OPERATIONS RESEARCH CENTER

working paper

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

FENCHEL AND LAGRANGE DUALITY ARE EQUIVALENT

by

T.L. Magnanti

OR 036-74 July 1974

Supported in part by the U.S. Army Research Office (Durham) under Contract No. DAHC04-73-C-0032.

ABSTRACT

A basic result in ordinary (Lagrange) convex programming is the saddlepoint duality theorem concerning optimization problems with convex inequalities and linear-affine equalities satisfying a Slater condition. This note shows that this result is equivalent to the duality theorem of Fenchel.

Among the most powerful tools in mathematical programming are the saddlepoint duality theories for both Fenchel and ordinary (Langrange) convex programs. The purpose of this note is to exhibit the equivalence of these two theories by showing that each can be used to develop the other.

The connection between Fenchel and Lagrange duality is certainly not surprising since each is based upon separating hyperplane arguments. Previously, Whinston [7] has shown how the Lagrangian theory can be derived from Frenchel's results when optimizing over R^n in the absense of equality constraints. Both theories also can be obtained from Rockafellar's general perturbation theory. The full equivalence between these theories, though, has not been stated in the open literature and seems to be unknown to most of the mathematical programming community. In fact, a number of recent comprehensive books in convex optimization including Luenberger [3], Rockafellar [5] and Stoer and Witzgall [6] do not explicitly mention or exploit this equivalence, but rather develop the two theories separately.

The Fenchel and Lagrange convex programs can be stated as

Fenchel

$$v=\inf_{x\varepsilon C_1\cap C_2} \{f_1(x)-f_2(x)\}$$

Lagrange

$$\bar{v}=\inf_{x\varepsilon C_o} \{f_o(x):g(x)\leq o,h(x)=o\}$$

Each C_j is a convex subset of R^n; $f_o(x)$ and $f_1(x)$ are convex functions defined respectively on C_o and C_1; $f_2(x)$ is a concave function defined on C_2; $g:C_o \to R^m$ is convex and $h:R^n \to R^r$ is linear-affine.

The saddlepoint dual problems are:

Fenchel Dual

$$d=\max_{\pi\varepsilon R^n} \{f_2^*(\pi) - f_1^*(\pi)\}$$

Lagrange Dual

$$\bar{d}=\max_{\substack{\pi\geq o\\ \alpha\varepsilon R^r\\ \pi\varepsilon R^n}} \inf_{x\varepsilon C_o} \{L(x;\pi,\alpha)\}$$

where

$$f_1^*(\pi) = -\inf_{x\varepsilon C_1} \{f_1(x)-\pi x\} \ , \ f_2^*(\pi) = \inf_{x\varepsilon C_2}\{\pi x - f_2(x)\}$$

are respectively conjugate convex and conjugate functions and

$$L(x;\pi,\alpha) = f(x)+\pi g(x)+\alpha h(x)$$

is the Lagrangian function. Note that the dual problems are formulated as maximizations. We shall consider conditions to insure that these maxima are attained.

Recall that $f_1^*(\pi) = +\infty$ and/or $f_2^*(\pi) = -\infty$ are possibilities. Such values for π will be inoperative for the maximization in the Fenchel dual and thus that problem is frequently written with $\pi\varepsilon C_1^* \cap C_2^*$ instead of $\pi\varepsilon R^n$ where

$$C_1^* = \{\pi\varepsilon R^n : f_1^*(\pi)<+\infty\} \quad \text{and} \quad C_2^* = \{\pi\varepsilon R^n : f_2^*(\pi)>-\infty\}.$$

Two of the most useful and delicate theorems connecting these primal-dual pairs are (for any convex C, $ri(C)$ denotes its relative interior):

Theorem 1: If $v>-\infty$ and $ri(C_1)\cap ri(C_2)\neq \phi$, then $v=d$.

Theorem 2: Assume $\bar{v}>-\infty$ and the Slater condition that there exists an $x^o \varepsilon ri(C_o)$ such that $g(x^o)<o$ and $h(x^o)=o$. Then $\bar{v}=\bar{d}$

The first theorem is Fenchel's original duality theorem [2] and the second is a consequence of a result due to Fan, Glicksberg and Hoffman [1] (see Chapter 28 of [5] or Chapter 6 of [6]). When there are no inequality constraints, theorem 2 remains valid by omitting the reference to $g(x)$ in the Slater condition. $\bar{v}=\bar{d}$ is equivalent to saying that the Kuhn-Tucker condition is satisfied: there exists $\hat{\pi}\geq o$ and $\hat{\alpha}$ such that $\bar{v}\leq L(x,\hat{\pi},\hat{\alpha})$ for all $x\varepsilon C_o$.

