CONVEX FUNCTIONS AND DUAL EXTREMUM PROBLEMS

A thesis presented

by

Ralph Tyrrell Rockafellar

to

The Department of Mathematics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Mathematics

Harvard University

Cambridge, Massachusetts

May, 1963

ACKNOWLEDGEMENT

The author especially wishes to thank Prof. G. Birkhoff of Harvard University, Dr. A. A. Goldstein of the Massachusetts Institute of Technology, and Prof. A. W. Tucker of Princeton University for their steadfast encouragement and support of his work. He is also grateful to Prof. A. M. Gleason, Prof. L. H. Loomis, and Prof. G. W. Mackey of Harvard University, who profoundly stimulated his interest in Mathematics as an undergraduate, and to Prof. J. V. Talacko of Marquette University, whose enthusiasm for linear and convex programming sparked the efforts which led to this Thesis. Miss Elaine Pearlstein deserves thanks for her patience and care in preparing the manuscript. Welcome financial assistance was provided by the National Science Foundation under Grant GP-595.

SECTION ONE

Introduction

Problems of minimizing convex functions on convex subsets of real vector spaces, which are called convex programs, have attracted attention in recent years, especially because of their importance in economics, network theory and other disciplines. Our aim here is to develop a new general theory of such problems in the finite-dimensional case, using an approach inspired by that of Fenchel [19]. Computational procedures will not be considered. We shall be interested rather in characterizing solutions in various ways, and in extending the duality principles which play so prominent a role in the study of certain classes of problems, such as linear programs. Before describing our results, we shall set the stage by reviewing some well known facts.

Let E and E^* be two copies of the finite-dimensional real vector space R^n, with

$$[x,x^*] = \xi_1\xi_1^*+\dots+\xi_n\xi_n^*$$

for $x = (\xi_1,\dots,\xi_n) \in E$ and $x^* = (\xi_1^*,\dots,\xi_n^*) \in E^*$.
(Instead of identifying E and E^*, we think of the elements of each space as corresponding to the linear functions on the other by means of the bilinear functional $[x,x^*]$.) Similarly, let F and F^* be two copies of R^m, with

$$[y,y^*] = \eta_1\eta_1^*+\dots+\eta_m\eta_m^*.$$

The ordinary vector partial orderings are to be used in these spaces; thus $x' \geq x$ is to mean that $\xi_j' \geq \xi_j$ for $j=1,\dots,n$. Let $A = ((\alpha_{ji}^*))$ be an m x n real matrix with transpose $A^* = ((\alpha_{ji}^*))$, where $\alpha_{ji}^* = \alpha_{ij}$.

Treating all vectors as "column vectors," we view $x \to Ax$ as a linear transformation from E to F with adjoint $y^* \to A^*y^*$ from F^* to E^*, so that

(1.1) $[Ax, y^*] = [x, A^*y^*]$ for all $x \in E$ and $y^* \in F^*$.

This notation will be assumed throughout the paper.

For fixed vectors $b^* \in E^*$ and $c \in F$, the canonical dual linear programs [23] are the two constrained extremum problems defined by

(1.2a) minimize $[x, b^*]$ subject to $x \geq 0$ and $Ax \geq c$,

(1.2b) maximize $[c, y^*]$ subject to $y^* \geq 0$ and $A^*y^* \leq b^*$.

The remarkable property of this pair of problems is that, whenever one has a solution, then so does the other and the minimum and maximum are equal. Moreover, if the constraints in both problems can be satisfied, then both have solutions. These facts, proved by Gale, Kuhn and Tucker [21] in 1951, constitute the duality theorem for linear programs. Sometimes the last fact is referred to separately as the existence theorem.

A minimax property of the solutions to the dual linear programs depends on the following notion.

DEFINITION 1-A

Let B and C^* be non-empty subsets of E and F^*, respectively, and let $L(x, y^*)$ be a real-valued function given for all $x \in B$ and $y^* \in C^*$. A pair of vectors $\langle x_0, y_0^* \rangle$ is said to be a (global) saddle-point for L (minimizing over B and maximizing over C^*) if $x_0 \in B$, $y_0^* \in C^*$, and

(1.3) $L(x_0, y^*) \leq L(x_0, y_0^*) \leq L(x, y_0^*)$ for all $x \in B$ and $y^* \in C^*$.

If a saddle-point exists, then

(1.4) $$\lambda_0 = \inf_{x\in B} \sup_{y^* \in C^*} L(x,y^*) = L(x_0,y_0^*) = \sup_{y^* \in C^*} \inf_{x \in B} L(x,y^*)$$

is called the minimax value of L. (It is well known that (1.3) implies (1.4); e.g. see [27, p.22].)

Associated with (1.2a) and (1.2b) is a third problem

(1.5) find a saddle-point of the function

$$L(x,y^*) = [x,b^*] + [c,y^*] - [Ax,y^*]$$

on $B = \{x \mid x \geq 0\}$ and $C^* = \{y^* \mid y^* \geq 0\}$.

It is customary to called L the Lagrangian function, because of an analogy with the classical theory of Lagrange multipliers which will be explained in a moment. It is known [23, p.77] that $\langle x_0, y_0^* \rangle$ is a solution to (1.5) if and only if x_0 is a solution to (1.2a) and y_0^* is a solution to (1.2b). Moreover, when such solutions exist, the minimax value of the Lagrangian function in (1.5) coincides with the common extreme value in the dual problems (1.2a) and (1.2b).

Finally, x_0 is a solution to (1.2a) and y_0^* is a solution (1.2b) if and only if $\langle x_0, y_0^* \rangle$ is a solution to the following system of inequalities:

(1.6) $$x \geq 0,\ y \geq 0,\ Ax-c \geq 0,\ b^*-A^*y^* \geq 0$$
$$[Ax-c, y^*] \leq 0,\ [x, b^*-A^*y^*] \leq 0$$

(see [23, proof of Theorem 5]). Because of the result mentioned above, (1.6) also characterizes the saddle-points in (1.5). We shall speak of relations (1.6) as the equilibrium conditions for the dual linear programs. Gale [20, p.19] uses similar terminology when discussing the significance of (1.6) in economics.

Problem (1.5) allows us to view the dual relationship between (1.2a) and (1.2b) as an expression of a "conflict of interest." Indeeed, a problem of finding saddle-points such as those defined in 1-A can always be interpreted as a game between two players (I) and (II) whose strategy spaces are B and C^* [27, p.16]. Assume that when (I) plays the strategy $x \epsilon B$, and (II) plays the strategy $y^* \epsilon C^*$, the result is a payment $L(x,y^*)$ to (II). A pair of strategies x_0, y_0^* satisfying (1.3) then corresponds to a state of equilibrium. As long as (I) plays x_0, he never has to pay more than $L(x_0,y_0^*)$, but he would risk a higher payment if he deviated from x_0. At the same time, (II) can guarantee himself at least $L(x_0,y_0^*)$ by keeping to y_0^*, but would risk receiving less if he played a different strategy.

Since the saddle-points in (1.5) can be found by solving (1.2a) and (1.2b) separately, we can think of the latter as the strategy problems to be solve by two opposing "players." The duality theorem for linear programs reflects the intuitive idea that neither "player" can determine his best strategy without implicitly determining at the same time the best strategy which can be used against him. The equality between the extrema in (1.2a) and (1.2b) and the minimax value in (1.5) expresses the fact that loss to one "player" is gain to the other. This game analogy provide further justification for speaking of the relations (1.6) as the equilibrium conditions.

Similar heuristic interpretations can be given even for classical constrained extremum problems. Suppose that $f, g_1, \dots, g_m$ are differentiable functions on E and consider the problem:

(1.8) minimize $f(x)$ subject to $g_1(x) = 0, \dots, g_m(x) = 0$.

In order that x be a solution to this problem, it is necessary (under suitable regularity assumptions) that x satisfy the given constraints and the directional derivative of f vanish at x in every direction in which the g_i are all constant, i.e. that

$$\sum_j \zeta_j \frac{\partial f}{\partial \xi_j}(x) = 0 \text{ for each } z = \langle \zeta_1, \dots, \zeta_n \rangle$$

$$\text{such that } \sum_j \zeta_j \frac{\partial g_i}{\partial \xi_j}(x) = 0 \text{ for } i=1,\dots,m.$$

By elementary linear algebra, this implies that x is a solution to (1.8) only if

(1.9) $g_i(x) = 0$ for $i=1,\dots,m$, and

$$\frac{\partial f}{\partial \xi_j}(x) = \sum_i \eta_i^* \frac{\partial g_i}{\partial \xi_j}(x) \text{ for } j=1,\dots,n$$

for certain scalars $\eta_1^*, \dots, \eta_m^*$.

The new unknowns in (1.9) are called Lagrange multipliers. The Lagrangian function for (1.8) is defined by

(1.10) $L(x,y^*) = f(x) - \sum_i \eta_i^* g_i(x)$ for $y^* = \langle \eta_1^*, \dots, \eta_m^* \rangle$.

By means of L, we can re-state (1.9) as

(1.9') $\frac{\partial L}{\partial \eta_i^*}(x,y^*) = 0$ for $i=1,\dots,m$ and

$$\frac{\partial L}{\partial \xi_j}(x,y^*) = 0 \text{ for } j=1,\dots,n.$$

These are necessary conditions for a stationary point of L. Actually, however, the solutions to the minimization problem (1.8) correspond not to minima of L, but to "local" saddle-points which can be defined much as in 1-A. Heuristically, in other words, they correspond to local equilibria in a conflict of interest situation. Relations (1.9) are the "equilibrium conditions." No obvious dual problem presents itself in this case, but an implicit dual is described informally by Courant and Hilbert [10, vol. I, p.231. ff.].

In 1951, Kuhn and Tucker [28] considered problems of the form

(1.11) minimize $f(x)$ subject to $x \geq 0$, $g_1(x) \geq 0, \ldots, g_m(x) \geq 0$,

where f is convex and the g_i are concave. (The set of vectors satisfying the inequalities is then convex in E.) Using certain regularity assumptions including differentiability, they showed that solutions of (1.11) correspond to the saddle-points of the Lagrangian function

$$L(x,y^*) = f(x) - \sum_i \eta_i^* g_i(x) \text{ for } x \geq 0 \text{ and } y^* \geq 0,$$

and can be characterized by a system of inequalities, i.e. a set of equilibrium conditions in our terminology. They did not devise a problem dual to (1.11), however.

Generally speaking, the question of the existence of duals to given problems has turned out to be difficult and often ambiguous. In the case of (1.11) where the g_i are linear, for example, several types of duals have been constructed, each having its own advantages and drawbacks (see [2, p.99], [11, F], [13]). For many other problems, no dual at all is known.

In this paper we hope to develop a duality theory which can be used to deduce most known results, including the strongest results for linear programs, and yet can be applied to a large variety of new problems. Let f be a finite-valued convex functions on a non-empty convex set B in E, and let g be a finite-valued concave function on a non-empty convex set C in F. The basic problem we shall consider is

(1.12a) minimize $f(x) - g(Ax)$ subject to $x \in B$ and $Ax \in C$.

According to a theorem of Fenchel [18], the function f on B corresponds to a conjugate convex function f^* on a convex set B^* in E^*; similarly, g on C corresponds to a conjugate concave function g^* on a convex set C^* in F^*. The problem dual to (1.12a) in our theory will be

(1.12b) maximize $g^*(y^*) - f^*(A^* y^*)$ subject to $y^* \in C^*$ and $A^* y^* \in B^*$.

(These problems will take on a somewhat simpler form later, due to a device of extending convex and concave functions to the whole space by means of infinite values.) We shall also associate with these dual problems a "game" problem:

(1.12c) find a saddle-point of the Lagrangian function

$$L(x,y) = f(x) + g^*(y^*) - [Ax, y^*] \text{ for } x \in B \text{ and } y^* \in C^* .$$

Finally, using a generalized concept of "differential" ("gradient") which is explained in §2, we shall define equilibrium conditions which appear formally as

(1.12d) $Ax = \partial g^*(y^*)$ and $A^* y^* = \partial f(x)$

Problem (1.12a) reduces to (1.2a) if one chooses

$$f(x) = [x, b^*] \text{ for } x \in B = x | x \geq 0 \quad ,$$

$$g(y) = 0 \text{ for } y \in C = y | y \geq c \quad .$$

Then (as we show in §3) (1.7b), (1.7c) and (1.7d) reduce to (1.2b), (1.5) and (1.6) respectively.

The first three problems above, but not the problem of solving the generalized "equilibrium conditions" were studied by Fenchel [19, p.105-115] in the case where $E = F$, $E^* = F^*$, and $A = I$ is the identity matrix. (An account of Fenchel's results is also given in Karlin's book [27, p.227-229].) Fenchel proved in this special case that, if problems (1.12a) and (1.12b) are "strongly consistent" in the sense that B and C have relative interior points in common, and dually for B^* and C^*, then both problems have solutions and the extrema are equal. He also proved that then the solutions of (1.12a) and (1.12b) correspond to saddle-points in (1.12c). A weaker theorem of Fenchel [19, p.106], [27, p.229], which says that the extrema in (1.12a) and (1.12b) are equal (though not necessarily attained) whenever the constraints in both problems can be satisfied, is not correct. (This will be established by counter-example in §4).

We shall show in §4, using very similar "strong consistency" conditions, that Fenchel's results can be extended to the general case in which the spaces can be of different dimensions and A need not be the identity matrix. A substitute for the false duality theorem will also be derived. Since "strong consistency" is sometimes too restrictive an assumption, e.g. in linear programming theory, we shall develop in §5 a weaker sufficient condition, which we call "stable consistency." This new type of consistency makes use of the regularity properties of certain classes of "stable" and "completely stable" convex functions. The latter includes, among others, all the quadratic convex functions and the "polyhedral" convex functions described by the author in [31]. For the case of (1.2a) in which f and g are "completely stable", we shall prove theorems that are in every respect as strong as those for linear programs. Moreover, as will be shown in §6, the class of "completely stable" convex functions is closed under a number of useful combinatorial operations. Therefore the strongest theorems can always be applied to (1.12a) when f and g have been constructed from known "completely stable" functions by means of such operations.

Various special cases of the theory will be considered in detail in §7 and §8. A new existence theorem for solutions to certain network problems, proved at the end of §8, deserves particular mention.

SECTION TWO

Convex Functions

The portions of the theory of convex functions that play a central role throughout this paper will be treated here. The facts about the closure, conjugate and differential of a convex function are of particular importance. The last of these three notions is new, at least in the generalized form in which we shall develop it; it is based on the familiar principle that, at most points, the graph of a convex function has one or more non-vertical tangents. The other two notions originate with Fenchel [18]. Through the device of extending all convex functions to the whole space by allowing them to take on the value $+\infty$, we are able to present Fenchel's results in a notationally simpler form. This device has already been used in a similar context by Hörmander [26] and Moreau [42].

Let C be a convex set in E (where E is a copy of R^n as explained in §1). The topological closure of C will be denoted by cl C. The <u>relative interior</u> ri C of C is its interior relative to the smallest linear manifold (translate of a subspace) containing it, while its <u>relative boundary</u> rb C is the set difference between cl C and ri C. It is well known that cl C and ri C are convex sets, and

(2.1) $\quad \mathrm{cl}(\mathrm{ri}\ C) = \mathrm{cl}\ C \text{ and } \mathrm{ri}(\mathrm{cl}\ C) = \mathrm{ri}\ C.$

In particular, a non-empty convex set always has a non-empty relative interior. (See [15, p.9-16].) These facts would not generally be true if E were not finite-dimensional. Relative interiors are considered in detail in Appendix A.

A <u>convex</u> <u>cone</u> is a convex set K such that $\lambda x \in K$ whenever $x \in K$ and $0 < \lambda \in R$.

DEFINITION 2-A

A function f, defined on a non-empty convex set C in E and having values in the extended real interval $-\infty \leq f(x) \leq \infty$, will be called a <u>convex</u> <u>function</u> on C if

(2.2) $$f(\lambda x_1+(1-\lambda)x_2) \leq \lambda\mu_1 + (1-\lambda)\mu_2 \text{ whenever}$$
$$x_1 \in C,\ x_2 \in C,\ f(x_1) \leq \mu_1 \in R,\ f(x_2) \leq \mu_2 \in R,\ 0 < \lambda < 1.$$

Unless f assumes both $+\infty$ and $-\infty$, in which case the ambiguous combination $\infty - \infty$ would arise, (2.2) can be simplified to

(2.2') $$f(\lambda x_1+(1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) \text{ for } x_1 \in C,\ x_2 \in C,\ 0 < \lambda < 1.$$

The following obvious conventions are to be used in (2.2') and elsewhere in calculations that involve $+\infty$ and $-\infty$:

(2.3) $$\lambda + \infty = \infty + \lambda = \infty \text{ if } -\infty < \lambda \leq \infty,$$
$$\lambda - \infty = -\infty + \lambda = -\infty \text{ if } -\infty \leq \lambda < \infty,$$
$$\lambda \cdot \infty = \infty \cdot \lambda = \begin{cases} \infty \text{ if } 0 < \lambda < \infty, \\ -\infty \text{ if } -\infty < \lambda < 0, \end{cases}$$
$$\lambda \cdot (-\infty) = (-\infty) \cdot \lambda = \begin{cases} -\infty \text{ if } 0 < \lambda < \infty, \\ \infty \text{ if } -\infty < \lambda < 0. \end{cases}$$

Combinations other than those in (2.3) will not appear. (The convention $0 \cdot \infty = \infty$ will, however, be introduced in a special context in §7.) It is easily verified that ordinary algebraic laws, such as the associativity of addition, are still satisfied when

(2.3) is used; at all events, every calculation in which we employ infinite values can be formalized by replacing expressions of form (2.2') by those of form (2.2) which involve only real numbers.

Given a convex function f_0 on C, we can define f on E by setting $f(x) = f_0(x)$ for $x \in C$ and $f(x) = \infty$ for $x \notin C$. Then f is a convex function on E; this follows trivially from Definition 2-A. Thus, without loss of generality, we can limit the discussion below to convex functions defined on all of E, provided we do allow such functions to be infinite-valued.

Properties of a convex function f on E can often be viewed as geometric properties of its upper graph set, which we define by

(2.4) $\quad \text{gph } f = \{\langle x,\mu\rangle \mid x \in E,\ f(x) \leq \mu \in R\} \subseteq E \oplus R.$

($E_1 \oplus E_2$ denotes the vector space whose elements are the pairs $\langle x_1, x_2\rangle$, $x_1 \in E_1$, $x_1 \in E_2$.) It is easy to see, for example, that the convexity of f is equivalent to the convexity of gph f in $E \oplus R$. Another important set which we associate with f is its effective domain:

$$\text{dom } f = \{x \mid f(x) < \infty\} \subseteq E.$$

Since dom f is essentially the projection of gph f into E, it is clear that dom f, too, is convex when f is convex.

One can always "close" a convex function in the following sense.

DEFINITION 2-B

The closure of a convex function f on E is the function cl f defined by

$$\text{cl } f(x) = \liminf_{z \to x} f(z) \text{ for each } x \in E.$$

If cl f = f, f is said to be <u>closed</u>. (In other words, a convex function f on E is closed if and only if it is lower semi-continuous.)

This definition is simpler than Fenchel's [18], because f is everywhere defined; a similar definition is given by Moreau [42]. Trivially, cl f ≤ f. Furthermore, cl f(x) ≤ $\mu \in R$ if and only if there exist sequences $x_k \to x$ and $\mu_k \to \mu$ such that $f(x_k) \leq \mu_k \in R$ for all k. By definition (2.4), this means that the point $\langle x, \mu \rangle \in$ gph(cl f) is a limit of points $\langle x_k, \mu_k \rangle \in$ gph f. Therefore

(2.5) $\quad$ gph(cl f) = cl(gph f).

Thus the closure operation for convex functions merely reflects the closure operation for convex sets. It follows at once from (2.5) that cl(cl f) = cl f, that cl f is closed convex function on E, and that f is closed if and only if gph f is closed. Definition 2-B implies that

(2.6) $\quad$ dom f $\subseteq$ dom(cl f) $\subseteq$ cl(dom f),

but simple examples show that f can be closed without dom f being closed.

The following properties are trivial extensions of properties proved by Fenchel [19, p.75 and p.78]:

(2.7) $\quad$ cl f(x) = f(x) whenever $x \notin$ rb(dom f),

(2.8) $\quad$ cl f(x) = $\lim_{\lambda \to 0^+} f(x + \lambda(x_0 - x))$ if $x_0 \in$ ri(dom f),

(2.8') $\quad$ f(x) = $\lim_{\lambda \to 0^+} f(x + \lambda(x_0 - x))$ if cl f = f and $x_0 \in$ dom f.

The first of these is a consequence of the well known fact that a finite-valued convex function on an open set is continuous (see [5, p.92] or [15, p.46]). By (2.7) and (2.8), a <u>closed</u> convex function on E is completely determined by its values on the relative interior of its effective domain. Very little can be said, however, about the behavior of a non-closed f on rb(dom f). For example, suppose E is two-dimensional and let $f(\xi_1,\xi_2) = 0$ for $\xi_1^2 + \xi_2^2 < 1$, $f(\xi_1,\xi_2) = \infty$ for $\xi_1^2 + \xi_2^2 > 1$, assigning arbitrary non-negative values to f for $\xi_1^2 + \xi_2^2 = 1$. Then f is a convex function on E. But if f is required to be closed, the arbitrariness disappears and only the value 0 can be assigned for $\xi_1^2 + \xi_2^2 = 1$.

We say that a convex function f on E is proper if $f(x) > -\infty$ for all $x \in E$ and $f(x) < \infty$ for at least one $x \in E$. Otherwise we say that f is improper. Geometrically, f is proper if and only if gph f is non-empty and contains no vertical lines. It is apparent from Definition 2-A that $f(x) = -\infty$ for all $x \in$ ri(dom f) when f is improper. Hence

$$\text{(2.9)} \qquad \text{cl } f(x) = \begin{cases} -\infty & \text{for } x \in \text{cl(dom } f) \\ \infty & \text{for } x \notin \text{cl(dom } f) \end{cases} \quad \text{if } f \text{ is improper.}$$

This implies, via (2.7), that cl f is proper if and only if f is proper. For the most part, we shall be concerned only with proper convex functions; improper functions, however, are important in several proofs (e.g. in 4-B).

We shall now describe the properties of Fenchel's fundamental conjugate operation [18], which induces a polar correspondence between the closed proper convex functions on E and those on the dual space E^*.

DEFINITION 2-C

The conjugate of a convex function f on E is the function f^*

on E^* defined by

(2.10) $\quad f^*(x^*) = \sup_x \{[x,x^*]-f(x)\}$ for each $x^* \in E^*$.

The <u>second conjugate</u> of f is the conjugate f^{**} of f^*,

$$f^{**}(x) = \sup_{x^*} \{[x,x^*]-f^*(x^*)\} \text{ for each } x \in E.$$

Observe that the supremum in the definition of f^*, while formally extended over all of E, could be expressed equivalently as a supremum over dom f (provided f is not identically $+\infty$). From Definitions 2-B and 2-C we have

$$\begin{aligned} (\mathrm{cl}\ f)^*(x^*) &= \sup_x \{[x,x^*]-\liminf_{z \to x} f(z)\} \\ &= \sup_x \lim_{z \to x} \sup \{[z,x^*]-f(z)\} \\ &= \sup_z \{[z,x^*]-f(z)\} = f^*(x^*), \end{aligned}$$

so that

(2.11) $\quad (\mathrm{cl}\ f)^* = f^*$.

It is immediate from (2.10) that

(2.12) $\quad f_1 \le f_2$ implies $f_1^* \ge f_2^*$.

There is a simple geometric idea behind the conjugate operation. Let $x^* \in E^*$, $\mu^* \in R$, and let $h(x) = [x,x^*]-\mu^*$. Then h is an affine convex function on E and gph h is a "non-vertical" closed half-space in $E \oplus R$. Moreover gph $f \subseteq$ gph h if and only if $f(x) \ge h(x)$ for all $x \in E$, i.e.

$$\mu^* \ge \sup_x \{[x,x^*]-f(x)\} = f^*(x^*).$$

If f is proper, it is plausible that the closed convex set cl(gph f) is the intersection of all such half-spaces gph h, which means by (2.5) that cl f is the supremum of all such affine functions h, i.e.

$$\mathrm{cl}\ f(x) = \sup_{x^*} \{[x,x^*]-f^*(x^*)\}.$$

By the last expression, by definition, is $f^{**}(x)$. Using the familiar separation theorems for convex sets, this argument can be formalized to obtain the following theorem. (We shall omit the formal proof, since it closely parallels Fenchel's in [18], except for notation. An infinite-dimensional version of the theorem may be found in Moreau [42].)

THEOREM 2-D

Let f be a proper convex function on E. Then f^* is a closed proper convex function on E^* and $f^{**} = \mathrm{cl}\, f$.

The theorem is almost, but not quite, true for improper convex functions. If f is identically $+\infty$, then trivially $f^*(x^*) = -\infty$ for all x^* and $f^{**}(x) = \mathrm{cl}\, f(x) = +\infty$ for all x. But if f assumes the value $-\infty$, then $f^*(x^*) = +\infty$ for all x^* and $f^{**}(x) = -\infty$ for all x, whereas cl f is given by (2.9). Thus, in the latter case, f^{**} and cl f agree on cl(dom f) but not elsewhere.

COROLLARY 2-E

Let f_1 and f_2 be proper convex functions on E. Then $f_1^* \geq f_2^*$ if and only if $\mathrm{cl}\, f_1 \leq \mathrm{cl}\, f_2$. In particular, the conjugate operation defines a one-to-one order-inverting correspondence between the closed proper convex functions on E and those on E^*.

Proof: This is immediate from (2.11), (2.12) and Theorem 2-D.

Conjugate pairs of closed proper convex functions also correspond to "best inequalities" of a certain type:

THEOREM 2-F

Let f be a closed proper convex function on E, with conjugate f^* on E^*. Then

(2.13) $\quad f(x) + f^*(x^*) \geq [x,x^*]$ for all $x \in E$ and $x^* \in E^*$.

Moreover, for each $x \in$ ri(dom f) there exists some x^* such that (2.13) holds with equality, and dually for each $x^* \in$ ri(dom f).

Proof: Inequality (2.13) is obvious from Definition 2-C. The sharpened final assertion was proved by Fenchel in [18].

If f is a differentiable convex function finite on all of E, its differential (or gradient) $\partial f(x)$ at x is given by

$$\partial f(x) = \left\langle \frac{\partial f}{\partial \xi_1}(x), \ldots, \frac{\partial f}{\partial \xi_n}(x) \right\rangle,$$

which can be interpreted as an element x^* of E^*. (Our terminology and notation agree with that of Dennis [11] in a similar context.) The affine function $h(z) = f(x) + [z-x, \partial f(x)]$ is then tangent to $f(z)$ at $z = x$, with $f(z) \geq h(z)$ for all $z \in E$. With this fact in mind, we extend the concept of differential to arbitrary proper convex function as follows.

DEFINITON 2-G

Let f be a proper convex function on E and let $x \in E$. We say that $x^* \in E^*$ is <u>a</u> <u>differential</u> <u>of</u> f <u>at</u> x, and write $x^* = \partial f(x)$, if

(2.14) $\quad f(z) \geq f(x) + [z-x, x^*]$ for all $z \in E$

The relation ∂f, which consists of all pairs $\langle x, x^* \rangle$ such that $x^* = \partial f(x)$, is called <u>the</u> <u>differential</u> <u>of</u> f.

Observe that we have only defined the expression $x^* = \partial f(x)$, and not $\partial f(x)$ by itself. This is due to the fact that f may have more than one (or no) differential at a given point x. (That can happen, for instance, if the graph of f has a vertex at x, or if x is not an interior point of dom f.) We could, however, interpret ∂f as a multiple-valued mapping. When f is actually differentiable, ∂f is single-valued and coincides with the ordinary differential described above. This is proved formally in Appendix C (see C-F), where the relationship between the generalized differentials of Definition 2-G and the classical theory of directional derivatives of convex functions is explained in detail. Our theoretical development is based directly on Definition 2-G and does not assume facts from the calculus (although these would be useful in determining the differentials in certain special applications).

THEOREM 2-H

Let f be a <u>closed</u> proper convex function on E. Then the following statements are equivalent:

(a) $x^* = \partial f(x)$,

(b) $x = \partial f^*(x^*)$,

(c) $f(x) + f^*(x^*) \leq [x,x^*]$.

In this sense the differential ∂f^* is the inverse of ∂f.

<u>Proof:</u> Re-writing (2.14) and applying Definition 2-C, we see that $x^* = \partial f(x)$ if and only if

$$[x,x^*] - f(x) \geq \sup_z [z,x^*]-f(z) = f^*(x^*).$$

Since f is proper by assumption, and f^* is proper by 2-D, f(x) must both be finite if either this inequality or inequality (c) holds. Hence (a) is equivalent to (c). A dual argument proves that (b) is equivalent to (c), because, by 2-D and the assumption that f is closed, f is the conjugate of f^*.

COROLLARY 2-I

A closed proper convex function f on E has differentials at all points $x \in$ ri(dom f), but no differentials at points $x \notin$ dom f.

Proof: This follows at once from 2-F and 2-H.

COROLLARY 2-J

Let f be a proper convex function on E such that $0 \in$ ri(dom f). Then

$$-\infty < \min_{x^*} f^*(x^*) < \infty.$$

(Following the usual convention, we indicate that an extremum is attained by replacing "inf" by "min" or "max".)

Proof: If $0 \in$ ri(dom f), then $0 \in$ ri(dom(clf)) by (2.1) and (2.6). Since cl f is a closed proper convex function, 2-I implies that $x^* = \partial(\text{cl } f)(0)$ for some $x^* \in E^*$. Since $(\text{cl } f)^* = f^*$ by (2.11), 2-H now implies that $0 = \partial f^*(x^*)$. This means by Definition 2-G that

$$f^*(z^*) \geq f^*(x^*) + [0, z^*-x^*] = f^*(x^*)$$

for all $z^* \in E^*$. Furthermore, $f^*(x^*)$ must be finite in this case because f^* is proper by 2-D.

The next two theorems provide formulas for conjugates in common situations.

THEOREM 2-K

Let f be a convex function on E. Then

(a) $h(x) = f(x-a)$, $a \in E$, implies $h^*(x^*) = f^*(x^*) + [a, x^*]$,

(b) $h(x) = f(x) + [x, a^*]$, $a^* \in E^*$, implies $h^*(x^*) = f^*(x^* - a^*)$,

(c) $h(x) = \lambda f(x)$, $0 < \lambda \in R$, implies $h^*(x^*) = \lambda f^*((1/\lambda)x^*)$,

(d) $h(x) = f(\lambda x)$, $0 \neq \lambda \in R$, implies $h^*(x^*) = f^*((1/\lambda)x^*)$,

(e) $h(x) = f(x) + \alpha$, $\alpha \in R$, implies $h^*(x^*) = f^*(x^*) - \alpha$.

Proof: These formulas are easy consequences of Definition 2-C, and have already been pointed out by Fenchel [19, p.93-94].

