A Graduate Seminar Report

On

Optimization with D.C. Data. (Difference of Convex Functions or Sets)

by

Alemayehu Geremew Negash

Advisor

Dr. Semu Mitiku

School of Graduate Studies

Addis Ababa University

June 2004
Addis Ababa

Table of Contents

	Page
Preface	**(i)**
1. Chapter 1	
Some Definitions and basic Concepts	1
2. Chapter 2	
Introduction	7
2.1. Motivation	7
2.2. Classification of Global Optimization	8
2.3. Some Typical examples	11
2.4. Properties d.c.functions	14
2.5. D.c.sets	23
3. Chapter 3	
Duality in d.c.optimization	25
3.1. Fenchel Duality	25
3.2. Lagrange Duality	27
3.3. DC duality. (Toland's result)	28
4. Chapter 4	
D.C. algorithm (DCA) for solving d.c.optimization	34
4.1. Global Optimality Criteria for d.c.optimality	36
4.2.Local Optimality Criteria for d.c.optimality	38
4.2.1.Necessary and sufficient condition for local d.c.optimization	38
4.2.2.D.C.duality transportation of a local minimizer	40
4.2.3.Main results of local duality in DCO	40
4.3.Subgradient – duality method for solving d.c.optimization	42
4.4.The simplified Primal-dual algorithm in solving a DC problem	47
4.5.Remarks on DCA	49
Reference	52

Preface

This seminar paper reviews the theory, methods and algorithm for solving d.c optimization, as has been developed in recent years.

In chapter one of this paper some fundamental concepts and definitions are discussed.

Chapter two focuses on the motivation, basic classification of global optimization, some typical examples of d.c.programming problems that are encountered in various fields and properties of d.c.functions and d.c.sets.

Chapter three is devoted to the concept of duality in the d.c optimization and chapter four discusses global and local optimality criteria and also solution methods for d.c optimization.

Finally I would like to express my heart felt thanks to my adviser Dr. Semu Mitiku for his valuable comments and suggestions during the preparation of this graduate report.

I am also grateful to my sister Frehiwot Geremew who is always willing to take care of me.

I sincerely express my gratitude to all those who have helped me in this endeavor.

Alemayehu Geremew

Addis Ababa University

June 2004

Chapter 1

Some Definitions and Basic Concepts

Scalar product and Norm

We shall use **R** to denote the real number system, and $\overline{R}$ to denote the extended real numbers.

In this material we consider the space $\mathbf{R}^n$, where $n \in \mathbf{N}$ and the topology we are using here is the usual topology in $\mathbf{R}^n$. In this spaces a scalar product is defined by

$$< x, y >= \sum_{i=1}^{n} a_i b_i, \quad x = (a_1, \ldots, a_n) \ , y = (b_1, \ldots, b_n)$$

The product has among other things the following important properties:

i) $\langle x, y \rangle = \langle y, x \rangle$

ii) $\langle x + y, z \rangle = \langle x, z \rangle + \langle x + z \rangle$

iii) $\langle \alpha x, y \rangle = \alpha \langle y, x \rangle = \langle x, \alpha y \rangle , \alpha \in \mathbf{R}.$

A norm is usually introduced by the scalar product in the following way:

$$\|x\| = \sqrt{< x, x >}$$

A norm in $\mathbf{R}^n$ is by definition a mapping $\|.\|$ from $\mathbf{R}^n$ to $\mathbf{R}$ that satisfies:

i) $\|x\| \geq 0 \ \forall x \in \mathbf{R}^n, \ \alpha \in \mathbf{R}$

ii) $\|\alpha x\| = |\alpha| \|x\| \ \forall x \in \mathbf{R}^n, \ \alpha \in \mathbf{R}$

iii) $\|x + y\| \leq \|x\| + \|Y\| \ \forall x, y \in \mathbf{R}^n$

Open and Closed Set

For $r > 0$ we define the sets

$$S(x_o, r) = \{x \in \mathbf{R}^n : \|x - x_o\| < r\}, \text{ (open ball)}$$

$$S[x_o, r] = \{x \in \mathbf{R}^n : \|x - x_o\| \leq r\}, \text{ (closed ball)}$$

Let $M \subseteq \mathbf{R}^n$. Then M is said to be open, if for each point $x_o \in M$ there exist an $\varepsilon > 0$ such that $S(x_o, \varepsilon) \subseteq M$. A point $x_o \in \mathbf{R}^n$ is called an accumulation point of M, if for each

$\varepsilon > 0$ there is a $y \in S(x_0, \varepsilon) \cap (M/\{x_0\})$. The set of all accumulation points of M is called the closure of M and will be denoted by $\overline{M}$. By the definition we have the closure of M is the set of all points $z \in \mathbf{R}^n$ for which a sequence (x^n) converges to z.

Given two points x, $y \in \mathbf{R}^n$, the set of all points $z = \lambda x + (1-\lambda)y$ such that $0 \le \lambda \le 1$ is called the (closed) line segment between x and y and denoted by [x,y]. A set $M \subseteq \mathbf{R}^n$ is called convex if it contains any line segment between two points of it; in other words, if $\lambda x + (1-\lambda)y \in M$ whenever x, $y \in M$, $0 \le \lambda \le 1$.

Given a function f: $X \to [-\infty, \infty] = \overline{\mathbf{R}}$ on a set $S \subseteq \mathbf{R}^n$, the sets

$$\text{dom } f = \{x \in S: f(x) < +\infty\}$$

$$\text{epi } f = \{(x, \alpha) \in S\text{x}\mathrm{R} : f(x) \le \alpha\}$$

are called the effective domain and the epigraph of f, respectively. If dom $f \neq \phi$ and $f(x) > -\infty$ for all $x \in S$ then we say that the function f is proper.

To define the closure, interior and relative interior points of any set C in $\mathbf{R}^n$ we use the following notions:

(i) The Euclidean unit ball $B \subseteq \mathbf{R}^n$ is $B = \{x \in \mathbf{R}^n : \|x\| \le 1\}$

$$= \{x \in \mathbf{R}^n : d(x,0) \le 1\}$$

where $d(x,y) = \|x - y\| = \sqrt{\langle x - y, x - y \rangle}$.

(ii) for any $a \in \mathrm{R}^n$, the ball with radius $\varepsilon > 0$ and center a is given by

$$\{x: d(x,a) \le \varepsilon\} = \{a + y: \|y\| \le \varepsilon\} = a + \varepsilon B.$$

(iii) for any set C in $\mathbf{R}^n$, the set of points x whose distance from C does not exceed ε is

$$\{x: \exists y \in C, d(x,y) \le \varepsilon\} = \bigcup \{y + \varepsilon B : y \in C\} = C + \varepsilon B.$$

Here are the definitions:

1) The closure cl C of C is $\text{cl } C = \bigcap \{C + \varepsilon B : \varepsilon > 0\}$.

2) The interior int C of C is $\text{int } C = \{x : \exists \varepsilon > 0, x + \varepsilon B \subset C\}$.

3) The relative interior ri C of C in $\mathbf{R}^n$ is defined as the interior which results when C is regarded as a subset of its affine hull aff C.

In other words,

$$\text{ri } C = \{x \in \text{aff } C : \exists \varepsilon > 0, (x + \varepsilon B) \cap (\text{aff } C) \subset C\}.$$

Convex Function and Affine Function

A function $f: S \to \overline{R}$ is called convex if its epigraph is a convex set in $R^n x R$. This is equivalent to saying that S is a convex set in $\mathbf{R}^n$ and for any x, y$\in$S and $0 \le \lambda \le 1$ we have

$$f(\lambda x + (1-\lambda) y) \le \lambda f(x) + (1-\lambda) f(y) \qquad (1)$$

whenever the right hand side is defined. In other words (1) must always hold unless $f(x) = -f(y) = \pm\infty$.

A function f is said to be concave on S if –f is convex. An affine function on $\mathbf{R}^n$ has the form: $f(x) = \langle a, x \rangle + \alpha$, with $a \in R^n, \alpha \in R$. A function is called strictly convex on S if it is convex and if the above inequality (1) is a strict inequality for all $x \ne y$ and $\lambda \in (0,1)$.

Let X be a normed space and X' is be its dual space. A linear mapping F: $X \to X'$ is said to be

a) positive definite if and only if $\langle Fx, x\rangle > 0 \; \forall x \in X \setminus \{0\}$.

b) positive semi definite if and only if $\langle Fx, x\rangle \ge 0 \; \forall x \in X$.

Obviously, from the positive definiteness of the linear mapping F the positive semi definiteness follows. With these two notions we have the following criterion for convexity, the proof of which is given in R.Deumlich [1]

Theorem1.1: Let X be a normed space, $U \subset X$ be an open and convex set and f: $U \to \mathbf{R}$ be twice continuously Frechet-differentiable. Then f is convex on U if and only if $f''(x_0)$ is Positive semi definite for all $x_0 \in U$, i.e.,

f is convex on U if and only if $f''(x_0)(x)(x) \ge 0$ for all $x \in \mathbf{X}$ and for all $x_0 \in U$.

Special case:

Let $X = \mathbf{R}^n$.

$$f''(x_0) = H(x_0) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) \right)_{i,j = 1,\dots,n}$$

Then f is convex if and only if $f''(x_0)(x)(x) = x^T H(x_0) x \ge 0$ for all $x \in \mathbf{R}^n$.

Thus the question of convexity of f is reduced to the question of positive definiteness of the Hesse-matrix $H(x_0)$.

Proposition 1.2: The Upper envelope (pointwise supremum) of an arbitrary family of convex functions is convex.

Proof: If $f(x) = \sup\{f_i(x): i\in I\}$, then epi $f = \bigcap_{i\in I} \text{epi} f_i$, and the intersection of a family of convex sets is a convex set.

Proposition 1.3: Let $K = (a,b) \subseteq \mathbf{R}$, $f: K \to \mathbf{R}$ be continuously differentiable on K. If f' is non decreasing (i.e. it is increasing on K), then f is convex on K.

Proof: See in [6].

Theorem 1.4: (Weierstrass approximation Theorem)

Let f be a continuous function on a compact convex set $K \subset \mathbf{R}^n$. Then given $\varepsilon > 0$ there exists a polynomial g defined on $\mathbf{R}^n$ such that $\|f - g\|_\infty \le \varepsilon$ i.e.,

$$\max_{x\in K} |f(x) - g(x)| \le \varepsilon.$$

Proof: See [1]

Subdifferential

Given an arbitrary function $f: R^n \to \overline{R}$ we consider the set of all affine functions h minorizing f. It is natural to restrict ourselves to proper functions, because an improper function either has no affine minorant (if $f(x) = -\infty$ for x) or is minorized by every affine function (f(x) is identical to $+\infty$)

Given a proper function f on $\mathbf{R}^n$, a vector $p \in \mathbf{R}^n$ is called a subgradient of f at a point x_0 if

$$\langle p, x - x_\circ \rangle + f(x_\circ) \le f(x) \forall x.$$

The set of all subgradients of f at x_0 is called a subdifferential of f at x_0 and is denoted by $\partial f(x_\circ)$. In such a case, we say $p \in \partial f(x) \subseteq R^n$. The multivalued mapping $\partial f: x \to \partial f(x)$ is called the subdifferential of f.

A function f from a set $S \subset R^n$ to R is said to be lower semicontnuous (l.s.c.) at a point $x \in S$ if

$$\liminf_{y\to x} f(y) \ge f(x),$$

that is the level sets $N_\alpha(f) = \{x \in S : f(x) \le \alpha\}$ are closed for all $\alpha \in \mathbf{R}$.

and it said to be upper semicontnuous (u.s.c.) at $x \in S$ if

$$\limsup_{y \to x} f(y) \le f(x).$$

A function that is lower and upper semi-continuous at x is continuous at x in the ordinary sense.

A convex function f on X is said to be essentially differentiable if it satisfies the following three conditions [6]:

i) $C = \mathrm{int}(\mathrm{dom}\, f) \ne \phi$

ii) F is differentiable on C, and

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.

Lagrange Method

The Lagrange Method for solving optimization problems with inequality constraints is described as follows:

Consider the optimization problem

$$(\mathbf{P}) \qquad f(x) \to \min, \; x \in S,$$

$$S = \{ x \in U : g(x) \le 0, h(x) = 0\},$$

where g: $U \to \mathbf{R}^m$, h: $U \to \mathbf{R}^p$ and

$$g(x) = (g_1(x), \ldots, g_m(x))^T \in \mathbf{R}^m \text{ and } h(x) = (h_1(x), \ldots, h_p(x))^T \in \mathbf{R}^p.$$

The Lagrange-function for **(P)** is defined by

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

$$= f(x) + \langle \lambda, g(x) \rangle + \langle \mu, h(x) \rangle, \qquad (x, \lambda, \mu) \in U x R_+^m x R^p$$

The numbers λ_i, $i \in \{1,\ldots, m\}$ and μ_j, $j \in \{1,\ldots, p\}$ are called Lagrange-multipliers.

Now we consider the following optimization problem:

$$(\mathbf{P}_{\lambda,\mu}) \qquad L(x, \lambda, \mu) \to \min, \; (x, \lambda, \mu) \in U x R_+^m x R^p.$$

As a connection between **(P)** and $(\mathbf{P}_{\lambda,\mu})$ we have the following theorem.

