Research Report CCS 261

EXTREME PRINCIPLES AND
OPTIMIZATION DUALITIES FOR
KHINCHIN-KULLBACK-LEIBLER
ESTIMATION

by

A. Charnes
W.W. Cooper*
L. Seiford

April 1976

*Graduate School of Business, Harvard University.

This research was partly supported by Project No. NR 047-021, 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 purpose 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

This paper presents a new extremal approach to deriving dual optimization problems with proper duality inequality which simplifies and generalizes the Fenchel-Rockafellar scheme. Our derivation proceeds in two stages, (i) inequality attainment, (ii) decoupling primal and dual variables. The power and convenience of this approach are exhibited through a new, much simpler derivation of the Charnes-Cooper results for Khinchin-Kullback-Leibler statistical estimation [1], the immediate establishment of the $C^2$ duality for general distributions and its extensions to general linear inequality constraints, plus the development of a new two-person zero-sum game connection.
0. Introduction

In [1] Charnes and Cooper established a duality with three mutually
exclusive and collectively exhaustive (MECE) duality states for constrained
$K^2L$ (Khinchin-Kullback-Leibler) statistical estimation with finite discrete
distributions. Their dual problems are

(I) \[
\max \nu(\delta) = -\delta^T \theta_m \left( \frac{\delta}{e^c} \right)
\]
with $\delta^T A = b^T$, $\delta \geq 0$
and

(II) \[
\inf \xi(z) = c^T e A z - b^T z
\]

Extensions to general distributions were developed in [2] by Ben-Tal and
Charnes using function space settings and some generalized Fenchel-Rockefellar
duality theory [3]. The "F-$R$" theory was also employed by Ben-Tal and
Charnes in [4] to develop a general class of non-linear inequality constrained
convex programming problems with unconstrained duals as in the above
"C$^2$" duality.

This paper presents a new extremal approach to deriving dual optimization
problems with proper duality inequality which simplifies and generalizes
the Fenchel-Rockafellar scheme. Our derivation proceeds in two stages,
(i) inequality attainment, (ii) decoupling primal and dual variables. The power
and convenience of this approach are exhibited through a new, much simpler
derivation of the Charnes-Cooper results for Khinchin-Kullback-Leibler
statistical estimation [1], the immediate establishment of the C$^2$ duality for
general distributions and its extensions to general linear inequality constraints,
plus the development of a new two-person zero-sum game connection.

The latter dualities as well as extensions to general classes of convex
functionals a la [4] are being developed elsewhere by Charnes, Cooper,
Seiford and Palacios.
1. The Extremal Principle

We restrict the presentation here to convex programming dualities. (The general case involves only the obvious deletion of the convexity (concavity) requirements.) There are two aspects to the principle. The first is the achievement of the duality inequality. The second is "decoupling" the variables of the dual problems.

Let \( K(\delta, x) \) be a real-valued function concave in \( \delta \) for \( (\delta, x) \in \Delta \otimes X \subseteq \mathbb{R}^m \times \mathbb{R}^s \). This function is chosen so that \( g(\delta) = \inf_{X} K(\delta, x) \) exists for each \( \delta \in \Delta \). (Thereby \( g(\delta) \) is a concave function of \( \delta \).)

For convexity re \( x \) and then decoupling, let \( x = T(z) \) map \( Z \), a convex set in \( \mathbb{R}^n \), into \( X \). Let \( K(\delta, x) = f(z) \), a convex function, for \( z \in Z \), \( \delta \in \Gamma \).

Thus \( \Gamma \otimes T(Z) \) is the decoupling set for \( (\delta, x) \). For convexity in \( \delta \) we require \( \Delta \otimes \Gamma \) to be a convex set.

