Journal of Global Optimization **11:** 253–285, 1997.

Solving a Class of Linearly Constrained Indefinite Quadratic Problems by D.C. Algorithms

LE THI HOAI AN and PHAM DINH TAO
Mathematical Modelling and Applied Optimization Group, LMI, INSA Rouen, CNRS, URA 1378 BP 8, F 76 131 Mont Saint Aignan Cedex, France.[*]

(Received: 5 December 1994; accepted: 20 January 1996)

Abstract. Linearly constrained indefinite quadratic problems play an important role in global optimization. In this paper we study d.c. theory and its local approach to such problems. The new algorithm, CDA, efficiently produces local optima and sometimes produces global optima. We also propose a decomposition branch and bound method for globally solving these problems. Finally many numerical simulations are reported.

Key words: Linearly constrained quadratic problems, d.c. optimization, d.c. optimization algorithm (DCA), local optimality, global optimality, decomposition branch and bound method, global algorithm.

1. Introduction

We consider the indefinite quadratic problem over a bounded polyhedral convex set:

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

where H is a symmetric indefinite $(q \times q)$ matrix, $l \in \mathbb{R}^q$, K is a nonempty bounded polyhedral set defined as $K = \{x \in \mathbb{R}^q : Ax \leq a, x \geq 0\}$ with A being an $(m \times q)$-matrix, $a \in \mathbb{R}^m$.

When

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

and the polytope is defined as

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

we have the problem

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

[*] Address for correspondence: lethi@insa-rouen.fr pham@insa-rouen.fr

Here, $\tilde{C}$ is a symmetric positive semi-definite $(n \times n)$ matrix, D is a symmetric negative semi-definite $(s \times s)$ matrix, $c \in \mathbb{R}^n$, $d \in \mathbb{R}^s$, and $\tilde{A}$ is a $(m \times n)$-matrix, B is a $(m \times s)$-matrix, A_1 is a $(r \times n)$-matrix, A_2 is a $(p \times s)$-matrix, $a \in \mathbb{R}^m$, $a_1 \in \mathbb{R}^r$, $a_2 \in \mathbb{R}^p$. Hence, 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):

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

with $\lambda_i > 0$.

We shall show in Section 3 that 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

$$\bar{\Omega} = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^s : Ax + By \le a, x \ge 0, y \ge 0\}$$

we have the linearly constrained concave quadratic problem which has been considered by several authors (see e.g. Rosen and Pardalos [27], Kalantari and Rosen [11], Pardalos et al. [13], Phillips and Rosen [26], 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 to be even more difficult to solve computationally than concave programs. Recently a decomposition branch and bound method was proposed in Phong-An-Tao [29] for dealing with (IQP_3) in the case where $\tilde{C} \not\equiv 0$. This method is based on normal rectangular subdivisions which exploit the separability of the concave part in the objective function. In general, the existing algorithms are efficient only if the number of the concave variables is small.

Clearly Problem (IQP_2) can be considered as a minimization of a d.c. function over a polytope for which some method developed in global approaches (see e.g. Tuy [31], Horst et al. [9]) 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. In convex approaches to nonconvex nondifferentiable optimization, Pham Dinh Tao has extensively studied subgradient methods for solving convex maximization problems ([14]–[18]) and d.c. optimization problems ([20]). Important developments and improvements from both theoretical and numerical points of view have been completed since [1], [2], [21]–[24]. These d.c. optimization algorithms (DCA) are actually among the rare algorithms which allow to solve large-scale d.c. optimization problems. DCA cannot guarantee globality of computed solutions. Nevertheless they have been successfully applied for various large scale concrete d.c. optimization problems ([1], [2], [21]–[24]).

The main purpose of this paper is to discuss the use of DCA for solving Problems (IQP_1) and (IQP_2). It should be noted that the d.c. objective function (of the d.c. optimization problem (P) hereafter) has infinitely many d.c. decompositions which may have an important influence on the qualities (robustness, stability, rate of convergence and global optimality of sought solutions) of DCA. We propose a “good” d.c. decomposition for which numerical experience indicates that DCA is efficient for solving (IQP_2). In contrast to global algorithms whose the 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 present some d.c. decompositions and corresponding DCA which seem to be efficient. We propose 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 [29] for the general case. We use these algorithms for checking globality of the solution computed by DCA when $s \leq 30$. Finally we provide extensive computational experiments on large-scale problems.

The paper is divided into four sections. In the next section we present DCA and their basic properties whose proofs can be found in An [1], Tao [20], Tao-An [21]. This section contains also main properties concerning polyhedral d.c. optimization (i.e., either g or h is polyhedral convex in (P)), in particular the finite convergence of DCA in polyhedral d.c. optimization. These results show that if $\tilde{C} \equiv 0$ in (IQP_2) our algorithm is finite. Section 3 is devoted to the solution of (IQP_1) and (IQP_2) by DCA. The decomposition method for globally solving these problems is developed in Section 4 and numerical results are reported in Section 5. Finally we present in Appendix the decomposition algorithm proposed in [29] for solving (IQP_3).

2. D.c. Optimization Algorithms

Let $X = \mathbb{R}^n$ be equipped with the canonical inner product $\langle , \rangle$. The dual space Y of X then can be identified with X itself. The Euclidean norm of X is $\|x\| = \langle x, x\rangle^{1/2}$. Denote $\Gamma_o(X)$ the cone of proper lower semi-continuous convex functions on X. The conjugate function g^* of $g \in \Gamma_o(X)$ belongs to $\Gamma_0(Y)$ and is defined by

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

For $\epsilon > 0$ and $x^o \in \operatorname{dom} g$, the symbol $\partial_\epsilon g(x^o)$ denotes the 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.

D.c. program is of the form

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

(in the sequel we mean $+\infty - (+\infty) = +\infty$). Such a function f is called d.c. function on X and g, h are its d.c. components.

Using the definition of the conjugate function 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

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

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

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

If α is finite then $\operatorname{dom} g \subset \operatorname{dom} h$ and the only values of $g - h$ in $\operatorname{dom} g$ intervene in the search of global and local solution for (P). This d.c. duality was first studied by Toland [30] in a more general framework. It can be considered as a logical generalization of Pham Dinh Tao's works concerning convex maximization [14]–[17].

Below are the fundamental results concerning the duality of d.c. optimization given in [1], [20], [21].

2.1. Duality and global optimality for d.c. optimization

THEOREM 1. *([1], [20], [21]) Let $\mathcal{P}$ and $\mathcal{D}$ be the solution sets of Problem (P) and (D) respectively. Then*

(i) $\partial h(x) \subset \partial g(x) \quad \forall x \in \mathcal{P}$.

(ii) $\partial g^*(y) \subset \partial h^*(y) \quad \forall y \in \mathcal{D}$.

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

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

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

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

COROLLARY 1. *([1], [8], [21])* x^* *is a global optimal solution to* (P) *if and only if*

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

This global optimality condition is impractical for deriving solution methods to Problem (P). The algorithms DCA which will be described in 2.3 are based on the local conditions for d.c. optimization.

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

A point x^* is said to be *local minimum* of $g - h$ if there exists a neighbourhood U of x^* such that $g(x^*) - h(x^*) \leq g(x) - h(x)$ for every $x \in U$. x^* is said to be *critical point* of $g - h$ if $\partial g(x^*) \cap \partial h(x^*) \neq \emptyset$.

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

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

For $x \in \text{dom}\, g$, $g'(x,d)$ denotes the directional derivative of g at x in the direction d.

$$g'(x,d) = \lim_{t\downarrow 0} \frac{g(x+td) - g(x)}{t}.$$

Let

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

THEOREM 2. *([1], [20], [30]) (i) If* x^* *is a local minimum of* $g - h$*, then* $x^* \in \mathcal{P}_l$*.*
(ii) $x^* \in \mathcal{P}_l$ *if and only if*

$$g'(x^*,d) - h'(x^*,d) \geq 0, \quad \forall d \in X.$$

(iii) Let x^* *be a critical point of* $g - h$*. If* g *and* h *are essentially differentiable and*

$$\langle \nabla g(x), x - x^* \rangle \geq \langle \nabla h(x), x - x^* \rangle$$

for every x *in a neighborhood* U *of* x^* *then*

$$g(x) - h(x) \geq g(x^*) - h(x^*), \quad \forall x \in U.$$

(iv) Let x^* *be a local minimum of* $g - h$*. If* g^* *is essentially differentiable then every point* $y^* \in \partial h(x^*)$ *is a local minimum of* $h^* - g^*$*.*

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

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

which is equivalent to the convex maximization one: $\inf\{\langle x^*, y\rangle - g^*(y) : y \in \partial h(x^*)\}$. Similarly, for each fixed $y^* \in Y$, by duality, we define the problem

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

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 following results concerning the local optimality in duality of d.c. optimization are the core for the complete form of DCA.

THEOREM 3. *([20], [21]) (i) $x^* \in \mathcal{P}_l$ if and only if there exists $y^* \in \mathcal{S}(x^*)$ such that $x^* \in \partial g^*(y^*), i.e., x^* \in (\partial g^* \circ \mathcal{S})(x^*)$.*
(ii) $y^* \in \mathcal{D}_l$ if and only if there exists $x^* \in \mathcal{T}(y^*)$ such that $y^* \in \partial h(x^*)$, i.e., $y^* \in (\partial h \circ \mathcal{T})(y^*)$.

These characterizations constitute the basis of DCA which will be studied in Subsection 2.3. In general DCA converges to a local solution of d.c. optimization problem. However it would be interesting to formulate sufficient conditions for local optimality.

THEOREM 4. *If a point x^* admits a neighbourhood U such that*

$$\partial h(x) \cap \partial g(x^*) \neq \emptyset \text{ for all } x \in U \tag{1}$$

then $g(x) - h(x) \geq g(x^*) - h(x^*)$ for all $x \in U$ (i.e., x^* is a local minimizer of $g - h$).