Theorem 1.5: (Lagrange Lemma)

Let **(P)** be given, $\lambda \in R^m_+, \mu \in R^p$, $x_0 \in S$ with $\langle \lambda, g(x_0) \rangle = 0$. If x_0 is a solution of $(\mathbf{P}_{\lambda,\mu})$, then x_0 is a solution of (P).

Proof: Because x_0 is an element of S and also is a minimum point of $(\mathbf{P}_{\lambda,\mu})$ and because $h(x_0) = 0$ and $\langle \lambda, g(x_0) \rangle \leq 0$ for all $x \in S$,

the inequality

$$\begin{aligned} f(x_0) &= f(x_0) + \langle \lambda, g(x_0) \rangle + \langle \mu, h(x_0) \rangle \\ &\leq f(x) + \langle \lambda, g(x) \rangle + \langle \mu, h(x) \rangle \\ &\leq f(x) \quad \text{for all } x \in S. \end{aligned}$$

which follows that

$$f(x_0) \leq f(x) \text{ for all } x \in S.$$

Hence x_0 is a solution of (P).

Chapter 2
Introduction

2.1. Motivation

Suppose we want to maximize the area of a rectangle whose perimeter P is given by $P = 1$. If the length of the rectangle is x_1 and its width is x_2, the problem is

maximize the area $A = x_1x_2$ subject to the perimeter $2x_1 + 2x_2 = 1$ where $x_1 > 0$, $x_2 > 0$.

Thus, we have the optimization problem

$$(P) \qquad f(x) = x_1x_2 \to \min, \quad x \in S,$$

$$S = \{x \in R_+^2 : 2x_1 + 2x_2 = 1, x_1 > 0, x_2 > 0\}$$

Or

$$(\mathbf{P^*}) \qquad -f(x) = x_1x_2 \to \max, \quad x \in S,$$

$$S = \{x \in R_+^2 : 2x_1 + 2x_2 = 1, x_1 > 0, x_2 > 0\}$$

Both f and –f are neither convex nor concave. So, **(P)** is a non-convex optimization problem. However we can rewrite the objective function as

$$f(x) = \frac{1}{2}(x_1^2 + x_2^2) - \frac{1}{2}(x_1 + x_2)^2$$
$$= g(x) - h(x)$$

where $f(x) = \frac{1}{2}(x_1^2 + x_2^2)$ and $g(x) = \frac{1}{2}(x_1 + x_2)^2$ are convex functions. Here the objective function f is expressed as a difference of the convex functions h and g.

Definition 2.1: Let K be a convex set in $\mathbf{R}^n$. We say that a function f is d.c. (difference of two convex functions) on K if it can be expressed as the difference of two convex functions on K, i.e., $f(x) = f_1(x) - f_2(x)$ for all $x \in K$, where f_1, f_2 are convex functions on K.

Examples (1) Convex and concave functions are d.c.functions.

(2) Consider the quadratic function $f(x) = \langle x, Qx \rangle$, where Q is a symmetric matrix. Set $x = Uy$ where $U=[U^1,...,U^n]$ is the matrix of normalized eigenvectors of Q, we have $U^T QU = \text{dia}(\lambda_1,\ldots,\lambda_n)$. Hence $f(x) = f(Uy) = \langle Uy, QUy \rangle = \langle y, U^T QUy \rangle$, so that

$$f(x) = f_1(x) - f_2(x) \text{ with } f_1(x) = \sum_{\lambda_i \geq 0} \lambda_i y_i^2, \quad f_2(x) = -\sum_{\lambda_i < 0} \lambda_i y_i^2, \quad y = U^{-1}x.$$

This shows that f is a d.c. function since f_1, f_2 is convex on $\mathbf{R}^n$.

If f(x) is a d.c. functions then the inequality (constraint) of the form $f(x) \leq 0$ or $f(x) \geq 0$ is called a d.c. inequality (constraint)

Definition 2.2: A d.c.optimization problem is a global optimization problem of the form

$$\text{(P)} \qquad f(x) \to x \in D, \ g_i(x) \leq 0 \ (i=1, \ldots, m)$$

where D is a convex set in $\mathbf{R}^n$, and each of the functions $f(x)$, $g_1(x),\ldots,g_m(x)$ is a d.c. function on $\mathbf{R}^n$.

Denoting the feasible set by S, we have $S = \{x \in D: g_i(x) \leq 0, i = 1, \ldots, m\}$

We can also write the problem as:

$$\min \{f(x): x \in S\}.$$

In the definition the minimum is understood in the global sense; that is, we wish to find a point $x^* \in S$ such that $f(x^*) \leq f(x)$ for all $x \in S$. A point $x^* \in S$ satisfying this condition is a global optimal solution (global minimizer), as opposed to a local optimal condition (local minimizer). A point $x^* \in S$ is said to be a local minimizer if there exists a neighborhood U of x^* satisfying $f(x^*) \leq f(x)$ for all $x \in S \cap U$.

In the next section we will give the classification of global optimization problem.

2.2. Classification of Global Optimization

Optimization problems can be categorized as linear and non-linear. Non-linear optimization problems can also be classified in to convex optimization problems and non-convex optimization problems. Non-convex optimization problems are usually named as global optimization problems. The classifications are

(1) **Linear Optimization problems** have the form:

(P) $f(x) = \langle c, x \rangle \rightarrow \max$

such that

$$Ax \leq b,\ x \geq 0$$

where c, $x \in \mathbf{R}^n$, $b \in \mathbf{R}^m$, A an mxn – matrix and $b \geq 0$.

Denoting $M = \{x \in R^n_+ : Ax \leq b,\ x \geq 0\}$, the set M becomes the intersection of a finite set of half spaces in $\mathbf{R}^n$, i.e., M is a polyhedral set or a polyhedron, if it is bounded and clearly M is a convex set.

So, we can rewrite the problem as:

(P) $f(x) = \langle c, x \rangle \rightarrow \max,\ x \in M.$

This problem is equivalent to:

(-P) $-f(x) = \langle c, x \rangle \rightarrow \min,\ x \in M.$

We know that $-f(x) = \langle -c, x \rangle$ is a linear function and hence a convex function on $\mathbf{R}^n$.

Thus, linear optimization problems are also convex minimization problem as stated below in (2).

(2) **Convex Optimization Problems** are of the form:

(P) $f(x) \rightarrow \min,\ x \in D,$

where f is a convex function defined on $\mathbf{R}^n$ and D is a closed convex set given by explicit system of convex inequalities: $g_i(x) \leq 0$ $(i = 1,\ldots,m)$.

In this class of optimization problems, any local minimizer of the problem is also a global minimizer.

(3) Non-convex optimization problems

The main feature of non-convex optimization problem is multi-extremality (that is the existence of many local minimizer with different objective values) and this multi-extremality often results from the non-convexity of the problem. Therefore, we are led to the following classification of non-convex problems.

(i) Concave minimization (convex maximization)

(P) $f(x) \rightarrow \min, \ x \in D,$

where f(x) is a concave function and D is a closed convex set given by an explicit system of convex inequalities: $g_i(x) \leq 0$ $(i = 1,\ldots,m)$.

This is the simplest class of global optimization problems since the search for the extremal points of the problem reduces to the search on the boundary points of the feasible set D.

(ii) Reverse convex programming

(P) $f(x) \rightarrow \min, \ x \in D,$

where f (x) is a convex function, D is a closed convex set, C is an open convex set, and D and C are given by explicit convex inequalities.

If $D = \{x: g(x) \leq 0\}$ and $C = \{x: h(x) < 0\}$ then the problem is

minimize f(x) subject to $g(x) \leq 0, \ h(x) \geq 0$

This differs from a conventional convex program only by the presence of the constraint $h(x) \geq 0$ (i=1,...,m), which is, reverse convex. The addition of reverse convex constraint destroys the convexity and some times the connectedness of the feasible sets. In this case the problem become a complicated non-convex global optimization problem.

(iii) D.C. Programming

(P) $f(x) \rightarrow \min, \ x \in D, \ h_i(x) \geq 0 \ (i = 1, \ldots, m)$

where each of the functions $f(x)$, $h_1(x)$, ..., $h_m(x)$ is a d.c. function given explicitly as a difference of two convex functions on the convex set D.

Clearly, reverse convex programs are special cases of d.c.programs. Conversely, at the expense of introducing additional variables any d.c.programming problem can be converted to a reverse convex program.

(iv) General Continuous optimization

This is the most general problem (P) when f(x), $g_i(x)$, i= 1,…,m are assumed only to be continuous. Any such problem can in theory be reduced to minimizing a linear function under a d.c. constraint. In practice, however, this reduction involves many operations, so continuous optimization problems are, generally speaking, the most difficult in the above hierarchy.

2.3.Some Typical Examples.

These examples are adopted from " Convex Analysis and Global Optimization" by Hoang Tuy

We will see in the subsequent sections that an overwhelming majority o f optimization problems of potential interest can be converted in to the form of d.c. programming problems. In this section we discuss some typical situations in which the d.c. structure occurs quite naturally, though some times in a hidden way.

1. (Production-transportation planning)

Consider k factories producing a certain good to satisfy the demands d_j (j =1,…,m) of m destination points. The production cost is g $(y_1,\dots,y_k) = \sum_{i=1}^{m} g_i(y_i)$, where $g_i(y_i)$ is the cost of producing y_i units of good at factory i and is a concave function because of economy of scale .The transportation cost is linear and equal to c_{ij} for every unit shipped from factory i to destination point j. In addition, there is a shortage penalty $h(z_1,\dots,z_m) = \sum_{j=1}^{m} h_j(z_j)$, with $h_j(z_j) \leq 0$ if $z_j \geq d_j$ and $h_j(.)$ is a decreasing non-negative function in the interval $[0,1)$. Usually, the penalty function h(.) is convex, so to minimize the total production-transportation cost, one must solve the following d.c. programming problem:

$$\text{minimize} \sum_{i=1}^{k} \sum_{j=1}^{m} c_{ij} x_{ij} + g(y) + h(z)$$

subject to

$$\sum_{i=1}^{k} x_{ij} = y_i \quad (i = 1,\dots,k)$$

$$\sum_{i=1}^{k} x_{ij} = z_j \ (j = 1,\dots,m)$$

$$x_{ij} \geq 0 \ \forall i, j$$

$$y_i, z_j \geq \forall i, j$$

It is d.c. optimization problem because

i) $\sum_{i=1}^{k}\sum_{j=1}^{m} c_{ij}x_{ij}$ is linear and hence convex function.

ii) h(z) is convex

iii) g(y) is concave, that is negative of a convex function and

iv) all the constraints are linear and hence convex constraints.

2. (Engineering design)

Global optimization problems of the following type frequently arise in engineering design. Consider a fabrication process in which the quality of a manufactured item is characterized by an n-dimensional parameter and an item is accepted only if when the parameter is contained in some region of acceptabity $S \subset \mathbf{R}^n$. If x is the nominal value of the parameter, and y is its actual value, then because of random fluctuation in the fabrication process, y will usually deviate from x and for every fixed x the production (expected proportion of accepted items) is an increasing function of the radius r(x) of the maximal ball around x that is contained in S (it is not hard to see that r(x) equals the distance from x to the boundary of S). Therefore, to maximize the production yield the designer should choose the nominal value of x so as to maximize r(x) over S. Since most often the acceptabity region is non-convex and r(x) is non-concave, this program max $\{r(x): x \in S\}$ (often referred to as the "design centering problem") is a difficult non-convex global optimization problem.

Consider design centering optimization problem

$$\text{maximize } r(x) \text{ subject to } x \in \mathbf{R}^n$$

where $r(x) = \inf\{\|x-y\| : y \notin M\}$, denoting $M = \mathbf{R}^n \backslash S$.

Modifying the problem as:

$$\text{maximize } r^2(x) \text{ subject to } x \in \mathbf{R}^n.$$

where $r^2(x) = \inf\{\|x-y\|^2 : y \notin M\}$, we have

$$\begin{aligned} r^2(x) &= \inf\{\|x-y\|^2 : y \notin M\} \\ &= \inf\{\|x\|^2 - 2x^T y + \|y\|^2 : y \notin M\} \\ &= \|x\|^2 - \sup\{2x^T y - \|y\|^2 : y \notin M\} \end{aligned}$$

So $r^2(x) = \|x\|^2$ - $h(x)$ with $h(x) = \sup\ \{2x^T y - \|y\|^2 : y \notin S\}$.

Since for each $y \notin M$ the function $x: \to 2x^T y - \|y\|^2$ is affine, h(x) is a convex function (proposition 1.2). Therefore, $r^2(x)$ is a d.c. function and the design-centering problem is to minimize the d.c. function $r^2(x)$ over the compact set S (which itself is a d.c. set in most cases).

