi_C(x)$, where χ_C is the indicator function of a nonempty polyhedral convex set C in X. (If in (P) g is polyhedral and h is not so, then we consider the dual problem (D), since g^* is polyhedral). Throughout this subsection we assume that the optimal value α of problem (P) is finite which implies that $\operatorname{dom} g \subset \operatorname{dom} h = C$. Thus, (P) is equivalent to the problem

$$(\tilde{P}) \qquad \alpha = \inf\{g(x) - \tilde{h}(x) : x \in X\},$$

where $\tilde{h}(x) = \max\{\langle a^i, x\rangle - \alpha^i : i \in I\}$ with $I = \{1,\dots,m\}$. By this way we can avoid $+\infty - (+\infty)$ in (P). Clearly, $\alpha = \inf\limits_{i\in I}\inf\limits_{x\in X}\{g(x) - (\langle a^i, x\rangle - \alpha^i)\}$. For each $i \in I$, let

$$(P_i) \qquad \beta^i = \inf\{g(x) - (\langle a^i, x\rangle - \alpha^i) : x \in X\}$$

whose solution set is $\partial g^*(a^i)$.

The dual problem $(\tilde{D})$ of $(\tilde{P})$ is:

$$(\tilde{D}) \qquad \alpha = \inf\{\tilde{h}^*(y) - g^*(y) : \ y \in \operatorname{co}\{a^i : i \in I\}\},$$

where α verifies $\alpha = \inf\{\tilde{h}^*(y) - g^*(y) : \ y \in \{a^i : i \in I\}\}$.

Also, let $J(\alpha) = \{i \in I : \beta^i = \alpha\}$ and $I(x) = \{i \in I : \langle a^i, x\rangle - \alpha^i = \tilde{h}(x)\}$. We have the following result ([1], [2], [81]).

Theorem 4. (i) *$x^* \in \mathcal{P}$ if and only if $I(x^*) \subset J(\alpha)$ and $x^* \in \cap\{\partial g^*(a^i) : i \in I(x^*)\}$.*

(ii) *$\mathcal{P} = \cup\{\partial g^*(a^i) : i \in J(\alpha)\}$. If $\{a^i : i \in I\} \subset \operatorname{dom}\partial g^*$, then $\mathcal{P} \neq \emptyset$.*

(iii) *$\tilde{h}(x) = \max\{\langle x, y\rangle - \tilde{h}^*(y) : y \in \operatorname{co}\{a^i : i \in I\}\} = \max\{\langle a^i, x\rangle - \tilde{h}^*(a^i) : \ i \in I\}$.*

(iv) *$J(\alpha) = \{i \in I : a^i \in \tilde{\mathcal{D}}$ and $\tilde{h}^*(a^i) = \alpha^i\}$; $\tilde{\mathcal{D}} \supset \{a^i : i \in J(\alpha)\}$.*

The proof of this theorem is very technical, the interested reader is therefore referred to [1], [2], [81] for a detailed analysis.

4.2. Finite convergence of DCA

From 4.1 we see that (globally) solving the polyhedral d.c. optimization problem $(\tilde{P})$ amounts to solving m convex programs (P_i) $(i \in I)$. For generating $\mathcal{P}$ one can first determine $J(\alpha)$ and then apply Theorem 4. In practice this can be effectively done if m is relatively small. In case when m is large we use the simplified DCA for solving (locally) Problem $(\tilde{P})$. Recall that (Lemma 1) the simplified DCA is well defined if and only if $\operatorname{co}\{a^i : i \in I\} \subset \operatorname{dom} \partial g^*$. Because of the finiteness of α, $\operatorname{dom} g \subset \operatorname{dom} h = C$ and $\operatorname{co}\{a^i : i \in I\} \subset \operatorname{dom} g^*$. The simplified DCA in this case is described as follows: Let x^o be chosen in advance. Set $y^k \in \partial\tilde{h}(x^k) = \operatorname{co}\{a^i : i \in I(x^k)\}$; $x^{k+1} \in \partial g^*(y^k)$ By setting $y^k = a^i, i \in I(x^k)$ the calculation of x^{k+1} is reduced to solve the convex program

$$(\tilde{P}_i) \qquad \min\{g(x) - \langle y^k, x\rangle : x \in X\}.$$

Note that if $y^k = a^i$ with $i \in J(\alpha)$, then, by Theorem 4, $x^{k+1} \in \mathcal{P}$.

Now let $\tilde{H}$ and G^* be two mappings defined respectively in $\operatorname{dom} \partial\tilde{h} = X$ and in $\operatorname{dom} \partial g^*$ such that $\tilde{H}(x) \in \partial\tilde{h}(x)$, $\forall x \in X$ and $G^*(y) \in \partial g^*(y)$ $\forall y \in \operatorname{dom} \partial g^*$. The simplified DCA with a fixed choice of subgradients is defined as:

$$y^k = \tilde{H}(x^k)\ ;\ x^{k+1} = G^*(y^k).$$

It is clear that for a polyhedral d.c. optimization problem range $\tilde{H}$ is finite if h is polyhedral convex, and range G^* is finite if g is polyhedral convex. In each of these cases the sequences $\{x^k\}$ and $\{y^k\}$ are discrete (i.e., they have only finitely many different elements).

Theorem 5 ([1], [2], [81]). (i) *The discrete sequences* $\{(g - \tilde{h})(x^k)\}$ *and* $\{(\tilde{h}^* - g^*)(y^k)\}$ *are decreasing and convergent.*

(ii) *The discrete sequences* $\{x^k\}$ *and* $\{y^k\}$ *have the same property: either they are convergent or cyclic with the same period* p*. In the latter case the sequences* $\{x^k\}$ *and* $\{y^k\}$ *contain exactly* p *limit points that are all critical points of* $g - h$*. Moreover, if* $\rho(g) + \rho(g^*) > 0$*, then these sequences are convergent.*

The proof of this theorem follows immediately from Theorem 3 and the discrete character of the sequences $\{x_k\}$ and $\{y_k\}$.

4.3. Natural choice of subgradients in DCA

Let $f \in \Gamma_o(X)$ and T be a selection of ∂f, i.e., $Tx \in \partial f(x)$, $\forall x \in \operatorname{dom} \partial f$. T is said to be a *natural choice* of subgradients of f if the conditions $Tx \in \operatorname{ri} \partial f(x)$ and $\partial f(x) = \partial f(x')$ imply that $Tx = Tx'$. The following results are useful to the proof of the finite convergence of DCA (applied to the polyhedral d.c. optimization) with the fixed choices of subgradients for h and g^*, and the natural choice for at least one polyhedral function among them. The natural choice has been successfully used in the subgradient-methods for computing bound norms of matrices ([70]-[73]) and the study of the iterative behaviour of cellular automatas ([75]).

Since $\tilde{h}(x) = \max\{\langle a^i, x\rangle - \alpha^i : i \in I\}$, one can take $\tilde{H}$ by setting $\tilde{H}(x) = \sum_{i \in I(x)} \lambda^i a^i$, where $\lambda^i, i \in I(x)$, satisfy:

(i) $\lambda^i > 0$, $\forall i \in I(x)$ and $\sum_{i \in I(x)} \lambda^i = 1$;

(ii) λ^i depends only on $I(x)$.

Lemma 3 ([75], [81]). (i) $\partial \tilde{h}(x) = \partial \tilde{h}(x') \Leftrightarrow I(x) = I(x')$.

(ii) $\tilde{H}$ *is a natural choice of subgradients of* $\tilde{h}$ *if and only if it is defined as above.*

Consider now DCA with a fixed choice of subgradient applied to the polyhedral d.c. optimization problem $(\tilde{P})$. If $\tilde{H}$ is a natural choice of $\tilde{h}$, then the following result strengthens that of Theorem 5.

Theorem 6 ([1], [2], [81]). *The simplified DCA with a fixed choice of subgradients is finite.*

5. Regularization techniques in d.c. programming

We consider the d.c. problem (P) and its dual problem (D) where α is finite. In this case, $\operatorname{dom} g \subset \operatorname{dom} h$ and $\operatorname{dom} h^* \subset \operatorname{dom} g^*$. As mentioned above, it is important to obtain various equivalent d.c. forms for the primal and dual problems. The Lagrangian duality in d.c. optimization ([1], [78], [82]) and regularization techniques partly answer this question.

Regularization techniques in d.c. optimization have been early introduced in [76] and extensively developed in our recent works [1], [80]. Besides three forms of regularization techniques, we present here some results corresponding to the computation of modulus of strong convexity. These results are essential to regularization techniques applied to DCA.

First, we introduce some related results of convex analysis:

Let $\varphi, \psi \in \Gamma_o(X)$, the *inf-convolution* of φ and ψ, denoted by $\varphi_\nabla \psi$, is defined by ([51], [93])

$$\varphi_\nabla\psi(x) = \inf\{\varphi(x_1) + \psi(x_2) : x_1 + x_2 = x\}.$$

One says that $\varphi_\nabla\psi$ is exact at $x = x_1 + x_2$ if $\varphi_\nabla\psi(x) = \varphi(x_1) + \psi(x_2)$.

Likewise, $\varphi_\nabla\psi$ is said to be exact if it is exact at every $x \in X$. The following result is useful to regularization techniques ([51], [93]):

Theorem 7. (i) *$\varphi_\nabla\psi$ is convex and* dom $\varphi_\nabla\psi = \operatorname{dom}\varphi + \operatorname{dom}\psi$.

(ii) $(\varphi_\nabla\psi)^* = \varphi^* + \psi^*$.

(iii) *If* $\operatorname{ri}(\operatorname{dom}\varphi) \cap \operatorname{ri}(\operatorname{dom}\psi) \neq \emptyset$*, then* $(\varphi + \psi)^* = \varphi^*_\nabla\psi^*$ *and the inf-convolution* $\varphi^*_\nabla\psi^*$ *is exact.*

(iv) $\partial\varphi(x_1) \cap \partial\psi(x_2) \subset \partial(\varphi_\nabla\psi)(x_1 + x_2)$.

Moreover, if $\partial\varphi(x_1) \cap \partial\psi(x_2) \neq \emptyset$*, then the inf-convolution* $\varphi_\nabla\psi$ *is exact at* $x = x_1 + x_2$*. Conversely, if* $\varphi_\nabla\psi$ *is exact at* $x = x_1 + x_2$*, then* $\partial(\varphi_\nabla\psi)(x_1 + x_2) = \partial\varphi(x_1) \cap \partial\psi(x_2)$.

Let $\varphi \in \Gamma_o(X)$. The function φ is said to be *strictly convex* on a convex subset C of dom f if

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

Likewise, φ is said to be *essentially strictly convex* if it is strictly convex on any convex subset of dom $\partial\varphi$.

Theorem 8 ([93]). *Let* $\varphi \in \Gamma_o(X)$*. Then the following conditions are equivalent*

(i) $\forall x \in \operatorname{dom}\varphi$*,* $\partial\varphi(x)$ *contains at most one element.*

(ii) φ *is essentially differentiable.*

In this case $\partial\varphi(x) = \{\nabla\varphi(x)\}$ *if* $x \in \operatorname{int}(\operatorname{dom}\varphi)$ *and* $\partial\varphi(x)$ *is empty otherwise.*

(iii) $\varphi \in \Gamma_o(X)$ *is essentially differentiable if and only if* φ^* *is essentially strictly convex.*

(iv) *Let* $\varphi, \psi \in \Gamma_o(X)$ *be such that* φ *is essentially differentiable and* $\operatorname{ri}(\operatorname{dom}\varphi^*) \cap \operatorname{ri}(\operatorname{dom}\psi^*) \neq \emptyset$*. Then* $\varphi_\nabla\psi$ *is essentially differentiable.*

We present now several types of regularization techniques.

5.1. Regularizing d.c. components of the dual d.c. program

Let $\theta \in \Gamma_o(X)$ such that

(i) $\operatorname{dom}\theta \supset \operatorname{dom} g$;

(ii) $\operatorname{ri}(\operatorname{dom}\theta) \cap \operatorname{ri}(\operatorname{dom} g) \neq \emptyset$ and $\operatorname{ri}(\operatorname{dom}\theta) \cap \operatorname{ri}(\operatorname{dom} h) \neq \emptyset$.

Clearly, $g + \theta$ and $h + \theta$ are also d.c. components of f. The following problem, by (i), is equivalent to (P):

$$(P_+) \qquad \lambda\alpha = \inf\{(\lambda g + \theta)(x) - (\lambda h + \theta)(x) : x \in X\},$$

where λ is a positive number. From (ii) the dual of (P_+) is formulated by (Theorem 7)

$$\begin{aligned}(D_+) \qquad \lambda\alpha &= \inf\{(\lambda g + \theta)^*(y) - (\lambda h + \theta)^*(y) : y \in Y\} \\ &= \inf\{(\lambda h)^*_\triangledown \theta^*(y) - (\lambda g)^*_\triangledown \theta^*(y) : y \in Y\},\end{aligned}$$

which is not equivalent to (D). This regularization, with suitable choices of θ, makes

- d.c. components in (P_+) are strongly convex: if θ is strongly convex, then so are $\lambda g + \theta$ and $\lambda h + \theta$, since

$$\rho(\lambda g + \theta) = \lambda\rho(g) + \rho(\theta);\ \rho(\lambda h + \theta) = \lambda\rho(h) + \rho(\theta),$$

- d.c. components in (D_+) are continuously differentiable.
  Indeed, for example if the sets $\operatorname{ri}(\operatorname{dom} g) \cap \operatorname{ri}(\operatorname{dom}\theta)$ and $\operatorname{ri}(\operatorname{dom} h) \cap \operatorname{ri}(\operatorname{dom}\theta)$ are nonempty and if θ^* is essentially differentiable, then according to Theorem 8, $(\lambda g)^*_\triangledown \theta^*$ and $(\lambda h)^*_\triangledown \theta^*$ are essentially differentiable. In particular, if θ is finite, strictly convex on the whole X, and coercive (i.e., $\lim_{\|x\| \to +\infty} \theta(x) = +\infty$), then $(\lambda g)^*_\triangledown \theta^*$ and $(\lambda h)^*_\triangledown \theta^*$ are differentiable on E (i.e., continuously differentiable on X), since the coerciveness of θ implies that $\operatorname{dom}\theta^* = Y$ (the dual space of X) and its strict convexity implies the essential differentiability of θ^* in virtue of Theorem 8.