Proof. We have $h(x^*) \geq h(x) + \langle x^* - x, y\rangle, \quad x \in X, \forall y \in \partial h(x)$. In particular $h(x) - h(x^*) \leq \langle x - x^*, y\rangle, \forall x \in U, \forall y \in \partial h(x) \cap \partial g(x^*)$. But $g(x) - g(x^*) \geq \langle x - x^*, y\rangle, \forall x \in U, \forall y \in \partial h(x) \cap \partial g(x^*)$. Hence $g(x) - g(x^*) \geq h(x) - h(x^*), \forall x \in U$. □

Polyhedral d.c. optimization will be extensively studied in Subsection 2.4. We now give some important results concerning local optimality for the class of locally polyhedral convex functions.

Recall that ([4]) a convex set C is *locally polyhedral* if, for every $x \in C$, there exists a polyhedral convex neighborhood of x relative to C. A convex function is said to be *locally polyhedral convex* if its epigraph is locally polyhedral convex. The indicator function of C is denoted by

$$\chi_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise} \end{cases}.$$

The local polyhedral convexity is a generalized notion of the polyhedral convexity. The former is intimately related to the diff-max property studied by Durier [4].

A function $\varphi \in \Gamma_o(X)$ is said to have *diff-max property* if for every $x \in \operatorname{dom}\varphi$ there is a neighbourhood U of x such that $\partial\varphi(u) \subset \partial\varphi(x)$, for every $u \in U$. It means that each point $x \in \operatorname{dom}\varphi$ is a local maximum for the subdifferential $\partial\varphi$ according to the inclusion relation. The next result due to Durier is worthy of attention.

THEOREM 5. *([4]) Let $\varphi \in \Gamma_o(X)$. The following are equivalent*

(i) φ has the diff-max property,

(ii) φ is locally polyhedral convex,

(iii) for every $x \in$ dom φ, there is a neighborhood V of x such that $\varphi_V = \varphi + \chi_V$ is polyhedral convex.

Moreover, for such a function,dom φ is locally polyhedral convex, φ is continuous relative to dom φ, and φ is subdifferentiable at each point of its effective domain.

REMARK 1. If $\varphi \in \Gamma_o(X)$ is finite on X then φ has the diff-max property if and only if for every $x \in X$ there is a neighbourhood U such that $\partial\varphi(u) \cap \partial\varphi(x) \neq \emptyset$ for all $u \in U$. This result has been earlier remarked in [8]. It can be proved by using the compactness of $\partial\varphi(x)$ for all $x \in X$ ([1], [21]).

COROLLARY 2. *Assuming h locally polyhedral convex, then*

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

is a necessary and sufficient condition for x^* to be a local minimum of $g - h$.

Proof. This just combines Theorems 1, 4 & 5.

REMARK 2. (i) The condition (1) constitutes a sufficient supplementary requirement for a critical point x^* of $g - h$ to be a local minimum one.

Its realization relies on the size and the width of $\partial g(x^*)$ and $\partial h(x^*)$ (int($\partial g(x^*)$) is nonempty for example) as well as on certain continuity of the (multivalued) mappings ∂g and ∂h ([1], [7], [8], [21]).

(ii) For a detailed study of local optimality in d.c. optimization, see [1], [7], [8], [21]. However, to our knowledge, Theorem 4 is one of the most general results relative to local minimum of d.c. functions.

2.3. ALGORITHMS FOR D.C. OPTIMIZATION (DCA)

The complete form of DCA is based upon Theorem 3. It allows approximating a point $(x^*, y^*) \in \mathcal{P}_l \times \mathcal{D}_l$. From a point x^o given in advance, the algorithm consist of constructing two sequences $\{x^k\}$ and $\{y^k\}$ defined by

$$y^k \in \mathcal{S}(x^k); \quad x^{k+1} \in \mathcal{T}(y^k). \tag{2}$$

From a practical point of view, although this algorithm uses a d.c. decomposition mentioned above, Problems $(S(x^k))$ and $(T(x^k))$ remain d.c. optimization programs. Calculation of y^k and x^{k+1} therefore is still a difficult task. In practice the following simplified form of DCA is used:

The Simplified Form of DCA:

The philosophy of simplified DCA is quite simple: it consists in the construction of 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^*).

Results concerning local and global optimality in d.c. optimization presented in the preceding subsections led us to the following description of simplified DCA. Namely, for $x^o \in X$ we define the two sequence $\{x^k\}$ and $\{y^k\}$ by taking

$$y^k \in \partial h(x^k); \quad x^{k+1} \in \partial g^*(y^k). \tag{3}$$

Interpretation of the simplified DCA:

The construction of the sequences $\{x^k\}$ and $\{y^k\}$ can be interpreted as follows:

According to [28], $x^{k+1} \in \partial g^*(y^k)$ if and only if $y^k \in \partial g(x^{k+1})$, i.e., $g(x^{k+1}) - \langle x^{k+1}, y^k\rangle \leq g(x) - \langle x, y^k\rangle, \forall x \in X$. In other words, x^{k+1} is a solution of the problem given as

$$\min\{g(x) - \langle x, y^k\rangle : x \in X\}. \tag{4}$$

This is equivalent to

$$\min\{g(x) - [h(x^k) + \langle x - x^k, y^k\rangle] : x \in X\} \quad (\mathrm{P}_k)$$

for each k fixed. But $y^k \in \partial h(x^k)$, i.e., $h(x) \geq h_k(x) = h(x^k) - \langle x - x^k, y^k\rangle, \forall x \in X$. So (P$_k$) is a convex optimization problem obtained from (P) by replacing h by its affine minorization function $h_k(x)$. Similarly, $y^k \in \partial h(x^k)$ means that y^k is a solution of the convex program (D$_k$)

$$\min\{h^*(y) - [g^*(y^{k-1}) + \langle x^k, y - y^{k-1}\rangle] : y \in Y\} \quad (\mathrm{D}_k)$$

which is obtained from (D) by using the affine minorization function of g^* defined by $x^k \in \partial g^*(y^{k-1})$. Here we can see a complete symmetry between Problems (P$_k$) and (D$_k$) as well as the sequences $\{x^k\}$ and $\{y^k\}$ relative to the duality of d.c. optimization.

It is clear that if h and g^* are essentially differentiable then the complete form and the simplified form of DCA are identical.

In general when Problem (P) is well defined (i.e. α is finite and the solution set of (P) is nonempty) we can construct such sequences $\{x^k\}$ and $\{y^k\}$. More precisely we have the following result:

LEMMA 1. *([1], [2]) Sequences $\{x^k\}$, $\{y^k\}$ in DCA are well defined if and only if*

$$\mathit{dom}\,\partial g \subset \mathit{dom}\,\partial h,\ \mathit{dom}\,\partial h^* \subset \mathit{dom}\,\partial g^*.$$

Convergence of Simplified DCA

For a convex function f we define $\rho(f) := \sup\{\rho \geq 0 : f - \frac{\rho}{2}\|\cdot\|^2 \text{ is convex}\}$. Let ρ_i and ρ_i^*, $(i = 1, 2)$ be real nonnegative numbers such that $0 \leq \rho_i < \rho(f_i)$ (resp. $0 \leq \rho_i^* < \rho(f_i^*)$) 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 $dy^k := y^{k+1} - y^k$. For $a, b \in X$, the line segment connecting them is denoted $[a, b]$. The following result is an improved version of the Convergence Theorem 3 and 4 in [20].

THEOREM 6. *([1], [20], [21]) Suppose that the sequences $\{x^k\}$ and $\{y^k\}$ are defined by the simplified DCA. Then we have*

$$\begin{aligned}(i)\quad (g-h)(x^{k+1}) \ &\leq\ (h^*-g^*)(y^k) - \max\left\{\frac{\rho_2}{2}\,\|dx^k\|^2, \frac{\rho_2^*}{2}\,\|dy^k\|^2\right\}\\ &\leq\ (g-h)(x^k) - \max\left\{\frac{\rho_1+\rho_2}{2}\,\|dx^k\|,\ \frac{\rho_1^*}{2}\|dy^{k-1}\|^2\right.\\ &\qquad \left. + \frac{\rho_2}{2}\|dx^k\|^2, \frac{\rho_1^*}{2}\|dy^{k-1}\|^2 + \frac{\rho_2^*}{2}\|dy^k\|^2\right\}.\end{aligned}$$

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

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

In this case

- *$(g-h)(x^{k+1}) = (h^*-g^*)(y^k)$ and x^k, x^{k+1} are critical points of $g-h$ satisfying*

$$y^k \in (\partial g(x^k) \cap \partial h(x^k)) \text{ and } y^k \in (\partial g(x^{k+1}) \cap \partial h(x^{k+1})),$$

- *y^k is a critical point of $h^* - g^*$ such that*

$$[x^k, x^{k+1}] \subset ((\partial g^*(y^k) \cap \partial h^*(y^k)),$$

- *$x^{k+1} = x^k$ 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 duality we have

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

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

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 critical points of* $h^* - g^*$ *satisfying*

$$x^{k+1} \in (\partial g^*(y^k) \cap \partial h^*(y^k)) \text{ and } x^{k+1} \in (\partial g^*(y^{k+1}) \cap \partial h^*(y^{k+1})),$$

• x^{k+1} *is a critical point of* $g - h$ *such that*

$$[y^k, y^{k+1}] \subset ((\partial g(x^{k+1}) \cap \partial h(x^{k+1})),$$

• $y^{k+1} = y^k$ *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 $\rho(g) + \rho(h) > 0$ *then* $\lim_{k\to+\infty}\{x^{k+1} - x^k\} = 0$.
If $\rho(g^*) + \rho(h^*) > 0$ *then* $\lim_{k\to+\infty}\{y^{k+1} - y^k\} = 0$.
Moreover

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

(iv) If α *is finite and the sequences* $\{x^k\}$ *and* $\{y^k\}$ *are bounded, then for every limit* x^* *of* $\{x^k\}$ *(respectively* y^* *of* $\{y^k\}$*) there exists a cluster point* y^* *of* $\{y^k\}$ *(respectively* 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. The proof can be done by the same way as in the proofs of Theorems 3 and 4 in Pham Dinh Tao [20] (see [1], [21]). □

REMARK 3. (i) In practice the simplified DCA usually yields a local minimizer which is also global ([1], [2], [21], [23]–[24]). Theorem 6 shows how strong convexity of d.c. components in primal and dual problems can influence DCA. To make the d.c. components (of the primal objective function $f = g - h$) strongly convex we usually apply the following process (the so-called proximal regularization technique)

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

In this case the d.c. components in the dual problem will be differentiable. Similarly 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 differentiable. For a detailed study of regularization techniques in d.c. optimization, see [1], [20], [21].
(ii) The main difference between the simplified and the complete DCA lies in the choice of y^k in $\partial h(x^k)$ and x^{k+1} in $\partial g^*(y^k)$. The convergence result of the complete DCA is thus improved: in Theorem 6, the nonemptiness of the subdifferentials intersection is replaced by a subdifferential inclusion ([1], [20], [21]). In other words the complete DCA yields a pair of elements $(x^*, y^*) \in \mathcal{P}_l \times \mathcal{D}_l$ (see Subsection 2.2). So the complete DCA converges to a local solution in polyhedral d.c. optimization problem (see Subsection 2.4).