3.(Location Planning)

In its classical version, Weber's problem of facility location consists in the following: A number of p facilities providing the same service have to be constructed to serve n users located at points a_i, i = 1,…,n on the plane. Each user will be served by one of the facilities; the transportation cost is the linear function of the distance. The problem is to determine the locations $x_k \in \mathbf{R}^2$, k = 1,…,p of the facilities so as to minimize the weighted sum of all transportation cost from each user to the facility that serves this user. To minimize the total cost, each user must be served by the closest facility, and that the distance from a user i to the closest facility is:

$$d_i(x_1,\ldots,x_p) = \min_{k=1,\ldots,p} \|a_i - x_k\|$$

the problem can be formulated more simply as the non-convex optimization problem

$$\text{minimize } \sum_{i=1}^{n} \omega_i d_i(x_1,\ldots,x_p) \text{ subject to } x_k \in \mathbf{R}^2, k=1,2,\ldots, p$$

where $\omega_i > 0$ is the weight assigned to user i. Since each $d_i(.)$ is a d.c. function (pointwise minimum of a set of convex functions $\|a_i - x_k\|$; see proposition 2.1(iv))), the problem is minimizing a d.c. function over $(\mathbf{R}^2)^p$.

Thus, a number of optimization problems of practical interest in economics and engineering, such as, production planning problems, engineering design problems, location planning problems, … involve d.c. functions and/or d.c. sets in their description.

2.4. Properties of D.C Functions

Before we pass directly to the discussion of methods for solving d.c. optimization problems, it is necessary to study the general d.c. structure. In this and the next section we discuss properties of d.c. functions and d.c. sets.

Let DC(K) denote the class of d.c functions on a convex set $K \subseteq \mathbf{R}^n$. Since a convex function $f: K \to \mathbf{R}$ is continuous on int K, provided int $K \neq \phi$, f is locally Lipschitz. Thus, the class DC(K) consists of locally Lipschitz functions.

Now let us discuss the important properties of d.c. functions.

Proposition 2.1: The class DC(K) is an algebra, which is closed under finite min-max combination.

(i) If $f(x), g(x) \in DC(K)$, then $f(x) \pm g(x) \in DC(K)$. Indeed, if $f(x) = f_1(x) - f_2(x)$ and $g(x) = g_1(x) - g_2(x)$, where $f_1(x)$, $f_2(x)$, $g_1(x)$, $g_2(x)$ are convex functions then $f(x) + g(x) = [f_1(x)+g_1(x)] - [f_2(x) + g_2(x)]$ where $f_1(x) + g_1(x)$ and $f_2(x) + g_2(x)$ are convex functions and $f(x) - g(x) = [f_1(x) + g_2(x)] - [f_2(x) + g_1(x)]$ where $f_1(x) + g_2(x)$ and $f_2(x) + g_1(x)$ are convex functions.

(ii) For each real number t, $tf \in DC(K)$ if $f \in DC(K)$. Indeed, if $f(x) = f_1(x) - f_2(x)$, where $f_1(x)$ and $f_2(x)$ are convex functions then for each $t \geq 0$, $tf(x) = tf_1(x) - tf_2(x)$ where $tf_1(x)$ and $tf_2(x)$ are convex functions and for each $t \leq 0$, $tf_1(x) = [-tf_2(x)] - [-tf_1(x)]$ where $-tf_1(x)$ and $-tf_2(x)$ are convex functions.

In general, from (i) and (ii), for $f_i \in DC(K)$, i=1,…, m we have

$$\sum_{i=1}^{m} \alpha_i f_i(x) \in DC(K), \text{ for any real numbers } \alpha_i.$$

Thus, the class DC (K) forms a linear space.

(iii) If f, g $\in$DC(K), then fg$\in$DC(K)

The proof will be based on Corollary 2.10.

Therefore, by (i), (ii), and (iii), for any convex set K, DC(K) is an algebra.

(iv) $g(x) = \max\{f_1(x),..., f_m(x)\}$ and

$h(x) = \min\{f_1(x),..., f_m(x)\}$ are d.c. functions where $f_i \in DC(K)$ for $i = 1,\ldots,m$.

Proof of the first part: Let $f_i(x) = g_i(x) - h_i(x)$ with g_i, h_i convex on K.

From the equality

$$f_i(x) = g_i(x) + \sum_{j\neq i} h_j(x) - \sum_{j=1}^{m} h_j(x) \text{ for any } x \in K,$$

we have

$$\max\{f_1(x),\ldots, f_m(x)\} = \max_i \{ g_i(x) + \sum_{j\neq i} h_j(x) \} - \sum_{j=1}^{m} h_j(x)$$

It follows that g(x) which is the difference of two convex functions since the upper envelope (the pointwise maximum) and the sum of finitely many convex functions are convex.

Proof of the second part: Let $f_i(x) = g_i(x) - h_i(x)$ with g_i, h_i convex on K.

From the equality

$$f_i(x) = \sum_{j=1}^{m} g_j(x) - \left(\sum_{j\neq i} g_j(x) + h_i(x)\right) \text{ for all } x \in K$$

we have

$$\min\{f_1(x),\ldots, f_m(x)\} = \sum_{j=1}^{m} g_j(x) - \max_i\{ \sum_{j\neq i} g_j(x) + h_i(x)\}$$

which is the difference of two convex functions as stated in the proof of the first part.

As a consequence of the considerations from (i) to (iv) we obtain for an arbitrary convex set K, DC(K) is an algebra, which is stable under the operations of pointwise maximum (upper envelope) and pointwise minimum (lower envelope) of finitely many d.c. functions.

Proposition 2.2: Every function $f \in C^2(\mathbf{R}^n)$, (a twice continuously differentiable function on $\mathbf{R}^n$), is d.c. on any compact convex set K.

Proof: We show that, given any compact convex set K $\subset \mathbf{R}^n$, the function g (x) = f (x) + $\frac{1}{2}\rho\|x\|^2$ becomes convex on K when ρ is sufficiently large (then f(x) = g(x) - $\rho\ \|x\|^2$ yields a d.c. representation of f). Indeed, since

$<U, \nabla^2 g(x)U> = <U, \nabla^2 f(x)U> + \rho\|U\|^2$, if ρ is so large that

$$-\min\{<U, \nabla^2 f(x)U>: x \in K, \|U\| = 1\} \le \rho$$

then $<U, \nabla^2 g(x)> \ge 0$ for all U, hence g (x) is convex (Theorem 1.1).

Since a polynomial has a continuous derivative of any order, it follows from Proposition 2.2. that:

Corollary 2.3: Any polynomial in $x \in \mathbf{R}^n$ is a d.c. function on $\mathbf{R}^n$.

A non-trivial problem, however, is how to represent effectively a polynomial as a difference of two convex polynomials.

Corollary 2.4: For any continuous function f(x) on a compact convex set K and for any $\varepsilon > 0$ there exists a d.c. function g(x), such that, $\max_{x \in K} |f(x) - g(x)| \le \varepsilon$

Proof: By Weierstrass theorem there exists a polynomial g(x) satisfying the required condition, and obviously g(x) $\in C^2$.

Thus DC(K) is dense in C(K), the Banach space of continuous functions on K, equipped with the sup norm.

Though d.c. functions occur frequently in practice, they often appear in a hidden, not directly recognizable, form. To identify d.c.functions in various situations we prove some further properties of these functions. First let us recall some facts from convex analysis.

Lemma 2.5: Let g: K→**R** be a function defined on a convex set K. If for every $x \in K$ there is a neighborhood U of x such that g is convex on U $\cap$ K, then g is convex on K.

Proof: It suffices to show that for any x, y $\in$ K the function $\varphi(t) = g(tx + (1-t)y)$, $0 < t < 1$, is convex. But from the hypothesis this function $\varphi(t)$ is convex in the neighborhood of every $t \in (0,1)$. That is there exists a real number $\delta > 0$ with $U_t = (t-\delta, t+\delta)$ such that $\varphi(t)$ is convex on U_t. Setting $a = t - \delta$ and $b = t + \delta$ we have $U_t = (a, b)$.

If $0 < \alpha < \beta$ and $t + \beta < b$, then the point $(t + \alpha, \varphi(t + \alpha))$ is below the segment joining $(t, \varphi(t))$ and $(t + \beta, \varphi(t + \beta))$, so

$$\frac{\varphi(t+\alpha)-\varphi(t)}{\alpha} \leq \frac{\varphi(t+\beta)-\varphi(t)}{\beta}$$

This shows that the function $t \mapsto \frac{\varphi(t+\beta)-\varphi(t)}{\beta}$ is non-decreasing as $\beta \downarrow 0$.

Hence it has a limit $\varphi'_+(t)$ (finite or $=-\infty$)

Further more, setting $s = t + \alpha$, $\beta = \alpha + \gamma$, we also have

$$\frac{\varphi(t+\alpha)-\varphi(t)}{\alpha} \leq \frac{\varphi(s+\gamma)-\varphi(s)}{\gamma} \qquad (1)$$

which implies $\varphi'_+(t) \leq \varphi'_+(s)$ for $t \leq s$, i.e., $\varphi'_+(t)$ is non-decreasing.

Similarly, the point $(t - \alpha, \varphi(t - \alpha))$ is below the segment joining $(t, \varphi(t))$ and $(t - \beta, \varphi(t - \beta))$, so

$$\frac{\varphi(t-\beta)-\varphi(t)}{-\beta} \leq \frac{\varphi(t-\alpha)-\varphi(t)}{-\alpha}$$

This again shows that the function $t \mapsto \frac{\varphi(t-\beta)-\varphi(t)}{-\beta}$ is non-increasing as $-\beta \uparrow 0$.

Hence it has a limit $\varphi'_-(t)$ (finite or $= +\infty$)

Further more, setting $p = t - \alpha$, $\beta = \alpha + \gamma$, we also have

$$\frac{\varphi(p-\gamma)-\varphi(p)}{-\gamma} \leq \frac{\varphi(t-\alpha)-\varphi(t)}{-\alpha}$$

which implies $\varphi'_-(p) \leq \varphi'_-(t)$ for $p < t$, i.e., $\varphi'_-(t)$ is non-decreasing.

Finally, as $\varphi'_+(t)$ is non-decreasing and $s - \alpha \leq s + \gamma$, by (1) we have

$$\frac{\varphi(s-\alpha)-\varphi(s)}{-\alpha} \leq \frac{\varphi(s+\gamma)-\varphi(s)}{\gamma}$$

and letting $-\alpha \uparrow 0$, $\gamma \downarrow 0$ yields $\varphi'_-(s) \leq \varphi'_+(s)$.

Further more, setting $t = t_1$, $s+\gamma = t_2$ in (1) and letting α, $\gamma \to 0$ yields

$$\varphi'_+(t_1) \le \varphi'_-(t_2) \text{ for } t_1 < t_2. \quad (2)$$

Hence from the proof above $\varphi(t)$ has left and right derivatives $\varphi'_-(t)$, $\varphi'_+(t)$ which are non-decreasing in a neighborhood of every $t \in (0,1)$. These derivatives are then non-decreasing on the whole interval (0,1) by (2). Therefore, by proposition (1.3) $\varphi(t)$ is convex on (0,1).

Lemma 2.6: Let K be either an open or a closed convex set having interior points. Let $x = x_0$ be a point of K and U be a convex neighborhood of x_0. Let F be a convex function on $K \cap U$. Then there exists a neighborhood U_1 of x_0 and a function F_1 defined and convex on K such that

$$F(x) = F_1(x) \text{ on } K \cap U_1.$$

Proof: Let U_2 be a small sphere $\|x - x_0\| < r$ such that F(x) is bounded on the closure of $K \cap U_2$. Let $G(x) = \beta\|x - x_0\| + F(x_0) - 1$, where β is a positive constant, chosen so large that $G(x) > F(x) + 1 > F(x)$ on the portion of the boundary of U_2 interior to K. Clearly $G(x) < F(x)$ holds for $x = x_0$, hence, for x on $K \cap U_1$ if U_1 is a suitably chosen neighborhood of x_0. If $x \in K$, define $F_1(x)$ to be max (F(x), G(x)) or G(x) according as is or is not in $U_2 \cap K$. Since max (F(x), G(x)) is convex on $U_2 \cap K$ and max (F(x), G(x)) = G(x) for x in a vicinity (relative to K) of the boundary of U_2 in K, it follows that $F_1(x)$ is convex on K. Finally, $F_1(x) = \max(F(x), G(x)) = F(x)$ for $x \in U_1 \cap K$.

Lemma 2.7: Let K be a closed, bounded convex set having $x = 0$ as an interior point. There exists a function h(x) defined and convex for all x such that $h(x) \le 1$ or $h(x) > 1$ according as $x \in K$ or $x \notin K$.

In fact, h(x) can also be chosen so as to satisfy $h(x) > 0$ for $x \neq 0$ and $h(cx) = ch(x)$ for $c > 0$. This function is then the supporting function of the polar convex set of K. The function h(x) is given by 0 or $\|x\| \Big/ \rho\left(\frac{x}{\|x\|}\right)$ according as $x = 0$ or $x \neq 0$, where, if u is a unit vector, ρ (u) is the distance from x = 0 to the point where the ray x = tu, t > 0 meets the boundary of K.

Definition.2.3: A function f: K $\rightarrow$ **R** defined on a convex set K$\subset$**R**^{n} is said to be locally d.c. if for every x$\in$K there exist a convex open neighborhood U of x and a pair of convex functions g, h on U such that f\U=g\U-h\U

Proposition 2.8: A locally d.c. function on an open or closed convex set K is d.c.on K.

Proof: The proof will be given for the case of an open convex set K. It will be clear from the proof and Lemma 2.6. how the proof should be modified for the case of a closed K.

