AD-A055 155 TEXAS UNIV AT AUSTIN CENTER FOR CYBERNETIC STUDIES F/G 12/1

A DUAL OPTIMIZATION FRAMEWORK FOR SOME PROBLEMS OF INFORMATION --ETC(U)

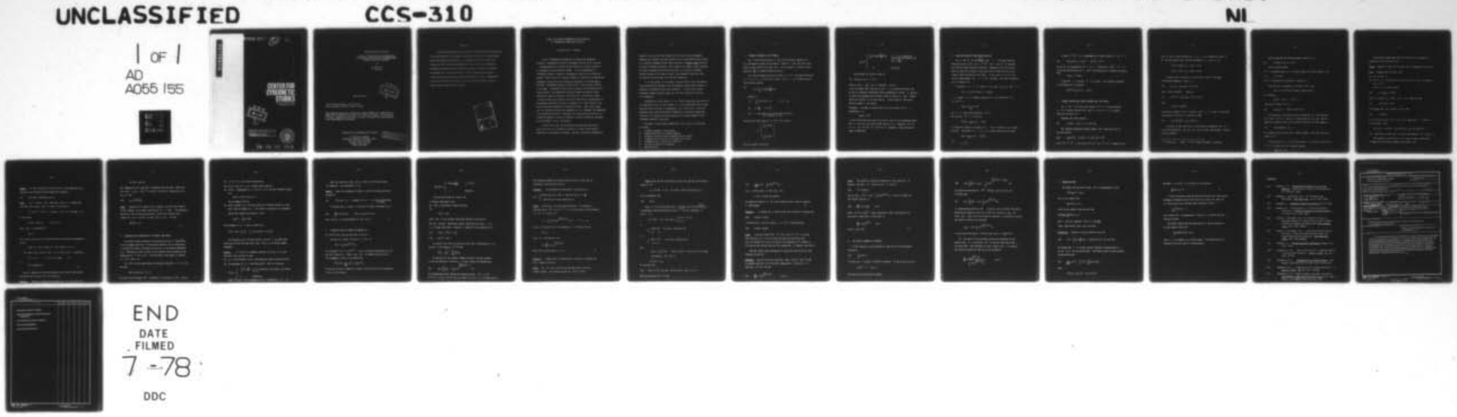
NOV 77 A BEN-TAL, A CHARNES N00014-75-C-0569

UNCLASSIFIED CCS-310 NL

1 OF 1

AD

A055 155



END

DATE

FILMED

7 -78

DDC

12 2

See 1473 in back

CENTER FOR CYBERNETIC STUDIES

The University of Texas
Austin, Texas 78712

Research Report CCS 310

A DUAL OPTIMIZATION FRAMEWORK FOR SOME PROBLEMS OF INFORMATION THEORY AND STATISTICS

by

A. Ben-Tal*
A. Charnes**

November 1977

*The Technion Institute, Haifa, Israel.
**The University of Texas, Austin, Texas.

This research was partly supported by Project NR047-021, ONR Contract N00014-75-C-0569 with the Center for Cybernetic Studies, The University of Texas. Reproduction in whole or in part is permitted for any purposes of the United States Government.

CENTER FOR CYBERNETIC STUDIES

A. Charnes, Director
Business-Economics Building, 203E
The University of Texas
Austin, Texas 78712
(512) 471-1821

Abstract

A new dual optimization framework for some problems of information theory and statistics is developed in the form of dual convex programming problems and their duality theory. It extends the work of Charnes and Cooper for finite discrete distributions to the case of general measures. Although the primal problem (constrained relative entropy) is an infinite dimensional one, the dual problem is a finite dimensional one without constraints and involving only exponential and linear terms. Applications range from mathematical statistics and statistical mechanics to traffic engineering, marketing and economics.

A DUAL OPTIMIZATION FRAMEWORK FOR SOME PROBLEMS OF INFORMATION THEORY AND STATISTICS

A. Ben-Tal and A. Charnes

In his "Mathematical Foundations of Statistical Mechanics" Khinchin[1] introduced the notion of conjugate function as the solution to the maximization of relative entropy subject to a single constraint on the mean of the distribution sought. In their developments of information theoretic methods in mathematical statistics, Kullback and Leibler[2] made a basis for treating problems of statistical estimation and hypothesis testing which was extensively developed in the monograph of Kullback "Information Theory and Statistics"[3]. More recently, Akaike in his paper "Information Theory and an Extension of the Maximum Likelihood Principle" has emphasized the great breadth and depth of these information theoretic methods by indicating their application to many classes of statistical problems, and also including the representation of the maximum likelihood principle as asymptotic, for large samples, to the decision theoretic approach of information theory. Again, more recent work in irreversible statistical mechanics by B.O. Koopman[4] has emphasized the importance and analytic convenience of a constrained entropy (or information) approach in deducing important statistical mechanics phenomena with a minimum of *ad hoc* hypothesis.