5.2. Regularizing d.c. components in the primal d.c. program

This regularization is introduced by considering the following d.c. program

$$(P_\triangledown) \qquad \lambda\alpha = \inf\{(\lambda g)_\triangledown \theta(x) - (\lambda h)_\triangledown \theta(x) : x \in X\},$$

where λ is a positive number and $\theta \in \Gamma_o(X)$ for which the following conditions hold:

(i) $\mathrm{ri}(\lambda \mathrm{dom}\, g^*) \cap \mathrm{ri}(\mathrm{dom}\, \theta^*) \neq \emptyset$, $\mathrm{ri}(\lambda \mathrm{dom}\, h^*) \cap \mathrm{ri}(\mathrm{dom}\, \theta^*) \neq \emptyset$,

(ii) $\mathrm{dom}\, \theta^* \supset \mathrm{dom}\, (\lambda h)^* = \lambda \mathrm{dom}\, h^*$.

In virtue of Theorem 7, one has $(\lambda g)^*(y) = \lambda g^*(\lambda^{-1} y)$. Thus, dom $((\lambda g)^*) = \lambda(\mathrm{dom}\, g^*)$. The condition (i) ensures that both $\lambda g_\nabla \theta$ and $\lambda h_\nabla \theta$ are in $\Gamma_o(X)$ (Theorem 7). The dual problem of (P_+) is then given as

$$(D_\nabla) \qquad \begin{aligned} \lambda\alpha &= \inf\{(\lambda h \nabla \theta)^*(y) - (\lambda g \nabla \theta)^*(y) : y \in Y\} \\ &= \inf\{((\lambda h)^* + \theta^*)(y) - ((\lambda g)^* + \theta^*)(y) : y \in Y\} \end{aligned}$$

which, by (ii), is equivalent to (D). As before, suitable choices of θ allow to obtain the continuously differentiable d.c. components in (P_∇) and strongly convex d.c. components in (D_∇).

5.3. Double primal-dual regularization of d.c. components

Let λ be a positive number and $\varphi, \theta \in \Gamma_o(X)$ such that

(i) $\mathrm{dom}\, g \subset \mathrm{dom}\, \varphi$;

(ii) $\mathrm{ri}(\mathrm{dom}\, \varphi) \cap \mathrm{ri}(\mathrm{dom}\, g) \neq \emptyset$, $\mathrm{ri}(\mathrm{dom}\, \varphi) \cap \mathrm{ri}(\mathrm{dom}\, h) \neq \emptyset$;

(iii) $\lambda \mathrm{dom}\, g^* + \mathrm{dom}\, \varphi^* \subset \mathrm{dom}\, \theta^*$.

We consider the primal problem (P_+)

$$(P_+) \qquad \lambda\alpha = \inf\{(\lambda g + \varphi)(x) - (\lambda h + \varphi)(x) : x \in X\},$$

which is equivalent to (P). We shall regularize (P_+) by applying inf-convolution with θ to d.c. components:

$$(P_{+\nabla}) \qquad \inf\{(\lambda g + \varphi)_\nabla \theta(x) - (\lambda h + \varphi)_\nabla \theta(x) : x \in X\}.$$

Under assumptions (ii), the dual problem of $(P_{+\nabla})$ takes the form (Theorem 7):

$$(D_{+\nabla}) \qquad \begin{aligned} \lambda\alpha &= \inf\{((\lambda h + \varphi)^* + \theta^*)(y) - ((\lambda g + \varphi)^* + \theta^*)(y) : y \in Y\} \\ &= \inf\{((\lambda h)^*_\nabla \varphi^* + \theta^*)(y) - ((\lambda g)^*_\nabla \varphi^* + \theta^*)(y) : y \in Y\}. \end{aligned}$$

By assumption (iii) this is equivalent to the dual problem (D_+) of (P_+). This double regularization, with suitable choices of φ and θ, allows to obtain both strongly convex and continuously differentiable d.c. components in the primal and the dual problems $(P_{+\nabla})$ and $(D_{+\nabla})$.

In practice, we usually take $\theta(x) = \frac{\mu}{2}\|x\|^2$, $\mu > 0$ which is both strongly convex and coercive. Such a regularization is called the *proximal regularization.*

Finally, the following result ([1], [80]), whose proof is omitted here, allows to compute the strong convexity modulus of regularized d.c. components which intervenes in the convergence theorem for DCA (Theorem 3).

Proposition 3. *If $\varphi \in \Gamma_o(X)$ and λ, μ be positive numbers, then*

(i) $\rho\left((\lambda\varphi)^*\right) = \lambda^{-1}\rho(\varphi^*)$,

(ii) $\nabla\left[(\lambda\varphi)^*\right]_\mu(x) = \left(I + \mu^{-1}\lambda\partial\varphi\right)^{-1}(\mu^{-1}x)$, $\forall x \in X$,

(iii) $\rho(\varphi)\rho(\varphi^*) \le 1$,

(iv) $\dfrac{\rho(\varphi)}{1+\lambda\rho(\varphi)} \le \rho\left(\varphi_\lambda\right) \le \dfrac{1}{\lambda}$.

6. Functions which are more convex, less convex

Let $f, g \in \Gamma_o(X)$. The function f is said to be *more convex than* g (or g is said to be *less convex than* f), and we write $f \succ g$ (or $g \prec f$) if $f = g + h$ with $h \in \Gamma_o(X)$. The binary relation $f \succ g$ has been introduced by Moreau [62]. It is almost a partial ordering on $\Gamma_o(X)$ except for the case that $f \succ g$ and $g \succ f$ only imply $f = g + h$ with h being affine on X.

Moreau proved the following interesting result

$$f \succ \frac{1}{2}\|\cdot\|^2 \Leftrightarrow f^* \prec \frac{1}{2}\|\cdot\|^2.$$

In a work related to Moreau decomposition theorem, Hiriart-Urruty and Plazanet ([62]) have obtained a characterization of convex functions g, h such that $g + h = \frac{1}{2}\|\cdot\|^2$ ([35])

$$g + h = \frac{1}{2}\|\cdot\|^2 \Leftrightarrow \exists F \in \Gamma_o(X) \quad \text{such that}$$
$$g = F_\triangledown \frac{1}{2}\|\cdot\|^2 \text{ and } h = F^*_\triangledown \frac{1}{2}\|\cdot\|^2.$$

An explicit formulation is given in [35] with the help of an operation on a convex function, which bears the name of deconvolution of a function by another one. Given φ and ψ in $\Gamma_o(X)$, the deconvolution of φ by ψ is the function denoted $\varphi \Box \psi$ and defined by

$$\varphi \Box \psi(x) = \sup\{\varphi(x+u) - \psi(u) \ : \ u \in \operatorname{dom} \psi\}.$$

It is worth to note the two main properties ([35])

- $\varphi \Box \psi$ is either in $\Gamma_o(X)$ or identically equals $+\infty$.
- $(\varphi \Box \psi)^* = (\varphi^* - \psi^*)^{**}$.

The following result is useful to d.c. programming.

Proposition 4. *Let $f \in \Gamma_o(X)$ and λ be a positive number. Then*

(i) $f \succ \frac{\lambda}{2}\|\cdot\|^2$ *if and only if* $f^* \prec \frac{1}{2\lambda}\|\cdot\|^2$.

(ii) $f \prec \frac{1}{2\lambda}\|\cdot\|^2$ *if and only if* $f = \varphi_\lambda$ *with* $\varphi \in \Gamma_o(X)$. *More precisely,* φ *can be taken as* $\varphi = \lambda^{-1}[\lambda f \Box \frac{1}{2}\|\cdot\|^2]$.

(iii) $f \succ \frac{\lambda}{2}\|\cdot\|^2$ *if and only if there is* $\varphi \in \Gamma_o(X)$ *such that* $f_\nabla \varphi = \frac{\lambda}{2}\|\cdot\|^2$. *In this case we have*

$$\varphi = \left[\frac{1}{2\lambda}\|\cdot\|^2 - f^*\right]^* = \frac{\lambda}{2}\|\cdot\|^2 \Box f$$

and

$$f = \left[\frac{1}{2\lambda}\|\cdot\|^2 - \varphi^*\right]^* = \frac{\lambda}{2}\|\cdot\|^2 \Box \varphi.$$

The proof of this proposition is based on the works by Moreau [62] and by Hiriart-Urruty *et al.* ([35]). The reader is referred to [35] for a simple proof.

Let g, $h \in \Gamma_o(X)$. Consider the corresponding d.c. proram

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

and its dual d.c. program

$$\text{(D)} \qquad \alpha = \inf\{h^*(y) - g^*(y) : \ y \in Y\}.$$

Theorem 1 shows that solving one of them implies solving the other. It is clear that (P) is a convex program if and only if g is more convex than h. So the following question is of great interest in d.c. programming:

Does the condition $g \succ h$ *imply* $h^* \succ g^*$?

In particular, if $g = \lambda f + h$ with $f, h \in \Gamma_o(X)$ and h being finite on the whole X (see the relation between DCA and the proximal point algorithm). In case of negation it is desired to find the class of functions g and h such that the above property holds true (or false). Except for the case where $h = \frac{\lambda}{2}\|\cdot\|^2$ for which a positive answer due to Moreau [62] has been obtained, the problem is completely open.

Remark that $g \succ h$ but not $h^* \succ g^*$ furnishes an example of (nonconvex) d.c. program that we can globally solve by using d.c. duality (Theorem 1).

7. Relation between DCA and the Goldstein-Levitin-Polyak gradient projection algorithm in convex programming

7.1. Relation between DCA and the Proximal Point Algorithm

Let $f \in \Gamma_o(X)$. Consider the following convex program

$$\alpha = \inf\{f(x) : \ x \in X\}. \tag{20}$$

The proximal point algorithm applied to (20) is described as:

$$x^{k+1} = (I + \lambda_k \partial f)^{-1} x^k \tag{21}$$

with x^o being an initial vector chosen in advance and $\lambda_k \geq c > 0$. We recall the well-known result concerning the convergence of the proximal point algorithm ([56]-[58], [94])

Proposition 5. *If α is finite, then $\lim_{k\to+\infty} \|x^{k+1} - x^k\| = 0$. If $\partial f^*(0)$ is nonempty, then*

- $\|x^{k+1} - x\| \leq \|x^k - x\|, \quad \forall x \in \partial f^*(0),$
- *The sequence $\{x^k\}$ converges to $x \in \partial f^*(0)$ and the sequence $\{f(x^k)\}$ is monotically decreasing to α.*

If $\partial f^*(0)$ is empty, then

- $\lim\limits_{k\to+\infty} \|x^k\| = +\infty,$
- $x^k / \sum\limits_{i=1}^{k-1} \lambda_i$ *converges to $-v$, where v is the element with minimum norm of the closed convex set cl(range ∂f) (cl stands for the closure).*

Different estimations w.r.t. the convergence of the sequence $\{f(x^k)\}$ have been pointed out in [32]. The Moreau-Yosida regularization f_λ of f

introduces the regularized convex program

$$\alpha_\lambda = \inf\{f_\lambda(x) : x \in X\}. \tag{22}$$

It is not too difficult to prove the following classical result ([32], [56]-[58], [94])

Proposition 6. (i) $\alpha_\lambda = \alpha,\ \forall\lambda > 0$; $f_\lambda(x) \le f(x),\ \forall x \in X$.

Moreover, $f_\lambda(x) = f(x)$ *if and only if* $0 \in \partial f(x)$.

(ii) $(\partial f_\lambda)^{-1}(0) = (\partial f)^{-1}(0),\ \forall\lambda > 0$.

The proximal point algorithm applied to (20) can be regarded as follows:

$$x^{k+1} = \arg\min\left\{f(x) + \frac{1}{2\lambda_k}\|x - x^k\|^2 : x \in X\right\}.$$

Since $\nabla f_\lambda = (1/\lambda)\left[I - (I + \lambda\partial f)^{-1}\right]$ (Theorem 7), we have

$$x^{k+1} = x^k - \lambda_k \nabla f_{\lambda_k}(x^k) = (I + \lambda_k \partial f)^{-1}(x^k),$$

i.e., the passage of x^k to x^{k+1} is exactly performing a step equal to λ_k in the gradient method applied to the convex program $\inf\{f_{\lambda_k}(x) :\ x \in X\}$. In particular, if $\lambda_k = c$ for every k, then the proximal point algorithm (PPA) is nothing but the gradient method with fixed step c applying to $\inf\{f_c(x) :\ x \in X\}$.

Let us study now the relation between PPA and DCA. Consider the problem (20) in the following equivalent d.c. program

$$\lambda\alpha = \inf\{g(x) - h(x) : x \in X\}, \tag{23}$$

where $\lambda > 0$, $g = \lambda f + h$ and $h \in \Gamma_o(X)$ is finite on the whole X. It is clear that (23) is a "false" d.c. program whose solution can be estimated by either a convex optimization algorithm applied to (21) or DCA applied to (23). The latter actually yields global solutions since $\partial h(x) \subset \partial g(x) = \lambda\partial f(x) + \partial h(x)$ implies $0 \in \lambda\partial f(x)$. This implication can be shown by using the support functions of $\partial h(x)$ and $\lambda\partial f(x) + \partial h(x)$ and the compactness of $\partial h(x)$. It is then important to know whether the dual d.c. program of (23)

$$\lambda\alpha = \inf\{h^*(y) - (\lambda f)^*_\nabla g^*(y) : y \in Y\} \tag{24}$$

is always a convex program for every $h \in \Gamma_o(X)$ being finite on the whole X. This question has already been studied in Section 6. According to Proposition 4, if $h = (\mu/2)\|\cdot\|^2$, then $h^* - (\lambda f)^*_{\nabla} h^* \in \Gamma_o(X)$, i.e., (24) is a convex program.