Let f be an arbitrary locally d.c. function on K. By the definition of locally d.c. functions on K for each x $\in$ K there is a neighborhood U of x and a convex function F defined on K such that the function f + F is convex on U. Let D_1 be an arbitrary convex compact set in K. Then by the compactness of D_1 we can find a finite set $\{x_1, x_2,\ldots,x_n\} \subseteq D_1$ together with convex neighborhoods $U_1, U_2,\ldots,U_n$ of these points covering D_1 such that there are convex functions $F_1, F_2,\ldots, F_n$ on K such that f + F_i is convex on U_i.

Let $$F = \sum_{i=1}^{n} F_i$$

The function F is convex and moreover the function

$$f + F = f + F_j + \sum_{i=1, i\neq j}^{n} F_i$$

is convex on the set U_j, j=1,2,…,n. Thus the function f + F is convex on D_1.Thus there exists a sequence of open bounded sets $D_1, D_2, \ldots$ with the properties that the closure of D_i is contained in D_{i+1} such that $K = \bigcup D_i$, and to each D_i there corresponds a function F_i defined and convex on K such that f + F_i is convex on D_i.

Now introduce a sequence of compact convex sets $C_1, C_2,\ldots$ such that $C_1 \subset D_1 \subset C_2 \subset D_2 \subset \ldots$. In particular, $K = \bigcup C_i$.

Then we construct by induction a sequence of functions $\{G_i\}$ with the following properties

(i) G_i defined and convex on K,

(ii) F = G_i is convex on D_{i+1}

(iii) $G_i = F_i$ on D_i

We shall start to construct the function G_1 with the requested properties (i) – (iii). With out loss of generality we may assume that x = 0 is an interior point of C_1 and h (x) be the function given by the Lemma 2.7. when K is replaced by C_1. Let $\beta > 0$ be chosen so that $F_2(x) - \beta \leq F_1(x)$ on C_1. Let

$$H(x) = \begin{cases} 0 & \text{for } x \in C_1 \\ \beta[h(x) - 1] & \text{for } x \notin C1 \end{cases}$$

The function H is defined and convex on K. Moreover H(x) = 0 for $x \in C_1$. By the choice of β

$$F_2(x) - \beta + H(x) \leq F_1(x) \text{ for } x \in C_1 \qquad (3)$$

Since h(x) > 1 for $x \notin C_1$, there is a $\beta > 0$ such that

$$F_2(x) - \beta + H(x) > F_1(x) \qquad (4)$$

for x belonging to the boundary of D_1.

Now we define the function G_1 in the following way

$$G(x) = \begin{cases} \max[F_1(x), F_2(x) - \beta + H(x)] & \text{for } x \in D_1 \\ F_2(x) - \beta + H(x) & \text{for } x \in D_2 \setminus D_1 \end{cases}$$

Clearly (3) implies (iii) for i = 1 and by (4) $G_1 + F_2 - \beta + H$ is convex on D_1. Observe that $D_2 \setminus D_1$

$$F(x) + G_1(x) = f(x) + F_2(x) - \beta + H$$

is convex on D_2. Finally $f + G_1$ is convex on D_2.

Replacing in those considerations 1 by k and 2 by k+1 we obtain by the same inductional step the existence of a sequence of functions G_i satisfying (i) – (iii).

Now put $F(x) = \lim_{i \to \infty} G_i(x)$. F(x) exists uniformly on a compact subset of K; infact, $F(x) = G_i(x)$ on C_k for all $i \geq k$. Hence, F (x) is defined and convex on K. Since f (x) + F(x) is convex on C_k, k = 1,2,…. it is convex on K; that is, f is a d.c. function on K.

In the proof it was essentially used that the space is locally compact.

Definition 2.4: A mapping $F = (f_1, \ldots, f_m): K \to \mathbf{R}^m$ defined on a convex set $K \subseteq \mathbf{R}^n$ is said to be d.c. on K, written $F \in DC(K)$, if $f_i \in DC(K)$ for every i = i,..., m.

Proposition 2.9: Let $K_1 \subset \mathbf{R}^n$, $K_2 \subset \mathbf{R}^m$ be convex sets such that K_1 is open or closed, K_2 is open. If $f_1: K_1 \to K_2$, $F_2: K_2 \to \mathbf{R}^k$ are d.c mappings then $F_2 o R_1$: $K_1 \to \mathbf{R}^k$ is also a d.c. mapping.

Proof: It suffices to show that $F = (f_1,...,f_m)$: $K_1 \to K_2$ is d.c. and g: $K_2 \to \mathbf{R}$ is convex then $g(f_1,..., f_m) \in DC(K_1)$. Let $x \in K_1$ and $y = F(x) \in K_2$. The convex function g(y) can be represented in a neighborhood U_2 of y as the pointwise supremum of a family of affine functions: $g(y) = \sup_t l_t$ where $l_t = a_{0t} + a_{1t}y_1 + a_{2t}y_2 + \ldots + a_{mt}y_m$ ($y_1,..., y_m$ are coordinates of y in $\mathbf{R}^m$) and $M = \sup_{i,t}|a_{it}| < +\infty$. Let $f_i(x) = f_i'(x) - f_i''(x)$ where f_i' and f_i'' are convex function on K in a neighborhood U_1 of x such that $F(U_1) \subset U_2$. Then

$$l_t(f_1,...,f_m) = a_{0t} + \sum_{i=1}^{m} a_{it} f_i' - \sum_{i=1}^{m} a_{it} f_i''$$

$$= [\ a_{0t} + \sum_{i=1}^{m} (M + a_{it}) f_i' + \sum_{i=1}^{m} (M - a_{it}) f_i''\ ] - M \sum_{i=1}^{m} (f_i' + f_i'')$$

$$= p_t - q$$

with p_t and q convex and q independent of t.

Then $g(f_1,...,f_m) = \sup_t l_t(f_1,...,f_m) = \sup_t (p_t - q) = \sup_t p_t - q = p - q$, i.e., $g(f_1,...,f_m)$ is locally d.c. on K_1. Hence, by Proposition 2.8., $gof \in DC(K_1)$.

Corollary 2.10: Let K_1, K_2 be as in Proposition 2.9. If F_1: $K_1 \to K_2$ is d.c. in $K_1 \subset \mathbf{R}^n$ and F_2: $K_2 \to \mathbf{R}^k$ is C^2 - smooth, then $F_2 o F_1$ is d.c. on K_1. In particular, if f(x) is d.c. on an open (or a closed) convex set K and $f(x) \neq 0$ for all $x \in K$, then $\frac{1}{f(x)}$ and $|f(x)|^{\frac{1}{m}}$ are d.c.on K and the product of two d.c. function on K is d.c. on K.

The above properties explain why the majority of functions of practical interest are d.c.

Now recall from convex analysis that: if h: $K \to \mathbf{R}_+$ is a convex (concave resp.) function on a convex subset K of $\mathbf{R}^m$ and if q: $\mathbf{R}_+ \to \mathbf{R}$ is a convex (concave, resp.) non-decreasing function, then q(h(x)) is a convex (concave, resp.) function on K.

Proposition 2.11: Let h: K→ $\mathbf{R}_+$ be a convex function on a convex compact subset K of $\mathbf{R}^m$. If q: $\mathbf{R}_+ \to \mathbf{R}$ is a convex non increasing function such that $q'_+(0) > -\infty$, then q(h(x)) is a d.c. function on K:

$$q(h(x)) = g(x) - kh(x),$$

where g is a convex function and k is a positive constant satisfying $k \geq |q'_+(0)|$ ($q'_+(t)$ denotes the right derivative of q(t) at point t).

Proof: We have $q'_+(0) \leq q'_+(t) \leq 0 \; \forall t \geq 0$, therefore $\tilde{q}(x)$ = q(t) + kt satisfies $\tilde{q}'_+(x) = q'_t(x) + k \geq q'_+(0) + k \geq 0, \forall t \geq 0$. So $\tilde{q}$ is a convex non-decreasing function and hence $\tilde{q}(h(x))$ is a convex function on K. S ince $\tilde{q}$ (h(x)) = q(h(x)) + kh(x), the result follows.

Analogously: Let h: K→ $\mathbf{R}_+$ be a convex function as in Proposition 2.11. If q: $\mathbf{R}_+ \to \mathbf{R}$ is a concave non-decreasing function such that $q'_+(0) < +\infty$, then q(h(x)) is a d.c. function on K:

$$q(h(x)) = kh(x) - g(x),$$

where g(x) is a convex function and k is a positive constant satisfying $k \geq |q'_+(0)|$

Proof: We have that $q'_+(0) \geq q'_+(0) \geq 0 \; \forall t \geq 0$, therefore $\tilde{q}$ (t) = kt − q(t) satisfies $\tilde{q}'_+(t) = k - q'_+(t) \geq k - q'_+(0) \geq 0 \; \forall t \geq 0$.

So $\tilde{q}$ is convex non-decreasing and hence $\tilde{q}$ (h(x)) is a convex function on K.

$\tilde{q}$ (h(x)) = kh(x) - q(h(x)), the result follows.

Letting g(x) = $\tilde{q}$ (h(x)), q(h (x)) = kh (x) - g(x).

Example: The function $\omega e^{-\theta\|x-a\|}$ (with $\omega > 0$, $\theta > 0$) is d.c., since it is equal to $q(\|x-a\|)$ and $q(t) = \omega e^{-\theta t}$ is convex decreasing function with $q'_t(0) = \theta\,\omega > -\infty$.

2.5.D.C.SETS

Definition 2.5: A set $M \subseteq \mathbf{R}^n$ is said to be a d.c. set if there exists a closed convex set $D \subseteq \mathbf{R}^n$ and an open convex set $C \subseteq \mathbf{R}^n$ such that $M = D \backslash C$. Since D and C are convex sets we can find convex functions g, h: $\mathbf{R}^n \rightarrow \mathbf{R}$ such that $D=\{x \in \mathbf{R}^n: g(x) \leq 0\}$ and $C=\{x \in \mathbf{R}^n: h(x) < 0\}$. We, therefore, have

$$\begin{aligned} M &= D \backslash C \\ &= D \cap C^c \\ &= \{x \in \mathbf{R}^n: g(x) \leq 0\} \cap \{x \in \mathbf{R}^n: h(x) \geq 0\} \\ &= \{x \in \mathbf{R}^n: g(x) \leq 0, -h(x) \leq 0\} \\ &= \{x \in \mathbf{R}^n: \max [g(x), -h(x)] \leq 0\} \end{aligned}$$

By Proposition 2.1 (iv), max $\{g(x), -h(x)\}$ is a d.c. function since g(x) is convex and $-h(x)$ is concave (see example 2.1). Therefore, we can find convex functions p(x) and q(x) which are defined on $\mathbf{R}^n$ such that

$$\max [g(x), -h(x)] = p(x) - q(x)$$

which implies that:

$$M = \{x \in \mathbf{R}^n: p(x) - q(x) \leq 0\} \tag{5}$$

From (5) we conclude that a d.c. set can also be defined by a d.c. inequality.

Conversely, if set S is defined by a d.c. inequality $S=\{x: g(x) - h(x) \leq 0\}$, with f(x) and h(x) convex, then clearly $S = \{x:(x, t) \in M \text{ for some } t\}$, where

$$M = \{(x, t) \in \mathbf{R}^n x \mathbf{R}: g(x) - t \leq 0, t - h(x) \leq 0\}.$$

Therefore, S can be obtained as the projection on $\mathbf{R}^n$ of the d.c. set $M \subset \mathbf{R}^{n+1}$. D.c sets are not different from arbitrary closed sets. This can be seen from the observation of Asplund (1973) that for any closed set $M \subset \mathbf{R}^n$, if $d(x, S) = \inf \{\|x - y\| : y \in S\}$ denotes the distance from x to S then the function $x :\rightarrow \|x\|^2 - d^2(x, S) = \sup \{2x^T y - \|y\|^2 : y \in S\}$ is convex. Thus, $S=\{x: d^2(x, S) \leq 0\}$ where $d^2(x, S)$ is a d.c. function.

Having (5) in mind, given any d.c. minimization problem subject to a d.c. constrained set we can reduce it to unconstrained d.c.minimization problem using the Lagrange approach for solving a constrained optimization problem.

Consider the d.c. optimization problem

$$(P) \qquad f(x) \to \min, \ x \in M \qquad (6)$$

where f is a d.c.function on R^n and M is a d.c. set. By (5), problem (P) is equivalent to

$$(P^*) \qquad f(x) \to \min, \ x \in M = \{x \in \mathbf{R}^n : p(x) - q(x) \leq 0\}$$

The Lagrange function for (P*) is defined by

$$L(x, \lambda) = f(x) + \lambda\,(p(x) - q(x))$$

The function $L(x, \lambda)$ is a d.c. function for all $\lambda \in \mathbf{R}^n_+$ since $\lambda\,(p(x) - q(x))$ is d.c. by proposition 2.1(ii) and the sum of d.c.functions is d.c. by Proposition 2.1(i). Then

$$(P^*_\lambda) \qquad L(x, \lambda) \to \min, \ (x, \lambda) \in \mathbf{R}^n \times \mathbf{R}_+$$

is a d.c.minimization problem without constraint and solving (P) reduces to solving (P^*_λ).