2.4. POLYHEDRAL D.C. OPTIMIZATION PROBLEMS AND FINITE CONVERGENCE OF DCA WITH FIXED CHOICES OF SUBGRADIENTS

2.4.1. *Polyhedral d.c. optimization problems*

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

$$h(x) = \max\{\langle a^i, x\rangle - \alpha^i : i = 1, \ldots, m\} + \chi_C(x)$$

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 then polyhedral.

Throughout this section 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

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

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

$$\alpha = \inf_{i \in I} \inf_{x \in X} \{g(x) - (\langle a^i, x\rangle - \alpha^i)\}. \tag{5}$$

For each $i \in I$, let

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

The solution set of this problem is $\partial g^*(a^i)$. Also, let

$$J(\alpha) = \{i \in I : \beta^i = \alpha\} \text{ and } I(x) = \{i \in I : \langle a^i, x\rangle - \alpha^i = \tilde{h}(x)\}.$$

THEOREM 7. *(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 $\{a^i : i \in I\} \subset dom\, \partial g^*$ then $\mathcal{P} \neq \emptyset$.

Proof. (i) Let $x^* \in \mathcal{P}$ and $i \in I(x^*)$. Then

$$\alpha = g(x^*) - \tilde{h}(x^*) = g(x^*) - (\langle a^i, x\rangle - \alpha^i)$$

which means that $i \in J(\alpha)$ and $x^* \in \cap\{\partial g^*(a^i)$. Thus

$$I(x^*) \subset J(\alpha) \text{ and } x^* \in \cap\{\partial g^*(a^i) : \quad i \in I(x^*)\}.$$

Conversely, if $i \in J(\alpha)$ and $x^* \in \partial g^*(a^i)$ then

$$\alpha = g(x^*) - (\langle a^i, x\rangle - \alpha^i) \geq g(x^*) - \tilde{h}(x^*)$$

which implies $\alpha = g(x^*) - \tilde{h}(x^*)$ and $i \in I(x^*)$.
(ii) is immediate from (i). □

LEMMA 2. *(i) $\tilde{h}^*(a^i) \leq \alpha^i, \forall i \in I$. Equality holds if and only if there exists $x \in X$ such that $i \in I(x)$.*
(ii) $\tilde{h}(x) = \max\{\langle x, y\rangle - \tilde{h}^*(y) : y \in co\{a^i : i \in I\}\} = \max\{\langle a^i, x\rangle - \tilde{h}^*(a^i) : i \in I\}$.

Proof. (i) From the definition of $\tilde{h}$ we have

$$\alpha^i \geq \langle a^i, x\rangle - h(x), \quad \forall x \in X, \forall i \in I.$$

Hence $\tilde{h}^*(a^i) \leq \alpha^i$. If there exists $x \in X$ such that $i \in I(x)$, then

$$\alpha^i = \langle a^i, x\rangle - \tilde{h}(x) \geq \tilde{h}^*(a^i)$$

which together with (i) implies $\tilde{h}^*(a^i) = \alpha^i$.

Conversely, suppose that $\tilde{h}^*(a^i) = \alpha^i$ for some $i \in I$. Since (see [28]) $\mathrm{dom}\, \partial\tilde{h}^* = \mathrm{dom}\, \tilde{h}^*$, there exists $x \in X$ such that $\tilde{h}(x) = \langle a^i, x\rangle - \alpha^i$. Hence $i \in I(x)$.
(ii) By the fact $\mathrm{dom}\, \tilde{h}^* = co\{a^i : i \in I\}$ (see [28]) we have

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

On the other hand, from (i)

$$\tilde{h}(x) = \max\{\langle a^i, x\rangle - \tilde{h}^*(a^i) : i \in I\}. \qquad □$$

By Lemma 2 we can write $(\tilde{P})$ as

$$\alpha = \inf\{\inf_{x\in X}\{g(x) - \langle x, y\rangle + \tilde{h}^*(y)\} : y \in \{a^i : i \in I\}\} \tag{6}$$

$$\alpha = \inf\{\inf_{x\in X}\{g(x) - \langle x, y\rangle + \tilde{h}^*(y)\} : y \in co\{a^i : i \in I\}\}. \tag{7}$$

Problem (7) is exactly the dual problem $(\tilde{D})$ of $(\tilde{P})$

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

while Problem (6) becomes

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

Note that, in general, for a convex set $M \subset X$ and $g, h \in \Gamma_o(X)$,

$$\inf\{g(x) - h(x) : x \in co(M)\} < \inf\{g(x) - h(x) : x \in M\}.$$

The following result concerning the solution set $\tilde{\mathcal{D}}$ of the dual problem $(\tilde{D})$ can be proven directly without using Theorem 1.

LEMMA 3. *$J(\alpha) = \{i \in I : a^i \in \tilde{\mathcal{D}}$ and $\tilde{h}^*(a^i) = \alpha^i\}$; $\tilde{\mathcal{D}} \supset \{a^i : i \in J(\alpha)\}$.*

Proof. Let $i \in J(\alpha)$ then

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

which implies $\tilde{h}^*(a^i) = \alpha^i$ and $a^i \in \tilde{\mathcal{D}}$. Conversely, let $i \in I$ such that $a^i \in \tilde{\mathcal{D}}$ and $\tilde{h}^*(a^i) = \alpha^i$. Then

$$\alpha = \tilde{h}^*(a^i) - g^*(a^i) = \alpha^i - g^*(a^i).$$

Thus

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

Hence $i \in J(\alpha)$. The inclusion $\{a^i : i \in J(\alpha)\} \subset \tilde{\mathcal{D}}$ is evident. □

REMARK 4. • Let $I' = \{i \in I : \exists x \in X, \langle a^i, x\rangle - \alpha^i = \tilde{h}(x)\}$. Clearly, the definition of $\tilde{h}$ involves the affine functions $\langle a^i, x\rangle - \alpha^i$ with $i \in I'$, i.e.,

$$\tilde{h}(x) = \max\{\langle a^i, x\rangle - \alpha^i : i \in I'\}.$$

In this case by Lemma 2, $\tilde{h}(a^i) = \alpha^i, \forall i \in I'$.
• From Theorem 1 applying to the dual problem we have
(i) $\mathcal{P} = \cup\{\partial g^*(x^*) : x^* \in \tilde{\mathcal{D}}\}$, since dom $\partial\tilde{h} = X$.
(ii) $\tilde{\mathcal{D}} \supset \cup\{co\{a^i : i \in I(x^*)\} : x^* \in \mathcal{P}\}$.

This result is stronger than that of Lemma 3.

2.4.2. *Finite convergence of DCA*

From 2.4.1 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 7. In practice this can be done effectively if m is relatively small. In the case where 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 $co\{a^i : i \in I\} \subset \operatorname{dom} \partial g^*$. Thanks to the finiteness of α one has $\operatorname{dom} g \subset \operatorname{dom} h = C$ and $co\{a^i : i \in I\} \subset \operatorname{dom} g^*$. The simplified DCA in this case is described simply as follows:

Let x^o be chosen in advance. Set

$$y^k \in \partial\tilde{h}(x^k) = co\{a^i : i \in I(x^k)\}; x^{k+1} \in \partial g^*(y^k).$$

By setting $y^k = a^i, i \in I(x^k)$ the calculation of x^{k+1} is reduced to solve the convex program

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

Note that if $y^k = a^i$ with $i \in J(\alpha)$ then, by Theorem 7, $x^{k+1} \in \mathcal{P}$.

Now let $\tilde{H}$ and G^* be two mappings respectively defined in $\operatorname{dom} \partial\tilde{h} = X$ and in $\operatorname{dom} \partial g^*$ such that

$$\tilde{H}(x) \in \partial\tilde{h}(x), \quad \forall x \in X \text{ and } G^*(y) \in \partial g^*(y) \quad \forall y \in \operatorname{dom} \partial g^*.$$

Then the simplified DCA with fixed choice of subgradients is defined as [21]

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

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 8. *(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 $\{y^k\}$ are of the same nature: 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.

Proof. Immediate from Theorem 6 and the discrete character of the above sequences. □

2.4.3. *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

• $Tx \in ri\partial f(x)$
• $\partial f(x) = \partial f(x') \Rightarrow Tx = Tx'$.

The following results are useful, in the sequel, to the proof of the finite convergence of DCA (applying 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 ([14]-[17]) and the study of iterative behaviour of cellular automatas ([19]).

LEMMA 4. *Let $f \in \Gamma_o(X)$, then for $x^o, x^1 \in X$ one has*
(i) $f^*(\sum_{i=1}^k \lambda^i y^i) = \sum_{i=1}^k \lambda^i f^*(y^i)$, whenever $y_1, \ldots, y_k \in \partial f(x^o)$, and $\lambda^i \geq 0$ such that $\sum_{i=1}^k \lambda^i = 1$.
(ii) $ri[\partial f(x^o)] \cap \partial f(x^1) \neq \emptyset \Rightarrow \partial f(x^0) \subset \partial f(x^1)$.