Let us describe now DCA applied to (23)

$$x^k \to y^k \in \partial h(x^k) = \mu x^k; \; x^{k+1} \in \partial g^*(y^k) = \nabla[(\lambda f)^*]_\mu(y^k).$$

So, according to Proposition 3

$$x^{k+1} = (I + \mu^{-1}\lambda\partial f)^{-1}(\mu^{-1}y^k) = (I + \mu^{-1}\lambda\partial f)^{-1}x^k,$$

which is exactly PPA (with constant parameter $\lambda_k = \mu^{-1}\lambda$) applied to (20).

7.2. Relation between DCA and the Goldstein-Levitin-Polyak gradient projection in convex programming

Consider now the constrained convex program

$$\text{(Q)} \qquad \alpha = \inf\{f(x) : \; x \in C\}$$

with $f \in \Gamma_o(X)$ and C being a nonempty closed convex set in X. Clearly, for every $\lambda > 0$, Problem (Q) is equivalent to the following problem which for simplicity we also denote by (Q)

$$\text{(Q)} \qquad \lambda\alpha = \inf\{\lambda f(x) : \; x \in C\}.$$

Assume that

- $\operatorname{dom} f$ contains two nonempty open convex sets Ω_1 and Ω_2 such that

  $$C \subset \Omega_1, \quad cl\Omega_1 \subset \Omega_2 \subset \operatorname{dom} f$$

  ($cl\Omega_1$ stands for the closure of Ω_1).
- λf admits a "false" d.c. decomposition on Ω_2:

  $$\lambda f(x) = \varphi(x) - (\varphi - \lambda f)(x), \quad \forall x \in \Omega_2,$$

  where φ is finite convex on Ω_2 such that $\psi = \varphi - \lambda f$ is convex on Ω_2 (see Section 6).

Consider the usual extensions to the whole X of the two functions φ and ψ:

$$\tilde{\varphi}(x) = \varphi(x) \text{ if } x \in cl\Omega_1, \ +\infty \text{ otherwise},$$
$$\tilde{\psi}(x) = \psi(x) \text{ if } x \in cl\Omega_1, \ +\infty \text{ otherwise}.$$

Since φ and ψ are finite convex and continuous relative to $cl\Omega_1$, the functions $\tilde{\varphi}$ and $\tilde{\psi}$ belong to $\Gamma_o(X)$ according to [93]. Problem (Q) then takes the standard form of a d.c. program

$$\lambda\alpha = \inf\{g(x) - h(x) : \ x \in X\} \tag{25}$$

with $g = \tilde{\varphi} + \chi_C \in \Gamma_o(X)$ and $h = \tilde{\psi} = \tilde{\varphi} - \lambda f$. Such a problem is called a *false* d.c. program. Let $x^* \in C$ be satisfied the necessary condition for local optimality (Theorem 2) $\partial h(x^*) \subset \partial g(x^*)$. That is equivalent to $\partial h(x^*) \subset \partial h(x^*) + \partial(\lambda f)(x^*) + \partial\chi_C(x^*)$, i.e., $0 \in \partial f(x^*) + \partial\chi_C(x^*)$, since $\partial h(x^*)$ is nonempty and bounded. Consequently, DCA and algorithms for convex programming can be used for solving Problem (25). Remark that $\varphi = h + \lambda f$ and $g = h + \lambda f + \chi_C$. So, λf and h are less convex than g. Nevertheless, as indicated in Section 6 it is important to know if g^* is less convex than h^*, i.e., if the dual d.c. program of (25) is a convex one.

Let us illustrate all reasonings above by an example. Let C be a nonempty bounded closed convex set in X and let $f \in \Gamma_o(X)$ be in $\mathcal{C}^2(\Omega)$ where Ω is a bounded open convex set containing C. Since C is nonempty and compact, there is an ε such that

$$\begin{aligned} C \subset \Omega_1 = \{x \in X : \ d(x, C) < \varepsilon\} &\subset cl\Omega_1 \\ = \{x \in X : \ d(x, C) \le \varepsilon\} &\subset \Omega_2 = \Omega. \end{aligned}$$

Thus, the condition (i) is fulfilled. Now let $\lambda > 0$ be satisfied the condition $\lambda \sup_{x\in\Omega} \|f''(x)\| \le 1$. Then the condition (ii) is verified with $\varphi(x) = \frac{1}{2}\|x\|^2, \ \forall x \in \Omega$. In this case DCA applied to (25) is given by (x^o chosen in advance)

$$y^k = x^k - \lambda\nabla f(x^k); \quad x^{k+1} = P_C(x^k - \lambda\nabla f(x^k)), \tag{26}$$

where P_C denotes the orthogonal projection mapping. One can recognize in (26) the Goldstein-Levitin-Polyak projection method ([89]). The following result is an immediate consequence of the DCA's general convergence theorem (Theorem 3).

Proposition 7. *If λf is differentiable and less convex than $\frac{1}{2}\|\cdot\|^2$ on an open convex set containing C, then the sequence $\{x^k\}$ generated by (26) satisfies the following properties*

(i) *The sequence $\{f(x^k)\}$ is decreasing. If α is finite, then $\lim_{k\to+\infty}\|x^{k+1}-x^k\|=0$.*

(ii) *If α is finite and if the sequence $\{x^k\}$ is bounded, then every limit point x^* of $\{x^k\}$ is a solution to (25). In this case $f(x^k)\downarrow f(x^*)$. If, in addition, (Q) admits a unique solution, then the whole sequence $\{x^k\}$ converges to this solution.*

Consider now the special case where $\Omega_2 = X$. Then $\frac{1}{2}\|\cdot\|^2$ is more convex than λf (on X), (see Section 6). It amounts to say that $f = \theta_\lambda$ with $\theta \in \Gamma_o(X)$ according to Proposition 4. In this case we take $\Omega_1 = \Omega_2$, so the iteration (26) becomes

$$x^{k+1} = P_C(I+\lambda\partial\theta)^{-1}x^k, \tag{27}$$

since

$$\nabla f = \frac{1}{\lambda}[I-(I+\lambda\partial\theta)^{-1}].$$

Proposition 8. *Assume that α is finite. If $\frac{1}{2}\|\cdot\|^2$ is more convex than λf, then the sequence $\{x^k\}$ generated by (27) is bounded if and only if (Q) admits a solution. In this case the whole sequence $\{x^k\}$ converges to a solution x^* to (Q).*

Proof. It is clear that the sequence $\{x^k\}$ generated by (27) satisfies (i) of Proposition 7. Since the solution set of (Q) and the set of fixed points of the nonexpansive mapping $P_C(I+\lambda\partial\theta)^{-1}$ are identical, the proof of Proposition 8 is straightforward by using the property (ii) of Proposition 7 and the following estimate

$$\|x^{k+1}-x^*\| \le \|x^k - x^*\|$$

for every solution x^* to (Q).

Remark 3. (i) It is worth noting the equivalence: $f\in\Gamma_o(X)$ is less convex than $\frac{\lambda}{2}\|\cdot\|^2$ if and only if f is differentiable and has a Lipschitzian gradient mapping with the constant $1/\lambda$ (see [38], [89]).

(ii) Propositions 7 and 8 suitably complement the convergence result of the Goldstein-Levitin-Polyak gradient projection algorithm ([89]).

8. Exact penalty, Lagrangian duality without gap and dimensional reduction in d.c. programming

8.1. Exact penalty in d.c. programming

Consider the problem

$$(\mathrm{P}_3) \qquad \alpha = \inf\{g(x) - h(x) : \ x \in C, \ f_1(x) - f_2(x) \leq 0\},$$

where g, h, f_1 and f_2 belong to $\Gamma_o(X)$ and C is a nonempty closed convex set in X. It is not a d.c. program but we can use the exact penalty technique to transform (P_3) into a d.c. program. Indeed, since $(f_1-f_2)^+ = \max(f_1, f_2) - f_2$, the usual exact penalty introduces the following non-differentiable d.c. program

$$(\mathrm{P}_3)_t \qquad \alpha_t = \inf\{g(x) + t\max(f_1(x), \ f_2(x)) - (h(x) + tf_2(x)) \ : \ x \in C\}.$$

The crucial point is to prove the effective exactness of such a penalty, i.e., the existence of a positive number t_0 such that (P_3) and $(\mathrm{P}_3)_t$ are equivalent for all $t \geq t_0$.

This technique has been successfully applied to a class of problems of type (P_3) ([6]):

$$\alpha = \inf\{f(x) : \ x \in K, \ g(x) \leq 0\}, \tag{28}$$

where K is a nonempty bounded polyhedral convex set in X and f, g are finite concave functions on K. Exact penalty involves the problems

$$\alpha_+(t) = \inf\{f(x) + tg^+(x) : \ x \in K\} \tag{29}$$

with $g^+(x) = \max(0, g(x))$.

If the vertex set of K, $V(K)$, is contained in $\{x \in K : \ g(x) \leq 0\}$ we set t_o be 0; otherwise, we always have $t_o \leq \frac{f(x^o)-\alpha_+(0)}{S}$ for every $x^o \in K, \ g(x^o) \leq 0$, where $S = \min\{g(x) : \ x \in V(K), \ g(x) > 0\}$.

Theorem 9. *Assume that the feasible set of* (28) *is nonempty and g is nonnegative on K. Then for $t > t_o$ the solution sets of Problems* (28) *and* (29) *are identical.*

The class of Problems (28) satisfying the assumption of Theorem 9 contains many important real-life ones ([7]): Convex maximization over the Pareto set, bilevel linear programs, linear programs with mixed linear complementarity constraints, mixed zero-one concave minimization programming, etc.

8.2. Lagrangian duality without gap in d.c. programming

Let us present now two important results concerning the Lagrangian duality without gap in d.c. programming. The first deals with the maximization of a gauge on the unit ball of another gauge.

Let $\psi \not\equiv 0$ and $\phi \not\equiv 0$ be two finite gauges on X such that $V = \phi^{-1}(0)$ is a subspace contained in $\psi^{-1}(0)$. Consider the following problem

$$(\mathrm{P}^1_{\max}) \qquad S_{\psi\phi}(I) = \sup\{\psi(x): \ \phi(x) \leq 1\},$$

which is equivalent to $-S_{\psi\phi}(I) = \inf\{-\psi(x): \ \phi(x) \leq 1\}$.

We consider the Lagrangian duality for this problem in the case where the feasible domain is written as $\{x \in X: \ (1/2)\phi^2(x) \leq (1/2)\}$. We can write $(P^1_{\max})$ in the form

$$(\mathrm{P}_{\max}) \qquad \alpha = -S_{\psi\phi}(I) = \inf\Big\{-\psi(x): \ \frac{1}{2}\phi^2(x) \leq \frac{1}{2}\Big\}.$$

We will establish, like in a case of convex optimization, the stability in the Lagrangian duality, namely, we will prove that there is no gap between the optimal value of the primal and dual problems. These results allow us to obtain equivalent d.c. forms of $(\mathrm{P}_{\max})$ and an explicit form of the graph of the objective function for the dual problem. They can also be used for checking the globality of the solution computed by DCA.

Let $U(\phi)$ and $S(\phi)$ denote the unit ball and its sphere, respectively, i.e.,

$$U(\phi) = \{x \in X : \phi(x) \leq 1\}; \quad S(\phi) = \{x \in X : \phi(x) = 1\}.$$

First, we observe that the solution set of Problem $(\mathrm{P}_{\max})$, denoted by $\mathcal{P}_{\max}$, is contained in $S(\phi)$, and the finiteness of $S_{\psi\phi}(I)$ implies $\phi^{-1}(0) \subset \psi^{-1}(0)$.

Lemma 4 ([78], [84]). *$\mathcal{P}_{\max} \neq \emptyset$ and Problem $(\mathrm{P}_{\max})$ is equivalent to*

$$(\mathrm{PR}_{\max}) \qquad \sup\{\psi(x): \ x \in V^{\perp}, \ \phi(x) \leq 1\},$$

with $\mathcal{P}_{\max} = V + \mathcal{PR}_{\max}$, where $\mathcal{PR}_{\max}$ is the solution set of $(\mathrm{PR}_{\max})$.

Proof. Since $X = V + V^{\perp}$, for every $x \in X$ we have $x = u + v$, $u \in V$, where $v \in V^{\perp}$. Thus, $\psi(x) = \psi(u+v) = \psi(v)$, $\phi(x) = \phi(u+v) = \phi(v)$, which implies that $P_{\max}$ is equivalent to $(\mathrm{PR}_{\max})$. Since $\{x \in V^{\perp} : \ \phi(x) \leq 1\}$ is compact, the solution set of $(\mathrm{PR}_{\max})$ is nonempty. Furthermore, $\mathcal{P}_{\max} = V + \mathcal{PR}_{\max}$. $\square$

The Lagrangian $L(x, \lambda)$ for $(\mathrm{P}_{\max})$ is given by

$$L(x,\lambda) = \begin{cases} -\psi(x) + \frac{\lambda}{2}(\phi^2(x) - 1) & \text{if } \lambda \geq 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

Clearly, $-\psi(x) + \chi_C(x) = \sup\{L(x,\lambda) : \ \lambda \geq 0\}$. Thus, $(\mathrm{P}_{\max})$ can be written as

$$\alpha = -S_{\psi\phi}(I) = \inf\{\sup\{L(x,\lambda) \ : \lambda \geq 0\} \ : x \in X\}.$$

For $\lambda \geq 0$ we have

$$(P_\lambda) \qquad \gamma(\lambda) = \inf\Big\{ -\psi(x) + \frac{\lambda}{2}(\phi^2(x) - 1) \ : \ x \in X \Big\}.$$

As for $(\mathrm{P}_{\max})$, solving (P_λ) amounts to solving

$$(\mathrm{PR}_\lambda) \qquad \gamma(\lambda) = \inf\Big\{ -\psi(x) + \frac{\lambda}{2}(\phi^2(x) - 1) \ : x \in V^{\perp} \Big\}$$

knowing that $\mathcal{P}_\lambda = V + \mathcal{PR}_\lambda$.