Clearly, then, the duality inequality is established

\[
(1) \quad g(\delta) \leq f(z), \quad \forall \delta \in \Delta \cap \Gamma, \forall z \in Z
\]

Further, the problems

(III) \( \sup g(\delta), \quad \delta \in \Delta \cap \Gamma \)

and

(IV) \( \inf f(z), \quad z \in Z \)

are dual convex programming problems. Existence or nonexistence of duality gaps is clearly dependent on the choices of \( \Delta \cap \Gamma \) and \( Z \).
2. The Dual Problems for Discrete $K^2$-L Estimation

Consider the function of two real variables $\delta, x$,

$$ K(\delta, x) = ce^x - \delta x, \text{ where } c, \delta \geq 0. $$

Then for $c > 0$ and $\delta \geq 0$,

$$ (2.1) \quad g(\delta) = \inf_{x} K(\delta, x) - \delta \ln \left( \frac{\delta}{e c} \right), $$

where $e$ is the base of the Napierian logarithms, and we set $\ln(0) = 0$.

Note that if $c = 0$ with $\delta \geq 0$, then we must restrict $x$ above, e.g. $x \leq 0$, in order that $g(\delta)$ exist.

Staying with $c_i > 0$ and introducing $\delta_i, x_i \geq 0$, we have

$$ (2.2) \quad v(\delta) = -\sum_i \delta_i \ln \left( \frac{\delta_i}{e c_i} \right) \leq \sum_i (c_i e^{x_i} - \delta_i x_i), $$

for all $\delta_i \geq 0$ and all $x$.

Next, suppose

$$ (2.3) \quad x_i = \sum_j a_{ij} z_j \quad \text{and} \quad \sum_i \delta_i a_{ij} = b_j, $$

for constants $a_{ij}$ and $b_j$.

Then, in vector-matrix notation,

$$ (2.4) \quad v(\delta) \leq \varphi(z) = c^T e^{A^T z} - b^T z $$

for $\delta \geq 0$, $\delta^T A = b^T$ and $z$ unconstrained.

Thus (I) and (II) are established as dual convex programs satisfying the duality inequality (2.4).

Further, they are established for the general discrete distribution case where the $\{x_i\}$, $\{\delta_i\}$, $\{z_j\}$ are sequences for which the sums in (2.2), (2.3) exist. For example, we could have $x, \delta, z$ in the sequence space $(\ell_1)$ and thereby, $A$ a linear transformation from $(\ell_1)$ into $(\ell_1)$. 
3. The Charnes-Cooper Duality States

Clearly, from (2.4), \( \xi(z) \) is bounded below provided that \( \delta \geq 0, \delta^T A = b^T \) has a solution. Suppose it does not. Consider the dual l.p. problems:

\[
\begin{align*}
(3.1) \quad & \max b^T z \quad \text{and} \quad \min \delta^T \cdot 0 \\
& \quad \text{subject to } Az \leq 0 \quad \delta^T A = b^T \quad \delta \geq 0
\end{align*}
\]

By the extended dual theorem [5], since \( z = 0 \) satisfies \( Az \leq 0 \), there exists a sequence \( z^n \) such that \( Az^n \leq 0 \) and \( b^T z^n \to \infty \). But then,

\[
\xi(z^n) = c^T e A z^n - b^T z^n \leq c^T e - b^T z^n \to -\infty
\]

Thus we have established duality state 1 e.g.:

\( \inf \xi(z) = -\infty \) iff \( \delta \geq 0, \delta^T A = b^T \) has no solution.

Next, suppose \( \xi(z) \) has a minimum. Since \( \xi(z) \) is differentiable and convex and \( z \) is unconstrained, the gradient must vanish at \( z^* \), a minimum, e.g.

\[
(3.2) \quad \frac{\partial \xi(z)}{\partial z^*_j} = \sum_i c^*_i e \sum_{j=1}^{n} a^*_{ij} z^*_j a_{ij} - b_j = 0, \quad j = 1, \ldots n
\]

But set \( \delta^*_i = c^*_i e \sum_{j=1}^{n} a^*_{ij} z^*_j \). Then \( \delta^*_i > 0 \) for all \( i \) and \( \delta^*^T A = b^T \).

Note that for this choice

\[
(3.3) \quad v(\delta^*) = -\sum_i \delta^*_i \cdot \text{e} \left( \delta^*_i \text{e} \right) = \sum_i \delta^*_i - \sum_i \delta^*_i \text{e} \left( \delta^*_i \text{e} \right)
\]

\[
= c^T e A z^* - b^T z^* = \min \xi(z)
\]

Hence \( v(\delta^*) = \max v(\delta), \delta \geq 0, \delta^T A = b^T \), and \( \min \xi(z) = \max v(\delta) \).