THEOREM 2-L

Suppose that $E = E_1 \oplus \ldots \oplus E_k$ and $E^* = E_1^* \oplus \ldots \oplus E_k^*$, with

$$[x, x^*] = [x_1, x_1^*] + \ldots + [x_k, x_k^*]$$

for $x = \langle x_1, \ldots, x_k \rangle \in E$ and $x^* = \langle x_1^*, \ldots, x_k^* \rangle \in E^*$. Let f_i be a proper convex function on E_i, with conjugate f_i^* on E_i^*, for $i=1,\ldots,k$, and let

$$f(x) = f(x_1, \ldots, x_k) = f_1(x_1) + \ldots + f_k(x_k).$$

Then f is a proper convex function on E, closed if and only if all the f_i are closed, and

$$f^*(x^*) = f^*(x_1^*, \ldots, x_k^*) = f_1^*(x_1^*) + \ldots + f_k^*(x_k^*)$$

on E^*. Moreover

$$\langle x_1^*, \ldots, x_k^* \rangle = x^* = \partial f(x) = \partial f(x_1, \ldots, x_k)$$

if and only if $x_i^* = \partial f_i(x_i)$ for $i=1,\ldots,k$.

<u>Proof:</u> Since the f_i are proper, f does not assume the value $-\infty$ and is not identically ∞; the convexity of f is then easy to verify using (2.2'). Thus f is a proper convex function on E. The assertion about closure is an obvious consequence of Definition 2-B. We calculate the conjugate of f directly from (2.10), obtaining

$$f^*(x^*) = \sup\left\{\sum [x_i,x_i^*]-\sum f_i(x_i) \mid x_i \in E_i, i=1,\ldots,k\right\}$$
$$= \sum_i \sup\left\{[x_i,x_i^*]-f_i(x_i) \mid x_i \in E_i\right\} = \sum_i f_i^*(x_i^*).$$

Finally, by Definition 2-G, $x^* = \partial f(x)$ if and only if

$$\sum_i f_i(z_i) = f(z) \geq f(x) + [z-x,x^*] = \sum_i (f_i(x_i)+[z_i-x_i,x_i^*])$$

for all $z_i \in E_i$, $i=1,\ldots,k$. This happens if and only if

$$f_i(z_i) \geq f_i(x_i) + [z_i-x_i,x_i^*]$$

for all $x_i \in E_i$, i.e. $x_i^* = \partial f_i(x_i)$, for $i=1,\ldots,k$.

A simple but important example of a convex function is the <u>convex characteristic function</u> $\check{\delta}_C$ of a convex set C, which we define by

$$(2.15) \qquad \check{\delta}_C(x) = \check{\delta}(x|C) = \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{if } x \notin C. \end{cases}$$

Clearly $\check{\delta}_C$ is proper if and only if C is non-empty, and is closed if and only if C is closed; in fact

$$(2.16) \qquad \text{cl}\ \check{\delta}(x|C) = \check{\delta}(x \mid \text{cl}\ C).$$

If a convex function f on E assumes no values other than 0 and ∞, then $f = \check{\delta}_C$, where $C = \text{dom}\ f$. This one-to-one correspondence between convex sets and certain convex functions leads to a useful principle: general facts about convex functions can always be specialized to facts about convex sets. For instance the conjugate of $\check{\delta}_C$ is by definition

$$\sup_x \{[x,x^*]-\check{\delta}(x|C)\} = \sup\{[x,x^*]|x\in C\} = \check{\sigma}_C(x^*)$$

(provided $C \neq \emptyset$), where $\check{\sigma}_C$ is called the convex support function of C. The well known theorems relating convex sets and their support functions can be deduced from 2-D. This is demonstrated in Appendix A.

Convex characteristic functions are very useful in extremum problems. Suppose that f is a convex function finite on all of E and that C is a non-empty closed convex set. Let

$$h(x) = f(x) + \check{\delta}(x|C) \text{ for all } x\in E.$$

Then h is a closed proper convex function on E by (2.7), (2.8') and (2.16), and dom h = C. Moreover

$$\inf\{f(x)\,|x\in C\} = \inf_x h(x).$$

In this manner constrained minimization can be treated uniformly as minimization on E, the constraints being incorporated into the effective domains of the functions involved. Operations defined for convex functions such as closure, conjugation, addition and convolution (see §6), then lead to an automatic calculus of constraints in extremum problems and their duals.

Concave functions will also be important in this paper.

DEFINITION 2-M

A function g on E is concave if -g is convex. The conjugate g^* of a concave function g on E is defined by

(2.17) $$g^*(x^*) = \inf_x \{[x,x^*]-g(x)\} \text{ for all } x^* \in E^*.$$

Notice from this definition that

(2.18) $$\text{if } f(x) = -g(x) \text{ then } f^*(x^*) = -g^*(-x^*),$$

rather than merely $f^* = -g^*$ as one might guess. Aside from this possibly misleading point, the theory of concave functions mirrors the theory of convex functions completely. We shall not explicitly state the obvious concave analogs of theorems and definitions that apply to convex functions, but these analogs will nevertheless be used in theoretical developments.

We find it conveninent to employ a notation dual to (2.15) for the <u>concave characteristic function</u> of a convex set C:

(2.19) $$\hat{\delta}_C(x) = \hat{\delta}(x|C) = \begin{cases} 0 \text{ if } x \in C, \\ -\infty \text{ if } x \notin C, \end{cases}$$

whose conjugate (if $C \neq \emptyset$) is the <u>concave support function</u> of C:

(2.20) $$\hat{\sigma}_C(x^*) = \hat{\sigma}(x^*|C) = \inf \{[x,x^*] | x \in C\} .$$

Of course $\hat{\delta}(x|C) = -\check{\delta}(x|C)$ and $\hat{\sigma}(x^*|C) = -\check{\sigma}(-x^*|C)$.

SECTION THREE

The Model Problems

We shall now define the problems whose theory is to be developed in this and later sections. It will be assumed in these problems that f is a closed proper convex function on E, with conjugate f^* on E^*, and that g is a closed proper concave function on F with conjugate g^* on F^*. By Theorem 2-D, f^* and g^* are also closed and proper, and their conjugates are in turn f and g. It will also be assumed, as in §1, that A is the matrix of a linear transformation from E to F, so that the transpose A^* of A is the matrix of a linear transformation from F^* to E^*. The notation is schematized in Figure 1.

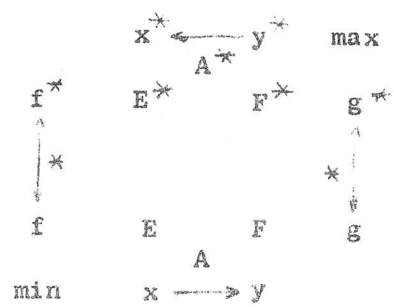

Figure 1

For each such system of elements we consider the following four problems:

(I) Convex program:

minimize $f(x) - g(Ax)$ on E.

(II) Concave program:

maximize $g^*(y^*) - f^*(A^*y^*)$ on F^*.

(III) Game, or saddle-point problem:

find a saddle-point of the Lagrangian function

$L(x,y^*) = f(x) + g^*(y^*) - [Ax,y^*]$, $x \in \text{dom } f$, $y^* \in \text{dom } g^*$.

(IV) "Equilibrium" problem:

find vectors satisfying the equilibrium conditions

$Ax = \partial g^*(y^*)$ and $A^*y^* = \partial f(x)$.

Once a problem arising in practice has been formulated as any one of the above, it automatically leads to three other problems. We shall say that (I) and (II) are dual to one another.

Clearly a solution to (IV) is a pair of vectors $\langle x_0, y_0^* \rangle$ satisfying the equilibrium conditions, while solutions to (III) were defined in 1-A. We say that x_0 is a solution to (I) if

(3.1) $$-\infty < f(x_0) - g(Ax_0) = \min_x \{f(x)-g(Ax)\} < \infty,$$

and that y_0^* is a solution to (II) if

(3.2) $$\infty > g^*(y_0^*) - f^*(A^*y_0^*) = \max_{y^*} \{g^*(y^*)-f^*(A^*y^*)\} > -\infty.$$

Thus we do not speak of solutions to (I) or (II) if the extrema are infinite.

Although (I) and (II) are formally unconstrained, they actually involve the implicit constraints:

(3.3a) $x \in \operatorname{dom} f$ and $Ax \in \operatorname{dom} g$,

(3.3b) $y^* \in \operatorname{dom} g^*$ and $A^* y^* \in \operatorname{dom} f^*$,

respectively. Indeed, $f(x) - g(Ax)$ is finite when x satisfies (3.3a), but has the value $+\infty$ when x does not; so that the minimization in (I) automatically subject to (3.3a) when this is possible at all. If $\inf_x f(x)-g(Ax) = +\infty$ in (I), this means that the implicit constraints cannot be satisfied. Similarly $\sup_{y^*} g^*(y^*)-f^*(A^* y^*) = -\infty$ if and only if the implicit constraints (3.3b) of (II) cannot be satisfied. It is important to keep this in mind when interpreting infinite extrema appearing in the duality theorems proved later.

Observe that the Lagrangian function $L(x,y^*)$ is convex and lower semi-continuous (by the closedness of f) on dom f for each $y^* \in \operatorname{dom} g^*$, and is concave and upper semi-continuous on dom g^* for each $x \in \operatorname{dom} f$. Restriction of L to the effective domains of f and g^* is necessary to avoid $\infty - \infty$.

An interesting insight into the equilibrium conditions can be gained in the following manner. Suppose, for the sake of argument, that f, f^*, g, g^* are all finite and differentiable everywhere. (This situation is studied in Appendix C.) Let $h(x) = f(x) - g(Ax)$. The differential (gradient) of h, determined by the ordinary methods of the calculus, is then

$$\partial h(x) = \partial f(x) - A^* \partial g(Ax).$$

The solutions to (I) can now be found from the equation $\partial h(x) = 0$, which amounts to

$$\partial f(x) = A^* y^*, \text{ where } y^* = \partial g(Ax).$$

But, by the concave analog of Theorem 2-H, $y^* = \partial g(Ax)$ if and only if $Ax = \partial g^*(y^*)$. These are therefore just the equilibrium conditions. Thus the equilibrium conditions in (IV) generalize the elementary idea that the solutions to extremum problems are found from the "partial differential equations" obtained by setting differentials equal to zero. Of course the differentials ∂f and ∂g^* which we have defined are not always single-valued, so that (IV) cannot be derived rigorously from (I) by the above argument in the non-differentiable case.

In view of Theorem 2-E, we can also express the general equilibrium conditions as a cyclic set of four conditions

(3.4) $$Ax = y, \quad \partial g(y) = y^*, \quad A^* y^* = x^*, \quad \partial f^*(x^*) = x.$$

Thus (IV) is solvable if and only if it is possible to "complete a circuit" via the four "mappings" indicated in Figure 2.

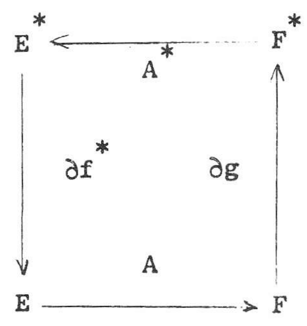

Figure 2

We shall now demonstrate how problems (I) through (IV) reduce to the corresponding problems of linear programming theory through a simple choice of elements. Let $b^* \in E$ and $c \in E$, and let

(3.5a) $f(x) = [x, b^*]$ if $x \geq 0$, $f(x) = \infty$ if $x \ngeq 0$,

(3.5b) $g(y) = 0$ if $y \geq c$, $g(y) = -\infty$ if $y \ngeq c$.

Then obviously

$$f(x) - g(Ax) = \begin{cases} [x, b^*] & \text{if } x \geq 0 \text{ and } Ax \geq c, \\ \infty & \text{otherwise,} \end{cases}$$

so problems (I) and (1.2a) coincide. Calculating the conjugates of f and g from definitions 2-C and 2-M, we obtain

$$f^*(x^*) = \sup_{x \geq 0} [x, x^*] - [x, b^*] = \begin{cases} 0 & \text{if } x^* \leq b^*, \\ \infty & \text{if } x^* \nleq b^*. \end{cases}$$

$$g^*(y^*) = \inf_{y \geq c} [y, y^*] = \begin{cases} [c, y^*] & \text{if } y^* \geq 0, \\ -\infty & \text{if } y^* \ngeq 0. \end{cases}$$

Therefore

$$g^*(y^*) - f^*(A^* y^*) = \begin{cases} [c, y^*] & \text{if } y^* \geq 0 \text{ and } A^* \leq b^*, \\ -\infty & \text{otherwise,} \end{cases}$$

and (II) turns out to be (1.2b). Moreover the Lagrangian function in (III) is just

$$L(x, y^*) = [x, b^*] + [c, y^*] - [Ax, y^*] \text{ for } x \geq 0 \text{ and } y^* \geq 0.$$

Hence (III) reduces to (1.5). Next we calculate the generalized differentials ∂f and ∂g^*. Applying definition 2-G to (3.5a), we see that $x^* = \partial f(x)$ if and only if

$$[z, b^*] + \delta(z \mid z \geq 0) \geq [x, b^*] + \delta(x \mid x \geq 0) + [z - x, x^*] \text{ for all } z.$$

(δ is defined in (2.15).) This is the same as

$$x \geq 0 \text{ and } [z, b^*-x^*] \geq [x, b^*-x^*] \text{ for all } z \geq 0,$$

which in turn is equivalent to

(3.6a) $\quad x \geq 0,\ b^*-x^* \geq 0, \text{ and } [x, b^*-x^*] \leq 0.$

By a similar argument, $y = \partial g^*(y^*)$ if and only if

(3.6b) $\quad y^* \geq 0,\ y-c \geq 0, \text{ and } [y-c, y^*] \leq 0.$

Substitution of $x^* = A^*y^*$ and $y = Ax$ into (3.6a) and (3.6b) transforms the equilibrium conditions in (IV) into (1.6). Other specializations of the model problems may be found in 7 and 8.

An elementary fact about (I) and (II) will now be proved.

THEOREM 3-A

$$\inf_x \{f(x)-g(Ax)\} \geq \sup_{y^*} \{g^*(y^*)-f^*(A^*y^*)\} .$$

Proof: By 2-D and its concave analog, it is always true that

$$f(x) + f^*(A^*y^*) \geq [x, A^*y^*] = [Ax, y^*] \geq g(Ax) + g^*(y^*).$$

The theorem is an immediate consequence of this.

Notice that 3-A provides a method of estimating the extrema in (I) and (II). Namely, if $x_1 \in E$ and $y_1^* \in F^*$ then by 3-A

(3.7) $\quad f(x_1) - g(Ax_1) \geq \lambda_0 \geq \lambda_0^* \geq g^*(y_1^*) - f^*(A^*y_1^*),$

where

(3.8) $\quad \lambda_0 = \inf_x \{f(x)-g(Ax)\}, \quad \lambda_0^* = \sup_{y^*} \{g^*(y^*)-f^*(A^*y^*)\} .$

In particular, if the implicit constraints (3.3a) and (3.3b) can be satisfied, then λ_0 and λ_0^* must both be finite.

Suppose an algorithm were known for solving (I) approximately, i.e. an algorithm which constructs a sequence of vectors x_k such that

$\lambda_k = f(x_k) - g(Ax_k)$ decreases to λ_0. The same algorithm could be used to construct a sequence of vectors y_k^* such that $\lambda_k^* = g^*(y_k^*) - f^*(A^* y_k^*)$ increases to λ_0^*. Then, as in (3.7),

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k \rightarrow \lambda_0 \geq \lambda_0^* \leftarrow \lambda_k^* \geq \ldots \geq \lambda_2^* \geq \lambda_1^*,$$

so at each stage one would have upper and lower bounds on both λ_0 and λ_0^*. If λ_0 and λ_0^* are finite and $\lambda_0 = \lambda_0^*$, by continuing the algorithm until $\lambda_k - \lambda_k^* < \varepsilon$ one could obtain approximate solutions x_k and y_k^* to (I) and (II), respectively, with errors $\lambda_k - \lambda_0 < \varepsilon$ and $\lambda_0^* - \lambda_k^* < \varepsilon$. (Similar uses of dual programs have been proposed by others, e.g. Duffin [14].)

The above discussion points out one good reason for wanting to know when 3-A actually holds with equality. An even stronger reason is given by the theorem that we shall prove next. The fact that extra hypotheses are needed to guarantee equality in 3-A is one of the results of §4.

THEOREM 3-B

The following conditions are equivalent to each other:

(a) x_0 and y_0^* are solutions to the dual programs (I) and (II), respectively, and $\inf_x \{f(x)-g(Ax)\} = \sup_{y^*} \{g^*(y^*)-f^*(A^*y^*)\}$,

(b) $\langle x_0, y_0^* \rangle$ is a solution to the game problem (III),

(c) $\langle x_0, y_0^* \rangle$ is a solution to the equilibrium problem (IV).

<u>Proof:</u> (a) implies (b): If (a) is true then certainly $x_0 \in \text{dom } f$ and $y_0^* \in \text{dom } g^*$ by definitions (3.1) and (3.2). Now by 2-D and its concave analog, we have, for each $x \in \text{dom } f$ and $y^* \in \text{dom } g^*$,

$$L(x_0,y^*) = f(x_0) + g^*(y^*) - [Ax_0,y^*] \le f(x_0) - g(Ax_0),$$

$$L(x,y_0^*) = f(x) + g^*(y_0^*) - [x,A^*y_0^*] \ge g^*(y_0^*) - f^*(A^*y_0^*).$$

But the right sides of these inequalities are equal according to (a). Therefore

$$(3.10) \qquad L(x_0,y_0^*) = f(x_0) - g(Ax_0) = g^*(y_0) - f^*(A^*y_0^*), \text{ and}$$

$\langle x_0,y_0^*\rangle$ is a saddle-point for L by definition 1-A. Hence (b) holds.

(b) implies (c): If $\langle x_0,y_0^*\rangle$ is a saddle-point for L, then by definition $x_0 \in$ dom f, $y_0^* \in$ dom g^*, and

$$f(x_0) + g^*(y^*) - [Ax_0,y^*] \le f(x_0) + g^*(y_0^*) - [Ax_0,y_0^*]$$

$$\le f(x) + g^*(y_0^*) - [Ax,y_0^*]$$

for all $x \in$ dom f and $y^* \in$ dom g^*. It follows that

$$g^*(y^*) \le g^*(y_0^*) + [Ax_0,y^*-y_0^*] \text{ for all } y^* \in F^*,$$

$$f(x) \ge f(x_0) + [x-x_0,A^*y_0^*] \text{ for all } x \in E.$$

These are just the equilibrium conditions, according to definition 2-G and its concave analog. Thus (c) holds.

(c) implies (a): Theorem 2-H allows us to express the equilibrium conditions as

$$(3.11) \qquad f(x) + f^*(A^*y^*) \le [x,A^*y^*] = [Ax,y^*] \le g(Ax) + g^*(y^*).$$

Hence if x_0 and y_0^* satisfy the equilibrium conditions we have

$$f(x_0) - g(Ax_0) \le g^*(y_0^*) - f^*(A^*y_0^*).$$

This yields (a) because of 3-A.

COROLLARY 3-C

If (III) has a solution at all, then the minimax value (1.4) of the Lagrangian function coincides with both the minimum in (I) and the maximum in (II). Moreover, the saddle-points can then be found by solving programs (I) and (II) or can be determined from the equilibrium conditions.

Proof: We noted this in the proof of Theorem 3-B, in (3.10).

Theorem 3-B says that problems (III) and (IV) are equivalent to one another, and that both are equivalent to a combined version of (I) and (II). The duality theorems proved later on are aimed at showing us what extent this combined version of (I) and (II) is equivalent to (I) and to (II) individually.

SECTION FOUR

General Equivalence and Duality Theorems

The theorems proved below are the central results of this paper. First of all, we shall study the general relationship between the extrema in programs (I) and (II). In the case already treated by Fenchel, where $E = F$, $E^* = F^*$, $A = I$ is the identity matrix, and both extrema are finite, Fenchel had asserted that the extrema would always be equal [19, p.106] (see also Karlin's account [27, p.229]). This is not true, as we shall prove by counter-example (4-C). (The error stemmed from assuming that a certain formula [19, p.95], [27, Theorem 7.6.1], was valid at all relative boundary points of the set where it was defined. C. Witzgall also observed the error recently and reported it to Karlin. Fenchel has pointed out to the author that a similar error occurs in another formula [19, p.97], [27, Theorem 7.6.2].) The precise nature of the possible discrepancy between the two extrema will be determined in Theorem 4-B. By a separate argument not relying on the formula in error, Fenchel proved (in the case mentioned above) that the extrema would be equal and attained if

$$\text{ri}(\text{dom } f) \cap \text{ri}(\text{dom } g) \neq \emptyset \text{ and } \text{ri}(\text{dom } f^*) \cap \text{ri}(\text{dom } g^*) \neq \emptyset$$

(see [19, p.108] or [27, p.228]). The extra hypothesis requires the implicit constraints to be consistent in a strong sense. We shall show that this result can be extended to the present case, although Fenchel's argument itself does not carry over. It will be demonstrated that the generalized "strong consistency" conditions also guarantee that problems (I) through (IV) are equivalent, i.e. that the solutions to all of them may be found by solving any one of them. Dual versions of the "strong consistency" conditions will also be derived.

"Strong consistency" can actually be replaced by a less restrictive notion of "stable consistency" which, however, is not as simple to apply. This will be proved in §5 (Theorem 5-J) using the results obtained here.

Given the convex program (I) defined in §3, we now consider an associated family of convex programs depending on a parameter $z \in F$:

(I') minimize $f(x) - g(Ax-z)$ on E.

When $z = 0$, program (I') coincides with (I). The lemma below describes the properties of the function $h(z)$ giving the infimum in program (I'). These properties will be crucial in later proofs.

LEMMA 4-A

For each $z \in F$, let

$$h(z) = \inf_x \{f(x)-g(Ax-z)\} .$$

Then h is a convex function on F and

(a) $\text{dom } h = A(\text{dom } f) - \text{dom } g = \{Ax-y \mid x \in \text{dom } f,\ y \in \text{dom } g\}$,

(b) $h^*(y^*) = f^*(A^*y^*) - g^*(y^*)$ for each $y^* \in F^*$,

(c) $h(0) = \inf_x \{f(x) - g(Ax)\}$,

(d) $h^{**}(0) = \sup_{y^*} \{g^*(y^*) - f^*(A^*y^*)\}$.

<u>Proof</u>: To prove h is convex, it will be enough to prove that

(4.1) if $h(z_1) < \mu_1 < \infty$, $h(z_2) < \mu < \infty$, $0 < \lambda < 1$, then

$$h(\lambda z_1+(1-\lambda)z_2) < \lambda\mu_1 + (1-\lambda)\mu_2,$$

since this implies the slightly stronger property required in Definition 2-A. By the definition of h, the hypothesis of (4.1) implies the existence of real numbers μ_{11}, μ_{12}, μ_{21}, μ_{22}, and vectors x_1 and x_2 in E, such that

$$\mu_1 = \mu_{11} - \mu_{12},\ f(x_1) < \mu_{11} < \infty,\ g(Ax_1-z_1) > \mu_{12} > -\infty,$$
$$\mu_2 = \mu_{21} - \mu_{22},\ f(x_2) < \mu_{21} < \infty,\ g(Ax_2-z_2) < \mu_{22} > -\infty.$$

Since f is convex and g is concave,

$$f(\lambda x_1+(1-\lambda)x_2) < \lambda\mu_{11} + (1-\lambda)\mu_{21},$$
$$g(A(\lambda x_1+(1-\lambda)x_2) - (\lambda z_1+(1-\lambda)z_2)=$$
$$g(\lambda(Ax_1-z_1) + (1-\lambda)(Ax_2-z_2)) > \lambda\mu_{12}+ (1-\lambda)\mu_{22}.$$

Therefore, for $x = \lambda x_1 + (1-\lambda)x_2$,

$$h(\lambda z+(1-\lambda)z_2) \le f(x) - g(Ax-(\lambda z_1+(1-\lambda)z_2))$$
$$< \lambda\mu_{11} + (1-\lambda)\mu_{21} - (\lambda\mu_{12}+(1-\lambda))\mu_{22}= \lambda\mu_1 + (1-\lambda)\mu_2$$

Thus h is convex on E as asserted.

Observe next that

$$\text{dom } h = \{z|h(z) < \infty\} = \{z|f(x)-g(Ax-z)< \infty \text{ for some } x\}$$
$$= \{z|Ax-z=y\in \text{dom } g \text{ for some } x\in \text{dom } f\}.$$

This verifies (a).

The conjugate of h, according to Definition 2-C, is given by the formula:

$$h^*(y^*) = \sup_z \{[z,y^*]-h(z)\}$$
$$= \sup_z \{[z,y^*]-\inf_x \{f(x)-g(Ax-z)\}\}$$
$$= \sup \{[z,y^*]-f(x)+g(Ax-z)|x\in E,\ z\in F\}.$$

Taking a supremum over all pairs $\langle x,z\rangle$ is the same as taking it over all pairs $\langle x,w\rangle$ where $z = Ax-w$. Therefore

$$h^*(y^*) = \sup\{[Ax-w,y^*]-f(x)+g(w) \mid x\in E, w\in F\}$$

$$= \sup\{([x,A^*y^*]-f(x))-([x,y^*]-g(w)) \mid x\in E, w\in F\}$$

$$= f^*(A^*y^*)-g^*(y^*)$$

by the definition of the conjugates f^* and g^*. Thus (b) is true.

Finally, (c) is obvious from the definition of h while (d) is immediate from (b) and the definition 2-C of h^{**}.

We now prove that the equality of the extrema in programs (I) and (II) depends on the behavior of the infimum in (I') as the parameter z approaches 0.

THEOREM 4-B (Weak Duality Theorem)

(a) $$\liminf_{Ax-y\to 0}\{f(x)-g(y)\} = \sup_{y^*}\{g^*(y^*)-f^*(A^*y^*)\}$$

except when, trivially, the left side is $+\infty$ and the right side is $-\infty$.

(b) $$\inf_x\{f(x)-g(Ax)\} = \limsup_{A^*y^*-x^*\to 0}\{g^*(y^*)-f^*(x^*)\}$$

except when, trivially, the left side is $+\infty$ and the right side is $-\infty$.

(The notation in programs (I) and (II) is assumed here.)

Proof: For the function h in Lemma 4-A, we have

(4.3) $$\liminf_{Ax-y\to 0} f(x)-g(y) = \liminf_{z\to 0} h(z) = \mathrm{cl}\, h(0).$$

Thus (a) holds if and only if $\mathrm{cl}\, h(0) = h^{**}(0)$, by part (b) of

Lemma 4-A. But h is a convex function, so this is true by Theorem 2-D and the remarks following it, except when cl $h(0) = \infty$ and $h^{**}(0) = -\infty$. This proves (a); (b) is proved by a dual argument, valid because f and g are in turn the conjugates of f^* and g^*. (Recall that f and g are closed and proper in (I), so that $f^{**} = f$ and $g^{**} = g$ by 2-D.)

The problem of determining the "lin inf" in (a) may be thought of as a weaker form of program (I), in which the implicit constraints (3.3a) need only be satisfied "in the limit". Theorem 4-B says that (II) is really the dual of this weaker problem, rather than the dual of (I), while (I) is really the dual of the corresponding weaker version of (II). The extrema in (I) and (II) will not be equal, therefore, unless the weaker versions of (I) and (II) are equivalent to (I) and (II) themselves. The next theorem furnishes examples where the problems fail to be equivalent; the third example, in particular, contradicts the result of Fenchel mentioned above.

THEOREM 4-C

The following situations are indeed possible in (I) and (II), with $\lambda_0 = \inf_x \{f(x)-g(Ax)\}$ and $\lambda_0^* = \sup_{y^*} \{g^*(y^*)-f^*(A^*y^*)\}$:

(a) $\infty > \lambda_0 > \lambda_0^* = -\infty$,

(b) $\infty = \lambda_0 > \lambda_0^* > -\infty$,

(c) $\infty > \lambda_0 > \lambda_0^* > -\infty$,

(d) $\infty = \lambda_0 > \lambda_0^* = -\infty$.

($\lambda_0=\lambda_0^*$ in all other cases by Theorem 3-A.)

Proof: In each case we taken $E = F = R^2$ and take A to be the identity matrix.

(a) Choose $f_1(x) = f_1(\xi_1,\xi_2) = \check{\delta}(\xi_1,\xi_2|\xi_1=0)$

$$g_1(x) = g_1(\xi_1,\xi_2) = \begin{cases} (\xi_1\xi_2)^{1/2} & \text{if } \xi_1\geq 0, \xi_2\geq 0, \\ -\infty & \text{otherwise.} \end{cases}$$

Then obviously

$$(4.4) \qquad \lambda_0 = \inf\left\{(\xi_1\ \xi_2)^{1/2} \,|\, \xi_1=0, \xi_2\geq 0\right\} = 0.$$

But if we let

$$(4.5) \qquad x_k = \langle 0,k^3\rangle\ ,\ y_k = \langle k^{-1},k^3\rangle\ ,\ k = 1,2,\ldots,$$

then $x_k - y_k \to 0$ but $f_1(x_k) - g_1(y_k) = -k$ for all k. Therefore, by part (a) of Theorem 4-B,

$$-\infty = \liminf_{x - y\to 0} \left\{f_1(x)-g_1(y)\right\} = \lambda_0^*.$$

(b) This is dual to (a).

(c) Let $f_3 = f_1$ and $g_3(x) = \min\left\{1,g_1(x)\right\}$. Again taking the sequences in (4.5), we have $f_3(x_k) - g_3(y_k) = -1$ for all k. But evidently $f_3(x) - g_3(y) \geq -1$ for all $x\in E$ and $y\in E$, so by part (a) of Theorem 4-B,

$$-1 = \liminf_{y-x\to 0} \left\{f_3(x)-g_3(y)\right\} = \lambda_0^*,$$

while $\lambda_0 = 0$ as in (4.4).

(d) Let $f_4 = f_1$, $g_4(\xi_1,\xi_2) = \xi_2 + \check{\delta}(\xi_1,\xi_2|\xi_1\geq 0,\xi_2\geq 0,\xi_1\xi_2\geq 1)$. This time $\operatorname{dom} f_4 \cap \operatorname{dom} g_4 = \emptyset$, so $\lambda_0 = \infty$ trivially. On the other hand, for the sequences in (4.8) we have $f_4(x_k)-g_4(y_k) = -k^3$ for all k, and therefore

$$-\infty = \liminf_{x-y \to 0} \left\{f_4(x)-g_4(y)\right\} = \lambda_0^*$$

by part (a) of Theorem 4-B.

A stronger theorem than 4-B will now be derived using additional assumptions.

DEFINITION 4-D

Program (I) will be called <u>strongly consistent</u> if there exists some $x \in \mathrm{ri}(\mathrm{dom}\ f)$ such that $Ax \in \mathrm{ri}(\mathrm{dom}\ g)$. Dually, program (II) will be called <u>strongly consistent</u> if there exists some $y^* \in \mathrm{ri}(\mathrm{dom}\ g^*)$ such that $A^*y^* \in \mathrm{ri}(\mathrm{dom}\ f^*)$.

THEOREM 4-E (Duality Theorem)

(a) If program (I) is strongly consistent, then

$$\inf_x \left\{ f(x)-g(Ax) \right\} = \max_{y^*} \left\{ g^*(y^*)-f^*(A^*y^*) \right\} .$$

(b) If program (II) is strongly consistent, then

$$\min_x \left\{ f(x)-g(Ax) \right\} = \sup_{y^*} \left\{ g^*(y^*)-f^*(A^*y^*) \right\} .$$

<u>Proof:</u> (a) Let h be the convex function in Lemma 4-A. Then, by 4-A(a) and two general facts about relative interiors proved in Appendix A (namely A-D and A-G),

$$\mathrm{ri}(\mathrm{dom}\ h) = \mathrm{ri}(A(\mathrm{dom}\ f)-\mathrm{dom}\ g) = \mathrm{ri}(A(\mathrm{dom}\ f))-\mathrm{ri}(\mathrm{dom}\ g)$$

$$= A(\mathrm{ri}(\mathrm{dom}\ f))-\mathrm{ri}(\mathrm{dom}\ g) = \left\{ Ax-y \mid x \in \mathrm{ri}(\mathrm{dom}\ f), y \in \mathrm{ri}(\mathrm{dom}\ g) \right\} .$$

The assumption that (I) is strongly consistent is therefore equivalent to the assumption that

(4.6) $\quad 0 \in \mathrm{ri}(\mathrm{dom}\ h)$.