It is clear that γ is a concave function and (P_λ) is a d.c. optimization problem.

The dual problem of $(\mathrm{P}_{\max})$ is

$$(D) \quad \beta = \sup\{\gamma(\lambda) \ : \ \lambda \geq 0\} = \sup\{\inf\{L(x,\lambda) \ : \ x \in X\} \ : \ \lambda \geq 0\}.$$

By the definition of Lagrangian we have

$$\begin{aligned} \alpha &= \inf\{\sup\{L(x,\lambda) \ : \ \lambda \geq 0\} \ : \ x \in X\} \\ &\geq \sup\{\inf\{L(x,\lambda) \ : \ x \in X\} \ : \ \lambda \geq 0\} = \beta. \end{aligned} \tag{30}$$

A point $(x^*, \lambda^*) \in X \times \mathbb{R}$ is said to be a saddle point of $L(x, \lambda)$ if

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*), \ \forall (x,\lambda) \in X \times \mathbb{R}.$$

Let us state now important results concerning characterization of solutions of the dual problem and the stability of the Lagrangian duality.

Theorem 10 ([78]). (i) *$\mathcal{P}_\lambda \neq \emptyset$ for every $\lambda > 0$ and* $\operatorname{dom}\gamma =]0, +\infty[$.

(ii) *$\mathcal{D} \neq \emptyset$ and $\gamma(\lambda) = \frac{-\lambda}{2} + \frac{K}{\lambda}$, where K is a negative constant (dependent only on ψ and ϕ).*

(iii) *$\mathcal{D} = \{\lambda^*\} = \{\sqrt{-2K}\}$, $\gamma(\lambda^*) = -\lambda^*$.*

(iv) *$\alpha = \beta$ and $\mathcal{P}_{\max} = \mathcal{P}_{\lambda^*}$.*

(v) *$(x^*, \lambda^*) \in \mathcal{P}_{\max} \times \mathcal{D}$ if and only if (x^*, λ^*) is a saddle point of $L(x, \lambda)$.*

(vi) *$\mathcal{P}_{\max} = \{\frac{x^*}{\phi(x^*)} : x^* \in \mathcal{P}_1\}$.*

Consider now Problem (P_λ) with $\lambda = 1$.

$$(\mathrm{P}_1) \qquad \gamma(1) + \frac{1}{2} = \inf\left\{\frac{1}{2}\phi^2(x) - \psi(x) : x \in X\right\}.$$

The following result allows determining a new d.c. optimization problem which is equivalent to $(\mathrm{P}_{\max})$.

Theorem 11 ([78]).

$$\mathcal{P}_1 = \{\psi(x^*)x^* : x^* \in \mathcal{P}_{\max}\}. \tag{31}$$

The second result concerns the minimization of a quadratic form on Euclidean balls (the so-called *trust region subproblem*) or spheres

$$\alpha_1 = \min\left\{q(x) = \frac{1}{2}x^T Ax + b^T x : \quad \|x\| \leq r\right\}, \tag{32}$$

$$\alpha_2 = \min\left\{q(x) = \frac{1}{2}x^T Ax + b^T x : \quad \|x\| = r\right\}, \tag{33}$$

where A is an $n \times n$ real symmetric matrix, $b \in I\!\!R^n$, r is a positive number and $\|\cdot\|$ denotes the Euclidean norm of $I\!\!R^n$.

We present first the results concerning the stability of the Lagrangian duality for Problem (32). To this end, we rewrite (32) as

$$\alpha_1 = \min\left\{q(x) : \frac{1}{2}\|x\|^2 \leq \frac{r^2}{2}\right\}.$$

Theorem 12 ([82]). (i) *$\alpha_1 = \beta_1$.*

(ii) *The dual problem has a unique solution λ^* and the solution set of the primal problem* (32) *is the set*

$$\{x^* \in \mathcal{P}_{\lambda^*} : \|x^*\| \leq r, \, \lambda^*(\|x^*\| - r) = 0\}.$$

As an immediate consequence we obtain the well-known optimality condition for (32) whose proof is not trivial.

Corollary 5 ([63], [77], [82]). *x^* is a solution to* (32) *if and only if there exists $\lambda^* \geq 0$ such that*

(i) $(A + \lambda^* I)$ *is positive semi-definite,*

(ii) $(A + \lambda^* I)x^* = -b$,

(iii) $\lambda^*(\|x^*\| - r) = 0, \|x^*\| \leq r$.

Such a λ^* is the unique solution to the dual problem.

By the same approach we obtain analogous results for Problem (33). The sole difference between these problems is that λ^* is not assigned to be nonnegative in the latter.

Remark that Problems (32) and (33) are among a few nonconvex optimization problems which possess a complete characterization of their solutions. These problems play an important role in optimization and numerical analysis ([1], [19], [31], [63], [77], [82], [83], [86], [98], [99], [100]).

8.3. Dimensional reduction technique in d.c. programming

Let $h \in \Gamma_o(X)$ and $h0^+$ be its recession function ([93]). The lineality space of h is denoted by [93]

$$L(h) = \{u \in X : \; h0^+(u) = -h0^+(-u)\}.$$

It is a subspace of the direction u in which h is affine:

$$h(x + \lambda u) = h(x) + \lambda\nu, \quad \forall x \in X, \; \forall \lambda \in I\!R,$$

where $h0^+(u) = -h0^+(-u) = \nu$. We have $\dim L(h) + \dim h^* = n$. More precisely, we have the following decomposition ([93])

$$X = V + V^{\perp},$$

where V is the subspace parallel to the affine hull of $\operatorname{dom} h^*$ (denoted by $\operatorname{aff}(\operatorname{dom} h^*)$), and $V^{\perp}$ is exactly $L(h)$. The function h can then be decomposed as

$$h(a + b) = h(a) + \langle w, b\rangle,$$

where $(a, b) \in V \times V^{\perp}$ and w is an arbitrary element of $\operatorname{aff}(\operatorname{dom} h^*)$. In other words, if P is an $n \times p$ matrix and Q is an $n \times q$ matrix such that the columns of P (resp. Q) constitue an orthonormal basis of V (resp. $V^{\perp}$), then we have

$$x = a + b = Pu + Qv,\ u \in I\!R^p,\ v \in I\!R^q;$$
$$h(x) = h(Pu + Qv) = h(Pu) + \langle Q^T w, v\rangle.$$

A d.c. program

$$\text{(P)} \qquad \alpha = \inf\{g(x) - h(x)\ :\ x \in X\}$$

is said to be weakly nonconvex, if $p = \dim h^*$ is small. In this case we have

$$\begin{aligned} f(x) = g(x) - h(x) &= g(Pu + Qv) - h(Pu + Qv) \\ &= g(Pu + Qv) - \langle Q^T w, v\rangle - h(Pu). \end{aligned}$$

The nonconvex part of $f(Pu + Qv)$ appears only in $h(Pu)$ which simply involves p dimensions. The dimensional reduction technique can improve DCA's qualities. It also permits global algorithms to treat large scale weakly nonconvex d.c. programs.

9. Applications

We present in this section the d.c. approach to some important nonconvex problems that DCA have been successfully applied to.

9.1. DCA for globally solving the trust region subproblem (TRSP)

Recall that TRSP is the problem of the form (32)

$$\alpha_1 = \min\left\{q(x) = \frac{1}{2}x^T Ax + b^T x :\ \|x\| \le r\right\}.$$

J. M. Martinez ([60]) has investigated the nature of local-nonglobal solutions of TRSP and shown the following property: TRSP has at most one local nonglobal solution. Moreover, being inspired by G.E. Forsythe & G.H. Golub's work ([28]), S. Lucidi *et al.* ([55]) have stated a very nice result: in TRSP the objective function can admit at most $2m+2$ different values at Kuhn-Tucker points, where m is the number of distinct negative eigenvalues of A.

Clearly, this problem is a d.c. program with different d.c. decompositions. Let us point out some examples, which are of particular interest.

(i) $f(x) = g(x) - h(x)$ with

$$g(x) = \frac{1}{2}\langle A_1 x, x\rangle + \langle b, x\rangle + \chi_C(x),$$

and

$$h(x) = \frac{1}{2}\langle A_2 x, x\rangle.$$

The matrices A_1 and A_2 are symmetric positive semidefinite related to the spectral decomposition of A: $A = A_1 - A_2$ with

$$A_1 = \sum_{i\in K} \lambda_i u_i u_i^T, \quad A_2 = -\sum_{i\in L} \lambda_i u_i u_i^T.$$

Here $\lambda_1 \leq \ldots \leq \lambda_n$ are the eigenvalues of A and $\{u_1, \ldots, u_n\}$ is an orthonormal basis of $\mathbb{R}^n$ constituted by corresponding eigenvectors of A, $K := \{i : \ \lambda_i \geq 0\}$, $L := \{i : \ \lambda_i < 0\}$, $C := \{x \in \mathbb{R}^n : \ \|x\| \leq r\}$.

(ii) $f(x) = g(x) - h(x)$ with

$$g(x) = \frac{1}{2}\langle (A + \rho I)x, x\rangle + \langle b, x\rangle + \chi_C(x),$$

and

$$h(x) = \frac{1}{2}\rho\|x\|^2.$$

The positive number ρ is chosen such that the matrix $A + \rho I$ is positive semidefinite, i.e., $\rho \geq -\lambda_1$.

(iii) $f(x) = g(x) - h(x)$ with

$$g(x) = \frac{\rho}{2}\|x\|^2 + \langle b, x\rangle + \chi_C(x),$$

and

$$h(x) = \frac{1}{2}\langle (\rho I - A)x, x\rangle.$$

The positive number ρ, in this case, should render the matrix $\rho I - A$ positive semidefinite, i.e., $\rho \geq \lambda_n$.

We can describe without any difficulty the corresponding DCA. However, for these decompositions, we see that DCA is explicit in the case (iii) only:

$$y^k = (\rho I - A)x^k \rightarrow x^{k+1} = P_C(x^k - \frac{1}{\rho}(Ax^k + b)).$$

Here P_C is the orthogonal projection mapping on C.

From the computational viewpoint, a lot of our numerical experiments proved the robustness and the efficiency of DCA with respect to the other well known algorithms, especially in the large scale trust-region subproblems ([83], [96], [99]).

9.2. DCA for solving the Multidimensional scaling problem (MDS)

The Multidimensional Scaling Problem (MDS) has a fundamental role in statistics by various applications in different fields such as social sciences ([13]), biochemistry ([21], [22]), psychology ([52]), mathematical psychology ([11], [97]), etc. The mathematical formulation of this important problem, due to Kruskal ([50]), is as follows:

Let ϕ be a norm on $\mathbb{R}^p$ and d be its corresponding metric. For an $n \times p$ matrix $\mathcal{X}$ we define

$$d_{ij}(\mathcal{X}) = d(\mathcal{X}_i^T, \mathcal{X}_j^T) = \phi(\mathcal{X}_i^T - \mathcal{X}_j^T),$$

where $\mathcal{X}_i$ is the ith row of $\mathcal{X}$.

Let two matrices $\Delta = (\delta_{ij})$ and $W = (w_{ij})$ of order n be given such that

$$\delta_{ij} = \delta_{ji} \geq 0\,,\ w_{ij} = w_{ji} \geq 0\,,\ \forall i \neq j;\ \delta_{ii} = w_{ii} = 0\ \forall i, j = 1, \ldots, n.$$

The normal case corresponds to $w_{ij} = 1$, $\forall i \neq j$. We shall refer to Δ as the dissimilarity matrix and W as the weight matrix. A metric MDS problem consists of finding in $\mathbb{R}^p$ n objects $\tilde{\mathcal{X}}_1, \ldots, \tilde{\mathcal{X}}_n$, representing an $n \times p$ matrix $\tilde{\mathcal{X}}$, such that the differences $\mid d(\mathcal{X}_i, \mathcal{X}_j) - \delta_{ij} \mid$ for every i and j are smallest.

Let $\mathcal{M}_{n,p}(\mathbb{R})$ denote the space of all real matrices of order $n \times p$. The mathematical form of metric MDS problem then can be written as

$$\text{(MDS)}\quad \sigma_* = \min\left\{\sigma(\mathcal{X}) = \frac{1}{2}\sum_{i<j} w_{ij}\theta(\delta_{ij} - d_{ij}(\mathcal{X}))\,:\ \mathcal{X} \in \mathcal{M}_{n,p}(\mathbb{R})\right\},$$

where $\theta(t) = |t|$ (resp. t^2) for $t \in I\!R$.

Other mathematical models have been discussed in [50], [102]. The classical metric MDS problem corresponds to $\theta(t) = t^2$ ([50]). The function σ is called the loss function. So, this problem is also sometimes referred to as the distance geometry problem. Semidefinite programming problems have been formulated by multidimensional and clustering techniques in [61]. Recently similar models have been used for the molecule problem [34], the protein structure determination problem ([116], [117]) and the protein folding problem ([30]). Some gradient-type methods were developed by J.B. Kruskal [48], [49], L. Gutman [33], J.C. Lingoes, E.E. Roskam [53] for solving this problem. These gradient-type methods fail to use if the loss function σ is nondifferentiable ([24], [25]).