Conversely, suppose \( \delta \geq 0, \delta^T A = b^T \) has a solution \( \delta > 0 \). Then

\[
(3.4) \quad \xi(z) = \sum_i \left\{ c^*_i e \sum_{j=1}^{n} a^*_{ij} z^*_j - \text{e} \sum_{j=1}^{n} a_{ij} z^*_j \right\}
\]
But the function of one variable \( x \), with \( c_i, \delta_i > 0 \),

\[
(3.5) \quad \mathcal{P}_i(x) = c_i e^{\delta_i x} - \delta_i x \rightarrow +\infty \text{ as } x \rightarrow \pm \infty
\]

Hence if \( z_i^n \) is a sequence for which

\[
(3.6) \quad \mathcal{P}(z^n) \rightarrow \inf_z \mathcal{P}(z)
\]

we cannot have \( iAz^n = \sum_j a_{ij}z_j^n \rightarrow \pm \infty \) with \( n \rightarrow \infty \).

Since the \( iAz^n \) are thereby bounded, we can choose a subsequence, again denoted \( z^n \), so that

\[
(3.7) \quad iAz^n \rightarrow a_i, \quad \text{all } i
\]

Applying the Farkas-Minkowski Closure Corollary (FMC²) [5] we conclude there exists \( z^* \) such that \( iAz^* = a_i, \) all \( i \).

But then

\[
(3.8) \quad \inf \mathcal{P}(z) = \lim_{n \rightarrow \infty} \mathcal{P}(z^n) = c^T a - \delta^T a = \mathcal{P}(z^*)
\]

i.e. \( \mathcal{P}(z) \) has a minimum at \( z^* \).

Thus state 3 is established:

\( \mathcal{P}(z) \) has a minimum iff \( \delta^T > 0, \delta^T a = b^T \) has a solution and \( \min \mathcal{P}(z) = \max v(\delta) \).

By exhaustion, then, state 2 is established:

\( \mathcal{P}(z) \) has only an infimum iff \( \sum_i \delta_i = 0 \) in every solution to \( \delta \geq 0, \delta^T a = b^T \).

To obtain the further properties in state 2, we note first that since there exists \( \delta \geq 0, \delta^T a = b^T \), we can write

\[
(3.81) \quad \mathcal{P}(z) = \sum_i \left[ c_i e^{(iAz)} - \delta_i (iAz) \right]
\]
Now suppose \( \zeta(z^n) \to \inf \zeta(z) \) and that (w.l.o.g.) \( (iA z^n) \to a_i, \ i \in D, \ (rA z^n) \to -\infty, \ r \notin D. \) (Clearly \( \delta_r = 0 \) for \( r \notin D \)).

By the FMC\(^2\), there exists \( \hat{z} \) such that \( iA \hat{z} = a_i, \ i \in D, \ rA \hat{z} \leq 1, \ r \notin D. \)

Then \( iA(z^n-\hat{z}) \to 0, \ i \in D \) and \( rA(z^n-\hat{z}) \to -\infty, \ r \notin D. \) By the FMC\(^2\), there exists \( \tilde{z} \) such that \( iA \tilde{z} = 0, \ i \in D, \ rA \tilde{z} \leq 1 \).

Then \( iA(\tilde{z} + p\tilde{z}) = a_i, \ i \in D, \ rA(\tilde{z} + p\tilde{z}) \to -\infty \) as \( p \to \infty \).

Thus \( \zeta(z + p\tilde{z}) \to \inf \zeta(z) \) as \( p \to \infty \).