Proof. (i) Let $y = \sum_{i=1}^k \lambda^i y^i$ with $y^i \in \partial f(x^0), \lambda^i \geq 0, \forall i = 1, \ldots, k$ and $\sum_{i=1}^k \lambda^i = 1$. Then $y \in \partial f(x^0)$, i.e., $f(x^0) + f^*(y) = \langle x^0, y\rangle$. On the other hand

$$f(x^0) + f^*(y^i) = \langle x^0, y^i\rangle, \forall i = 1, \ldots, k.$$

Thus

$$f(x^0) + \sum_{i=1}^k \lambda^i f^*(y^i) = \langle x^0, y\rangle.$$

Hence (i).
(ii) We suppose that $ri[\partial f(x^0)] \cap \partial f(x^1) \neq \emptyset$. Let y^0 be an element of this intersection. Since $y^0 \in ri\partial f(x^0)$, for every $y \in \partial f(x^0)$ there exists $y' \in \partial f(x^0)$ such that $y^0 = \alpha y + (1-\alpha)y', 0 < \alpha < 1$. Thus by virtue of (i)

$$f^*(y^0) = \alpha^* f^*(y) + (1-\alpha) f^*(y).$$

On the other hand, $y^0 \in \partial f(x^1)$ implies $f(x^1) + f^*(y^0) = \langle x^1, y^0\rangle$, from which follows

$$f(x^1) + \alpha f^*(y) + (1-\alpha) f^*(y') = \alpha\langle x^1, y\rangle + (1-\alpha)\langle x^1, y'\rangle.$$

This means that

$$\alpha[f(x^1) + f^*(y)] + (1-\alpha)[f(x^1) + f^*(y')] = \alpha\langle x^1, y\rangle + (1-\alpha)\langle x^1, y'\rangle.$$

Note that by the definition of f^* we always have $f(x^1) + f^*(y) \geq \langle x^1, y\rangle$ and $f(x^1) + f^*(y') \geq \langle x^1, y'\rangle$. Thus

$$f(x^1) + f^*(y) = \langle x^1, y\rangle \text{ and } f(x^1) + f^*(y') = \langle x^1, y'\rangle$$

which implies that both y and y' are elements of $\partial f(x^1)$. □

Recall $\tilde{h}(x) = \max\{\langle a^i, x\rangle - \alpha^i : i \in I\}$. Thus one can take $\tilde{H}$ by setting

$$\tilde{H}(x) = \sum_{i \in I(x)} \lambda^i a^i$$

where $\lambda^i, i \in I(x)$ satisfying
(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 5. *(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.*

Proof. Since $ri(\partial\tilde{h}(x)) = \{\sum_{i \in I(x)} \lambda^i a^i : \lambda^i > 0 \,\forall i \in I(x)\}$ ([3]), it is sufficient to show that $\partial\tilde{h}(x) = \partial\tilde{h}(x')$ implies $I(x) = I(x')$. To do this, by the symmetry, we need to show only that if $k \in I(x)$ then $k \in I(x')$. Note that $k \in I(x)$ implies $a^k \in \partial\tilde{h}(x')$. i.e.,

$$\langle a^k, x'\rangle = \tilde{h}(x') + \tilde{h}^*(a^k).$$

In view of Lemma 2, $\tilde{h}^*(a^k) = a^k$. Thus

$$\langle a^k, x'\rangle - a^k = \tilde{h}(x'). \qquad \square$$

Consider now DCA with fixed choice of subgradient applying to the polyhedral d.c. optimization presented in 2.4.1. If $\tilde{H}$ is a natural choice of $\tilde{h}$, then the following result strengthens that of Theorem 8.

THEOREM 9. *The simplified DCA with fixed choice of subgradients is finite.*

Proof. Take $p = \min\{r : \exists k \geq 0,\, x^{k+r} = x^k\}$ (p is the period of $\{x^k\}$) and $q = \min\{r : x^{p+r} = x^r\}$. Then $x^{p+q} = x^q$. In virtue of Theorem 6 we have

$$(g-h)(x^q) = (g-h)(x^{p+q}) \leq (g-h)(x^{p+q-1}) \leq \cdots \leq (g-h)(x^q)$$

which implies

$$y^{q+i} \in \partial\tilde{h}(x^{q+i+1}) \text{ for every } i = 1, \ldots, p-1.$$

By Lemma 4 one can write

$$\partial\tilde{h}(x^{q+i}) \subset \partial\tilde{h}(x^{q+i+1}) \text{ for every } i = 1, \ldots, p-1.$$

i.e.,

$$\partial\tilde{h}(x^{q+i}) = \partial\tilde{h}(x^{q+i+1}) \text{ for every } i = 1, \ldots, p-1.$$

Thus $y^k = y^q$ and $x^k = x^{q+1}, \forall k \geq q$. $\square$

We consider now the problem of maximizing a convex function φ on a polytope C, i.e., $g = \chi_C$ and $h = \varphi$ in (P):

$$\text{(PM)} \quad \min\{\chi_C(x) - \varphi(x) : x \in X\}.$$

Clearly (PM) is a polyhedral d.c. problem. Let $\{x^k\}$ and $\{y^k\}$ be generated by the simplified DCA (with fixed choice of subgradients) such that x^k is a vertex of C, then according to Theorems 8 and 9 we obtain after a finite number of iterations (x^*, y^*) such that:

(i) x^* is a vertex of C such that $\nabla\varphi(x^*) \in \partial\chi_C(x^*)$,
(ii) $\nabla\varphi^*(y^*) \in \partial\chi_C^*(y^*) = \nabla\varphi(x^*)$.
From Theorem 2 (property (iv)) y^* is a local minimum of $\varphi^* - \chi_C^*$ (i.e., $\partial\chi_C^*(y^*) = \nabla\varphi^*(y^*)$ by Corollary 2) then the vertex x^* is a local minimum of $\chi_C - \varphi$ (i.e., a local maximum of φ on C). But we have from (ii)

$$x^* = \nabla\varphi^*(y^*) \in \partial\chi_C^*(y^*),\ i.e.,\ y^* \in \partial\chi_C(x^*).$$

So we can state the following result

PROPOSITION 1. *Let x^* be a vertex of C computed by DCA as above. If $\nabla\varphi(x^*) \in$ int$(\partial\chi_C(x^*))$ then x^* is a local maximum of φ on C.*

Proof. It is immediate from the above reasoning since χ_C^* is then differentiable at y^*. □

Since the sufficient condition in Proposition 1 is almost always satisfied, one can say that in general the simplified DCA (with fixed choice of subgradients and with $\{x^k\}$ contained in the vertex set of C) converges after a finite number of iterations to a local solution of (PM). Similarly it is worth noting that complete DCA (with fixed choice of subgradients) applying to (PM) (always) converges after a finite number of iterations to a local solution of (PM) ([1], [21]).

3. Solving Problems (IQP$_1$) and (IQP$_2$) by DCA

In this section we use the simplified DCA presented in Subsection 2.3 for solving Problems (IQP$_1$) and (IQP$_2$). Denote by g and h the d.c. components of the objective function of the problem being considered. As indicated before, we try to choose g and h such that the sequences $\{x^k\}$ and $\{y^k\}$ in (3) are easy to calculate, i.e., either $\{y^k\}$ is explicitly defined and the solution of (P_k) is inexpensive or $\{x^k\}$ is explicitly defined and the solution of (D_k) is inexpensive.

3.1. Problem (IQP$_2$)

One can write (IQP$_2$) in the form

$$\min\left\{\frac{1}{2}\langle w, \mathcal{C}w\rangle + \langle t, w\rangle - \frac{1}{2}\langle w, \mathcal{D}w\rangle : \right.$$
$$\left. w \in \Omega = \{w \in \mathbb{R}^{n+s} : \mathcal{A}w \le \bar{a}, w \ge 0\}\right\} \tag{8}$$

where $\mathcal{C}$ and $\mathcal{D}$ are $(n+s) \times (n+s)$ matrices

$$\mathcal{C} = \begin{pmatrix} \tilde{C} & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathcal{D} = \begin{pmatrix} 0 & 0 \\ 0 & -D \end{pmatrix}, \quad w = \begin{pmatrix} x \\ y \end{pmatrix}, \quad t = \begin{pmatrix} c \\ d \end{pmatrix}$$

and

$$\mathcal{A} = \begin{pmatrix} \tilde{A} & B \\ A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \quad \bar{a} = \begin{pmatrix} a \\ a_1 \\ a_2 \end{pmatrix}.$$

Clearly $\mathcal{C}$ and $\mathcal{D}$ are positive semi-definite matrices. Then (8) is a d.c. optimization problem of the form (P) with the following "natural" d.c. decomposition:

$$g(w) := \frac{1}{2}\langle w, \mathcal{C}w\rangle + \langle t, w\rangle + \chi_\Omega(w), \quad h(w) := \frac{1}{2}\langle w, \mathcal{D}w\rangle \tag{9}$$

where χ_Ω, as before, stands for the indicator function of Ω.

First, we observe that h is differentiable and $\nabla h(w) = \mathcal{D}w, \forall w \in \mathbb{R}^{n+s}$. Then, to apply the simplified DCA, we have to solve, at each iteration k, a problem of the form (4) given by

$$\min\{g(w) - \langle w, \mathcal{D}w^k\rangle : w \in \mathbb{R}^{n+s}\}$$

for computing w^{k+1}.

Our algorithm can be formulated as follows:

ALGORITHM 1. Let $w^o \in \mathbb{R}^{n+s}$ be given. At each iteration $k \geq 0$ compute w^{k+1} by solving the convex quadratic program

$$(Q_1^k) \quad \min\left\{\frac{1}{2}\langle, w, \mathcal{C}w\rangle + \langle t - \mathcal{D}w^k, w\rangle : w \in \Omega\right\}.$$

The stopping criterion is $\|w^{k+1} - w^k\| \leq \varepsilon$.

REMARK 5. (i) The main subroutine in this algorithm is for solving Problem (Q_1^k) in the (x, y)-space. The dimensions of the variable x and y do not affect the complexity for DCA.
(ii) From Theorem 6 we see that if either g or h is strongly convex then the sequence $\{(g-h)(w^k)\}$ is strictly decreasing and $\lim_{k\to+\infty} \|w^{k+1} - w^k\| = 0$. Thus if both C and $(-D)$ are only positive semi-definite then we use the proximal regularization technique (see Remark 3) for finding a "good" d.c. decomposition. More precisely, in this case we take

$$\begin{aligned} g(w) &:= \frac{1}{2}\langle w, (\rho I + \mathcal{C})w\rangle + \langle t, w\rangle + \chi_\Omega(w), \\ h(w) &:= \frac{1}{2}\langle w, (\rho I + \mathcal{D})w\rangle \end{aligned} \tag{10}$$

with any positive number ρ. The simplified DCA applied to (8) with the decomposition (10) gives exactly Algorithm 1 where $\mathcal{C}$ and $\mathcal{D}$ are replaced by $\rho I + \mathcal{C}$ and $\rho I + \mathcal{D}$ respectively. In practice the choice of ρ may have an important influence on the qualities of this algorithm. Numerical experiments show that the algorithm is efficient if ρ is small enough $(\rho = 0.0001)$.