In 1977 J. de Leeuw [24] proved that the classical metric MDS problem is equivalent to the convex maximization problem given as

$$(S_{\max}) \qquad \max\left\{\frac{\rho(\mathcal{X})}{\eta(\mathcal{X})} : \eta(\mathcal{X}) \neq 0\right\} \Leftrightarrow \max\left\{\rho(\mathcal{X}) : \eta(\mathcal{X}) \leq 1\right\},$$

where

$$\eta(\mathcal{X}) = \left[\sum_{i<j} w_{ij} d_{ij}^2(\mathcal{X})\right]^{1/2} \quad \text{and} \quad \rho(\mathcal{X}) = \sum_{i<j} w_{ij}\delta_{ij} d_{ij}(\mathcal{X})$$

are two seminorms in $\mathcal{M}_{n,p}(I\!R)$. So, the classical MDS problem is reduced to the problem of maximization of the semi-norm ρ on the unit ball defined by the seminorm η. For solving the metric Euclidean MDS with $\theta(t) = t^2$ de Leeuw proposed an algorithm [24] (see also [25]), called the majorization method, similar to Guttman's C-matrix method whose convergence result has been stated by following the same scheme of the subgradient method (for maximizing a semi-norm on the unit ball of another semi-norm) which has been studied extensively in the general framework of convex analysis by Pham Dinh Tao ([70]-[74]). De Leeuw merely used the subgradient method ([70]-[74]) for solving the classical metric MDS problem with noneuclidean norm ϕ ([24]). In the Euclidean case de Leeuw observed judiciously that his elegant adapted algorithm is unlike the subgradient method since $\lim_{k\to+\infty} \|\mathcal{X}^{k+1} - \mathcal{X}^k\|$ is not necessarily zero in the former. In fact, such a property can be obtained by applying regularization techniques to subgradient methods (see Section 5). De Leeuw's algorithm has been so successful that it has become a reference method to the classical metric Euclidean MDS problem ($\theta(t) = t^2$). Remark that de Leeuw's approach fails to use in the case where $\theta(t) = |t|$.

(MDS) with $\theta(t) = t^2$ in fact is a d.c. optimization problem of the form

$$(34) \qquad (S_1) \qquad \min\left\{\frac{1}{2}\eta^2(\mathcal{X}) - \rho(\mathcal{X}) : \quad \mathcal{X} \in \mathcal{M}_{n,p}(I\!R)\right\},$$

since

$$\begin{aligned}\sigma(\mathcal{X}) &= \frac{1}{2}\sum_{i<j} w_{ij} d_{ij}^2(\mathcal{X}) + \sum_{i<j} w_{ij}\delta_{ij}^2 - \sum_{i<j} w_{ij}\delta_{ij} d_{ij}(\mathcal{X}) \\ &= \frac{1}{2}\eta^2(\mathcal{X}) - \rho(\mathcal{X}) + \sum_{i<j} w_{ij}\delta_{ij}^2.\end{aligned}$$

It has been proved in [84] that the seminorms η and ρ can be expressed as

$$\eta(\mathcal{X}) = \frac{1}{2}\langle T\mathcal{X}, \mathcal{X}\rangle, \quad \rho(\mathcal{X}) = \frac{1}{2}\langle B(\mathcal{X})\mathcal{X}, \mathcal{X}\rangle,$$

where the matrices $T = (t_{ij})$ and $B(\mathcal{X}) = (b_{ij}(\mathcal{X}))$ are defined by

$$t_{ij} = \begin{cases} -w_{ij} & \text{if } i \neq j, \\ \sum_{j=1}^{n} w_{ij} & \text{if } i = j. \end{cases} \qquad b_{ij} = \begin{cases} -w_{ij}\delta_{ij}s_{ij}(\mathcal{X}) & \text{if } i \neq j, \\ -\sum_{j=1, j\neq i}^{n} b_{ij} & \text{if } i = j. \end{cases}$$

$$s_{ij}(\mathcal{X}) = \begin{cases} 1/(\|\mathcal{X}_i^T - \mathcal{X}_j^T\|) & \text{if} \mathcal{X}_i \neq \mathcal{X}_j, \\ 0 & \text{otherwise.} \end{cases}$$

The stability of the Lagrangian duality for Problem ($S_{\max}$) (Subsection 8.2) permits finding various d.c. decompositions for this problem, in particular, the equivalence of ($S_{\max}$) and (S_1). Moreover, the expression of the dual objective function can be used to check the globality of solutions computed by DCA.

Solving Metric MDS Problem by DCA:

Let $\mathcal{A}$ denote the set of matrices in $\mathcal{M}_{n,p}(I\!R)$ whose rows are identical, i.e.,

$$\mathcal{A} := \{\mathcal{X} \in \mathcal{M}_{n,p}(I\!R) : \mathcal{X}_1 = \cdots = \mathcal{X}_n\}.$$

It is a subspace of $\mathcal{M}_{n,p}(I\!R)$. Its orthogonal subspace is $\mathcal{A}^\perp = \{\mathcal{Y} \in \mathcal{M}_{n,p}(I\!R) : \sum_{i=1}^{n} \mathcal{Y}_i = 0\}$.

We can assume, without loss of generality, the irreducibility of the weight matrix W, otherwise the MDS problem can be decomposed into a number of smaller problems. We will give below the description DCA applied to (S_1) and $(S_{\max})$ for solving the metric MDS problem with $\theta(t) = t^2$ and the Euclidean norm ϕ. Problems (S_1) and $(S_{\max})$ then take the form (P_1) and $(P_{\max})$, where the semi-norms ρ and η replace the gauges ψ and ϕ, respectively (Subsection 8.2)

From the stability results obtained in the last section (Theorem 11) solving Problem (P_1) amounts to solving $(P_{\max})$

(35)

$$(P_{\max}) \quad \max\{\rho(\mathcal{X}) : \eta(\mathcal{X}) \leq 1\} \Leftrightarrow \min\{\chi_C(\mathcal{X}) - \rho(\mathcal{X}) : \mathcal{X} \in \mathcal{M}_{n,p}(I\!R)\},$$

where $\mathcal{C} = \{\mathcal{X} \in \mathcal{M}_{n,p}(I\!R) : (1/2)\eta^2(\mathcal{X}) \leq 1/2\}$. The last is a d.c. program.

Now we describe DCA (with and without regularization) for solving Problems (P_1) and $(P_{\max})$. First, observe that for both (P_1) and $(P_{\max})$ we can restrict ourselves within to $\mathcal{A}^\perp$, where each of these problems has always a solution ([1], [84]).

The DCA applied to (P_1) and $(P_{\max})$ is reduced to calculating subdifferentials of the functions $\rho, ((1/2)\eta^2)^*$, $\chi_C^*, [(\lambda/2)\eta^2 + (\mu/2)\|.\|^2]^*$ and $[\lambda\chi_C + (\mu/2)\|.\|^2]^*$.

We have first calculated these subdifferentials for the general case and then particularized the results in the normal case ($w_{ij} = 1, \ \forall i \neq j$) ([1], [84]).

Let $\varepsilon > 0$, $\lambda > 0$, $\mu > 0$ and $\mathcal{X}^o \in \mathcal{A}^\perp$ be given.

Let $k = 0, 1, \ldots$ until $\|\mathcal{X}^{k+1} - \mathcal{X}^k\| \leq \varepsilon$.

DCAP$_1$: Take $\mathcal{X}^{k+1} = \frac{B(\mathcal{X}^k)\mathcal{X}^k}{n}$

RDCAP$_1$: Take $\mathcal{X}^{k+1} = \frac{(\lambda B(\mathcal{X}^k)+\mu I)\mathcal{X}^k}{\mu+\lambda n}$

DCAP$_{\max}$: Take $\mathcal{X}^{k+1} = \frac{B(\mathcal{X}^k)\mathcal{X}^k}{\eta(B(\mathcal{X}^k)\mathcal{X}^k)}$

RDCAP$_{\max}$: Take $\mathcal{X}^{k+1} = \begin{cases} \frac{(\lambda B(\mathcal{X}^k)+\mu I)\mathcal{X}^k}{\mu}, & \text{if } \eta((\lambda B(\mathcal{X}^k) +\mu I)\mathcal{X}^k) \leq \mu, \\ \frac{(\lambda B(\mathcal{X}^k)+\mu I)\mathcal{X}^k}{\eta((\lambda B(\mathcal{X}^k)+\mu I)\mathcal{X}^k)}, & \text{otherwise.} \end{cases}$

Recall that $\eta(\mathcal{Y}) = \sqrt{n}\|\mathcal{Y}\|$ in $\mathrm{DCAP_{max}}$ and $\mathrm{RDCAP_{max}}$.

The original d.c. decomposition for the objective function of (P_1) in (34) seems to be better than the corresponding one relative to $(\mathrm{P_{max}})$ in (35). As direct consequences of the above description of DCA, we have:

(i) Without computing the denominator in each step, DCAP_1 (resp. $\mathrm{RDCAP_1}$) is less expensive than $\mathrm{DCAP_{max}}$ (resp. $\mathrm{RDCAP_{max}}$), and the difference of the cost is proportional to the dimension $n \times p$ of (MDS).

(ii) On the practical point of view, the first two algorithms are obivously more stable than the last ones, especially in large-scale MDS problems.

These observations have been justified by many numerical simulations. Let us point out now the fact that de Leeuw's algorithm for solving the metric Euclidean MDS problem (in the case where the weight matrix is supposed to be irreducible) is a special case of DCA applied to the d.c. program (P_1). Indeed DCAP_1 is given by (for a given $\mathcal{X}^o \in \mathcal{A}^\perp$)

$$\mathcal{X}^{k+1} = T^+ B(\mathcal{X}^k)\mathcal{X}^k \in \mathcal{A}^\perp,$$

where the computation of $X = T^+Y$ with $Y \in \mathcal{A}^\perp$ amounts solving the nonsingular linear system $(T + (1/n)ee^T)\mathcal{X} = \mathcal{Y}$ ([1], [84]) We rediscover thus de Leeuw's algorithm [24], [25].

Normal metric Euclidean MDS as parametrized trust region subproblems and its solution by the parametrized DCA

The particular structure of the seminorms η and ρ have enabled us to formulate the normal metric Euclidean MDS problem (P_1) as parametrized trust region subproblems and to devise the parametrized DCA for solving (P_1). This approach is promising because the DCA is robust and efficient for globally solving the trust region subproblem ([83]).

It is worth noting that if the rugularization parameter is fixed, then the corresponding parametrized DCA is reduced to $\mathrm{RDCAP_{max}}$.

9.3. DCA for solving a class of linearly constrained indefinite quadratic problems

We consider the indefinite quadratic problem over a bounded polyhedral convex set which plays an important role in global optimization:

$$(\mathrm{IQP}_1) \qquad \min\left\{\frac{1}{2}\langle Hx, x\rangle + \langle l, x\rangle \ : \ x \in K\right\}.$$

Here H is a symmetric indefinite $q \times q$-matrix, $l \in I\!\!R^q$, K is a nonempty bounded polyhedral set defined as $K = \{x \in I\!\!R^q : \ Ax \leq a, \ x \geq 0\}$ with

A being an $m \times q$-matrix, $a \in I\!R^m$. When

$$H = \begin{pmatrix} \tilde{C} & 0 \\ 0 & D \end{pmatrix}$$

and the polytope is defined as

$$\Omega = \{(x,y) \in I\!R^n \times I\!R^s : \tilde{A}x + By \leq a,\ A_1 x \leq a_1,\ A_2 y \leq a_2,\ x \geq 0,\ y \geq 0\}$$

we have the problem

$$(\mathrm{IQP}_2) \quad \min \Big\{ F(x,y) = \frac{1}{2}\langle \tilde{C}x, x\rangle + \langle c, x\rangle + \frac{1}{2}\langle Dy, y\rangle + \langle d, y\rangle : \ (x,y) \in \Omega \Big\}.$$

Here, $\tilde{C}$ is a symmetric positive semi-definite $n \times n$-matrix, D is a symmetric negative semi-definite $s \times s$-matrix, $c \in I\!R^n$, $d \in I\!R^s$, and $\tilde{A}$ is an $m \times n$-matrix, B is an $m \times s$-matrix, A_1 is an $r \times n$-matrix, A_2 is a $p \times s$-matrix, $a \in I\!R^m$, $a_1 \in I\!R^r$, $a_2 \in I\!R^p$. So the objective function of (IQP_2) is decomposed in a sum of a convex part and a concave part. A special case of (IQP_2) is the problem where D is diagonal (i.e., the concave part is separable):

$$(\mathrm{IQP}_3) \quad \min \Big\{ f(x,y) = \frac{1}{2}\langle \tilde{C}x, x\rangle + \langle c, x\rangle + \sum_{i=1}^{s}[d_i y_i - \frac{1}{2}\lambda_i y_i^2] : \ (x,y) \in \Omega \Big\}$$

with $\lambda_i > 0$.

Problem (IQP_1) is in fact a problem of the form (IQP_3). Likewise, Problem (IQP_2) can be equivalently transformed into a problem of the form (IQP_3) where the concave variable is separable.