But

\[
(3.82) \quad \zeta(z + p\tilde{z}) = \sum_{i \in D} \{ c_i e^{iA\tilde{z}} - \delta_i (iA\tilde{z}) \} + \sum_{r \notin D} c_r e^{rA(\tilde{z} + p\tilde{z})}
\]

and

\[
(3.83) \quad \lim_{p \to \infty} \zeta(z + p\tilde{z}) = \inf \zeta(z) = \sum_{i \in D} \{ c_i e^{iA\tilde{z}} - \delta_i (iA\tilde{z}) \} = \sum_{i \in D} c_i e^{iA\tilde{z}} - b^T \tilde{z} = \zeta_D(\tilde{z})
\]

Now \( \zeta_D(z) = \sum_{i \in D} \{ c_i e^{iA\tilde{z}} - \delta_i (iA\tilde{z}) \} \) has a minimum at say \( \bar{z} \), since \( \delta_i > 0, \ i \in D \) and \( \sum_{i \in D} \delta_i (iA) = b^T \).

Thus \( \zeta_D(z) \leq \zeta_D(\bar{z}) \). If \( \zeta_D(\bar{z}) < \zeta_D(\tilde{z}) \), then

\[
\lim_{p \to \infty} \zeta(z + p\tilde{z}) = \zeta_D(\tilde{z}) < \zeta_D(\bar{z}) = \inf \zeta(z), \ a \ contradiction.
\]

So \( \zeta_D(\bar{z}) = \min \zeta_D(z) = \inf \zeta(z) \).

Then \( v(\delta) = \inf \zeta(z) \) where \( \delta_i = \sum_{i \in D} c_i e^{iA\tilde{z}}, \ i \in I \)

Hence \( v(\delta) = \max v(\delta) \).

Thus a \( \delta^* \) exists and by uniqueness \( \delta^* = \delta \).

The further properties of state 2 are thereby established, e.g.:

\[
(3.9) \quad \inf \zeta(z) = \max v(\delta) = \min \zeta_D(z) \]
As noted by Charnes and Cooper in (1), the duality state may be characterized by means of the linear program:

\[
\begin{align*}
\max \ & \ u^T e - \delta \\
\text{s.t.} \ & \ A^T \delta = b^T \\
\ & \ \delta \geq 0
\end{align*}
\]

State 1 corresponds to infeasibility, state 2 to \( \mu^* = 0 \), state 3 to \( \mu^* > 0 \).

4. The C² Duality For General Distributions

From the (two real variable) result in (2) and (2.1), we can write

\[
\begin{align*}
-\delta(t) \circ \left[ \frac{\delta(t)}{\text{ec}(t)} \right] & \leq c(t) e^{x(t)} - \delta(t)x(t), \\
\text{where} \ & \ c(t) > 0, \ \delta(t) \geq 0.
\end{align*}
\]

Taking \( \mu(t) \) as a non-negative Radon-Stieltjes measure, then, assuming existence of the integrals,

\[
\begin{align*}
\phi(t) = -\int \delta(t) \circ \left[ \frac{\delta(t)}{\text{ec}(t)} \right] \, d\mu(t) \leq \int [c(t)e^{x(t)} - \delta(t)x(t)] \, d\mu(t)
\end{align*}
\]

Let

\[
\begin{align*}
x(t) = \int A(t, s)z(s) \, d\psi(s), \quad \text{and} \\
\int \delta(t) A(t, s) \, d\mu(t) = b(s)
\end{align*}
\]

where \( \psi(s) \) is a Radon-Stieltjes measure, and convergence of integrals is assumed. Then (assuming Fubini's theorem holds),

\[
\begin{align*}
\phi(t) = -\int \delta(t) \circ \left[ \frac{\delta(t)}{\text{ec}(t)} \right] \, d\mu(t) \leq \phi(z), \quad \text{and} \\
\phi(z) = \int [c(t) \exp \left( \int A(t, s) z(s) \, d\psi(s) \right) \, d\mu(t) - \int b(s)z(s) \, d\psi(s)]
\end{align*}
\]

when (4.1 and 4.2) hold.

Clearly with \( \delta(t) = f_1(t) \), \( c(t) > 0 \) (except on a set of \( \mu \)-measure zero).