In the case where $\tilde{C} \equiv 0$ (8) is d.c. polyhedral optimization problem. If in addition $\mathcal{D}$ is positive definite (i.e., D is negative definite) we have

PROPOSITION 2. *Algorithm 1 with fixed choice of subgradients converges almost always to a local minimum of (8) after a finite number of iterations.*

Proof. Immediate from Proposition 1 and from the fact: a positive definite quadratic form and its conjugate are differentiable. □

3.2. PROBLEM (IQP_1)

We will present here some d.c. decompositions of the objective function in (IQP_1) for which the function h is always differentiable and the gradient of h is given explicitly. Then, as in the solution of (IQP_2), the use of the simplified DCA amounts to solving, at each iteration k, a problem of the form (4). Besides the spectral decomposition of H presented hereafter, the following direct d.c. decomposition seems to be suitable:

$$g(x) := \frac{1}{2}\langle (H + \rho I)x, x\rangle + \langle l, x\rangle + \chi_K(x); \quad h(x) := \frac{\rho}{2}\|x\|^2 \tag{11}$$

where ρ is a positive number such that $(H + \rho I)$ is positive semi-definite. Since $\nabla h(x) = \rho x$, we have:

ALGORITHM 2. Let $x^o \in \mathbb{R}^q$ be given and let ρ be a positive number such that $(H + \rho I)$ is positive semi-definite. At each iteration $k \geq 0$ compute x^{k+1} by solving the convex quadratic program

$$\min\left\{\frac{1}{2}\langle (H + \rho I)x, x\rangle + \langle l - \rho x^k, x\rangle : x \in K\right\}.$$

The stopping criterion is $\|x^{k+1} - x^k\| \leq \varepsilon$.

Nevertheless the “good” d.c. decomposition (9) suggests us to decompose the objective function of (IQP_1) in the form (9). For this some processes have been studied in [1]. Among them it is worth to note the following d.c. decomposition: $H = W + V$ where

$$W_{ij} = H_{ij} \quad \forall i, j \in N \text{ and } i \neq j \tag{12}$$

$$W_{ii} = \begin{cases} \sum_{i\neq j} H_{ij} + \alpha_1 & \text{if } i \in I^- \text{ and } \sum_{i\neq j} H_{ij} > 0 \\ -\sum_{i\neq j} H_{ij} + \alpha_2 & \text{if } i \in I^- \text{ and } \sum_{i\neq j} H_{ij} \leq 0 \\ H_{ii} + \alpha_3 & \text{if } i \in I^+ \text{ and } (\sum_{j\neq i} H_{ij}) - H_{ii} \leq 0 \\ \sum_{j\neq i} H_{ij} + \alpha_4 & \text{if } i \in I^+ \text{ and } (\sum_{j\neq i} H_{ij}) - H_{ii} > 0 \end{cases} \tag{13}$$

and

$$V_{ij} = 0 \quad \forall i, j \in N \text{ and } i \neq j \tag{14}$$

$$V_{ii} = \begin{cases} H_{ii} - \sum_{i \neq j} H_{ij} - \alpha_1 & \text{if } i \in I^- \text{ and } \sum_{i \neq j} H_{ij} > 0 \\ H_{ii} + \sum_{i \neq j} H_{ij} - \alpha_2 & \text{if } i \in I^- \text{ and } \sum_{i \neq j} H_{ij} \leq 0 \\ -\alpha_3 & \text{if } i \in I^+ \text{ and } (\sum_{j \neq i} H_{ij}) - H_{ii} \leq 0 \\ H_{ii} - \sum_{j \neq i} H_{ij} - \alpha_4 & \text{if } i \in I^+ \text{ and } (\sum_{j \neq i} H_{ij}) - H_{ii} > 0 \end{cases} \tag{15}$$

with $\alpha_i \geq 0, i = 1, \ldots, 4$ such that W is positive semi-definite. For instance a possible choice of the α_i is that making U diagonally dominant ([32]). We can now write (IQP$_1$) in the form (IQP$_3$):

$$\min \left\{ \frac{1}{2}\langle Wx, x\rangle + \langle l, x\rangle + \frac{1}{2}\langle Vx, x\rangle : x \in K \right\} \tag{16}$$

and then use the decomposition (9) for solving (16). More precisely, taking

$$g(x) := \frac{1}{2}\langle Wx, x\rangle + \langle l, x\rangle + \chi_K(x); \quad h(x) := \frac{1}{2}\langle -Vx, x\rangle \tag{17}$$

DCA gives rise to Algorithm 3 which consists of solving

$$\min \left\{ \frac{1}{2}\langle Wx, x\rangle + \langle l + Vx^k, x\rangle : x \in K \right\}$$

at each iteration k for computing x^{k+1}.

In practice it seems that the smaller are the α_i the more efficient are DCA for solving (IQP$_1$).

Finally let us present now the d.c. decomposition based on the spectral decomposition of H. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_q$ be the eigenvalues of H whose corresponding eigenvectors $\{u_1, \ldots, u_q\}$ constitute an orthogonal basis of $\mathbb{R}^q$. We have

$$H + P\Delta P^T \tag{18}$$

where the diagonal matrix Δ is diag$(\lambda_1, \ldots, \lambda_q)$ and P the orthogonal matrix whose columns are $\{u_1, \ldots, u_q\}$.

The first d.c. decomposition of the objective function in (IQP$_1$) is obtained by writing

$$H = P\Delta_1 P^T + P\Delta_2 P^T = H_1 + H_2$$

where Δ_1 (resp. Δ_2) is the diagonal positive semi-definite part (resp. the diagonal negative semi-definite part) of Δ, i.e.

$$(\Delta_1)_{ii} = \lambda_i \text{ if } \lambda_i \geq 0, 0 \text{ otherwise for } i = 1, \ldots, q$$

$$(\Delta_2)_{ii} = \lambda_i \text{ if } \lambda_i < 0, 0 \text{ otherwise for } i = 1, \ldots, q$$

DCA applied to the following d.c. decomposition

$$g(x) := \frac{1}{2}\langle H_1 x, x\rangle + \langle l, x\rangle + \chi_K(x); \quad h(x) := \frac{1}{2}\langle -H_2 x, x\rangle \tag{19}$$

is called Algorithm 4.

In parallel by using the change of variables $y = P^T x$, we can transform (IQP$_1$) into the form (IQP$_3$).

4. A Decomposition Method for Globally Solving (IQP$_1$) and (IQP$_2$)

We shall present in the Appendix the decomposition branch and bound method developed in Phong-An-Tao [29] (denoted ALGG) to solve Problem (IQP$_3$). There the separability of the concave part is crucial. In this section we show how to use ALGG for solving (IQP$_1$) and (IQP$_2$).

4.1. PROBLEM (IQP$_1$)

By the d.c. decompositions (11), (17) in Subsection 3.2 one can transform Problem (IQP$_1$) into the form (IQP$_3$). Observing that the decomposition (11) can be also formulated as (17) where $W = H + \rho I$ and $V_{ii} = -\rho$ for all i, in the sequel we shall consider only the decomposition (17). We have

$$(\mathrm{IQP}_1) \Leftrightarrow \min\left\{\frac{1}{2}\langle Wx, x\rangle + \langle l, x\rangle - \frac{1}{2}\sum_{i=1}^{q} v_i x_i^2 : x \in K\right\} \tag{20}$$

where $v_i = -V_{ii}, i = 1, \dots, q$. Then ALGG can be applied to solve (IQP$_1$) when q is not large. The rectangle R_0 (cf. Appendix) is now defined as

$$R_0 = \{x : 0 \le x_i \le L_i^0\}, \quad i = 1, \dots, q$$

where L_i^0 are the optimal values of q linear programs

$$\max\{x_i : x \in K\}, \quad i = 1, \dots, q.$$

The convex program (RCP) in ALGG (cf. Appendix)

$$(\mathrm{RCP}) \quad \min\left\{\frac{1}{2}\langle \tilde{C}x, x\rangle + \langle c, x\rangle + \phi_R(y) : (x, y) \in \Omega, y \in R\right\}$$

is now replaced by

$$(\mathrm{RCP}_1) \quad \min\left\{\frac{1}{2}\langle Wx, x\rangle + \langle l, x\rangle + \phi_R^1(x) : x \in K \cap R\right\}$$

where

$$\phi_R^1(x) = \sum_{i=1}^{q} \phi_{Ri}^1(x_i); \phi_{Ri}^1(x_i) = -\frac{1}{2} v_i(l_i + L_i)x_i + \frac{1}{2} v_i l_i L_i.$$

According to these modifications, we obtain the modified version ALGG:

ALGORITHM ALGG 1. Initialization: Solve q linear programs:

$$\max\{x_i : x \in K\}, \quad i = 1, \ldots, q$$

to get optimal values L_i^0, $i = 1, \ldots, q$ and set $R_0 = \{x : 0 \leq x_i \leq L_i^0\}$, $i = 1, \ldots, q$. Compute $\phi_{R_0}^1$ and solve the convex program

$$(\mathrm{R_0CP_1}) \quad \min\left\{\frac{1}{2}\langle Wx, x\rangle + \langle l, x\rangle + \phi_R^1(x) : x \in K \cap R_0\right\}$$

to obtain an optimal solution x^{R_0} and the optimal value $\beta(R_0)$. Set $\mathcal{R} = \{R_0\}, \beta_0 = \beta(R_0), \alpha_0 = f(x^{R_0})$ and $x^0 = x^{R_0}$.
Iteration $k = 0, 1, 2, \ldots,$:

k.1. Delete all $R \in \mathcal{R}_k$ with $\beta(R) \geq \alpha_k$. Let $\mathcal{P}_k$ be the set of remaining rectangles. If $\mathcal{P}_k = \emptyset$ stop: x^k is a global optimal solution.

k.2. Otherwise, select $R_k \in \mathcal{P}_k$ such that

$$\beta_k := \beta(R_k) = \min\{\beta(R) : R \in \mathcal{P}_k\}$$

and subdivide R_k into R_{k1}, R_{k2} according to the normal rectangular subdivision process "w-subdivision" (cf. Appendix).

k.3. For each R_{k1}, R_{k2} compute $\phi_{R_{ki}}^1$ and solve

$$(\mathrm{R}_{ki}\mathrm{CP}_1) \quad \min\left\{\frac{1}{2}\langle Wx, x\rangle + \langle l, x\rangle + \phi_R^1(x) : x \in K \cap R\right\}$$

to obtain $x^{R_{ki}}$ and $\beta(R_{ki})$.

k.4. Set x^{k+1} to the best of the feasible solutions known so far and update α_{k+1}.

k.5. Set $\mathcal{R}_{k+1} := (\mathcal{P}_k \backslash R_k) \cup \{R_{k1}, R_{k2}\}$ and go to the next iteration.