If h is improper, (4.6) implies that $h(0) = -\infty = h^{**}(0)$ by (2.7) and the remarks after Theorem 2-D. In this case, (a) follows immediately from parts (c) and (d) of 4-A. On the other hand, suppose h is proper. Then $h(0) = h^{**}(0)$ again by (4.6), (2.7) and 2-D, and hence "inf" = "sup" in (a) by Lemma 4-A. Moreover (4.6) implies, via 2-J, that

$$-\infty < \min_{y^*} h^*(y^*) < \infty.$$

Since $h^*(y^*) = f^*(A^*y^*) - g^*(y^*)$ by 4-A(b), we can therefore replace "sup" by "max" in (a). Part (b) is proved by a dual argument.

COROLLARY 4-F

If programs (I) and (II) are both strongly consistent, then

$$\infty > \min_x \{f(x)-g(Ax)\} = \max_{y^*} \{g^*(y^*)-f^*(A^*y^*)\} > -\infty$$

and, in particular, both programs have solutions.

THEOREM 4-G (Equivalence Theroem)

Suppose that (I) is strongly consistent. Then the following conditions on x_0 are equivalent:

(a) x_0 is a solution to the convex program (I),

(b) there exists some y_0^* such that $\langle x_0, y_0^* \rangle$ is a saddle-point for the Lagrangian function in (III),

(c) there exists some y_0^* such that x_0 and y_0^* satisfy the equilibrium conditions in (IV).

Moreover the vectors y_0^* occurring in these conditions are precisely the solutions to the concave program (II).

The dual theorem, in which (II) is assumed to be strongly consistent instead of (I), is also valid.

Proof: Duality Theorem 4-E(a) says that (a) is equivalent to the longer statement:

(a') there exists some y_0^* such that x_0 and y_0^* are solutions to (I) and (II) respectively, and

$$\inf_x \{f(x)-g(Ax)\} = \sup_{y^*} \{g^*(y^*)-f^*(A^*y^*)\} .$$

But (a'), (b), and (c) are equivalent by Theorem 3-B.

THEOREM 4-H (Minimax Theorem)

If (I) and (II) are strongly consistent, then the minimax value of the Lagrangian function in (II) exists and equals both the minimum in (I) and the maximum in (II).

Proof: This is immediate from 4-F, 3-B and 3-C.

In view of the importance of the strong consistency conditions in applying the last several results, one would like to know what property of (I) is equivalent to the strong consistency of (II), and dually. This question is answered below.

THEOREM 4-I

(a) Choose arbitrary vectors $x_0 \in$ dom f and $y_0 \in$ dom g, and let $h(x) = f(x_0+x)-g(y_0+Ax)$. Then program (II) is strongly consistent if and only if, for each x such that $h(\lambda x)$ is a finite non-increasing function of λ for $0 < \lambda < \infty$, $h(\lambda x)$ is actually constant for $-\infty < \lambda < \infty$.

(b) Choose arbitrary vectors $y_0^* \in$ dom g^* and $x_0^* \in$ dom f^*, and let $h(y^*) = g^*(y_0^*+y^*) - f^*(x_0^*+A^*y^*)$. Then program (I) is strongly consistent if and only if, for each y^* such that $h(\lambda y^*)$ is finite non-decreasing function of λ for $0 < \lambda < \infty$, $h(\lambda y^*)$ is actually constant for $-\infty < \lambda < \infty$.

Proof: We shall deduce this from a general fact proved in Appendix B (Theorem B-F) about the effective domain of the conjugate of a convex function. Let

(4.7) $\quad k(x,y) = f(x) - g(y),\ M = \{\langle x,y\rangle \mid y=Ax\}.$

Then k is a closed proper convex function on $E \oplus F$, $\langle x_0,y_0\rangle \in \operatorname{dom} k$, and M is a subspace of $E \oplus F$. The subspace of $E^* \oplus F^*$ orthogonal to M is

$$\begin{aligned} M^* &= \{\langle x^*,y^*\rangle \mid [x,x^*]+[y,y^*]=0 \text{ for all } \langle x,y\rangle \in M\} \\ &= \{\langle x^*,y^*\rangle \mid [x,x^*]=-[Ax,y^*] \text{ for all } x \in E\} \\ &= \{\langle x^*,-y^*\rangle \mid x^*=A^*y^*\}. \end{aligned}$$

On the other hand, by Theorem 2-L and (2.18) (or by direct calculation)

$$k^*(x^*,y^*) = f^*(x^*) - g^*(-y^*),\ \operatorname{dom} k^* = \{\langle x^*,-y^*\rangle \mid x^* \in \operatorname{dom} f^*,\ y^* \in \operatorname{dom} g^*\},$$

and hence

$$\operatorname{ri}(\operatorname{dom} k^*) = \{\langle x^*,-y^*\rangle \mid x^* \in \operatorname{ri}(\operatorname{dom} f^*), y^* \in \operatorname{ri}(\operatorname{dom} g^*)\}$$

(see Theorem A-H in Appendix A). Therefore M^* intersects $\operatorname{ri}(\operatorname{dom} k^*)$ if and only if

$$A^*y^* = x^* \in \operatorname{ri}(\operatorname{dom} f^*) \text{ for some } y^* \in \operatorname{ri}(\operatorname{dom} g^*),$$

i.e. if and only if (II) is strongly consistent. Now we apply the general theorem B-F indicated above. According to this theorem, M^* intersects $\operatorname{ri}(\operatorname{dom} k^*)$ if and only if, for each $\langle x,y\rangle \in M$ such that $k(x_0+\lambda x, y_0+\lambda y)$ is a finite non-increasing function for $0 < \lambda < \infty$.

$k(x_0+\lambda x, y_0+\lambda y)$ is actually constant for $-\infty < \lambda < \infty$. Substitution of (4.7) into this statement proves (a). A dual argument proves (b).

SECTION FIVE

Stable Consistency

Theorems 4-E, 4-F, 4-G, and 4-H in the last section all depend on assumptions of strong consistency, ad defined in 4-D. The strong consistency conditions, which require that the implicit constraints can be satisfied "with some to spare", were dualized in Theorem 4-I in order to make them easier to apply. Nevertheless, these conditions are inconvenient in certain situations because they are too restrictive. The duality theorem for linear programs, for instance, needs no strong consistency assumptions; hence it does not completely follow when Theorem 4-F is applied to linear programs as formulated in §3. We shall prove here, however, that "strong consistency" can be replaced by a far weaker (but rather more complicated) notion of "stable consistency", which takes advantage of the special properties of a class of "stable" convex functions. In this way we obtain duality and equivalence theorems (see 5-J) which generalize the simpler ones in §4, and at the same time are powerful enough to contain the linear programming theorems and other presently known results as easy corollaries (see §7).

We begin by considering a useful class of functions whose definition is motivated by an interesting argument of Fenchel [19, p.113-115] for deriving the linear programming duality theorem. Fenchel's argument is weak in two respects. In the first place, it rests on the invalid result discussed in §4. Secondly, it assumes that a "piecewise linear" convex function which is bounded below always attains its minimum. This is not altogether clear, as the following example indicates. Let E be two-dimensional, and let

$$f(\xi_1,\xi_2) = \xi_1 + \check{\delta}(\xi_1,\xi_2 | \xi_1 \geq 0,\ \xi_2 \geq 0,\ \xi_1\xi_2 \geq 1).$$

Although f would seem to be "piecewise linear", its minimum is approached at best along a hyperbolic path. Evidently, "piecewise linearity" ought to somehow take effective domains into account. Now for convex sets, the nearest thing to "piecewise linearity" is the familiar "polyhedral" property. A convex set C is said to be polyhedral if it can be represented as the intersection of finitely many closed half-spaces:

$$(5.1) \qquad C = \left\{ x \mid [x,a_i^*] \leq \alpha_i,\ i = 1,\ldots,n \right\}.$$

It is known [22] that C is polyhedral if and only if it can be expressed as the convex hull of finitely many points and rays:

$$(5.2) \qquad C = \left\{ x=\lambda_1 b_i + \ldots + \lambda_k b_k + \mu_i a_1 + \ldots \mu_m a_m \mid \lambda_j \geq 0, \sum \lambda_j = 1, \mu_i \geq 0 \right\}.$$

This suggests the following notion as a replacement for "piecewise linearity."

DEFINITION 5-A

A polyhedral convex function on E is a function f whose upper graph set gph f is a polyhedral convex set in $E \oplus R$.

Applying (5.1) to gph f, one can readily see that f is polyhedral if and only if it can be represented in the form

$$(5.3) \qquad f(x) = \max \left\{ [x,b_j^*] - \beta_j^* \mid j=1,\ldots,k \right\} + \check{\delta}(x \mid [x,a_i^*] \leq \alpha_i, i=1,\ldots,m).$$

(In the improper case where $k = 0$, set f identically $-\infty$ where the characteristic function is zero in this formula.) Dually, by (5.2), f is polyhedral if and only if it can be represented in the form

$$(5.4) \qquad f(x) = \min\{\lambda_1\beta_1+\dots+\lambda_k\beta_k+\mu_1\alpha_1+\dots\mu_m\alpha_m + \check{\delta}(\lambda_1,\dots,\lambda_k;\mu_1,\dots,\mu_m \mid x=\sum\lambda_j b_j+\sum\mu_i a_i,\ \lambda_j\geq 0,\ \sum\lambda_j=1, \mu_i\geq 0)\},$$

This fact was already used by the author elsewhere [31]. (Needless to say, polyhedral convex functions can arise from more complicated mixed expressions as well.)

THEOREM 5-B

If f is a polyhedral convex function, then dom f is a polyhedral convex set. Moreover f and dom f are closed.

Proof: Obvious from representation (5.3).

THEOREM 5-C

If f is a polyhedral convex function on E, then f^* is a polyhedral convex function on E^*, and $f^{**} = f$ when f is proper. Furthermore, the fundamental representations (5.3) and (5.4) are dual to one another.

Proof: By calculating f^* directly from Definition 2-C using a representation (5.4) for f, one obtains a representation (5.3) for f^*. The fact that $f^{**} = f$ when f is proper follows from Theorems 2-D and 5-B.

One of the most common polyhedral convex functions in extremum problems is

$$\check{\delta}(x \mid x \geq 0) = \begin{cases} 0 \text{ if } x \geq 0, \\ \infty \text{ if } x \ngeq 0. \end{cases}$$

It should be noted that the functions f and g used to formulate the linear programs in §3 (see (3.5a) and (3.5b)) are polyhedral.

It is reasonable to expect that the strong consistency assumptions in §4 can be weakened when polyhedral functions are involved. Actually, we shall be able to bring about this weakening for a far broader class of functions.

DEFINITION 5-D

Let f be a closed proper convex function on E. For each $x \in \operatorname{dom} f$ and each subspace M in E, let

(5.5) $$f_{x,M}(z) = f(z) + \check{\delta}(z-x \mid M) \text{ for all } z \in E.$$

(Then $f_{x,M}$ is a proper convex function on E which may be thought of as the restriction of f to the linear manifold x+M.) Let M^* denote the subspace of E^* orthogonal to M, i.e.

(5.6) $$M^* = \left\{ z^* \mid [z,z^*] = 0 \text{ for all } z \in M \right\}.$$

We shall say that f is <u>stable</u> if

(5.7) $$(f_{x,M})^*(x^*) = \min_{z^* \in M^*} \left\{ f^*(x^*+z^*) - [x,z^*] \right\} \text{ for all } x^* \in E^*,$$

for every $x \in \operatorname{dom} f$ and subspace M. If f and f^* are both stable, we shall say that f is <u>completely stable</u>.

The lemma below shows that this definition is "reasonable."

LEMMA 5-E

Let f be a closed proper convex function on E and let $x \in \operatorname{dom} f$. Let M, M^*, and $f_{x,M}$ be as in Definitin 5-D. Then

(5.8) $$(f_{x,M})^*(x^*) = \operatorname{cl}_{x^*} \inf_{z^* \in M^*} \{ f^*(x^*+z^*) - [x,x^*] \} \text{ for all } x^* \in E^*.$$

Moreover (5.8) can be strengthened to (5.7) if the linear manifold $x + M$ intersects ri(dom f).

<u>Proof:</u> Let $h(x^*) = \inf \{ f^*(x^*+z^*) - [x,z^*] \mid z^* \in M^* \}$. Then h is a convex function on E^*, as one can show by the argument used in Lemma 4-A. Furthermore,

$$h^*(z) = \sup_{x^*} \{ [z,x^*] - h(x^*) \} = f_{x,M}(z)$$

by straightforward calculation. Since the latter function is proper, Theorem 2-D and the remarks after it imply that

$$(f_{x,M})^* = h^{**} = \operatorname{cl} h,$$

which is just formula (5.8). The final statement of the lemma must be proved by a different argument. Let $x^* \in E^*$ and let

$$g(z) = [z,x^*] - \check{\delta}(z-x \mid M) \text{ for all } z \in E.$$

Then g is a closed proper concave function on E and

$$g^*(z^*) = [x,z^*-x^*] - \check{\delta}(z^*-x^* \mid M^*) \text{ for all } z^* \in E^*$$

(as one readily calculates from Definition 2-M). Moreover $\operatorname{dom} g = \operatorname{ri}(\operatorname{dom} g) = x + M$. If the linear manifold $x + M$ intersects ri(dom f), we can apply Theorem 4-E(a) to f and g (with E=F, $E^*=F^*$, A=I) to obtain:

(5.9) $$\inf_z \left\{f(z)-g(z)\right\} = \max_{y^*} \left\{g^*(y^*)-f^*(y^*)\right\}.$$

The left side of (5.9) is $-(f_{x,M})^*(x^*)$ by definition, while the right side is

$$-\min_{y^*} \left\{f^*(y^*)-[x,y^*-x^*]+\check{\delta}(y^*-x^*|M^*)\right\}$$

$$= \quad -\min_{z^*\in M^*} \left\{f^*(x^*+z^*)-[x,z^*]\right\}.$$

Hence (5.9) is equivalent to (5.7).

Thus a closed proper convex function f fails to be stable only when (5.8) cannot be strengthened to (5.7) for some x and M such that $x\in$ dom f but $(x+M)\cap$ ri(dom f) = $\emptyset$.

THEOREM 5-F

Let f be a closed proper convex function on E such that ri(dom f) = dom f. Then f is stable.

Proof: In this event, the exceptional cases just mentioned cannot occur.

Remark: Certainly ri(dom f) = dom f if dom f is a linear manifold, and in particular if dom f = E. If both dom f and dom f^* are linear manifolds, then f is completely stable. The properties that f must have in order that dom f^* be a linear manifold, or dom $f^* = E^*$, are given in Appendix B (see Corollaries B-C and B-E). Completely regular convex functions (see C-I and C-M in Appendix C) are completely stable. All quadratic convex functions are completely stable (see B-E and also Appendix C). In § 6 we shall show that complete stability is preserved by various operations, such as addition.

THEOREM 5-G

Every proper polyhedral convex function on E is completely stable.

Proof: If f is proper and polyhedral, then f is closed by Theorem 5-B and f^* is polyhedral by Theorem 5-C. Hence f^* can be represented as in (5.4). By choosing a basis for M^*, we can then also represent the function h in the proof of Lemma 5-E as in (5.4). It follows that "inf" can be replaced by "min" in the formula for h, and that h is polyhedral. But the latter implies that h is closed, so (5.8) can be strengthened to (5.7). Thus f is stable. Now f^* must also be stable, since it is also polyhedral and proper. Hence f is completely stable.

We now define the condition which, as will be seen below, can be substituted for strong consistency.

DEFINITION 5-H

Suppose that the functions f and g in program (I) can be expressed by

(5.10) $f(x) = f_0(x) + f_1(x)$ for all $x \in E$,

$g(y) = g_0(y) + g_1(y)$ for all $y \in F$,

where f_0 is a stable convex function on E, f_1 is a closed proper convex function on E, g_0 is stable convex function on F, and g_1 is a closed proper concave function on F. Suppose that there exists some x such that

(5.11) $x \in \text{dom } f_0 \cap \text{ri}(\text{dom } f_1)$ and $Ax \in \text{dom } g_0 \cap \text{ri}(\text{dom } g_1)$.

Then we shall say program (I) is stably consistent. Stable consistency is defined similarly for program (II).

THEOREM 5-I

Strong consistency implies stable consistency. Also, if f and g are themselves stable in (I) (in particular if f and g are polyhedral), then (I) is stably consistent whenever it is merely consistent, i.e. whenever there exists some x such that

(5.12) $x \in \text{dom } f$ and $Ax \in \text{dom } g$.

Proof: If (I) is strongly consistent, choose f_0 and g_0 identically zero in (5.10). Then f_0 and g_0 are stable (by either 5-F or 5-G) and (5.11) holds. Hence (I) is stably consistent. On the other hand, if f and g are stable, choose f_1 and g_1 identically zero in (5.10). Then (5.11) coincides with (5.12).

The main result of this section is the following.

THEOREM 5-J

"Strong consistency" can be replaced by "stable consistency" in Theorems 4-E, 4-F, 4-G and 4-H.

COROLLARY 5-K (Fundamental Theorem for "Completely Stable" Programs)

Suppose that f and g are completely stable (or, in particular, polyhedral) in (I). Suppose that any one of the following holds:

(a) $\infty > \inf_x \{f(x)-g(Ax)\} > -\infty$,

(b) $\infty > \sup_{y^*} \{g^*(y^*)-f^*(A^*y^*)\} > -\infty$,

(c) $\infty > \inf_x \{f(x)-g(Ax)\}$ and $\sup_{y^*} \{g^*(y^*)-f^*(A^*y^*)\} > -\infty$.

Then both (I) and (II) have solutions and

(5.13) $\infty > \min_x \{f(x)-g(Ax)\} = \max_{y^*} \{g^*(y^*)-f^*(A^*y^*)\} > -\infty$.

Moreover the conclusions of 4-H, 4-G and the dual of 4-G are then valid.

Proof of the Corollary: (5.12) holds if and only if the infimum in (I) is not $+\infty$, as pointed out in §3. When f and g are completely stable, it follows from 5-I and 5-J (applied to 4-E(a)) that (a) is equivalent to

$$\infty > \inf_x \{f(x)-g(Ax)\} = \max_{y^*} \{g^*(y^*)-f^*(A^*y^*)\} > -\infty.$$

Since f^* and g^* are also stable (by definition of complete stability), a dual argument shows that (b) is equivalent to

$$\infty > \max_x \{f(x)-g(Ax)\} = \sup_{y^*} \{g^*(y^*)-f^*(A^*y^*)\} > -\infty.$$

In particular (a) and (b) imply each other, so each one actually implies (5.13). Moreover (c) implies both (a) and (b), and hence implies (5.13), by Theorem 3-A. The last assertion of the corollary is a direct consequence of 5-I and 5-J.

This corollary contains all the linear programming theorems described earlier (see also the beginning of 7).

Two lemmas are needed in the proof of Theorem 5-J.

LEMMA 5-L

Let f_0 be a stable convex function on E and let f_1 be any other closed proper convex function on E. If dom $f_0 \cap$ ri(dom f_1) $\neq \emptyset$, then $f_0 + f_1$ is a closed proper convex function and

$$(f_0+f_1)^*(x^*) = \min_{z^*} \{ f_0^*(x^*-z^*)+f_1^*(z^*) \} \text{ for all } x^* \in E^*.$$

Proof: It is elementary that $f_0 + f_1$ is a proper convex function. Let $x \in$ dom $f_0 \cap$ ri(dom f_1) and let M be the subspace of E such that $x + M$ is the smallest linear manifold containing dom f_1. Then

(5.14) $\quad f_0(z) + f_1(z) = f_0'(z) + f_1(z)$ for all $z \in E$,

where

$$f_0'(z) = f_0(z) + \check{\delta}(z-x|M).$$

Moreover

(5.15) $\quad$ ri(dom f_0') $\cap$ ri(dom f_1) $\neq \emptyset$.

Assume for a moment that this has been proved. Then, by an argument very similar to the one used in the last half of Lemma 5-E, one can prove that

(5.16) $$(f_0'+f_1)^*(x^*) = \min_{y^*} \{ (f_0')^*(x^*-y^*)+f_1^*(y^*) \}.$$

On the other hand,

(5.17) $$(f_0')^*(x^*-y^*) = \min_{u^* \in M^*} \{ f_0^*(x^*-y^*+u^*)-[x,u^*] \}$$

by definition of the stability of f_0, with M^* as in (5.6).

Combining (5.14), (5.16) and (5.17), we get

$$(5.18) \qquad (f_0+f_1)^*(x^*) = \min_{y^*} \min_{u^* \in M^*} \{ f_0^*(x^*-y^*+u^*)+f_1^*(y^*)-[x,u^*] \} .$$

But by definition of x, M and M^*,

$$[z,u^*] = [x,u^*] \text{ for all } z \in \operatorname{dom} f_1 \text{ and } u^* \in M^* .$$

Therefore, for all $y^* \in E^*$ and $u^* \in M^*$,

$$f_1^*(y^*-u^*) = \sup_z \{ [z,y^*-u^*]-f_1(z) \}$$

$$= -[x,u^*]+\sup_z \{ [z,y^*]-f_1(z) \} = f_1^*(y^*)-[x,u^*].$$

Applying this fact in (5.18), and replacing $y^* - u^*$ by z^*, we obtain the desired formula in the Lemma. Also, straightforward calculation of $((f_0+f_1)^*)^*$ from this formula yields $f_0 + f_1$ again; hence $f_0 + f_1$ is closed by Theorem 2-D. The proof of the lemma will therefore be complete as soon as (5.15) has been verified. By definition,

$$(5.19) \qquad x \in \operatorname{dom} f_0' \subseteq x + M, \; x \in \operatorname{ri}(\operatorname{dom} f_1).$$

Let $z \in \operatorname{ri}(\operatorname{dom} f_0')$. (Such a z exists because a non-empty convex set has a non-empty relative interior [15, p.16].) Then there exists some λ, $0 < \lambda < 1$, such that

$$z_0 = \lambda x + (1-\lambda)z \in \operatorname{ri}(\operatorname{dom} f_1),$$

because, by (5.19) and the definition of M, $z \in x + M$ and x is an interior point of dom f_1 relative to the linear manifold x + M. But also $z_0 \in \operatorname{ri}(\operatorname{dom} f_0')$. This follows from the general fact that if C is convex, $x \in C$, $z \in \operatorname{ri} C$, $0 < \lambda < 1$, then $\lambda x + (1-\lambda)z \in \operatorname{ri} C$ (see [15, p.9].) Hence (5.15) is true as asserted.

LEMMA 5-M

Let f be a stable convex function on E, let $x_0 \in \operatorname{dom} f$ and let

$x_0^* \in E^*$. Let

$$f'(x) = f(x) + \check{\delta}(x \mid [x,x_0^*] = [x_0,x_0^*]).$$

Then f' is a stable convex function on E and

$$(5.20) \qquad (f')^*(x^*) = \min_{-\infty<\lambda<\infty} \{ f^*(x^*+\lambda x_0^*) - \lambda[x_0,x_0^*] \} \text{ for all } x^* \in E^*.$$

Proof: Let $f_0 = f$ and $f_1(x) = \check{\delta}(x \mid [x,x_0^*] = [x_0,x_0^*])$.

Then $f' = f_0 + f_1$, where f_0 is stable convex function, and f_1 is a closed proper convex function. Also, ri(dom f_1) = dom f_1, and this set intersects dom f_0 according to its definition. Lemma 2-L therefore implies that f' is a closed proper convex function. Furthermore, one sees readily that

$$f_1^*(x^*) = \begin{cases} \lambda[x_0,x_0^*] & \text{if } x^* = \lambda x_0^*, \ -\infty<\lambda<\infty, \\ \infty & \text{otherwise.} \end{cases}$$

Formula (5.20) now follows from Lemma 5-L. We must show next that f' is stable. Let

$$x \in \text{dom } f' = (x_0+M_0) \cap \text{dom } f,$$

where $M_0 = z \mid [z,x_0^*]=0$ is the subspace of E orthogonal to x_0^*. Let M be any other subspace of E. Then

$$f'_{x,M} = f_{x,M'}, \text{ where } M' = M \cap M_0,$$

since the linear manifolds $x + M_0$ and $x_0 + M_0$ coincide. Hence by the stability of f

$$(5.21) \qquad (f'_{x,M})^*(x^*) = (f_{x,M'})^*(x^*) = \min_{u^* \in (M')^*} f^*(x^*+u^*) - [x,u^*].$$

But it follows from the definition of M' that

$$(M')^* = \{ z^*+\lambda x_0^* \mid z^* \in M^*, \ -\infty<\lambda<\infty \},$$

while $[x,x_0^*] = [x_0,x_0^*]$ by definition. Therefore the last term in (5.21) becomes

$$\min\left\{f^*(x^*+\lambda x_0^*+z^*)-\lambda[x_0,x_0^*]-[x,z^*]\,|\,z^*\in M^*,\ -\infty<\lambda<\infty\right\}$$

$$= \min_{z^*\in M^*}\left\{(f')^*(x^*+z^*)-[x,z^*]\right\}$$

by (5.20). Thus f' satisfies the definition of stability.

<u>Proof of Theorem 5-J</u>: Suppose program (I) is stably consistent but not strongly consistent. We shall show that then (I) and (II) can be replaced by an "equivalent" dual pair of programs (I') and (II') which are still stably consistent, and are "significantly nearer" to being strongly consistent.

Let $x_0^*\in$ dom f^* and $y_0^*\in$ dom g^* and let

$$h(y^*) = g^*(y_0^*+y^*) - f^*(x_0^*+A^*y^*).$$

Then h is a proper concave function on F^* such that $0\in$ dom h. Since (I) is not strongly consistent, there exists by Theorem 4-I(b) some $y_1^*\in F^*$ such that $h(\lambda y_1^*)$ is a non-decreasing function of $\lambda > 0$, but is not constant for $-\infty < \lambda < \infty$. Let x_0 be a vector satisfying (5.11). Let

(5.21) $$f'(x) = f(x) + \check{\delta}(x \mid [x,A^*y_1^*] = [x_0,A^*y_1^*]),$$

$$g'(y) = g(y) + \hat{\delta}(y \mid [y,y_1^*] = [Ax_0,y_1^*]).$$

We claim that then

(5.22) $$f'(x) - g'(Ax) = f(x) - g(Ax) \text{ for all } x\in E.$$

To prove this, we show first that

(5.23) $$[x,A^*y_1^*] \le [y,y_1^*] \text{ for all } x\in \text{dom } f \text{ and } y\in \text{dom } g.$$

By definition of y_1^*, we have

(5.24) $$\infty > g^*(y_0^*+\lambda y_1^*) - f^*(x_0^*+\lambda A^*y_1^*) \ge g^*(y_0^*) - f^*(x_0^*)$$

for all $\lambda > 0$. Thereom 2-F and its concave analog imply that, for all $x \in \text{dom } f$ and $y \in \text{dom } g$,

$$g^*(y_0^*+\lambda y_1^*) \leq [y, y_0^*+\lambda y_1^*] - g(y),$$
$$f^*(x_0^*+\lambda A^* y_1^*) \geq [x, x_0^*+\lambda A^* y_1^*] - f(x).$$

Combining this with (5.24) and manipulating it algebraically, we get

$$\lambda([y,y_1^*]-[x,A^*y_1^*]) \geq$$
$$([x,x_0^*]-f(x)-f^*(x_0^*)) - ([y,y_0^*]-g(y)-g^*(y_0^*)).$$

But the expression on the right is finite when $x \in \text{dom } f$ and $y \in \text{dom } g$, and the inequality holds for all $\lambda > 0$. Therefore (5.23) is true. Now we prove (5.22). Whenever the left side of (5.22) is less than $+\infty$, both sides are finite and equal by definition (5.21). Hence, to prove (5.22), we need only show that if the right side of (5.22) is less then $+\infty$, then the left side is also. Let x be a vector such that $f(x) - g(Ax) < \infty$. Then $x \in \text{dom } f$ and $Ax \in \text{dom } g$. The vector x_0 also has the latter property, since it satisfies the even stronger condition (5.11). Consequently

$$[x,A^*y_1^*] \leq [Ax_0,y_1^*] = [x_0,A^*y_1^*] \leq [Ax,y_1^*],$$

by (5.23), so that

$$[x,A^*y_1^*] = [x_0,A^*y_1^*] = [Ax_0,y_1^*] = [Ax,y_1^*].$$

Therefore, by (5.21), the left side of (5.22) must also be less then $+\infty$ for this x, as we wanted to prove. Thus (5.22) is true.

Given the function in (5.10), we now let

(5.25) $$f_0'(x) = f_0(x) + \check{\delta}(x|[x,A^*y_1^*] = [x_0,A^*y_1^*]), \quad f_1' = f_1,$$
$$g_0'(y) = g_0(y) + \hat{\delta}(y|[y,y_1^*] = [Ax_0,y_1^*]), \quad g_1' = g_1.$$

Then f_0' and g_0' are stable by Lemma 5-M, and

(5.26) $\quad f' = f_0' + f_1'$ and $g' = g_0' + g_1'$

just as in (5.10). Furthermore, the vector x_0, which satisfies (5.11) by assumption, also satisfies

(5.27) $\quad x_0 \in \operatorname{dom} f_0' \cap \operatorname{ri}(\operatorname{dom} f_1')$ and $Ax_0 \in \operatorname{dom} g_0' \cap \operatorname{ri}(\operatorname{dom} g_1')$.

It follows now from 5-L that f' is a closed proper convex function on E and g' is a closed proper concave function on F. Hence we can consider the dual programs

(I') $\quad$ minimize $f'(x) - g'(Ax)$ on E,

(II') $\quad$ maximize $(g')^*(y^*) - (f')^*(A^*y^*)$ on F^*.

In view of (5.22), programs (I) and (I') amount to the same thing. Program (I') is also stably consistent, according to the above remarks. We shall compare programs (II) and (II') next. By (5.26), (5.27) and Lemma 5-L, we have

$$(f')^*(x^*) = \min_{z^*} \left\{ (f_0')^*(x^*-z^*)+(f_1')^*(z^*) \right\}.$$

Applying Lemma 5-M to (5.25), we can re-express this as

$$(f')^*(x^*) = \min_{z^*} \min_{-\infty<\lambda<\infty} \left\{ f_0^*(x^*+\lambda A^*y_1^*-z^*)-\lambda[x_0,A^*y_1^*]+f_1^*(z^*) \right\}.$$

But, from (5.10) and Lemma 5-L,

$$f^*(x^*) = \min_{z^*} \left\{ f_0^*(x^*-z^*)+f_1^*(z^*) \right\}.$$

Therefore

(5.28) $$(f')^*(x^*) = \min_{-\infty<\lambda<\infty} \left\{ f^*(x^*+\lambda A^*y_1^*)-\lambda[x_0,A^*y_1^*] \right\}.$$

Actually, the expression in brackets is a non-increasing function of λ. This is proved as follows. By (5.23) and the choice of x_0,

$$[x, A^* y_1^*] \le [Ax_0, y_1^*] = [x_0, A^* y_1^*] \text{ for all } x \in \text{dom } f.$$

Hence, for $-\infty < \lambda < \infty$ and arbitrary $\mu > 0$,

$$\begin{aligned} & f^*(x^* + (\lambda+\mu)A^* y_1^*) - (\lambda+\mu)[x_0, A^* y_1^*] \\ & = -(\lambda+\mu)[x_0, A^* y_1^*] + \sup_x \{[x, x^* + (\lambda+\mu)A^* y_1^*] - f(x)\} \\ & \le -(\lambda+\mu)[x_0, A^* y_1^*] + \mu[x_0, A^* y_1^*] + \sup_x \{[x, x^* + \lambda A^* y_1^*] - f(x)\} \\ & = f^*(x^* + \lambda A^* y_1^*) - \lambda[x_0, A^* y_1^*]. \end{aligned}$$

A dual argument proves that

$$(5.29) \qquad (g')^*(y^*) = \max_{-\infty<\lambda<\infty} \{ g^*(y^* + \lambda y_1^*) - \lambda[Ax_0, y_1^*] \},$$

where the expression in brackets is a non-decreasing function of λ.

