GRADUATE SEMINAR REPORT
ON

ABSTRACT DUALITY FOR PRACTITIONERS

By
ZELALEM TESHOME WALE

Advisor
Prof. Dr. rer. nat. habil. R. Deumlich

School of Graduate Studies
Addis Ababa University

June,1997
Addis Ababa

ACKNOWLEDGEMENT

I would like to thank my instructor and advisor Prof.Dr. R. Deumlich for his several valuable and helpful suggestions during the preparation of this Graduate seminar report. I also grateful to my colleagues who helped me in typing this Graduate seminar report.I would like to give special mention to Ato Geta Techanie,who was always willing to type this seminar report . Ato wondimagegnehu Geremew in drawing the graphs. Ato Abate Tibebu for his kind cooperation and Ato Mehari G/Eghizhabher in printing this report .

CONTENT

1. THE PROBLEM AND THE GENERAL APPROACH

We consider a constrained optimization problem characterized by a non empty set U of admissable control variables u, an objective function φ :U$\rightarrow$R and constraint functions $C_{1,,}, C_2, \ldots C_m$:U$\rightarrow$R; then we want to solve the optimization problem.

$$\varphi(u) \rightarrow \sup,\ u \in S$$
$$S := \{u \in U: C_j(u) = 0, \text{ for } j = 1,\ldots,p\} \qquad (P)$$

We will call (P) the primal problem and u$\in$ S is called ***feasible***.

Denotation : $C(u):(C_1\,(u), C_2(u), \ldots, C_m(u)) \in R^m$ is the vector of constraint values at u$\in$U.

We define the Lagrangian or Lagrange function by

$$L(u,\lambda) := \varphi(u) - \lambda^T C(u) \qquad (u,\lambda)\in U\times R^m \qquad (1)$$

Remark:: The minus sign in (1) comes from the fact that we start from amaximizationproblem.

Definition: The Lagrange problem associated with $\lambda \in R^m$ fixed is the optimization problem in U

$$L(u,\lambda) \rightarrow \sup,\ u \in U \qquad (P_\lambda)$$

Example: If U is a compact set on which L(.,λ)is upper semicontinuous,then there is a $u_\lambda \in U$ such that

$$L(u_\lambda,\lambda) = \sup\{L(u,\lambda): u \in U\}$$

that is u_λ is a solution of (P_λ).
A particular and important case is when U is finite.

Solution: Let f:=L(., λ) be upper semicontius.Then -f = -L(., λ) is lower semicontinuous.This implies M(-f,U)$\neq\varnothing$ .That is there exists u_λ such that
$-f(u_\lambda) = \min\{-f(u) : u \in U\} = -\max\{f(u): u \in U\}$.
Thus $f(u_\lambda) = \max\{f(u): u \in U\}$
Therefore $L(u_\lambda,\lambda) = \sup\{L(u,\lambda): u \in U\}$
When U is a finite set $\max\{L(u, \lambda): u \in U\}$ exists.

Examples:

1. THE KNAPSACK PROBLEM

Our first example is when U is a finite set or a countable set. Suppose one has a knapsack and one considers putting in it n objects (tooth brush,book, etc,...) each of which has a price p^i and a volume V^i. Then one wants to make the knapsack of maximal price, knowing that it has a limited volume V.

For each object two decisions are possible: take it or not ;Uis the set of all possible such decisions for the n objects,that is U is identified with the set $\{1,2,...,2^n\}$.To each $u \in U$, are associated the objective and constraint values, namely the sum of respectively the prices and volumes of all objects taken.The problem is solvable in a finite time, but quite a large time if n is large.

Let us now consider the Lagrange problem : $\lambda \in R$ and the number λ is a penality coefficient or the price of space in the knapsack .The Lagrange function for each object i is $P^i - \lambda V^i$:

If $P^i - \lambda V^i > 0$; take the object.

If $P^i - \lambda V^i < 0$; leave it

If $P^i - \lambda V^i = 0$; do what you want.

Preliminary observation

-From its interpretation λ should be non negative : there is no point in giving a bonus to bulky objects like TV sets .

-There always exists an optimal decision in the Lagrange problem (P_λ) and it is well defined (unique) except when λ is one, of the n values P^i/V^i.

The knapsack problem can be given an analytical flavour: assign to the control variable u^i the value 1 if the i^{th} object is taken, 0 if not. Then we have to solve:

$$\max \sum_{i=1}^{n} P^i u^i \quad \text{subject to } u^i \in \{0,1\}, \quad \sum_{i=1}^{n} v^i u^i \le V$$

In order to fit with the equality -constrained form (P) , a nonnegative slack variable u^o is inserted and we obtain

$$\max \varphi(u) \text{ where } \varphi(u) := \sum_{i=1}^{n} P^i u^i$$

$$C(u) := \sum_{i=1}^{n} v^i u^i + u^o - V = 0 \qquad (2)$$

$$u^o \ge 0,\ u^i \in \{0,1\}, \qquad \text{for } i = 1,....,n$$

which certainly has a solution.

Associated with $\lambda \in R$, the associated Lagrange problem is then

$$\sup\{L(u,\lambda): u \in U\} = \sup\{\varphi(u) - \lambda^T C(u): u \in U\}$$
$$= \sup\left\{\sum_{i=1}^{n} P^i u^i - \lambda\left(\sum_{i=1}^{n} v^i u^i + u^o - V\right) : u^o \ge 0, u^i \in \{0,1\}, i = 1, ..., n\right\}$$
$$= \sup\left\{\lambda V - \lambda u^0 + \sum_{i=1}^{n}\left(P^i - \lambda v^i\right) u^i u^o \ge 0, u^i \in \{0,1\}, i = 1, ..., n\right\}$$

If $\lambda < 0$, then $\sup L(., \lambda) = \infty$. Thus for any $\lambda \ge 0$ and $u^o = 0$, we will have

$$L(u_\lambda, \lambda) = \sum_{i \in I(\lambda)} \left(P^i - \lambda v^i\right), \qquad C(u_\lambda) = \sum_{i \in I(\lambda)} v^i - V$$

where $I(\lambda) = \{i: P^i - \lambda v^i \ge 0\}$

2. ENTROPY MAXIMIZATION

The set of control variables is now an infinite dimensional space.Say $U=L_1(\Omega,R)$ where Ω is an interval of R. We want to solve the optimization problem

$$\int_\Omega \psi(x, u(x))dx \rightarrow \sup,\ u \in S \tag{3}$$

$$S=\{u \in U: \int_\Omega \gamma_j(x, u(x))dx = 0, j = 1, \ldots, m\}$$

The (m+1) functions $\psi,\gamma_1,\ldots,\gamma_m$ sent $(x,t)\in\Omega\times R$ to R. The Lagrangian is the integral over Ω of the function

$$l(x, u, \lambda) := \psi(x, u(x)) - \sum_{j=1}^{m} \lambda_j\gamma_j(x, u(x))$$

Indeed without going into technical details from functional analysis, suppose that the maximization of l with respect to $u\in L_1(\Omega,R)$ makes sense, and that the standard optimality conditons hold: given $\lambda\in R^m$ we must solve for each $x\in\Omega$ the equation in $t\in R$

$$\frac{\partial l}{\partial u}(x, t, \lambda) = \frac{\partial\psi}{\partial u}(x, t) - \sum_{j=1}^{m} \lambda_j \frac{\partial\gamma_j}{\partial u}(x, t) = 0 \tag{4}$$

Take a solution giving the largest l (if there are several), call $u_\lambda(x)$ the result.

The favourable cases are those where $t \rightarrow \psi(x,t)$ is strictly concave and each function $t \rightarrow \gamma_j(x,t)$ is affine. A typical example in one in which ψ is an entropy (usually not depending on x), for example

$$\psi(x, t) = \begin{cases} -t\log t & t > 0 \\ 0 & t = 0 \\ -\infty & t < 0 \end{cases} \tag{5}$$

or

$$\psi(x, t) = \begin{cases} \log t & t > 0 \\ -\infty & t \le 0 \end{cases} \tag{6}$$

Being affine, the constraints have the general form

$$\int_\Omega a_j(x)u(x)dx - b_j = o \qquad \text{where } a_j\in L_\infty(\Omega,R)$$

Incase of (5)

$$l(x, t, \lambda) = -t\log t - \sum_{j=1}^{m} \lambda_j(a_j(x)u(x) - b_j)$$

$$\frac{\partial l}{\partial u}(x, t, \lambda) = -\log t - 1 - \sum_{j=1}^{m} \lambda_j(a_j(x)$$

$$\frac{\partial l}{\partial u}(x, t, \lambda) = 0 \Rightarrow \log t = -1 - \sum_{j=1}^{m} \lambda_j(a_j(x)$$

Hence

$$t = \exp\left(-1 - \sum_{j=1}^{m} \lambda_j(a_j(x)\right)$$

Therefore,

$$u_\lambda(x) = \exp\left(-1 - \sum_{j=1}^{m} \lambda_j(a_j(x)\right) \tag{7}$$

Incase of (6)

The optimalty conditions (4) for u_λ give

$$u_\lambda(x) = \frac{1}{\sum_{j=1}^{m} \lambda_j(a_j(x)} \tag{8}$$

2. THE NECESSARY THEORY

Throughout this section , our leading motivation is as follows : we want to solve the u problem (P). Our first aim is to formulate a more explicit λ problem here after called the *dual problem* , which will turn out to be very well posed.Then we will examine questions regarding its solvability and its relevance: to what extent does the dual problem really solve the primal u problemthat we stated from?

The data (U,φ,C) will still be viewed as totally unstructured , up to the point where we need to require some specfic properties from them.

2.1 Preliminary results : *The dual problem*

To slove our problem , we must find a feasible u in (P) (that is $u \in S$). This turns out to be also sufficient , and the reason lies in what is probably the practical link between (P) and (P_λ)

Theorem 1: (H Everett) Fix $\lambda \in R^m$, suppose that (P_λ) has an optimal solution $u_\lambda \in U$ and set $c_\lambda := C(u_\lambda)$. Then u_λ is also an optimal solution of

$$\varphi(u) \to \max,\ u \in C \tag{2.1}$$

$$C_0 = \{\ u \in U:\ c(u) = c_\lambda \in R^m\}$$

Proof: Take an arbitrary $u \in U$. By the definition of u_λ

$$L(u_\lambda,\lambda) \geq L(u,\lambda) \qquad \text{for all } u \in U.$$

That is,

$$\varphi(u_\lambda) - \lambda^T C(u_\lambda) = L(u_\lambda,\lambda) \geq L(u,\lambda) = \varphi(u) - \lambda^T C(u)$$

If in addition ,

$$C(u) = c_\lambda,$$

then u is feasible in (2.1)and hence

$$\varphi(u) - \lambda^T C(u) = \varphi(u) - \lambda^T c_\lambda \leq \varphi(u_\lambda) - \lambda^T C(u_\lambda)$$

From this follows

$$\varphi(u) \leq \varphi(u_\lambda)$$

Hence u_λ is an optimal solution of (2.1).

Corrolary 2: If for some $\lambda \in R^m$ (P_λ) happens to have a solution u_λ which is feasible in (P), then u_λ is a solution of (P).

Proof: Put $c_\lambda = 0$. Then by the Everett's theorem u_λ is also an optimal solution of

$$\varphi(u) \to \max, u \in S$$
$$S := \{u \in U: C(u) = 0\}$$

Hence u_λ is a solution of (P).■

If u_λ solves (P), then we have to solve the system of equations

$$C_j(u_\lambda) = 0 \qquad j = 1,\ldots,m. \tag{2.2}$$

In case of Corollary 2 it suffices to solve this system.
We consider a problem which helps a great deal
Find λ such that there is $u \in U$ maximizing $L(., \lambda)$ and satisfying $C(u) = 0$ (2.3)

Definition 3. (Dual function) In the Lagrange problem (P_λ), the optimal value is called the dual funcion, denoted by θ,

$$\theta(\lambda) := \sup\{L(u,\lambda) : u \in U\} \qquad \lambda \in R^m \tag{2.4}$$

If there is u_λ such that

$$L(u_\lambda,\lambda) \geq L(u,\lambda) \qquad \text{for all } u \in U,$$

then $\theta(\lambda) = L(u_\lambda,\lambda)$

Theorem 4 .(*Weak duality theorem*) For all $\lambda \in R^m$ and all u feasible in (P) there holds

$$\theta(\lambda) \geq \varphi(u) \tag{2.5}$$

proof: For any u feasible in (P), we have $C(u)=0$. Thus

$$\varphi(u) = \varphi(u) - 0 = \varphi(u) - \lambda^T C(u) = L(u,\lambda) \leq \theta(\lambda)$$

by definition 3. Therefore.

$$\theta(\lambda) \geq \varphi(u)$$

Thus from the above we conclude that each value of the dual function gives an upper bound on the primal optimal value; a sensible idea is then to find the best such upper bound.

Definition 5: (*Dual problem*) The optimization problem in λ

$$\theta(\lambda) \to \inf, \lambda \in R^m \tag{2.6}$$

is called the dual problem associated with (U, φ, C) of (P).

Theorem 6: If μ is such that the associated Lagrange problem $(P\mu)$ is maximized at a feasible u_o, that is μ solves (2.3) , then μ is also a solution of the dual problem (2.6)
Conversely , if μ solves (2.6) and if (2.3) has a solution at all , then μ is such a solution .

Proof: The property $C(u_o) = 0$, means $\theta(\mu) = L(u_o,\mu) = \varphi(u_o) - \mu^T C(u_o)$ but $\varphi(u_o) \leq \theta(\lambda)$ for all $\lambda \in R^m$ by the weak duality theorem. Hence

$$\theta(\mu) = \varphi(u_o) \leq \theta(\lambda) \text{ for all } \lambda \in R^m$$

That means

$$\theta(\mu) = \inf\{\theta(\lambda) : \lambda \in R^m\}$$

Thus μ is a solution of the dual problem(2.6).

Conversely (2.3) has a solution λ^* : some $u^* \in U$ satisfies $C(u^*) = 0$ and

$$L(u^*,\lambda^*) = \sup\{L(u,\lambda^*) : u \in U\} = \theta(\lambda^*)$$

but

$$L(u^*,\lambda^*) = \varphi(u^*) - (\lambda^*)^T C(u^*) = \varphi(u^*) - 0 = \varphi(u^*).$$

Thus $\theta(\lambda^*) = \varphi(u^*) \leq \theta(\lambda)$ for all $\lambda \in R^m$
Hence $\theta(\lambda^*)$ has to be the minimal value $\theta(\mu)$ of the dual function and we write.

$$L(u^*, \mu) = \varphi(u^*) - (\lambda^*)^T C(u^*) = \varphi(u^*) = \theta(\mu)$$

Therefore μ is a solution of (2.3), since u^* maximizes $L(.,\mu)$ and $C(u^*) = 0$.■

Existence of a solution to (2.3) means full success of the dual approach in the following sense:

Corrolary 7: Suppose (2.3) has a solution.Then for any solution λ of the dual problem (2.6), the primal solutions are those u maximizing $L(., \lambda)$ that are feasible in (P).

Proof: First we claim that when (2.3) has a solution.
λ minimizes θ iff λ solves (2.3), and there is some $u \in U$ such that $\varphi(u) = L(u,\lambda) = \theta(\lambda)$.
Proof of the claim :
$\Rightarrow$: λ minimizes θ means $\theta(\lambda) = \inf\{\theta(\mu): \mu \in R^m\}$
by theorem 6 (converse part) λ is a solution of (2.3)
By the definition of λ solves (2.3), there is some $u \in U$ such that $C(u) = 0$ and
$L(u,\lambda) = \sup\{L(u^*,\lambda): u^* \in U\}$. This implies $\theta(\lambda) = \sup\{L(u^*,\lambda): u^* \in U\} = L(u,\lambda) = \varphi(u)$
$\Leftarrow$: $C(u) = 0$ and $\varphi(u) = L(u,\lambda) = \sup\{L(u^*,\lambda): u^* \in U\} = \theta(\lambda)$.
By weak duality theorem

$$\theta(\mu) \geq \varphi(u) \quad \text{for all } \mu \in R^m$$

....