In the next sections we shall consider only unconstraint d.c. minimization problems.

Chapter 3
Duality in d.c.optimization

A powerful technique in modern optimization is the dualization of various concepts (polar sets, conjugate functions, dual programs, etc.). This technique relies on the dual nature of convex sets which can be described in a two way, either by specifying their points or by specifying their supporting hyperplanes (which correspond to linear functionals, i.e., points in the dual space.)

Several important duality concepts have been developed and studied for non-convex optimization in the last decades [Toland [7]].

In fact, for each non-convex program given in space X we would like to be able to define a dual program in the dual space X^* such that a close relationship exists between certain properties of primal and corresponding properties of the dual and information on the dual may help to better understand the primal and ideally, solve the primal more efficiently.

3.1: Fenchel Duality

Let X be a normed space and X^* be the dual space of X. Let $f: X \to \mathbf{R} \cup \{+\infty\}$ with $\operatorname{dom} f = \{x \in X : f(x) \in \mathbf{R}\} \neq \phi$, and Let $g: X \to \mathbf{R} \cup \{+\infty\}$ with $\operatorname{dom} g = \{x \in X: g(x) \in \mathbf{R}\} \neq \phi$.

Then (i) the function $f^*: X^* \to \mathbf{R} \cup \{+\infty\}$, where $\operatorname{dom} f^* = \{x \in X^*: f^*(\mathbf{x}^*) \in \mathbf{R}\}$, is said to be the upper Fenchel-conjugate of f if

$$f^*(x^*) = \sup_{x \in X} \{< x, x^* > - f(x)\}, x^* \in X^* \tag{1}$$

(ii) the function $g_*: X^* \to \mathbf{R} \cup \{+\infty\}$ where $\operatorname{dom} g_* = \{x \in X^*: g_*(\mathbf{x}^*) \in \mathbf{R}\}$, is said to be the lower Fenchel-conjugate of g if

$$g_*(x^*) = \inf_{x \in X} \{< x, x^* > - g(x)\}, x^* \in X^* \tag{2}$$

Fenchel's inequalities

i) $f(x) + f^*(x^*) \geq \langle x, x^* \rangle$ for all $x \in \text{dom } f$ and for all $x^* \in \text{dom } f^*$.

ii) $g(x) + g_*(x^*) \leq \langle x, x^* \rangle$ for all $x \in \text{dom } g$ and for all $x^* \in \text{dom } g_*$.

The proof is immediate from the definitions (1) and (2)

Example 1: Let X be a Hilbert Space and let $f(x) = \|x\|^2 = \langle x, x \rangle, x \in X$.

Then from $X = X^*$ we have

$$f^*(x^*) = \sup_{x \in X}\{\langle x, x^* \rangle - f(x)\}, x^* \in X^* = X$$

Since

$$0 \leq \left\|x^* - 2x\right\|^2 = \langle x^* - 2x, x^* - 2x \rangle = \langle x^*, x^* \rangle - 4 \langle x^*, x \rangle + 4 \langle x, x \rangle$$

$$= 4[\frac{1}{4} \langle x^*, x^* \rangle - (\langle x, x^* \rangle - \langle x, x \rangle)] \text{ for all } x \in X$$

This implies

$$\langle x, x^* \rangle - \langle x, x \rangle \leq \frac{1}{4} \langle x^*, x^* \rangle \text{ for all } x \in X$$

Thus

$$f^*(x^*) = \tfrac{1}{4} \langle x^*, x^* \rangle, x^* \in X^*.$$

Example 2: Let X be a normed space and let $f(x) = \|x\|, x \in X$. Then we have

$$f^*(x^*) = \sup_{x \in X}\{\langle x, x^* \rangle - f(x)\} = \begin{cases} 0, & \text{if } \|x^*\| \leq 1 \\ \infty & \text{otherwise} \end{cases}, \quad x^* \in X^*$$

because of the inequality $\left|\langle x, x^* \rangle\right| \leq \|x\|\left\|x^*\right\|$ we have

$$\langle x, x^* \rangle - \|x\| \leq \|x\|\left\|x^*\right\| - \|x\| = \|x\|[\left\|x^*\right\| - 1] \leq \begin{cases} 0 & \text{if } \|x^*\| \leq 1 \\ \infty & \text{else} \end{cases}$$

Example 3: Let X be a vector space and let $f(x) = \frac{1}{2} < x, Ax >$ where A is a symmetric matrix.

Then we have

$$f^*(x^*) = \sup_{x \in X} \{< x, x^* > - f(x)\}, x^* \in X^*$$

Now let $h(x) = < x, x^* > - \frac{1}{2} < x, Ax >$

Then $\nabla h(x) = x^* - Ax$ and $\nabla h(x) = 0$ implies $x = A^{-1}x^*$. Thus the maximum of h will be reached at $x = A^{-1}x^*$.

Then the maximum (supremum) of h is

$$h(x) = h(A^{-1}x^*). = \tfrac{1}{2} < A^{-1}x^*, x^* > .$$

Therefore,

$$f^*(x^*) = \tfrac{1}{2} < x^*, A^{-1}x^* >, x^* \in X^*.$$

Recall the subdifferential of a function f at x^0 is the set

$$\partial g(x^0) = \{p \in X^* : g(x) \geq g(x^0) + < x - x^0, p >: x \in X\}.$$

Theorem 3.1: Let f: $\mathbf{R}^n \to \overline{\mathbf{R}}$ and let f* be its conjugate. Then $u^* \in \partial f(u)$ if and only if $f(u) + f^*(u^*) = \langle u, u^* \rangle$.

Proof: See in [6]

Corollary 3.2: If f: $\mathbf{R}^n \to \overline{\mathbf{R}}$ is a closed proper convex function, ∂f^* is the inverse of ∂f in the sense of multivalued mappings, i.e., $x \in \partial g^*(x^*)$ if and only if $x^* \in \partial g(x)$. This observation important in the subsequent sections.

3.2: Lagrange Duality

Let X be a normed space, let $U \subseteq X$, let $f : U \to \mathbf{R}$ and let g: $U \to \mathbf{R}^m$. Then we consider the following optimization problem

(P) $\qquad f(x) \to \min, x \in S$ such that

$$S = \{x \in U : g(x) \leq 0\}.$$

The Lagrange-function with regard to (P) is given by

$$L(x,\lambda) = f(x) + \langle \lambda, g(x) \rangle, \ (x,\lambda) \in U x \mathbf{R}_{+}^{m}.$$

Now we define

$$\psi(\lambda) = \inf_{x \in U} L(x,\lambda), \ \lambda \in S^*,$$
$$S^* = \{\lambda \in R_{+}^{m} : \inf_{x \in U} L(x,\lambda) > -\infty\}.$$

Then we consider the following optimization problem

(DL) $$\psi(\lambda) \rightarrow \max, \lambda \in S^*.$$

Theorem 3.3: (Weak duality)

Let $S \neq \phi$, $S^* \neq \phi$. Then

$$\psi(\lambda) \leq f(x) \text{ for all } x \in S \text{ and for all } \lambda \in S^*, \tag{3}$$

Or

$$\sup_{\lambda \in S^*} \psi(\lambda) \ \leq \ \inf_{x \in S} f(x) \tag{4}$$

Proof: From $x \in S$ we get $g(x) \leq 0$. Since $\lambda \in S^*$ we have $\lambda \geq 0$ and therefore $\langle \lambda, g(x) \rangle \leq 0$.

Then we have for all $x \in S$ and for all $\lambda \in S^*$

$$\psi(\lambda) = \inf_{x \in U} L(x,\lambda) \leq \ L(x,\lambda) = f(x) + \langle \lambda, g(x) \rangle \leq f(x)$$

From this the claim holds.

Since inequality (3) is true we call (DL) the Lagrange- dual problem of (P).

3.3: DC duality (Tolland's result)

Although in d.c.optimization problems convexity is present only partially or in the reverse direction, we can build up a meaningful duality principle for d.c.optimization and dually schemes for minimization of d.c.functions which was formulated by J.F.Toland ([7], [8], [11] and [12])

Let X and X^* be pair of linear spaces in duality and let $<,>: X x X^* \rightarrow \mathbf{R}$ denote the corresponding bilinear form, which determines the duality between X and X^*. If f: $X \rightarrow \mathbf{R}$ is linear functional, then we shall consider the optimization problem P:

(P) $$\inf_{x \in X} f(x)$$

We might consider the maximization problem

$$(-P) \qquad \sup_{x \in X}(-f(x)) \text{ instead.}$$

Let Y, Y^* be another pair of linear spaces in duality and let $\Phi : X \times Y \to \mathbf{R}$ be a functional with the following property:

$$\Phi(x,0) = -f(x) \text{ for all } x \in X \tag{5}$$

Then the Lagrange function L is defined on $X \times Y^*$ by

$$-L(x,u^*) = \sup_{u \in Y} \{< u, u^* > - \Phi(x,u)\} \tag{6}$$

For each $u^* \in Y^*$ put the primal optimization problem:

$$-L(u^*) = \sup_{x \in X} L(x,u^*) \tag{7}$$

Theorem 3.4:

$$\inf_{x \in X} f(x) \leq \inf_{u^* \in Y^*} L(u^*)$$

Proof: By (5),

$$\Phi(x,0) = -f(x) \text{ for all } x \in X$$

$$\Rightarrow \sup \Phi(x,0) = \sup(-f(x)) \text{ for all } x \in X$$

$$\Rightarrow \sup \Phi(x,0) = -\inf f(x) \text{ for all } x \in X$$

$$\Rightarrow \Phi(x,0) \leq -\inf f(x) \text{ for all } x \in X \tag{8}$$

And by (6), $-L(x, u^*) = \sup_{u \in Y} \{< u, u^* > - \Phi(x,u)\}$

For $u = 0$ we have

$$L(x,u^*) \leq \Phi(x,0)$$

$$\Rightarrow \Phi(x,0) = \sup_{u^* \in Y^*} L(x,u^*) \tag{9}$$

From (8) and (9) we get

$$-\inf_{x \in X} f(x) \geq \Phi(x,0) = \sup_{u^* \in Y^*} L(x,u^*), \forall x \in X$$

$$-\inf_{x \in X}(x) \geq \sup_{u^* \in Y^*} L(x,u^*)$$

$$\inf_{x\in X}(x) \le -\sup_{x\in X}\sup_{u^*\in Y^*} L(x,u^*) = \inf_{u^*\in Y^*}\left(-\sup_{x\in X}(x,u^*)\right) = \inf_{u^*\in Y^*} L(u^*) \text{ by (7)}$$

Thus,

$$\inf_{x\in X} f(x) \le \inf_{u^*\in Y^*} L(u^*).$$

For each $x \in X$ define $\Phi_x : Y \to \mathbf{R}$ by $\Phi_x(u) = \Phi(x,u)$ for each $u \in Y$.

Theorem3.5: If for all $x \in X$,

$$\Phi_x^{**}(0) = \Phi_x(0) \tag{10}$$

then

$$\inf_{x\in X} f(x) = \inf_{u^*\in Y^*} L(u^*) \tag{11}$$

Proof: By (5) we have $-f(x) = \Phi_x(0) = \Phi(x,0)\ \forall x \in X$

$$\Rightarrow\ \sup(-f(x)) = \sup \Phi_x(0)\ \ \forall x \in X$$

$$\Rightarrow\ -\inf(f(x)) = \sup \Phi_x(0) = \sup \Phi_x^{**}(0)\ \ \forall x \in X \text{ by (10)} \tag{12}$$

but $\Phi_x^{**}(0) = -\sup_{u^*\in Y^*}\{<u^*,0> - \Phi_x^*(u^*)\} = \sup_{u^*\in Y^*}(-\Phi_x^*(u^*))$ (13)

and

$$-\Phi_x^*(u^*) = -\sup_{u\in Y}\{<u,u^*> - \Phi_x(u)\} = L(x,u^*) \text{ by (6)} \tag{14}$$

Combining (13) and (14) and using (7) we get

$$\Phi_x^{**}(0) = -L(u^*) \tag{15}$$

So, using (12) and (15) we have

$$\inf_{x\in X} f(x) = \inf_{u^*\in Y^*} L(u^*)$$

Corollary 3.6: If $\partial\Phi_x(0) \ne \phi$ for all x in X, or if $\Phi_x : Y \to \mathbf{R}$ is convex and lower semi-continuous for all x in X, then (11) holds.

Proof: Since $\partial\Phi_x(0) \ne \phi$, let $p \in \partial\Phi_x(0)$. Then

$$\Phi_x(0) + \Phi_x^*(p) = <0,p>$$

$$\Rightarrow -\Phi_X(0) = \Phi_X^*(p) \qquad (16)$$

But

$$\Phi_X^{**}(0) = \sup_{p \in Y^*} \{ < p,0 > - \Phi_X^*(p) \}$$
$$= \sup_{p \in Y^*} \Phi_X(0) \qquad \text{by (16)}$$
$$= \Phi_X(0)$$

Therefore, by Theorem 3.5, the result follows.

Now put the dual optimization problem D to be:

$$\text{(D)} \qquad \inf_{u^* \in Y^*} L(u^*)$$

and an element $u_\circ^* \in Y^*$ will be called a solution of D if

$$L(u_\circ^*) = \inf_{u^* \in Y^*} L(u^*) \in \mathbf{R}.$$

The duality result in Theorem 3.5. is true whether solutions for (P) or (D) exist or not. In the next theorem we examine the relationship between solution of (P) and of (D).