When $\tilde{C} \equiv 0$ in (IQP_3) and the polytope is defined as

$$\bar{\Omega} = \{(x,y) \in I\!R^n \times I\!R^s : Ax + By \leq a,\ x \geq 0,\ y \geq 0\},$$

we have the linearly constrained concave quadratic problem which have been considered by several authors (see e.g. Rosen and Pardalos [92], Kalantari and Rosen [45], Pardalos et al. [66], Phillips and Rosen [88], etc.). In this case the global minimum point is always attained at least at a vertex of the convex polytope $\bar{\Omega}$. This property is no longer true when $\tilde{C} \not\equiv 0$. Hence, Problem (IQP_3) with $\tilde{C} \not\equiv 0$ is likely even more difficult to be numerically solved than concave programs. Recently, a decomposition branch-and-bound method for dealing with (IQP_3) in the case where $\tilde{C} \not\equiv 0$ was proposed in Phong-An-Tao [103]. This method based on normal rectangular subdivisions which exploit the separability of the concave part in the objective function. In general, the existing algorithms are efficient if the number of the concave variables is small.

Clearly, Problem (IQP_2) can be considered as minimization of a d.c. function over a polytope for which some method developed in global approaches (see e.g., Horst *et al.* [41], Tuy [112]) can be applied. For solving (IQP_2) in the case where the number of variables is large, we should avoid the inherent difficulties of this global optimization problem by using local approaches. We proposed a “good” d.c. decomposition for which numerical experience indicates that the DCA is efficient for solving (IQP_2). In contrast to global algorithms whose complexity increases exponentially with the dimension of the concave variable, DCA has the same behaviour with respect to both dimensions of convex variables and concave variables. Consequently, they solve these problems when the number of concave variables is large. For solving (IQP_1) we presented some d.c. decompositions and their corresponding DCA which seem to be efficient. We proposed also a decomposition branch-and-bound method for globally solving (IQP_1) and (IQP_2). These methods are just a modification of the one in Phong-An-Tao [103] for the general case. We used these algorithms for checking the globality of the solution computed by DCA when $s \le 30$. Computational experiments proved that

• The global algorithms only run until $s = 30$. The DCA is much faster than the global algorithm (until 30 times).

• In most problems the DCA gave global solutions.

• The DCA terminated very rapidly; the average number of iterations is 15.

9.4. A branch-and-bound method via D.C. Optimization Algorithms and Ellipsoidal technique for Nonconvex Quadratic Problems

For simplicity we present only the algorithm for solving box constrained nonconvex quadratic problems [3] (see [8]) for the general nonconvex quadratic program):

$$\text{(QB)} \qquad \min\Big\{f(x) := \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle : \\ -\infty < l_i \le x_i \le u_i < +\infty,\ i = 1, \dots, n\Big\},$$

with A being an $n \times n$ symmetric matrix, b, $x \in I\!R^n$.

A special case of (QB) is the unconstrained quadratic zero-one problem which has many important applications, particularly in combinatorial optimization, looks as follows

$$\min \left\{ f(x) := \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle : \ x \in \{0,1\}^n \right\}. \tag{36}$$

Indeed, writing (36) in the form

$$\min \left\{ \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle + \frac{t}{2}(\langle e, x\rangle - \langle x, x\rangle) : \ x \in \{0,1\}^n \right\}, \tag{37}$$

we have

$$(37) \Leftrightarrow \min \left\{ \frac{1}{2}\langle x, (A - tI)x\rangle + \langle b + \frac{t}{2}e, x\rangle : \ x \in [0,1]^n \right\}, \tag{38}$$

where e is the vector of ones and t is a real scalar such that $A - tI$ is negative semidefinite.

Several algorithms have been proposed for globally solving these problems (see e.g. [10], [42], [43], [67], [68], [90]). The procedures often used in the existing methods are to relax the constraints and perturb the objective function. The branching is usually based upon rectangular and/or simplicial partitions.

Ellipsoidal techniques have been used to state estimations for optimal values in nonconvex quadratic programming ([114], [115]). In [114] an approximation algorithm was proposed for finding an ε-approximate solution. It was shown that such an approximation can be found in polynomial time for fixed ε and t, where t denotes the number of negative eigenvalues of the quadratic term.

We have proposed a combined branch-and-bound algorithm and DCA for solving (QB). The aim of DCA is here twofold. The first use deals with the lower bound computation scheme. We apply DCA to solve the problem (QE) which is obtained from (QB) by replacing the rectangle constraint with an ellipsoid containing the selected rectangle. Our main motivation for the use of this technique is that DCA is very efficient for globally minimizing a quadratic form over an euclidean ball ([1], [83]). The second use concerns the upper bound computation process. For finding a good quality of feasible solutions we apply DCA to solving (QB) from good starting points which are obtained during computing lower bounds.

The global algorithm we are going to describe starts with an ellipsoid E^0 containing the feasible region K^0 of (QB). If a minimizer of f over E^0 is feasible then we stop the process. Otherwise, we divide K^0 into two rectangles (rectangular bisection) and construct two ellipsoids such that each of them contains one of the just divided rectangles. To improve the lower bound we compute minimums of f on each newly generated ellipsoid. The procedure is then repeated by replacing K^0 by a rectangle corresponding to the smallest lower bound and E^0 with an ellipsoid containing this rectangle. As this procedure repeats infinitely many times two infinite nested sequences of rectangles and ellipsoids are generated and shrink to a singleton. The convergence of the algorithm thus is ensured.

For computing lower bounds, we are concerned with the construction of a sequence of ellipsoids E^k such that:

(i) E^k contains the current rectangle K^k,

(ii) the volume $\text{vol}(E^k)$ decreases to zero as $k \to +\infty$,

(iii) the lower bounds (supplied by the solution of (QE) defined below) are sufficiently tight.

The lower bound of f in our algorithm is improved by the solution of the problem

$$\text{(QE)} \qquad \min\left\{\frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle : \, x \in X\right\}.$$

Again by the transformation L one has $x \in X$ if and only if $y = Lx \in S$, so the last problem is equivalent to the trust region subproblem

$$\text{(39)} \quad \min\left\{\frac{1}{2}\langle y, D^{-1}AD^{-1}y\rangle + \langle D^{-1}q - D^{-1}AD^{-1}c, y\rangle : \, \|y\|^2 \le n\right\}.$$

Clearly, $D^{-1}AD^{-1}$ is an $n \times n$ symmetric matrix. The DCA is used for globally solving (39).

The main contribution of our method is to give a good combination between the local and global approaches: first, we have provided a simple and efficient algorithm using the information of the bounding procedure for finding good local minima of box constrained quadratic problem; secondly, we have used the elippsoidal technique for bounding based on an efficient algorithm (DCA) for the ball constrained quadratic problem. We have studied also the new branching procedure in rectangular partition in order to well adapt to the new ellipsoidal bounding technique.

Preliminary experiments showed the efficiency as well as the limit of our algorithm. The advantages are that either an ε-optimal solution or a good feasible point is obtained rapidly by the DCA and thereby a considerable number of generated ellipsoids are deleted from further consideration. Therefore, in general the algorithm is efficient for a proper choice of tolerance ε (nevertheless this choice does not present always exactly the quality of the solution obtained, see, e.g., Table 6 in [3]). Moreover, it is very efficient for some classes of problems: it could treat until the dimension 1000 (Tables 1, 2, 3 in [3]). Again, the DCA for subproblems in bounding procedures are explicit and not expensive. The limit is that, as often happened in a branch-and-bound algorithm, the lower bound is improved slowly. That is why we introduced the second process of lower bounding in such a case. In the study of the performance of our algorithm, an open question is that which ellipsoid can be chosen to warrant the best initial lower bound? Meanwhile our ellipsoid technique seems to be quite suitable.

9.5. DCA for globally solving nonconvex quadratic programming

We consider the general nonconvex quadratic programming

$$\text{(QP)} \qquad \inf\left\{f(x) = \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle : \ Bx \le c\right\},$$

with A being an $n \times n$ symmetric matrix, B being an $m \times n$ matrix and $b \in I\!R^n, \ c \in I\!R^m$.

As in the TRSP, Kuhn-Tucker points of nonconvex quadratic program (QP) enjoy the following nice property due to Bomze & Danninger ([15]): *the objective function f can be assume at most $2^m + 1$ different values at Kuhn-Tuker points of* (QP).

We associate with the DCA (for solving nonconvex quadratic programs) an escaping procedure based on the trust region method ([63], [64], [77], [98]) for globally solving (QP). This is proceeded as follows:

Let x^* be a Kuhn-Tucker point computed by the DCA and let ε be a positive number. Consider the linear quadratic equations

$$(40) \qquad \begin{cases} f(x) = f(x^*) - \varepsilon, \\ Bx + v^2 = c, \end{cases}$$

where $v = (v_1, \dots, v_m) \in I\!R^m$ and v^2 denotes the vector $v = (v_1^2, \dots, v_m^2)$.

If the Kuhn-Tucker point x^* is not a global solution to (QP), then for ε sufficiently small there exists a solution $(\bar{x}, \bar{v})$ to (40) which can be computed by the trut region methods with the starting point (x^*, v^*), where ${v^*}^2 = c - Bx^*$. In this case we must restart DCA with the new initial point $\bar{x}$.

Finally, the DCA (with at most $2^m + 1$ restarting procedures) converges to a global solution of (QP).

9.6. D.c. approach for linearly constrained quadratic zero-one programming problems

We are interested in solving the well-known linearly constrained quadratic zero-one programming

$$(01QP) \quad \alpha = \min \left\{ f(x) := \frac{1}{2} x^T C x + c^T x : \; Ax \le b, \, x \in \{0,1\}^n \right\},$$

where C is an $(n \times n)$ symmetric matrix, c, $x \in I\!R^n$, A is an $(m \times n)$ matrix and $b \in I\!R^m$.

This problem, which is a key one in combinatorial optimization, is NP-hard and has been extensively investigated by several authors. A special case of this is linear zero-one programming which has many important applications in economics, planning and various kinds of engineering design. We propose a new algorithm in a continuous approach for the following equivalent concave minimization problem

$$(CCQP) \quad \alpha = \min \left\{ \frac{1}{2} x^T (C - \lambda I) x + (c + \frac{\lambda}{2} e)^T x : \; Ax \le b, \, x \in [0,1]^n \right\},$$

where e is the vector of ones and $\lambda \ge \lambda_o > 0$ (see Section 8.1 and [7]).

Clearly, (CCQP) is a d.c. program to which some global methods can be applied (see e.g. [41], [112]), when the dimension is small. For large scale setting it would be interesting to use the DCA for solving (CCQP).

In order to guarantee that solutions computed by the DCA are global we introduce a branch-and-bound scheme. With this purpose, first we transform (CCQP) into the d.c. program with the separable concave part

$$(DCQP) \quad \alpha = \min\{g(x) - h(x) : \; x \in K\}.$$

Here $K = \{x \in I\!R^n : Ax \leq b, 0 \leq x_i \leq 1, i = 1, \ldots, n\}$,

$$g(x) := \frac{1}{2}x^T(C + (\rho - \lambda)I)x + (c + \frac{\lambda}{2}e)^T x; \quad h(x) := \frac{1}{2}\rho\sum_{i=1}^{n} x_i^2, \tag{41}$$

and ρ is a positive number such that the matrix $(\rho - \lambda)I + C$ is semi positive definite. The branch-and-bound algorithm is in fact considered for the problem (DCQP). It aims to find a good starting point for the DCA and, naturally, to prove the globality of the solutions obtained by the DCA. This combination of two approaches improves the qualities of the DCA and accelerates the convergence of the branch-and-bound algorithm.

We note that the equivalence between (01QP) and (CCQP) is not unusual. Neverthelesse the existing algorithms for solving the last problem are not often applied in the solution of the former. The reason is that, in practice a branch-and-bound scheme (in a continuous approach) is often used for finding an ε-solution of the problem being considered, since it is very difficult to obtain an exact solution. Unfortunetly, such an ε-solution of (CCQP) may be not integer. However, the DCA applied to (CCQP) with t sufficiently large provides a point in $\{0, 1\}^n$. This advantage of the DCA makes our method efficient in finding an integer ε-solution of (DCQP) in large scale setting.

Preliminary computational experiences [5] showed that the DCA is very efficient: it provides either an optimal solution or a good approximation to it within some first iterations. As for the branch-and-bound algorithm, it works very well in linear zero-one programming. In case of quadratic programming, when $n > 40$, to ensure the globality the algorithm needed much more iterations. Another procedure for improving lower bounds in this case is currently under the research.

9.7. DCA for optimizing over the efficient set

Let C be a $p \times n$ real matrix, and K be a nonempty compact convex set of $I\!R^n$. Let f be a real valued function on $I\!R^n$. An optimization problem over the efficient set is that of the form

$$\max\{f(x) : x \in K_E\}. \tag{42}$$

Here K_E denotes the Pareto set of the following multiple objective programming problem:

$$\text{(MP)} \qquad \max\{Cx : x \in K\}.$$

We recall that a point x^0 is said to be an *efficient point* of Problem (MP) when $x^0 \in K$, and whenever $Cx \geq Cx^0$ for some $x \in K$, then $Cx = Cx^0$.

Following [12] for each x we define $p(x)$ as

$$(L_x) \qquad p(x) := \max\{e^T C(y - x) : Cy \geq Cx, y \in K\},$$

where e is the row vector in $\mathbb{R}^p$ whose all entries are 1. It is clear that the effective domain of $-p$ is the projection on $\mathbb{R}^n$ of $\{(x, y) \in \mathbb{R}^n \times K \ : \ Cx - Cy \leq 0\}$ which contains K. This function has the following properties:

Lemma 5 ([6], [7]). (i) *p is nonnegative on K, and $x \in K_E$ if and only if $p(x) = 0, \ x \in K$.*

(ii) *p is upper semicontinuous concave function on the whole space.*

(iii) *If, in addition, K is polyhedral convex, then p is polyhedral concave on K.*

In virtue of Lemma 5, Problem (P) can be written as

$$\text{(P')} \qquad \max\{f(x) : p(x) \leq 0, \ x \in K\}.$$

Since $p(x) \geq 0$ for every $x \in K$, and $p(x) = 0, \ x \in K$ if and only if $x \in K_E$, this function is an exact penalty function (with respect to K). Thus, the corresponding penalized problem is

$$\min\{tp(x) - f(x) : x \in K\}. \tag{43}$$

According to Theorem 9, if f is convex, then there is a nonnegative number t_o such that (42) and (43) are equivalent for $t > t_o$.