-7-
corresponding to \( f_2(t) \), in Kullback's notation \([6]\), we have here exhibited
the Charnes-Cooper dual programs for the case of general distributions. Note
for special choice of \( \nu \) that \( s \) takes on only a finite number of values as in
Rockefeller's or in Ben-Tal's and Charnes' work. This is the important
case in statistical estimation of a finite number of constraints. Thereby \( \mathcal{P}(z) \)
involves only an unconstrained finite vector \( z \).

5. A Game-Theoretic Connection

We can also connect the \( C^2 \)-duality with a saddle-point or two-person
zero-sum game. For, let

\[
(5.1) \quad K(\delta, x) = \sum_i (c_i e^{x_i} - \delta_i x_i)
\]

be the payoff to the "\( \delta \)-player" from the "\( x \)-player." Let the
respective strategy sets \( \Delta, X \) be defined by

\[
\Delta = \{ \delta : \delta^T A = b^T, \delta \geq 0 \}
\]

\[
X = \{ x : x = Az \}
\]

Clearly

\[
(5.3) \quad \sup_{\Delta} v(\delta) = \sup_{\Delta} \sum_i \delta_i \min_i \left( \frac{\delta_i}{c_i} \right) \leq \sup_{\Delta} \inf_X K(\delta, x)
\]

But

\[
(5.4) \quad \sup_{\Delta} \inf_X K(\delta, x) \leq \inf_{X} \sup_{\Delta} K(\delta, x)
\]

whenever both sides have meaning.

Also,

\[
\inf_X \sup_{\Delta} K(\delta, x) = \inf_X \sup_{\Delta} (c^T A z - \delta^T A z)
\]

\[
= \inf_z (c^T A z - b^T z), \text{ when } \Delta \neq \phi,
\]

\[
= \inf_z \mathcal{P}(z)
\]

From the \( C^2 \) duality theory, then \( \sup v(\delta) = \inf \mathcal{P}(z) \)
Thus,

Theorem: \( \sup_{\Delta} \inf_{X} K(\delta, x) = \inf_{X} \sup_{\Delta} K(\delta) \) whenever \( \Delta \neq \emptyset \).

Or more sharply,

\[ \max \inf K(\delta, x) = \inf \max K(\delta, x) \] whenever \( \Delta \neq \emptyset \).

6. A More General Inequalities Form

Suppose we wish to consider discrete \( K^2L \) estimation of the form

\[ \text{(6.1)} \quad \max v(\delta) = -\delta^T \alpha_n \left[ \begin{array}{c} \delta \\ \varepsilon \end{array} \right] \]

subject to

\[ \delta^T A^1 = b^1 \]
\[ \delta^T A^2 \leq b^2 \]
\[ \delta \geq 0 \]

(6.2) The constraints may be considered equivalently as

\[ \delta^T A^1 = b^1 \]
\[ \delta^T A^2 + \gamma^T = b^2 \]
\[ \delta^T, \gamma^T \geq 0 \]

By the remark on \( c = 0 \) in \( \delta \), we may write a duality inequality as

\[ \text{(6.3)} \quad -\sum_{i} \delta_i \alpha_i \left[ \begin{array}{c} \delta_i \\ \varepsilon_{i} \end{array} \right] + 0 \leq \sum_{i \in I_i} (c_i \epsilon x_i - \delta_i x_i) - \sum_{r \in I_r} \gamma_r y_r \]

with (6.2) holding as well as

\[ \text{(6.4)} \quad y_r \leq 0, \quad i \in I_r \]

To decouple we must choose representations of the \( x_i, y_r \), so that (6.2) represents the "T" and so that \( Z \) comprehends (6.4). In vector form we want

\[ \text{(6.5)} \quad \delta^T x + \gamma^T y = b^1 T z^1 + b^2 T z^2 \]
It suffices to have

\[(6.6) \quad (\delta^T, \gamma^T) \begin{bmatrix} A^1 & A^2 \\ 0 & I \end{bmatrix} \begin{bmatrix} z^1 \\ z^2 \end{bmatrix} = (b^1T, b^2T) \begin{bmatrix} z^1 \\ z^2 \end{bmatrix}\]

or,

\[(6.7) \quad \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} A^1 & A^2 \\ 0 & I \end{bmatrix} \begin{bmatrix} z^1 \\ z^2 \end{bmatrix} \quad \text{with} \quad z^2 \leq 0\]