In all of this work the extremization problem has been solved explicitly only (as in Khinchin's case) for a single linear equality constraint in non-negative variables. Not until the work of Charnes and

Cooper[5,6], has the fact been brought out that dual convex programming problems are involved, and that the dual of the constrained entropy problem is in terms of exponential and linear functions in unconstrained variables. The work of Charnes and Cooper, while tying in the method to other problems of traffic engineering and economics, as well as providing a complete characterization of the duality states, has encompassed explicitly only the case of finite discrete distributions (measures).

It is the purpose of this paper to extend these results and develop a dual optimization framework that can adequately handle these classes of problems of information theory and statistics. In particular we have developed a complete duality theory for the case of general as well as finite measures.

Although our primal problem is an infinite dimensional one (with finitely many constraints) the dual problem is a finite dimensional one, without constraint and involving only exponential and linear terms. As we show elsewhere, such a dual optimization framework with convenient analytical functions in an unconstrained dual seems to be a unique property of the information theoretic functional.

The paper is conveniently summarized by the titles of its sections as follows:

1. A formal statement of the problem.
2. Some preliminaries from Convex Analysis.
3. Linearly constrained convex programs and their duals.
4. Conjugates and subgradient of integral functionals.
5. A complete duality theory for problem (A).
6. The case of probability measures.
7. Generalizations.

1. A FORMAL STATEMENT OF THE PROBLEM

Let T be an arbitrary set, F the σ-field of Borel subsets of T, dt a non-negative regular Borel measure (rBm) on T and $M(T)$ the linear space of real-valued finite rBm's on T. For an element $\mu \in M(T)$ we shall denote by $\frac{d\mu}{dt}$ its Radon-Nikodym derivative.

For a given summable positive function $c: T \to R$; continuous functions $F_i: T \to R$ $(i = 1,\dots,m)$ and real scalars $\theta_i (i = 1,\dots,m)$ we seek to solve the following problem

(A)
$$\inf \int_T u(t)\log[\frac{u(t)}{c(t)}]dt$$

subject to

(1) $\int_T u(t)F_i(t)dt = \theta_i \qquad i = 1,\dots,m.$

(2) $u = \frac{d\mu}{dt}, \quad \mu \in M(T)$

(3) μ non-negative and absolutely continuous (with respect to dt).

Consider the linear operator $A: M(T) \to R^m$ given by

$$\mu \longrightarrow \begin{bmatrix} \int_T F_1(t)d\mu \\ \vdots \\ \int_T F_m(t)d\mu \end{bmatrix}$$

and the integral functional

$$J(\mu) \triangleq \begin{cases} \int_T u(t)\log[\frac{u(t)}{c(t)}]dt & \text{if } \mu \text{ is an absolutely continuous non-negative rBm and } u = \frac{d\mu}{dt}. \\ \infty & \text{otherwise.} \end{cases}$$

Then Problem (A) can be written as

(P) $\inf\{J(\mu): A\mu = \theta,\ \mu \in S\}$

where $\theta = (\theta_1,\dots,\theta_m)^T$ and $S = \mathrm{dom}J \triangleq \{\mu : J(\mu)<\infty\}$.

It will be shown later (Section 5) that J is a convex functional, and so (P) is a *linearly constrained convex optimisation problem.* In Section 3 we study such programs and introduce a duality theory for them. Before doing so we collect in the next section certain material from Convex Analysis needed in the sequel.

Throughout the paper we assume that the linear system $A\mu = \theta$ is *irreducible*, i.e.

$$\text{Range } A = R^m.$$

In the finite dimensional case (A is an m×n matrix) this assumption means that A is of full row rank so that none of its m equations $A^i\mu = \theta_i$ $(i = 1,\dots,m)$ (A^i the i-th row of A) is redundant. Hence the terminology "irreducible".