Remark that (43) is a polyhedral d.c. program if f is concave. Moreover, if f is linear and the vertex set of K, $V(K)$, is not contained in K_E, then for $t > (U_0 - f(x^0))/M_0$, with $U_0 := \max\{f(x) : x \in K\}$ and $M_0 := \min\{p(x) : x \in V(K), \ p(x) > 0\} > 0$, the sequence $\{x^k\}$ generated by DCA are in K_E provided the starting point x^0 is efficient.

Numerous computational experiments proved that ([6])

1. *For DCA:*

• DCA terminates very rapidly; average number of iterations is 3.

• DCA works with problems where the number of the criteria may be large. The running times are relatively insensitive to the increase of the number of the criteria.

• The optimal value computed by DCA depends on the initial point. In general the optimal value is close to the global optimal value if the initial point x^0 is chosen as

$$x^0 = \gamma\bar{x} + (1-\gamma)x^{max}, \quad 0.5 < \gamma < 1,$$

where

$$\bar{x} = \text{ argmax}\{\bar{\lambda}Cx : x \in K\}, \text{ with } \bar{\lambda} \in V(S_0),$$
$$x^{max} = \text{ argmax}\{f(x) : x \in K\}.$$

Although this initial point in general is not efficient, the computed solutions always are efficient.

• The algorithms are unstable if the penalty parameter is large.

2. *For global algorithms* (a combination of DCA and a based branch-and-bound global algorithm)

• The algorithm is efficient only if the number of the criteria is small, since the running time is very sensitive to the increase of the number of the criteria.

• In general, after a reasonable number of iterations a global solution is reached. To detect its globality, however, we need much more iterations. The reason is that the upper bounds are not sufficiently sharp, and sometimes they are improved very slowly.

9.8. Linear and nonlinear complementarity problems. Difference of subdifferentials complementarity problem

Consider the problem of the form

$$\text{(NLC)} \qquad \begin{cases} \text{Find } (x,y) \in I\!R^n \times F(x) \text{ such that} \\ y \geq 0 \text{ and } \langle x, y\rangle = 0, \end{cases}$$

where F is an operator defined on $I\!R^n$.

If $F(x) = (f_1(x), \ldots, f_n(x))^T$ is a d.c. vector-function (i.e., its components f_i are d.c.), then it is well-known that ([44]) (NLC) is equivalent to the following nonconvex program type (3) (Introduction)

$$0 = \inf\left\{\sum_{i=1}^{n} \min(x_i, f_i(x)) : \; x \geq 0, \; f_i(x) \geq 0, \; i = 1, \ldots, n\right\}.$$

The linear complementarity problem (LCP) corresponds to the case where F is affine, ($F(x) = Ax + b$) ([20]). (LCP) is equivalent to the following d.c. programs which are, nonconvex quadratic programs except (47) with $p \neq 1$ and (48):

$$\alpha = \inf \left\{ \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle : \ x \geq 0 \right\}, \tag{44}$$

with A being symmetric,

$$0 = \inf \left\{ \frac{1}{2}\langle x, (A + A^T)x\rangle + \langle b, x\rangle : \ x \geq 0, \ Ax + b \geq 0 \right\}, \tag{45}$$

$$0 = \inf \left\{ \|y - (Ax + b)\|^2 + x^T y : \ x \geq 0, \ y \geq 0 \right\}, \tag{46}$$

$$0 = \inf \left\{ \sum_{i=1}^{n} \min(x_i, A_i x + b_i) : \ x \geq 0 \right\}, \tag{47}$$

where $\rho > 1$ and $p > 1$,

$$0 = \inf \ \left\{ \sum_{i=1}^{n} \min(x_i, A_i x + b_i) : \ x \geq 0 \right\}. \tag{48}$$

The relations between solutions to these problems are straightforward.

(i) If A is symmetric then $\bar{x}$ is a solution to (LCP) if and only if $\bar{x}$ is a Kuhn-Tucker point of (44).

(ii) $\bar{x}$ is a solution to (LCP) if and only if the optimal value of (45) (resp. (48)) is zero and $\bar{x}$ is a solution to (45) (resp. (48)).

(iii) $\overline{x}$ is a solution to (LCP) if and only if the optimal value of (46) (resp. (47)) is zero and $(\overline{x}, \overline{y} = A\overline{x} + b)$ is a solution to (46) (resp. (47)).

It is worth noting that (48) is polyhedral d.c. program (polyhedral concave minimization over a polyhedral convex set).

Recently A. Friedlander *et al.* [29] have studied the solution of the horizontal linear complementarity problem:

Given two $n \times n$ matrices Q, R and $b \in I\!R^n$, find $x, y \in I\!R^n$ such that

$$Qx + Ry = b, \quad x^T y = 0, \quad x, y \geq 0 \tag{49}$$

by using box constrained minimization algorithms applied to the following equivalent problem

$$0 = \inf \left\{ \rho \|Qx + Ry - b\|^2 + \left(x^T y\right)^p : \; x \geq 0, \; y \geq 0 \right\}, \tag{50}$$

with arbitrary constants $\rho > 1$ and $p > 1$. It is clear that (LCP) (resp. (50)) is a particular case of (49) (resp. (50)). Observe that when $p = 2$ we obtain easily an interesting d.c. decomposition of the objective function.

Problem (49) is said to be monotone if $Qu + Rv = 0$ implies $u^T v \geq 0$. That is equivalent, in case Q (resp. R) is nonsingular, to $Q^{-1}R$ (resp. $R^{-1}Q$) negative semidefinite. Likewise (LCP) is monotone if and only if A is positive semidefinite.

A. Fridlander *et al.* [29] have stated interesting conditions under which Kuhn-Tucker necessary optimality conditions for (50) are also sufficient. In particular if (50) is monotone, then there is identity between Kuhn-Tucker points and solutions to (50).

In our new non-standard approaches ([85]) we used the DCA for solving these d.c. programs. Furthermore we have studied the difference of subdifferentials (of convex functions) complementarity problem

$$F(x) = \partial g(x) - \partial h(x)$$

with g and h belonging to $\Gamma_o(\mathbb{R}^n)$. This problem is equivalent (under technical hypothesis) to

$$0 \in \partial\left(g + \chi_{\mathbb{R}^n_+}\right)(x) - \partial h(x). \tag{51}$$

It has been proved (Section 3) that the DCA converges to a solution of (51).

Conclusion

We have completed a thorough study on the convex analysis approach to d.c. programming. The DCA, despite its local character, seemed to be robust and efficient with respect to the other related algorithms and led quite often to global solutions. Escaping procedures have also been introduced in nonconvex quadratic programs and the trust region subproblem. For these problems the DCA (with a finite number of restarting procedures) converges to a global solution.

Our further developments will be consecrated to globalizing the DCA by either studying new escaping procedures or combining the DCA with global algorithms in a deeper way. To our knowledge, these approaches are not much studied in the literature. We are convinced that they will constitute new trends in nonconvex programming.

References

1. Le Thi Hoai An, *Analyse numérique des algorithmes de l'optimisation d.c.. Approches locales et globales. Code et simulations numériques en grande dimension. Applications,* Thèse de Doctorat de l'Université de Rouen (1994).
2. Le Thi Hoai An and Pham Dinh Tao, *Solving a class of linearly constrained indefinite quadratic problems by D.c. algorithms,* to appear in Journal of Global Optimization.
3. Le Thi Hoai An and Pham Dinh Tao, *A Branch-and-Bound method via D.C. optimization algorithm and ellipsoidal technique for box constrained nonconvex quadratic programming problems,* to appear in Journal of Global Optimization.
4. Le Thi Hoai An and Pham Dinh Tao, *DCA with escaping procedure using a trust region algorithm for globally solving nonconvex quadratic programs,* Technical Report, LMI-CNRS URA 1378, INSA-Rouen, 1996.
5. Le Thi Hoai An and Pham Dinh Tao, *A continuous approach for large-scale linearly constrained quadratic zero-one programming*, submitted.
6. Le Thi Hoai An, Pham Dinh Tao and Le Dung Muu, *Numerical solution for optimization over the efficient set by d.c. optimization algorithm*, Operations Research Letters **19** (1996), 117-128.
7. Le Thi Hoai An, Pham Dinh Tao and Le Dung Muu, *Exact penalty in d.c. programming*, submitted to Journal of Convex Analysis.
8. Le Thi Hoai An, Pham Dinh Tao and Le Dung Muu, *A combined d.c. optimization - ellipsoidal branch-and-bound algorithm for solving nonconvex quadratic programming problems,* to appear in Journal of Combinatorial Optimization.
9. Auslender, *Optimisation, Méthodes Numériques*, Ed. Masson. Paris, 1976.
10. F. Barahona, M. Junger and R. G. Reinelt, *Experiments in quadratic 0-1 programming,* Mathemetical Programming **44** (1989), 127-137.
11. R. Beals, D. H. Krantz and A. Tversky, *Foundation of multidimensional scaling*, Psychol. Rev. **75** (1968), 12-142.
12. H. P. Benson, *Optimization over the efficient set,* J. of Mathematical Analysis and Applications **98** (1984), 562-580.
13. W. Bick, H. Bauer, P. J. Mueller and O. Gieseke, *Multidimensional scaling and clustering techniques (theory and applications in the social science)*, Institut für angewandte Sozialforchung, Universitat zu Köln, 1977.
14. I. M. Bomze and G. Danninger, *A global optimization algorithm for concave quadratic problems*, SIAM J. Optimization **3**(4) (1993), 826-842.

15. I. M. Bomze and G. Danninger, *A finite algorithm for solving general quadratic problems,* J. Global Optimization **4** (1994), 1-16.

16. H. Brezis, *Opérateurs Maximaux Monotones*, Mathematics Studies 5, North Holland, 1973.

17. A. Bronsted, *An Introduction to Convex Polytopes*, Springer Verlag New York, 1983.

18. F. H. Clarke, *Optimization and Nonsmooth Analysis*, Classics in Applied Mathematics, SIAM, 1990.

19. J. R. Clermont, M. E De la Lande and Pham Dinh Tao *Analysis of plane and axisym- metric flows of incompressible fluids with the stream tube method: Numerical simulation by Trust region algorithm,* Inter. J. for Numer. Method in Fluids **13** (1991), 371-339.

20. R. W. Cottle, J. S. Pang and R. E. Stone, *The Linear Complementarity Problem,* Academic Press, Boston, 1992.

21. G. M. Crippen, *A novel approach to calculation of conformation: distance geometry*, J. Computational physics **24** (1977), 96-107.

22. G. M. Crippen, *Rapid calculation of coordinates from distance measures*, J. Computational physics **26** (1978), 449-452.

23. G. M. Crippen and T. F. Havel, *Distance Geometry and Molecular Conformation*, Research Studies Press Ltd., Taunton, Somerset, England, 1988.

24. J. De Leeuw, *Applications of convex analysis to multidimensional scaling*, Recent developments in statistics, J.R. Barra and al., editors, North-Holland Publishing Company, pp. 133-145, 1977.

25. J. De Leeuw, *Convergence of the majorization methode for multidimansional scaling*, Journal of classification **5** (1988), 163-180.

26. R. Durier, *On locally polyhedral convex functions*, in "Trends in Mathematical Mathematics, Birkhäuser Verlag (1988), pp. 55-66.

27. I. Ekeland and R. Temman, *Analyse Convexe et Problèmes Variationnels*, Dunod, Paris, 1974.

28. G. E. Forsythe and G.H. Golub, *On the stationary values of a second-degree polynomial on the unit sphere*, J. Soc. Indust. Appl. Math. **13** (1965), 1050-1068.

29. A. Friedlander, J. M. Martinez and S. A. Santos, *Solution of linear complementarity problems using minimization with simple bounds,* J. of Global Optimization **6** (1995), 253-267.

30. L. M. Gierasch and J. King, eds., *Protein folding: deciphering the second half of the genetic code,* AAAS, 1989.

31. G. H. Golub and U. Von Matt, *Quadratically constrained least squares and quadratic problems,* Numer. Math. **59** (1991), 561-580.

32. O. Güler, *On the convergence of the proximal point algorithm for convex minimization,* SIAM J. Control and Optimization **29** (1991), 403-419.

33. L. Guttman, *A general nonmetric technique for finding the smallest coordinate space for a configuration of point*, Psychometrika **33** (1968), 469-506.

34. B. Hendrickson, *The Molecule Problem: Exploiting structure in global optimization*, SIAM

J. Optimization **5** (1995), 835-857.

35. J. B. Hiriart-Urruty and Ph. Plazanet, *Moreau's decomposition theorem revisited*, (Dedicated to J. J. Moreau on the occasion of his retirement).

36. J. B. Hiriart-Urruty, *Generalized differentiability, duality and optimization for problems dealing with differences of convex functions,* Lecture Note in Economics and Math. Systems **256** (1985), 37-70.

37. J. B. Hiriart-Urruty, *From convex optimization to non convex optimization. Part I: Necessary and sufficent conditions for global optimality*, Nonsmooth Optimization and Related Topics, Ettore Majorana International Sciences, Series 43, Plenum Press, 1988.

38. J. B. Hiriart-Urruty and C. Lemarechal, *Testing necessary and sufficient conditions for global optimality in the problem of maximizing a convex quadratic over a convex polyhedron*, Seminar of numerical analysis, Université Paul Sabatier-Toulouse, 1990.