We have $\theta(\mu) \geq \theta(\lambda)$ for all $\mu \in R^m$
Therefore λ minimizes θ

Now follows the proof of the corrolary

For any u_0 solving (P) we have $C(u_0)=0$ and $\varphi(u_0) \geq \varphi(u^*)$ for all $u^* \in U$

By weak duality theorem

$\theta(\lambda) \geq \varphi(u_0)$, Since u_0 is feasible in (P).But λ minimizes θ

$\theta(\lambda)=\varphi(u)$.Hence

$\varphi(u) \geq \varphi(u^*)$.In particular $u^*=u_0$. We have

$\varphi(u^*) \geq \varphi(u_0)$.

We get

$\varphi(u_0)=\varphi(u)$.

Therefore

$L(u_0,\lambda)=\varphi(u_0)=\theta(\lambda)$

.

2.2 First properties of the dual problem

Let us summarize our development so far

$$\lambda \text{ solves } (2.2) \Rightarrow \lambda \text{ solves } (2.3) \Rightarrow \lambda \text{ solves } (2.6) \qquad (2.7)$$

Thus to solve (2.2) ,we must solve the perfectly well stated problem (2.6).

Property:

–If no primal solution is thus produced ,the task is hopeless from the very beginning:no other value of λ could give a primal solution.

Proof: Let λ solves (2.6) and (2.3) has a solution. By theorem 6 λ is a solution of (2.3). This implies there is $u^* \in U$ such that

$$L(u^*,\lambda)=\sup\{L(u,\lambda) : u \in U\} \text{ and } C(u^*)=0$$

We have

$$L(u^*,\lambda)=\varphi(u^*)=\theta(\lambda)$$

Since by weak duality theorem, we have

$$\theta(\lambda) \geq \varphi(u_0)$$

for all feasible u_0. Thus

$$\varphi(u^*) \geq \varphi(u_0)$$

for all feasible u_0. Then we conclude that no other value of λ could give a primal solution.

–If the technique works,no primal solution can be missed they all solve the Lagrange problem associated with any dual optimum;this is Corollary 7.

Remark :- From now on,we will use the word dual for every thing concerning (2.6): we have to minimize the dual function with respect to the dual variable λ (u being primal variable). The symmetry between (P) and (2.6) becomes more suggestiive if the primal constaints are incorporated into the objective function :

P can be formulated as

$$\sup\{\varphi(u) - I_C(u) : u \in U\},$$

where C is the domain described by C(u) = 0,

I_c being its indicator function :

$$I_c(u) = \begin{cases} 0, & \text{if } C(u) = 0 \\ \infty, & \text{otherwise} \end{cases}$$

The knapsack problem and the entropy maximization give good illustrations of the logical chain(2.7).

Example 1 (The knapsack problem)

Take the case n=1,for example

max u	subject to	$2u \leq 1$, $u \in \{0,1\}$

(2.8)

$\lambda \in \mathbf{R}$:

If $\lambda < 0$, then $\theta(\lambda) = \infty$ (by definition)

If $\lambda \geq 0$, then $\theta(\lambda) = \sup\{u - \lambda(2u-1) : u \in \{0,1\}\}$

$= \sup\{u(1-2\lambda) + \lambda : u \in \{0,1\}\}$

If $0 \leq \lambda \leq 1/2$, then $1-2\lambda \geq 0$ .In this case we take u=1

We have

$$\theta(\lambda) = 1(1-2\lambda) + \lambda = 1 - \lambda$$

If $\lambda > 1/2$,then $1-2\lambda < 0$.In this case we choose u=0 .We have then

$$\theta(\lambda) = \lambda .$$

Hence

$$\theta(\lambda) = \begin{cases} \infty & if \quad \lambda < 0 \\ 1-\lambda & if \quad 0 \leq \lambda \leq \frac{1}{2} \\ \lambda & if \quad \lambda \geq \frac{1}{2} \end{cases}$$

$$C_\lambda = C(u_\lambda) = \begin{cases} 1 & \text{if } 0 \leq \lambda \leq \frac{1}{2} \\ \text{ambiguously } \pm 1 & \text{if } \lambda = \frac{1}{2} \\ -1 & \text{if } \lambda > \frac{1}{2} \end{cases}$$

If $\lambda < 0$, then there is no u_λ,and c_λ .

For (2.3) , $\lambda = 1/2$ gives the doubleton {-1,1}.

Here none of the converse inclusions hold in (2.7) because

$\lambda = 1/2$ solves (2.6), since $\theta(1/2) = 1/2$

but $\lambda = 1/2$ doesnot solve (2.3) since $c_\lambda \pm 1 \neq 0$ and (2.2) and (2.3) have no solutions.

Example 2 : Take the entropy (5)

$$\psi(x,t)=\begin{cases} -t\log t & \text{if } t>0 \\ 0 & \text{if } t=0 \\ -\infty & \text{if } t<0 \end{cases}$$

From (7)

$$u_\lambda(x)=\exp\left(-1-\sum_{j=1}^{m}\lambda_j a_j(x)\right)$$

and

$$\theta(\lambda)=L(u_\lambda,\lambda)=\lambda^T b+\int_\Omega \exp\left(-1-\sum_{j=1}^{m}\lambda_j a_j(x)\right)dx$$

and

$$C_l(u_\lambda)=\int_\Omega a_l(x)\exp\left(-1-\sum_{j=1}^{m}\lambda_j a_j(x)\right)dx-b_l$$

for $l=1,\ldots,m$.

Here θ is convex and $-C=\nabla\theta$.

Here full equivalence holds in (2.7). If λ solves (2.6)

$$\theta(\lambda)=\inf\{\theta(\mu):\mu\in R^m\} \text{ and } \nabla\theta(\lambda)=0,$$

since λ minimizes θ .This implies $-C(u_\lambda)=0$.Hence $C(u_\lambda)=0$.,so we have λ solves (2.3).

If λ solves (2.3),then clearly $C_j(u_\lambda)=0$ for $j=1,2,\ldots,m$. Therefore λ solves (2.2)

Definition : The subdifferential of f at x is a set of vectors $s\in R^n$ satisfying

$$f(y)\geq f(x)+\langle s,y-x\rangle \quad \text{for all } y\in R^n$$

Denotation: $\partial f(x):=\{s\in R^n : f(y)\geq f(x)+\langle s,y-x\rangle \text{ for all } y\in R^n\}$

Proposition 8 :If not identically $+\infty$ the dual function is in $\overline{\text{Conv}}\ R^n$ (the set of lower semicontinous and convex functions on R^n). Furthermore,for any u_λ solution of the Lagrange problem (P_λ)the corrosponding $-C_\lambda$ is a subdifferential of θ at λ .

Proof : For each λ and μ

$$\theta(\mu)=\sup\{L(u,\mu):u\in U\}\geq L(u_\lambda,\mu) \quad \text{for some } u_\lambda\in U$$

That is

$$\begin{aligned}\theta(\mu) &\geq \varphi(u_\lambda)-\mu^T C(u_\lambda)\\ &=\varphi(u_\lambda)-\lambda^T C_\lambda+(\lambda-\mu)^T C_\lambda\\ &=\theta(\lambda)+(\mu-\lambda)^T(-C(u_\lambda))\end{aligned}$$

Thus

$$\theta(\mu)\geq\theta(\lambda)+(\mu-\lambda)^T(-C(u_\lambda))$$

so

$$-C(u_\lambda)\in\partial\theta(\lambda).$$

Convexity of θ:

$$\theta(t\lambda+(1-t)\mu) = \sup\{t\,\varphi(u)+(1-t)\varphi(u)-(t\lambda+(1-t)\mu)^T C(u):\ u\in U\}$$
$$= \sup\{t(\varphi(u)-\lambda^T C(u))+(1-t)(\varphi(u)-\mu^T C(u)):\ u\in U\}$$
$$\le \sup\{t(\varphi(u)-\lambda^T C(u)):\ u\in U\} + \sup\{(1-t)(\varphi(u)-\mu^T C(u)):\ u\in U\}$$
$$= t\sup\{(\varphi(u)-\lambda^T C(u)):\ u\in U\} + (1-t)\sup\{(\varphi(u)-\mu^T C(u)):\ u\in U\}$$
$$= t\,\theta(\lambda)+(1-t)\,\theta(\lambda)$$

This implies

$$\theta(t\lambda+(1-t)\mu) \le t\,\theta(\lambda)+(1-t)\,\theta(\mu).$$

Hence is convex.

For any $\lambda\in R^m$ and $\mu\in R^m$

$$\theta(\lambda) \ge \theta(\mu) + (\lambda-\mu)^T(-C_\mu).$$ This implies

$$\inf\theta(\lambda) \ge \theta(\mu) + (\lambda-\mu)^T(-C_\mu).$$

as $\lambda\to\mu$.we have $\inf\theta(\lambda) \ge \theta(\mu)$ that is

$$\liminf_{\lambda\to\mu}\theta(\lambda) \ge \theta(\mu).$$

Hence by the definition of lower semicontinous (closed function) θ is closed.So $\theta \in \overline{\text{Conv}}\ R^n$

Note that the case $\theta = \infty$ is quite possible:take with U =R, $\varphi(u) = u^2$,$C(u) = u$, no matter how λ is choosen,the Lagrange function $L(u,\lambda) = u^2 - u$ goes to $+\infty$ with $|u|\to+\infty$

Thus the dual problem enjoys several important properties

–It makes sense to minimize on R^m a lower semicontinuous function such as θ .

–The weak duality theorem (4) tells us that θ is bounded below unless there is no feasible u (in which the primal problem doesnot makes sense any way.)

–The set of minimizers of a convex function such as θ is well described,any suitably designed minimization algorithm will converge to such a miniimum.(if any).

The dual approach amounts to exploring a certain primal set parametrized by U^* namely

$U^*=\{u\in U: u$ solves the Lagrange problem $(P_\lambda)$for some $\lambda\}$ (2.9)

Usually $U^*$is properly contained in U.If U contains a feasible point,the dual approach will produce a primal solution ; how much it breaks down depends on the problem.

As an example

Consider the knapsack example (2.8), it has the unique dual solution $\mu = 1/2$, but the primal optimal value is 0.

At $\mu = 1/2$, the Lagrange problem has two solutions (when $u^0 = 0$)

$(u,u^0)=(0,0)$ and $(u,u^0)=(1,0)$

with constraint values -1 and 1:none of them is feasiible, of course on the other hand the non slackened solution u=0 is feasible with respect to the inequality constraint.

This last example illustrates the following important concept:

Definition 9 (Duality gap): The difference between the optimal primal and dual values

$$\inf\{\theta(\lambda): \lambda\in\mathbf{R}^m\} - \sup\{\varphi(u): C(u)=0\} \tag{2.10}$$

when it is not zero,is called the duality gap.

Property : The number

$$\inf\{\theta(\lambda): \lambda\in\mathbf{R}^m\} - \sup\{\varphi(u): C(u)=0\}$$

is always non negative.

Proof: Since by weak duality theorem

$$\theta(\lambda) \geq \varphi(u) \text{ for all } \lambda\in\mathbf{R}^m$$

and all u feasible in (P) , this implies

$$\inf\{\theta(\lambda): \lambda\in\mathbf{R}^m\} \geq \sup\{\varphi(u): C(u)=0\}$$

Hence

$$\inf\{\theta(\lambda): \lambda\in\mathbf{R}^m\} - \sup\{\varphi(u): C(u)=0\} \geq 0$$

Theorem 10: The presence of duality gap definitely implies that U^* defined in (2.9) contains no feasible point.

Proof: suppose U^*contains a feasible point u_0 , by definition of U^*: $C(u_0) = 0$ and u_0 solves the Lagrange problem (P_λ) for some λ : say μ. Thus

$$L(u_0,\mu) = \mathrm{Sup}\{L(u,\mu): u\in U\}$$

but

$$\varphi(u_0) - \mu^T C(u_o) = \varphi(u_0) = L(u_0,\mu) = \mathrm{Sup}\{L(u,\mu): u\in U\} = \theta(\mu).$$

We have

$$\varphi(u_0) = \theta(\mu).$$

Since by weak duality theorem $\theta(\lambda) \geq \varphi(u_0)$ for all $\lambda\in R^m$.

We get

$$\theta(\mu) = \inf\{\theta(\lambda):\lambda\in R^m\} \text{ and } \varphi(u_0) = \theta(\mu) \geq \varphi(u) \text{ for all u feasible in (P).}$$

That means

$$\varphi(u_0) = \mathrm{Sup}\{\varphi(u): u\in U\}.$$

So the duality gap is zero . Hence a contradiction . Therefore U^* contains a feasiible point.

–The absence of duality gap that is having 0 in (2.10) is not quite sufficient to be on the safe side,each of the two exremization problems in (2.10) must have a solution.

–In the latter case,we are done : Corollary 7 applies.

To close this section, we emphasize once more the fact that all the results so far are valid without any specific assumption on the primal data (U, φ, C). Such assumptions become relevant only questions like: does the primal problem have a solution? Does the dual problem have a solution?These questions are going to be addressed now.

2.3 PRIMAL DUAL OPTIMALITY CHARACTERIZATIONS

Suppose the dual problem is solved: some $\lambda \in R^m$ has to be found, minimizing the dual function θ on the whole space; a solution of the primal problem (P) is still to be found.
A dual solution λ is characterized by $0 \in \partial\theta(\lambda)$ because

$$\theta(\mu) \geq \theta(\lambda) \text{ for all } \mu \in R^m \quad \text{implies} \quad \theta(\mu) \geq \theta(\lambda) + \langle 0, \mu-\lambda \rangle \text{ for all } \mu \in R^m.$$

Hence $0 \in \partial\theta(\lambda)$
Consider the (possibly empty) optimal set in the Lagrange problem (P_λ):

$$U(\lambda) := \{u \in U : L(u,\lambda) = \theta(\lambda)\} \tag{2.11}$$

Remembering Proposition (8),
for $u \in U(\lambda)$, $-C(u) \in \partial\theta(\lambda)$. Since $\partial\theta(\lambda)$ is convex, $\text{Conv}\{-C(u): u \in U(\lambda)\} \subseteq \partial\theta(\lambda)$.

Definition 11: (Filling proprty) With $U(\lambda)$ defined above in (2.11),then we say that the filling property holds at each $\lambda \in R^m$ when

$$\partial\theta(\lambda) = -\text{Conv}\{C(u) : u \in U(\lambda)\} \tag{2.12}$$

Observe that the right hand side in (2.12) is then closed , since the left hand side is.

Lemma 12: Suppose that U in (P) is a compact set,on which φ is uppersemicontinuous,and each C_j is continuous.Then the filling property (2.12) holds for each $\lambda \in R^m$.

Proof : - φ is lower semicontinous and $\sum_{j=1}^{m} \lambda_j C_j$ is lower semicontinous, this implies

$$-\varphi + \sum_{j=1}^{m} \lambda_j C_j$$

is lower semicontinous. This means

$$\varphi - \sum_{j=1}^{m} \lambda_j C_j$$

is uppersemicontinuous. Since U is a compact set, then $L(.,\lambda)$ has a maximum for each $\lambda \in R^m$.
This implies dom $\theta = R^m$.
Define $L_u(\lambda) := \varphi(u) - \lambda^T C(u)$ indexed by $u \in U$ is an affine function .Hence it is a convex function.

$$\theta(\lambda) = \text{Sup}\{L_u(\lambda) : u \in U\} < \infty \qquad \text{for all } \lambda \in R^m .$$

this implies θ is convex (Compare prop. 2.1.2 of chap. IV). As defined earlier

$$U(\lambda) := \{u \in U : L_u(\lambda) := \theta(\lambda)\}$$
$$u \to L_u(\lambda) = \varphi(u) - \lambda^T C(u)$$

is upper semicontinous.

$$\theta(\lambda) = \text{Sup}\{L_u(\lambda) : u \in U\} < \infty, \qquad \text{for all } \lambda \in R^m .$$

Then by theorem 4.4.2 of Chapter VI, we have

$$\partial\theta(\lambda) = \text{Conv}\{\partial L_u(\lambda) : u \in U(\lambda)\}.$$

Hence

$$\partial\theta(\lambda) = \{-C(u)\}$$

Finally we have

$$\partial\theta(\lambda) = \mathrm{Conv}\{-C(u): u\in U(\lambda)\}$$
$$\partial\theta(\lambda) = -\mathrm{Conv}\{C(u): u\in U(\lambda)\} .$$

Therefore the filling property (2.12) holds at each $\lambda.\in R^m$.

The essential result concerning primal dual relationships can now be stated.