2. SOME PRELIMINARIES FROM CONVEX ANALYSIS

Let E and E^* be real vector spaces, and $<,>$ a bilinear function defined on pairs (x,x^*), $x \in E$, $x^* \in E^*$. Let E and E^* be equipped with locally convex Hausdorff topologies, compatible with the bilinear form, so that every element of one space can be identified with a continuous linear functional on the other. In this case E and E^* are called *paired spaces* and $<,>$ is the *pairing*. (For more information see [7, Chapter IV].

A function $f: E \to R$ is *convex* if for every $x_1, x_2 \in E$ and $0 < \lambda < 1$

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y),$$

f is *proper* if it is bounded below and it is not identically $+\infty$.
If, for all x,

$$f(x) = \liminf_{z \to x} f(z)$$

then f is *lower semi-continuous* (l.s.c).
The function $f^*: E^* \to R$ given by

$$f^*(x^*) = \sup_x\{<x,x^*> - f(x)\}$$

is called the (convex) *conjugate of* f. This is always a l.s.c. convex function. Conversely if f is a l.s.c. proper convex function then

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

i.e.

$$f = f^{**}.$$

A vector $x^* \in E^*$ is a *subgradient of* a convex function f *at* x if

(4) $\quad f(z) \geq f(x) + \langle z-x, x^* \rangle \quad$ for all $z \in E$.

The set of all subgradients of f at x is denoted by $\partial f(x)$. If f is finite and differentiable at x, with $f'(x)$ denoting its (Frechet) derivative, then,

$$\partial f(x) = \{ f'(x) \}.$$

A function g is *concave* if $-g$ is convex. The (concave) conjugate of such a function is defined by

$$g^*(y^*) \triangleq \inf_y \{\langle y^*, y \rangle - g(y)\}.$$

3. LINEARLY CONSTRAINED CONVEX PROGRAMS AND THEIR DUALS

Let E and F be real vector spaces, $A: E \to F$ a linear operator, $h: E \to R$ a convex function with $\mathrm{dom}\, h = S$ and $g: F \to R$ a concave function with $\mathrm{dom}\, g = B$.

Consider the primal problem

(I) $\quad \inf\{h(x) - g(Ax) : x \in S, Ax \in B\}.$

The *Fenchel-Rockafellar duality theory* [8] associates with (I) the dual problem

(II) $\quad \sup\{g^*(x^*) - h^*(A^*x^*) : x^* \in B^*, A^*x^* \in S^*\}$

where $A^*: F^* \to E^*$ is the adjoint of A, E^* and F^* are the spaces paired with E and R (with the pairing $\langle\cdot,\cdot\rangle_E$, $\langle\cdot,\cdot\rangle_F$) respectively and h^*, g^* are the convex (resp. concave) conjugates of h and g, i.e.

$$h^*(\cdot) = \sup\{\langle x,\cdot\rangle_E - h(x) : x \in S\}$$

$$g^*(\cdot) = \inf\{\langle y,\cdot\rangle_F - g(y) : y \in B\}.$$

The main result concerning (I) and (II) is that if the *super-consistency assumption* holds, i.e.

(5) $\quad \exists\, x \in S$ such that $Ax \in \operatorname{int} B$

then $\inf(\mathrm{I}) = \max(\mathrm{II})$. Dually if

(5') $\quad \exists\, x^* \in S^*$ such that $A^*x^* \in \operatorname{int} B^*$.

Then

$$\min(\mathrm{I}) = \sup(\mathrm{II})^{\dagger}.$$

Furthermore, whenever $\min(\mathrm{I}) = \max(\mathrm{II})$, a pair $\bar{x}$, x^* solves (I) and (II) respectively if and only if (see [8], p.185).

(6) $\quad \bar{x} \in \partial h^*(A^*x^*)$, $A\bar{x} \in \partial g^*(x^*)$.

It should be mentioned that in the absence of assumption (5), (5') or similar assumption, one still has the so called *weak duality* relation: $\inf(\mathrm{I}) \geq \sup(\mathrm{II})$.

---

† We write 'min' ('max') if the infimum (supremum) is attained.

Here we consider the following special case of (I):

(P) $\inf\{h(x) : Ax = b,\ x \in S\}$

and we further assume:

(i) E is a Banach space, F is a Hilbert space (with inner product $\langle , \rangle$),

(ii) $\operatorname{int} S \neq \emptyset$,

(iii) ["irreducibility assumption"] Range $A = F$.