Therefore

$$\begin{aligned} (5.30) \qquad & (g')^*(y^*) - (f')^*(A^* y^*) \\ & = \max_{-\infty<\lambda<\infty} \{ g^*(y^* + \lambda y_1) - f^*(A^*(y^* + \lambda y_1)) \} \end{aligned}$$

where, again, the expression in brackets is a non-decreasing function of λ. It follows from this that

$$\sup_{y^*} \{ (g')^*(y^*) - (f')^*(A^* y^*) \} = \sup_{y^*} \{ g^*(y^*) - f^*(A^* y^*) \}$$

and that both of these suprema are attained whenever one is attained. Thus program (II') is "equivalent" to program (II).

We want to show next that (I') and (II') are in a certain sense simpler than (I) and (II). Consider the function h defined at the beginning of the proof. Let

$$M = \{ y^* \mid h(z^* + \lambda y^*) = h(z^*) \text{ for all } z^* \in F^* \text{ and } -\infty<\lambda<\infty \}.$$

It is not difficult to see that M is a subspace of F^*. Next, for the same x_0^* and y_0^* used in the definition of h, let

$$h'(y^*) = (g')^*(y_0^*+y^*) - (f')^*(x_0^*+A^*y^*),$$

and let

$$M' = \left\{ y^* \middle| h'(z^*+\lambda y^*)=h'(z^*) \text{ for all } z^*\epsilon F^* \text{ and } -\infty<\lambda<\infty \right\}.$$

By the results above, we have

$$h'(y^*) = \max_{-\infty<\lambda<\infty} h(y^*+\lambda y_1^*) \text{ for all } y^*\epsilon E^*.$$

It is apparent from this that $M'\supseteq M$. Also, $y_1^*\epsilon M'$. But $y_1^*\notin M$, by the choice of y_1^*. Thus the dimension of the subspace M' is strictly larger than that of M.

Now if program (I') is strongly consistent, Theorem 4-E(a) implies that the extrema in (I') and (II') are equal, and that the second is attained. This must also be true then of the original programs (I) and (II), as we have observed earlier. On the other hand, (I') is itself stably consistent; hence, if (I') is not strongly consistent, we can use the same procedure to construct another pair of programs (I") and (II") equivalent to (I') and (II'), and so forth. The procedure can be carried out at most finitely many times, since F^* is finite-dimensional and a certain subspace of F^* is replaced by a strictly larger one at each iteration. Therefore we must eventually reach a point where Theorem 4-E(a) can be applied. This proves that "strong consistency" can be replaced by "stable consistency" in 4-E(a). The same thing now follows for 4-E(b), 4-F, 4-G and 4-H, since these results needed strong consistency only because their proofs depended on 4-E(a).

SECTION SIX

Combinatorial Operations

Convex functions appearing in extremum problems frequently arise from other convex functions through certain combinatorial operations. For example, as pointed out in §2, the effective domain of a given convex function f can be restricted to a convex set C by adding f and the characteristic function of C. Therefore, in calculating the dual (II) of a program (I), it is often necessary to apply the conjugate operation to a sum of convex functions. We have already done this in special cases in §5 (see 5-E, 5-L, 5-M). The general case will be studied here.

The strongest results in the duality theory that we have developed (namely 5-J and 5-K) depend on the use of stable and completely stable convex functions. The definition of stability is rather complicated, so that one cannot easily check whether a given function satisfies it. It will be proved below, however, that the class of completely stable functions is closed under addition, direct sum constructions (as in Theorem 2-L) and other operations. Consequently, any function constructed by such operations from polyhedral, quadratic and completely regular convex functions, among others (see 5-G and the remark after 5-F), will be completely stable. This resembles the situation in the calculus, where one relies on a combinatorial argument to prove that algebraic functions are continuous and differentiable where defined.

A formula for the conjugate of a sum of convex functions has already been given by Fenchel [19, p.95] (see also [27, p.222]), but this formula is slightly incorrect. (The error was pointed out in §4; it was the source of the false duality theorem.) The conjugate of a sum of convex functions is obtained from the conjugates of the individuals functions by an operation useful in itself. This operation, which we denote by #, has recently been investigated in a one-dimensional case by Bellman and Karush [1].

DEFINITION 6-A

Let $f_1,\dots,f_k$ be proper convex functions on E. Then $f_1+\dots+f_k$ and $f_1\#\dots\#f_k$ are the functions on E defined by

(6.1) $$(f_1+\dots+f_k)(x) = f_1(x)+\dots+f_k(x),$$

(6.2) $$(f_1\#\dots\#f_k)(x) = \inf\left\{ f_1(x_1)+\dots+f_k(x_k) \mid x_1+\dots+x_k = x \right\}.$$

The operation # will be called <u>(minimal)</u> <u>convolution</u>.

The term "convolution" is suggested by the fact that, when only two functions are involved, one has

(6.3) $$(f_1\#f_2)(x) = \inf_z \left\{ f_1(x-z)+f_2(z) \right\} .$$

THEOREM 6-B

Let $f_1,\dots,f_k$ be proper convex functions on E. Then $f_1+\dots+f_k$ and $f_1\#\dots\#f_k$ are also convex functions on E, although not necessarily proper, and

(6.4) $$\mathrm{dom}(f_1+\dots+f_k) = \mathrm{dom}\, f_1 \cap \dots \cap \mathrm{dom}\, f_k,$$

(6.5) $$\mathrm{dom}(f_1\#\dots\#f_k) = \mathrm{dom}\, f_1+\dots+\mathrm{dom}\, f_k = \left\{ x_1+\dots+x_k \mid x_i \in \mathrm{dom}\, f_i \right\}.$$

Addition and convolution are commutative and associative where they are defined (i.e for proper functions).

Proof: We shall omit the proof, since it consists merely in checking the definitions.

THEOREM 6-C

Let $f_1,\ldots,f_k$ be proper convex functions on E. Then

$$(f_1\#\ldots\#f_k)^* = f_1^* +\ldots+ f_k^*.$$

If $f_1,\ldots,f_k$ are also closed, and $f_1+\ldots+f_k$ is not identically $+\infty$, then

$$(f_1+\ldots+f_k)^* = \mathrm{cl}(f_1^*\#\ldots\#f_k^*).$$

Proof: We use the definitions directly

$$(f_1\#\ldots\#f_k^*)(x^*) = \sup_x\left\{[x,x^*]- \inf_{x_1+\ldots+x_k=x}\left\{f_1(x_1)+\ldots+f_k(x_k)\right\}\right\}$$

$$= \sup_x \sup_{x_1+\ldots+x_k=x}\left\{([x_1,x^*]-f_1(x_1))+\ldots+([x_k,x^*]-f_k(x_k))\right\}$$

$$= \sup_{x_1}\left\{[x_1,x^*]-f_1(x_1)\right\} +\ldots+\sup_{x_k}\left\{[x_k,x^*]-f_k(x_k)\right\}$$

$$= f_1^*(x^*)+\ldots+f_k^*(x^*).$$

Applying this fact now to $f_1^*,\ldots,f_k^*$, we have

$$(f_1^*\#\ldots\#f_k^*)^* = f_1^{**}+\ldots f_k^{**}.$$

If $f_1,\ldots,f_k$ are closed and $f_1+\ldots+f_k$ is proper, it now follows from Theorem 2-D that

$$(f_1+\ldots+f_k)^* = (f_1^{**}+\ldots+f_k^{**})^* = (f_1^*\#\ldots\#f_k^*)^{**} = \mathrm{cl}(f_1^*\#\ldots\#f_k^*).$$

COROLLARY 6-D

If $f_1,\ldots,f_k$ are closed proper convex functions on E then $f_1+\ldots+f_k$ is closed.

Proof: As pointed out above, in this case $f_1+\dots+f_k$ is the conjugate of $f_1^*\#\dots\#f_k^*$, and hence is closed by Theorem 2-D and the remarks following it. (This fact may also be proved using the definition of closure directly.)

It is not always true that $f_1\#\dots\#f_k$ is closed when $f_1,\dots f_k$ are closed. In fact $cl(f_1\#\dots\#f_k)$ may not agree with $f_1\#\dots\#f_k$ at all points of the effective domain of the latter. An example of such misbehavior is readily constructed from the example in Theorem 4-C(c). It is true, incidentally, that

$$cl(f_1\#\dots\#f_k)(x) = (f_1\#\dots\#f_k)(x) \text{ if }$$
$$x\in ri(\text{dom } f_1)+\dots+ri(\text{dom } f_k).$$

This follows from (2.7), (6.5) and a fact proved in Appendix A (namely A-D).

Before showing that Theorem 6-C can be strengthened in certain important cases, we shall illustrate some uses of + and #. When the characteristic functions of non-empty convex sets $C_1,\dots,C_k$ are combined by + and #, the sets themselves are combined as in (6.4) and (6.5). Thus

(6.7) $\quad \check{\delta}_{C_1}+\dots+\check{\delta}_{C_k} = \check{\delta}_{C_1\cap\dots\cap C_k},$

(6.8) $\quad \check{\delta}_{C_1}\#\dots\#\check{\delta}_{C_k} = \check{\delta}_{C_1+\dots+C_k}.$

Since the support function $\check{\sigma}_C$ is the conjugate of $\check{\delta}_C$ (see Appendix A), it follows from Theorem 6-C that

(6.9) $\quad \check{\sigma}_{C_1+\dots+C_k} = \check{\sigma}_{C_1}+\dots+\check{\sigma}_{C_k},$

and, if $C_1, \dots, C_k$ are closed and have a point in common,

(6.10) $$\check{\sigma}_{C_1 \cap \dots \cap C_k} = cl(\check{\sigma}_{C_1} \# \dots \# \check{\sigma}_{C_k}).$$

If f is a closed proper convex function on E and C is a non-empty closed convex set intersecting dom f, we have

(6.11) $$(f + \check{\delta}_C)^* = cl(f^* \# \check{\sigma}_C)$$

by Theorem 6-C.

If $K = \{x \mid x \leq 0\}$ and f is a proper convex function on E, then by (6.3)

(6.12) $$(f \# \check{\delta}_K)(x) = \inf\{f(z) \mid z \geq x\} \text{ for each } x \in E.$$

This is the largest convex function $h \leq f$ such that $h(x_1) \leq h(x_2)$ whenever $x_1 \leq x_2$ in E. Furthermore

$$(\check{\delta}_K)^*(x^*) = \check{\delta}(x^* \mid x^* \geq 0),$$

so that by Theorem 6-C

$$(f \# \check{\delta}_K)^*(x^*) = f^*(x^*) + \check{\delta}(x^* \mid x^* \geq 0).$$

More generally, if C is a non-empty convex set in E,

$$(f \# \check{\delta}_C)(x) = \inf\{f(x-z) \mid z \in C\} \text{ for each } x \in E.$$

Closures, too, can be expressed in terms of convolution. Suppose $\sigma(x) = \|x\|$ is some norm on E, and for each $\varepsilon \geq 0$ let

$$\delta_\varepsilon(x) = \check{\delta}(x \mid \|x\| \leq \varepsilon).$$

Then, for any proper convex function f on E,

$$(f \# \delta_\varepsilon)(x) = \inf\{f(z) \mid \|x-z\| \leq \varepsilon\}.$$

Convolution with δ_ε can be viewed as a "smoothing" or "smearing" operation. It is clear that $(f \# \delta_\varepsilon)(x)$ increases, if anything, as ε approaches 0. In fact

$$f \# \delta_0 = f, \quad \lim_{\varepsilon \to 0^+} (f \# \delta_\varepsilon)(x) = cl\, f(x).$$

Notice that, for $\sigma(x) = \|x\|$ and a non-empty convex set C,

$$(\sigma \# \check{\delta}_C)(x) = \inf \{ \|x-z\| \mid z \in C \}$$

gives the distance of x from C. If C_1 and C_2 are non-empty closed convex sets, then minimizing

$$f = \check{\delta}_{C_1} + (\sigma \# \check{\delta}_{C_2})$$

on E is the same as determining the distance between C_1 and C_2.

We now prove some special results involving stable convex functions.

THEOREM 6-E

Let f be a stable (completely stable) convex function on E. Then the following functions are also stable (completely stable):

(a) $h(x) = f(x+a)$, $a \in E$,

(b) $h(x) = f(x) + [x, a^*]$, $a^* \in E^*$,

(c) $h(x) = \lambda f(x)$, $0 < \lambda \in R$,

(d) $h(x) = f(\lambda x)$, $0 \neq \lambda \in R$,

(e) $h(x) = f(x) + \alpha$, $\alpha \in R$.

Proof: Elementary consequences of Definition 5-D and Theorem 2-K.

THEOREM 6-F

If $f_1, \ldots, f_k$ are stable convex functions on E such that $f_1 + \ldots + f_k$ is proper, then $f_1 + \ldots + f_k$ is stable. Moreover, then $f_1^* \# \ldots \# f_k^*$ is closed, and for all $x^* \in E^*$

(6.13) $$(f_1^* \# \ldots \# f_k^*)(x^*) = \min \{ f_1^*(x_1^*) + \ldots + f_k^*(x_k^*) \mid x_1^* + \ldots + x_k^* = x^* \} .$$

Proof: We shall prove the second half first. Suppose $k = 2$ and

and let $x^* \in E^*$. Let

$$f(x) = f_1(x) - [x, x^*] \text{ and } g(x) = -f_2(x).$$

Then f is stable convex (by 6-E), g is stable concave, and

$$(6.14) \qquad -\inf_x f(x)-g(x) = \sup_x [x^*, x]-(f_1+f_2)(x) = (f_1+f_2)^*(x^*).$$

Also, dom f ∩ dom g ≠ ∅ by the assumption that $f_1 + f_2$ is proper.
Hence by Theorems 5-I, 5-J and 4-E(a) (with E=F, $E^*=F^*$ and A=I),

$$(6.15) \qquad -\inf_x f(x)-g(x) = -\max_{y^*} g^*(y^*)-f^*(y^*)$$

$$= \min_{y^*} f^*(y^*)-g^*(y^*) \quad .$$

But by elementary calculation

$$f^*(y^*) = f_1^*(x^*+y^*) \text{ and } g^*(y^*) = -f_2^*(-y^*).$$

Combining this with (6.14) and (6.15) we get

$$(6.16) \qquad (f_1+f_2)^*(x^*) = \min_{y^*} f_1^*(x^*-y^*)+f_2^*(y^*) \quad .$$

The second half of the theorem now follows from Theorem 6-C for k = 2, and for general k by induction. We now prove the first half. Let $x \in \mathrm{dom}(f_1+\ldots+f_k)$ and let M be any subspace of E. Let

$$f_{k+1}(z) = \check{\delta}(z-x|M).$$

Then f_{k+1} is also a stable convex function on E (by Theorem 5-F) and, by elementary calculation with M^* as in (5.6),

$$f_{k+1}^*(z^*) = [x, z^*] + \check{\delta}(z^*|M^*).$$

Moreover

$$((f_1+\ldots+f_k)_{x,M})(z) = (f_1+\ldots+f_k+f_{k+1})(z)$$

according to definition (5.5). Now 6-C and the second half of the theorem imply that

$$((f_1+\cdots+f_k)_{x,M})^*(x^*) =$$

$$\min\{f_1^*(x_1^*)+\ldots+f_k^*(x_k^*)+f_{k+1}^*(x_{k+1}^*) \mid x_1^*+\ldots+x_k^*+x_{k+1}^*=x^*\}$$

$$= \min_{z^*}\{(f_1+\ldots+f_k)^*(x^*-z^*)+f_{k+1}^*(z^*)\}$$

$$= \min_{z^*\in M^*}\{(f_1+\ldots+f_k)^*(x^*+z^*)-[x,z^*]\}.$$

Thus $f_1+\ldots+f_k$ is stable by Definition 5-D

THEOREM 6-G

Suppose that $E = E_1 \oplus \ldots \oplus E_k$ and $E^* = E_1^* \oplus \ldots \oplus E_k^*$ as in Theorem 2-L. Let f_i be a stable (completely stable) convex function on E_i with conjugate f_i^* on E_i^*, for $i = 1,\ldots,k$. Let

$$f(x) = f(x_1,\ldots,x_k) = f_1(x_1) + \ldots + f_k(x_k).$$

Then f is a stable (completely stable) convex function on E.

Proof: Let $f_i'(x) = f_i'(x_1,\ldots,x_k) = f_i(x_i)$ for $i = 1,\ldots,k$. We shall show that each f_i' is stable on E. Fix i and $x\in$ dom f_i', and let M be a subspace of E. Let $x^*\in E^*$. Let

$$g(z) = [z,x^*] - \delta(z-x|M).$$

Then g is a stable concave function on E (by 5-F). Furthermore, let A_i be the matrix such that $z \to A_i z = z_i$ is the projection of E onto E_i. Then $f_i'(z_i) = f_i(A_i z)$, while f_i is stable on E_i by assumption. Also, $x\in$ dom g and $A_i x\in$ dom f_i by the choice of x and g. It follows now from Theorems 5-I, 5-J and 4-E(a) that

$$\text{(6.17)} \qquad \sup_z\{g(z)-f_i(A_i z)\} = \min_{z_i^*}\{f_i^*(z_i^*)-g^*(A_i^* z_i^*)\}.$$

Due to the choice of elements, the left side of (6.17) is just

$$((f_i')_{x,M})^*(x^*)$$

(see definition (5.5)). On the other hand, one calculates that

$$A_i^* z_i^* = \langle 0,\ldots,z_i^*,0\ldots\rangle \text{ for all } z_i^* \in E_i^*,$$

$$(f_i')^*(z^*) = \begin{cases} f_i^*(z_i) & \text{if } z = \langle 0,\ldots,z_i^*,0,\ldots\rangle, \\ \infty & \text{otherwise,} \end{cases}$$

$$g^*(z^*) = [x,z^*] - \check{\delta}(z^*|M^*),$$

with M^* as in (5.6). The right side of (6.17) is therefore the same as

$$\min_{z^* \in M^*} \left\{ (f_i')^*(x^*+z^*) - [x,z^*] \right\}.$$

Hence (6.17) is the desired equation (5.7), and f_i' is stable by Definition 5-D, as asserted. The rest of the proof is easy. Since obviously $f = f_1' + \ldots + f_k'$, Theorem 6-F says that f is stable. If the f_i are actually completely stable, i.e. if $f_1^*,\ldots,f_k^*$ are also stable, then the same argument shows, by Theorem 2-L, that

$$f^*(x^*) = f^*(x_1^*,\ldots,f_k^*) = f_1^*(x_1^*) + \ldots + f_k^*(x_k^*)$$

is stable. Hence f is completely stable when the f_i are completely stable.

THEOREM 6-H

Let $f_1,\ldots,f_k$ be completely stable convex functions on E. Then $f_1+\ldots+f_k$ and $f_1\#\ldots\#f_k$ are also completely stable, and

(6.18) $$(f_1\#\ldots\#f_k)(x) = \min\left\{ f_1(x_1)+\ldots+f_k(x_k) \,|\, x_1+\ldots+x_k = x \right\},$$

whenever proper.

<u>Proof</u>: Suppose $f = f_1\#\ldots\#f_k$ is proper. Then f^* is proper (by 2-D) and $f^* = f_1^*+\ldots+f_k^*$ by Theorem 6-C. By the definitions of complete

stability, each f_i^* is stable. Also, $f_i^{**} = f_i$ for $i = 1,\ldots,k$ by Theorem 2-D because the f_i, being stable, are closed and proper. Theoren 6-F therefore implies that f^* is stable, that f is closed, and that (6.18) is true. Now let $x \in \operatorname{dom} f$, $x^* \in E^*$, and let M be a subspace of E. We must verify that (5.7) holds. Let $F = E \oplus \ldots \oplus E$ and $F^* = E^* \oplus \ldots \oplus E^*$ (k times), and let A be the matrix giving the linear transformation

$$y = \langle x_1, \ldots, x_k \rangle \rightarrow Ay = x_1 + \ldots + x_k$$

from F to E. Let

$$h(y) = f_1(x_1) + \ldots + f_k(x_k), \quad g(z) = [z, x^*] - \check{\delta}(z - x \mid M).$$

Then h is a stable concave function on F by Theorem 6-G, while g is a stable concave function on E by Theorem 5-F. Also, by (6.5) and the choice of x, there does exist at least one $y \in \operatorname{dom} h$ such that $Ay \in \operatorname{dom} g$. Hence by Theorems 5-I, 5-J and 4-E(a),

$$(6.19) \qquad \inf_y \{h(y) - g(Ay)\} = \max_{z^*} \{g^*(z^*) - h^*(A^* z^*)\}.$$

The left side of (6.19) is the same as

$$\inf_z \inf_{x_1 + \ldots + x_k = z} \{f_1(x_1) + \ldots + f_k(x_k) + \check{\delta}(z - x \mid M) - [z, x^*]\}$$

$$= -\sup_z \{[z, x^*] - (f_1 \# \ldots \# f_k)(z) - \check{\delta}(z - x \mid M)\}$$

$$= -(f_{x,M})^*(x^*).$$

But also, with M^* as in (5.6),

$$h^*(y^*) = f_1^*(x_1^*) + \ldots + f_k^*(x_k^*) \quad \text{(by Theorem 2-L)},$$

$$g^*(z^*) = [x, z^* - x^*] - \check{\delta}(z^* - x^* \mid M^*),$$

$$A^* z^* = \langle z^*, \ldots, z^* \rangle \text{ for all } z^* \in E^*.$$

Hence the right side of (6.19) is

$$-\min_{z^*} \{h^*(A^* z^*)-g^*(z^*)\}$$

$$= -\min_{z^*} \{(f_1^*+\ldots+f_k^*)(z^*)-[x,z^*-x^*]+\check{\delta}(z^*-x^*|M^*)\}$$

$$= -\min_{u^* \in M^*} \{f^*(x^*+u^*)-[x,u^*]\}.$$

This finishes the proof that $f_1\#\ldots\#f_k$ is completely stable. Now suppose instead that $f = f_1+\ldots+f_k$ is proper. Then f is stable, and $f^* = f_1^*\#\ldots\#f_k^*$, by Theorems 6-C and 6-F. Also, f^* is proper by Theorem 2-D. But $f_1^*,\ldots,f_k^*$ are completely stable by definition. Hence, in particular, f^* is stable by the first part of the proof. Therefore f is completely stable.

Remark: As a special case, suppose that f is a quadratic convex function on E and that C is a non-empty polyhedral convex set in E. Then f and $\check{\delta}_C$ are completely stable (see Theorem 5-G and the remarks after Theorem 5-F). Hence $f + \check{\delta}_C$ and $f \# \check{\delta}_C$ are completely stable by the above theorem. More generally, suppose that the functions f and g in the model convex program (I) are constructed from completely stable convex functions by means of the operation in Theorems 6-E, 6-G and 6-H. Then f and g are completely stable, and Corollary 5-K is valid for (I). Thus, through 5-K and the above results, we have extended the linear programming theorems, without weakening them at all, to a much larger class of problems.

In the next theorem we describe an important case where Theorem 6-C can be sharpened and at the same time generalize the rule that "the differential of the sum is the sum of the differentials."

THEOREM 6-I

Let $f_1,\ldots,f_r$ be stable convex functions on E and let $f_{r+1},\ldots,f_k$ be closed proper convex functions on E. (Either set of functions may be empty.) Suppose that

(6.20) $\quad \text{dom } f_1 \cap \cdots \cap \text{dom } f_r \cap \text{ri}(\text{dom } f_{r+1}) \cap \cdots \cap \text{ri}(\text{dom } f_k) \neq \emptyset.$

Then $f_1^*\#\ldots\#f_k^*$ is closed and proper, and, for all $x^* \in E^*$,

(6.21) $\quad (f_1+\ldots+f_k)^*(x^*) = (f_1^*\#\ldots\#f_k^*)(x^*)$

$$= \min\left\{ f_1^*(x_1^*)+\ldots+f_k^*(x_k^*) \mid x_1^*+\ldots+x_k^*=x^* \right\}.$$

Moreover, then $x^* = \partial(f_1+\ldots+f_k)(x)$ if and only if there exist x_i^* such that $x_i^* = \partial f_i(x)$, $i = i,\ldots,k$, and $x^* = x_1^*+\ldots+x_k^*$.

Proof: Let $F = E\oplus\ldots\oplus E$, $F^* = E^*\oplus\ldots\oplus E^*$ (k times). Fix $x^* \in E^*$ and let

$$h(x) = -[x,x^*] \text{ for all } x\in E,$$

$$g(y) = -f_1(x_1)-\ldots-f_k(x_k) \text{ for all } y = \langle x_1,\ldots,x_k\rangle \in F.$$

Then h is closed proper and convex, while g is closed proper and concave. Let A be the matrix inducing the linear transformation

$$x \longrightarrow Ax = \langle x,\ldots,x\rangle$$

from E to F. Then

(6.22) $\quad -\inf_x \{h(x)-g(Ax)\} = -\inf_x \{-[x,x^*]+(f_1+\ldots+f_k)(x)\}$

$$= (f_1+\ldots+f_k)^*(x).$$

In fact, the convex program on the left of (6.22) is stably consistent. Namely, let

$$h_0(x) = -[x,x^*],\ h_1(x) = 0 \text{ for all } x\in E,$$

$$g_0(y) = -f_1(x_1)-\dots-f_r(x_r),\ g_1(y) = -f_{r+1}(x_{r+1})-\dots-f_k(x_k).$$

These functions satisfy the conditions of Definition 5-H in view of (5.20). (See 6-G and Theorem A-H in Appendix A.) Now we can apply Theorems 5-J and 4-E(a) to (6.22) obtaining

$$(6.23)\qquad (f_1+\dots+f_k)^*(x^*) = -\max_{y^*}\{g^*(y^*)-h^*(A^*y^*)\}.$$

Inasmuch as

$$h^*(z^*) = \check{\delta}(z^*|z^*=-x^*),$$

$$g^*(y^*) = -f_1^*(-x_1^*)-\dots-f_k^*(-x_k^*) \text{ (see 2-L and 2.18)},$$

$$A^*y^* = x_1^*+\dots+x_k^* \text{ for } y^* = \langle x_1^*,\dots,x_k^*\rangle \in F^*,$$

the left side of (6.23) is

$$\min_{y^*}\{h^*(A^*y^*)-g^*(y^*)\}$$

$$= \min\{f_1^*(-x_1^*)+\dots+f_k^*(-x_k^*)\,|\,-x_1^*-\dots-x_k^*=-x^*\}$$

$$= (f_1^*\#\dots\#f_k^*)(x^*).$$

This proves (6.21), and $f_1^*\#\dots\#f_k^*$ must now be closed and proper by Theorem 2-D, since it is the conjugate of $f_1+\dots+f_k$. The final statement of the theorem is deomonstrated as follows. By Theorem 2-H, $x^* = \partial(f_1+\dots+f_k)(x)$ if and only if

$$(6.24)\qquad 0 \geq (f_1+\dots+f_k(x) + (f_1+\dots+f_k)^*(x^*)$$

$$= f_1(x)+\dots+f_k(x) + \min\{f_1^*(x_1^*)+\dots+f_k^*(x_k^*)\,|\,x_1^*+\dots+x_k^*=x^*\}.$$

Since $0 \leq f_i(x) + f_i^*(x_i^*)$ for all x_i^* by Theorem 2-F, for $i = 1,\dots k$, (6.24) holds if and only if there exist $x_1^*,\dots,x_k^*$ such that

$$x^* = x_1^*+\dots+x_k^* \text{ and } 0 \geq f_i(x) + f_i^*(x_i^*) \text{ for } i = 1,\dots,k.$$

But by 2-H the latter means that $x_i^* = \partial f_i(x)$, so the proof of the Theorem is complete.

SECTION SEVEN

Some Applications of the General Theory

Some uses of the preceding theory will now be demonstrated by applying it to the various types of convex programs that have attracted attention in the literature. In each example we begin by specifying a closed proper convex function on E or E^* and a closed proper concave function on F or F^*. The conjugates of these functions are calculated from Definitions 2-C and 2-M, often with the aid of the formulas in §6, and their generalized differentials are derived from Definition 2-G and Theorem 2-H, or by the methods of Appendix C. This completely determines two dual programs (I) and (II), a "game" (III) and an "equilibrium" problem (IV), as explained in §3. We then describe the corresponding strong consistency and stable consistency conditions, and explain how the corresponding special versions of theorems 4-E, 4-F, 4-G, 4-H, 5-J or 5-K are related to known results.

It is convenient to start by reviewing the linear programming case (see §3).

EXAMPLE 7-A (Linear Programming)

Let $b^* \in E^*$ and $c \in F$. Let

$$f(x) = [x,b^*] + \check{\delta}(x|x \geq 0), \quad g(y) = \hat{\delta}(y|y \geq c),$$
$$f^*(x^*) = \check{\delta}(x^*|x^* \leq b^*), \quad g^*(y^*) = [c,y^*] + \hat{\delta}(y^*|y^* \geq 0).$$

Dual Programs:

(I) minimize $[x,b^*] + \check{\delta}(x|x \geq 0, Ax \geq c)$,

(II) maximize $[c,y^*] + \hat{\delta}(y^*|y^* \geq 0, A^*y^* \leq b^*)$.

<u>Lagrangian Function:</u>

$$L(x,y^*) = [x,b^*] + [c,y^*] - [Ax,y^*] \text{ for } x \geq 0,\ y^* \geq 0.$$

<u>Equilibrium Conditions:</u>

$$x \geq 0,\ y^* \geq 0,\ Ax-c \geq 0,\ b^*-A^*y^* \geq 0,\ [Ax-c,y^*] \leq 0,\ [x,b^*-A^*y^*] \leq 0.$$

The functions in this example are polyhedral, so that the fundamental theorem 5-K for "completely stable programs" is applicable. This yields all the facts about linear programs that were discussed in §1.

We shall make use next of the correspondence between convex sets and positively homogeneous functions, which is explained in Appendix A.

EXAMPLE 7-B (Homogeneous Programming)

Let B^* and C be non-empty closed convex sets in E^* and F^*, respectively. Let

$$f(x) = \check{\sigma}(x|B^*),\quad g(y) = \hat{\delta}(y|C),$$
$$f^*(x^*) = \check{\delta}(x^*|B^*),\quad g^*(y^*) = \hat{\sigma}(y^*|C).$$

(See Appendix A for the notation.)