Theorem 3.7: If $x^* \in X$ solves (P) and $u_\circ^* \in \partial\Phi_{x^*}(0)$, then $u_\circ^*$ solves (D).

Further more

$$\Phi(x^*,0) - L(x^*,u_\circ^*) = 0 \qquad (17)$$

$$L(u_\circ^*) + L(x^*,u_\circ^*) = 0 \qquad (18)$$

Proof: Since Theorem 3.4. always holds it will suffice to show that

$$L(u_\circ^*) = f(x^*).$$

Let $f(x^*) = \inf_{x \in X} f(x) = \alpha \in \mathbf{R}$. Then

Because $u_\circ^* \in \partial\Phi_{x^*}(0)$, we have

$$\Phi(x^*,u) = \Phi_{x^*}(u) \geq \Phi_{x^*}(0) + < u - 0, u_\circ^* > \quad \text{(subgradient inequality)}$$
$$= \Phi(x^*,0) + < u, u_\circ^* >$$
$$= -f(x^*) + < u, u^* > \qquad \text{by (5)}$$

$$= -\alpha + < u, u_\circ^* > \forall u \in Y \text{ (since } f(x^*) = \alpha)$$

$$\Rightarrow \alpha \geq < u, u_\circ^* > -\Phi(x^*, u) \forall u \in Y$$

$$\alpha \geq \sup_{u \in Y}\{< u, u_\circ^* > -\Phi(x^*, u)\}$$

$$\alpha \geq -L(x^*, u_\circ^*) \text{ by (6)}$$

Therefore,

$$-L(u_\circ^*) = \sup_{x^* \in X} L(x^*, u_\circ^*) \geq -\alpha \quad \text{implies}$$
$$L(u_\circ^*) \leq \alpha = f(x^*)$$

By Theorem 3.4, it follows that $L(u_\circ^*) = f(x^*)$ and $u_\circ^*$ solves D.

Since

$$\Phi(x^*, 0) = -f(x) = -\alpha \text{ and } L(u_\circ^*) = \alpha, \text{we get}$$

$$\Phi(x^*, 0) + L(u_\circ^*) = 0 \tag{19}$$

Since

$u_\circ^* \in \partial\Phi_{x^*}(0)$, we have

$$\Phi_{x^*}(0) + \Phi_{x^*}^*(u_\circ^*) = 0 \tag{20}$$

In other words, since

$$-L(x^*, u_\circ^*) = \sup_{u \in Y}\{< u, u_\circ^* > - \Phi(x^*, u)\} = \Phi_{x^*}^*(u_\circ^*),$$

from (20) we obtain

$$\Phi(x^*, 0) - L(x^*, u_\circ^*) = 0 \tag{21}$$

Subtracting (21) from (19) the result (18) follows.

Thus both (18) and (19) holds. This completes the proof of the Theorem.

Remark: We have used the Lagrangian L to define a dual optimization problem not to define a minimax problem equivalent to (P).

Let $\Gamma_\circ(X)$ denote the set of all proper semicontnuous convex functions defined on X.

Theorem 3.8: Let g, $h \in \Gamma_\circ(X)$ and consider the d.c.optimization problem

$$(P) \qquad \lambda = \inf_{x \in X}\{g(x) - h(x)\}$$

and $g^*, h^* \in \Gamma_\circ(X^*)$. Then

$$\inf_{x \in X}\{g(x) - h(x)\} = \lambda = \inf_{x^* \in X^*}\{h^*(x^*) - g^*(x^*)\} \tag{22}$$

Proof: $\lambda = \inf_{x \in X}\{g(x) - h(x)\}$

$$= \inf_{x \in X}\{g(x) - \sup_{x^* \in X^*}\{< x, x^* > - h^*(x^*)\} \quad \text{by}(1)$$

$$= \inf_{x \in X}\inf_{x^* \in X^*}\{g(x) + h^*(x^*) - < x, x^* >\}\}$$

$$= \inf_{x^* \in X^*}\{h^*(x^*) + \inf_{x \in X}\{g(x) - < x, x^* >\}\}$$

$$= \inf_{x^* \in X^*}\{h^*(x^*) - \sup_{x \in X}\{< x, x^* > - g(x)\}\}$$

$$= \inf_{x^* \in X^*}\{h^*(x) - g^*(x^*)\} \quad \text{by}(1)$$

Hence, the result follows.

Let us define the dual optimization problem (Q) of (P) by

$$(Q) \qquad \lambda = \inf_{x^* \in X^*}\{h^*(x) - g^*(x^*)\}$$

and let Ω denote the solution set (P) and Δ denote the solution set of (Q). Now it is obvious that there is symmetry between (P) and (Q) and the dual of (Q) is exactly (P). Note that the finiteness of λ merely implies that

$$\text{dom } g \subset \text{dom } h \text{ and dom } h^* \subset \text{dom } g^* \tag{23}$$

Chapter 4

D.C algorithm (DCA) for solving d.c.optimization

So far we have been concerned with the motivation and theoretical foundation of d.c. optimization. The remaining sections will be devoted to methods and algorithms for solving d.c. optimization problem. We have two classes of algorithms that enable us to solve d.c. optimization problems.

I. The class of algorithms of combinatorial optimization

This class is introduced by Hoang Tuy ([2], [3]) and it is made of algorithms of " cutting plane" type (i.e., using hyper planes for localizing an optimal solution.)

Although these algorithms are applicable only to problems of small size, due to their complexity and their cumbersomeness concerning their effective programming; in principle they lead to an optimal solution.

II. The class of algorithms of convex and nonconvex optimization

Here based on the duality in d.c. optimization we deal with subgradient methods. The sub gradient algorithms are easy to program and infact practical problems of large size are also well treated by it. However we cannot claim whether the local optimal solution obtained are effectively global optimal solution unless an initial vector x^0 can be obtained by the algorithms of the first class.

On the other hand the algorithms of the first class can be furnished with a " good " local optimal solution by the algorithms of the second class, which accelerate their convergence.

This seminar focuses only on the duality in d.c. optimization (DCO) and on methods of sub gradient for solving DCO problems.

DCA is an iterative method. The essence of an iterative method of optimization is that we have a trial solution x* to the problem and look for a better one.

Definition 4.1: An 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 exist a neighborhood U of x* such that

$$g(x^*) - h(x^*) \leq g(x) - h(x) \; \forall x \in U \qquad (1)$$

Adopting the convention $+\infty - (+\infty) = +\infty$, property (1) is equivalent to

$$g(x^*) - h(x^*) \le g(x) - h(x) \forall x \in U \cap dom\, g$$

Definition 4.2:A point x* is a critical point of g - h if $0 \in \partial g(x^*) - \partial h(x^*)$

From definition 4.2 we can conclude that a point x* is said to be a critical point of g-h if

$$\partial g(x^*) \cap \partial h(x^*) \neq \phi$$

Now for each $x^* \in X$, let $\omega(x^*)$ denote the problem

$$\omega(x^*) \qquad \inf_{y \in \partial h(x^*)} \{h^*(y) - g^*(y)\} \qquad (2)$$

Because $h(x^*) + h^*(y) = \langle x^*, y\rangle$ for all y in $\partial h(x^*)$, (2) is equivalent to

$$\inf_{y \in \partial h(x^*)} \{-h(x^*) + \langle x^*, y\rangle - g^*(y)\}$$

which is equivalent to

$$-h(x^*) + \inf_{y \in \partial h(x^*)} \{\langle x^*, y\rangle - g^*(y)\}$$

which is also equivalent to

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

since $-h(x^*)$ is a constant.

Let $W(x^*)$ denote the solution set of $\omega(x^*)$.

Similarly and dually, for each $y^* \in X^*$, we define the problem $\eta(y^*)$ by

$$\eta(y^*) \qquad \inf_{x \in \partial g^*(y^*)} \{g(x) - h(x)\} \qquad (3)$$

which is equivalent to

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

because $g^*(y^*) + g(x) = \langle x, y^*\rangle$ for all x in $\partial g^*(y^*)$.

Let its solution set is denoted by $N(y^*)$.

The idea of d.c.algorithm (DCA) is that: given a point x^0 in dom g, it constructs two sequences $\{x^k\}$ and $\{y^k\}$(candidates to the primal and dual solutions respectively) as follows

$$\begin{array}{lll} x^0 & \mapsto & y^0 \in W(x^0) \\ x^1 \in N(y^0) & \mapsto & y^1 \in W(x^1) \\ x^2 \in N(y^1) & \mapsto & y^2 \in W(x^2) \\ & \vdots & \\ x^{k+1} \in N(y^k) & \mapsto & y^{k+1} \in W(x^{k+1}) \end{array}$$

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

Lemma 4.1:Sequence $\{x^k\}$, $\{y^k\}$ in the DCA are well defined if and only if

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

Since for $\varphi \in \Gamma_0(X)$ *we have* ri $(\text{dom}\,\varphi) \subset \text{dom}\,\partial\varphi \subset \text{dom}\varphi$ (where $\text{ri}(\text{dom}\varphi)$ stands for the relative interior of dom φ) (see Rockafellar (1972)[6] page 227) we can say, under the essential assumption (23) of Chapter 3, that the DCA is well defined.

4.1:Global optimality criteria for d.c.optimization

The core of global optimization is: given a solution x*, check whether it is a global optimal solution, and if it is not, find a better feasible solution. One way to deal with this issue is devising criteria for recognizing a global solution.

Based on the theory of the subdifferential and of the conjugate of convex functions we proof the following basic result of global duality in DCO.

Theorem 4.2: Let $g, h \in \Gamma_0(X)$, then we have

i. $\partial h(x^*) \subset \partial g(x^*), \forall x^* \in \Omega$

ii. $\partial g^*(y^*) \subset \partial h^*(y^*), \forall y^* \in \Delta$

iii. $\bigcup_{y^* \in \Delta} \partial g^*(y^*) \subset \Omega$ and equality hold if h^* is suddifferentiable on Ω (for instance $\Omega \subset \text{ri}(\text{dom}\,h)$ where ri stands for relative interior of domh.

iv. $\bigcup_{x^*\in\Omega} \partial h(x^*) \subset \Delta$ and equality hold if g* is subdifferentiable on Δ (for instance if $\Delta \subset \text{ri}(\text{dom}\, g^*)$)

Proof: By duality, we only need to show properties (i) and (iv).

Let $x^* \in \Omega$. $\forall y^* \in \partial h(x^*)$ we have

$$\lambda = g(x^*) - h(x^*)$$

$$= g(x^*) - \langle x^*, y^* \rangle + h^*(y^*) (\text{since } y^* \in \partial h(x^*) \Leftrightarrow h(x^*) + h^*(y^*) = \langle x^*, y^* \rangle)$$

$$\geq h^*(y^*) - g^*(y^*) \ (\text{since } g(x^*) + g^*(y^*) \geq \langle x^*, y^* \rangle) \tag{4}$$

But λ is infimum, i.e.,

$$\lambda \leq h^*(y) - g^*(y) \forall y \in \text{dom} h^* \quad (\text{by (22 of Chapter 3)})$$

In particular for $y = y^*$, we have $\lambda \leq h^*(y^*) - g^*(y^*)$ (5)

From (4) and (5) we conclude

$$\lambda = h^*(y^*) - g^*(y^*) \text{ and } y^* \in \Delta \tag{6}$$

Thus,

$$\lambda = g(x^*) - \langle x^*, y^* \rangle + h^*(y^*) = h^*(y^*) - g^*(y^*) \forall y^* \in \partial h(x^*)$$

$$\Rightarrow g(x^*) + g^*(y^*) = \langle x^*, y^* \rangle \forall y^* \in \partial h(x^*)$$

$$\Rightarrow y^* \in \partial g(x^*) \forall y^* \in \partial h(x^*) \tag{7}$$

From (7) we conclude that $y^* \in \partial g(x^*) \quad \forall y^* \in \partial h(x^*)$ and hence

$$\partial h(x^*) \subset \partial g(x^*) \quad \forall x^* \in \Omega \tag{8}$$

Thus, we have proved (i).

Since by (6), $y^* \in \Delta$, we conclude for each $x^* \in \Omega$ we have $\partial h(x^*) \subset \Delta$, which implies

$$\bigcup_{x^*\in\Omega} \partial h(x^*) \subset \Delta \tag{9}$$

On the other hand if g* is subdifferential on Δ, then for all y* in Δ there exists an x* in $\partial g^*(y^*) \subseteq \partial h^*(y^*)$ (by (ii) of the same theorem) which implies x* is in $\partial h^*(y^*)$ and hence y* is in $\partial h(x^*)$ for all x^* in Ω. This in turn implies $\Delta \subseteq \partial h(x^*)$ for each x*in Ω.

Which again shows

$$\Delta \subseteq \bigcup_{x^* \in \Omega} \partial h(x^*) \qquad (10)$$

Therefore, by (9) and (10), the equality holds.

The containment in (ii) and (iv) shows that solving the primal d.c. program (P) implies solving the dual program (Q) and vice versa. The global optimality criteria in (i) is difficult to use for devising solutions method to problem (P).

So, the DCA is based on local optimality criteria.