Note that (P) corresponds to (I) with $B = \{b\}$ and

$$g(\cdot) = \delta(\cdot|B) \overset{\Delta}{=} \text{the indicator function of } B.$$

This implies:

$$g^*(x^*) = \langle b, x^* \rangle , \quad B^* = F^* ;$$

therefore, the dual of (P) is

(D) $\sup\{\langle b, x^* \rangle - h^*(A^* x^*) : A^* x^* \in S^*\}$.

Unfortunately, the superconsistency assumption (5) does not hold here since $\operatorname{int} B = \emptyset$. However, we shall make use of a less familiar regularity condition ([9], p.50), which for the pair (P)-(D) reduces to

(7) $0 \in \operatorname{core}(A(S) - b)$.

This condition also implies that $\inf(P) = \max(D)$. We recall that for a subset $Q \subset F$

$$\operatorname{core} Q \overset{\Delta}{=} \{q \in Q : \forall v \in F,\ \exists \varepsilon > 0 \text{ such that } q + \lambda v \in Q \text{ for all } \lambda \in [-\varepsilon, \varepsilon]\}.$$

If Q is a convex set with nonempty interior

$$\operatorname{core} Q = \operatorname{int} Q.$$

The following lemma shows that the irreducibility assumption is essential for the validity of (7).

Lemma 1. Regularity condition (7) holds only if assumption (iii) holds.

Proof. Suppose that (7) holds, then

(a) $\exists x \in S \ni Ax = b$;

(b) $\forall v \in F$, $\exists \varepsilon > 0$ such that, for every $\lambda \in [-\varepsilon, \varepsilon]$, $\exists x \in S$ satisfying $Ax + \lambda v = b$.

Since F is a Hilbert space

(8) $$F = \text{Range } A + N(A^*),$$

where N denotes null space, thus if (iii) does not hold

(9) $$\exists \bar{v} \in N(A^*),\ \bar{v} \neq 0.$$

At the same time , (8) implies, by (a), that

(10) $$b \perp N(A^*).$$

Let $\bar{x}$ be a solution of $Ax + \lambda\bar{v} = b$ for some given $\lambda > 0$ (such $\bar{x}$ exists by (b)). Now

$$0 < \lambda \langle\bar{v},\bar{v}\rangle = \langle\bar{v}, b - A\bar{x}\rangle = \langle\bar{v}, b\rangle - \langle A^*\bar{v}, \bar{x}\rangle = 0 \text{ by (9) and (10).}$$

This contradiction shows that (iii) must hold whenever (7) is valid. □

The regularity condition (7) is not easy to check, therefore we introduce in the following lemma a much simpler one.

<u>Lemma 2</u>. If (P) satisfies (i) (ii) and (iii), then condition (7) is implied by the following *strict feasibility* assumption:

(ii) $\exists x \in \text{int}S$ satisfying $Ax = b$.

<u>Proof</u>. Let $\hat{x}$ satisfy (ii). Note that, since S is convex and $\text{int}S \neq \emptyset$, $\text{int}S = \text{core}S$. Now $\hat{x} \in \text{core}S$ if and only if

$$\hat{x} + \lambda x \in S \quad \text{for all} \quad x \quad \text{and all} \quad \lambda \in [-\varepsilon\ \varepsilon] \quad \text{for some } \varepsilon > 0.$$

In particular

$$\bar{x} \overset{\Delta}{=} \hat{x} + \lambda x(v) \in S \qquad \forall \lambda \in [-\varepsilon, \varepsilon]$$

where $x(v)$ is a solution of

$$Ax = v.$$

(That such a solution exists follows from the irreducibility assumption.) Further

$$b - A\bar{x} = b - A(\hat{x} + \lambda x(v)) = b - A\hat{x} + \lambda Ax(v) = 0 + \lambda v.$$

The latter shows that for every $v \in F$, there exist x satisfying

$$Ax + \lambda v = b \qquad \lambda \in [-\varepsilon, \varepsilon] \quad \text{for some} \quad \varepsilon > 0,$$

i.e.

$$0 \in \text{core}(A(S) - b).$$

□

We will summarize the results concerning the linearly constrained problem (P) and its dual (D) in the following

<u>Theorem 1</u> Consider problem (P) and assume that (i), (ii) and (iii)

$$\inf\,(P) = \max\,(D).$$

(2) Whenever min (P) = max (D), a necessary and sufficient conditions for a pair $x_o \in S$, $x_o^* \in S^*$ to solve (P) and (D), respectively, are $Ax_o = b$ and