REMARK 6. From the numerical point of view, we see that the speed of the convergence of ALGG1 with the decomposition (11) very much depends on the value ρ. Numerical experiments show that the nearer $\rho > 0$ is to

$$\bar{\rho} = \inf\{\lambda > 0 : \lambda I + H \text{ is positive definite}\}$$

the more efficient is the algorithm. This suggests us to calculate the smallest eigenvalue $\lambda_1(H)$ of matrix H for finding ρ.

4.2. PROBLEM (IQP$_2$)

Using the decomposition (11) for the objective function of Problem (IQP$_2$) we have

$$(\mathrm{IQP_2}) \Leftrightarrow \min\left\{F(x,y) = \frac{1}{2}\langle \tilde{C}x, x\rangle + \langle c, x\rangle + \frac{1}{2}\langle (D + \rho I)y, y\rangle + \langle d, y\rangle - \frac{1}{2}\rho \sum_{i=1}^{s} y_i^2 : (x,y) \in \Omega\right\} \tag{21}$$

where ρ is a positive number such that $(D + \rho I)$ is positive semi-definite. So $F(x, y)$ is decomposed in a sum of a convex part

$$F_1(x, y) = \frac{1}{2}\langle \tilde{C}x, x\rangle + \langle c, x\rangle + \frac{1}{2}\langle (D + \rho I)y, y\rangle$$

and a concave part

$$F_2(y) = \langle d, y\rangle - \frac{1}{2}\rho \sum_{i=1}^{s} y_i^2$$

which is separable.

This interesting decomposition allow us to use ALGG for solving (IQP$_2$). By (21), the only difference between Problems (IQP$_3$) and (IQP$_2$) lies on the fact in (IQP$_2$) the convex term $F_1(x, y)$ of the objective function is defined in (x, y)-space. Then the convex program (RCP) in ALGG is replaced by

$$\text{(RCP}_2\text{)} \quad \min\{F_1(x, y) + \bar{\phi}_R(y) : (x, y) \in \Omega, y \in R\}.$$

Also, the convex envelope over a rectangle R of the concave function F_2 is now defined as

$$\bar{\phi}_R(y) = \sum_{i=1}^{s} \phi_{R_i}(y_i) = \sum_{i=1}^{s} \left\{ \left[d_i - \frac{1}{2}\rho(l_i + L_i)\right]y_i + \frac{1}{2}\rho l_i L_i \right\}.$$

Hence, we have

ALGORITHM ALGG 2. Initialization: Compute the smallest eigenvalue $\lambda_1(D)$ of matrix D. Set $\rho = -\lambda_1(D) + 0.01$. Solve s linear programs:

$$\max\{y_i : (x, y) \in \Omega\}, \quad i = 1, \ldots, s$$

to get optimal values L_i^0, $i = 1, \ldots, s$ and $R_0 = \{y : 0 \le y_i \le L_i^0\}$. Compute $\bar{\phi}_{R_0}$ and solve the convex program

$$\text{(R}_0\text{CP}_2\text{)} \quad \min\{F_1(x, y) + \bar{\phi}_{R_0}(y) : (x, y) \in \Omega, y \in R_0\}$$

to obtain an optimal solution (x^{R_0}, w^{R_0}) and the optimal value $\beta(R_0)$. Set $\mathcal{R} = \{R_0\}, \beta_0 = \beta(R_0)$, $\alpha_0 = f(x^{R_0}, w^{R_0})$ and $(x^0, y^0) = (x^{R_0}, w^{R_0})$.

Iteration $k = 0, 1, 2, \ldots$:

k.1. Delete all $R \in \mathcal{R}_k$ with $\beta(R) \ge \alpha_k$. Let $\mathcal{P}_k$ be the set of remaining rectangles. If $\mathcal{P}_k = \emptyset$ stop: (x^k, y^k) is a global optimal solution.

k.2. Otherwise, select $R_k \in \mathcal{P}_k$ such that

$$\beta_k := \beta(R_k) = \min\{\beta(R) : R \in \mathcal{P}_k\}$$

and subdivide R_k into R_{k1}, R_{k2} according to the normal rectangular subdivision process "w-subdivision" (cf. Appendix).

k.3. For each R_{k1}, R_{k2} compute $\phi^1_{R_{ki}}$ and solve

$$(\mathrm{R}_{ki}\mathrm{CP}_2) \quad \min\{F_1(x,y) + \bar{\phi}_{R_{ki}}(y) : (x,y) \in \Omega, y \in R_{ki}\}$$

to obtain $(x^{R_{ki}}, w^{R_{ki}})$ and $\beta(R_{ki})$.

k.4. Set (x^{k+1}, y^{k+1}) to the best of the feasible solutions known so far and update α_{k+1}.

k.5. Set $\mathcal{R}_{k+1} := (\mathcal{P}_k \backslash R_k) \cup \{R_{k1}, R_{k2}\}$ and go to the next iteration.

Clearly both (RCP) and (RCP_2) are considered in the same (x,y)-space. On the other hand the calculation of $\lambda_1(D)$ when s is moderate size is not expensive. Thus Problem (IQP_2) seems to be not more difficult to solve computationally than Problem (IQP_3).

5. Numerical Results

In this section we present some computational tests on the performance of our algorithms for different sets of test problems. Our experiments are composed of two parts. In the first we study the performance of DCA and the global algorithms for problems (IQP_1), (IQP_2) and (IQP_3). In the second we provide a comparison between DCA (with two different decompositions) and an active set method (in the local approach) for the general problem (IQP_1).

The stopping criterion of DCA was actually $er \leq 10^{-7}$ where

$$er = \begin{cases} \|x^{k+1} - x^k\|^2/\|x^k\|^2 & \text{if } \|x^k\| > 1 \\ \|x^{k+1} - x^k\|^2 & \text{otherwise} \end{cases}. \tag{22}$$

5.1. The performance of DCA and the global algorithms

In the first experiment the algorithms have been coded in PASCAL under a Unix system and run on SUN SPARC-2 station with double precision. We solved 48 randomly selected problems and the problem taken from Floudas and Pardalos [6] (Problem 1, Table 1). We used the Lemke algorithm for minimizing the convex quadratic problems over a polytope. The elements of matrices A, B and vectors a, c, d are generated with their signs, so that the feasible region was nonempty and bounded. (For simplicity we take $A_1 \equiv A_2 \equiv 0$, i.e., the feasible region is $\bar{\Omega}$). A positive definite matrix $\tilde{C}$ is constructed following Moré and Sorensen ([12]). More precisely we set $\tilde{C} = Q\tilde{D}Q^T$ for some orthogonal matrix Q and a diagonal matrix $\tilde{D}$. The orthogonal matrix Q of the form $Q_1(P_2)Q_3$ where

$$Q_j = I - 2\,\frac{w_j w_j^T}{\|w_j\|^2}, \quad j = 1,2,3$$

and the components w_j are random numbers in (-1,1). The matrix $(-D)$ is constructed by the same procedure. For an indefinite matrix $H = Q\tilde{D}Q^T$ (in Problems 40–49), the diagonal elements of matrix $\tilde{D}$ are random numbers in (–10, 10).

Table 1. The performance of Algorithm 1 and ALGG for solving (IQP_3)

Pb	n	s	m	Algorithm 1			ALGG		
				iter	time	value	iter	time	value
1	10	10	10	3	0.10	−49318.01796	5	1.20	−49318.01796
2	10	10	10	2	0.08	−474.9335	6	1.42	−474.9335
3	50	10	10	3	1.27	−8477.3949	5	5.25	−8477.3949
4	50	10	10	4	2.43	−18410.6544	4	7.77	−18410.6544
S	50	10	10	3	1.82	−1870719.9497	2	3.13	−1870719.9497
6	50	10	10	14	6.57	−1411.2272	71	107.00	−1411.2272
7	50	10	20	12	16.50	−709.4954	64	236.20	−1053.2883
8	50	20	10	14	15.95	−19745.0002	28	53.60	−19745.0002
9	100	10	10	8	13.00	−17995.4845	9	29.95	−17995.4843
10	100	10	20	5	21.03	−338032.6704	6	47.82	−338032.6704
11	100	10	20	3	10.90	−37068.4762	10	104.47	−37068.4762
12	100	10	20	5	18.85	−12903.3843	5	37.07	−23172.4911
13	100	20	10	5	8.98	−48122.9213	50	235.90	−48122.9213
14	100	20	15	13	68.95	−13033.1788	65	494.48	−21909.2309
15	150	15	20	10	100.8-	−165808.8504	22	445.62	−165808.8485
16	150	15	20	8	113.25	−100712.8212	36	853.00	−100712.8210
17	150	20	20	3	29.48	−461601.5248	25	568.00	−461601.5248
18	150	20	20	3	43.00	−559466.4654	6	207.32	−559446.4654
19	150	20	20	4	35.13	−822692.6431	6	121.93	−822692.6431
20	150	30	20	4	21.02	−625589.0117	7	162.85	−1031057.3468
21	150	30	20	4	39.92	−155964.5034	6	136.58	−155964.4849
22	150	30	20	3	19.47	−137832.6798	5	116.90	−146291.6683
23	200	20	20	12	276	−21845.4611	66	2840.65	−21845.4611
24	200	30	20	30	994.77	−137806.3506	118	7497.70	−137806.2335

In the globally algorithms, the deletion rule $\beta(R) \geq \alpha_k$ was replaced by $\beta(R) \geq (\alpha_k - \epsilon|\alpha_k|)$ so that these algorithms terminate whenever an ϵ-optimal solution $\bar{x}$ has been obtained. Table 1 provides the computational results of Algorithm 1 and the global algorithm ALGG ([29]) for 24 tested problems in the form (IQP_3) when $s \leq 30$.