Below we show that either theorem 1 or 2 can be used to derive the other. The following elementary result concerning relative interiors will be useful.

Lemma 1: Let $C = \{(y^o,y^1,y^2)\varepsilon R^{1+m+r} : y^o\geq f_o(x), y^1\geq g(x)$ and $y^2=h(x)$ for some $x\varepsilon C_o\}$. If $x^o\varepsilon ri(C_o)$, $y^o>f_o(x^o)$, $y^1>g(x^o)$ and $y^2=h(x^o)$, then $y=(y^o,y^1,y^2)\varepsilon ri(C)$.

Proof: The following condition ([5] theorem 6.4) characterizes the relative interior of any convex set S:

$z\varepsilon ri(S)$ if and only if for each $\bar{z}\ \varepsilon S$ there is a $\mu>1$ such that $\mu z+(1-\mu)\bar{z}\ \varepsilon S$.

From our hypothesis C is a convex set; given any $x\varepsilon C_o$ and $\bar{y}=(\bar{y}^o,\bar{y}^1,\bar{y}^2)$ with $\bar{y}^o\geq f_o(x)$, $\bar{y}^1\geq g(x)$ and $\bar{y}^2=h(x)$, we must find a $\mu>1$ such that $\mu y+(1-\mu)\bar{y}\varepsilon C_o$.

By convexity and $x^o\varepsilon ri(C_o)$ there is a $\lambda>1$ such that $\mu x^o+(1-\mu)x\varepsilon C_o$ for all $1<\mu\leq\lambda$. Since $h(\cdot)$ is linear-affine

$$\mu y^2+(1-\mu)\bar{y}^2 = \mu h(x^o)+(1-\mu)h(x)=h(\mu x^o+ (1-\mu)x) \tag{1}$$

and since both $g(\cdot)$ and $f_o(\cdot)$ are continuous on $ri(C_o)$ the definition of y^o and y^1 shows that

$$\mu y^o+(1-\mu)\bar{y}^o>f_o(\mu x^o+(1-\mu)x) \tag{2}$$

and $$\mu y^1+(1-\mu)\bar{y}^1>g(\mu x^o+(1-\mu)x) \tag{3}$$

for some $1<\mu<\lambda$. (1)-(3) imply that $\mu y+(1-\mu)\bar{y}\varepsilon C$ so that $y\varepsilon ri(C)$..

<u>Theorem 3</u>: Theorem 1 and Theorem 2 are equivalent.

<u>Proof</u>; (Theorem 2 $\Rightarrow$ Theorem 1) Letting $C_o=C_1xC_2xR$ and $x=(x^1,x^2,x^3)$ the Fenchel problem is restated as

$$v= \inf_{x\varepsilon C_o} \{f_1(x^1)-f_2(x^2):x^1= x^3,x^2 = x^3\}$$

This is an ordinary convex problem with the associations $f_o(x)=f_1(x^1)-f_2(x^2)$ and $h(x) = \begin{pmatrix}x^1-x^3\\x^2-x^3\end{pmatrix}$. Its Lagrangian dual problem is

$$\bar{d} = \max_{\alpha^j\varepsilon R^n}\ \inf_{x\varepsilon C_o} \{f_1(x^1)-f_2(x^2)+\alpha^1x^1+\alpha^2x^2-(\alpha^1+\alpha^2)x^3\} . \tag{4}$$

The hypothesis $ri(C_1)\cap ri(C_2)\neq\emptyset$ implies the Slater condition for this problem and consequently that $v=\bar{d}$ by Theorem 2. Since $x^3\varepsilon R^n$, the infimum in the dual problem is $-\infty$ whenever $\alpha^1\neq-\alpha^2$. Therefore the dual maximization occurs for some $\pi= -\alpha^1=\alpha^2$ and the Lagrangian dual (4) can be written as the Fenchel dual, giving $d=\bar{d}=v$ to prove Theorem 1.