<u>Dual Programs:</u>

(I) minimize $\check{\sigma}(x|B^*) + \check{\delta}(x\ |Ax \in C)$,

(II) maximize $\hat{\sigma}(y^*|C) + \hat{\delta}(y^*|A^*y^* \in B^*)$,

<u>Lagrangian Function:</u>

$$L(x,y^*) = \check{\sigma}(x|B^*) + \hat{\sigma}(y^*|C) - [Ax,y] \text{ for}$$

$$x \in \operatorname{dom} \check{\sigma}_{B^*} \text{ and } y^* \in \operatorname{dom} \hat{\sigma}_C.$$

<u>Equilibrium Conditions:</u>

$$Ax \in C \text{ and } [Ax,y] \le \hat{\sigma}(y^* | C),$$

$$A^* y^* \in B^* \text{ and } [x, A^* y^*] \ge \check{\sigma}(x | B).$$

To justify these equilibrium conditions, we note from Theorem 2-H that, for the present choice of f, $A^* y^* = \partial f(x)$ if and only if

$$\check{\sigma}(x | B^*) + \check{\delta}(A^* y_0^* | B^*) \le [x, A^* y^*].$$

This leads to the second condition; the first condition is derived similarly. In view of the definition of support functions, the equilibrium conditions have an interesting geometric meaning: the hyperplane $[y, y^*] = \mu^*$ in F is to be tangent to C at the point Ax, and the hyperplane $[x, x^*] = \mu$ in E^* is to be tanget to B^* at the point $A^* y^*$, where $\mu^* = \hat{\sigma}(y^* | C)$ and $\mu = \check{\sigma}(x | B^*)$.

Observe that 7-B specializes to 7-A when $B^* = \{x^* | x^* \le b^*\}$ and $C = \{y | y \ge c\}$. More generally, 5-K can be applied to these problems if the sets B^* and C are polyhedral. In the case where

(7.1) $\quad \operatorname{dom} \check{\sigma}_{B^*} = \{x | x \ge 0\}$ and $\operatorname{dom} \hat{\sigma}_C = \{y^* | y^* \ge 0\}$,

the dual programs (I) and (II) have been studied by Eisenberg [16]. It follows immediately from the well known properties of support functions of convex sets (see Theorem A-A) that, given any $x_0^* \in B^*$ and $y_0^* \in C$, (7.1) implies:

(7.2a) $\quad x_0^* + \lambda x^* \in B^*$ for all $\lambda > 0$ if and only if $x^* \le 0$,

(7.2b) $\quad y_0 + \lambda y \in C$ for all $\lambda > 0$ if and only if $y \ge 0$.

We can easily determine from this the duals of the strong consistency conditions (see Theorem 4-I). Namely, let $x_0 = 0$ and $y_0 \in C$. Then, by (7.2b)

$$f(x_0+\lambda x) - g(y_0+\lambda Ax) = \check{\delta}(\lambda x|B^*) + \check{\delta}(y_0+\lambda Ax|C)$$

is a finite, non-increasing function of $\lambda > 0$ if and only if $\check{\sigma}(x|B^*) \leq 0$ and $Ax \geq 0$, but it is not constant for $-\infty < \lambda < \infty$ unless $x = 0$. Therefore, according to Theorem 4-I, program (II) is strongly consistent if and only if

(7.3a) $\check{\sigma}(x|B^*) \leq 0$ and $Ax \geq 0$ imply $x = 0$.

By a similar argument, program (I) is strongly consistent if and only if

(7.3b) $\hat{\sigma}(y^*|C) \geq 0$ and $A^*y^* \leq 0$ imply $y^* = 0$.

Conditions (7.3a) and (7.3b) are the ones employed by Eisenberg. More general conditions equivalent to strong or stable consistency can also be derived when (7.1) is not assumed, and of course one can always apply the definition of strong consistency itself, determining the relative interiors of the characterizations in Theorem A-C.

EXAMPLE 7-C

Let $b^* \in E^*$ and $c \in F$, with $b^* \nleq 0$ and $c \nleq 0$. Let

$$f(x) = \check{\delta}(x|x\geq 0, [x,b^*]=1), \quad g^*(y^*) = \hat{\delta}(y^*|y^*\geq 0, [c,y^*]=1).$$

The conjugates of these functions are calculated as follows. Let $f_1(x) = \check{\delta}(x|[x,b^*]=1)$ and $f_2(x) = \check{\delta}(x|x\geq 0)$. Then f_1 and f_2 are polyhedral (and hence completely stable by 5-G) and $f = f_1 + f_2$ is proper (since $b^* \nleq 0$). Therefore by Theorem 6-I

$$f^*(x^*) = \min_{z^*} \left\{ f_1^*(x^*-z^*) + f_2^*(z^*) \right\} .$$

Elementary calculation from Definition 2-C shows that

$$f_2^*(x^*) = \check{\delta}(x^*|x^* \le 0), \quad f_1^*(x^*) = \begin{cases} \lambda^* \text{ if } x^* = \lambda^* b^*, & -\infty < \lambda^* < \infty, \\ \infty \text{ otherwise.} \end{cases}$$

Therefore

$$f^*(x^*) = \min\{\lambda^* + \check{\delta}(\lambda^* | \lambda^* b^* \ge x^*) \mid -\infty < \lambda^* < \infty\}.$$

Also, it is easy to shown from 2-G and 2-H that $x^* = \partial f_1(x)$ if and only if $[x, b^*] = 1$ and $x^* = \lambda^* b^*$ for some λ^*, $-\infty < \lambda^* < \infty$, while $x^* = \partial f_2(x)$ if and only if $x \ge 0$, $x^* \le 0$ and $[x, x^*] \ge 0$. Hence by Theorem 6-I, $x^* = \partial f(x)$ if and only if $x \ge 0$, $[x, b^*] = 1$, $x^* - \lambda^* b^* \le 0$ and $[x, x^* - \lambda^* b^*] \le 0$ for some λ^*.

Similarly one shows that

$$g(y) \; (= g^{**}(y)) = \max\{\lambda + \hat{\delta}(\lambda | \lambda c \le y) \mid -\infty < \lambda < \infty\},$$

and that $y = \partial g^*(y^*)$ if and only if $y^* \ge 0$, $[c, y^*] = 1$, $y - \lambda c > 0$ and $[y - \lambda c, y^*] \le 0$ for some λ.

Since f and g^* are polyhedral, so are f^* and g by Theorem 5-C.

Note that

$$\text{dom } f^* = \{x^* \mid x^* \le \lambda^* b^* \text{ for some } \lambda^*\},$$

$$\text{dom } g = \{y \mid y \ge \lambda c \text{ for some } \lambda\}.$$

<u>Dual</u> <u>Programs</u>:

(I) minimize $-\lambda + \check{\delta}(\lambda \mid Ax \ge \lambda c) + \check{\delta}(x | x \ge 0, [x, b^*] = 1)$,

(II) maximize $-\lambda^* + \hat{\delta}(\lambda^* | A^* y^* \le \lambda^* b^*) + \hat{\delta}(y^* | y^* \ge 0, [c, y^*] = 1)$.

<u>Lagrangian</u> <u>Function</u>:

$$L(x, y^*) = -[Ax, y^*] \text{ for } x \ge 0, \; [x, b^*] = 1, \text{ and } y^* \ge 0, \; [c, y^*] = 1.$$

<u>Equilibrium</u> <u>Conditions</u>:

$$x \ge 0, \; [x, b^*] = 1, \; \lambda^* b^* - A^* y^* \ge 0, \; [x, \lambda^* b^* - A^* y^*] \le 0,$$

$$y \ge 0, \; [c, y^*] = 1, \; Ax - \lambda c \ge 0, \quad [Ax - \lambda c, y^*] \le 0.$$

Since the functions are polyhedral, 5-K is applicable. It yields, among other things, a duality theorem proved elsewhere by the author [31, Theorem 2]. When $b^* = \langle 1,\ldots,1\rangle$ and $c = \langle 1,\ldots,1\rangle$, the dual programs are the strategy problems for a <u>matrix game</u>. In this case the implicit constraints in (I) and (II) can always be satisfied, and hence by 5-K both problems always have solutions. Also, 5-K and 4-H say then that the minimax value of the Lagrangian function always exists; this is the von Neumann minimax theorem [30]. For other special cases of 7-C, see [31].

We can generalize 7-C in the same way we generalized 7-A by 7-B.

EXAMPLE 7-D

Let B and C^* be non-empty closed convex sets in E and F^* respectively. Let

$$f(x) = \check{\delta}(x|B), \quad g(y) = \hat{\sigma}(y|C^*),$$
$$f^*(x^*) = \check{\sigma}(x^*|B), \quad g^*(y^*) = \hat{\delta}(y^*|C^*).$$

<u>Dual Problems</u>

(I) minimize $-\hat{\sigma}(Ax|C^*)+\check{\delta}(x|B)$,

(II) maximize $-\check{\sigma}(A^*y^*|B)+\hat{\delta}(y^*|C^*)$,

<u>Lagrangian Function</u>:

$$L(x,y^*) = -[Ax,y^*] \text{ for } x\in B \text{ and } y^*\in C^*.$$

<u>Equilibrium Conditions</u>

$$x\in B,\ y\in C,\ [x,A^*y^*] \geq \check{\sigma}(A^*y^*|B),\ [Ax,y^*] \leq \hat{\sigma}(Ax|C^*).$$

The equilibrium conditions are justified by the argument used in 7-B, and they have a similar geometric meaning. Some applications of problem (III) here are worth mentioning. If B and C^* are polyhedral, Theorems 4-H and 5-K yield Wolfe's minimax theorem for polyhedral games [35] when applied to (III). If B and C^* are bounded, then dom $\check{\delta}_B = E^*$ and dom $\hat{\delta}_{C^*} = F$; the strong consistency conditions are trivially satisfied in this case, and another well known minimax result (see [27, p.28]) then follows from 4-H. More generally, a minimax theorem can be stated for sets B and C^* which are not necessarily compact but satisfy conditions based on Theorem 4-I. These conditions resemble those derived in 7-B.

It frequently happens in applications that the model problems of §3 are "separable" in the sense described below. The case where all the component spaces are one-dimensional is of particular importance in network theory, as will be explained in detail in §8.

EXAMPLE 7-E (General Decomposition Principle)

Suppose, much as in Theorem 2-L, that

$$E = E_1 \oplus \dots \oplus E_r, \quad F = F_1 \oplus \dots \oplus F_s.$$
$$E^* = E_1^* \oplus \dots \oplus E_r^*, \quad F^* = F_1^* \oplus \dots \oplus F_s^*.$$

Correspondingly, suppose that A is partitioned into submatrices A_{ij} such that

$$x_j \longrightarrow A_{ij} x_j$$

is a linear transformation from E_j to F_i, so that

$$x = \langle x_1, \dots x_r \rangle \longrightarrow Ax = \left\langle \sum_j A_{1j} x_j, \dots, \sum_j A_{sj} x_j \right\rangle.$$

Denote the transpose of A_{ij} by A_{ji}^*. For $j = 1,\ldots,r$, let f_j be a closed proper convex function on E_j with conjugate f_j^* on E_j^*. For $i = 1\ldots,s$ let g_i be a closed proper concave function on F_i with conjugate g_i^* on F_i^*. Let

$$f(x) = f_1(x_1)+\ldots+f_r(x_r), \quad g(y) = g_1(y_1)+\ldots+g_s(y_s),$$
$$f^*(x)= f_1^*(x_1^*)+\ldots+f_r^*(x_r^*), \quad g^*(y^*)=g_1^*(y_1^*)+\ldots+g_s^*(y_s^*).$$

(See Theorem 2-L).

<u>Dual</u> <u>Programs</u>:

(I) minimize $\sum_j f_j(x_j) - \sum_i g_i(\sum_j A_{ij}x_j)$, for $x_j \in E_j$, $j=1,\ldots,r$,

(II) maximize $\sum_i g_i^*(y_i^*) - \sum_j f_j^*(\sum_i A_{ji}^* y_i^*)$, for $x_i \in E_i$, $i=1,\ldots,s$

<u>Lagrangian</u> <u>Function</u>:

$$L(x_1,\ldots,x_r,y_1^*,\ldots,y_s^*) = \sum_j f_j(x_j) + \sum_i g_i^*(y_i^*) - \sum_{ij} [A_{ij}x_j,y_i^*]$$

for $x_j \in \text{dom } f_j, j=1,\ldots,r$, and $y_i^* \in \text{dom } g_i^*, i=1,\ldots,s$.

<u>Equilibrium</u> <u>Conditions</u>

$$\sum_j A_{ij}x_j = \partial g_i^*(y_i^*) \text{ and } \sum_i A_{ji}^* y_i^* = \partial f(x_i), i=1,\ldots,r, \text{ and } j=1,\ldots,s.$$

The strong consistency conditions for (I) and (II) respectively, are:

(7.4a) there exist $x_j \in \text{ri}(\text{dom } f_j)$, $j=1,\ldots,r$, such that

$$\sum_j A_{ij}x_j \in \text{ri}(\text{dom } g_i),\ i=1,\ldots,s;$$

(7.4b) there exist $y_i^* \in \text{ri}(\text{dom } g_i^*)$, $i=1,\ldots,s$, such that

$$\sum_i A_{ji}^* y_i^* \in \text{ri}(\text{dom } f_j^*),\ j=1,\ldots,r.$$

This is a consequence of a general fact about relative interiors proved in Appendix A (see A-H). On the other hand, if the f_j and g_i

are all completely stable then, in view of 6-G, we can apply the powerful theorem 5-K to the above problems. The much broader conditions for stable consistency are satisfied in the following situation. Suppose, just as in Definition 5-H, that $f_j = f_{j0} + f_{j1}$ for $j = 1,\dots,r$, and $g_i = g_{i0} + g_{i1}$ for $i = 1,\dots,s$, where the f_{jo} and g_{i0} are stable. Then (I) is stably consistent if

(7.5) there exist $x_j \in \operatorname{dom} f_{j0} \cap \operatorname{ri}(\operatorname{dom} f_{j1})$, $j = 1,\dots,r$, such that $\sum_j A_{ij} x_j \in \operatorname{dom} g_{i0} \cap \operatorname{ri}(\operatorname{dom} g_{i1})$, $i = 1,\dots,s$.

To prove this, we note first that

$$h_1(x) = h_1(x_1,\dots,x_r) = f_{11}(x_1)+\dots+f_{r1}(x_r)$$

is a closed proper convex function on E by 2-L, while

$$h_0(x) = h_0(x_1,\dots,x_r) = f_{10}(x_1)+\dots+f_{r0}(x_r)$$

is a stable convex function on E by Theorem 6-G. Obviously $f = h_0 + h_1$. Similarly

$$k_1(y) = k_1(y_1,\dots,y_s) = g_{11}(y_1)+\dots+g_{s1}(y_s)$$

is a closed proper concave function on F,

$$k_0(y) = k_0(y_1,\dots,y_s) = g_{10}(y_1)+\dots+g_{s0}(y_s)$$

is a stable concave function on E, and $g = k_0 + k_1$. Hence by Definition 5-H, (I) is stably consistent is there exists some x such that

$$x \in \operatorname{dom} h_0 \cap \operatorname{ri}(\operatorname{dom} h_1) \text{ such that } Ax \in \operatorname{dom} k_0 \cap \operatorname{ri}(\operatorname{dom} k_1).$$

This is equivalent to (7.5). (A-H is used here to calculate $\operatorname{ri}(\operatorname{dom} h_1)$ and $\operatorname{ri}(\operatorname{dom} k_1)$.) Therefore (7.5) guarantees, as asserted, that (I) is stably consistent.

EXAMPLE 7-F (Convex Programming with Linear Constraints)

Let h be a closed proper convex function on E such that $\{x|x\geq 0\}\subseteq \operatorname{dom} h$, and let $c\in F$. Let

$$f(x) = h(x) + \check{\delta}(x|x\geq 0), \quad g(y) = \hat{\delta}(y|y\geq c),$$

$$f^*(x^*) = \min\{h^*(x^*)|z^*\geq x^*\}, \quad g^*(y^*) = [c,y^*] + \hat{\delta}(y^*|y^*\geq 0).$$

(The formula for f^* follows from 6-I; see also (6.12).) Then $x^* = \partial f(x)$ if and only if $x \geq 0$ and there exists some $z^* \geq x^*$ such that $z^* = \partial h(x)$ and $[x,z^*-x^*] \leq 0$. Also, $y = \partial g^*(y^*)$ if and only if $y^* \geq 0$, $y \geq c$ and $[y^*,y-c] \leq 0$. (This may be proved as an Example 7-C.)

Dual Programs:

(I) minimize $h(x) + \check{\delta}(x|x\geq 0, Ax\geq c)$

(II) maximize $[c,y^*] + \hat{\delta}(y^*|y^*\geq 0) - \min\{h^*(z^*)|z^*\geq A^*y^*\}$.

Lagrangian Function:

$$L(x,y^*) = h(x) - [Ax-c,y^*] \text{ for } x \geq 0 \text{ and } y^* \geq 0.$$

Equilibrium Conditions:

$$x \geq 0,\ Ax - c \geq 0,\ [Ax-c,y^*] \leq 0,$$

$$y^* \geq 0, z^* = \partial h(x),\ z^*-A^*y^* \geq 0,\ [x,z^*-A^*y^*] \leq 0.$$

We prove first from Theorem 4-I(a) that, given an arbitrary vector $x_0 \geq 0$, (II) is strongly consistent if and only if

(7.6) the only vector x such that $x \geq 0$, $Ax \geq 0$ and $h(x_0+\lambda x)$ is non-increasing function of $\lambda > 0$, is $x = 0$.

Namely, for each $x\in E$ consider

$$k(\lambda) = f(x_0+\lambda x) - g(c+\lambda Ax) = h(x_0+\lambda x) + \check{\delta}(\lambda|x_0+\lambda x\geq 0,\ \lambda Ax\geq 0).$$

This is a finite non-increasing function of $\lambda > 0$ if and only if $x \geq 0$, $Ax \geq 0$ and $h(x_0+\lambda x)$ is non-increasing. On the other hand, $k(\lambda)$ is constant for $-\infty < \lambda < \infty$ if and only if $x = 0$ (in fact is is not even finite for all λ unless $x = 0$). Thus (7.6) specializes the condition in 4-I(a) equivalent to the strong consistency of (II) as asserted. Next we show that (I) is stably consistent if

(7.7) there exists some $x \in ri(\mathrm{dom}\ h)$ such that $x \geq 0$ and $Ax \geq c$.

Let $f_0(x) = \delta(x|x>0)$, $f_1(x) = h(x)$, $g_0(y) = \delta(y|y\geq c)$, $g_1(y) = 0$ for all y. Then f_0 and g_0 are stable because they are polyhedral (Theorem 5-G). These functions obviously satisfy the requirements of Definition 5-H, and (7.7) is the specialization of (5.11). According to Theorem 5-J, stable consistency can be used in place of strong consistency in the duality and equivalence theorems of §4. Thus, for example, if (7.7) holds we can conclude from 4-G that (II) has a solution whenever (I) has a solution, that the pairs of solutions to (I) to satisfy the equilibrium relations, and so forth. Note especially that, when (7.6) holds, 4-E(b) implies that (I) has a solution whenever it is merely consistent, i.e whenever

(7.7') there exists some $x \geq 0$ such that $Ax \geq c$.

Problem (I) has been studied by Dorn [13] under restriction that (in the present terminology)

(7.8) $\{x|x\geq 0\} \subseteq ri(\mathrm{dom}\ h)$

and that h be <u>differentiable</u> throughout ri(dom h). In this case

∂h is single-valued on ri(dom h) (see Appendix C). The dual program given by Dorn, however, is not (II) but

(II') maximize $[c,y^*] + h(x) - [x,\partial h(x)]$

subject $y^* \geq 0$, $A^* y^* \leq \partial h(x)$, $x \in$ ri(dom h).

(This is usually not a convex or concave program.) Dorn proved (under the above restrictions) that if (I) has a solution then (II') has a solution and the extrema in (I) and (II') are equal. We shall show that (II') and (II) are closely related, and that Dorn's theorem is included in the above results. Let $x \in$ ri(dom h) and let $z^* = \partial h(x)$. Then by Theorems 2-F and 2-H

$$h^*(z^*) = [x,\partial h(x)] - h(x).$$

Hence (II') can be re-expressed as

(II") maximize $[c,y^*] - h^*(z^*)$ subject to $y^* \geq 0$, $A^* y^* \leq z^*$

and $z^* = \partial h(x)$ for some $x \in$ ri(dom h).

If the last condition were "$z^* \in$ dom h^*", then (II") would be equivalent to (II). In general, the set of z^* satisfying the last condition in (II") is a subset of dom h^* (which contains ri(dom h^*) in various cases, for instance when h is finite on all of E; see Appendix C). Thus Dorn's dual program (II') is "smaller" then (II); so that, conceivably, the extrema in (I) and (II) could be equal when the extrema in (I) and (II') are not, or (II) might provide solutions not contained in (II'). When (I) actually has a solution, such discrepancies cannot arise. Indeed, because of (7.8), the stable consistency condition (7.7) for (I) is the same as mere consistency (7.7'). Hence if (I) has a solution, 4-E(a) implies that (II) also has a solution and that the extrema in (I) and (II) are equal. Moreover then, by 4-G, pairs of solutions satisfy the above equilibrium relations; it follows from this that the solutions to (II) are also solutions to (II'') (and hence to (II')). Thus Dorn's theorem is implied by the special versions of 4-E(a) and 4-G that use condition (7.7). The latter are more general in that they do not require (7.8) or differentiability. The special versions of 4-E(b), 4-F and the dual of 4-G that use (7.6) are new.

Dennis [11, F] has treated (I) under the more restrictive assumption that

$$h(x) = f_1(x_1) + f_2(x_2) \text{ for } x = (x_1, x_2) \in E_1 \oplus E_2 = E,$$

where f_1 is linear and f_2 is strictly convex as well as differentiable. The dual problem constructed by Dennis is almost the same as (II) above, but it is based on the Legendre transformation (see Appendix C) rather than on the conjugate operation.

The dual programs in 7-F can be embedded in a symmetric pair of dual programs resembling those below. For the sake of variety, however, we shall state these symmetric programs in a particularly strong form, rather than in a particularly general form.

EXAMPLE 7-G

Suppose r = s = 2 in 7-E. Let $b_1^* \in E_1^*$, $b_2^* \in E_2^*$, $c_1 \in F_1$, $c_2 \in F_2$. Let h be a convex function on E_2 such that

(7.9a) h is finite on all of E_2, and

(7.9b) $\lim_{\lambda \to \infty} h(\lambda x_2)/\lambda = \infty$ for all $x_2 \neq 0$.

Then h is closed and proper by (7.9a) and (2.7), and dom h = E. Moreover (7.9b) is equivalent to the assumption that dom $h^* = E^*$ (see Corollary B-C in Appendix B). Therefore h^* satisfies the same conditions. In particular, h is completely stable by Theorem 5-F. Similarly let k be a concave function on F_2 such that

(7.10a) k is finite on all of F_2, and

(7.10b) $\lim_{\lambda\to\infty} k(\lambda y_2)/\lambda = -\infty$ for all $y_2 \neq 0$.

Then k is completely stable, and k^* satisfies the same conditions.

Let

$$f_1(x_1) = [x_1, b_1^*] + \check{\delta}(x_1 | x_1 \geq 0),$$
$$f_2(x_2) = [x_2, b_2^*] + \check{\delta}(x_2 | x_2 \geq 0) + h(x_2).$$

Then f_1 and f_2 are completely stable by 5-G and 6-H, and

$$f_1^*(x_1^*) = \check{\delta}(x_1^* | x_1^* \leq b_1^*),$$
$$f_2^*(x_2^*) = \min \{ h^*(z_2^*) \mid x_2^* \leq b_2^* + z_2^* \}.$$

(This may be proved from 6-I, as in earlier examples.)

Similarly we take

$$g_1(y_1) = \hat{\delta}(y_1 | y_1 \geq c_1),$$
$$g_2(y_2) = \max \{ k(z_2) | y_2 \geq c_2 + z_2 \},$$
$$g_1^*(y_1^*) = [c_1, y_1^*] + \hat{\delta}(y_1^* | y_1^* \geq 0),$$
$$g_2^*(y_2^*) = [c_2, y_2^*] + \hat{\delta}(y_2^* | y_2^* \geq 0) + k^*(y_2^*).$$

These functions, too, are completely stable.

Dual Programs:

(I) minimize $[x_1, b_1^*] + [x_2, b_2^*] + h(x_2) - k(z_2)$

subject to $x_1 \geq 0$, $x_2 \geq 0$,

$$A_{11}x_1 + A_{12}x_2 \geq c_1,$$

$$A_{21}x_1 + A_{22}x_2 \geq c_2 + z_2.$$

(II) maximize $[c_1, y_1^*] + [c_2, y^*] + k^*(y_2^*) - h^*(z_2^*)$

subject to $y_1^* \geq 0$, $y_2^* \geq 0$,

$$A_{11}^* y_1^* + A_{12}^* y_2^* \leq b_1^*$$

$$A_{21}^* y_1^* + A_{22}^* y_2^* \leq b_2^* + z_2^*$$

Lagrangian Function:

$$L(x_1, x_2; y_1^*, y_2^*) = h(x_2) + k^*(y_2^*)$$

$$+ \sum_j [x_j, b_j^*] + \sum_i [c_i, y_i^*] - \sum_{ij} [A_{ij}x_j, y_i^*]$$

for $x_1 \geq 0$, $x_2 \geq 0$, $y_1^* \geq 0$, $y_2^* \geq 0$.

Equilibrium Conditions:

the constraints of (I) and (II) along with

$$z_2^* = \partial h(x_2), \quad z_2 = \partial k^*(y_2^*),$$

$$[A_{11}x_1 + A_{12}x_2 - c_1, y_1^*] \leq 0, \quad [A_{21}x_1 + A_{22}x_2 - c_2 - z_2, y_2^*] \leq 0,$$

$$[x_1, b_1^* - A_{11}^* y_1^* - A_{12}^* y_2^*] \leq 0, \quad [x_2, z_2^* + b_2^* - A_{21}^* y_1^* - A_{22}^* y_2^*] \leq 0.$$

Since the functions are completely stable, 5-K applies to the above problems (see the discussion following 7-E). Thus, for example, if the infimum in (I) is finite, or if the supremum in (II) is finite, or if the constraints in both (I) and (II) are consistent, then both problems have solutions and the extrema are equal. Notice that the above problems reduce to linear programs when E_2 and F_2 are zero-dimensional. When merely F_2 is zero-dimensional, these problems are very similar to problems treated by Ghouila-Houri [2, p.99-102]. Ghouila-Houri's result could be applied here to prove that (I) has a solution if and only if (II) has a solution, in which case the extrema are equal and the solutions may be determined from the equilibrium relations. The present result is more general in that F_2 need not be zero-dimensional, and it is stronger because it guarantees that the extrema are attained whenever they are finite.

The quadratic case of 7-G is especially worth mentioning. This occurs when

$$h(x) = \tfrac{1}{2}[x, Bx] \text{ and } k^*(y^*) = \tfrac{1}{2}[Cy^*, y^*]$$

for a positive definite matrix B and a negative definite matric C. Then

$$h^*(x^*) = \tfrac{1}{2}[B^{-1}x^*, x^*] \text{ and } k(y) = \tfrac{1}{2}[y, C^{-1}y],$$

where B^{-1} and C^{-1} are the inverses of B and C. (See the end of Appendix C.) Cottle [38] has recently studied a symmetrically dual pair of quadratic programs of a different form, in which the duality resembles that Dorn's papers [12] and [13] (see the detailed discussion following 7-F). Cottle's results, too, could be deduced easily from the theory developed here.

We would like now to give a general interpretation of (I) and (II) in 7-G. For this purpose we assume that

(7.11) $\quad 0 = h(0) = \min h$ and $0 = k(0) = \max k$.

Actually (7.11) imposes no significant restriction. Namely, we can always change h by an additive constant to make $h(0) = 0$;

this would not essentially change the problems. On the other hand, take any a_2^* such that $a_2^* = \partial h(0)$ (this is possible by 2-I): let

$$h'(x_2) = h(x_2) - [x_2, a_2^*], \quad b'^*_2 = b_2^* + a_2^* .$$

Then $h'(0) = \min h'$ (from Definition 2-G) and h' still satisfies (7.9a) and (7.9b). Moreover the problems are not essentially changed when h and b_2^* are replaced by h' and b'^*_2. A simple argument also works for k. Thus (7.11) is just a "normalization". It follows from (7.11) that, dually,

(7.12) $\quad 0 = h^*(0) = \min h^*$ and $0 = k^*(0) = \max k^*$.

Now consider the function k_0 on F_2 defined by

$$k_0(y_2) = \max \{k(z_2) \mid y_2 \geq z_2\} .$$

Due to (7.11), k_0 has the following properties:

(7.13a) $\quad k_0(y_2) \leq 0$ for all y_2, $k_0(y_2) = 0$ if $y_2 \geq 0$,

(7.13b) $\quad k_0(z_2) \leq k_0(y_2)$ if $z_2 \leq y_2$.

We can re-express (I), using k_0, as

(I') $\quad$ minimize $[x_1, b_1^*] + [x_2, b_2^*] + h(x_2) - k_0(A_{21}x_1 + A_{22}x_2 - c_2)$

subject to $x_1 \geq 0$, $x_2 \geq 0$, $A_{11}x_1 + A_{12}x_2 \geq c_1$.

The last term in the minimand is zero by (7.13a) whenever

(7.14) $\quad A_{21}x_1 + A_{22}x_2 \geq c_2$

is satisfied. Otherwise it contributes a non-negative amount (positive if k attains its maximum only at 0) which may be thought of as a "penalty" for violating (7.14). According to (7.13b), the greater the violation, the greater the penalty. Thus (I') is much like a modifed version of 7-F in which it is desirable, but not absolutely necessary, to satisfy all the constraints exactly.

When F_2 is zero-dimensional, the penalty for violating a constraint is always infinite. A completely dual interpretation can be given for program (II), in which the penalties arise from h^*. In the quadratic case, the penalties have a "distance" meaning.

We shall now apply the general theory to the simple problem of minimizing a convex function on a convex set.

EXAMPLE 7-H

Let $E = F$, $E^* = F^*$, $A = I$. Let f be a closed proper convex function on E and let C be a non-empty closed convex set in E.

Taking

$$g(x) = \hat{\delta}(x|C),$$

we have

$$g^*(x^*) = \inf_x \{[x,x^*]-\hat{\delta}(x|C)\} = \inf \{[x,x^*]|x\epsilon C\} = \hat{\sigma}(x^*|C),$$

which is the concave support function of C (see Appendix A). Then $x = \partial g^*(x^*)$ if and only if $x\epsilon C$ and $[x,x^*] \leq \hat{\sigma}(x^*|C)$. (See 7-B).

Dual Programs:

(I) minimize $f(x) + \check{\delta}(x|C)$ on E,

(II) maximize $\hat{\sigma}(x^*|C) - f^*(x^*)$ on E^*.

Lagrangian Function:

$$L(x,x^*) = f(x) + \hat{\sigma}(x^*|C) - [x,x^*] \text{ for}$$

$$x\epsilon \text{ dom } f,\ x^* \epsilon \text{ dom } \hat{\sigma}_C.$$

Equilibrium Conditions:

$$x\epsilon C,\ x^* = \partial f(x),\ [x,x^*] \leq \hat{\sigma}(x^*/C).$$

Program (I) is strongly consistent if and only if

$ri(\text{dom } f) \cap ri\ C \neq \emptyset$.