Theorem 13: Let the filling property (2.12) hold (for example make the assumptions of Lemma 12) denote by

$$C(\lambda) := \{ C(u):\ u \in U(\lambda) \} \tag{2.13}$$

1. A dual optimum μ is characterized by the existence of $k \le m+1$ points $u_1, u_2, .., u_K$ in $U(\mu)$ and convex multipliers $\alpha_1, \alpha_2, \ldots, \alpha_K$ such that

$$\sum_{i=1}^{k} \alpha_i C(u_i) = 0$$

2. In particular, if $C(\mu)$ is convex for some optimal μ, then for any optimal ν, the feasible points in $U(\nu)$ make up all the solutions of (P).

Proof: when the filling property holds

$$\partial\theta(\mu) = -\mathrm{Conv}\{C(u): u\in U(\mu)\}$$

1. Since μ is an optimal solution

$$0 \in \partial\theta(\mu).$$

This implies

$$0\in -\mathrm{Conv}\{C(u):\ u\in U(\lambda)\}.$$

That is

$$0 = -\sum_{i=1}^{k} \alpha_i C(u_i) \text{ where } k \le m+1 \quad \text{and} \qquad \sum_{i=1}^{k} \alpha_i = 1, \alpha_i \ge 0,$$

$i = 1,\ldots,k$ and $u_i \in U(\mu)$, hence

$$0 = \sum_{i=1}^{k} \alpha_i C(u_i)$$

2. If $C(\mu)$ is convex for some optimal μ, then

$$\mathrm{Conv}\{C(u):\ u\in U(\mu)\} = \{C(u):\ u\in U(\mu)\}$$
$$0\in -\mathrm{Conv}\{C(u):\ u\in U(\mu)\} \text{ implies } 0\in -\{C(u):\ u\in U(\mu)\}.$$

This means there is $v\in U(\mu)$ such that $C(v) = 0$. Hence 0 is the constraint value of some v maximizing $L(.,\mu)$, that means μ solves (2.3). By Corollary 7 the primal solutions are those u maximizing $L(.,\mu)$ that are feasible in (P).

This is the first instance where convexity comes into play, to guarantee that the set U^* of (2.9) contains a feasible point. Convexity is therefore crucial to rule out duality gap.

Example : Consider the knapsack problem

$$\max u \qquad 2u \leq 1 \qquad u\in\{0,1\}$$

At the unique dual optimum $\mu = 1/2$, $C(1/2) = \{-1,1\}$ is not convex .

Duality gap = $\inf\{\theta(\lambda):\lambda \in R^m\} - \sup\{\varphi(u): C(u) = 0\}$

$$\inf\{\theta(\lambda):\lambda \in R^m\}=1/2$$

$$\sup\{\varphi(u): C(u)=0\} = \sup\{u: 2u+u_0-1 = 0, u_0 \geq 0 , u\in\{0,1\}\} = 0$$

Duality gap = 1/2 – 0 = 1/2.

$$U(1/2) = \{u\in\{0,1\} : L(u,\lambda)=\theta(\lambda)\}$$
$$=\{(0,0),(1,0)\} \qquad (\text{where } u_0 = 0)$$

and the convex multipliers $a_1= a_2= 1/2$ do make up the point

$1/2\,(0,0) +1/2\,(1,0) = (1/2\,,0)$, which satisfies the constraint $2u \leq 1$.

In practice, the convexity property needed for theorem 13 implies that the original problem itself has the required convex structure: roughly speaking,the only cases in which there is no duality gap are those described by the following result.

Corollary 14: Suppose that the filling property (2.12) holds. In either of

(I) For some dual solution μ, the associated Lagrange function $L(., \mu)$ is maximized at a unique u_0; then u_0 is the unique solution of (P).

(II) In (P) U is convex, φ concave and $C:U\to R^m$ is affine, there is no duality gap; for every dual solution ν (if any), the feasible points in $U(\nu)$ make up all the solutions of the primal problem (P)

Proof:

(I) $\qquad \theta(\mu).= \inf\{\theta(\lambda):\lambda\in R^m\}$

$$\theta(\mu) = L(u_0,\mu) \text{ and } C(u_0) = 0$$
$$= \varphi(u_0) - \mu^T C(u_0) = \varphi(u_0).$$

That is $\varphi(u_0) = \theta(\mu) \geq \varphi(u)$ for all u feasible in (P)

Hence $\varphi(u_0) \geq \varphi(u)$ for all u feasible in (P)

$$\inf\{\theta(\lambda):\lambda\in R^m\}- \sup\{\varphi(u):C(u) = 0,u\in U\} = \theta(\mu) - \varphi(u) = \theta(\mu) - \theta(\mu) = 0$$

This implies no duality gap.

(II) φ concave implies that $-\varphi$ is convex and $C(u) = \langle a,u\rangle + b$, being affine.

$$U(\lambda)=\{u\in U:L(u,\lambda) =\theta(\lambda)\}.$$

First we show that $U(\lambda)$ is convex.

Let $C(u_1),C(u_2)\in C(\lambda)$ where $u_1,u_2\in U(\lambda))$ .By definition :

$$L(u_1,\lambda) = \theta(\lambda) = \varphi(u_1)-\lambda^T C(u_1)$$
$$L(u_2,\lambda) = \theta(\lambda) = \varphi(u_2)-\lambda^T C(u_2).$$

Since $-\varphi$ is convex,

$$-\varphi(tu_1+(1-t)u_2) \leq -t\varphi(u_1)+-(1-t)\varphi(u_2) \qquad t\in[0,1].$$

That is

$$\varphi(tu_1+(1-t)u_2)\geq t\varphi(u_1)+(1-t)\varphi(u_2) .$$

Since C is affine,

$$C(tu_1+(1-t)u_2) = \langle a, tu_1+(1-t)u_2\rangle + b$$
$$= t[\langle a,u_1\rangle+b] + (1-t)[\langle a,u_2\rangle+b]$$
$$= tC(u_1) + (1-t)C(u_2).$$

Hence

$$-\lambda^T[C(tu_1+(1-t)u_2] = -\lambda^T[tC(u_1)+(1-t)C(u_2)].$$

we get

$$\varphi(tu_1+(1-t)u_2)-\lambda^T[C(tu_1+(1-t)u_2] \geq t[\varphi(u_1)-\lambda^T C(u_1)] + (1-t)[\varphi(u_2)-\lambda^T C(u_2)]$$
$$= t\theta(\lambda)+(1-t)\theta(\lambda)$$
$$= \theta(\lambda) \qquad \text{(X)}$$

Since U is convex $tu_1 + (1-t)u_2 \in U$, and since $\theta(\lambda) = \sup\{L(u,\lambda): u\in U\}$
We have equality in (X). This implies $tu_1+(1-t)u_2 \in U(\lambda)$.

Convexity of $C(\lambda)$:
Let $C(u_1),C(u_2)\in C(\lambda)$. Then $(1-t)C(u_1)+tC(u_2) = C((1-t)u_1+tu_2)\in C(\lambda)$. Then by the above theorem 0 is the constraint value of some u maximizing $L(.,\mu)$, then U^* contains a feasible point. Hence there is no duality gap.
Moreover from theorem (13) for every dual solution ν, the feasible points in $U(\nu)$ make up all the solutions of the primal problem (P).

2.4. EXISTENCE OF DUAL SOLUTIONS

The question is now whether the dual problem (2.6) has a solution $\mu\in R^m$. This implies

$$\theta(\mu) = \inf\{\theta(\lambda):\lambda\in R^m\} = \min\{\theta(\lambda):\lambda\in R^m\}$$

Denotation: $C(u)$: $\{\gamma\in R^m: \gamma=C(u) \text{ for some } u\in U\}$

First of all denote by Γ the affine hull of C(U):

$$\Gamma := \{\gamma\in \mathbf{R}^m : \gamma = \sum_{i=1}^{k} \alpha_i C(u_i), u_i\in U, \ \sum_{i=1}^{k}\alpha_i = 1, \ k = 1, 2, \ldots\}$$

and by Γ_o is the subspace parallel to Γ. Fixing $u_0\in U$, $C(u_0)\in\Gamma$, $\Gamma - C(u_0)$ is a subspace of R^m.
$C(u)\in\Gamma$ implies $C(u)-C(u_0)\in\Gamma_o$ for all $u\in U$, hence

$$L(u,\lambda+\mu) = \varphi(u)-(\mu+\lambda)^T C(u)$$
$$= \varphi(u)-\lambda^T C(u)-\mu^T C(u)$$
$$= L(u,\lambda)-\mu^T C(u)$$

If $\mu\in\Gamma_o^{\perp}$ then $\langle\nu,\mu\rangle = 0$ for all $\nu\in\Gamma_o$. In particular for $\nu = C(u) - C(u_0)$ we have

$$\mu^T(C(u)-C(u_0)) = 0 \quad \text{iff} \quad \langle\mu,C(u)\rangle = \langle\mu,C(u_0)\rangle$$

Thus

$$L(u,\lambda+\mu)=L(u,\lambda)-\mu^T C(u_0) \text{ for all } (\lambda,\mu)\in R^m\times\Gamma_o^{\perp}$$

For $\theta(\lambda+\mu)$:

$$\theta(\lambda+\mu) = \operatorname{Sup}\{L(u,\lambda+\mu): u\in U\} = \operatorname{Sup}\{[L(u,\lambda)-\mu^T C(u_0)]: u\in U\}$$

$$= \text{Sup}\,\{L(u,\lambda) : u\in U\} - \mu^T C(u_0)$$
$$= \theta(\lambda) - \mu^T C(u_0) \quad \text{for all } (\lambda,\mu)\in R^m\times\Gamma_o^{\perp} \tag{2.14}$$

that is θ is affine in the subspace $\Gamma_o^{\perp}$.

This observation clarifies two cases :

1. If $0\notin\Gamma$ then $\Gamma\neq\Gamma_o$

Let $\mu_0 \neq 0$ be the projection of the origin onto Γ. In (2.14), fix λ and take $\mu = t\mu_0$ with $t\to\infty$. Because $C(u_0)\in\Gamma$ We have with (2.14)

$$\theta(\lambda+t\mu_0) = \theta(\lambda) - t\mu_0^T C(u_0) = \theta(\lambda) - t\|\mu_0\|^2 \to -\infty .$$

In this case, the primal problem cannot have a feasible point (since not bounded below) by the weak duality theorem; to become meaningful, (P) should have its constraints perturbed say to $C(u) = \mu_0$.

If $0\in\Gamma$, then the dual optimal set is $\Lambda + \Gamma_o^{\perp}$, $\Lambda \subseteq\Gamma_o$, where the (possibly empty) optimal set of

$$\inf\{\ \theta(\lambda): \lambda\in\Gamma_o = \text{aff}C(U)\} \tag{2.15}$$

In a way, (2.15) is the relevant dual problem to consider admitting that Γ_o is known. Alternatively, the essence of the dual problem would remain unchanged if we assume $\Gamma = \Gamma_o = R^m$,

$$\text{Inf}\,\{\theta(\lambda):\lambda\in R^m\} = \inf\{\theta(\lambda):\lambda\in\Gamma_o = R^m\}$$

In 2.3, some convexity structure came into play for solving the primal problem by duality,the same phenomenon occurs here.The next result contains the essential conditions for the existence of the dual solution.

Proposition 15 : Assume $\theta \neq \infty$ with the definition of $C(U)$

1. If $0 \notin \overline{\text{Conv}}\ C(U)$, then $\text{Inf}\{\ \theta(\lambda): \lambda\in R^m\} = \infty$

2. If $0\in\text{Conv}\ C(U)$, then $\text{Inf}\{\ \theta(\lambda) : \lambda\in R^m\ \} > -\infty$

3. If $0\in\text{riConv}C(U)$, then the dual problem has a solution.

Proof:

1.The closed convex set $\overline{\text{Conv}}\ C(U)$ is separated from $\{o\}$.

[Compare Theorem 4.1.3 of Chapter III].

For some $\mu_0 \neq 0$ and $\delta>0$, $\langle\mu_0,C(u)\rangle \geq \delta > 0$ for all $u\in U$. This implies $-\mu_0^T C(u) \leq -\delta < 0$ for all $u\in U$. Since $\theta \neq \infty$ then there is $\mu\in R^m$ such that $\theta(\mu) < \infty$ and write for all $u\in U$:

$$L(u,\mu+t\mu_o) = \varphi(u) - \mu TC(u) - t\mu_o^T C(u)$$
$$\leq \theta(\mu) - t\mu_o^T C(u) \qquad t>0$$
$$\leq \theta(\mu) - t\delta$$

Since u was arbitrary then we get

$$\theta(\mu+t\mu_0) \leq \theta(\mu) - t\,\delta$$

Thus

$$\inf\{\ \theta(\lambda):,\ \lambda\in R^m\ \}\ .$$

2. $0\in\text{Conv}\ C(U)$

this implies

$$0 = \sum_{i=1}^{p} \alpha_i C(u_i) \qquad \sum_{i=1}^{p} \alpha_i = 1 \qquad \alpha_i \geq 0$$

Then write for all $\lambda\in R^m$

$$\theta(\lambda)\geq\varphi u_i)-\alpha_i\lambda^T C(u_i) \quad \text{for } i=1,2,\ldots P.$$

$$\theta(\lambda)=\sup\{L(u,\lambda): u\in U\}$$

thus

$$\alpha_i\theta(\lambda)\geq\alpha_i\varphi(u_i)-\alpha_i\lambda^T C(u_i) \qquad \text{for } i=1,2,\ldots P, \text{ since } \alpha_i\geq 0.$$

Hence

$$\sum_{i=1}^{p}\alpha_i\theta(\lambda)\geq\sum_{i=i}^{p}\alpha_i\varphi(u_i)-\sum_{i=1}^{p}\alpha_i\lambda^T C(u_i)\geq\min\{\varphi(u_i) : 1\leq i\leq p\}$$

We have

$$\inf\{\theta(\lambda) : \lambda\in R^m\}>\infty$$

3. We know from theorem 2.1.3 of chapter (III) that, if $0\in$ riConv C(U), then 0 is also in the relative interior of some simplex contained in Conv C(U). Using the notation of (2.15), this means that there are finitely many points in U, say $u_1,u_2,\ldots,u_p$ and $\delta>0$ such that

$$0\in \text{aff } C(U)=\Gamma=\Gamma_0$$

$$B(0,\delta)\cap\Gamma_0\subseteq \text{Conv}\{C(u_1),C(u_2),\ldots,C(u_p)\}.$$

For all $\lambda\in R^m$still by definition of $\theta(\lambda)$

$$\theta(\lambda)\geq\varphi(u_i)-\lambda^T C(u_i) \quad \text{for } i=1,2,\ldots,P \qquad 2.16)$$

If $\Gamma_0=\{0\}$, that is

$C(U)=\{0\}$, $L(u,\lambda)=\varphi(u)$ for all $u\in U$,

so θ is a constant, since

$$\theta(\lambda)=\sup\{L(u,\lambda) : u\in U\}=\sup\{\varphi(u) : u\in U\}$$

and this constant is not ∞ , by assumption ($\theta\neq\infty$)

If $\Gamma_0\neq\{0\}$then there exists $\lambda\neq 0$such that $\lambda\in\Gamma_0$, in (2.16)

$$\gamma := -\delta\,\lambda/\|\lambda\|\in B(0,\delta)\cap\Gamma_0$$

since $\|\gamma\|=\delta\in B(0,\delta)$ and Γ_0 is a vector space , $-\delta/\|\lambda\|$ is a constant.Hence $\gamma\in B(0,\delta)\cap\Gamma_0$

Since $B(0,\delta)\cap\Gamma_0\subseteq \text{Conv}\{C(u_1),\ldots,C(u_p)\}$.We have

$\gamma\in \text{Conv}\{C(u_1),C(u_2),\ldots,C(u_p)\}$.thsi means γ is a convex combination of the $C(u_i)$'s.