(12) $$x_o \in \partial h^*(A^* x_o^*) .$$

<u>Proof</u>. Condition (11) implies (7), by Lemma 2, and the latter implies $\inf(P) = \max(D)$ by the above cited result [9 , p. 50]. The optimality conditions (12) are the specialization of (6) to our special case. Indeed, for $g(\cdot) = \delta(\cdot|\{b\})$ we have $g^*(\cdot) = \langle b,\cdot\rangle$ so that

$$\partial g^*(x^*) = \{b\} .$$

□

4. CONJUGATES AND SUBGRADIENTS OF INTEGRAL FUNCTIONALS

The duality theory presented in the previous section is stated in terms of the conjugate function of the objective function, and its subgradient. In Problem (A) however, the objective function is the *integral functional* $J(\mu)$. Therefore, we shall collect in this section results concerning the computation of J^* and $\partial J(\cdot)$. For more details the reader is referred to [10] and [11] .

Let $C(T)$ be the vector space of continuous functions $x: T \to R$ with the norm

$$\|x\| = \max\{|x(t)| : t \in T\}.$$

We recall that the space $M(T)$ of Section 1 is the dual of $C(T)$. Further let $f: T \times R \to R$ be a function satisfying

(a) $\forall t \in T$, $f(t,\cdot)$ is l.s.c. proper convex function.

(b) $f(t,x)$ is measurable in t for all $x \in R$ and the (nonempty convex) set

$$D(t) = \{x \in R: f(t,x) < \infty\}$$

has a nonempty interior.

(c) $D(t)$ is *fully l.s.c.* (see [10,p.457] this condition holds e.g. when $D(t)$ does not depend on t. The latter is enough for our purposes).

Define the integral functional on $C(T)$:

$$I_f(x) = \int_T f(t,x)dt.$$

The conjugate of I_f is then, by definition

$$I^*(\mu) = \sup \{ \int_T x d\mu - \int_T f(t,x(t))dt: x \in C(T)\}.$$

The following result follows directly from [10]. We remark that conditions (a) and (b) above imply that $f(t,x)$ is, so-called, *normal integrant*.

Lemma 3 If $f(t,x)$ is a summable function, for every $x \in R$, and satisfies (a), (b) and (c) then

(A) I_f is well-defined, finite, continuous and convex function on $C(T)$;

(B) the conjugate of I_f is the function $I^*: M(T) \to R$ given by:

$$I^*(\mu) = \begin{cases} \int_T f^*(t, \frac{d\mu}{dt}) & \text{if } \mu \text{ is absolutely continuous with respect to } dt \\ \infty & \text{otherwise,} \end{cases}$$

where $f^*(t,x^*)$ is the conjugate of $f(t,\cdot)$ evaluated at x^*. □

Next, one can derive from [10, Cor. 4B], the following formula for computing the subgradient of I_f.

<u>Lemma 4</u> Under the assumption of Lemma 3, and the following additional assumption

(d) $\{f(t,u(t)+x)$ is summable for all x in some neighbourhood of zero,

it follows that $\mu \in \partial I_f(u)$ if and only if almost everywhere (a.e.)

(14) $\frac{d\mu}{dt} \in \partial f(t,u(t))$. almost everywhere (a.e)

Here $\partial f(t,u)$ is the subgradient of $f(t,\cdot)$ at u. □

5. A COMPLETE DUALITY THEORY FOR PROBLEM (A)

We return to the setting described in Section 1.

Consider the integral functional $I: C(T) \to R$,

$$I(x) = \int_T c(t)e^{x(t)-1}dt .$$

The integrand $f(t,x) = c(t)^{x(t)-1}$ clearly satisfies assumptions (a), (b) and (d) of Section 4. (Recall that $c(t)$ is summable and positive.) The conjugate of $f(t,\cdot)$ is by definition

$$f^*(t,x^*) = \sup_{x \in R} \{xx^* - c(t)e^{x-1}\} .$$

The sup can be easily computed by equating the derivative of the supremand to zero, so one obtains

$$f^*(t,x^*) = \begin{cases} x^* \log \left[\dfrac{x^*}{c(t)}\right] & \text{if } x^* \geq 0 \\ \infty & \text{otherwise} \end{cases}$$

(we use the convention $0 \log 0 = 0$).