Theorem 4-I(a) implies that given any $x_0 \in$ dom f and $x_1 \in C$, (II) will be strongly consistent if and only if the following condition is satisified:

(7.16) for each x such that $x_1 + \lambda x \in C$ for all $\lambda > 0$ and $f(x_0+\lambda x)$ is finite and non-increasing in $0 < \lambda < \infty$, it is actually true that $x_1 + \lambda x \in C$ and $f(x_0+\lambda x) = f(x_0)$ for $-\infty < \lambda < \infty$.

The results of §4 may be specialized using these conditions. In view of the fundamental nature of (I), we shall state two of these results as theorems.

THEOREM 7-I

Let f and C be as above, and suppose that condition (7.16) is satisfied for some $x_0 \in$ dom f and $x_1 \in C$. Then f attains a minimum on C (possibly $+\infty$).

Proof: This specializes part of 4-E(b).

Remark: The theorem would not generally be true if (7.16) were weakened by removing the statement that $x_1 + \lambda x \in C$ for $-\infty < \lambda < 0$. For example, let $E = R^2$ and let

$$C = \{ x = \langle \xi_1, \xi_2 \rangle \mid \xi_2 \geq \xi_1^2 \}, \quad f(x) = f(\xi_1 \xi_2) = \xi_1. \text{ Take}$$

$x_0 = x_1 = \langle 0,0 \rangle$. Then the weaker version of (7.16) is satisfied, but f is not even bounded below on C.

THEOREM 7-J

Let f and C be as above and suppose that (7.15) holds.

Then f attains a minimum on C at the point x if and only if there exists some $x^* = \partial f(x)$ such that the linear function $h(z) = [z,x^*]$ attains a minimum on C at x.

Proof: The last condition is just the condition that there exist some x^* such that x and x^* satisfy the equilibrium conditions. The theorem is therefore a consequence of 4-G. (Notice further from 4-G that the vectors x^* in question are precisely the solutions of (II).)

Programs (I) and (II) are of the same type when C is a convex cone, since the support function of a convex cone is the characteristic function of another convex cone (see Appendix A). This case is treated in the next theorem.

THEOREM 7-K

Let h be a closed proper convex function on E. Let K be a non-empty closed convex cone in E and let

$$K^* = \left\{x^* \mid [x,x^*] \geq 0\right\} \text{ for all } x \in K.$$

(Then K^* is a non-empty closed convex cone in E^*.) Let $a \in E$ and $a^* \in E^*$. Suppose that

(a) $a + x \in$ ri(dom h) for some $x \in$ ri K,

and suppose either that

(b) $a^* + x^* \in$ ri(dom h^*) for some $x^* \in$ ri K^*,

or that

(b') there exists some $x_0 \in$ dom h such that, for each $x \in K$ such that $h(x_0+\lambda x) - \lambda[x,a^*]$ is a finite non-increasing function of λ for $0 < \lambda < \infty$, it is actually true that

$-x \in K$ and $h(x_0+\lambda x) = \lambda[x,a^*]$ for $-\infty < \lambda < \infty$.

Then

$$(7.17) \qquad \infty > \min_{x \in K} \{h(a+x)-[x,a^*]\} + \min_{x^* \in K^*} \{h^*(a^*+x^*)-[a,x^*]\} = [a,a^*].$$

Moreover the minima in (7.17) occur at x and x^* if and only if

$$(7.18) \qquad x \in K,\ x^* \in K^*,\ [x,x^*] = 0,\ a^* + x^* = \partial h(a+x).$$

<u>Proof</u>: Let $f(x) = h(a+x) - [a,x^*]$ and $C = K$. Then by Theorem 2-K

$$f^*(x^*) = h^*(a^*+x^*) - [a,x^*] - [a,a^*],$$

while

$$\hat{\sigma}(x^*|C) = \inf \{[x,x^*] | x \in K\} = \delta(x^*|K^*)$$

(see (A.17)). Consider the corresponding programs (I) and (II) in 7-H. Conditions (a) and (b) are the conditions that (I) and (II) be strongly consistent, while (b') is an equivalent version of (b) derived from (7.16). But

$$(7.19) \qquad \infty > \min_x \{f(x)-\hat{\delta}_C(x)\} = \max_{x^*} \{\hat{\sigma}_C(x^*)+f^*(x^*)\} > -\infty$$

by 4-F when (I) and (II) are both strongly consistent. For the present choice of f and C, (7.19) is the same as (7.17). The equilibrium conditions in 7-H reduce likewise to (7.18), so the final assertion of the theorem is a consequence of 4-G.

COROLLARY 7-L

Let K and K^* be as in Theorem 7-K. Let h be a convex function finite on all of E such that, for some x_0,

$$(7.20) \qquad \lim_{\lambda \to \infty} h(x_0+\lambda x)/\lambda = \infty \text{ for all } x \neq 0.$$

Then the conclusions of Theorem 7-K are true for every $a \in E$ and $a^* \in E^*$.

Proof: In this case dom h = E, and h is closed by (2.7). Furthermore, as proved in Appendix B (see B-C), condition (7.20) guarantees that dom $h^* = E^*$. Therefore (a) and (b) of Theorem 7-K are satisfied for every $a \in E$ and $a^* \in E^*$.

This rather striking fact is essentially a generalization of Theorem 2-F. (The latter deals with the cases where K = 0 and K = E.) Particularly interesting applications occur when K is the non-negative orthant in E (and K^* is the non-negative orthant in E^*) or when K is a subspace of E (and K^* is the subspace of E^* orthogonal to K). A weaker version of 7-K in the subspace case, using the Legendre transformation rather than the conjugate operation (see Appendix C), has recently been proved by Duffin [14]. Observe that the quadratic convex function

$$h(x) = \tfrac{1}{2}[x, Bx],$$

where B is a positve definite matrix, satisfies the hypothesis of 7-L; the conjugate of this function is

$$h^*(x^*) = \tfrac{1}{2}[B^{-1}x, x^*]$$

(see the end of Appendix C).

Theorem 7-K was obtained by taking C to be a convex cone in 7-H. We shall now consider convex sets defined by inequalities; for such sets C, problem (I) in 7-H is the type of convex program studied by Kuhn and Tucker [28]. It is necessary, first of all to determine the support functions of such sets.

THEOREM 7-M

Let $g_1,\dots,g_r,g_{r+1},\dots,g_k$ be convex functions finite on all of E, where $g_1,\dots,g_r$ are affine, and suppose there exists at least one $x\in E$ such that

(7.21) $$g_1(x) \geq 0,\dots,g_r(x) \geq 0,\ g_{r+1}(x) > 0,\dots,g_k(x) > 0.$$

Then

$$C = \{x \mid g_i(x) \geq 0 \text{ for } i=1,\dots,k\}$$

is a non-empty closed convex set and

(7.22) $$\hat{\sigma}(x^* \mid C) = \max\left\{\sum_{i=1}^{k} \lambda_i g_i^*(x_i^*) \,\middle|\, \lambda_i \geq 0,\ \sum_{i=1}^{k} \lambda_i x_i^* = x^*\right\}.$$

(The convention $0\cdot\infty = \infty$ is to be used in this formula.)

Moreover

(7.23a) $$x\in C \text{ and } [x,x^*] \leq \hat{\sigma}(x^* \mid C)$$

holds if and only if, for some choice of vectors x_i^* and scalars λ_i,

(7.23b) $$x^* = \lambda_1 x_1^* +\dots+\lambda_k x_k^*,\quad \lambda_1 g_1(x) +\dots+ \lambda_k g_k(x) \leq 0,$$
$$x_i^* = \partial g_i(x),\ \lambda_i \geq 0,\ g_i(x) \geq 0 \text{ for } i=1,\dots,k.$$

<u>Proof</u>: A concave function is continuous on any open set where it is finite [15, p.46]; hence the g_i are all actually continuous on E (and in particular are closed). Let

$$h_i(x) = \hat{\delta}(x \mid g_i(x)\geq 0) \text{ for } i=1,\dots,k.$$

Each h_i is a closed concave function on E because the g_i are closed and concave, and is proper because (7.21) can be satisfied. Also,

(7.24) $$\hat{\delta}(x \mid C) = h_1(x) +\dots+ h_r(x) + h_{r+1}(x) +\dots+ h_k(x).$$

The functions $h_1,\dots,h_r$ are polyhedral, and hence stable, since $g_1,\dots,g_r$ are affine. Therefore if

(7.25) $$\text{dom } h_1 \cap \dots \cap \text{dom } h_r \cap \text{ri}(\text{dom } h_{r+1}) \cap \dots \cap \text{ri}(\text{dom } h_k) \neq \emptyset,$$

we can conclude that

(7.26) $$\hat{\sigma}(x^* | C) = \max\left\{ \sum_{i=1}^{k} h_i^*(z_i^*) \;\middle|\; \sum_{i=1}^{k} z_i^* = x^* \right\}$$

by the concave analog of Theorem 6-I. We must verify (7.25). Trivially,

(7.27) $$\text{dom } h_i = \{x \mid g_i(x) \geq 0\} \quad \text{for } i=1,\dots,k.$$

The fact that

(7.28) $$\text{ri}(\text{dom } h_i) = \{x \mid g_i(x) > 0\} \quad \text{for } i=r+1,\dots,k.$$

is proved as follows. Fix $i > r$, and let $g_i(x_0) > 0$, $g_i(x_1) \geq 0$; this is possible by the hypotheses. Then

$$g_i(\lambda x_0 + (1-\lambda)x_1) \geq \lambda g_i(x_0) + (1-\lambda) g_i(x_1) > 0 \text{ for } 0 < \lambda < 1.$$

This shows that

$$\text{cl}\{x \mid g_i(x) > 0\} \supseteq \{x \mid g_i(x) \geq 0\} = \text{dom } h_i.$$

Since g_i is continuous, the reverse inclusion is obvious, and the set of x for which $g_i(x) > 0$ is open. Therefore

$$\text{ri}(\text{dom } h_i) = \text{ri}(\text{cl}\{x \mid g_i(x) > 0\}) = \{x \mid g_i(x) > 0\}$$

by (2.1). Now (7.25) follows from (7.27), (7.28) and the assumption that (7.21) can be satisified. We show next that (7.26) implies the desired formula (7.22). According to (the concave analog of) a general result in Appendix B (Theorem B-I),

(7.29) $$\inf\{[z_i^*, x] \mid g_i(x) \geq 0\} = \begin{cases} \max\limits_{0<\mu<\infty} g_i^*(\mu z_i^*)/\mu & \text{if } z_i^* \neq 0, \\ 0 \text{ if } z_i^* = 0, \end{cases}$$

provided dom $g_i = E$ and sup $g_i > 0$. These conditions are satisfied here for $i=1,\ldots,k$, except if g_i happens to be a non-negative constant function; in the latter case, however, (7.29) is true trivially. Furthermore, the left side of (7.29) is the same as

$$\inf \{ [z_i^*, x] - h_i(x) \} = h_i^*(z_i^*).$$

Adopting the convention $0\cdot\infty = \infty$, we can therefore re-express (7.29) as

$$(7.30) \qquad h_i^*(x_i^*) = \max \{ \lambda g_i^*(x_i^*) \mid \lambda_i \geq 0,\ \lambda_i x_i^* = z_i^* \}.$$

Formula (7.22) is now obtained by substituting (7.30) into (7.26). Finally, due to (7.22), condition (7.23a) is satisified if and only if there exist vectors x_i^* and scalars λ_i such that

$$(7.31) \qquad x^* = \lambda_1 x_1^* + \ldots + \lambda_k x_k^* \quad \sum \lambda_i [x, x_i^*] \leq \sum \lambda_i g_i^*(x_i^*),$$

$$\text{where } \lambda_i \geq 0 \text{ and } g_i(x) \geq 0 \text{ for } i=1,\ldots,k.$$

We must show that, given x and x^*, (7.31) can be satisified if and only if (7.23b) can be satisified. If (7.23b) holds, then

$$(7.32) \qquad g_i(x) + g_i^*(x_i^*) \geq [x, x_i^*]$$

for $i=1,\ldots,k$ by the concave analog of Theorem 2-H so that

$$(7.33) \qquad [x, x_i^*] \leq g_i^*(x_i^*) \text{ if } g_i(x) = 0.$$

But the conditions in (7.23b) imply that $\lambda_i = 0$ unless $g_i(x) = 0$, while $g_i^*(x_i^*)$ is finite by (7.32) (since a proper concave function cannot take on the value $+\infty$). Therefore (7.31) holds also. To prove the converse we note first that, if (7.31) can be satisfied, it can be satisfied along with the additional requirement

$$(7.34) \qquad x_i^* = \partial g_i(x) \text{ for each } i \text{ such that } \lambda_i = 0.$$

On the other hand,

$$g_i^*(x_i^*) \leq [x, x_i^*] - g_i(x) \text{ for } i=1,\ldots,k$$

by the concave analog of 2-F, so (7.31) implies that

$$g_i(x) = 0 \text{ and } g_i(x) + g_i^*(x_i^*) = [x,x_i^*] \text{ if } \lambda_i > 0.$$

But the second equation says that $x_i^* = \partial g_i(x)$ by 2-H. This shows that (7.23b) can be satisfied if (7.31) can be satisfied, and completes the proof of the lemma.

Remark: Suppose that $r = k$ above, i.e. that the g_i are all affine functions:

$$g_i(x) = [x,a_i^*] - \alpha_i, \quad a_i^* \in E^*, \quad \alpha_i \in R.$$

Then we have

$$g_i^*(x_i^*) = \alpha_i + \delta(x_i^* \mid x_i^* = a_i^*), \quad i=1,\ldots,k.$$

Formula (7.22) reduces in this event to

$$\inf \{ [x,x^*] \mid [x,a_i^*] \geq \alpha_i, i=1,\ldots,k \}$$
$$\max \{ \sum \lambda_i \alpha_i \mid \lambda_i \geq 0, \sum \lambda_i a_i^* = x^* \}$$

for each x^* such that either side is finite. This is a well known alternate form of the linear programming duality theorem. ("inf" can be replaced by "min" because C is polyhedral.) Conversely, we may view the linear programming result essentially as a general formula for the support function of a polyhedral convex set.

We shall now specialize 7-H to convex sets of the above type.

EXAMPLE 7-N

Assume that the hypothesis of Theorem 7-M is satisfied, and let C be the convex set defined there. For simplicity, let f be convex function finite on all of E. Then, by 7-M, problems 7-H become the following.

Dual Programs:

(I) minimize $f(x) + \check{\delta}(x|g_i(x)\geq 0, i=1,\ldots,k)$,

(II) maximize $\max\left\{\sum_i \lambda_i g_i^*(x_i^*) \,|\, \lambda_i \geq 0, \sum \lambda_i x_i^* = x^*\right\} - f^*(x^*)$.

(The convention $0\cdot\infty = \infty$ is to be used in (II).)

Lagrangian Function:

$$L(x,x^*) = f(x) + \max\left\{\sum_i \lambda_i g_i^*(x_i^*) \,|\, \lambda_i \geq 0, \sum_i \lambda_i x_i^* = x^*\right\} - [x,x^*]$$

for $x\in E$, $x^*\in E^*$. (See the remarks below.)

Equilibrium Conditions:

$$x^* = \partial f(x), \text{ and } x_i^* = \partial g_i(x),\ \lambda_i \geq 0,\ g_i(x) \geq 0, \text{for } i=1,\ldots,k,$$

$$x^* = \sum_i \lambda_i x_i^*, \quad \sum_i \lambda_i g_i(x) \leq 0.$$

Programs (I) and (II) can also be expressed as

(I') minimize $f(x)$ subject to $g_1(x) \geq 0,\ldots,g_k(x) \geq 0$,

(II') maximize $\sum_i \lambda_i g_i^*(x_i^*) - f^*(\sum_i \lambda_i x_i^*)$

subject to $\lambda_i \geq 0$, $x_i^* \in E^*$, $i=1,\ldots,k$.

which is an extremum problem for

(7.35) $X^* = \langle \lambda_1,\ldots,\lambda_k, x_1^*,\ldots,x_k^*\rangle$.

(Program (II') also contains the implicit constraints

(7.36) $x_i^* \in \operatorname{dom} g_i^*, i=1,\ldots,k, \sum_i \lambda_i x_i^* \in \operatorname{dom} f^*$.)

Since f is finite on all of E, it is not really necessary to restrict x^* in the Lagrangian function (inasmuch as $\infty - \infty$ cannot arise). In any case, problem (III) for this L is the same as

(III') find a saddle-point $\langle x_0, X_0^*\rangle$ of

$$L'(x,X^*) = f(x) - \sum_i \lambda_i([x,x_i^*] - g_i^*(x_i^*))$$

for $x\in E$ and X^* (as in (7.35)) satisfying $\lambda_i \geq 0$ and

$x_i^* \in \operatorname{dom} g_i^*$, $i=1,\ldots,k$.

We shall compare this with the Kuhn-Tucker saddle-point problem in a moment.

Program (I) is strongly consistent because of the assumption that f is finite on all of E and that (7.21) can be satisified. Given any x_0 satisfying the constraints of (I), program (II) is strongly consistent if and only if

(7.37) for each x such that $f(x_0+\lambda x)$ is a non-increasing function $\lambda > 0$ and $g_i(x_0+\lambda x) \geq 0$ for all $\lambda > 0$ and $i=1,\ldots,k$, it is actually true that $f(x_0+\lambda x) = f(x_0)$ $g_i(x_0+\lambda x) \geq 0$, $i=1,\ldots,k$, for $-\infty < \lambda < \infty$.

This specializes (7.16). Incidentally, the strong consistency of (II) is also equivalent to the condition that

(7.38) there exist $x_i \in \mathrm{ri}(\mathrm{dom}\, g_i^*)$ and $\lambda_i > 0$, $i=1,\ldots k$, such that $\lambda_1 x_1+\ldots+\lambda_k x_k \in \mathrm{ri}(\mathrm{dom}\, f^*)$.

The proof of this fact, which we shall not use, is an extension of the argument in Theorem A-D. The theorems of §4 now yield the following results.

THEOREM 7-O

(a) The maximum in (II') (possibly $-\infty$) is always attained, and it always equals the infimum in (I').

(b) If (7.37) holds, the infimum in (I') is finite and attained.

Proof: This is implied by 4-E, since (I) is strongly consistent by the assumptions in 7-N.

THEOREM 7-P

The following condtions in x_0 are equivalent to one another (with X_0^* as in (7.35)):

(a) x_0 is a solution to (I'),

(b) there exists some X_0^* such that $\langle x_0, X_0^* \rangle$ is a saddle-point in (III'),

(c) there exists some X_0^* such that $\langle x_0, X_0^* \rangle$ satisfies the equilibrium conditions in 7-N (with $x^* = \lambda_1 x_1^* + \ldots + \lambda_k x_k^*$).

More the vectors X_0^* satisfying (b) and (c) then precisely the soltutions of (II').

Proof: This follows from 4-G, since (I) is strongly consistent in 7-N and (III') is equivalent to problem (III) for 7-N.

The dual program (II') which we have given for (I') is finite-dimensional and does not require differentiability; moreover it is "independent" of (I), in that it does not involve the vector x in (I) as one of the unknowns. In contrast, the dual programs constructed by Hanson [25] and Wolfe [36], while finite-dimensional, require differentiability and are not "independent" of (I') (See below.) The dual program which Charnes, Cooper and Kortanek [8] consider, while "independent" of (I) and free of differentiability assumptions, is an infinite-dimensional linear program.

Assume temporarily that the functions in (I) are actually differentiable. Hanson's dual program is then

(II_1) maximize $\sum_i \lambda_i([x,\partial g_i(x)]-g_i(x)) - ([x,\partial f(x)]-f(x))$

subject to $\lambda_1 \geq 0,\ldots,\lambda_k \geq 0$, $x\in E$, $\sum_i \lambda_i \partial g_i(x) = \partial f(x)$,

and Wolfe's dual program is

(II_2) maximize $f(x) - \sum \lambda_i g_i(x)$

subject to $\lambda_1 \geq 0,\ldots,\lambda_k \geq 0$, $x\in E$, $\sum_i \lambda_i \partial g_i(x) = \partial f(x)$.

(We have made some inessential changes in the form of these problems, in order to make then conform to the present context.) Let

$$x^* = \partial f(x), \quad x_i^* = \partial g_i(x), \; i=1,\ldots,k.$$

Then

$$(7.39) \qquad [x,\partial f(x)] - f(x) = f^*(x^*), \; [x,\partial g_i(x)] - g_i(x) = g_i^*(x^*), \; i=1,\ldots,k,$$

by Theorems 2-F and 2-H (and their concave analogs). Hence (II_1) can be expressed as

(II_3) maximize $\sum_i \lambda_i g_i^*(x_i^*) - f^*(\sum_i \lambda_i x_i^*)$

subject to $\lambda_i \geq 0$ and $x_i^* \in E^*$ for $i=1,\ldots,k$, and the condition that

$\sum_i \lambda_i x_i^* = \partial f(x)$, $x_i^* = \partial g_i(x), i=1,\ldots,k$, for some $x\in E$.

Similarly, the maximand in (II_2) is

$$[x,\partial f(x)] - f^*(x^*) - \sum \lambda_i([x,\partial g_i(x)] - g_i^*(x_i^*))$$

$$= \sum \lambda_i g_i(x_i^*) - f^*(x^*) + [x,\partial f(x)-\sum_i \lambda_i \partial g_i(x)]$$

by (7.39), so (II_2) is also equivalent to (II_3). It is clear that (II_3) is more restrictive than (II') in general. However, under the assumption that (I') has a solution, (II_3) and (II') are equivalent (even in the non-differentiable case). This follows from Theorem 7-P and the nature of the equilibrium conditions in 7-N.

Kuhn and Tucker [28] associate the following problem with (I'):

(III") find a saddle-point for

$$L''(x;\lambda_1,\dots,\lambda_k) = f(x) - \sum_i \lambda_i g_i(x)$$

subject to $x \epsilon E$, and $\lambda_1 \geq 0,\dots,\lambda_k \geq 0$.

This problem is simpler and more convenient than (III'). Actually, the two problems are equivalent. Saddle-points in (III') are the solutions to the equilibrium conditions in 7-N, according to Theorem 7-P. But (III"), too, is equivalent to solving these equilibrium conditions. In the differentiable case, this was proved by Kuhn and Tucker in [28]; the arguments in the general case is similar. Thus the Kuhn-Tucker saddle-point theorem (according to which, under the assumptions in 7-N, x_0 is a solution to (I') if and only if $x_0;\lambda_1,\dots,\lambda_k$ is a saddle-point in (III") for some choice of $\lambda_1 \geq 0$, $\lambda_1 \geq 0$) can be deduced from 7-P. The original version of this theorem in [28] requires differentiability, however an extension identical to the present one is given by Ghouila-Houri in [2, p.81]. Other versions not requiring differentiability may be found in [17] and [27, p.201].

SECTION EIGHT

Completely Separable Problems and Monotone Relations

Problems (I), (II), (III) and (IV) will be called completely separable if they can be expressed in the manner of Example 7-E with all the component spaces one-dimensional. The linear programming problems, for instance, are completely separable; so are certain problems of importance in network theory. In analyzing such problems here, we shall find that the equilibrium conditions play an especially significant role. It turns out that the generalized differentials of one-dimensional closed proper convex functions can be described axiomatically as "maximal increasing relations," so that (IV), which up to this point has been a derived problem, arises of its own accord in the completely separable case as the problem of solving a system of such "relations".

Completely separable problems, by their definition, have the following form. Let f_j be a closed proper convex function on R, with conjugate f_j^* on R, for j=1,...,n. Let g_i be a closed proper concave function on R, with conjugate g_i^* on R, for i=1,...,m. Let $A = ((\alpha_{ij}))$ be an m x n matrix with transpose $A^* = ((\alpha_{ji}^*))$. Then, as in 7-E, we have:

(I_0) Convex program:

$$\text{minimize } \sum_j f_j(\xi_j) - \sum_i g_i\left(\sum_j \alpha_{ij}\xi_j\right) \text{ for } \xi_j \in R,\ j=1,\ldots,n.$$

(II_0) Concave program:

$$\text{maximize } \sum_i g_i^*(\eta_i^*) - \sum_j f_j^*\left(\sum_j \alpha_{ji}^*\eta_i^*\right) \text{ for } \eta_i^* \in R,\ i=1,\ldots,m.$$

(III_0) Game problem:

Find a saddle-point for

$$L(\xi_1,\ldots,\xi_n;\eta_1^*,\ldots,\eta_m^*) = \sum_j f_j(\xi_j) + \sum_i g_i^*(\eta_i^*) - \sum_{ij} \eta_i^* \alpha_{ij} \xi_j .$$

subject to $\xi_j \in \text{dom } f_j$, $j=1,\ldots,n$ and $\eta_i^* \in \text{dom } g_i$, $i=1,\ldots,m$.

(IV_0) Equilibrium problem:

find $\xi_1,\ldots,\xi_n$ and $\eta_1^*,\ldots,\eta_m^*$ satisfying

$$\sum_j \alpha_{ij} \xi_j = \partial g_i^*(\eta_i^*) \text{ for } i=1,\ldots,m \text{ and}$$

$$\sum_i \alpha_{ji}^* \eta_i^* = \partial f_j(\xi_j) \text{ for } j=1,\ldots,n.$$

These become the linear programming problems in 7-A if, given real numbers β_j^*, $j=1,\ldots,n$, and γ_i, $i=1,\ldots,m$, we choose

(8.1) $$f_j(\lambda) = \lambda\beta_j^* + \check{\delta}(\lambda|\lambda\geq 0), \quad g_i(\lambda) = \hat{\delta}(\lambda|\lambda\geq\gamma_i),$$
$$f_j^*(\lambda^*) = \check{\delta}(\lambda^*|\lambda^*\leq\beta_j^*), \quad g_i^*(\lambda^*) = \gamma_i\lambda^* + \hat{\delta}(\lambda^*|\lambda^*\geq 0).$$

The completely separable programs (I_0) and (II_0) are consistent when their implicit constraints can be satisfied, i.e. when

(8.2a) there exist $\xi_j \in \text{dom } f_j$, for $j=1,\ldots,n$ such that $\sum_j \alpha_{ij}\xi_j \in \text{dom } g_i$ for $i=1,\ldots,m$,

(8.2b) there exist $\eta_i^* \in \text{dom } g_i^*$ for $i=1,\ldots,m$ such that $\sum_i \alpha_{ji}^*\eta_i^* \in \text{dom } f_j^*$ for $j=1,\ldots,n$,

respectively. As we pointed out more generally in the remarks following 7-E, (I_0) and (II_0) are strongly consistent, respectively, if and only if

(8.3a) there exist $\xi_j \in \text{ri}(\text{dom } f_j)$ for $j=1,\ldots,n$ such that $\sum_j \alpha_{ij}\xi_j \in \text{ri}(\text{dom } g_i)$ for $i=1,\ldots,m$, (8.3b) there exist $\eta_i^* \in \text{ri}(\text{dom } g_i^*)$ for $i=1,\ldots,m$,

such that $\sum_i \alpha_{ji}^* \eta_i^* \in \text{ri}(\text{dom } f_j^*)$ for $j=1,\ldots,n$.

We now consider an intermediate pair of conditions. Suppose that h is a closed proper convex function on R. Define

$$\text{dom } \partial h = \left\{ \lambda \mid \lambda^* = \partial h(\lambda) \text{ for some } \lambda^* \right\},$$

$$\text{range } \partial h = \left\{ \lambda^* \mid \lambda^* = \partial h(\lambda) \text{ for some } \lambda \right\},$$

and similarly for h^*. Then by 2-H and 2-I we have

(8.4a) $\text{ri}(\text{dom } h) \subset \text{dom } \partial h = \text{range } \partial h^* \subset \text{dom } h$,

(8.4b) $\text{ri}(\text{dom } h^*) \subset \text{range } \partial h = \text{dom } \partial h^* \subset \text{dom } h^*$.

The analogous relations hold, of course, for closed proper concave functions. Problem (IV_0) involves implicit constraints which we can express, using this notation, as

(8.5a) there exist $\xi_j \in \text{dom } \partial f_j$ for $j=1,\ldots,n$

such that $\sum_j \alpha_{ij} \xi_j \in \text{dom } \partial g_i$ for $i=1,\ldots,m$,

(8.5b) there exist $\eta_i^* \in \text{dom } \partial g_i^*$ for $i=1,\ldots,m$,

such that $\sum_i \alpha_{ji}^* \eta_i^* \in \text{dom } \partial f_j^*$ for $j=1,\ldots,n$.

In view of (8.4), these are weaker than condtions (8.3) but stronger than conditions (8.2). We shall prove below that (8.5a) and (8.5b) are precisely the conditions that (I_0) and (II_0) respectively, be <u>stably consistent</u>. It is necessary first to examine the properties of differentials of convex functions in the one-dimensional case.

THEOREM 8-A

Let f be a closed proper convex function on R. Then the <u>right</u> and <u>left derivatives</u> of f can be defined by

$$(8.6)\qquad \left.\begin{aligned} f'_+(\xi) &= \lim_{\lambda\downarrow\xi} (f(\lambda)-f(\xi))/(\lambda-\xi) \\ f'_-(\xi) &= \lim_{\lambda\uparrow\xi} (f(\lambda)-f(\xi))/(\lambda-\xi) \end{aligned}\right\} \text{if } \xi\in \text{dom } f,$$

$$f'_+(\xi) = f'_-(\xi) = \begin{cases} \infty \text{ if } \xi > \lambda \text{ for all } \lambda\in \text{dom } f, \\ -\infty \text{ if } \xi < \lambda \text{ for all } \lambda\in \text{dom } f, \end{cases}$$

and they satisfy

$$(8.7)\qquad f'_-(\xi_1) \le f'_+(\xi_1) \le f'_-(\xi_2) \le f'_+(\xi_2) \text{ whenever } \xi_1 \le \xi_2.$$

Furthermore, for all $\xi\in R$,

$$(8.7a)\qquad \sup_{\lambda<\xi} f'_+(\lambda) = f'_-(\xi),\ \inf_{\lambda>\xi} f'_+(\lambda) = f'_+(\xi),$$

$$(8.7b)\qquad \sup_{\lambda<\xi} f'_-(\lambda) = f'_-(\xi),\ \inf_{\lambda>\xi} f'_-(\lambda) = f'_+(\xi).$$

Finally,

$$(8.8)\qquad \xi^* = \partial f(\xi) \text{ if and only if } \xi^*\in R \text{ and } f'_-(\xi) \le \xi^* \le f'_+(\xi).$$

Proof: Since f is closed and dom f is an interval in the present case, f is actually continuous on cl(dom f). This follows from (2.8). Secondly, it is well known that the difference quotient $(f(\lambda)-f(\xi))/(\lambda-\xi)$ is a non-decreasing function of each of its arguments as long as λ and ξ are in the interval dom f and $\lambda \ne \xi$ (see [4, p.19] or [15, p.47]); this is trivially true also if either $\lambda\notin$ dom f or $\xi\notin$ dom f (but not both). All the assertions of the theorem, except (8.8), follow directly from these two facts. To prove (8.8) we observe that, according to Definition 2-G, $\xi^* = \partial f(\xi)$ if and only if $f(\lambda) \ge f(\xi) + (\lambda-\xi)\xi^*$ for all $\lambda\in R$. But this happens if and only if $\xi\in$ dom f and

$$\sup_{\lambda<\xi} (f(\lambda)-f(\xi))/(\lambda-\xi) \le \xi^* \le \inf_{\lambda>\xi} (f(\lambda)-f(\xi))/(\lambda-\xi).$$

By the remarks just made, the left and right sides of this inequality are $f'_-(\xi)$ and $f'_+(\xi)$, respectively, when $\xi \in \text{dom } f$. On the other hand, when $\xi \notin \text{dom } f$ neither side of (8.8) can be satisfied. This proves the theorem.