This implies

$$\theta(\lambda)\geq\sum_{i=1}^{p}\alpha_i\varphi(u_i)-\lambda^T\sum_{i=1}^{p}\alpha_i C(u_i)=\sum_{i=1}^{p}\alpha_i\varphi(u_i)-\lambda^T\gamma$$

$$=\sum_{i=1}^{p}\alpha_i\varphi(u_i)+\delta\|\lambda\|\geq\min_{1\leq i\leq p}\varphi(u_i)+\delta\|\lambda\|$$

We conclude that $\theta(\lambda)\to\infty$ Γ_0 if λ grows unboudedly in Γ_0 : the closed function θ (Propo.8) does have a minimum on Γ_0 and on R^m as well, in veiw of (2.15).

figure 1 summarizes the different situations revealed by our above analysis

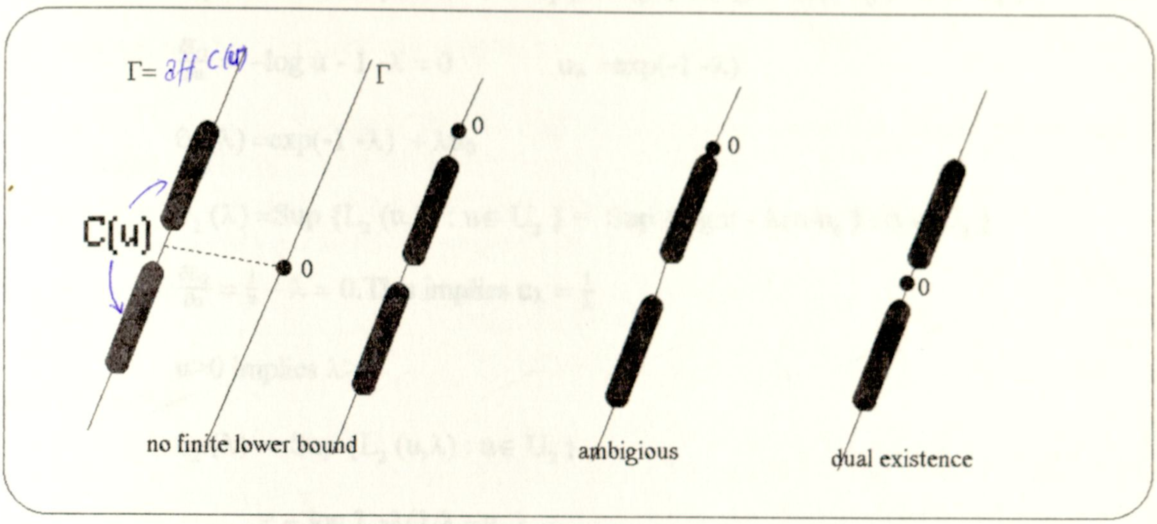


Fig 1 Various situations for dual existence.

Example 1

In view of the weak duality theorem 4, existence of a feasible u in (P) implies boundedness of θ from below. To confirm the pridiction of (III), that this is not necessary, take a variant of the knapsack problem (2.8), in which the knapsack should be completely filled : we want to solve

$$\text{Sup } u \ , \quad 2u=1 \quad , u \in \{0,1\}$$

$$2u-1=0 : \quad u=0, \quad 2u-1=-1\neq 0$$

$$u=1, \quad 2u-1=1\neq 0$$

The problem has no feasible point the supremal value is

$$\sup\varnothing = -\infty.$$

Nevertheless the dual problem is the sam,e

$$\theta(1/2) = 1/2 \leq \theta(\lambda) \ \text{ for all } \lambda \in R^m$$

$$\text{Duallity gap } = \inf\{\theta(\lambda), \lambda\in R^m\} - \sup\{\varphi(u) : C(u) = 1, u \in\{0,1\}\}$$

$$= 1/2 - (-\infty) = \infty.$$

The entropy problem provides few examples to show that the topological operations also play their role in proposition 15.

Examples:

Take $\Omega = \{0\}$ and one equality constraint : Our entropy problem is

$$\text{Sup } \varphi(u) \ , \ u\in U \quad , \ u-u_0= 0 \quad \text{where } u_0 \text{ is fixed} .$$

we consider two particular cases :

$$U_1=[0,\infty.) \qquad \varphi_1(u)= -u \log u$$

$$U_2=(0,\infty) \qquad \varphi_2(u)= \log u$$

$$C(u)= u-u_0$$

$$C(U_1)=[-u_0,\infty.) \quad \text{and} \quad C(U_2)=(-u_0,\infty.)$$

both $C(U_1)$ and $C(U_2)$ are convex.

Dual functions

$$\theta_1(\lambda) = \text{Sup}\,\{L_1(u,\lambda) : u\in U_1\} = \text{Sup}\,\{-u \log u - \lambda(u-u_0) : u \in U_1\}$$

$$\frac{\partial L_1}{\partial u} = -\log u - 1 - \lambda = 0 \qquad u_\lambda = \exp(-1-\lambda)$$

$$\theta_1(\lambda) = \exp(-1-\lambda) + \lambda u_0$$

$$\theta_2(\lambda) = \text{Sup}\,\{L_2(u,\lambda) : u\in U_2\} = \text{Sup}\,\{\log u - \lambda(u-u_0) : u \in U_2\}$$

$$\frac{\partial L_2}{\partial u} = \frac{1}{u} - \lambda = 0. \text{This implies } u_\lambda = \frac{1}{\lambda}$$

$u>0$ implies $\lambda>0$

$$\theta_2(\lambda) = \text{Sup}\,\{L_2(u,\lambda) : u\in U_2\}$$

$$= -\log\lambda - \lambda(1/\lambda - u_0)$$

$$= -\log\lambda + \lambda u_0 - 1$$

As predicted by proposition 15:
The cases $u_0 > 0$ and $u_0 < 0$ implies :
If $u_0 > 0$, then $-u_0 < 0$. This implies $0\in \text{ri}C(u_1)$ and $0\in \text{ri}C(u_2)$,since $\theta_1,\theta_2 \neq \infty$, by proposition 15 the dual problem has a solution.
If $u_0 < 0$,then $-u_0 > 0$. This implies

$$C(U_1) = [-u_0,\infty) \; ; \; C(U_2) = (-u_0,\infty)$$

$$\text{Conv}\,C(U_1) = [-u_0,\infty) \; ; \; \text{Conv}\,C(U_2) = (-u_0,\infty)$$

$$0 \notin \overline{\text{Conv}}\,C(U_1) \quad \text{and} \quad 0 \notin \overline{\text{Conv}}\,C(U_2)$$

then we have

$$\inf\{\theta(\lambda) : \lambda\in R\} = -\infty$$

If $u_0=0$, then

$$C(U_1) = [0,\infty) \text{ and } C(U_2) = (0,\infty)$$

We have $0\in \text{bd}\,C(U_i)$ for $i = 1,2$
If $u_0 = 0$ then

$$\theta_1(\lambda) = \exp(-1-\lambda) \text{ and } \theta_2(\lambda) = -\log\lambda - 1$$

$\inf\{\theta_1(\lambda): \lambda\in R\}= 0$

$\inf\{\theta_2(\lambda): \lambda\in R\}= -\infty$, there is no dual solution.

ILLUSTRATIONS

3.1 THE MINIMAX POINT OF VIEW

Define :

- Some set V (playing the role of R^m, the dual of the space of constraint values) whose elements $v\in V$ play the role of $\lambda\in R^m$.
- Some bivariate function l:UxV→ $R\cup\{\infty\}$, playing the role of the Lagrangian L.

V and l must satisfy the following property,for all $u \in U$

$$\inf \{l(u,v):v\in V\} = \begin{cases} \varphi(u) & \text{if } C(u)=0 \\ -\infty & \text{otherwise} \end{cases} \quad - \tag{3.1}$$

then (P) is equivalent to

$$\sup_{u\in U} \{\inf_{v\in V} l(u,v)\} \tag{3.2}$$

Let $\theta(v) := \text{Sup}\{l(u,v):u\in U\}$.

The whole business of the general pattern is therefore to invert the inf - and sup-operations, replacing (P) = (3.2) by (2.6) of section 2, which reads

$$\inf \{\theta(v) :v\in V\} \tag{3.3 }$$

In order to make (3.3) easy, an extra requirement is therefore:

the function θ and the set V are closed convex . (3.4)

$L_u(\lambda) = L(u,\lambda) = \varphi(u)- \lambda^T C(u) = \varphi(u) - <\lambda,C(u)>$ affine with respect to the dual varible λ.

If $\theta \neq \infty$, then there is μ such that $\theta(\mu) < \infty$. By proposition 2.1.2 of chapter IV, θ is closed convex.

Remark :

In the dual approach,the basic idea is thus to replace the sup-inf problem (P) = (3.2) by its Sup-inf problem (P) = (3.2) by its inf -sup form (2.6)=(3.3).

3.2 INEQUALITY CONSTRAINTS

Suppose (P) has rather the form

$$\sup_{u\in U} \{\varphi(u): C_j(u)\leq 0 \text{ for } j=1,2,...P\} \tag{3.5}$$

In the knapsack example, non-negative slack variables were included,so as to recover the equality-constrained framework;then the dual function was ∞ outside the non negative orthant.

From the same Lagrange function as before,

$(u,\mu) \in U\text{x}V$ such that $(u,\mu)\to L(u,\mu):=\varphi(u)- \mu^T C(u)$

but take

$V:=\{\mu\in R^p: \mu_j\geq 0 , \text{ for } j=1,2,...p\}$

This construction satisfies (3.1) : If u violates some constraint, say $C_i(u)>0$,fix all $\mu_j = 0$ except μ_i, which is sent to ∞, the Lagrange function is :

$$L(u,\mu) = \varphi(u) - \mu_i C(u) \ \rightarrow \ -\infty$$

The dual problem associated with (3.5) is therefore

$$\inf_{\mu\in(R^+)^P} \ \{\sup\{L(u,\mu) : u\in U\}$$

Remarks :

1. $(R^+)^P$ is a closed convex set.(3.4) is preserved if θŸŸ}
2. More generally,a problem such as

$$\sup_{u\in U} \{\varphi(u) : a_i(u)=0 \text{ for } i=1,2,\dots,P\}$$

Can be dualized as follows: supremize over $u\in U$ the Lagrange function

$$L(u,\lambda,\mu) := \varphi(u) - \sum_{i=1}^{m} \lambda_i a_i(u) - \sum_{j=1}^{P} \mu_j C_j(u)$$

To obtain the dual function

$$\theta : R^m x R^p \rightarrow R\cup\{\infty\} \ \text{ such that}$$

$$\theta(\lambda,\mu) = \text{Sup}\ \{L(u,\lambda,\mu) : u\in U\}$$

then solve the dual problem

$$\text{Inf}\ \{\ \theta(\lambda,\mu) : (\lambda,\mu)\in R^m\ x(R^+)^P\ \} \tag{3.6}$$

Depending on the context,it may be simpler to impose directly dual nonnegativity constraints,or to use slack variables. Let us review the main results of Section 2 in an inequality-constrained context.

A. Modified Everett's Theorem

For $\mu\in(R^+)^P$,let u_μ maximize the Lagrangian $\varphi(u) - \mu^T C(u)$ associated with (3.5).Then u_μ solves

$$\begin{array}{ll} \max \varphi(u) & u\in U \\ C_j(u) \le C_j(u_\mu) & \text{if } \mu_j > 0 \\ C_j(u) \ \text{unconstrained} & \text{otherwise} \end{array} \tag{3.7}$$

A way to prove this is to use a slackened formulation of (3.5)

$$\sup_{u,v} \{\varphi(u) : u\in U, v\in(R^+)^P, C(u)=v\} \tag{3.8}$$

Then the original Everett theorem directly applies; observe that any v_j maximizes the associated Lagrangian if $\mu_j= 0$. Furthermore,if u_μ , is feasible, the last line of (3.7) can be replaced by $C_j(u) \le 0$. [$C_j(u_\mu) \le 0$].We see that, if u_μ is feasible in (3.5) and satisfies $C_j(u_\mu) = 0$ if $\mu_j>0$, then u_μ solves (3.5).

B. Modified filling property

For convinient primal dual relationship in problems with inequality constraints, the filling assumption (2.12) is not sufficient. Let again

$$\theta(\mu) = \text{Sup}\ \{L(u,\mu) : \ u\in U\}$$

be the dual function associated with (3.5), the basic result reproducing Section 2.3 is then as follows

[redacted]

PROPOSITION 1: With the above notation, assume that

$$\partial\theta(\mu) = -Conv\{C(u) : L(u,\lambda) = \theta(\lambda) \quad \text{for all } \mu\in(R^+)^p$$

and

$$\theta(\mu) < \infty \text{ for some m with } \mu_j > 0 \text{ for } j=1,2,...p \tag{3.9}$$

then

λ solves the problem iff there are $k \le P+1$ primal points $u^1, u^2, ..., u^k$ maximizing $L(.,\lambda)$, and $\alpha=(\alpha_1,\alpha_2, ...,\alpha_p) \in \Delta_K$ such that

$$\sum_{i=1}^{k} \alpha_i C_j(u^i) \le 0 \text{ and } \lambda_j \sum_{i=1}^{k} \alpha_i C_j(u^i) = 0 \quad \text{for } j=1,2,....,k$$

Proof:

λ solves the dual problem

$$\theta(\lambda) = \inf\{\theta(\mu) : \mu\in R^P\}$$

$$\theta(\lambda) \le \theta(\mu) \quad \text{for all } \mu\in R^P$$

this implies

$$\theta(\lambda) + I_{(R^+)^p}(\lambda) \le \theta(\mu) + I_{(R^+)^p}(\mu) \quad \text{for all } \mu\in R^P$$

That is

$$0 \in \partial(\theta + I_{(R^+)^p})(\lambda)$$

So we need to compute the subdifferential of this sum of functions. With μ as stated in (3.9)

$$\theta(\mu) < \infty \quad (\mu_1,\mu_2,...,\mu_p) \ \mu_j > 0 \text{ for } j=1,2,...p$$

there is a ball $B(\mu,\delta)$ contained in $(R_*^+)^P$: and any such ball intersects ri dom θ . Then

$$B(\mu,\delta) \cap (R_*^+)^P \ne \varnothing$$

$$B(\mu,\delta) \cap \text{ri dom } \theta \ne \varnothing$$

Thus

$$\text{ri dom } \theta \cap (R_*^+)^P \ne \varnothing$$

Hence equality holds in Corollary 3.1.2 of Chapter XI, that is

$$\partial(\theta + I_{(R^+)^p}) = \partial\theta + N_{(R^+)^p}$$

holds on dom $\theta \cap (R^+)^P$

but

$$N_{(R^+)^p}(\lambda) = \{s \in (R_-)^p \mid \langle s, \lambda\rangle\}$$

For $\lambda \ne 0$, $0 \in \partial(\theta + I_{(R^+)^p})(\lambda)$ implies $0 \in \partial\theta(\lambda) + N_{(R^+)^p}(\lambda)$

This implies

$$0 = -\sum_{i=1}^{k} \alpha_i C(u^i) + s \text{ , where } s \in (R_-)^p \text{ such that } \langle s , \lambda\rangle = 0$$

This implies

$$s_j = \sum_{i=1}^{k} \alpha_i C_j(u^i) \leq 0 \text{ and}$$

$$\sum_{j=1}^{k} \lambda_j s_j = 0 \text{ implies that } \lambda_j s_j = 0 \text{ for } j = 1, \ldots, k$$

Therefore

$$\lambda_j \sum_{i=1}^{k} \alpha_i C_j(u^i) = 0 \text{ for } j = 1, \ldots, k$$

C) EXISTENCE OF DUAL SOLUTIONS

Finally let us examine how proposition 15 can handle inequality constraints using slack variables, the image set associated with (3.8) is

$$C'(U) = C(U) + (R^+)^p$$

where C(U) is defined in (2.14). Clearly , aff $C'(U) = (R)^p$ and we obtain

(i) If $0 \notin \overline{Conv}\,[C(U) + (R^+)^p]$ then $\inf \theta = -\infty$

(ii) If $0 \in \text{Conv } C(U) + (R^+)^p$, $\inf \theta > -\infty$. This comes from the relation

$$\text{Conv } C'(U) = \text{Conv}[C(U) + (R^+)^p] = \text{Conv } C(U) + (R^+)^p.$$

(iii) If $0 \in \text{ri}[\text{Conv } C(U) + (R^+)^p]$, then the dual solution exists.