4.2: Local Optimality Criteria for d.c. optimization

Let us consider now local solutions of problems (P) and (Q). We begin by defining sets:

$$\Omega^* = \{x^* \in X : \partial h(x^*) \subset \partial g(x^*)\}$$
$$\Delta^* = \{y^* \in X^* : \partial g^*(y^*) \subset \partial h^*(y^*)\}$$

Also, $\text{dom}\,\partial g = \{x \in X : \partial g(x) \neq \phi\}$ and $\text{range}\,\partial g = \bigcup\{\partial g(x) : x \in \text{dom}\,\partial g\}$

4.2.1: Necessary and sufficient conditions for local optimality for DCO

Theorem 4.3:

(i) If x^* is a local minimizer of $g - h$, then $x^* \in \Omega^*$.

(ii) Let x^* be a critical point of g-h and $y^* \in \partial g(x^*) \cap \partial h(x^*)$. Let U be a neighborhood of x^* such that $U \cap \text{dom}\, g \subset \text{dom}\,\partial h$. If for any x in $U \cap \text{dom}\, g$ there is a 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,

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

Proof: (i) Let x^*be a local minimizer of g-h. Then there exists a neighborhood U of x^* such that

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

Now let $y^* \in \partial h(x^*)$. Then

$$h(x) - h(x^*) \geq \langle x - x^*, y^* \rangle \ \forall x \in U \cap \text{dom}\, g.$$

which follows that

$$g(x) - g(x^*) \geq \langle x - x^*, y^* \rangle \ \forall x \in U \cap \text{dom}\, g.$$

Due to the convexity of g this implies that $y^* \in \partial g(x^*)$.

Therefore, $x^* \in \Omega^*$.

(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,

$$g(x^*) - h(x^*) = h^*(y^*) - g^*(y^*) \tag{11}$$

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

$$h^*(y) - g^*(y) \geq h^*(y^*) - g^*(y^*) \tag{12}$$

Since y is in $\partial h(x)$ we have

$$h(x) + h^*(y) = \langle x, y \rangle \leq g(x) + g^*(y)$$

which implies

$$g(x) - h(x) \geq h^*(y) - g^*(y) \tag{13}$$

Combining (11), (12), and (13) we get

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

Property (ii) of the theorem establishes sufficient condition for local d.c. optimality.

Corollary 4.4:Let x* be a point that admits a neighborhood U such that $\partial h(x) \cap \partial g(x^*) \neq \phi \, \forall x \in U \cap \operatorname{dom} g$. Then x* is a local minimizer of g - h. Moreover, $g(x) - h(x) \geq g(x^*) - h(x^*) \, \forall x \in U \cap \operatorname{dom} g$.

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 $g(x^*) + g^*(y) = \langle x^*, y \rangle \leq h(x^*) + h^*(y)$.

So $h^*(y) - g^*(y) \geq g(x^*) - h(x^*)$. 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 suppositions of Theorem 4.3(ii) are satisfied. Thus, the result holds.

4.2.2: D.C.duality transportation of a local minimizer.

It may happen that it is easier to solve locally the dual d.c.program (Q) than the primal d.c.program (P). So it is useful to state results relative to the d.c. duality transportation of a local minimizers.

Corollary 4.5:Let $x^* \in \text{dom}\,\partial h$ be a local minimizer of g - h, and $y^* \in \partial h(x^*)$. If

$$y^* \in \text{int}(\text{dom}\, g^*) \text{ and } \partial g^*(y^*) \subset U \tag{14}$$

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

Proof: According to (i) of Theorem 4.3. we have $y^* \in \partial h(x^*) \subset \partial g(x^*)$.

So $x^* \in \partial g^*(y^*) \cap \partial h^*(y^*)$. Under the assumption (14) and the upper semi continuity of ∂g^*, y* admits a neighborhood $V \subset (U \cap \text{dom}\, g^*)$ such that ([6])

$\partial g^*(V) \subset U$. More precisely, $\partial g^*(V) \subset (U \cap \text{dom} g)$ since we have [6]

range $\partial g^* = \text{dom}\,\partial g$ and $\text{dom}\,\partial g \subset \text{dom}\, g$. Using the dual property (in the d.c. duality) in (ii) of Theorem 4.3. 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 (14) [6].

By the symmetry of the d.c.duality, Corollary 4.5. has its corresponding dual part.

4.2.3: Main results of local duality in DCO.

The Main results of local duality in DCO shall be shown below. These results constitute the basis of our subgradient method for solving DCO problems.

Theorem 4.6.

(i) $x^* \in \Omega^*$ if and only if there exists $y^* \in W(x^*)$ such that $x^* \in \partial g^*(y^*)$;

i.e., $x^* \in (\partial g^* o W)(x^*)$

(ii) $x^* \in \Delta^*$ if and only if there exists $x^* \in N(x^*)$ such that $y^* \in h(x^*)$;

i.e., $y^* \in (\partial h o N)(y^*)$.

Proof: By duality we need to show (i).

If $x^* \in \Omega^*$ then $\langle x^*, y\rangle - g^*(y) = g(x^*)\ \forall y \in \partial h(x^*)$ because $\partial h(x^*) \subset \partial g(x^*)$.

Now $y^* \in W(x^*)$ is equivalent to

$$\begin{aligned}\langle x^*, y^*\rangle - g^*(y^*) &= \inf_{y\in\partial h(x^*)} \{\langle x^*, y\rangle - g^*(y)\} \\ &= g(x^*)\ \forall y \in \partial h(x^*) \\ &= \langle x^*, y\rangle - g^*(y)\ \forall y \in \partial h(x^*)\end{aligned}$$

Hence $y^* \in W(x^*)$ is equivalent to $y^* = y$ for all y in $\partial h(x^*)$.

Therefore, $W(x^*) = \partial h(x^*)$ and hence $x^* \in \partial g^*(y^*)$ for some $y^* \in W(x^*)$.

Conversely let $x^* \in X$ such that $x^* \in \partial g^*(y^*)$ for some $y^* \in W(x^*)$. Because $y^* \in W(x^*)$, we have

$$-g^*(y) + g^*(y^*) - \langle x^*, y^*\rangle \geq -\langle x^*, y\rangle\ \forall y \in \partial h(x^*)$$

But $x^* \in \partial g^*(y^*)$, then $g(x^*) = \langle x^*, y^*\rangle - g^*(y^*)$.

It follows that

$$-g^*(y) - g^*(x^*) \geq -\langle x^*, y\rangle\ \forall y \in \partial h(x^*)$$

By using the definition of g*,

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

So, it follows that

$$-\sup_{x\in X}\{\langle x, y\rangle - g(x)\} - g(x^*) \geq -\langle x^*, y\rangle\ \forall x \in \partial h(x^*)$$

$$\inf_{x\in X}\{g(x) - \langle x, y\rangle\} - g(x^*) \geq -\langle x^*, y\rangle\ \forall x \in \partial h(x^*)$$

$$g(x) - \langle x, y\rangle - g(x^*) \geq -\langle x^*, y\rangle\ \forall x \in \partial h(x^*)$$

since $g(x) - \langle x, y\rangle \geq \inf_{x\in X}\{g(x) - \langle x, y\rangle\}$

Thus we obtain the following inequality:

$$g(x) - g(x^*) \geq \langle x - x^*, y\rangle, \forall x \in X, \forall y \in \partial h(x^*)$$

which follows that

$$y \in \partial g(x^*) \forall y \in \partial h(x^*)$$

and hence $\partial h(x^*) \subseteq \partial g(x^*)$ and hence $x^* \in \Omega^*$.

So the proof is achieved.

From the observation in the proof we can deduce the following:

(1) If $x^* \in \Omega^*$, then we have

(i) $W(x^*) = \partial h(x^*)$

(ii) $h^*(y) - g^*(y) = g(x^*) - h(x^*), \ \forall y \in \partial h(x^*)$

(2) If $y^* \in \Delta^*$, then we get

(iii) $N(y^*) = \partial g^*(y^*)$

(iv) $g(x) - h(x) = h^*(y^*) - g^*(y^*), \ \forall x \in \partial g^*(y^*)$.

In the description of the subgradient methods, if the initial vector x^0 is close enough to an element Ω, we shall see the role of the problems $\omega(x^*)$ and $\eta(y^*)$, and also a possible justification of the convergence of these methods towards a couple (x^*, y^*) of optimal solutions (P) and (Q) respectively.

4.3.Subgradient -duality method for solving d.c.optimization

Here we shall see the role of the problems $\omega(x^*)$ and $\eta(y^*)$, and also a possible justification of the convergence of these methods towards a couple (x^*, y^*) of optimal solutions of (P) and (Q) respectively when the initial vector x^0 is close enough to an element of Ω.

Recall the following facts from convex analysis. (Rockafellar [6])

Lemma 4.7.

Let $h \in \Gamma_o(X)$ and let x^k be a sequence, which possesses the following properties:

(1) $x^k \to x^*$

(2) There exists a bounded sequence $y^k \in \partial h(x^k)$

(3) $\partial h(x^*)$ is non empty

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

Proof: In fact let $y^* \in \partial h(x^*)$, we have

$$h(x^k) \geq h(x^*) + \langle x^k - x^*, y^* \rangle$$

But $y^k \in \partial h(x^k)$, then

$$h(x^*) \geq h(x^k) + \langle x^* - x^k, y^k \rangle$$

This is equivalent to

$$h(x^k) \leq h(x^*) + \langle x^k - x^*, y^k \rangle$$

As $x^k \to x^*$, we have $\lim_{k\to\infty} \langle x^k - x^*, y^* \rangle = 0$ and $\lim_{k\to\infty} \langle x^k - x^*, y^k \rangle = 0$ because the sequence y^k is bounded. It follows that $\lim_{k\to\infty} h(x^k) = h(x^*)$.

Recall from the algorithm considered on page 36 of this chapter, $x^{k+1} \in N(y^k)$ and $y^k \in W(x^k)$.

Theorem 4.8.

(i) $g(x^{k+1}) - h(x^{k+1}) \leq h^*(x^k) - g^*(y^k) \leq g(x^k) - h(x^k)$

(ii) $g(x^{k+1}) - h(x^{k+1}) = g(x^k) - h(x^k)$ if and only if $x^k \in \partial g^*(y^k)$ and $y^k \in \partial h(x^{k+1})$

In this case we have $x^k \in \Omega^*$ and $y^k \in \Delta^*$.

Proof: (i) Because $x^{k+1} \in \partial g^*(y^k)$, we have

$$g(x^{k+1}) - \langle x^{k+1} - x^k, y^k \rangle \leq g(x^k)$$

which implies

$$g(x^{k+1}) - \langle x^{k+1} - x^k, y^k \rangle - h(x^k) \leq g(x^k) - h(x^k) \quad (15)$$

But $y^k \in \partial h(x^k)$ as $y^k \in W(x^k)$, then

$$h(x^k) + \langle x^{k+1} - x^k, y^k \rangle \leq h(x^{k+1})$$

which is equivalent to

$$\begin{aligned} g(x^{k+1}) - h(x^{k+1}) &\leq g(x^{k+1}) - \langle x^{k+1} - x^k, y^k \rangle - h(x^k) \\ &\leq g(x^k) - h(x^k) \qquad \text{(by (15))} \end{aligned}$$

and

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

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

(ii) Necessity

Let $g(x^{k+1}) - h(x^{k+1}) = g(x^k) - h(x^k)$ (17)

From (i) and because $y^k \in \partial h(x^k)$, we have

$$g(x^{k+1}) - h(x^{k+1}) - (g(x^k) - h(x^k)) \le g(x^{k+1}) - g(x^k) - \langle x^{k+1} - x^k, y^k \rangle \le 0 \quad (18)$$

But (17) implies

$$g(x^{k+1}) - h(x^{k+1}) - (g(x^k) - h(x^k)) = 0$$

and hence (18) becomes

$$0 \le g(x^{k+1}) - g(x^k) - \langle x^{k+1} - x^k, y^k \rangle \le 0$$

which follows that

$$g(x^{k+1}) - g(x^k) = \langle x^{k+1} - x^k, y^k \rangle \quad (19)$$

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

$$g(x^{k+1}) + g^*(y^k) = \langle x^{k+1}, y^k \rangle \quad (20)$$

Subtracting (19) from (20), we have

$$g(x^k) + g^*(y^k) = \langle x^k, y^k \rangle$$

In others words $x^k \in \partial g^*(y^k)$. This implies, by Theorem 4.6, that $x^k \in \Omega^*$

because $y^k \in W(x^k)$ by construction.

It remains to prove that $y^k \in \Delta^*$.

Using (17) and (19), we get

$$\langle x^k, y^k \rangle - h(x^k) = \langle x^{k+1}, y^k \rangle - h(x^{k+1})$$

which implies that $x^{k+1} \in N(y^k)$ and $y^k \in \partial h(x^{k+1})$. Then, Theorem 4.6. gives $y^k \in \Delta^*$ because $y^k \in \partial h(x^k)$ by construction.