Remark: The conjugate of a one-dimensional closed proper convex function can be calculated using (8.8) and Theorem C-H in Appendix C.

THEOREM 8-B

Let f be a closed proper convex function on R. Then

(a) f is stable if and only if $\text{dom } \partial f = \text{dom } f$. Moreover, this is true except when either dom f has a lower end-point $\alpha > -\infty$ at which $f(\alpha) < \infty$ but $f'_+(x) = -\infty$, or dom f has an upper-end point $\beta < \infty$ at which $f(\beta) < \infty$ but $f'_-(\beta) = \infty$.

(b) f is completely stable if and only if, in addition to being stable, $\text{range } \partial f = \text{dom } f^*$. Moreover, the latter is true except when the graph of f has a proper non-vertical asymptotic line, i.e. except when, for some $\alpha \in R$, $f(\lambda) - \alpha\lambda$ approaches a finite minimum without attaining it as $\lambda \to \infty$ or as $\lambda \to -\infty$.

Proof: (a) Since E = R is one-dimensional here, its only subspaces are $M = \{0\}$ and $M = R$. It follows from Definition 5-D and the remark after 5-E that f is stable on R if and only if

$$(8.9) \qquad \sup_{\xi} \{\xi\xi^* - (f(\xi) + \delta(\xi-\xi_0 \mid 0))\} = \min_{\lambda^*} \{f^*(\xi^* + \lambda^*) - \xi_0\lambda^*\}$$

for each $\xi_0 \in \text{dom } f$ and $\xi^* \in R$. The left side of (8.8) is trivially just $\xi_0\xi^* - f(\xi_0)$, so (8.9) holds for all ξ^* if and only if there exists some λ_0^* such that

$$\xi_0\xi^* - f(\xi_0) = f^*(\xi^*+\lambda_0^*) - \xi_0\lambda_0^*,$$

i.e. if and only if there exists some ξ_0^* such that

$$(8,10) \qquad f(\xi_0) + f^*(\xi_0^*) = \xi_0\xi_0^*.$$

But (8.10) holds if and only if $\xi_0^* = \partial f(\xi_0)$, by Theorems 2-F and 2-H. Therefore f is stable if and only if dom f $\subseteq$ dom ∂f. The reverse inclusion has already been pointed out in (8.4a), so this proves the first part of (a). Now, since dom f is an interval, (8.4a) implies that dom ∂f = dom f unless the latter contains an end-point not contained in the former. But if α is a lower end-point of dom f, then $f'_-(\alpha) = -\infty$. Hence $\alpha \in$ dom ∂f by (8.8) except when also $f'_+(\alpha) = -\infty$. A similar observation for upper end-points completes the proof of (a).

(b) If f is stable, then, by definition, f is completely stable if and only if f^* is stable. The first part of (b) therefore follows from (a) and (8.4b). Also, by the dual of the argument in (a), f^* is stable if and only if

$$(8.9') \qquad \xi\xi_0^* - f^*(\xi_0^*) = \min_\lambda \; f(\xi+\lambda)-\lambda\xi_0^* \quad ,$$

for all $\xi \in R$ and $\xi_0^* \in$ dom f^*, i.e. if and only if

$$(8.10) \qquad -f^*(\xi_0^*) = \min_\lambda \{f(\lambda)-\lambda\xi_0^*\}$$

for all $\xi_0^* \in$ dom f^*. But, by definition,

$$-f^*(\xi_0^*) = -\sup_\lambda \{\lambda\xi_0^*-f(\lambda)\} = \inf_\lambda \{f(\lambda)-\lambda\xi_0^*\} .$$

Hence f^* will be stable if and only if $f(\lambda) - \lambda\xi_0^*$ attains its infimum in λ for each ξ_0^* such that the infimum is finite. Since f is lower semi-continuous (because it is closed by assumption), the infimum in question is always attained unless it is approached asymptotically as $\lambda \to \infty$ or as $\lambda \to -\infty$.

COROLLARY 8-C

If f is a closed proper convex function on R such that dom ∂f is closed and range ∂f is closed then f is completely stable.

Proof: This follows immediately from the theorem and from (8.4a) and (8.4b).

COROLLARY 8-D

Suppose that $f_1, \ldots, f_r$ are closed proper convex functions on R which are non-asymptotic, and are not infinitely steep at end-points included in their effective domains, as described in Theorem 8-B. Suppose that $g_1, \ldots, g_m$ are closed proper concave functions having the analogous properties. Then the fundamental theorem 5-K for completely stable programs can be applied to the completely separable programs (I_0) and (II_0).

Proof: In this case the functions in (I_0) are all completely stable. But then, according to the remarks following 7-E, (I_0) is a completely stable program and 5-K is applicable.

THEOREM 8-E

Program (I_0) is stably consistent if and only if (8.5a) holds. Program (II_0) is stably consistent if and only if (8.5b) holds.

<u>Proof</u>: We shall prove first that (8.5a) guarantees the stable consistency of (I_0). Due to the remarks after 7-E, it will be enough to prove that: if h is any closed proper convex function on R, then there exists a stable convex function h_0 on R and a closed proper convex function h_1 on R such that

(8.11) $\quad h = h_0 + h_1$ and $\text{dom}\,\partial h \subseteq \text{dom}\,h_0 \cap \text{ri}(\text{dom}\,h_1)$.

We prove this by exhausting several possible cases. If $\text{dom}\,\partial h = \text{dom}\,h$, then h is stable by Theorem 8-B(a). Then we can let $h_0 = h$ and let h_1 be identically zero. If $\text{dom}\,\partial h = \text{ri}(\text{dom}\,h)$, we can let $h_1 = h$ and let h_0 be identically zero. Because of (8.4a), we are now left only with the following possibility:

(8.12) $\quad \text{dom}\,h = \{\lambda \mid \alpha \leq \lambda \leq \beta\}$

where $\alpha \in R$, $\beta \in R$, $\alpha < \beta$, but

(8.13) $\quad \text{dom}\,\partial h = \{\lambda \mid \alpha \leq \lambda < \beta\}$ or $\text{dom}\,\partial h = \{\lambda \mid \alpha < \lambda \leq \beta\}$.

We shall only consider the first case in (8.13); the argument for the other case is similar. Since $\alpha \in \text{dom}\,\partial h$ there exists some $\alpha^* \in R$ with $\alpha^* = \partial h(\alpha)$. Let

$$h_0(\lambda) = \check{\delta}(\lambda \mid \lambda \geq \alpha),$$

$$h_1(\lambda) = \begin{cases} h(\lambda) \text{ if } \lambda \geq \alpha, \\ h(\alpha) + (\lambda - \alpha)\alpha^* \text{ if } \lambda \leq \alpha. \end{cases}$$

Since by definition of α^*

$$h(\lambda) \geq h(\alpha) + (\lambda - \alpha)\alpha^* \text{ for all } \lambda \geq \alpha,$$

while $h(\lambda) = \infty$ for $\lambda < \alpha$, it is readily seen that h_1 is a proper convex function; moreover

$$\text{dom}\,h_1 = \{\lambda \mid -\infty < \lambda < \beta\},$$

so h_1 must be closed by (2.7) because h is closed. Inasmuch as dom f and dom ∂f are given by (8.12) and the first half of (8.13), respectively, h_0 and h_1 satisfy (8.11). This finishes the proof of the sufficiency of condition (8.5a). In proving its necessity we shall actually prove a more general fact. Suppose that the general program (I) satisfies the definition 5-H of stable consistency, and assume the notation given there. We shall show that then

(8.14) there exists some $x \in \text{dom } \partial f$
such that $Ax \in \text{dom } \partial g$

Of course, (8.14) reduces to (8.5a) when (I) is completely separable. From definition 5-H, we have $f = f_0 + f_1$, where f_0 is stable, f_1 is closed and proper, and

$$\text{dom } f_0 \cap \text{ri}(\text{dom } f_1) \neq \emptyset.$$

Therefore, by Theorem 6-I, $x^* = \partial f(x)$ if and only if $x^* = x_0^* + x_1^*$ for certain vectors x_0^* and x_1^* such that $x_0^* = \partial f_0(x)$ and $x_1^* = \partial f_1(x)$. This implies in particular that

$$\text{dom } \partial f = \text{dom } \partial f_0 \cap \text{dom } \partial f_1 .$$

But $\text{ri}(\text{dom } f_1) \subseteq \text{dom } \partial f_1$ by 2-I, while $\text{dom } \partial f_0 = \text{dom } f_0$ because f_0 is stable. The latter follows, namely, from specializing the definition 5-D of the stability of f_0 to the subspace $M = \{0\}$, much as in the proof of Theorem 8-B(a). Hence

$$\text{dom } f_0 \cap \text{ri}(\text{dom } f_1) \subseteq \text{dom } \partial f.$$

A dual argument demonstrates that

$$\text{dom } g_0 \cap \text{ri}(\text{dom } g_1) \subseteq \text{dom } \partial g.$$

Therefore condition (5.11) implies (8.14), and the first half of of the theorem is proved. The second half follows now by duality.

COROLLARY 8-F

Theorems 4-E, 4-F, 4-G and 4-H may be applied to the completely separable problems (I_0), (II_0), (III_0) and (IV_0) using (8.5a) in place of the condition that (IV_0) be strongly consistent and (8.5b) in place of the condition that (II_0) be strongly consistent.

Proof: This combines the present theorem with 5-J.

Remark: Theorem 8-E cannot, in general, be extended to programs which are not completely separable; in other words, (8.14) does not usually guarantee the stable consistency of program (I). This may be seen from the example in Theorem 4-C(c), where $x = \langle 0,0 \rangle$ satisfies (8.14) but the extrema in (I) and (II) are not equal. The latter would be impossible, by 5-J and 4-E(a), if (I) were stably consistent.

The differentials of one-dimensional closed proper convex functions were described in Theorem 8-A in terms of one-sided derivatives. We shall now characterize such differentials abstractly, using the following concept.

DEFINITION 8-G

Let r be a non-empty set of ordered pairs $\langle \xi, \xi^* \rangle$ where $\xi \in R$ and $\xi^* \in R$. It will be convenient to introduce the notation that

$$\xi^* = r(\xi) \text{ means } \langle \xi, \xi^* \rangle \in r.$$

We shall say that r is an <u>increasing relation</u> if

(8.15) $\xi_1^* \leq \xi_2^*$ whenever $\xi_1^* = r(\xi_1)$, $\xi_2^* = r(\xi_2)$ and $\xi_1 < \xi_2$.

An increasing relation is <u>maximal</u> if it cannot be embedded in any properly larger increasing relation, i.e. if the set of ordered pairs is maximal with respect to property (8.15).

The <u>domain</u> and <u>range</u> of an increasing relation r are defined by

$$\text{dom } r = \{\xi \mid \xi^* = r(\xi) \text{ for some } \xi^*\},$$

$$\text{range } r = \{\xi^* \mid \xi^* = r(\xi) \text{ for some } \xi\}.$$

(<u>Decreasing relations</u> are defined by reversing the inequality in (8.15).)

It is easy to prove, by means of Zorn's Lemma, that every increasing relation is contained in a maximal one. This concept has been important in recent developments in network theory; indeed, our "maximal increasing relations" are "resistors" in the terminology of Minty [51]. We shall see below (Theorem 8-I) that such relations correspond precisely to continuous "increasing" curves which are unbounded (i.e. not limited to a bounded region of the plane). Moreover, they are precisely the differentials ∂f of closed proper convex functions f on the real line (8-J). In this sense, maximal increasing relations are almost functional relations (which is why we have chosen to use functional notation to describe them). For example, the graph of a continuous non-decreasing function on the real line is a maximal increasing relation (see 8-H). Given the graph of an increasing "step function" on the real line, one also obtains a maximal increasing relation by supplying the vertical segments which connect the "steps". If there are finitely many "steps," this is the differential of a polyhedral convex function. It is a "step resistor" in Minty's terminology. "Resistor" is perhaps too narrow a designation for such relations, even in network theory, since other elements such as sources of current or potential, or diodes, can also be characterized as certain types of maximal increasing relations (see Millar [50], Dennis [11, p.3-6] and Berge [2, p.165]). The fact that maximal increasing relations give rise to one-dimensional convex functions (as we are about to prove) is well known in network theory, at least in special cases. It has been an important tool in proving the existence of solutions to non-linear network problems, as we shall explain later.

THEOREM 8-H

(a) Let Φ be a non-decreasing extended-real-valued function defined for all $\xi \in R$, such that $\Phi(\xi)$ is not always $-\infty$ or always $+\infty$. Let

$$\Phi^+(\xi) = \inf_{\lambda > \xi} \Phi(\lambda) = \lim_{\lambda \searrow \xi} \Phi(\lambda),$$

$$\Phi^-(\xi) = \sup_{\lambda < \xi} \Phi(\lambda) = \lim_{\lambda \nearrow \xi} \Phi(\lambda)$$

for all $\xi \in R$, and define r by

(8.16) $\quad \xi^* = r(\xi)$ if and only if $\xi^* \in R$ and $\Phi^-(\xi) \le \xi^* \le \Phi^+(\xi)$.

Then r is a maximal increasing relation.

(b) Conversely, let r be any maximal increasing relation. Define Φ as follows. For each $\xi^* \in$ dom r, let $\Phi(\xi)$ denote some particular $\xi^* \in R$ such that $\xi^* = r(\xi)$. Let $\Phi(\xi) = -\infty$ if $\xi < \lambda$ for all $\lambda \in$ dom r, and let $\Phi(\xi) = \infty$ if $\xi > \lambda$ for all $\lambda \in$ dom r. Then Φ has the properties required in (a) and r is given by (8.16).

Proof: (a) Let r be defined by (8.16). Suppose that $\xi_1^* = r(\xi_1)$ and $\xi_2^* = r(\xi_2)$, with $\xi_1 < \xi_2$. Then

$$\xi_1^* \leq \Phi^+(\xi_1) \leq \Phi^-(\xi_2) \leq \xi_2^*.$$

Thus (8.15) is satisfied, and r is an increasing relation. Next choose any $\xi \in R$ and $\xi^* \in R$ such that $\xi^* \neq r(\xi)$. Then either $\xi^* > \Phi^+(\xi)$ or $\xi^* < \Phi^-(\xi)$. In the first case, by definition of Φ^+, there exists some $\xi_0 > \xi$ such that $\xi^* > \Phi(\xi_0) = \xi_0^* \in R$. Now $\xi_0^* = r(\xi_0)$ and $\xi < \xi_0$ but $\xi^* > \xi_0^*$, so that r cannot be extended to include the pair $\langle \xi, \xi^* \rangle$ without violating (8.15). A similar argument works in the second case. Thus r is maximal.

(b) We show first that Φ is defined for all $\xi \in R$. This is true unless there exists some $\xi_0 \notin$ dom r such that $\xi_0 > \xi_1$ for some $\xi_1 \in$ dom r and $\xi_0 < \xi_2$ for some $\xi_2 \in$ dom r. In this event, by (8.15),

$$(8.16) \qquad -\infty < \sup \{\xi^* \mid \xi^* = r(\xi) \text{ for some } \xi < \xi_0\}$$
$$\leq \inf \{\xi^* \mid \xi^* = r(\xi) \text{ for some } \xi > \xi_0\} < \infty.$$

If ξ_0^* is any real number lying between the extrema in (8.17), we can add $\langle \xi_0, \xi_0^* \rangle$ to r without violating (8.15). This contradicts the maximality for r. Therefore Φ is defined for all $\xi \in R$, as asserted. It is immediate from (8.15) that Φ is non-decreasing; trivially, Φ is not identically $-\infty$ or identically $+\infty$. Thus Φ satisfies the hypothesis of (a). Moreover (8.15) implies that

$\Phi^-(\xi) \leq \xi^* \leq \Phi^+(\xi)$ whenever $\xi^* = r(\xi)$. Since r is already a maximal increasing relation, it follows from (a) that r coincides with the relation defined as in (8.16) by Φ.

The next theorem explains the exact sense in which the maximal increasing relations are the unbounded continuous increasing curves in the plane.

THEOREM 8-I

Let k and k_* be continuous finite non-decreasing functions, defined for all $\lambda \in R$, such that $k + k_*$ is strictly increasing and unbounded above or below. Define r by

(8.18) $\quad \xi^* = r(\xi)$ if and only if $\xi^* = k_*(\lambda)$ and $\xi = k(\lambda)$

for some $\lambda \in R$.

Then r is a maximal increasing relation. Moreover, every maximal increasing relation can be represented in this manner.

<u>Proof</u>: Since k and k_* are non-decreasing, (8.15) is satisfied when r is defined by (8.18). We show next that r is maximal. Suppose $\xi \in R$, $\xi^* \in R$ and $\xi^* \neq r(\xi)$. Choose $\lambda \in R$ such that

(8.19) $\quad \xi + \xi^* = k(\lambda) + k_*(\lambda)$.

This is possible by the hypothesis. Let $\xi_0 = k(\lambda)$ and $\xi_0^* = k_*(\lambda)$; then $\xi_0^* = r(\xi_0)$ by definition (8.18). Since $\xi^* \neq r(\xi)$, it follows from (8.19) that $\xi^* - \xi_0^* = -(\xi - \xi_0) \neq 0$. Therefore $\langle \xi, \xi^* \rangle$ could not be added to r without violating (8.15), so that r is maximal as asserted.

Conversely, suppose now that r is a maximal increasing relation. Then, for each $\lambda \in R$, there exists a unique $\xi \in R$ and a unique $\xi^* \in R$ such that $\xi^* = r(\xi)$ and $\xi + \xi^* = \lambda$. We shall prove the uniqueness first. Suppose that $\xi_1 + \xi_1^* = \lambda = \xi_2 + \xi_2^*$, where $\xi_1^* = r(\xi_1)$ and $\xi_2^* = r(\xi_2)$. Then $\xi_2^* - \xi_1^* = -(\xi_2 - \xi_1)$, and hence $\xi_2^* = \xi_1^*$ and $\xi_2 = \xi_1$ by (8.15). In proving the existence, we may suppose that r has been represented as in (8.16) for some function Φ. The properties of Φ imply that $\xi + \Phi(\xi)$ is strictly increasing and unbounded above or below. Hence, given any $\lambda \in R$, there exists some $\xi_0 \in R$ such that $\xi + \Phi(\xi) \geq \lambda$ for $\xi > \xi_0$ and $\xi + \Phi(\xi) \leq \lambda$ for $\xi < \xi_0$. Then

$$\Phi^-(\xi_0) \leq \lambda - \xi_0 \leq \Phi^+(\xi_0),$$

so that $\lambda - \xi_0 = r(\xi_0)$ by (8.16). Therefore ξ_0 and $\xi_0^* = \lambda - \xi_0$ have the required properties. Now, for each $\lambda \in R$, let $k(\lambda)$ and $k_*(\lambda)$ be the unique real numbers ξ and ξ^* such that $\xi^* = r(\xi)$ and $\xi + \xi^* = \lambda$. Then (8.18) holds. Moreover k and k_* are non-decreasing by (8.15). Inasmuch as

$$k(\lambda) + k_*(\lambda) = \lambda \text{ for all } \lambda \in R$$

by definition, $k + k_*$ is strictly increasing and unbounded above or below. Furthermore, $k + k_*$ has no jumps, so k and k^* are actually continuous. This completes the proof of the theorem.

THEOREM 8-J

If f is a closed proper convex function on R, then ∂f is a maximal increasing relation. Conversely, if r is a maximal increasing relation, then there exists a closed proper convex function f on R, unique up to an additive constant, such that $\Gamma = \partial f$.

<u>Proof</u>: The first assertion follows immediately from Theorems 8-A and 8-H. To prove the converse part of the theorem, we can assume that Γ is given by (8.16) for some function Φ having the properties specified in 8-H. Choose any $\xi_0 \in \operatorname{dom} f$. Then $\Phi(\xi) < \infty$ for all $\xi < \xi_0$ and $\Phi(\xi) > -\infty$ for all $\xi > \xi_0$. Define

$$(8.20) \qquad f(\xi) = \int_{\xi_0}^{\xi} \Phi(\mu)d\mu \text{ for all } \xi \in R.$$

The integral makes sense (at least as a Lebesque integral) despite the infinite values, due to the choice of ξ_0 and the fact that Φ is non-decreasing. For the same reasons, $f(\xi) > -\infty$ for all ξ and $\operatorname{dom} f = \left\{\xi \mid f(\xi) < -\infty\right\}$ is an interval, non-empty since $f(\xi_0) = 0$. Also, f is continuous on cl(dom f) by the ordinary properties of integrals, so that f is lower semi-continuous on R. Thus we shall know that f is a closed proper convex function on R as soon as we have proved that f is convex. It is enough to prove that f is convex on the interval dom f, where the integral is finite. Let $\xi_1 \in \operatorname{dom} f$, $\xi_2 \in \operatorname{dom} f$, $\xi_1 < \xi_2$, $0 < \lambda < 1$. Let $\xi_3 = (1-\lambda)\xi_1 + \lambda\xi_2$. Then $\Phi(\xi_3)$ is finite and

$$(1-\lambda)f(\xi_1)+\lambda f(\xi_2)-f(\xi_3) = (1-\lambda)(f(\xi_1)-f(\xi_3))+\lambda(f(\xi_2)-f(\xi_3))$$

$$= (1-\lambda)\int_{\xi_3}^{\xi_1} \Phi(\mu)d\mu + \lambda \int_{\xi_3}^{\xi_2} \Phi(\mu)d\mu$$

$$\geq (1-\lambda)(\xi_1-\xi_3)\Phi(\xi_3) + \lambda(\xi_2-\xi_3)\Phi(\xi_3)$$

$$= [(1-\lambda)\lambda(\xi_1-\xi_2)+\lambda(1-\lambda)(\xi_2-\xi_1)]\Phi(\xi_3) = 0$$

because Φ is non-decreasing. Thus

$$f((1-\lambda)\xi_1+\lambda\xi_2) \le (1-\lambda)f(\xi_1) + \lambda f(\xi_2).$$

This proves that f is convex. Finally, given any $\xi \in$ dom f and $\lambda > \xi$ we have

$$f(\lambda) - f(\xi) = \int_{\xi}^{\lambda} \Phi(\mu)d\mu \ge (\lambda-\xi)\Phi^{+}(\xi)$$

because Φ is non-decreasing. Fron this, and from the dual fact for $\lambda < \xi$, we conclude that

$$f'_{-}(\xi) \le \Phi^{-}(\xi) \text{ and } \Phi^{+}(\xi) \le f'_{+}(\xi)$$

for all $\xi \in$ dom f. But this is trivially ture also for $\xi \notin$ dom f. Since r is maximal, it follows now from (8.8) and (8.16) that $r = \partial f$.

The fact that f is unique up to an additive constant is not obvious, because f is not necessarily differentiable in the ordinary sense. Suppose that $\partial f_1 = r = \partial f_2$, where f_1 and f_2 are closed proper convex functions on R. Then ri(dom f_1) = ri(dom f_2) = D, say, by (8.4a), and $f'_{1+}(\xi) = f'_{2+}(\xi)$ and $f'_{1-}(\xi) = f'_{2-}(\xi)$ for all $\xi \in D$ by (8.8). We may assume that D is an open interval, for if it is a single point then trivially f_1 and f_2 differ by at most an additive constant. Let $h(\xi) = f_1(\xi) - f_2(\xi)$ for $\xi \in D$. Then h has left and right derivatives at all points of the open interval D, and these are all zero. Thus h is actually a differentiable function on D whose derivative vanishes identically, so that h is constant. This shows that, for some $\alpha \in R$, $f_2(\xi) = f_1(\xi) + \alpha$ for all $\xi \in D$. But f_1 and f_2 are closed, so the same formula must hold for all $\xi \in R$ by (2.8) and the definition of D. This proves the theorem.

Remark: It is true in general that, if f_1 and f_2 are closed proper convex functions on an n-dimensional space E such that $\partial f_1 = \partial f_2$, then $f_2 = f_1 + \alpha$ for some $\alpha \in R$. One can prove this by an argument similar to that given above, using Theorem C-C.

Maximal decreasing relations correspond in an entirely analogous manner to the differentials ∂g of closed proper concave functions on R.

We shall now apply the results proved above to the following problem:

(V_0) Given maximal increasing relations $r_1,\ldots,r_n$, maximal decreasing relations $s^*_1,\ldots,s^*_m$, and an $m \times n$ real matrix $((\alpha_{ij}))$, find real numbers $\xi_1,\ldots,\xi_n$ and $\eta^*_1,\ldots,\eta^*_m$ such that

$$\sum_j \alpha_{ij}\xi_j = s^*_i(\eta^*_i) \text{ and } \sum_i \eta^*_i \alpha_{ij} = r_j(\xi_j)$$

for $i=1,\ldots,m$ and $j=1,\ldots,n$.

THEOREM 8-K

The general problems (IV_0) and (V_0) are identical. Indeed, given (V_0), there exist closed proper convex and concave functions f_j and g_i on R such that

$$\partial f_j = r_j \text{ for } j=1,\ldots,n \text{ and } \partial g^*_i = s^*_i \text{ for } i=1,\ldots,m.$$

Then $\langle \xi_1,\ldots,\xi_n; \eta^*_1,\ldots,\eta^*_m \rangle$ is a solution to (V_0) if and only if it is a saddle-point in (III_0); moreover this happens if and only if $\langle \xi_1,\ldots,\xi_n \rangle$ is a solution to (I_0), $\langle \eta^*_1,\ldots,\eta^*_m \rangle$ is a solution to (II_0), and the extrema in (I_0) and (II_0) coincide.

Proof: The first part follows from Theorem 8-J, while the rest is a specialization of Theorem 3-B.

COROLLARY 8-L

In order that (V_0) have a solution, it is necessary and sufficient that

(8.21a) there exist $\xi_j \in \text{dom } r_j$ for $j=1,\ldots,n$ such that $\sum_j \alpha_{ij}\xi_j \in \text{range } s_i^*$ for $i=1,\ldots,m$,

and also

(8.21b) there exist $\eta_i^* \in \text{dom } s_i^*$ for $i=1,\ldots,m$ such that $\sum_i \eta_i^*\alpha_{ij} \in \text{range } r_j$ for $j=1,\ldots,n$.

Proof: The necessity of the condition is trivial. Its sufficiency is an immediate consequence of 8-F, 4-F and 8-K.

Problem (V_0) appears in network theory with $((\alpha_{ij}))$ as the incidence matrix of the network, ξ_j as the flow in the jth branch, and η_i^* as the potential at the ith node. Birkhoff and Diaz [3] proved an existence theorem for (V_0) in this context. This theorem may be viewed as a special case of 8-L in which the r_j are actually strictly increasing continuous functions unbounded above or below, with $\text{dom } r_j = R$ and $r_j(0) = 0$, and each s_i^* is either a continuous non-increasing function with $\text{dom } s_i^* = R$ and $s_i^*(0) = 0$, or is of the form:

$$\lambda = s_i^*(\lambda^*) \text{ if and only if } \lambda^* = \beta_i^* \text{ (and } \lambda \text{ is arbitrary)},$$

where β_i^* is a given constant. Under these assumptions, conditions (8.21a) and (8.21b) are always trivially satisfied. A somewhat more general result along these lines has been proved by Dwyer [48]. An existence theorem for networks due to Minty [51, Theorem 8.1] follows from 8-L when the s_i^* all vanish identically, i.e. when

$$\lambda^* = s_i^*(\lambda) \text{ if and only if } \lambda = 0 \ (\lambda^* \text{ arbitrary}).$$

These authors all characterize the solutions of (V_0) (in the various special cases) as the solutions of some problem of form (I_0) or (II_0) (or both). (See also Millar [50, Theorem 2], Berge [49] and [50, p.162ff.].) Minty also considers a problem related to (III_0). Aside from these results in network theory, and the linear programming case, 8-K and 8-L are new.

APPENDIX A

Support Functions and Relative Interiors of Convex Sets

As Fenchel has pointed out [19, p.101], the conjugate correspondence 2-E between convex functions includes the classical correspondence between convex sets and their support functions [4, p.23]. After stating this fact here in a form convenient for reference, we shall use it to characterize relative interiors. Some new results will then be obtained about the behavior of relative interiors under certain operations, such as the addition of convex sets. Familarity with the material in § 2 will be assumed.

A convex function f on E is said to be <u>positively homogeneous</u> if

(A.1) $f(\lambda x) = \lambda f(x)$ for all $x \in E$ and $\lambda > 0$.

This happens if and only if gph f is a convex cone. Since the closure of a convex cone is again a convex cone, it follows from (2.5) that cl f is positively homogeneous if f is.

Let C be a non-empty convex set in E. The <u>convex support function</u> $\check{\sigma}_C$ of C is then defined by

(A.2a) $\check{\sigma}_C(x^*) = \check{\sigma}(x^* \mid C) = \sup\left\{[x, x^*] \mid x \in C\right\}$

$= \sup_x \left\{[x, x^*] - \check{\delta}_C(x)\right\}$ for each $x^* \in E^*$.

Dually, the concave support function $\hat{\sigma}_C$ of C is

(A.2b) $\hat{\sigma}_C(x^*) = \hat{\sigma}(x^* \mid C) = \inf\left\{[x, x^*] \mid x \in C\right\}$

$= \inf_x \left\{[x, x^*] - \hat{\delta}_C(x)\right\}$ for each $x^* \in E^*$.

Notice that, by definition,

(A.3) $\check{\sigma}_C = \check{\delta}_C^*$ and $\hat{\sigma}_C = \hat{\delta}_C^*$,

i.e. the support functions of convex sets are precisely the conjugates of the characteristic functions defined in (2.15) and (2.19). Obviously $C_1 \supseteq C_2$ if and only if $\check{\delta}_{C_1} \leq \check{\delta}_{C_2}$, so that by 2-E

(A.4) $\check{\sigma}_{C_1} \geq \check{\sigma}_{C_2}$ if and only if cl $C_1 \supseteq$ cl C_2.

THEOREM A-A

(a) If C is a non-empty convex subset of E, then $\check{\sigma}_C$ is a closed, proper, positively homogeneous convex function on E^*, and

(A.5) $$\text{cl } C = \left\{ x \mid [x,x^*] \leq \check{\sigma}(x^*|C) \text{ for all } x^* \in E^* \right\}$$

(b) If h is any proper, positively homogeneous, convex function on E^*, then cl h = $\check{\sigma}_C$ where

$$C = \left\{ x \mid [x,x^*] \leq h(x^*) \text{ for all } x^* \in E^* \right\}$$

is a non-empty closed convex subset of E.

Proof: (a) The fact that $\check{\sigma}_C$ is a closed proper convex function is a specialization of Theorem 2-D, in view of (A.3), while the positive homogeneity is obvious from the definitions. Moreover by (A.3), 2-D and (2.16),

$$\check{\sigma}(x|\text{cl } C) = \check{\sigma}_C^{**}(x) = \sup_{x^*} \left\{ [x,x^*] - \check{\sigma}(x^*|C) \right\},$$

which says that $x \in$ cl C if and only if $0 \geq [x,x^*] - \check{\sigma}(x^*|C)$ for all x^*.

(b) Observe first that, by 2-D with the roles of E and E^* reversed, cl h = f^*, where

(A.6) $$f(x) = \sup_{x^*} \left\{ [x,x^*] - h(x^*) \right\}$$

is a closed proper convex function on E. Substituting $x^* = \lambda z^*$ (for an arbitrary fixed $\lambda > 0$) and applying the fact that h is positively homogeneous, we see that

$$f(x) = \sup_{z^*} \left\{ [x,\lambda z^*] - h(\lambda z^*) \right\} = \lambda \sup_{z^*} \left\{ [x,z^*] - h(z^*) \right\} = \lambda f(x).$$

This is true for every $\lambda > 0$, so $f(x)$ is either 0 or ∞ for each $x \in E$. Thus $f = \check{\delta}_C$, and hence cl $h = \check{\delta}_C$ by (A.3), for some convex set C (closed and non-empty because f is closed and proper). The asserted formula for C is now a consequence of (A.6).