Table 2 indicates the performance of Algorithm 1 and ALGG2 when $s \leq 30$ for 15 problems in the form (IQP_2).

Table 3 contains the computational results of Algorithm 2, Algorithm 3 and ALGG1 when $n \leq 30$ for 10 problems in the form (IQP_1).

The initial point of Algorithm 1 is chosen as

$$w_i^o = 0, i = 1, \ldots, n, w_{i+n}^o = 0.4L_i^o, i = 1, \ldots, s. \tag{23}$$

In Algorithms 2 and 3 we started at the same point $x_i^o = 0.4L_i^o, i = 1, \ldots, n$.

Table 2. The performance of Algorithm 1 and ALGG2 for solving (IQP_2)

Pb	n	s	m	Algorithm 1			ALGG2		
				iter	time	value	iter	time	value
25	50	10	10	2	0.85	−1965.9444	4	8.42	−3525.0121
26	100	10	20	3	28.80	−90.3198	2	30.22	−90.3198
27	100	10	20	9	41.92	−4183.5794	9	102.68	−4183.4422
28	100	10	20	7	41.02	−689.7835	11	129.80	−964.1756
29	150	20	20	4	18.80	−4200835.4290	12	248.47	−4200835.4290
30	150	20	20	4	47.02	−3049374.3695	8	201.28	−3049374.3695
31	150	20	20	2	9.07	−15764602.2436	6	136.42	−15764602.2437
32	150	30	20	3	20.87	−7190983.3576	103	2968.42	−7274068.8458
33	150	30	20	5	51.05	−2473171.0127	23	644.90	−2473177.1578
34	150	30	20	10	102.72	−36290.2666	19	515.55	−36290.2666
35	30	100	20	5	22.48	−125633.58021			
36	200	100	20	5	114.10	−4071717.3418			
37	100	150	20	30	314.35	−2175466.0546			
38	150	50	20	15	200.53	−131666.07454			
39	100	100	20	24	260.82	−230037.8888			

For ALGG, ALGG1 and ALGG2 we took $\epsilon = 10^{-3}$. In ALGG we used w-subdivision which was shown to be the best among three types of normal rectangular subdivision given in [29].

The abbreviations in these tables are the following ones: Pb – Problem; iter – number of iteration, time – CPU time in seconds; value – value optimal computed by algorithm.

5.2. Comparison between DCA and the active set method

In the second experiment we solved 20 problems which is the form (IQP_1) by Algorithms 2, 4 ad the active set method. The algorithms have been coded in MATLAB and run on SUN SPARC-10 station with double precision. The data was generated as in Subsection 5.1. For minimizing the convex quadratic problems over a polytope in Algorithms 2 and 4 we also used the active set method. We employed the function EIG in MATLAB for computing the eigenvalues and eigenvectors of matrix H in Algorithm 4.

Comments

• From the results in the tables 1 and 2 we see that Algorithm 1 with the choice of starting point (23) is very efficient: in most problems (19 over 24 for (IQP_3) and 7 over 10 for (IQP_2)) its computed solution is a global solution.

Table 3. The performance of Algorithms 2, 3 and ALGG1 for solving (IQP_1)

Pb	q	m	Algorithm 2			Algorithm 3			ALGG		
			iter	time	value	iter	time	value	iter	time	value
40	10	10	47	0.38	−50.90	3	0.13	−54.37	184	32.28	−54.69
41	20	15	22	3.75	−822.60	5	0.95	−822.60	930	715	−822.60
42	30	10	6	1.32	−6022.13	6	1.67	−6022.12	500	739.42	−7696.05
43	30	15	6	1.83	−495.12	4	1.23	−544.23	250	415.20	−544.23
44	30	20	80	35.95	−232.60	6	2.82	−549.73	252	480.25	−658.40
45	30	30	90	44.62	−646.22	99	52.30	−646.99	501	1295.07	−646.97
46	100	20	49	220.72	−222162.56	26	107.35	−222162.56			
47	150	20	9	96.57	−405676.43	8	93.93	−395513.83			
48	200	10	4	55.53	−23545293769.60	3	30.35	−23545293803.98			
49	200	20	138	2727.98	−1216527.09	36	579.97	−1216527.09			

Table 4. The performance of Algorithms 3, 4 and active set method for solving (IQP_1)

Pb	q	m	Algorithm 2		Algorithm 4		Active set method
			iter	value	iter	value	value
50	10	5	10	−155.28	4	−44.35	−0.11
51	10	10	17	−218.04	8	−218.04	−0.02
52	15	10	11	−126.62	6	−126.63	0.89
53	20	20	21	−122.64	7	−122.66	−0.02
54	30	10	14	−549.52	4	−549.58	6.09
55	30	15	24	−363.63	9	−363.67	−6.98
56	30	20	18	−198.45	7	−167.11	4.96
57	40	10	14	−8794.4	7	−8794.4	−0.08
58	40	20	19	−862.97	7	−862.97	1.81
59	50	20	8	−3825.2	5	−3825.2	6.16
60	60	20	26	−2610.5	9	−2617.1	10.33
61	70	20	14	−8865.8	5	−8870.7	−6.05
62	100	20	15	−338930	6	−192800	0.06
63	100	20	18	−826670	10	−394260	0.12
64	100	50	17	−6534.2	5	−6475.6	8.97
65	150	20	20	−220760	4	−185220	−1.66
66	150	30	28	−170300	5	−170330	15.66
67	150	50	13	−22546	11	−24760	4.69
68	200	20	13	−109500	5	−109510	1.71
69	200	30	19	−152620	8	−152650	−8.29

• Table 3 shows that in general Algorithm 3 is more efficient than Algorithm 2. Note that the choice of α_i, $i = 1, \ldots, 4$ (resp. ρ) for algorithms 3 (res. algorithm 2) is very important.

• Table 4 indicates that the solutions provided by the active set method are very bad. Moreover we observe that Algorithm 4 is faster than Algorithm 2 while the approximate optimal value given by the latter is smaller than that provided by the former. • DCA terminates very rapidly; the average number of iterations is 7, 26, 19 and 7 for Algorithms 1, 2, 3 and 4 respectively.

• DCA can work with problems where the number of both convex and concave variable may be large.

Acknowledgements

We would like to thank the Referee for his careful reading and constructive criticism. His useful comments clarified the presentation and corrected some inaccuracies.

Appendix

A Decomposition Method for Solving Problem (IQP$_3$) ([29])

One considers Problem (IQP$_3$)

$$(\text{IQP}_3) \quad \min\left\{f(x,y) = f_1(x) + f_2(y) = \frac{1}{2}\langle\tilde{C}x, x\rangle + \langle c, x\rangle + \sum_{i=1}^{s}\left[d_i y_i - \frac{1}{2}\lambda_i y_i^2\right] : \quad (x,y) \in \Omega\right\}$$

with $\lambda_i > 0$.

The method presented here should be efficient for large-scale (IQP$_3$) problems, when the number of variables that enter the concave part of the objective function is small in comparison with the total number of variables. The separability of the concave part motivates the use of rectangular subdivision. First a rectangular domain $R_0 \subset \mathbb{R}^s$ is constructed that contains the projection of Ω in the y-space. This rectangle is then divided into smaller and smaller subrectangles. For each rectangle R a convex underestimating function $f_1(x)+\phi(y)$ of the original objective function $f(x,y)$ is constructed and the convex minimization problem

$$\min\{f_1(x) + \phi(y) : (x,y) \in \Omega, y \in R\}.$$

is solved. The solution of this convex program gives both a lower and upper bound for the optimal value of the problem

$$\min\{f_1(x) + f_2(y) : (x,y) \in \Omega, y \in R\}.$$

The branch-and-bound procedure is then applied to discard regions which cannot contain any global minimizer and eventually to locate an optimal solution.

To construct the smallest rectangular domain $R_0 \subset \mathbb{R}^s$ which contain the projection of Ω on the y-space, one solves s linear programming problems

$$\max\{y_i \quad \text{s.t.} \quad (x,y) \in \Omega\}, \quad i = 1, \ldots, s$$

to get optimal values L_i^0, $i = 1, \ldots, s$. The rectangular domain can then be expressed as

$$R_0 = \{y : 0 \le y_i \le L_i^0\}.$$

a) **Lower bounding**

Let $R = \{y : l_i \le y_i \le L_i\}$ be a rectangle in $\mathbb{R}^s$. As usual, one has the convention that the infinitum of an empty set is $+\infty$.

A standard method for lower bounding in branch and bound algorithms is to use convex underestimators of the objective function. Since concave function $f_2(y) = \sum_{i=1}^{s} q_i(y_i)$ is separable, its convex envelope over a rectangle R is simply the sum of affine function $\phi_{Ri}(y_i)$ that agrees with q_i at the endpoints of the segment $[l_i, L_i]$, i.e. the function (cf. [11], [27], [25], etc.)

$$\phi_R(y) = \sum_{i=1}^{s} \phi_{Ri}(y_i) \tag{24}$$

where $\phi_{Ri}(y_i)$ is given explicitly by

$$\phi_{Ri}(y_i) = \left[d_i - \frac{1}{2}\lambda_i(l_i + L_i)\right] y_i + \frac{1}{2}\lambda_i l_i L_i. \tag{25}$$

So $f_1(x) + \phi_R(y)$ is a convex underestimating function of $f(x, y)$ over the domain $\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^s : (x, y) \in \Omega, y \in R\}$. The solution to the convex program

$$\text{(RCP)} \quad \min\{f_1(x) + \phi_R(y) : (x, y) \in \Omega, y \in R\}$$

provides a point (x^R, w^R) such that

$$f_1(x^R) + \phi_R(w^R) \leq \min\{f(x, y) : (x, y) \in \Omega, y \in R\} \leq f(x^R, w^R) \tag{26}$$

i.e. $\beta(R) = f_1(x^R) + \phi_R(w^R)$ is a lower bound for f over R and $f(x^R, w^R)$ is an upper bound for the global optimal value f_*.

b) **Normal rectangular subdivision (NRS)**

The concept of a normal rectangular subdivision as introduced by Tuy (see e.g. Horst-Tuy [10] (Definition VII.7)).