It follows from Lemma 3 that

(A) $I(x)$ is a continuous convex functional

(B)

$$I^*(\mu) = J(\mu)$$

where $J(\mu)$ is the integral functional defined in Section (1).

The last relation immediately implies (see Section 2) that J is a l.s.c. convex functional. Moreover, in regard of the continuity of I

(15) $$J^*(\mu) = I^{**}(u) = I(u)$$

(16) $$\operatorname{dom} J^* = C(T).$$

We further note that the adjoint of the linear transformation A of Section 1 is the mapping $A^*: R^m \to C(T)$

$$(x_1^*, \dots, x_m^*)^T \to \sum_{i=1}^m x_i^* F_i(t).$$

We have now all the elements needed to specify the dual problem of (A), as described in Section 3. In fact, problem (D) becomes here

(B) $$\sup \left\{ \sum_{i=1}^m x_i^* \theta_i - \int_T c(t) e^{\sum_{i=1}^m x_i^* F_i(t) - 1} dt \right\},$$

an *unconstrained finite dimensional* concave program. (The relation $A^* x^* \in S^*$ is here $x^* \in R^m$ since, by (16), $S^* = C(T)$, i.e. the whole space.)

The relations between the primal problem (A) and its dual (B) are expressed in the following results.

<u>Theorem 2</u> The supremum of Problem (B) is attained only if

$$(17)\quad \begin{cases} \exists \text{ positive (a.e.) rBm } \mu, \text{ whose derivative } u = \dfrac{d\mu}{dt} \\ \text{satisfies the linear equations (1).} \end{cases}$$

<u>Proof</u>. Since (B) is an unconstrained problem, its supremum is attained (say at $\bar{x}^*$) only if $\bar{x}^*$ is a critical point of the supremand. I.e. $\bar{x}^*$ is a solution of

$$\theta_i - \int_T c(t)F_i(t)e^{\Sigma \bar{x}_i^* F_i(t)-1}\, dt = 0 \qquad i = 1,\ldots,m$$

which, by the definition of the mapping A, is nothing else but

$$A\bar{\mu} = \theta$$

where $\bar{\mu}$ is the measure with

$$\bar{u} = \frac{d\bar{\mu}}{dt} = c(t)e^{\Sigma \bar{x}_i^* F_i(t)-1} .$$

Hence (17) is satisfied by $\mu = \bar{\mu}$. □

<u>Theorem 3</u>. Problem (B) is bounded above if and only if Problem (A) has a feasible solution.

<u>Proof</u>. The "if" part follows from the weak duality relation $\inf(A) \geq \sup(B)$. We proceed to prove the "only if" part.

Suppose that (A) has no feasible solution and consider the following subset of R^m

$$K = \{y \in R^m : \ y = A\mu \quad \text{for some } \mu \in M(T) \text{ satisfying (3)} \} .$$

By our assumption then

(19) $\quad \theta \notin K.$

Since K is a closed convex cone, it follows from (19) that there exists a hyperplane, passing through the origin, which strictly separates θ and K. I.e.

$$\exists\, z \in R^m, \ z \neq 0 \ \text{ such that } \begin{cases} z'y \leq 0 & \forall y \in K \\ z'\theta > 0 \end{cases}$$

or

$$\begin{cases} z'(A\mu) \leq 0 & \forall \mu \in M(T) \text{ satisfying (3)} \\ z'\theta > 0 \end{cases}$$

or

$$\begin{cases} \langle A^* z, \mu \rangle \leq 0 & \forall \mu \in M(T) \text{ satisfying (3)} \\ z'\theta > 0 \end{cases} .$$

Now

$\langle A^* z, \mu \rangle = \int_T (A^* z) d\mu$ and the latter is non-positive for *every* non-negative rBm only if

$A^* z \leq 0$ (a.e).

We conclude that

(20) $\quad \exists 0 \neq z \in R^m$ such that $A^* z \leq 0$ (a.e.) and $z'\theta > 0$.