The dual problems are now

\[(I') \quad \max \quad v(\delta) = 5^T \delta n \begin{bmatrix} A \\ \delta \end{bmatrix} \quad \text{with} \quad \begin{bmatrix} 5^T, \gamma^T \end{bmatrix} \begin{bmatrix} A^1 & A^2 \\ 0 & I \end{bmatrix} = \begin{bmatrix} b^1T, b^2T \end{bmatrix} \]

\[\delta, \gamma \geq 0\]

and

\[(II') \quad \inf \quad \mathcal{C}(z) = c^T e (A^1z^1 + A^2z^2) - b^1T z^1 - b^2T z^2 \quad \text{with} \quad -z^2 \geq 0\]

As might be expected, there are three duality states. These may be established as above and characterized by the solution of the \(\delta\)-constraint system (6.1). These results have been considered by a different method by F. Palacios (a student of Charnes).

The duality generalization to arbitrary statistical distribution functions is, of course, immediate, as follows.
The dual problems are

(III') \[ \sup: v(f_1) = -\int f_1 \frac{\partial l}{\partial f_2} \, d\mu \]

with \[ \int f_1(t) A^1(t, s) \, d\mu(t) = b^1(s), \quad s \in S_1 \]

\[ \int f_1(t) A^2(t, s) \, d\mu(t) + g(s) = b^2(s), \quad s \in S_2 \]

\[ f_1(t), g(s) \geq 0, \quad t \in T, \quad s \in S_2 : \]

\[ (f_2 \text{ is given and greater than zero a.e.}) \]

and

(IV') \[ \inf: \zeta(z) = \int f_2(t) \exp \left[ \int A^1(t, s) z^1(s) \, d\lambda^1(s) + \int A^2(t, s) z^2(s) \, d\lambda^2(s) \right] \, du - \]

\[ - \int b^1(s) z^1(s) \, d\lambda^1(s) - \int b^2(s) z^2(s) \, d\lambda^2(s) \]

with \[ z^2(s) \leq 0, \quad s \in S_2 \]

where, of course, the duality inequality

\[ (6.8) \quad v(f_1) \leq \zeta(z) \] holds for \( f_1(t), g(s), z^2(s) \)

constrained as in (III'), (IV') and for suitable convergence of the integrals.

7. Concluding Comments

The Charnes-Cooper extremal principle herein enunciated is more general than the Fenchel-Rockafellar construction with the Legendre transformation (their \( K(\delta, x) \) concave function \( \alpha "\delta" \) is always linear in \( \delta \)).

As illustrated above, the principle's simplicity and concreteness should make it a useful tool for the practicing analyst.
A special instance of the game-theoretic connection is being employed by Charnes, Cooper and Schinnar in current research to elucidate economic phenomena in a cartel economy.

The generalizations to functions other than our \( \psi(\delta), \chi(z) \) and the precise characterizations of the "non-discrete" dualities are also to be published elsewhere. These are important for problems of statistical inference with linear inequality constraints since the dual problem is not so encumbered. Further they apply directly to problems involving the evolution of macroscopically irreversible systems from microscopically reversible systems, as developed, for example, by B. O. Koopman [7].
REFERENCES


This paper presents a new extremal approach to deriving dual optimization problems with proper duality inequality which simplifies and generalizes the Fenchel-Rockafellar scheme. Our derivation proceeds in two stages, (i) inequality attainment, (ii) decoupling primal and dual variables. The power and convenience of this approach are exhibited through a new, much simpler derivation of the Charnes-Cooper results for Khinchin-Kullback-Leibler statistical estimation [1], the immediate establishment of the C^2 duality for general distributions and its extensions to general linear inequality constraints, plus the development of a new two-person zero-sum game connection.
<table>
<thead>
<tr>
<th>KEY WORDS</th>
<th>LINK A</th>
<th>LINK B</th>
<th>LINK C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremal Principles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimization Dualities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematical Statistics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khinchin-Kullback-Leibler Estimation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Convex Programming</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noncooperative Games</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>