PROPOSITION 2 :If there are $k \leq p+1$ points $u^1, u^2, \ldots, u^k$ in U and convex multipliers $\alpha_1, \alpha_2, \ldots, \alpha_k$ such that

$$\sum_{i=1}^{k} \alpha_i C_j(u^i) < 0 \quad \text{for } j = 1, 2, \ldots, p$$

then the dual function associated with the optimization problem (3.5) has a minimum over the nonnegative orthant $(R^+)^p$.

proof: If there are $k \leq p+1$ such that

$$\sum_{i=1}^{k} \alpha_i C_j(u^i) < 0 \text{ for } j = 1, 2, \ldots, p$$

This implies

$$\sum_{i=1}^{k} \alpha_i C_j(u^i) + \beta_j = 0 \text{ for } \beta_j > 0 \text{ ,for } j = 1, 2, \ldots, p$$

This implies $0 \in \text{Conv } C(U) + (R_*^+)^p$

that is

$$\sum_{i=1}^{k} \alpha_i C_j(u^i) + \beta_j = 0 \text{ for } j = 1, 2, \ldots, p$$

it follows

$$0 \in \text{Conv } C(U) + (R_*^+)^p = \text{ri}[\text{Conv } C(U) + (R_*^+)^p]$$

Then by proposition 15 (iii) the dual solution exists.

3.3 DUALIZATION OF LINEAR PROGRAMS

As a particular case of Sec.3.2.Suppose that U=R^n, with some scalar product <,>, and that (3.5) is

$$\text{Sup } \langle q,u\rangle \quad u\in U \tag{3.10}$$

$$\langle a_j,u\rangle \le b_j \quad \text{for } j=1,2,\ldots P$$

Where q and each a_j are in R^n ,b=(b_1 ,b_2 ,...,b_p) $\in R^P$ .The Lagrange function

$$L(u,\mu) := \langle q,u\rangle - \sum_{j=1}^{p} \mu_j[\langle a_j, u\rangle - b_j]$$

$$= \langle q-\sum_{j=1}^{p} \mu_j a_j,\ u\rangle + \sum_{j=1}^{p} \mu_j b_j$$

$\theta(\mu) = \infty$ if $\mu \notin R^P$ or if

$$q-\sum_{j=1}^{p} \mu_j a_j \neq 0 .$$

The dual problem is therefore

$$\inf\{\sum_{j=1}^{p} \mu_j b_j : \mu_j \ge 0 \text{ for } j=1,2,\ldots,p \quad \sum_{j=1}^{p} \mu_j a_j = q\} \tag{3.11}$$

If $\mu \in \text{Dom } \theta$,

$$\theta(\mu) = \sum_{j=1}^{p} \mu_j b_j$$

therefore any $u\in R^m$ maximizes the Lagrangian.(that is any u is u_μ)

Remark: The dual constraints that appear in (3.6) and (3.11) simply express that the study can be restricted to dom θ, since θ must be minimized instead of (3.11) ,for example, the dual of (3.10) can obviously be formulated as the unconstrained minimization of the function

$$\begin{cases} \sum_{j=1}^{p} \mu_j b_j \ \text{ if } \ \sum_{j=1}^{p} \mu_j a_j = q \ \text{ and } \ \mu \in (R^+)^p \\ \infty \qquad\qquad \text{otherwise} \end{cases}$$

Assume that <.,.> is the usual dot product and that our primal problem is now

$$\text{Sup } \{ q^T u : Au=b,\ u_i \ge 0 \text{ for } i=1,2,\ldots,n\} \tag{3.12}$$

Here A is an mxn matrix, $b\in R^m$. It is natural to take as primal set U:=$(R^+)^n$, so that the dual space is apriori R^m, m being the number of equalities. Then the Lagrange function

$$L(u,\lambda) := (q^T - \lambda^T A)u - \lambda^T b$$

The maximal value is finite if and only if the vector $q-A^T\lambda$ has all its coordinates non positive. The dual of (3.12) is

$$\inf \{b^T\lambda : \lambda\in R^m,\ A^T\lambda - q \in (R^+)^n\} \tag{3.13}$$

3.4 DUALIZATION OF QUADRATIC PROGRAMS

Let again U be R^n,equipped with the dot product for simplicity, and consider

$$\text{Sup}\{q^T u - 1/2 u^T Q u : Au \le b\} \tag{3.14}$$

Here $q \in R^n$, $b \in R^P$ and the nxn matrix Q is symmetric and positive definite. This problem has a unique solution if the feasible domain is nonempty .

Choose the Lagrange function

$$L(u,\mu) := q^T u - \tfrac{1}{2} u^T Qu - \mu^T(Au-b) = -1/2<Qu,u> + <q-A^T\mu,u> + b^T\mu$$

Maximum is attained : $-Qu + q - A^T\mu = 0$.

This implies $Qu = q - A^T\mu$

Hence $u_\mu = Q^{-1}(q - A^T\mu)$ (3.15)

The corrosponding constaint value is

$$C(u_\mu) = Au_\mu - b = AQ^{-1}(q - A^T\mu) - b.$$

Inserting the value of u_μ into L, we obtain the dual function

$$\theta(\mu) = L(u_\mu, \mu) = (q-A^T\mu)^T u_\mu - 1/2\, u^T_\mu Qu_\mu + b^T\mu$$

$$= (q-A^T\mu)^{T}\, {}^{-}Q^{1}(q-A^T\mu) - 1/2[Q^{-1}(q-A^T\mu)]^T Q[Q^{-1}(q-A^T\mu)] + b^T\mu$$

$$= (q-A^T\mu)^T Q^1 (q-A^T\mu) - 1/2\,(q-A^T\mu)^T Q^{-1}(q-A^T\mu) + b^T\mu$$

$$= 1/2(q-A^T\mu)^T Q^1 (q-A^T\mu) + b^T\mu$$

to be minimized over $(R^+)^P$ and

$$\theta(\mu) = AQ^{-1}(A^T\mu - q) + b = -C(u_\mu).$$

On the other hand,suppose that the constraints in (3.14) are equalities: Au=b; then the dual minimization is performed on the whole of R^P. If they exist, the dual solutions of $\nabla\theta(\mu) = 0$, which is the linear system

$$\nabla\theta(\mu) = AQ^{-1}(A^T\mu - q) + b = 0 \qquad (3.16)$$

Existence of such a solution implies first that $AQ^{-1}(A^T\mu - q) = -b$. Hence $-b \in \text{im}A$. Since im A is a subspace, $b \in$ im A this implies there exists a $u \in U$ such that Au = b. That is there exist a primal feasible u. In this case we claim that the dual problem has one solution at least,and that all such solutions make via (3.15) a unique point (the unique primal solution).

This will illustrate Corollary 7. The key lies in the following facts:

(I) The two subspaces im A^T and ker A are two orthogonal generators of R^n. This implies

(II)When applying the positive definite operator Q or Q^{-1} to one of them, we never obtain a vector in the other

$$Q^{-1}(\text{im } A^T) \cap \ker A = \{0\} = \text{im } A^T \cap Q(\ker A) \qquad (3.17)$$

Hence:

(III) We have further decompositions of R^n into two subspaces:

$$Q^{-1}(\text{im } A^T) \oplus \ker A = R^n = \text{im } A^T \oplus Q(\ker A) \qquad (3.18)$$

Then the proof our claim goes as follows:

Existence : If $b \in$ im A, take $u_0 \in R^n$ such that $b = -AQ^{-1}u_0$; then use (3.18) to write

$$q + u_0 = A^T\mu_0 + Qv_0, \text{ with } v_0 \in \ker A.$$

In summary (3.16) becomes

$$0 = AQ^{-1}(A^T\mu - q - u_0) = AQ^{-1}A^T(\mu - \mu_0) - Av_0$$

which has the obvious solution $\mu = \mu_0$:

Since

$$R^n = \text{im } A^T \oplus Q(\ker A)$$

$$q + u_0 = A^T \mu_0 + Q v_0 \qquad v_0 \in \ker A$$

$$0 = AQ^{-1}(A^T \mu - q) + -AQ^{-1} u_0$$

$$= AQ^{-1}(A^T \mu - q - u_0)$$

$$= AQ^{-1}[A^T \mu - (A^T \mu_0 + Q v_0)]$$

$$= AQ^{-1} A^T (\mu - \mu_0) - A v_0$$

Uniqueness: If μ_1 and μ_2 solve (3.16), with corrosponding u_1 and u_2 maximizing L over R^n.

$$L(u_1, \mu_1) = \sup\{L(u, \mu_1) : u \in U\},$$

$$L(u_2, \mu_2) = \sup\{L(u, \mu_2) : u \in U\}$$

$$AQ^{-1}(A^T \mu_1 - q) + b = AQ^{-1}(A^T \mu_2 - q) + b = 0.$$

Hence

$$AQ^{-1}(A^T \mu_1 - q) = AQ^{-1}(A^T \mu_2 - q).$$

We have

$$AQ^{-1} A^T (\mu_1 - \mu_2) = 0.$$

$$u_1 = Q^{-1}(q - A^T \mu_1)$$

$$u_2 = Q^{-1}(q - A^T \mu_2)$$

$$0 = AQ^{-1} A^T (\mu_1 - \mu_2)$$

$$= AQ^{-1} A^T \mu_1 - AQ^{-1} A^T \mu_2$$

$$= A[Q^{-1} A^T \mu_1 - Q^{-1} q + Q^{-1} q - Q^{-1} A^T \mu_2]$$

$$= A[-Q^{-1}((q - A^T \mu_1) + Q^{-1}(q - A^T \mu_2)]$$

$$= A[u_2 - u_1].$$

This implies $u_2 - u_1 \in \ker A$. But by (3.17)

$$u_2 - u_1 \in \ker A \cap . Q^{-1}(\text{im } A^T) = \{0\}.$$

We get $u_2 = u_1$

The essence of the above "proof" is the fact that $<\mu, \nu> := \mu^T A Q^{-1} A^T \nu$ is a scalar product in $\ker A^T = \text{im } A$.

When (3.16) has no solution, which means that the primal feasible set is empty, θ is not bounded from below.

$$C(U) = \{\gamma \in R^m \mid \gamma = Au \text{ for some } u \in U\}$$

Applying proposition 15(I) knowing that C(U) is a closed convex affine manifold.

4. CLASSICAL DUAL ALGORITHMS

In this section we review the most classical algorithms aimed at solving the dual problem (2.6). We must minimize a convex function (the dual functin θ). We assume that for each $\lambda \in R^m$, there is a $u_\lambda \in U$ such that

$$\theta(\lambda) = L(u_\lambda, \lambda) = \text{Sup}\{L(u, \lambda) : u \in U\}$$

4.1 SUBGRADIENT OPTIMIZATION

Algorithm 4.1.1 (Basic subgradient Algorithm).A sequence $\{t_k\}$ is given,with $t_k>0$ for k=1,2,....

Step 0. (Initialization) Choose $\lambda_1 \in R^n$ and obtain $s_1 \in \partial\theta(\lambda_1)$

Step 1. If $s_k = 0$, stop. Otherwise set

$$\lambda_{k+1} = \lambda_k - t_k s_k / \|s_k\| \tag{4.1}$$

Step 2. Obtain s_{k+1} $\in \partial\theta(\lambda_{k+1})$ Replace k by k+1 and loop to step 1

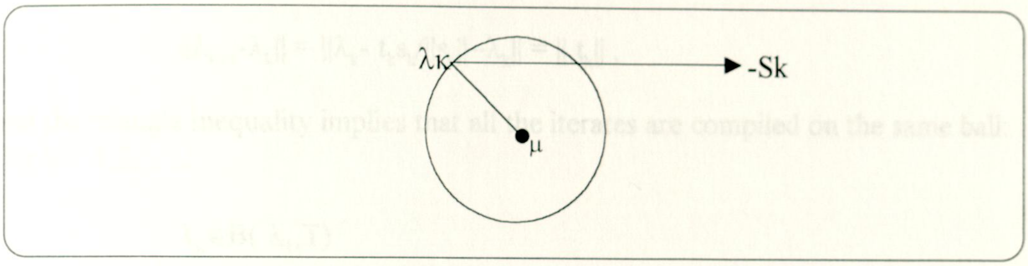


Fig 2. An antisubgradient gets closer to any better point

Each subgradient is obtained via some u_λ solving the Lagrange problem at the corresponding λ, since there exists $u_\lambda \in U$ such that

$$L(u_\lambda, \lambda) = \text{Sup}\{L(u,\lambda): u \in U\}.$$

Hence $-C(u_\lambda) \in \partial\theta(\lambda)$.

If

$$\theta(\mu) - \theta(\lambda_k) < 0,$$

then

$$\theta(\mu) \geq \theta(\lambda_k) + s_k^T(\mu - \lambda_k) \text{ for all } \mu,$$

since

$$s_k \in \partial\theta(\lambda_k).$$

This implies

$$s_k^T(\mu - \lambda_k) \leq \theta(\mu) - \theta(\lambda_k) < 0.$$

That is $< -s_k, \mu - \lambda_k > > 0$.

Let φ be the angle between the direction of move - s_k and the direction $\mu-\lambda_k$ then

$$\cos\varphi = \frac{\langle -s_k, \mu-\lambda_k\rangle}{||-s_k||\,||\mu-\lambda_k||} > 0 \qquad 0 \le \varphi \le \pi,$$

that is φ is acute.

If our move along the direction of $-s_k$ is small enough,we get closer to μ. From this interpretation,the step size t_k should be so small and we will require

$$t_k \to 0 \quad \text{for } k \to \infty \tag{4.2}$$

On the otherhand

$$||\lambda_{k+1}-\lambda_k|| = ||\lambda_k - t_k s_k/||s_k|| - \lambda_k|| = ||\, t_k||\,,$$

and the triangle inequality implies that all the iterates are compiled on the same ball: for k = 1,2,......

$$\lambda_k \in B(\lambda_1, T)$$

where

$$T = \sum_{k=1}^{\infty} t_k$$

since

$$\begin{aligned} ||\lambda_k-\lambda_1|| &= ||\lambda_k - \lambda_{k-1} + \lambda_{k-1} - \ldots + \lambda_2 - \lambda_1|| \\ &\le ||\lambda_k - \lambda_{k-1}|| + \qquad + ||\lambda_2 - \lambda_1|| \\ &= t_{k-1} + t_{k-2} + \ldots + t_1\,. \end{aligned}$$

This implies

$$||\lambda_k - \lambda_1|| \le \sum_{k=1}^{\infty} t_k$$

Hence

$$||\lambda_k - \lambda_1|| \le T$$

Thus $||\lambda_k-\lambda_1|| \le T$ for all k=1,2,.....