Note that Problem (B) is in fact

$$(21) \qquad \sup_{x \in R^m} \left\{ x'\theta - \int_T c(t) e^{A^* x - 1} dt \right\} .$$

Let z be the vector in (20), then, with

$$x = Mz \quad (M \text{ positive scalar}),$$

the objective function in (21) can be made arbitrary large by choosing M large enough. □

<u>Theorem 4</u> If Problem (A) is feasible then the infinimum is attained and

$$(22) \qquad \min(A) = \sup(B).$$

If Problem (A) is strictly feasible, i.e. (17) is satisfied, then

$$(23) \qquad \min(A) = \max(B).$$

<u>Proof.</u> Since for Problem (A): $S^* = C(T)$ and $B^* = R^m$ it follows that condition (5') trivially holds, and hence the conclusion (22). Now, for Problem (A), the strict feasibility assumption (11) reduces to (17) and so (23) follows from the first conclusion in Theorem 1 and (22). □

The last result gives the optimality conditions for the pair of dual Problems (A) and (B).

<u>Theorem 5</u>. Let (A) be strictly feasible. Then $\bar{\mu} \in M(T)$ and $\bar{x}^* \in R^m$ are optimal solution of (A) and (B), respectively, if and only if $\bar{\mu}$ satisfies (1), (2), (3) and

$$(24) \qquad \frac{d\bar{\mu}}{dt} = c(t) e^{\Sigma \bar{x}_i^* F_i(t) - 1} \qquad \text{(a.e.)} \ .$$

Proof. The optimality conditions expressed in the second part of Theorem 1 are here: (i) feasibility of $\bar{\mu}$ and (ii)

(25) $\bar{\mu} \in \partial J^*(Ax_o^*)$.

Now, by (15), $J^*(u) = I(u) = \int_T c(t)e^{u(t)-1}dt$ so that, by Lemma (4), (25) holds if and only if

(26) $\frac{d\bar{\mu}}{dt} \in \partial f(t,Ax_o^*)$ (a.e.)

where $f(t,u) = c(t)e^{u-1}$ whose subgradient simply coincides with its derivative. Hence (26) is equivalent to

$$\frac{d\bar{\mu}}{dt} = c(t)e^{A^*\bar{x}^*-1} \quad \text{(a.e.)}$$

which is just (24). □

6. THE CASE OF PROBABILITY MEASURES

In this section we study problems of type (A) with the additional constraint

(27) $\int_T u(t)dt = 1$

in which case μ becomes a probability measure. Of course one can write

$$F_{m+1}(t) \equiv 1, \quad \theta_{m+1} = 1$$

and derive the following dual problem:

$$(28) \qquad \sup \{ \sum_{i=1}^{m} x_i^* \theta_i + x_{m+1}^* - \int_T c(t) e^{\sum_{i=1}^{m} x_i^* F_i(t) + x_{m+1}^* - 1} dt \}$$

an unconstrained problem in R^{m+1}. However, one can derive the following dual

$$(B') \qquad \sup \{ \sum_{i=1}^{m} x_i^* \theta_i - \log \int_T c(t) e^{\Sigma x_i^* F_i(t)} dt \}.$$

an *unconstrained problem in* R^m. To derive (B') from (28), one merely maximizes the objective function in (28) with respect to x_{m+1}^* for fixed $(x_1^*, \ldots, x_m^*)$, which can be done by equating its derivative to zero. The analytic solution thus obtained

$$x_{m+1}^* = -\log \int_T c(t) e^{\Sigma x_i^* F_i(t) - 1} dt$$

is then substituted again in (28) and the result is indeed (B').

All the result of the previous section can be applied to the present case. If, in condition (17), we add the requirement that u satisfies (27), then Theorems 2,3, and 4 remain valid. In Theorem 5 the optimality condition (24) has to be replaced by

$$\frac{d\bar{\mu}}{dt} = \frac{c(t) e^{\Sigma \bar{x}_i^* F_i(t)}}{\int_T c(t) e^{\Sigma \bar{x}_i^* F_i(t)} dt} \quad . \qquad \text{(a.e.)}$$

7. GENERALIZATIONS

The reason that the dual Problem (B) is *unconstrained* is that

$$S^* \triangleq \operatorname{dom} J^* = C(T).$$

This in turn holds since

$$\operatorname{dom} f^*(t,\cdot) = R ;$$

which in turn results from the fact that

(29) $\text{Range}\ \frac{d}{dx} f(t,x) = R.$

Here f was the integrand $f(x,t) = x \log \frac{x}{c(t)}$.