Sufficiency

Suppose that $x^k \in \partial g^*(y^k)$ and $y^k \in \partial h(x^{k+1})$. Then we have

$$g(x^k) + g^*(y^k) = \langle x^k, y^k \rangle$$

and

$$h(x^{k+1}) + h^*(y^K) = \langle x^{k+1}, y^k \rangle$$

But by the construction $x^{k+1} \in \partial g^*(y^k)$ and $y^k \in \partial h(x^k)$, then

$$g(x^{k+1}) + g^*(y^k) = \langle x^{k+1}, y^k \rangle$$

and

$$h(x^k) + h^*(y^K) = \langle x^k, y^k \rangle$$

From these relations, it can easily be concluded that

$$g(x^{k+1}) - h(x^{k+1}) = g(x^k) - h(x^k).$$

So the proof (i) and (ii) are achieved.

Theorem 4.9:

If λ is finite, then we have

(i) $\lim_{k\to\infty} \{g(x^k) - h(x^k)\} = \lim_{k\to\infty} \{h^*(y^k) - g^*(y^k)\} = \mu \geq \lambda$

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

Proof: Because λ is finite, property (i) is a simple consequence of Theorem 4.8(i) and Squeezing Theorem for Limits. Taking the limit to be μ; $\mu \geq \lambda$ as λ is the infimum value.

Using (16) and (i), we get

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

which implies

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

That is,

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

Theorem 4.10: If λ is finite and if the sequences x^k and y^k are bounded, then for every cluster point 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:

(i) $x^* \in \Omega^*$ and $g(x^*) - h(x^*) = \mu$

(ii) $y^* \in \Delta^*$ and $h^*(y^*) - g^*(y^*) = \mu$

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

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

Proof: Suppose now that λ is finite and that the sequences x^k and y^k are bounded. The set of cluster points of x^k and of y^k are non-empty because x^k and y^k are supposed to be bounded.

Let x* be a cluster point of x^k; for the sake of simplicity in notations we shall write:

$$\lim_{k\to\infty} x^k = x^*$$

We can suppose (by extracting a subsequence if necessary) that the sequence y^k converges to a point $y^* \in \partial h(x^*)$. Property (ii) of Theorem 4.9 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 now $\theta(x, y) = g(x) + g^*(y)$ for $(x,y) \in X \times X^*$. It is clear that $\theta \in \Gamma_0(X \times X^*)$.

Then, because of the lower semi continuity of θ we obtain

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

That is,

$$\theta(x^*, y^*) = g(x^*) + g^*(y^*) = \langle x^*, y^* \rangle.$$

So, property (iii) is proved and $y^* \in \partial g(x^*)$.

According to Lemma 4.7. we have $\lim_{k\to\infty} h(x^k) = h(x^*)$ 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 (ii) of Theorem 4.9. we have:

$$\lim_{k\to\infty} \{g(x^k) - h(x^k)\} = \lim_{k\to\infty} g(x^k) - \lim_{k\to\infty} h(x^k)\} = \lim_{k\to\infty} g(x^k) - h(x^*) = \mu$$

$$\lim_{k\to\infty}\{h^*(y^k)-g^*(y^k)\}=\lim_{k\to\infty}h^*(y^k)-\lim_{k\to\infty}g^*(y^k)=h^*(y^*)-\lim_{k\to\infty}g^*(y^k)=\mu$$

It then suffices to show that

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

Since $\lim_{k\to\infty}g(x^k)$ and $\lim_{k\to\infty}g^*(y^k)$ exists, property (ii) of Theorem 4.9. implies

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

Further, because of the lower semi continuity of g and g*,

$$\lim_{k\to\infty}g(x^k)=\liminf_{k\to\infty}g(x^k)\geq g(x^*),$$

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

The former equalities imply that these last inequalities are in fact equalities. The proof of Theorem 4.10 is complete.

4.4.The Simplified Primal-dual algorithm in solving a DC program

The simplified DCA constructs two sequences $\{x^k\}$ and $\{y^k\}$ (candidate to primal and dual solutions) that 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* (respectively y*) of the sequence $\{x^k\}$ (respectively $\{y^k\}$) is the critical point of g-h (respectively h*-g*).

These conditions suggests constructing two sequences $\{x^k\}$ and $\{y^k\}$, starting from a given point $x^0\in \text{dom } g$, by setting

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

The simplified DCA is interpreted as follows: At each iteration k we do the following

$$x^k\in\partial g^*(y^{k-1}):\to y^k\in\partial h(x^k)$$

$$=\text{argmin}\ \{h^*(y)-[g^*(y^{k-1})+\langle x^k, y-y^{k-1}\rangle]: y\in X^*\} \qquad (D_k)$$

$$y^k\in\partial h(x^k)\ :\to x^{k+1}\in\partial g^*(y^k)$$

$$=\text{argmin}\ \{g(x)-[h(x^k)+\langle x-x^k, y^k\rangle]: x\in X\} \qquad (P_k)$$

Problem (P_k) are a convex program that is obtained from (P) by using the affine minorization of h defined by $y^k \in \partial h(x^{k-1})$. Similarly, the convex program (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 a 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 complete and the simplified form of the DCA are identical if g* and h are essentially differentiable. The only difference between the simplified DCA and the complete DCA lies on the choice of y^k in $\partial h(x^k)$ and x^{k+1} in $\partial g^*(y^k)$.

The complete DCA theoretically provides an x* such that $\partial h(x^*) \subset \partial g(x^*)$. In practice, except for the cases where the convex maximization problems ($\omega(x^k)$ and $\eta(y^k)$) are easy to solve, one generally uses the simplified DCA. It is worth nothing 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 restating it from a new initial point $x^0 = x^*$ with $y^0 \in \partial h(x^0)$ such that $y^0 \notin \partial g(x^0)$. In fact, since

$$g(x^1) + g^*(y^0) = \langle x^1, y^0 \rangle \le h(x^1) - h(x^0) + \langle x^0, y^0 \rangle$$

and $\langle x^0, y^0 \rangle < g(x^0) + g^*(y^0)$ because $y^0 \notin \partial g(x^0)$, we have

$$g(x^1) + g^*(y^0) < h(x^1) - h(x^0) + g(x^0) + g^*(y^0).$$

Hence,

$$g(x^1) - h(x^1) < g(x^0) - h(x^0).$$

We have given the main concept of the d.c.programming and the DCA.

4.5. Remarks On DCA

1) A d.c. function f has infinitely many d.c. decompositions.

For example, if $f = g - h$, then $f = (g + \theta) - (h + \theta)$ for $\theta \in \Gamma_0(X)$ 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 the DCA relative to these d.c. decompositions of the objective function f. It is useful to find a suitable d.c. decomposition of f since it may have an important influence on the efficiency of the DCA for its solution.

2) If the sequences $\{x^k\}$ and $\{y^k\}$ are constructed in a simpler following manner which is called simplified DCA:

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

then we should replace in Theorem 3.8 (respectively Theorem 3.9) $x^* \in \Omega^*$ (respectively $y^* \in \Delta^*$) by $\partial g(x^*) \cap \partial h(x^*) \neq \phi$ (respectively $\partial g^*(y^*) \cap \partial h^*(y^*) \neq \phi$)

This is the fact that $y^k \in W(x^k)$ (respectively $x^{k+1} \in N(y^k)$ which allows to prove $x^* \in \Omega^*$ (respectively $y^* \in \Delta^*$) in Theorem 3.8 (respectively Theorem 3.9).

3) If g is strongly convex (i.e. there exists $\rho_1 > 0$ such that

$$g(x_2) \geq g(x_1) + \langle x_2 - x_1, y_1 \rangle + \rho_1 \|x_2 - x_1\|^2$$

for every x_1 and x_2 in X and every $y_1 \in \partial g(x_1)$), then we have

$$g(x^{k+1}) - h(x^{k+1}) \leq g(x^k) - h(x^k) - \rho_1 \|x^{k+1} - x^k\|^2$$

This result (which is also true for sequence x^k and y^k constructed as above in (2) is immediate from the proof of Theorem 4.8.

In the same way if h is strongly convex with constant $\rho_2 > 0$, then

$$g(x^{k+1}) - h(x^{k+1}) \leq g(x^k) - h(x^k) - \rho_2 \|x^{k+1} - x^k\|^2$$

Evidently if both the functions g and h are strongly convex then

$$g(x^{k+1}) - h(x^{k+1}) \leq g(x^k) - h(x^k) - (\rho_1 + \rho_2) \|x^{k+1} - x^k\|^2$$

In the above statements equality holds ($g(x^{k+1}) - h(x^{k+1}) = g(x^k) - h(x^k)$) if and only if $x^{k+1} = x^k$.

4) If g or h is strongly convex then $\lim_{k\to\infty}\|x^{k+1} - x^k\| = 0$. Moreover if the set of cluster points of x^k is finite then the whole sequence x^k converges.

5) With out loss of generality, we can assume that f and g are strongly convex. In fact we can write:

$$g(x) - h(x) = (g + \varphi)(x) - (h + \varphi)(x)$$
$$= g_1(x) - g_2(x)$$

and then it suffices to take φ strongly convex (e.g. $\varphi(x) = \frac{1}{2}\|x\|^2$).

This procedure could render the differentiability of the functions g* and h* in the dual problem (Q).

Parallely we can apply to primal problem (P) the following regularization procedure:

$$(P') \qquad \lambda = \inf_{x\in X}\{(g\,_\nabla\,\theta)(x) - (h\,_\nabla\,\theta)(x)\}$$

(where $\theta \in \Gamma_0(X)$ and ∇ being the infimal convolution operator) whose dual problem is:

$$(Q') \qquad \lambda = \inf_{y\in X^*}\{(h^* + \theta^*)(y) - (g^* + \theta^*)(y)\}$$

It is clear that problems (Q) and (Q′) are equivalent. This is not the case for (P) and (P′). (see Theorem 4.2). However the regularization performed in (P′) can simplify the programmation of our algorithm.

6) Remark that we can regularize both the primal and the dual problems:

we denote by $P(\varphi)$ the following problem: ($\varphi \in \Gamma_0(X)$)

$$P(\varphi) \qquad \lambda = \inf_{x\in X}\{(g\,_\nabla\,\varphi)(x) - (h\,_\nabla\,\varphi)(x)\}$$

which is equivalent to (P). By regularizing $P(\varphi)$ as above-mentioned, we obtain

$$P(\varphi)' \qquad \lambda = \inf_{x\in X}\{((g + \varphi)_\nabla\,\theta)(x) - ((h + \varphi)_\nabla\,\theta)(x)\}$$

whose dual problem is:

$$\lambda = \inf_{y \in X^*} \{((h+\varphi)^* + \theta^*)(y) - ((g+\varphi)^* + \theta^*)(y)\}$$

The latter is equivalent to the dual problem $Q(\varphi)$ of $P(\varphi)$.

$$Q(\varphi) \qquad \lambda = \inf_{y \in X^*} \{((h+\varphi)^*)(y) - ((g+\varphi)^*)(y)\}$$

which can take under some well-known condition, the following form:

$$\lambda = \inf_{y \in X^*} \{((h *_{\nabla} \varphi^*)(y) - ((g *_{\nabla} \varphi^*)(y)\}$$

From a practical viewpoint, we should be prudent in evaluating the complexity and the performance of our algorithm with these regularizations. The interesting problem of finding a better (or optimal) decomposition of f-g remains open.

5. Reference

[1]. **Deumlich, R.:** Optimization and Theory of approximation Textbook for the lecture, Addis Ababa 1997.

[2]. **Tuy, H.:** Convex Analysis and Global Optimization, Kluwer Academic publishers, Dordrecht, 1998.

[3]. **Konno, H.; Thach, P.T., and Tuy, H.:** Optimization on low rank non-convex structures, Klumer Academic publishers, Dordrecht/Boston/London, 1997.

[4]. **Pallaschke, D. and Rolewicz,S.:** Foundations of Mathematical – Optimization – Convex Analysis without Linearity, Kluwer Academic publishers, 1997.

[5].**Semu, M.K:** On Minimal pairs of Compact Convex Sets and Of Convex Functions, 2002.

[6]. **R.T.Rockafellar:** Convex Analysis, Princeton University Press; Princeston, 1972.

[7]. **Pham Dinh Tao and El Bernoussi Souad (1988):** Duality in D.C. (Difference of Convex functions) Optimization. Subgradient Methods. International Series of Numerical Mathematics, 84, Birkh ä user, Basel, 1988, pp.276 – 294.

[8]. **Pham Dinh Tao and Le Thi Hoai An:** A D.C. Optimization Algorithm for Solving the Trust-Region Sub problem. Society for Industrial and Applied Mathematics, 8 N º 2, 1998, pp.476-505

[9]. **Philip Hartman:** On Functions Representable As A Difference of Convex Functions pp.707-713

[10]. **Thai Quynh Phong:** An algorithm for solving general D.C. programming Problems.
Operation Research Letters 15 (1994) pp.73-79

[11]. **J.F.Toland:** Duality in Non-convex Optimization, Journal of Mathematical Analysis and Applications, 66(1978), pp.399-415

[12]. **John F.Toland:** On Subdifferential Calculus and Duality in Non-convex Optimization, Analyse non-convexe [1977.pau] Bull.Soc.Math.France.M e' moiré 60, 1979, pp.177-183.

[13]. **Hoang Tuy:** D.C.OPTIMIZATION; Theory, Methods and Algorithms, Institute of Mathematics, P.O.Box 631, Bo Ho, Hanoi, Vietnam.