Therefore $\lambda_k \in B(\lambda_1, T)$. And moreover we will require

$$\sum_{k=1}^{\infty} t_k = \infty \tag{4.3}$$

LEMMA 1: Let $\{t_k\}$ be a sequence of positive numbers satisfying (4.2) and (4.3) and set

$$\tau_n := \sum_{k=1}^{n} t_k \qquad \rho_n := \sum_{k=1}^{n} t_k^2 \tag{4.4}$$

Then ($\tau_n \to \infty$ and) $\frac{\rho_n}{\tau_n} \to 0$ when $n \to \infty$

Proof: See Hiriat-urruty Vol II Convex Analysis And Minimization Algorithms

LEMMA 2: Let $\theta: R^m \to R$ be convex and fix $\nu \in R^m$. For all $\lambda \in R^m$ such that $\theta(\lambda) > \theta(\nu)$ for all $s \in \partial\theta(\lambda)$
Set

$$d(\lambda) := \frac{<s, \lambda - \nu>}{||s||} > 0 \tag{4.5}$$

Given M>0, there exists L>0 such that

$$d(\lambda) \le M \Rightarrow [0<] \ \theta(\lambda) - \theta(\nu) \le L d(\lambda)$$

Proof:- For $s \in \partial\theta(\lambda)$ we have $\theta(\mu) \ge \theta(\lambda) + <s, \mu - \lambda>$ for all $\mu \in R^m$. In particular for $\mu = \nu$ we get
$\theta(\nu) - \theta(\lambda) \ge <s, \nu - \lambda>$. Since $\theta(\lambda) > \theta(\nu)$, we have $<s, \lambda - \nu> > 0$.
That is $\frac{<s, \lambda - \nu>}{||s||} > 0$ and $s \ne 0$ if s=0, then in the above $\theta(\nu) \ge \theta(\lambda)$ which is a contradiction to $\theta(\lambda) > \theta(\nu)$. Let $\mu(\lambda)$ be the projection of λ onto the hyperplane:

$$H = \{ \mu \in R^m | <s, \mu - \lambda> = 0 \}$$

By calculation we get

$$||\mu(\lambda) - \nu|| = d(\lambda) \text{ and } \theta(\mu(\lambda)) \ge \theta(\lambda) + <s, \mu(\lambda) - \lambda> = \theta(\lambda) + 0 = \theta(\lambda)$$

That is $\theta(\mu(\lambda)) \ge \theta(\lambda)$.
If $d(\lambda) \le M$, then $\mu(\lambda) \in B(\nu, M)$. Since $\theta: R^m \to R$, we have $\text{dom}\theta = R^m$ and $\text{ridom}\theta = R^m$. Moreover $B(\nu, M)$ is closed and bounded subset of R^m, hence $B(\nu, M)$ is compact. In addition to this $B(\nu, M)$ is convex. Then (Theorem 3.1.2 of Chapter IV) there exists $L \ge 0$ such that

$$|\theta(\lambda) - \theta(\mu)| \le L|| \lambda - \mu|| \text{ for all } \lambda, \mu \in B(\nu, M).$$

Since $\mu(\lambda)$ and $\nu \in B(\nu, M)$, we have

$$| \theta(\mu(\lambda)) - \theta(\nu)| \le L|| \mu(\lambda) - \nu||.$$

$$\theta(\mu(\lambda) \ge \theta(\lambda) \text{ and } \theta(\lambda) > \theta(\nu)$$

implies

$$|(\mu(\lambda)) - \theta(\nu)| = \theta(\mu(\lambda)) - \theta(\nu)$$

Thus

$$0 < \theta(\mu(\lambda)) - \theta(\nu) \le L||\mu(\lambda) - \nu||.$$

hence we have

$$\theta(\lambda) \le \theta(\mu(\lambda)) \le \theta(\nu) + L||\mu(\lambda) - \nu||.$$

Therefore

$$\theta(\lambda)\leq\theta(\nu) + Ld(\lambda).$$

Finally we introduce the sequence of best values generated by Algorithm 4.1.1:

$$\bar{\theta}_k = \min\{\theta(\lambda_i): i=1,2,....,k\}$$

Which is needed because $\{\theta(\lambda_i)\}$ is not monotonic.The whole issue is whether these best function values tend to the infimum of θ over R^m.(a number in $R\cup\{-\infty\}$)

THEOREM 3: Apply Algorithm 4.1.1 to the convex function $\theta : R^m \rightarrow R$ and let the stepsizes satisfy (4.2),(4.3).Then

$$\bar{\theta}_k \rightarrow \inf\{\theta(\lambda) : \lambda \in R^m\} \quad \text{When } k\rightarrow \infty$$

Proof:-Assume for the contradiction the existence of $\mu \in R^m$ and $\varepsilon>0$ such that

$$[\theta(\lambda_k)\geq] \qquad \bar{\theta}_k \geq \theta(\mu) + \varepsilon \qquad \text{for all } k \qquad (*)$$

then

$$\begin{aligned}||\mu-\lambda_{k+1}||^2 &= ||\mu-\lambda_k + \lambda_k-\lambda_{k+1}||^2 \\ &= ||\mu-\lambda_k||^2 + 2(\mu-\lambda_k)^T(\lambda_k-\lambda_{k+1}) + ||\lambda_k-\lambda_{k+1}||^2 \\ &= ||\mu-\lambda_k||^2 - 2t_k\frac{(\mu-\lambda_k, -\lambda_k+\lambda_{k+1}>}{t_k} + ||\lambda_k-\lambda_{k+1}||^2 \\ &= ||\mu-\lambda_k||^2 - 2t_k d(\lambda_k) + t_k^2\end{aligned}$$

Where the notation (4.5) is used: the triple (μ,λ_k,s_k) enters the framework of Lemma 2.For $n\geq k$, set

$\delta_n := \min\{d(\lambda_k): k=1,2,...n\}$.Thus we have $-\delta_n \geq -d(\lambda_k)$,k=1,2,...,n.

Hence

$$\begin{aligned}||\mu-\lambda_{k+1}||^2 &= |\mu-\lambda_k||^2 - 2t_k d(\lambda_k) + t_k^2|| \\ &\leq ||\mu-\lambda_k||^2 - 2t_k\delta_n + t_k^2 \qquad \text{for } k=1,2,....,n\end{aligned}$$

Summing from 1 to n

$$\sum_{k=1}^{n} ||\mu-\lambda_{k+1}||^2 \leq \sum_{k=1}^{n} ||\mu-\lambda_k||^2 - \sum_{k=1}^{n} 2t_k\delta_n + \sum_{k=1}^{n} t_k^2$$

Hence $\delta_n \rightarrow 0$ by Lemma 1.

Thus we have an infinite subset K of integers such that

$\lim_{k\in K} d(\lambda_k)=0$.Now apply Lemma 2:$\{d(\lambda_k)\}$ where $k\in K$ is bounded and

$\lim_{k\in K} [\theta(\lambda_k)-\theta(\mu)] = 0$,which is a contradiction to (*),since $\{d(\lambda_k)\}$ is bounded then there exists

$M\in R$ such that $d(\lambda_k)\leq M$ for all $k\in K$.That is $\theta(\lambda_k)-\theta(\mu)\leq Ld(\lambda_k)$.

Hence

$$0\leq\lim_{k\in K}[\theta(\lambda_k)-\theta(\mu)]\leq\lim_{k\in K} Ld(\lambda_k)=0.$$

Then we get

$$\lim_{k\in K}[\theta(\lambda_k)-\theta(\mu)]=0.$$ But by our assumption $\theta(\lambda_k)-\theta(\mu) \geq \varepsilon$ for all k.

The above implies

$$\lim_{k\in K}[\theta(\lambda_k)-\theta(\mu)] \geq \varepsilon > 0$$ a contradiction.

Algorithm 4.1.1 is often called a subgradient method

4.2 THE BASIC CUTTING PLANE ALGORITHM

To ease our exposition, we assume now that a compact convex set $C \subseteq R^m$ is known to contain a dual solution. Then we can write our dual problem

$\inf\{\theta(\lambda): \lambda \in R^m\}$ as

$$\begin{cases} \min \theta, \theta \in R, \lambda \in C \\ \theta \geq L(u,\lambda) \quad \text{for all } u \in U \end{cases} \qquad (4\cdot6)$$

(4.6)

The problem would be easy if it had only finitely many constraints that is if U were a finite set.

ALGORITHM 4.2.1 (Basic Cutting Plane Algorithm)

The basic idea of the basic cutting plane algorithm is to accumulate the constraints one after the other in (4.6) and taking the advantage of the fact that the constraint index u can be restricted to the smaller set U^* of (2.9).

The compact convex set $C \neq \varnothing$ and the stopping tolerence $\varepsilon \geq 0$ are given:

Step 0 Choose $\lambda_0 \in C$ and solve the Lagrange problem at λ_0 to obtain $u_0 := u_{\lambda 0}$. Set k=0

Step 1 Solve the following relaxation of (4.6)

$$\begin{cases} \min \theta \quad \theta \in R, \lambda \in C \\ \theta \geq L(u,\lambda) \quad \text{for i=0,1,2,.....k-1} \end{cases} \qquad (4.7)$$

to obtain a solution (θ_k, λ_k)

Step 2 Solve the Lagrange Problem $(P_{\lambda k})$ to obtain a next primal point $u_k := u_{\lambda k}$

Step 3 f $\theta(\lambda_k) \leq \theta_k + \varepsilon,$ (4.8)

then stop. Otherwise replace k by k+1 and loop to step 1.

For the Lagrange function

$$L(u_i,\lambda)=\varphi(u_i)-\lambda^T C(u_i) = L(u_i,\lambda_i) + [-C(u_i)]^T(\lambda-\lambda_i) \qquad (4.9)$$

Letting $s(\lambda_i)=-C(u_i)$

we have

$L(u_i,\lambda)=\theta(\lambda_i) + [s(\lambda_i)]^T(\lambda-\lambda_i)$ valid for all ,λ and i=1,2,........

Thus (4.7) can be written

$$\begin{cases} \min \theta \quad \theta \in R, \ \lambda \in C \\ \theta \geq \theta(\lambda_i) + [s(\lambda_i)]^T(\lambda - \lambda_i) \quad \text{for } i=0,1,2,\ldots,k-1 \end{cases}$$

$$-C(u_i) = s(\lambda_i) \in \partial\theta(\lambda_i) \ . \tag{4.10}$$

When the hope that (4.10) does approximate (4.6), we rely upon on the property proved in theorem 1.3.8 of Chapter XI:

For all $\lambda \in R^m$, $\theta(\lambda) = L(u_\lambda, \lambda)$ where $u_\lambda \in U$ exists.

Hence

$$\mathrm{dom}\theta = R^m .$$

Therefore ri dom$\theta = R^m$ .Theorem 1.3.8 of Chap.XI: $\theta \neq \infty$ implies $\theta \in \mathrm{Conv}\ R^n$.

S: $R^m \to R^n$ satisfying $s(\lambda) \in \partial\theta(\lambda)$ for all $\lambda \in$ ri domθ.

Then

$$\theta(\lambda) = \sup_{\mu \in R^m} [\theta(\mu) + (s(\mu))^T(\lambda - \mu)$$

for all $\lambda \in R^m$ where $s(\mu)$ is an arbitrary element of $\partial\theta(\mu)$.

The convergence of Algorithm 4.2.1 is easy to establish:

Theorem 4:-With $C \subseteq R^m$ convex compact and $\theta : R^m \to R$ convex,consider the optimal value

$$\bar{\theta}_c := \min\{\theta(\lambda) : \lambda \in C\}$$

in Algorithm 4.2.1,$\theta_k \leq \bar{\theta}_c$ for all k and the following cnvergence properties hold:

-If $\varepsilon > 0$, the stop occurs at some iteration k_ε with λ_{k_ε} satisfying

$$\theta(\lambda_{K_\varepsilon}) \leq \bar{\theta}_c + \varepsilon \tag{4.11}$$

-If $\varepsilon = 0$,the sequences $\{\theta_k\}$ and $\{\theta(\lambda_k)\}$ tend to $\bar{\theta}_c$ when $k \to \infty$

Proof: Because (4.6) is less constrained than (4.7) =(4.10),the inequality $\theta_k \leq \bar{\theta}_c$ is clear and the optimality condition (4.11) hold when the stopping criterion (4.8) is satisfied.Now suppose for contradiction that ,for k=1,2,....

$$\theta(\lambda_k) - \varepsilon > \theta_k \geq \theta(\lambda_i) + [s(\lambda_i)]^T(\lambda_k - \lambda_i) \quad \text{for all } i<k$$

The above implies

$$-\varepsilon > \theta_k - \theta(\lambda_k) \geq \theta(\lambda_i) - \theta(\lambda_k) + [s(\lambda_i)]^T(\lambda_k - \lambda_i) \qquad \text{for all } i<k$$

or

$$-\varepsilon > \theta(\lambda_i) - \theta(\lambda_k) - \|s(\lambda_i)\| \|\lambda_k - \lambda_i\|,$$

since

C compact implies bounded ,θ convex and dom $\theta = R^m$ by theorem 3.1.2 of Chapter IV there exists $L \geq 0$ such that

$$|\theta(\mu) - \theta(\lambda)| \leq L \|\mu - \lambda\| \text{ for all } \mu, \lambda \in C$$

since $\lambda_i, \lambda_k \in C$

$$\|\theta(\lambda_k) - |\theta(\lambda_i)\| \leq L^1 \|\lambda_k - \lambda_i\| \text{ for } i<k=1,2,\ldots, \text{ since it holds for all.}$$

Let L= max $\{L^1, M\}$. Then

$$-L\|\lambda_k - \lambda_i\| \le -M\|\lambda_k - \lambda_i\| \le -\|s(\lambda_i)\|\|\lambda_k - \lambda_i\|$$

and

$$-|\theta(\lambda_k) - \theta(\lambda_K)| \ge -L\|\lambda_k - \lambda_i\|$$

From

$$-\varepsilon > \theta(\lambda_k) - \theta(\lambda_K) - \|s(\lambda_i)\|\|\lambda_k - \lambda_i\|$$

we get

$$-\varepsilon > -L\|\lambda_k - \lambda_i\| + -L\|\lambda_k - \lambda_i\| \quad \text{for all } i < k = 1,2,\ldots$$

or

$$-\varepsilon > -2L\|\lambda_k - \lambda_i\| \quad \text{for all } i < k=1,2,\ldots..$$

or

$$\varepsilon < 2L\|\lambda_k - \lambda_i\| \quad \text{for all } i < k=1,2,\ldots$$

or

$$\|\lambda_k\| + \|\lambda_i\| > \varepsilon/2L$$

or

$$\|\lambda_k\| > \varepsilon/2L - \|\lambda_i\| \quad \text{for all } i < k=1,2,\ldots$$

Let $\|c\| \le a$ for all $c \in C$, choose $\varepsilon=4aL + 1$

$$\|\lambda_k\| > \varepsilon/2L - \|\lambda_i\| \ge 2a + 1/2L - a = a + 1/2L$$

a contradiction.

Because $\varepsilon>0$ was arbitrary and $\theta_K \le \bar{\theta}_c \le \theta(\lambda_k)$ for all k .We have

$$\theta(\lambda_k) - \theta_K \to 0 \quad \text{as } k \to \infty$$

Which can happen only if $\varepsilon = 0$, t hen $\bar{\theta}_c$ is the common limit of $\{\theta_k\}$ and $\{\theta(\lambda_k)\}$.
If $\varepsilon=0$,Algorithm 4.2.1 $\theta(\lambda_k) \le \theta_K$ is not true, hence loops forever. When U is a finite set, then the algorithm stops any way ,even if $\varepsilon=0$.This is implied by the following result.

Proposition 5 Here no assumption is made on C in Algorithm 4.2.1.The u_k generated at step 2 is different from all points $u_0, u_1, \ \ , u_{k-1}$ unless the stop in step 3 is going to operate.

***Proof*:** If $u_k = u_i$ for some $i \le k-1$,then

$$\theta(\lambda_k) = L(u_k, \lambda_k) = L(u_i, \lambda_k) \le \theta_k \ .$$

Thus stop

The presence of a compact set C in the basic cutting plane Algorithm 4.2.1 is not only motivated by theorem 4.More importantly (4.10) would have no solution if C were unbounded.

Theorem 6 : Let U be convex, φ concave , C affine and suppose that Algorithm 4.2.1 applied to the convex function $\theta: R^m \rightarrow R$, can be used with (4.12) instead of (4.10).When the stop occurs denote $\beta \in \Delta_k$ an optimal solution of

$$\max_{\beta \in \Delta_k} \{ \sum_{i=0}^{k-1} \alpha_i \varphi(u_i) : \sum_{i=1}^{k} \alpha_i C(u_i) = 0 \tag{4.13}$$

Then $u(\beta)$ is an ε solution of the primal problem (P).