These observations lead to the following:

<u>Proposition</u> Consider a primal problem of the form

$$\text{(30)} \qquad \inf \{ \int_T f(t, \frac{d\mu}{dt})dt: \ \mu \text{ satisfies (1), (2), and (3)} \}$$

And assume that f is a normal convex integrand, differentiable in x for all $t \in T$ and satisfying (29). Then Problem (30) has the following unconstrained dual:

$$\text{(31)} \qquad \sup_{x^* \in R^m} \{ \theta' x^* - \int_T f^*(t, \sum_{i=1}^m x_i^* F_i(t))dt \} \ .$$

where

$$f^*(t,y) = y\Gamma(t,x^*) - f(t,\Gamma(t,x^*))$$

and where $x = \Gamma(t,x^*)$ is a solution of the equation

$$\frac{d}{dx} f(t,x) = x^* . \qquad \square$$

If $f(x,t)$ satisfies condition (a)-(d) of Section (4), then results analogous to Theorems 4 and 5 are valid for the dual pair (30)-(31). It is also easy to see from the proof of Theorem 3 that, if

$$\lim_{y\to-\infty} f^*(t,y) < \infty ,$$

then Problem (31) is bounded above if and only if, Problem (30) has a feasible solution.

We finally remark that the constraints (1), can be replaced by the more general constraint

$$\int_T u(t)K(t,s)dt = b(s),$$

where b is an element of an Hilbert space. The generalization of Theorems 2-5 to this case is straightforward.

REFERENCES

[1] Khinchin, A.I., Mathematical Foundation of Statistical Mechanics (translated by G. Gamov), Dover Publications Inc., New York, 1949.

[2] Kullback, S. and Leibler, R.A., "On Information and Sufficiency" Ann. Math. Stat., Vol.22 (1951) 79-86.

[3] Kullback, S., Information Theory and Statistics, Wiley, 1959.

[4] Koopman, B.O., "Relaxed Motion in Irreversible Molecular Statistics", Stochastic Processes in Chemical Physics, R. Schuler (ed.), Wiley, 1969.

[5] Charnes, A. and Cooper, W.W., "An Extremal Principle for Accounting Balance of a Resource-Value Transfer Economy: Existence, Uniqueness and Computations," Rend. Acad.Naz. Lincei, April 1974, 556-561.

[6] Charnes, A. and Cooper, W.W., " Constrained Kullback-Leibler Estimation; Generalized Coble-Douglas Balance, and Unconstrained Convex Programming", Rend. Acad. Naz. Lincei, April 1975, 568-576.

[7] Bourbaki, N., Espaces Vectoriels Topologiques, Hemann & Cie., Paris, 1955.

[8] Rockafellar, R.T., "Duality and Stability in Extremum Problems Involving Convex Functions", Pacific J. Math., 21, No.1, 1976, 167-186.

[9] Rockafellar, R.T., Conjugate Duality and Optimization, SIAM Regional Conference Series in Appl. Math., No.16, 1974.

[10] Rockafellar, R.T., "Integrals which are Convex Functionals, II", Pacific J. Math., 39, No.2, 1971, 439-468.

[11] Rockafellar, R.T., "Convex Integral Functionals and Duality". Contributions to Nonlinear Functional Analysis, E.H. Zarantonello, Ed., Academic Press, New York, 1971.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Center for Cybernetic Studies The University of Texas	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

6 A Dual Optimization Framework for Some Problems of Information Theory and Statistics.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

9 Research rept.,

5. AUTHOR(S) (First name, middle initial, last name)

10 A. Ben-Tal
A. Charnes

6. REPORT DATE 11 November 1977 12 27p.	7a. TOTAL NO. OF PAGES 24	7b. NO. OF REFS 11
8a. CONTRACT OR GRANT NO. 15 N00014-75-C-0569 b. PROJECT NO. NR047-021 14 CCS-310 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Center for Cybernetic Studies Research Report CCS 310 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Research (Code 424) Washington, D.C.

13. ABSTRACT

A new dual optimization framework for some problems of information theory and statistics is developed in the form of dual convex programming problems and their duality theory. It extends the work of Charnes and Cooper for finite discrete distributions to the case of general measures. Although the primal problem (constrained relative entropy) is an infinite dimensional one, the dual problem is a finite dimensional one without constraints and involving only exponential and linear terms. Applications range from mathematical statistics and statistical mechanics to traffic engineering, marketing and economics.

406197

CL

Unclassified

Security Classification

KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
Information Theory Duality						
Khinchin-Kullback-Leibler Statistical Estimation						
Dual Relative Entropy Problems						
Statistical Mechanics						
Mathematical Statistics						

DD FORM 1 NOV 65 1473 (BACK)

S/N 0102-014-6600

Unclassified

Security Classification