Let $R = \{y : l_i \leq y_i \leq L_i\}$ be a rectangle and let $\phi_R(y)$ be the above defined convex underestimator of $f_2(y)$ over R. Denote by (x^R, w^R) and $\beta(R)$ an optimal solution and the optimal value, respectively, of the convex program (RCP).

Consider now a rectangular subdivision process in which a rectangle is subdivided into subrectangles by means of a finite number of hyperplanes parallel to certain facets of the orthant $\mathbb{R}^s_+$. Such a process generates a family of rectangles which can be represented by a tree with root R_0 and such that a node is a successor of another one if and only if it represents an element of the partition of the rectangle corresponding to the latter node. An infinite path in this tree corresponds to an infinite nested sequence of rectangles $R_h, h = 0, 1, \ldots$ For each h let $(x^h, w^h) = (x^{R_h}, w^{R_h}), \phi_h(y) = \phi_{R_h}(y)$.

DEFINITION 1. A nested sequence R_h is said to be normal if

$$\underline{\lim}_{h\to\infty} |f_2(w^h) - \phi_h(w^h)| = 0. \tag{27}$$

A rectangular subdivision process is said to be normal if any infinite nested sequence of rectangles that it generates is normal.

Suppose now that an NRS process has been defined. One can construct the following branch and bound algorithm for solving (IQP_3).

c) **Algorithm ALGG**
Initialization: Compute the enclosing rectangle R_0 by solving s linear programs. Compute ϕ_{R_0} and solve the convex program

$$(\mathrm{R}_0\mathrm{CP}) \quad \min\{f_1(x) + \phi_{R_0}(y) : (x, y) \in \Omega, y \in R_0\}$$

to obtain an optimal solution (x^{R_0}, w^{R_0}) and the optimal value $\beta(R_0)$. Set $\mathcal{P}_0 = \{R_0\}, \beta_0 = \beta(R_0), \alpha_0 = f(x^{R_0}, w^{R_0})$ and $(x^0, y^0) = (x^{R_0}, w^{R_0})$.
Iteration $k = 0, 1, 2, \ldots$:

k.1. Delete all $R \in \mathcal{R}_k$ with $\beta(R) \geq \alpha_k$. Let $\mathcal{P}_k$ be the set of remaining rectangles. If $\mathcal{P}_k = \emptyset$ stop: (x^k, y^k) is a global optimal solution.

k.2. Otherwise, select $R_k \in \mathcal{P}_k$ such that

$$\beta_k := \beta(R_k) = \min\{\beta(R) : R \in \mathcal{P}_k\}.$$

and subdivide R_k into R_{k1}, R_{k2} according to the chosen normal rectangular subdivision process.

k.3. For each R_{k1}, R_{k2} compute $\phi_{R_{ki}}$ and solve

$$(\mathrm{R}_{ki}\mathrm{CP}) \quad \min\{f_1(x) + \phi_{R_{ki}}(y) : (x, y) \in \Omega, y \in R_{ki}\}$$

to obtain $(x^{R_{ki}}, w^{R_{ki}})$ and $\beta(R_{ki})$.

k.4. Set (x^{k+1}, y^{k+1}) to the best of the feasible solutions known so far and update α_{k+1}.

k.5. Set $\mathcal{P}_{k+1} := (\mathcal{P}_k \backslash R_k) \cup \{R_{k1}, R_{k2}\}$ and go to the next iteration.

Normal rectangular subdivision process Some methods for constructing normal rectangular subdivision (NRS) process are discussed in [29]. We present here the w-subdivision process which was shown to be the best among three types of normal rectangular subdivision given in [29].

***w*-subdivision**: (Falk and Soland [5])
For the selected $R_k, \beta(R_k) < f(x^k, y^k)$, hence,

$$f_2(w^k) - \phi_k(w^k) > 0.$$

Choose an index i_k satisfying

$$i_k \in \arg\max_i \{f_{2i}(w_i^h) - \phi_{ki}(w_i^k)\}$$

and subdivide R_k into two subrectangles

$$R_{k,1} = \{y \in R_k : y_{i_k} \leq w_{i_k}^k\}, \quad R_{k,2} = \{y \in R_k : y_{i_k} \geq w_{i_k}^k\}.$$

THEOREM 10. *(i) If the Algorithm terminates at iteration k then (x^k, y^k) every accumulation point of which is a global optimal solution of (IQP$_2$), and*

$$\alpha_k \searrow f_*, \quad \beta_k \nearrow f_*.$$

References

1. Le Thi Hoai An (1994), 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.
2. Le Thi Hoai An, Pham Dinh Tao & Le Dung Muu (1996), D.c. optimization approach for optimizing over the efficient set. *Operations Research Letters* 19, 117–128.
3. A. Bronsted (1983), *An Introduction to Convex Polytopes*, Springer-Verlag, New York.
4. R. Durier (1988), On locally polyhedral convex functions. In *Trends in Mathematical Mathematics*, Birkhäuser, Verlag, 55–66.
5. J.E. Falk & R.M. Soland (1969), An algorithm for separable nonconvex programming problems, *Management Science* 15, 550–569.
6. C.A. Floudas & P.M. Pardalos (1987), A collection of test problems for constrained global optimization algorithms. In G. Goos and J. Hartmanis (eds.), *Lecture Notes in Computer Sciences*, 455, Springer-Verlag.
7. J.B. Hiriart Urruty (1985), Generalized differentiability, duality and optimization for problems dealing with differences of convex functions. In *Lecture Note in Economics and Math. Systems* 256, 37–70.
8. J.B. Hiriart Urruty (1988), From convex optimization to nonconvex optimization. Part I: Necessary and sufficient conditions for global optimality. *Nonsmooth Optimization and Related Topics, Ettore Majorana International Sciences*, Series 43, Plenum Press.
9. R. Horst, T.Q. Phong, N.V. Thoai, and J. De Vries (1991), On solving a d.c. programming problem by a sequence of linear programs. *J. of Global Optimization* 1, 183–203.
10. R. Horst and H. Tuy (1993), *Global optimization (Deterministic approaches)*, Springer-Verlag, Berlin.
11. B. Kalantari & J.B. Rosen (1987), Algorithm for global minimization of linearly constrained concave quadratic functions, *Mathematics of Operations Research* 12, 544–561.
12. J.J. More and D.C. Sorensen (1983), Computing a trust region step. *SIAM J. Sci. Stat. Comput.* 4, 553–572.
13. P.M. Pardalos, J.H. Glick & J.B. Rosen (1987), Global optimization if indefinite quadratic problems. *Computing* 39, 281–291.
14. Pham Dinh Tao (1975), 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, no. 236.
15. Pham Dinh Tao (1981), Contribution à la théorie de normes et ses applications à l'analyse numérique. Thèse de Doctorat d'Etat Es Science, Université Joseph Fourier-Grenoble.
16. Pham Dinh Tao (1984), Convergence of subgradient method for computing the bound norm of matrice. *Linear Alg. and its Appl.* 62, 163–182.
17. Pham Dinh Tao (1985), Algorithmes de calcul d'une forme quadratique sur la boule unité de la norme maximum, *Numer. Math.* 45, 377–440.
18. Pham Dinh Tao (1986), 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.
19. Pham Dinh Tao (1988), Iterative behaviour, Fixed point of a class of monotone operators. Application to non symmetric threshold function. *Discrete Mathematics* 70, 85–105.
20. Pham Dinh Tao (1988), Duality in d.c. (difference of convex functions) optimization. Subgradient methods. *Trends in Mathematical Optimization*, K.H. Hoffmann et al. (ed.), International Series of Numer Math. 84, Birkhauser.

21. Pham Dinh Tao et Le Thi Hoai An (1994), Optimisation d.c. (différence de deux fonctions convexes). Théorie, algorithmes et applications. *Rapport de Recherche, LMI, CNRS URA 1378, INSA-Rouen*.
22. Pham Dinh Tao and Le Thi Hoai An (1994), Polyhedral d.c. optimization. Theory, Algorithms and Applications. *Technical Report, LMI-CNRS URA 1378, INSA-Rouen*.
23. Pham Dinh Tao et Le Thi Hoai An (1996), D.C. (difference of convex functions) optimization algorithms (DCA) for globally minimizing nonconvex quadratic forms on Euclidean balls and spheres. *Operations research letters* 19, 207–216.
24. Pham Dinh Tao et Le Thi Hoai An, D.c. optimization algorithm for solving the trust region problem. *To appear in SIAM Journal of Optimization*.
25. A.T. Phillips & J.B. Rosen (1988), A parallel algorithm for constrained concave quadratic global minimization, *Mathematical Programming* 42, 412–448.
26. A.T. Phillips & J.B. Rosen (1990), A parallel algorithm for partially separable non-convex global optimization: Linear constraints, *Annals of Operations Research* 25, 101–118.
27. J.B. Rosen & P.M. Pardalos (1986), Global minimization of large scale constrained quadratic problem by separable programming, *Mathematical Programming* 34(2), 163–174.
28. R.T. Rockafellar (1970), *Convex Analysis*, Princeton University, Princeton.
29. Thai Quynh Phong, Le Thi Hoai An & Phan Dinh Tao (1996), On the global solution of linearly constrained indefinite quadratic minimization problems by decomposition branch and bound method. *RAIRO, Recherche Opérationnelle* **30**(1), 31–49.
30. J.F. Toland (1979), A duality principle for non convex optimization and calculus of variations, *Arch. rational. Mech. Analisis* 71.
31. H. Tuy (1987), Global Optimization of a Difference of two convex functions, *Mathematical Programming Study* 30, 150–182.
32. R. Varga (1962) *Matrix Iterative Analysis*, Prentice Hall.