(Theorem 1 $\Longrightarrow$ Theorem 2). Letting C_1 be the set C of Lemma 1 and letting

$C_2 = \{(y^o,y^1,y^2)\varepsilon R^{1+m+r} : y^1\leq o \text{ and } y^2=o\}$ the Lagrange problem is written as

$$v= \inf \{y^o: y=(y^o,y^1,y^2)\varepsilon C_1\cap C_2\} \quad .$$

If x^o is the Slater point of Theorem 1 and $y^o>f_o(x^o)$, $o>y^1>g(x^o)$, $y^2=h(x^o)$, Lemma 1 implies that

$$(y^o,y^1,y^2)\varepsilon ri(C_1)\cap ri(C_2) \quad .$$

Identifying $f_1(y)=y^o$ and $f_2(y)=o$, the Fenchel dual to this formulation is

$$d=\max_{\pi^o,\pi^1,\pi^2} \{\inf_{y\varepsilon C_1}(y_o-\pi^oy^o-\pi^1y^1-\pi^2y^2)+ \inf_{y\varepsilon C_2} (\pi^oy_o+\pi^1y^1+\pi^2y^2)\}$$

and by Theorem 2, $v = \bar{d}$.

Since the second infimum is $-\infty$ if $\pi^o\neq o$ or $\pi^1_j>o$ for some $j=1,\ldots,m$ and is o otherwise, this dual problem reduces to

$$\bar{d}=\max_{\substack{\pi^1\leq o\\ \pi^2}} \inf_{y\varepsilon C_1} \{y^o-\pi^1y^1-\pi^2y^2\} = \max_{\substack{\pi\geq o\\ \alpha}} \inf_{x\varepsilon C_o} \{f_o(x)+\pi g(x)+\alpha h(x)\}=d$$

so that $v = d$ to prove Theorem 2.

Observe that the sets C_1 and C_2 used above are those typically constructed in a separating hyperplane approach to Lagrange duality. Also, the approach utilized here can be applied to relate other versions of the saddlepoint theories, for example when the dual problems do not have optimizing solutions. Other extensions can be obtained. For instance, the problem

$$v = \min \{ \sum_{j=1}^{k} f_j (A^j x) : A^j x \in C_j \}$$

where C_j are convex sets, A^j are real matrices with n columns, and f_j is convex on C_j can be expressed in Lagrange form as

$$\min_{x \in C_o} \{ \sum_{j=1}^{k} f_j (A^j x^j) : A^j x^j = A\bar{x} \}$$

where $C_o = C_1 x C_2 x \ldots x C_k x R^n$ and $x = (x^1, \ldots, x^k, \bar{x})$. Writing the Lagrange dual and simplifying provides the dual problem

$$d = \max_{\pi^j} \{ \sum_{j=1}^{k} \inf_{x \in C_j} (f_j(x) + \pi^j x) : \sum_{j=1}^{k} \pi^j A^j = o \}$$

and $v = d$ if there is an $x^o \in R^n$ with $A^j x^o \in ri(C_j)$. When $r=2$ and A^1 is the identity matrix this is the dual problem of Rockafellar [5]. These duality correspondences can be carried even further by appending inequality and equality constraints to the above formulation.

Acknowledgement: This material is an outgrowth of a section of report [4]. I would like to thank W. Oettli for several comments concerning this report and for suggesting this note.

References:

1. Ky Fan, I. Glicksberg and A. J. Hoffman, "Systems of inequalities involving convex functions, "Proc. Amer. Math. Soc. 8 (1957), 617-622

2. W. Fenchel, "Convex Cones, Sets and Functions", lecture notes, Princeton University, 1951

3. D. Luenberger, Optimization by Vector Space Methods, John Wiley & Sons, New York, 1969

4. T. Magnanti, "A linear approximation approach to duality in nonlinear programming, "Tech. Report OR 016-73, Operations Research Center, M.I.T., April 1973

5. R. T. Rockafellar, Convex Analysis, Princeton University Press, 1970

6. J. Stoer and C. Witzgall, Convexity and Optimization in Finite Dimensions I, Springer-Verlag, 1970

7. A. Whinston, "Some applications of the conjugate function theory to duality", in Nonlinear Programming (J. Abadie ed.), North-Holland Publishing Co., (1967), 77-98.