Proof: $\beta \in \Delta_k$ is an optimal solution of (4.13) implies

$$\sum_{i=0}^{k-1} \beta_i C(u_i) = 0$$

and

$$\sum_{i=0}^{k-1} \beta_i \varphi(u_i) \ge \sum_{i=0}^{k-1} \alpha_i \varphi(u_i) \qquad \text{for all } \alpha_i \text{ such that}$$

$$\sum_{i=0}^{k-1} \alpha_i = 1 \, , \, \alpha_i \ge 0$$

$$u(\beta) = \sum_{i=0}^{k-1} \beta_i u_i$$

Since C is affine ,we have $C(u(\beta)) = 0$ and

$$\theta_k = \sum_{i=1}^{k-1} \beta_i \varphi(u_i) \le \varphi(u(\beta)$$

Since $-\varphi$ is convex $\varphi(u(\beta)) \ge \theta_k$

When the stop occurs

$$\theta(\lambda_k \ ) \le \theta_k + \varepsilon$$

By weak duality theorem

$$\varphi(u(\beta)) \le \theta(\lambda_k \ ) \le \theta_k + \varepsilon$$

By the above

$$\theta_k \le \varphi(u(\beta))$$

Hence we have

$$\theta_k \le \varphi(u(\beta)) \le \theta_k + \varepsilon$$

This implies $u(\beta)$ is an ε solution of (P)

5 PUTTING THE METHOD IN PERSPECTIVE

5.1 THE PRIMAL FUNCTION

Here we consider the following function

$$P(\gamma) := -\sup(\varphi(u) : u \in U, C(u) = \gamma \} \qquad (5.1)$$

Theorem :- Suppose dom $\theta \neq \emptyset$.Then

$$P^*(\lambda) = \theta(-\lambda) \quad \text{for all } \lambda \in R^m$$

Proof: By definition

$$\begin{aligned} P^*(\lambda) &= \sup_{\gamma}\{\lambda^T \gamma - P(\gamma)\} \\ &= \sup_{\gamma}\{\lambda^T \gamma + \sup\{\varphi(u) : C(u) = \gamma, : u \in U\}\} \\ &= \sup_{\gamma}\{\sup[\lambda^T \gamma + \varphi(u) : C(u) = \gamma, u \in U] : \gamma \in R^m \} \\ &= \sup\{\lambda^T C(u) + \varphi(u) : u \in U\} \\ &= \theta(-\lambda) \end{aligned}$$

REMARK

Dom θ = $-$ dom P^* ; Dom $\theta \neq \emptyset$ implies $\theta \neq \infty$

Dom $\theta = -$ dom P^* is the set of slopes minorizing P. The condition in question :

for some $(\theta, \mu) \in R \times R^m$

$P(\gamma) \geq \theta + <\mu, \gamma>$ for all $\gamma \in R^m$

or $-P(\gamma) \leq -\theta - <\mu, \gamma>$ for all $\gamma \in R^m$

or $\sup\{\varphi(u) : C(u) = \gamma, u \in U]) \leq -\theta - <\mu, \gamma>$

or $\varphi(u) + \mu^T C(u) \leq -\theta$ for all $u \in U$

or $\theta(-\mu) \leq -\theta < \infty$

1.This property holds for example when φ is bounded from above on U.That is

$\varphi(u) \leq \alpha$ for all $u \in U$

or $\varphi(u) - 0C(u) \leq \alpha$ for all $u \in U$

or $\theta(0) \leq \alpha < \infty$

$f \neq \infty$,and there is an affine function minorizing f on R^n (*)

For f satisfying (*)

By Chapter X of Theorem 1.1.2 $f^* \in \overline{Conv R^n}$

$\theta(-\gamma) = p^*(\gamma)$ or $\theta^*(-\gamma) = p^{**}(\gamma)$ and $p^{**}(\gamma) = \overline{conv}\, R^n$ by chapter X (1.3.4)

Let U be a bounded subset of $\mathbf{R}^n$,

$$\varphi(u) = \langle q,u\rangle \quad \text{linear} \qquad (5.2)$$
$$C(u) = Au - b \quad \text{is affine}$$

Proposition 1:-Consider the dual problem associated (U,φ ,C) in (5.2).Its infimal value is the supremal value in

$$\text{Sup}\{<q,u>: u\in \overline{\text{Conv}}U\ ,\ Au-b=0\} \qquad (5.3)$$

Furthermore,assume that this dual problem has some optimal solution μ ;then the solutions of (5.3) are those $u\in(\overline{\text{Conv}}U)(\mu)=\{u\in \overline{\text{Conv}}U : L(u,\mu)=\theta(\mu)\}$ that satisfy Au=b.

Proof:-In the case of (5.2),we recognize in (5.1) the definition of an image function:

$$\begin{aligned} P(\gamma - b) &= -\text{Sup}\{\varphi(u) : u \in U ; C(u) = \gamma - b\} \\ &= \inf\{-\varphi(u) : u \in U ; Au - b = \gamma - b\} \\ &= \inf\{-\varphi(u) : u \in U ; Au = \gamma\} \\ &= \inf\{I_U(u) - <q , u> : Au = \gamma\} \quad \text{for all } \gamma \in \mathbf{R}^m \end{aligned}$$

Let us compute its biconjugate.Thanks to the boundedness of U,theorem 2.1.1 of Chapter X and using various Calculus rules in 1.3.1 of Chapter X, we obtain

$$P^*(\lambda)=(I_U-<q,.>)^*(A^*\lambda) +\lambda^T b =\sigma_U(A^*\lambda +q) +\lambda^T b \quad \text{for all } \lambda\in R^m.$$

The support function $\sigma_U = \sigma_{\overline{Conv}U}$ is finite everywhere;Theorem 2.2.1 of Chapter X applies and using again Chapter X of 1.3.1,we conclude

$$\begin{aligned} (\overline{\text{Conv}}\ P)(\gamma - b) &= \text{Inf}\{I_{\overline{\text{Conv}}U}(u) - <q , u> : Au = \gamma\} \\ &= -\ \text{Sup}\{<q , u> : u \in \overline{\text{Conv}}\ U , Au = \gamma\} \text{ for all } \gamma \in R^m \end{aligned}$$

To finish the proof,

$$\text{Sup}\{<q,u>: u\in \overline{\text{Conv}}\ U, Au -b=0\} .$$

$\overline{\text{Conv}}U$ is closed and bounded implies compact

φ is linear,C is affine implies L is upper semicontinuous,filling property holds at each $\lambda\in R^m$ and Corolary 14(ii) $\overline{\text{Conv}}U$ convex $-\varphi$ is convex, C is affine hence by Lemma 12 no duality gap.

5.2 AUGMENTED LAGRANGIANS

We saw in section 2 that an important issue was uniqueness of the Lagrange problem (P_λ)

$$U(\lambda) := \{u \in U : L(u , \lambda) = \theta(\lambda)\}$$
$$C(\lambda) := \{C(u) \in R^m : u \in U(\lambda)\}$$

More precisely,the important property was single valuedness of the multifunction $\lambda\rightarrow C(\lambda)$. Such a property would imply,under the filling property (2.12)

$$\partial\theta(\lambda) = -\text{Conv}\{C(u) : u \in U(\lambda)\}$$

For a given $t \geq 0$, the problem

$$\left| \begin{array}{ll} \text{Sup } [\varphi(u) - \frac{1}{2}t\|C(u)\|^2] & u \in \mathbf{U} \\ C(u) = 0 \in R^m & \end{array} \right. \tag{5.4}$$

is equivalent to (P): it has the same feasible set,and the same objective function there . The Lgrangian associated with (5.4) is

$$L_t(u,\lambda) := \varphi(u) - \frac{1}{2}t\|C(u)\|^2 - \lambda^T C(u) = L(u,\lambda) - \frac{1}{2} t\|C(u)\|^2$$

called the augumented Lagrangian associated with (P). $L = L_0$ is the "ordinary" Lagrangian.Corrospondingly,we have the "augumented dual function".

$$\theta_t(\lambda) := \text{Sup}\{ L_t(u,\lambda) : u \in \mathbf{U}\} \tag{5.5}$$

A mere application of the weak duality theorem to the primal dual pair(5.4),(5.5) gives

$$\theta_t(\lambda) \geq \varphi(u) - \frac{1}{2} t\|C(u)\|^2 = \varphi(u)$$

since u feasible implies $C(u) = 0$, for all t, all $\lambda \in R^m$ and all u feasible in (P). On the otherhand the inequality

$$L_t(u,\lambda) := L(u,\lambda) - \frac{1}{2}t\|C(u)\|^2 \leq L(u,\lambda) \quad \text{extends to} \quad \theta_t \leq \theta .$$

$$\begin{aligned} \text{Duality gap} &= \inf_\lambda \theta_t(\lambda) - \sup_{u \in U} \{ \varphi(u) - \frac{1}{2} t\|C(u)\|^2 : C(u) = 0\} \\ &= \inf_\lambda \theta_t(\lambda) - \sup_{u \in U} \{\varphi(u) : C(u) = 0\} \\ &\leq \inf_\lambda \theta(\lambda) - \sup_{u \in U} \{\varphi(u) : C(u) = 0\} \end{aligned}$$

This implies the augumented Lagrangian approach cannot worsen the duality gap.

Theorem 2 :With P defined by (5.1). Suppose there are $t_0 \geq 0$ and $\mu \in R^m$ such that

$$P(\gamma) \geq P(0) - \mu^T \gamma - \frac{1}{2} t_0\|\gamma\|^2 \quad \text{for all } \gamma \in R^m . \tag{5.7}$$

Then for all $t \geq t_0$, there is no duality gap associated with the augumented Lagrangian L_t and actually

$$-P(0) = \theta_t(\mu) \leq \theta_t(\lambda) \text{ for all } \lambda \in R^m$$

Proof:- When (5.7) holds,

$$P(\gamma) \geq P(0) - \mu^T \gamma - \frac{1}{2}t_0\|\gamma\|^2$$

$t \geq t_0$ implies $-t \leq -t_0$.

Then we get

$$-\frac{1}{2}t_0\|\gamma\|^2 \geq -\frac{1}{2}t\|\gamma\|^2.$$

Thus

$$P(\gamma) \geq P(0) - \mu^T \gamma - \frac{1}{2}t\|\gamma\|^2 \text{ for all } t \geq t_0,.$$

Then we have for all γ :

$$\begin{aligned}
-P(0) &\geq -P(\gamma) - \mu^T\gamma - \frac{1}{2}t_0\|\gamma\|^2 \\
&= \operatorname*{Sup}_{u\in U}\{\varphi(u) : C(u) = \gamma\} - \mu^T\gamma - \frac{1}{2}t_0\|\gamma\|^2 \\
&= \operatorname*{Sup}_{u\in U}\{\varphi(u) - \mu^T\gamma - \frac{1}{2}t_0\|\gamma\|^2 : C(u) = \gamma\} \\
&= \operatorname*{Sup}_{u\in U}\{L_t(u,\mu) : C(u) = \gamma\}
\end{aligned}$$

Since γ is arbitrary we conclude that

$$-P(0) \geq \operatorname*{Sup}_{u,\gamma}\{L_t(u,\mu) : C(u) = \gamma\} = \theta_t(\mu).$$

By the definition of

$$-P(0) = \operatorname*{sup}_{u\in U}\{\varphi(u) : C(u) = 0\}$$

$$\begin{aligned}
\text{Duality gap} &= \inf_{\lambda}\theta_t(\lambda) - \operatorname*{sup}_{u\in U}\{\varphi(u) - \frac{1}{2}t\|C(u)\|^2 : C(u) = 0\} \\
&= \inf_{\lambda}\theta_t(\lambda) - \operatorname*{sup}_{u\in U}\{\varphi(u) : C(u) = 0\} \\
&= \inf_{\lambda}\theta_t(\lambda) - (-P(0)) \\
&= 0 \ .
\end{aligned}$$

since $-P(0) \geq \theta_t(\mu)$.

Hence we have $-P(0) = \theta_t(\mu)$. That means

$$-P(0) = \theta_t(\mu) = \inf_{\lambda} \theta_t(\lambda).$$

This result is to be compared with our comment in section 5.1.(iii):

Corolary:- μ satisfies (5.7) iff it minimizes θ_{t_0}

Proof:-μ satisfies (5.7) implies

$$P(\gamma) \geq P(0) - \mu^T \gamma - \frac{1}{2}t_0\|\gamma\|^2 \text{ for all } \gamma \in R^m .$$

By the above theorem μ minimizes θ_{t_0} for all $\gamma \in R^m$

If μ minimizes θ_{t0} ,then $\theta_{t0}(\mu) \leq .\theta_t(\lambda)$ for all $\lambda \in R^m$ and no duality gap implies $\partial P(0) \neq \varnothing$

Therefore

$$P(\gamma) \geq P(0) + \langle \mu, \gamma - 0 \rangle \text{ for all } \gamma \in R^m.$$

or

$$P(\gamma) \geq P(0) - \langle -\mu, \gamma \rangle + \tfrac{1}{2} t_0 \|\gamma\|^2 \quad \text{for all } \gamma \in R^m,$$ where $t_0 = 0$.

Thus holds (5.7). The primal function associated with (5.4) is

$$-P_t(\gamma) := \sup_{u \in U} \ \varphi(u) - \tfrac{1}{2} t \|C(u)\|^2 : C(u) = \gamma\}$$
$$= -P(\gamma) - \tfrac{1}{2} t \|\gamma\|^2$$

and (5.7) just means ,

$$\partial P_{to}(0) \neq \emptyset$$

since

$$P(\gamma) \geq P(0) - \mu^T \gamma - \tfrac{1}{2} t_0 \|\gamma\|^2 \quad \text{for all } \gamma \in R^m$$

Implies

$$P(\gamma) + \tfrac{1}{2} t_0 \|\gamma\|^2 \geq P(0) + \langle -\mu, \gamma \rangle \quad \text{for all } \gamma \in R^m.$$

That is

$$P_{to}(\gamma) \geq P_{to}(0) + \langle -\gamma, \mu \rangle \text{ for all } \gamma \in R^m$$

since there exists $-\mu \in R^m$,by the definition of differential

$$\partial P_{to}(0) \neq \emptyset$$

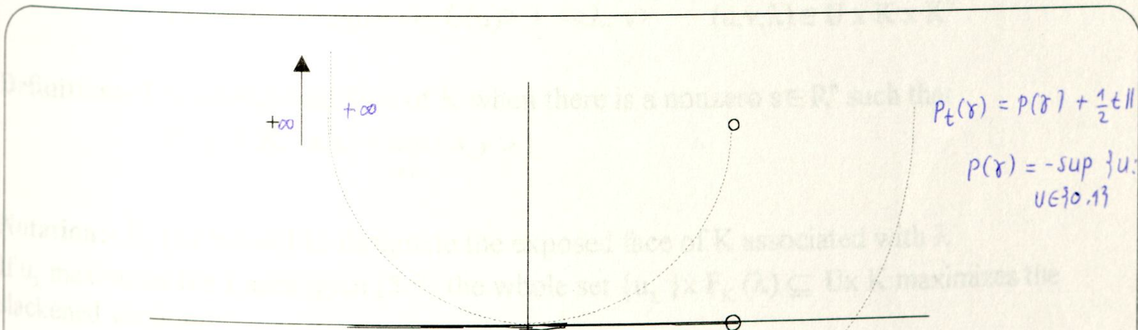
EXAMPLE

Consider the simple knapsack problem

$$\max u \qquad 2u \leq 1 \qquad u \in \{0,1\}$$

whose optimal value is 0.

The minimal value of θ is 0.5.The primal function is plotted below.The graph of P_t is obtained by bending upwards the graph of P that is by adding $\frac{1}{2} t\|.\|^2$; for $t \geq t_0 = 2$,the discontinuity at $\gamma = 1$ is lifted high enough to yield (5.7) with $\mu = 0$.In view of theorem 2,the duality gap is suppressed.



$$p(\gamma) = \begin{cases} -1 & if \quad \gamma \geq 1 \\ 0 & if \quad \gamma \in [-1, 1) \\ \infty & if \quad \gamma < -1 \end{cases} \quad ; \quad p_t(\gamma) = \begin{cases} -1 + \frac{1}{2}t\|\gamma\|^2 & if \quad \gamma \geq 1 \\ 0 + \frac{1}{2}t\|\gamma\|^2 & if \quad \gamma \in [-1, 1) \\ \infty & if \quad \gamma < -1 \end{cases}$$

Fig 3 The primal function in a knapsack problem

5.3 THE DUALIZATION SCHEME IN VARIOUS SITUATIONS

(a) CONSTRAINTS WITH VALUES IN A CONE

Consider abstractly a primal problem posed under the form

$$\begin{cases} \text{Sup } \varphi(u) \quad u \in \mathbf{U}, \\ C(u) \in K \end{cases} \tag{5.8}$$

In (P) we had $K=\{0\} \subseteq R^m$.

In (3.5) $C(u) \in -(R^+)^P$. More generally we take here for K a closed convex cone in some finite dimensional vector space call it R^m ,equipped wlth the scalar product<.,.> and K^0 will be the polar cone of K.

The Lagrange function is then defined as

$$L(u,\lambda) = \varphi(u) - \langle \lambda, C(u) \rangle \qquad (u,\lambda) \in \mathbf{U} \text{x } K^0 \tag{5.9}$$

and the dual function is $\theta = \text{Sup}\{L(u,.) : u \in \mathbf{U}\}$ as before.The dualization enters the frame work of section 3.1,as is shown by an easy check of (3.1)

$$\begin{aligned} \inf\{L(u,\lambda) : \lambda \in K^0\} &= \inf\{\varphi(u) - \langle \lambda, C(u) \rangle : \lambda \in K^0\} \\ &= \varphi(u) + \inf\{\langle -\lambda, C(u) \rangle : \lambda \in K^0\} \\ &= \varphi(u) - \text{Sup}\{\langle \lambda, C(u) \rangle : \lambda \in K^0\} \\ &= \varphi(u) - I_K(C(u)) \quad \text{by definition} \end{aligned}$$

The weak duality theorem

$\theta(\lambda) \geq \varphi(u)$ for all $(u,\lambda) \in \mathbf{U}\text{x}K^0$ with u feasible follows. Slack variables can be used in (5.8).

$$C(u) \in K \text{ if and only if } C(u) = v \in K$$

The slackened Lagrangian is

$$L(u,v,\lambda) := \varphi(u) - \langle \lambda, C(u) \rangle + \langle\langle \lambda, v \rangle \qquad (u,v,\lambda) \in \mathbf{U} \text{ x K x } K^0$$

Definition:- F is an exposed face of K when there is a nonzero $s \in R^n$ such that

$$F = \{x \in K : \langle s,x \rangle = \underset{y \in K}{Sup} \langle s, y \rangle\}$$

Notation:- $F_K(\lambda)$ is used to designate the exposed face of K associated with λ

If u_λ maximizes the Lagrangian (5.9) the whole set $\{u_\lambda\} \text{x } F_K(\lambda) \subseteq \mathbf{U}\text{x K}$ maximizes the slackened version.