39. J. B. Hiriart Urruty and C. Lemarechal. *Convex Analysis and Minimization Algorithms I,* Springer-Verlag Berlin Heidelberg, 1993.

40. R. Horst, N. V. Thoai & H. P. Benson, *Concave minimization via conical partitions and polyhedral outer approximation*, Mathematical Programming **50** (1991), 259-274.

41. R. Horst, T. Q. Phong, N. V. Thoai and J. De Vries. *On solving a d.c. programming problem by a sequence of linear programs,* J. of Global Optimization **1** (1991), 183-203.

42. R. Horst and P. T. Thach, *A Decomposition method for quadratic minimization problems with integer variables,* in P. M. Pardalos (ed.) Advances in Optimization and Parallel Computing, Elsevier Science Publishers, Amsterdam (1992), pp. 143-163.

43. R. Horst and N. V. Thoai, *A new algorithm for solving the general quadratic programming problem*, Computational Optimization and Applications **5** (1996), 39-48.

44. R. Horst and H. Tuy, *Global Optimization (Deterministic Approaches)*, second edition, Springer-Verlag, Berlin New York, 1993.

45. B. Kalantari and J. B. Rosen, *Algorithm for global minimization of linearly constrained concave quadratic functions*, Mathematics of Operations Research **12** (1987), 544-561.

46. H. Konno, *Maximization of a convex quadratic function under linear constraints*, Mathematical Programming **11** (1976), 117-127.

47. H. Kono, P. T. Thach and H. Tuy, *Optimization on Low Rank Nonconvex Structures,* Kluwer, Dordrecht-Boston-London, 1997.

48. J. B. Kruskal, *Multidimensional scaling by optimizing goodness-of-fit to a non metric hypothesis*, Psychometrika **29** (1964), 1-28.

49. J. B. Kruskal, *Nonmetric multidimensional scaling: a numerical method*, Psychometrika **29** (1964), 115-229.

50. J. B. Kruskal and M. Wish, *Multidimensional Scaling*, Newbury Park, CA. Sage, 1978.

51. P. J. Laurent, *Approximation et Optimisation*, Hermann, Paris, 1972.

52. W. J. M. Levet, J. P. Van De Geer and R. Plomp, *Tridiac comparaisons of mulsical intervals*, Bristish. J. Math. Statist. Psychol. **19** (1966), 163-179.

53. J. C. Lingoes and E. E. Roskam, *A mathematical and emprical analysis of two multidimensional scaling algorithms* Psychometrika **38** (1973), monograph supple-

ment.

54. P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal. **16** (1979), 964-979.

55. S. Lucidi, L. Palagi and M. Roma, *On some properties of quadratic programs with a convex quadratic constraint*, to appear.

56. P. Mahey, Nguyen Van Hien and Pham Dinh Tao, *Proximal techniques for convex programming,* Technical Report, Lab ARTEMIS, IMAG, CNRS-Grenoble, 1992.

57. P. Mahey and Pham Dinh Tao, *Partial regularization of the sum of two maximal monotone operators,* Math. Modell. Numer. Anal. (M^2AN) **27** (1993), 375-395.

58. P. Mahey and Pham Dinh Tao, *Proximal decomposition of the graph of maximal monotone operator*, SIAM J. Optimization **5** (1995), 454-468.

59. B. Martinet, *Détermination approchée d'un point fixe d'une application pseudo-contractante,* CRAS, Série A **274** (1972), 163-165.

60. J. M. Martinez, *Local minimizers of quadratic functions on Euclidean balls and spheres,* SIAM J. Optimization **4**, No 1 (1994), 159-176.

61. B. Mirkin, *Mathematical Classification and Clustering*, in Book Series "Nonconvex Optimization and its Applications", P.M. Pardalos and R. Horst eds., Kluwer Academic Publishers (to appear).

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

63. J. J. More and D.C. Sorensen, *Computing a trust region step,* SIAM J. Sci. Stat. Comput. **4** (1983), 553-572.

64. J. J. More, *Generalization of the trust region problem*, Technical report MCS P349-0193, Argonne national laboratory, 1993.

65. A. M. Ostrowski, *Solutions of Equations and Systems of Equations*, Academic Press, New York, 1966.

66. P. M. Pardalos and J. B. Rosen, *Constrained Global Optimization: Algorithms and Applications*, Lecture Notes in Computer Science, 268, Springer-Verlag, Berlin, 1987.

67. P. M. Pardalos and G. Rodgers, *Computational aspects of a branch and bound algorithm for quadratic zero-one programming*, Computing **45** (1990), 131-144.

68. P. M. Pardalos and G. Rodgers, *Parallel branch and bound algorithms for unconstrained quadratic zero-one programming,* in: Impacts of Recent Computer Advances on Operations Research (eds R. Sharda et al. (1989), North-Holland, pp. 131-143.

69. J. P. Penot, *Calcul de sous-différentiel et optimisation*, J. of Funct. Analysis **27** (1978), 248-276.

70. Pham Dinh Tao, *Eléments homoduaux d'une matrice A relatif à un couple des normes* (ϕ, ψ). *Applications au calcul de* $S_{\phi\psi}(A)$, Séminaire d'analyse numérique, Grenoble, n^o236, 1975.

71. Pham Dinh Tao, *Contribution à la théorie de normes et ses applications à l'analyse numérique*, Thèse de Doctorat d'Etat Es Science, Université Joseph Fourier- Grenoble, 1981.

72. Pham Dinh Tao, *Convergence of subgradient method for computing the bound norm of matrice,* Linear Alg. and Its Appl. **62** (1984), 163-182.

73. Pham Dinh Tao, *Algorithmes de calcul d'une forme quadratique sur la boule unité de la norme du maximum*, Numer. Math. **45** (1985), 377-404.

74. Pham Dinh Tao, *Algorithms for solving a class of non convex optimization problems. Methods of subgradients,* Fermat days 85. Mathematics for Optimization, J. B. Hiriart Urruty (ed.), Elsevier Science Publishers B. V. North-Holland, 1986.

75. Pham Dinh Tao, *Iterative behaviour, fixed point of a class of monotone operators. Application to non symmetric threshold function,* Discrete Mathematics **70** (1988), 85-105.

76. Pham Dinh Tao, *Duality in d.c. (difference of convex functions) optimization. Subgradient methods*, Trends in Mathematical Optimization, International Series of Numer Math. **84** (1988), Birkhauser, 277-293.

77. Pham Dinh Tao and S. Wang, *Training multi-layered neural network with a Trust region based algorithm,* Math. Modell. Numer. Anal. **24** (4) (1990), 523-553.

78. Pham Dinh Tao and Le Thi Hoai An, *Stabilité de la dualité lagrangienne en optimisation d.c. (différence de deux fonctions convexes)*, C.R. Acad. Paris, **318**, Série I (1994), 379-384.

79. Pham Dinh Tao and Le Thi Hoai An, *Difference of convex functions optimization algorithms (DCA) for globally minimizing nonconvex quadratic forms on Euclidean balls and spheres,* to appear in Operations Research Letters.

80. Pham Dinh Tao and Le Thi Hoai An, *Optimisation d.c. (différence de deux fonctions convexes). Dualité et Stabilité. Optimalités locale et globale. Algorithmes de l'optimisation d.c. (DCA),* Technical Report, LMI-CNRS URA 1378, INSA-Rouen, 1994.

81. Pham Dinh Tao and Le Thi Hoai An, *Polyhedral d.c. optimization. Theory, algorithms and applications*, Technical Report, LMI-CNRS URA 1378, INSA-Rouen, 1994.

82. Pham Dinh Tao and Le Thi Hoai An, *Lagrangian stability and global optimality on nonconvex quadratic minimization over Euclidean balls and spheres,* Journal of Convex Analysis **2** (1995), 263-276.

83. Pham Dinh Tao and Le Thi Hoai An, *D.c. optimization algorithms for solving the trust region subproblem,* to appear in SIAM J. Optimization.

84. Pham Dinh Tao and Le Thi Hoai An, *D.c. Programming approach for solving the multidimensional scaling problem,* Submitted to Proceedings on "Semidefinite Programming and Interior-Point Approaches for Combinatorial Optimization Problems" of the American Mathematical Society in the Fields Series.

85. Pham Dinh Tao and Le Thi Hoai An, *DCA for solving linear and nonlinear complementarity problems,* Technical Report, LMI-CNRS URA 1378, INSA-Rouen, 1996.

86. Pham Dinh Tao, Thai Quynh Phong, R. Horaud and Longquan, *Stability of Lagrangian duality for nonconvex quadratic programming. Solution methods and applications in computer vision*, Mathematical Modelling and Numerical Analysis (M^2AN) **30**, n^0 6, 1996.

87. A. T. Phillips and J. B. Rosen, *A parallel algorithm for constrained concave quadratic global minimization,* Mathematical Programming **42** (1988), 412-448.

88. A. T. Phillips and J. B. Rosen, *A parallel algorithm for partially separable nonconvex global minimization: Linear constraints,* Annals of Operations Research **25** (1990), 101-118.

89. B. Polyak, *Introduction to Optimization. Optimization Software,* Inc., Publication Division, New York, 1987.

90. S. Poljak, F. Rendl and H. Wolkowicz, *A recipe for semidefinite relaxation for (0,1)-quadratic programming,* Journal of Global Optimization **7** (1995), 51-73.

91. F. Rendl and H. Wolkowicz, *A semidefinite framework to trust region subproblems with application to large scale minimization*, CORR Report 94-32, Department of Combinatorics and Optimization, University of Waterloo, 1994.

92. J. B. Rosen and P.M. Pardalos, *Global minimization of large scale constrained quadratic problem by separable programming,* Mathematical Programming **34** (2) (1986), 163-174.

93. R. T. Rockafellar, *Convex Analysis*, Princeton University, Princeton, 1970.

94. R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control and Optimization **14**, N^o 5 (1976).

95. R. T. Rockafellar, *The theory of Subgradients and its Applications to Problems of Optimization. Convex and Nonconvex Functions*, Heldermann Verlag Berlin 1981.

96. S. A. Santos and D. C. Sorensen, *A new matrix-free algorithm for the large-scale trust-region subproblem*, Technical Report 95-20, Department of Computational and Applied Mathematics, Rice University, 1995.

97. R. N. Shepard, *Representation of structure in similarity data; problem and prospects*, Psychometrika **39** (1974), 373-421.

98. D. C. Sorensen, *Newton's method with a model trust region modification,* SIAM J. Numer. Anal. **19** (2) (1982), 409-426.

99. D. C. Sorensen, *Implicit application of polynomial filters in a K-step Arnoldi method*, SIAM Journal on Matrix Analysis and Applications **13** (1992), 357-385.

100. D. C. Sorensen, *Minimization of a large scale quadratic function subject to an ellipsoidal constraint*, to appear in SIAM Journal on Optimization.

101. J. E. Spingarn, *Partial inverse of a monotone operator*, Appl. Math. Optim. **10** (1983), 247-265.

102. Y. Takane, F. W. Young and J. De Leeuw, *Nonmetric individual differences multidimensional scaling : an alternating least squares method with optimal scaling features*, Psychometrika **42** (1977), 7-67.

103. Thai Quynh Phong, Le Thi Hoai An and Pham Dinh Tao, *On the global solution of linearly constrained indefinite quadratic minimization problems by decomposition branch-and-bound method,* RAIRO, Recherche Opérationnelle **30** (1) (1996), 31-49.

104. P. T. Thach, *D.c. sets, d.c. functions and nonlinear equations*, Mathematical Programming **58** (1993), 415-428.

105. P. T. Thach, *Global optimality criterion and duality with a zero gap in nonconvex optimization problems*, SIAM J. Math. Anal. **24** (1993), 1537-1556.

106. P. T. Thach, *A nonconvex duality with zero gap and applications*, SIAM J. Optimization **4** (1994), 44-64.

107. J. F. Toland, *Duality in nonconvex optimization,* Journal Mathematical Analysis and Applications **66** (1978), 399-415.

108. J. F. Toland, *On subdifferential calculus and duality in nonconvex optimization*, Bull. Soc. Math. France, Mémoire **60** (1979), 173-180.

109. J. F. Toland, *A duality principle for non convex optimization and calculus of variations*, Arch. rational. Mech. Analisis **71** (1979).

110. H. Tuy, *Concave programming under linear constraints,* Translated Soviet Mathematics **5** (1964), 1437-1440.

111. H. Tuy, *A general deterministic approach to global optimization via d.c. programming*, in J. B. Hiriart Urruty ed., Fermat Days 1985: Mathematics for Optimization, North-Holland, Amsterdam, (1986), pp. 137-162.

112. H. Tuy, *Global optimization of a difference of two convex functions*, Mathematical Programming Study **30** (1987), 150-182.

113. H. Tuy, *Introduction to Global Optimization*, Les Cahiers du Gerad, Groupe d'études et de recherche en analyse des décision, Montréal, Québec, 1994.

114. S. A. Vavasis, *Approximation algorithms for indefinite quadratic programming,* Mathematical Programming **57** (1992), 279-311.

115. Y. Ye, *On affine scaling algorithms for nonconvex quadratic programming,* Mathematical Programming **56** (1992), 285-300.

116. K. Wütrich, *The development of nuclear magnetic resonance spectroscopy as a technique for protein structure determination*, Accounts Chemical Res. **22** (1989), 36-44.

117. K. Wütrich, *Protein structure determination in solution by nuclear magnetic resonance spectroscopy*, Science **243** (1989), 45-50.

Mathematical Modelling and Applied Optimization Group,
Laboratory of Mathematics-
INSA Rouen - CNRS URA 1378.
BP 08, 76131 Mont Saint Aignan, France
E-mail adress: `pham@insa-rouen.fr`
`lethi@insa-rouen.fr`