Everett's theorem becomes : a maximizer u_λ of the Lagrangian (5.9) solves the primal problem

$$\underset{u \in U}{Sup}\{\varphi(u) : C(u) - C(u_\lambda) \in K \times F_k(\lambda)\}$$

If $C(u_\lambda) \in F_K(\lambda)$,the feasible set in this problem contains that of (5.8).Thus corollary 2 becomes: If some u_λ maximizing the Lagrangian is feasible and satisfies the complementary slackness $\langle\lambda, C(u_\lambda)\rangle=0$,then this u_λ is optimal in (5.8)

(b) PENALIZATION OF THE CONSTRAINTS

We replace (P) by

$$\mathrm{Sup}\{[\varphi(u)-\tfrac{1}{2}t\|C(u)\|^2, u \in \mathbf{U}\} \tag{5.10}$$

for a given parameter $t > 0$

Introduce the additional variable $v \in R^m$ and formulate (5.10) as

$$\begin{cases} \mathrm{Sup}\ [\varphi(u)-\tfrac{1}{2}t\|v\|^2], u \in \mathbf{U},\ v \in R^m \\ C(u) = v \end{cases} \tag{5.11}$$

Note the difference with (5.4):the extra variable v is free,while it was frozen to 0 in the case of the augmented Lagrangian.We can define the Lagrangian

$$L_t^*(u,v,\lambda):=\varphi(u)-\tfrac{1}{2}t\|v\|^2-\lambda^T[C(u)-v] \quad \text{for all } (u,v,\lambda) \in \mathbf{U} \times R^m \times R^m$$

The corosponding dual function

$$\theta_t^*(\lambda):= \mathrm{Sup}\{L_t^*(u,v,\lambda): (u,v)\in \mathbf{U}\mathbf{x}R^m\}= \theta(\lambda)+\tfrac{1}{2t}\|\lambda\|^2$$

Observe that θ_t^* always has a minimum (provided that $\theta\neq\infty$). Thus we have $\theta_t^* \neq\infty$.

$C(\mathbf{U})=R^m$ implies Conv $C(\mathbf{U})=R^m$

$0\in \mathrm{ri}\, R^m$ implies $\inf\{\theta_t^*(\lambda):\lambda\in R^m\}=\theta_t^*(\mu)$ there exists $\mu\in R^m$,

θ_t^* has a minimum always.

Consider the primal function associated with (5.11)

$$-P_t^*(\gamma):= \mathrm{Sup}\ \{\varphi(u)-\tfrac{1}{2}t\|C(u)-\gamma\|^2: u\in \mathbf{U}\} \tag{5.12}$$

Proposition 3: Let $t > 0$

$P(\gamma):=-\ \mathrm{Sup}\ \{\varphi(u): u\in \mathbf{U}:C(u)=\gamma\}$

$-P_t^*(\gamma) := \mathrm{Sup}\ \{\varphi(u)-\tfrac{1}{2}t\|C(u)-\gamma\|^2: u\in \mathbf{U}\}$

There holds

$$P_t^*(\gamma)=\inf\{P(z)+\tfrac{1}{2}t\|z-\gamma\|^2:z\in R^m\}$$

Proof:-Associativity of infima is much used.By the definition of P

$$\begin{aligned}\inf\{P(z)+\tfrac{1}{2}t\|z-\gamma\|^2:z\in R^m\} &=\inf_z[\inf_u\{\varphi(u):C(u)=z\}+\tfrac{1}{2}t\|z-\gamma\|^2\} \\ &=\inf_z[\inf_u\{\{-\varphi(u)+\tfrac{1}{2}t\|z-\gamma\|^2:C(u)=z\} \\ &=\inf_{z,u}\{-\varphi(u)+\tfrac{1}{2}t\|z-\gamma\|^2:C(u)=z\} \\ &=\inf_u[-\varphi(u)+\tfrac{1}{2}t\|C(u)-\gamma\|^2]=P_t^*(\gamma)\end{aligned}$$

(C) MIXED OPTIMALITY CONDITIONS FOR CONSTRAINED MINIMIZATION PROBLEM

Consider now a problem where both forms occur;

$$\left| \begin{array}{ll} \inf f(x) & x \in C_0 \\ Ax=b & \\ C_j(x) \leq 0 & \text{for } j=1,2,\ldots p \end{array} \right. \qquad (5.13)$$

Here $f \in \overline{\text{Conv}}\ R^m, C_0 \subseteq R^m, C_0 \neq \emptyset$ is a closed convex set $, C_0 \cap \text{Dom } f \neq \emptyset$

$A: R^n \to R^m$ linear

$C_j: R^n \to R$ is convex

Thus we now accept an extended-valued function , but the constraint functions are still assumed finite valued.

The problem enters the general frame work of the present chapter if we take the Lagrangian

$$L(x,\lambda,\mu) = f(x)+\lambda^T (Ax-b)+\mu^T C(x)$$

and the corrosponding dual function

$$\theta(\lambda,\mu) = -\inf\{L(x,\lambda,\mu), x \in C_0\}$$

The control space is now $U=C_0 \cap \text{dom } f$.The whole issue is then whether there is a dual solution,and whether the filling property (2.12) holds;altogether ,these properties will guarantee the existence of a saddle point,that is of a primal-dual pair.

In view of section 2.3 a relevant question is now the following:Suppose we have found a dual solution (λ,μ),can we construct a primal solution from it?for this we need the filling property.

5.4 FENCHEL'S DUALITY

For two closed functions g_1 ,g_2 ,the optimal value in the primal problem

$$m=\inf\{g_1(x)+g_2(x): x \in R^n\} \qquad (5.14)$$

is opposite to that in the "dual problem"

$$\inf\{ g_1^*(s)+g_2^*(-s): s \in R^n\} \qquad (5.15)$$

Under some appropriate assumption.For example m is a finite number and

$$\text{ri dom } g_1 \cap \text{ri dom } g_2 \neq \emptyset \qquad (5.16)$$

Since ri dom $g_1 \cap$ ri dom $g_2 \neq \emptyset$ .We have dom $g_1 \cap$ dom $g_2 \neq \emptyset$. Then by theorem 2.3.1 of Chapter X

$$(g_1^*+g_2^*)^*(0)=-\min\{ g_1^*(s)+g_2^*(-s): s \in R^n\} \quad \text{where } m=\inf\{g_1(x)+g_2(x): x \in R^n\}$$

Hence (5.15) has a solution.

The construction (5.14)$\to$(5.15) is called Fenchel's duality,whose startig idea is to conjugate the sum g_1+g_2 in (5.15)

Propositon 4:- Let g_1 and g_2 be functions of Conv R^n satisfying (5.16).If s is an arbitrary solution of (5.15),the (possibly empty) solution set of (5.14) is

$$\partial g_1^{\bullet}(s) \cap \partial g_2^{\bullet}(-s) \qquad (5.17)$$

Proof:- Let $f_1= g_1^*$ and $f_2= g_2^*$.The infimal convultion $g_1^* \underset{\vee}{+} g_2^*$ is exact at 0= s+-s, and the subdifferential

$$\partial(g_1^* \underset{\vee}{+} g_2^*)(0) = \partial g_1^* \cap \partial g_2^* \quad \text{by (3.4.5) of chapter XI}$$

$$(g_1+g_2)^* = g_1^* \underset{\vee}{+} g_2^* \quad \text{by theorem 2.3.2. of chapter X we get}$$

$$\partial(g_1 + g_2)^*(0) = \partial g_1^*(s) \cap \partial g_2^*(-s)$$

But

$$\partial(g_1 + g_2)^{\bullet}(0) = \min\{g_1(x) + g_2(x) : x \in R^n\} \text{ by 1.4.6 of Chapter X}$$

$$= \partial g_1^*(s) \cap \partial g_2^*(-s)$$

(a) FROM FENCHEL TO LAGRANGE

As was done in section 5.3 in different situations the approach of this chapter can be applied to Frenchel's duality.It suffices to formulate (5.14) as

$$\inf\{g_1(x_1) + g_2(x_2): x_1 - x_2 = 0\}$$

This is a minimization prblem posed in $R^n x R^n$,with constraint values in R^n, which lends itself to Lagrangian duality:taking the dual variable $\lambda \in R^n$,we form the Lagrangian

$$L(x_1,x_2,\lambda) = g_1(x_1) + g_2(x_2) + \lambda^T(x_1 - x_2)$$

The associated closed convex dual function (to be minimized)

$$\theta(\lambda) = -\inf_{x_1,x_2} [g_1(x_1) + g_2(x_2) + \lambda^T(x_1 - x_2)$$

$$= -[\inf_{x_1}[g_1(x_1) + \lambda^T x_1] + \inf_{x_2}[g_2(x_2) - \lambda^T x_2]$$

$$= -[\inf_{x_1}[g_1(x_1) + < \lambda, x_1 >] + -[\inf_{x_2}[g_2(x_2) - < \lambda, x_2 >$$

$$= g_1^*(-\lambda) + g_2^*(\lambda)$$

So

$$\theta(\lambda) = g_1^*(-\lambda) + g_2^*(\lambda)$$

(b) FROM LAGRANGE TO FENCHEL

Conversely suppose we would like to apply Frenchel duality to the problems encountered in this capter.This can be done at least formally,with the help of appropriate functions in (5.14)

(i) $g_2(x) = I_{\{0\}}(Ax-b)$ models affine equality constraints,as in convex instances of (P).

(ii) $g_2(x) = I_K(C(x))$,where K is a closed convex cone,plays the same role for the problem of section 5.3(a) inequality constraints corrospond to the non positive orthant K of R^P

(iii) $g_2(x) = \frac{1}{2} t ||Ax - b||^2$ is associated to penalized affine constraints as in section 5.3(b).

(iv) In the case of Augumented Lagrangian (5.4),we have a sum of three functions

$$\inf\{-\varphi(x) + \frac{1}{2} t||Ax-b||^2 + I_{\{0\}}(Ax-b): x \in R^n\}$$

Many other situations can be imagined,let us consider the case of affine constraints in some detail.

(c) QUALIFICATION PROPERTY

With g_1 and f_2 closed convex function. $A:R^n \rightarrow R^m$ linear and R^m is equipped with the usual dot product,consider the following form of (5.14):

$$\inf\{g_1(x)+f_2(Ax-b):x\in R^m\} \tag{5.18}$$

The role of the control set **U** is played by domg_1 and f_2 can be an indicator function;g_2 of (5.14) is given by

$$g_2(x):=f_2(Ax-b)$$

By using proposition (III) 2.1.12:

A linear implies A is affine.Let ri dom $g_1\cap$ri dom $g_2\neq\emptyset$,then there exists $x\in$ ri dom g_1 and $x\in$ri dom g_2=ri dom$f_2(Ay-b)$.$x\in$ri dom$f_2(Ay-b)$ implies $Ax-b\in$ri domf_2.

That is

$$\text{there exists } x\in \text{ri dom } g_1 \text{ such that } Ax-b\in \text{ri dom } f_2 \tag{5.19}$$

With $f_2=I_{\{0\}}$,hence $g_2(x)=I_{\{0\}}(Ax-b)$.Thus

dom $f_2=\{0\}$

ri dom $f_2=\{0\}$

U=dom g_1 is a convex set

Let B(x)=Ax-b.Hence B is an affine mapping.By proposition 2.1.12 of Chapter III,

ri[B(**U**)]=B[ri(**U**)]=A[ri(**U**)]-b. $f_2=I_{\{0\}}$ means $0\in A(\text{ri }[\text{dom}g_1])-b$ =ri[A(dom g_1)-b]

that is

$$0\in \text{ri }[A(\text{dom}g_1)-b]$$

Let C(**U**)=A(dom g_1)-b =A(**U**)-b since **U** is a convex set,A**U**-b is a convex set.Hence Conv(A(**U**)-b)=A(**U**)-b.By proposition 15(3) the dual problem has a solution

(d) PRIMAL DUAL RELATIONSHIPS

Proposition 4 characterizes the solutions of the primal problem(5.18)(if any):they are those x satisfying simultaneously

- $x\in\partial g_1^*(-A^T\lambda)$ with λ solving

$$\min\{f_2^*(\mu)+\theta(\mu):\mu\in R^m\}$$

:equivalently,$0\in\partial g_1(x)+A^T\lambda$, that is x minimizes the Lagrangian

$$L(.,\lambda)=g_1(x)+\mu^T(Ax-b)$$

- $x\in\partial g_2^*(A^T\lambda)$ with λ solving

$$\min\{f_2^*(\mu)+\theta(\mu):\mu\in R^m$$

and g_2^* is given by

$g_2^*(-s)=\min\{\{f_2^*(\lambda)+b^T\lambda : A^T\lambda=-s\}$

using the Calculus rule XI 3.3.1 , this means $Ax-b\in\partial f_2^*(\lambda)$

$$g_j^*(s)=\min\{\mu C_j^*(\tfrac{1}{\mu}s) : \mu > 0\} \quad \text{for } j=1,2,\ldots p.$$

Then (5.20) has actually two (groups of) minimization variables:$(s_1,s_2,\ldots,s_p)$ and $(\mu_1,\mu_2,\ldots\mu_p)$.We minimize with respect to $\{s_j\}$ first.The value (5.20) is the infimum over $\mu\in(R_*^+)^p$ of the function

$$\inf\{g_0^*(s_0)+\sum_{j=1}^{p} \mu_j C_j^*(\tfrac{1}{\mu_j}s_j) : \sum_{j=1}^{p} s_j = 0\}$$

This is the value at s=0 of the infimal convolution

$$g(s):=(g_0^* \underset{\vee}{+} \pi_1^* \ldots \underset{\vee}{+} \pi_p^*)(s) \quad \text{Where } s\in R^n$$

and where for j=1,2,..p,

$$\pi_j^*(s):=\mu_j C_j^*(\tfrac{1}{\mu_j}s) \text{ where } s\in R^n \text{ is the conjugate of}$$
$$\pi_j(s):=\mu_j\ C_j\ (x)\ ,$$

since

$$\pi_j^*(s)=\operatorname{Sup}_s\{\langle x,s\rangle-\mu_j\ C_j(s)\}$$
$$=\mu_j \operatorname{Sup}_s\{\langle \tfrac{1}{\mu_j}x, s\rangle - C_j(s)\}$$
$$=\mu_j C_j^*(\tfrac{1}{\mu_j}x)$$

Theorem 2.3.2 of Chapter X says

$$g(s)=(g_0^* \underset{\vee}{+} \pi_1^* \ldots \underset{\vee}{+} \pi_p^*)(s)$$
$$=(g_0+\pi_1+\ldots+\pi_p)^*(s)$$

By Chapter X (1.4.6) we have

$$(g_0+\pi_1+\ldots+\pi_p)^*(0)= -\inf\{g_0+\pi_1+\ldots+\pi_p)(x): x\in R^n\}$$
$$= -\inf\{g_0(x)+\sum_{j=1}^{p} \mu_j C_j(x) : x \in R^n\}$$

Let us conclude :there is a two way corrospondence between Lagrange and Fenchel duality schemes,eventhough they start from different primal problems;the difference is a matter of taste.The Lagrangian approach may be deemed more natural and flexible;in particular,it is often efficient when the initial optimization problem contains variables,say $\gamma_j=C_j(x)$,which one wants to single out for some reason.On the otherhand,Fenchel's approach is often more direct in theoretical developments.

References

1 Hiriart-urruty, J.B., Lemarechal, C.: Convex Analysis and Minimization Algorithms volume I and II Springer Verlag Berlin, Heidelberg, Newyork, 1993.

2. Deumlich, R., Optimization and Theory of Approximation, Addis Ababa University