COROLLARY A-B

The closed proper convex characteristic functions on E correspond one-to-one with the closed proper positively homogeneous convex functions on E^* under the conjugate operation.

Of course convex sets in E^* correspond dually to positively homogeneous convex functions on E. There are also analogous correspondences for convex sets and positively homogeneous concave functions.

It is interesting to note that a proper convex function is <u>both</u> a characteristic function and positively homogeneous if and only if it is of the form $\check{\delta}_K$, where K is a non-empty convex cone. The conjugate of such a function must again be the characteristic function of a non-empty convex cone, for the properties of being a characteristic function and positively homogeneous are dual to one another by the above results. In fact if K is a non-empty convex cone in E, then

(A.7a) $\quad \check{\delta}_K^* = \check{\delta}_K = \check{\delta}_{K^-}$, where $K^- = \{x^* \mid [x, x^*] \leq 0$ for all $x \in K\}$.

The closed convex cone $K^- \subseteq E^*$ in (A.7a) is called the <u>polar</u> of K. Applying (A.5) to (A.7a), we conclude that the polar of K^- is in turn cl K. In particular, the well known polar correspondence between non-empty closed convex cones (see [6, p.52]) is a special case of the conjugate correspondence. For concave functions, one has the dual result

(A.7b) $\quad \hat{\delta}^*_K = \hat{\sigma}_K = \hat{\delta}_{K^+}$, where $K^+ = \{x^* \mid [x,x^*] \geq 0$ for all $x \in K\}$.

In particular, when $K = M$ is a subspace of E, then

(A.7c) $\quad \check{\delta}^*_M = \check{\sigma}_M = \check{\delta}_{M^*}$, where $M^* = \{x^* \mid [x,x^*] = 0$ for all $x \in M\}$.

Useful properties of the relative interior of a convex set are described by the next theorem.

THEOREM A-C

Let C be a non-empty convex set in E and let $x \in E$. Then if x has one of the following properties it has them all:

(a) $x \in \text{ri } C$,

(b) For each $z \in C$ there exists some $\varepsilon > 0$ such that $x - \varepsilon(z-x) \in C$,

(c) $\hat{\sigma}(x^* \mid C) \geq [x,x^*]$ for each $x^* \in E^*$ such that $\check{\sigma}(x^* \mid C) \leq [x,x^*]$.

Proof: (a) implies (b): Let L be the smallest linear manifold containing C, so that ri C is the interior of C relative to L. If $x \in \text{ri } C$ and $z \in L$, we will have $x + \lambda(z-x) \in \text{ri } C \quad C$ for $-\varepsilon \leq \lambda < \varepsilon$, provided $\varepsilon > 0$ is small enough. In particular, this will be true when $z \in C$.

(b) implies (c): If $\hat{\sigma}(x^* \mid C) < [x,x^*]$, then by definition (A.2b) there must exist some $z \in C$ such that $[z,x^*] < [x,x^*]$. Applying (b), we get

$$\check{\sigma}(x^* \mid C) \geq [x - \varepsilon(z-x), x^*] > [x,x^*]$$

by definition (A.2a).

(c) implies (a): Suppose $x \notin \text{ri } C$. We shall show that then (c) cannot hold. If actually $x \notin \text{cl } C$, there exists some $x_0^* \in E^*$ such that $[x, x_0^*] > \check{\sigma}(x_0^* | C)$ by Theorem A-A. But trivially $\check{\sigma}(x_0^* | C) \geq \hat{\sigma}(x_0^* | C)$, so (c) is violated by x_0^*. Therefore we can suppose $x \in \text{rb } C$. It is well known that a non-empty open convex set can be separated from a point not in it by a hyperplane [15, p.20]. Since ri C is an open convex set relative to a certain linear manifold which also contains x, and ri $C \neq \emptyset$ [15, p.16], there exists (by an elementary argument based on the fact just mentioned) some hyperplane in E containing x but disjoint from ri C. Thus there exists some $x_0^* \in E^*$ such that

$$[x, x_0^*] \geq \check{\sigma}(x_0^* | C) > \hat{\sigma}(x_0^* | C).$$

In this event x cannot have property (c). Thus (c) implies (a) as asserted.

If C_1 and C_2 are convex sets in E, so is

$$C_1 + C_2 = \{x_1 + x_2 \,|\, x_1 \in C_1,\ x_2 \in C_2\}$$

and by direct calculation

$$\text{(A.8)} \qquad \begin{aligned} \check{\sigma}(x^* | C_1 + C_2) &= \check{\sigma}(x^* | C_1) + \check{\sigma}(x^* | C_2), \\ \hat{\sigma}(x^* | C_1 + C_2) &= \hat{\sigma}(x^* | C_1) + \hat{\sigma}(x^* | C_2) \end{aligned}$$

where $C_1 \neq \emptyset \neq C_2$. It is easy to show that

$$\text{(A.9)} \qquad \text{cl}(C_1 + C_2) \supseteq \text{cl } C_1 + \text{cl } C_2,$$

but examples are known where the inclusion is proper. More can be proved, however, for relative interiors.

THEOREM A-D

If C_1 and C_2 are convex sets in E, then $\text{ri}(C_1 + C_2) = \text{ri } C_1 + \text{ri } C_2$.

Proof: We can suppose $C_1 \neq \emptyset \neq C_2$, for the theorem is trivial otherwise. First we prove ri C_1 + ri $C_2 \subseteq$ ri(C_1+C_2) using condition (c) of Theorem A-C Let $x_1 \in$ ri C and $x_2 \in$ ri C_2. If $[x_1+x_2, x^*] \geq \check{\sigma}(x^*|C_1+C_2)$, then

$$[x_1, x^*] + [x_2, x^*] \geq \check{\sigma}(x^*|C_1) + \check{\sigma}(x^*|C_2)$$

by (A.8). Since in particular $x_1 \in C_1$ and $x_2 \in C_2$, this implies

$$[x, x^*] \geq \check{\sigma}(x^*|C_1) \text{ and } [x_2, x^*] \geq \check{\sigma}(x^*|C_2),$$

and hence by Theorem A-C

$$[x, x^*] \leq \hat{\sigma}(x^*|C_1) \text{ and } [x_2, x^*] \leq \hat{\sigma}(x^*|C_2).$$

Therefore

$$[x_1+x_2, x^*] \leq \hat{\sigma}(x^*|C_1) + \hat{\sigma}(x^*|C_2) = \hat{\sigma}(x^*|C_1+C_2)$$

by (A.8). Thus $x_1 + x_2 \in$ ri(C_1+C_2) according to condition (c) of Theorem A-C. The reverse inclusion will follow from (2.1) and (A.9). Namely

$$\text{ri } C_1 + \text{ri } C_2 \supseteq \text{ri}(\text{ri } C_1 + \text{ri } C_2) = \text{ri}(\text{cl}(\text{ri } C_1 + \text{ri } C_2))$$

$$\supseteq \text{ri}(\text{cl}(\text{ri } C_1) + \text{cl}(\text{ri } C_2)) = \text{ri}(\text{cl } C_1 + \text{cl } C_2) \supseteq \text{ri}(C_1 + C_2).$$

This completes the proof of the theorem.

The above fact has not previously appeared in the literature. It is very closely related, however, to the next theorem, which is essentially one of Fenchel's results [19, p.48].

THEOREM A-E (Fenchel's Separation Theorem)

Let C_1 and C_2 be non-empty convex sets in E. Then ri $C_1 \cap$ ri $C_2 = \emptyset$ if and only if there exists some $x^* \in E^*$ and $\alpha \in R$ such that

$$[x_1, x^*] \leq \alpha \leq [x_2, x^*] \text{ for all } x_1 \in C_1,\ x_2 \in C_2.$$

with strict inequality for at least one $x_1 \in C_1$ or for at least one $x_2 \in C_2$.

Proof: Using the notation $-C = \{-x \mid x \in C\}$, we see from Theorem A-D that $\text{ri } C_1 \cap \text{ri } C_2 = \emptyset$ if and only if

(A.10) $\quad 0 \notin \text{ri } C - \text{ri } C_2 = \text{ri } C_1 + \text{ri}(-C_2) = \text{ri}(C_1 - C_2)$,

It is obvious that

(A.11) $\quad \check{\sigma}(x^* \mid -C) = -\hat{\sigma}(x^* \mid C), \; \check{\sigma}(x^* \mid -C) = -\hat{\sigma}(x^* \mid C),$

which by (A.8) implies

(A.12) $\quad \check{\sigma}(x^* \mid C_1 - C_2) = \check{\sigma}(x^* \mid C_1) - \hat{\sigma}(x^* \mid C_2),$

$$\hat{\sigma}(x^* \mid C_1 - C_2) = \hat{\sigma}(x^* \mid C_1) - \check{\sigma}(x^* \mid C_2).$$

Furthermore, according to Theorem A-C, $C \notin \text{ri } C$ if and only if

$$0 \geq \check{\sigma}(x^* \mid C) > \hat{\sigma}(x^* \mid C)$$

for some $x^* \in E^*$. By (A.12), condition (A.10) is therefore equivalent to the existence of some $x^* \in E^*$ such that

$$0 \geq \check{\sigma}(x^* \mid C_1) - \hat{\sigma}(x^* \mid C_2) > \hat{\sigma}(x^* \mid C_1) - \check{\sigma}(x^* \mid C_2).$$

Thus $\text{ri } C_1 \cap \text{ri } C_2 = \emptyset$ if and only if, for some x^*,

$$\check{\sigma}(x^* \mid C_1) \leq \hat{\sigma}(x^* \mid C_2) \text{ but } \hat{\sigma}(x^* \mid C_1) < \check{\sigma}(x^* \mid C_2).$$

The theorem now follows from the definitions of the support functions.

COROLLARY A-F

Let C_1 and C_2 be non-empty convex sets in E. Then $\text{ri } C_1 \cap \text{ri } C_2 \neq \emptyset$ if and only if $\check{\sigma}(-x^* \mid C_1) \leq \hat{\sigma}(-x^* \mid C_2)$ for every x^* such that $\check{\sigma}(x^* \mid C_1) \leq \hat{\sigma}(x^* \mid C_2)$.

Proof: This follows from the last condition in the proof above, since

(A.13) $\quad \check{\sigma}(-x^* \mid C) = -\hat{\sigma}(x^* \mid C)$ and $\hat{\sigma}(-x^* \mid C) = -\check{\sigma}(x^* \mid C)$

according to the definitions.

The next result is new, except for special cases.

THEOREM A-G

Let $x \to Ax$ be a linear transformation from E to F. Let C be a convex set in E and let $A(C) = \{Ax \mid x \in C\}$. Then A(C) is a convex set in F and

$$\text{ri}(A(C)) = A(\text{ri } C).$$

<u>Proof</u>: The convexity of A(C) is well known, and easy to check. Also, the formula is trivial if $C = \emptyset$. Suppose therefore that $C \neq \emptyset$. and let $M = \{x \mid Ax = 0\}$. Then M is a subspace of E and $Ax \in A(C)$ if and only if $x \in C+M$. Now from condition (b) of Theorem A-C, we see that $Ax \in \text{ri}(A(C))$ if and only if, for each $Az \in A(C)$, there exists some $\varepsilon > 0$ such that $Ax - \varepsilon(Az - Ax) \in A(C)$. Hence ri(A(C)) is the set of vectors x with the property that, for each $z \in C+M$, there exists some $\varepsilon > 0$ such that $x - \varepsilon(z-x) \in C+M$. But x has this property if and only if $x \in \text{ri}(C+M)$, by Theorem A-C. Therefore

$$\text{ri}(A(C)) = A(\text{ri}(C+M)).$$

Moreover

$$\text{ri}(C+M) = \text{ri } C + \text{ri } M = \text{ri } C + M$$

by Theorem A-D and the fact that M is a subspace. But by the definition of M,

$$A((\text{ri } C)+M) = A(\text{ri } C).$$

This proves the theorem.

Incidentally, the theorem is not true if "ri" is replaced by "cl". This can be seen from geometric considerations; for instance, the projection of a half-hyperbola into a line perpendicular to one of its asymptotes will not be a closed set.

Finally, we explain how closures and relative interiors of convex sets behave under direct sum constructions.

THEOREM A-H

Let $E = E_1 \oplus \ldots \oplus E_k$, and let C_i be a convex set in E_i for $i = 1, \ldots k$. Let

$$C = \{x = \langle x_1, \ldots, x_k \rangle \mid x_i \in C_i, i=1, \ldots, k\}.$$

Then C is a convex set in E and

$$\text{ri } C = \{x = \langle x_1, \ldots, x_k \rangle \mid x_i \in \text{ri } C_i, i=1, \ldots k\},$$

$$\text{cl } C = \{x = \langle x_1, \ldots, x_k \rangle \mid x_i \in \text{cl } C_i, i=1, \ldots, k\}.$$

<u>Proof</u>: For $i = 1 \ldots, k$, let

$$C_i' = \{\langle 0, \ldots, x_i, 0, \ldots \rangle \mid x_i \in C_i\}.$$

Then C_i' is convex in E and, by the definition of relative interior

$$\text{ri } C_i' = \{\langle 0, \ldots, x_i, 0, \ldots \rangle \mid x_i \in \text{ri } C_i\}.$$

Moreover $C = C_1' + \ldots + C_k'$, so C is convex and by Theorem A-D

$$\text{ri } C = \text{ri } C_1' + \ldots + \text{ri } C_k' = \{\langle x_1, \ldots, x_k \rangle \mid x_i \in \text{ri } C_i, i=1, \ldots, k\}.$$

The statement about cl C is trivial.

APPENDIX B

Effective Domains and Level Sets

This Appendix is really an extension of §2, although some background material is drawn from Appendix A. It is devoted to the study of convex sets associated with a convex function and its conjugate. The results obtained are entirely new, and are meant to be counted among the main results of this paper.

Let f be a proper convex function on E with conjugate f^* on E^*. The first problem we shall consider is that of characterizing the convex set dom f^* by expressing its support function (see Appendix A) in terms of f. This will enable us to determine the properties of f dual to various useful properties of dom f^* such as boundedness.

Secondly, we shall calculate the support function of the <u>zero level set</u>

(B.1) $\quad \text{lev } f = \{x \mid f(x) \le 0\} \subseteq E$

in terms of f^*. It is immediate from Definitions 2-A and 2-B that lev f is a convex set, and that lev f is closed when f is closed. Other level sets can be represented as zero level sets, for example

(B.2) $\quad \text{lev}(f-\alpha) = \{x \mid f(x) \le \alpha\}, \quad \text{lev}(f-g) = \{x \mid f(x) \le g(x)\},$

where $\alpha \in R$ and g is concave. An investigation of such sets is worthwhile because of their importance in constrainted extremum problems.

THEOREM B-A

Let f be a proper closed convex function on E and let $x_0 \in$ dom f. Then the convex support function of the convex set dom f^* in E^* is given by

$$\check{\delta}(x|\mathrm{dom}\ f^{*}) = \sup_{\lambda>0}(f(x_0+\lambda x)-f(x_0))/\lambda = \lim_{\lambda\to\infty}(f(x_0+\lambda x)-f(x_0))/\lambda.$$

Proof: The second equality is due to the fact that $(f(x_0+\lambda x) - f(x_0))/\lambda$ is a non-decreasing function of $\lambda > 0$ as long as $x_0 + \lambda x \epsilon$ dom f (see [15, p.47]). To prove the first equality, we observe first that

$$\check{\delta}(x^{*}|\mathrm{dom}\ f^{*}) = \inf_{\lambda>0}(f(x_0)+f^{*}(x^{*})-[x_0,x^{*}])/\lambda.$$

This follows from the inequalities in Theorem 2-F. But the support function of a convex set is the conjugate of its characteristic function, as pointed out in Appendix A. Therefore

$$\check{\delta}(x|\mathrm{dom}\ f^{*}) = \sup_{x^{*}}\left\{[x,x^{*}]-\inf_{\lambda>0}\left\{f(x_0)+f^{*}(x^{*})-[x_0,x^{*}]/\lambda\right\}\right.$$

$$= \sup_{x^{*}}\sup_{\lambda>0}(1/\lambda)(-f(x_0)+[x_0+\lambda x,x^{*}]-f^{*}(x^{*}))$$

$$= \sup_{\lambda>0}(1/\lambda)(-f(x_0)+\sup_{x^{*}}\left\{[x_0+\lambda x,x^{*}]-f^{*}(x^{*})\right\})$$

$$= \sup_{\lambda>0}(1/\lambda)(-f(x_0)+f^{**}(x_0+\lambda x)).$$

Since f is closed and proper, $f^{**} = f$ by Theorem 2-D; hence this is the desired formula.

COROLLARY B-B

Let f be a closed proper convex function on E and let $x_0 \epsilon$ dom f. Then dom f^{*} is bounded if and only if

$$\lim_{\lambda\to\infty}(f(x_0+\lambda x)-f(x_0))/\lambda < \infty \text{ for all } x\epsilon E$$

Proof: A non-empty convex set is bounded if and only if its support functions are finite everywhere.

COROLLARY B-C

Let f be a closed proper convex function on E and let $x_0 \in$ dom f. Then dom $f^* = E^*$ if and only if

$$\lim_{\lambda \to \infty}(f(x_0+\lambda x)-f(x_0))/\lambda = \infty \text{ for all non-zero } x \in E.$$

Proof: According to (A.4), cl(dom f^*) = E^* if and only if

$$\check{\sigma}(x|\text{dom } f^*) = \check{\sigma}(x|E^*) = \check{\delta}(x|x=0).$$

But cl(dom f^*) = E^* if and only if dom $f^* = E^*$, since dom f^* is convex.

COROLLARY B-D

Let f be a closed proper convex function on E and let $x_0 \in$ dom f. Then $x \in E$ and $\alpha \in R$ have the property that

$$[x,x^*] = \alpha \text{ for all } x \in \text{dom } f^*$$

if and only if

$$f(x_0+\lambda x) = f(x_0)+\lambda\alpha \text{ for all } z \in \text{dom } f \text{ and } -\infty < \lambda < \infty.$$

Proof: The first condition is equivalent to

$$\alpha = \check{\sigma}(x|\text{dom } f^*) = -\hat{\sigma}(-x|\text{dom } f^*),$$

which by Theorem B-A is in turn equivalent to

$$\text{(B.3)} \qquad \alpha = \sup_{\lambda>0}(f(x_0+\lambda x)-f(x_0))/\lambda = -\sup_{\lambda>0}(f(x_0-\lambda x)-f(x_0))/\lambda.$$

The second condition in the corollary certainly implies (B.3). Conversely, if (B.3) holds then

$$(f(x_0+\lambda x)-f(x_0))/\lambda \leq \alpha \leq -(f(x_0-\lambda x)-f(x_0))/\lambda$$

for all $\lambda > 0$. This implies

$$f(x_0+\lambda x)-f(x_0) \leq \lambda\alpha \text{ and } f(x_0-\lambda x)-f(x_0) \leq -\lambda\alpha$$

for all $\lambda > 0$, in other words

$$h(\lambda) = f(x_0+\lambda x)-\lambda\alpha \leq f(x_0) \text{ for } -\infty < \lambda < \infty.$$

Now h is a convex function for $-\infty < \lambda < \infty$ whose maximum is $h(0)$. On the other hand, convexity implies that

$$h(0) = h((\tfrac{1}{2})\lambda+(\tfrac{1}{2})(-\lambda)) \leq (\tfrac{1}{2})(h(\lambda)+h(-\lambda))$$

for all λ, so we must have $h(\lambda) = h(0)$ for all λ. Therefore

$$f(x_0+\lambda x)-\lambda\alpha = f(x_0) \text{ for all } \lambda,$$

which is the second condition in the corollary.

COROLLARY B-E

Let f be a closed proper convex function on E and let $x_0 \in \text{dom } f$. Then dom f^* is a linear manifold if and only if, for each x such that

$$\lim_{\lambda\to\infty}(f(x_0+\lambda x)-f(x_0)/\lambda = \alpha < \infty,$$

it is actually true that

$$(f(x_0+\lambda x)-f(x_0))/\lambda = \alpha \text{ for } -\infty < \lambda < \infty.$$

Proof: By Theorem B-A and Corollary B-D, the latter condition means that

$$[x,x^*] = \alpha \text{ for all } x^* \in \text{dom } f^*$$

whenever

$$\infty > \alpha = \check{\delta}(x|\text{dom } f^*) = \sup\{[x,x^*]|x^* \in \text{dom } f^*\}.$$

Thus the condition says that every linear function bounded above on dom f^* must be constant there. This is certainly true if dom f^* is a linear manifold. On the other hand, suppose dom f^* is not a linear manifold. Then dom f^* has a relative boundary point x_0^*. Applying Fenchel's separation theorem (Theorem A-E) to $C_1^* = \text{dom } f^*$ and $C_2^* = \{x_0^*\}$ with the roles of E and E^* reversed, we can find some

$x \in R$ and $\alpha \in R$ such that

$$[x, x_1^*] \leq \alpha \leq [x, x_0^*] \text{ for all } x_1^* \in \text{dom } f^*$$

with strict equality for at least one $x_1^* \in$ dom f^*. But $x_0^* \in$ cl(dom f^*), so this linear function $h(x^*) = [x, x^*]$ is not constant on dom f^*, although it is bounded there. Hence the condition is both necessary and sufficient.

The next theorem dualizes some useful intersection properties of ri(dom f^*).

THEOREM B-F

Let f be a closed proper convex function on E and let $x_0 \in$ dom f. Let M^* be a subspace of E^* and let

$$M = \{x \mid [x, x^*] = 0 \text{ for all } x^* \in M^*\}$$

be the subspace of E orthogonal to M^*. Let $x_0^* \in E^*$. Then the linear manifold $x_0^* + M^*$ intersects ri(dom f^*) if and only if, for each $x \in M$ such that $f(x_0 + \lambda x) - \lambda[x, x_0^*]$ is a finite non-increasing function of $\lambda > 0$, actually $f(x_0 + \lambda x) - \lambda[x, x_0^*]$ is constant for $-\infty < \lambda < \infty$.

Proof: We shall apply Theorem A-F. Let $C_1^* =$ dom f^* and $C_2^* = x_0^* + M^*$. Since $x_0^* + M^*$ is a linear manifold, it intersects ri(dom f^*) if and only if ri $C_1^* \cap$ ri $C_2^* \neq 0$. Also,

$$\hat{\delta}(x \mid C_2^*) = \hat{\delta}(x \mid x_0^* + M^*) = \hat{\delta}(x \mid M) + [x, x_0^*],$$

$$\check{\delta}(x \mid C_1^*) = \check{\delta}(x \mid \text{dom } f^*) = \sup_{\lambda > 0} (f(x_0 + \lambda x) - f(x_0))/\lambda.$$

(See Theorem B-A.) A necessary and sufficient condition that $x_0^* + M^*$ intersect ri(dom f^*), according to Theorem A-F, is therefore that

$$(f(x_0-\lambda x)-f(x_0))/\lambda \le \hat{\delta}(-x|M) - [x,x_0^*] \text{ for all } \lambda > 0$$

for each $x \in E$ such that

$$f(x_0+\lambda x) - f(x_0)/\lambda \le \hat{\delta}(x|M) + [x,x_0^*] \text{ for all } \lambda > 0.$$

This is equivalent to the condition that:

(B.4) $$f(x_0+\lambda x) - \lambda[x,x_0^*] \le f(x_0) \text{ for all } \lambda < 0,$$

for each $x \in M$ such that

(B.5) $$f(x_0+\lambda x) - \lambda[x,x_0^*] \le f(x_0) \text{ for all } \lambda > 0.$$

But (B.4) and (B.5), taken together are equivalent to $f(x_0+\lambda x) - \lambda[x,x_0^*]$ being constant for $-\infty < \lambda < \infty$ (by an argument already used in the proof of Corollary B-D). On the other hand, (B.5) by itself is equivalent to $f(x_0+\lambda x) - \lambda[x,x_0^*]$ being a finite non-increasing function of $\lambda > 0$. The latter condition trivially implies (B.5), so to prove this assertion it will be enough to prove that if $h(\lambda)$ is a convex function such that

(B.6) $$h(\lambda) \le h(0) \text{ for all } \lambda > 0,$$

then h is non-increasing for $\lambda > 0$. Let $0 \le \lambda_1 \le \lambda_2 < \lambda$, and set

(B.7) $$\mu_1 = (\lambda-\lambda_2)/(\lambda-\lambda_1), \quad \mu_2 = (\lambda_2-\lambda_1)/(\lambda-\lambda_1).$$

Then $\mu_1 \ge 0$, $\mu_2 \ge 0$, $\mu_1 + \mu_2 = 1$, and $\mu_1\lambda_1 + \mu_2\lambda = \lambda_2$.

Therefore by (B.6) and the convexity of h

$$h(\lambda_2) \le \mu_1 h(\lambda_1) + \mu_2 h(\lambda) \le \mu_1 h(\lambda_1) + \mu_2 h(0).$$

Substituting (B.7) into this inequality and taking the limit as $\lambda \to \infty$, we get $h(\lambda_2) \le h(\lambda_1)$. Thus we have verified that $h(\lambda_1) \ge h(\lambda_2)$ whenever $0 \le \lambda_1 \le \lambda_2$, and the proof of the theorem is complete.

In particular, by choosing $M^* = \{0\}$ (and hence $M = E$) one obtains from Theorem B-F a characterization of the points x_0^* in $\mathrm{ri}(\mathrm{dom}\, f^*)$.

We turn now to the study of the zero level set lev f defined in (B.1).

THEOREM B-G

Let f be a closed proper convex function on E. Then

(B.8) $$\check{\delta}(x^* \mid \mathrm{lev}\, f) = \mathrm{cl}_{x^*} \inf_{0<\lambda<\infty} f^*(\lambda x^*)/\lambda$$

except when lev f = $\emptyset$, in which case the function on the right is improper.

<u>Proof</u>: Let $h(x^*) = \inf\{f^*(\lambda x^*)/\lambda \mid 0 < \lambda < \infty\}$ for each x^* in E^*. We shall prove first that h is convex according to Definition 2-A. It will be enough to show that

(B.9) if $h(x_1^*) < \mu_1 < \infty$, $h(x_2^*) < \mu_2 < \infty$, $0 < \mu < 1$,

then $h(\mu x_1^* + (1-\mu)x_2^*) < \mu\mu_1 + (1-\mu)\mu_2$.

The hypothesis of (B.9) implies the existence of positive real numbers λ_1 and λ_2 such that

(B.10) $f^*(\lambda_1 x_1^*)/\lambda_1 < \mu_1$ and $f^*(\lambda_2 x_2^*)/\lambda_2 < \mu_2$.

Let $x^* = \mu x_1^* + (1-\mu)x_2^*$ and set

$$\lambda = \lambda_1\lambda_2/(\mu\lambda_1 + (1-\mu)\lambda_2) > 0,$$

$$\lambda_1' = \lambda\mu/\lambda_1 > 0, \quad \lambda_2' = \lambda(1-\mu)/\lambda_2 > 0.$$

Then $\lambda_1' + \lambda_2' = 1$, and therfore by (B.10) and the definition of h,

$$h(\mu x_1^* + (1-\mu)x_2^*) \le f^*(\lambda x^*)/\lambda = f^*(\lambda_1'(\lambda_1 x_1^*) + \lambda_2'(\lambda_2 x_2^*))/\lambda$$

$$\le (\lambda_1'/\lambda)f^*(\lambda_1 x_1^*) + (\lambda_2'/\lambda)f^*(\lambda_2 x_2^*) = (\mu/\lambda_1)f^*(\lambda_1 x_1^*) + ((1-\mu)/\lambda_2)f^*(\lambda_2 x_2^*)$$

$$< \mu\mu_1 + (1-\lambda)\mu_2.$$

Thus h is convex as claimed. Now we calculate the conjugate of h. Since $f^{**} = f$ by Theorem 2-D,

$$h^*(x) = \sup_{x^*}\left\{[x,x^*] - \inf_{0<\lambda<\infty} f^*(\lambda x^*)/\lambda\right\}$$

$$= \sup_{0<\lambda<\infty}\left\{(1/\lambda)\sup_{x^*}\left\{[x,\lambda x^*] - f^*(\lambda x^*)\right\}\right\}$$

$$= \sup_{0<\lambda<\infty} f(x)/\lambda = \check{\delta}(x|f(x)\le 0) = \check{\delta}(x|\text{lev } f).$$

But a convex function is proper is and only if its conjugate is proper, by 2-D and the remarks following it. Therefore h is proper if and only if lev $f \neq \emptyset$, and, when this happens,

$$\text{cl } h(x^*) = h^{**}(x^*) = \check{\delta}(x^*|\text{lev } f)$$

by 2-D and (A.3). This proves (B.8).

COROLLARY B-H

Let f be a closed proper convex function on E and let $\alpha > \inf f$. Then $\{x|f(x) \le \alpha\}$ is a compact convex set if and only if 0 is in the interior of dom f^*. When this happens, the closure operation in (B.8) can be omitted.

Proof: $\alpha > \inf f$ if and only if $0 > \inf(f-\alpha)$. Also, $(f-\alpha)^* = f^* + \alpha$, so that $\text{dom}(f-\alpha)^* = \text{dom } f^*$. Therefore we need only prove this for the case where $\alpha = 0$. Since now $0 > \inf f$ by assumption, the closed convex set lev f is non-empty. It is bounded if and only if the effective domain of its convex support function is all of E^*. According to Theorem B-G this occurs if and only if, for all $x^* \in E^*$, there exists some $\lambda > 0$ such that $\lambda x^* \in \text{dom } f^*$. The latter is equivalent to the condition that 0 is in the interior of dom f^*, because dom f^* is convex. The final statement of the corollary is valid, because a convex function whose effective domain is the whole space is automatically closed (see(2.7)).

In applying formula(B.8), it is important to know when the closure operation can be omitted and when the infimum is attained. Corollary B-H gives one case where the first holds. We now describe a simple case where both hold.

THEOREM B-I

Let f be a convex function, finite on all of E, such that inf f < 0. Then

$$\check{\sigma}(x^*|\text{lev } f) = \begin{cases} \min_{0<\lambda<\infty} f^*(\lambda x^*)/\lambda \text{ if } x^* \neq 0, \\ 0 \text{ if } x^* = 0. \end{cases}$$

Proof: The hypothesis implies that f is a closed proper convex function on E and that lev f $\neq \emptyset$. The formula is trivially true for $x^* = 0$. In view of Theorem B-G, it will therefore suffice if we prove that:

(B.11) whenever $\{x_k^*\} \subseteq E^*$ and $\{\lambda_k > 0\}$ are sequences such that $x_k^* \to x^* \neq 0$ and $f^*(\lambda_k x_k^*)/\lambda_k \to \mu \in R$, then $f^*(\lambda x^*)/\lambda \leq \mu$ for some $\lambda > 0$.

Under the hypothesis of (B.11) we have

(B.12) $$[x, x^*] = \lim_{k\to\infty} (1/\lambda_k)([x, \lambda_k x_k^*] - f^*(\lambda_k x_k^*))$$

$$\leq \liminf_{k \to \infty} (1/\lambda_k) f(x) \text{ for all } x \in E,$$

by Theorem 2-F. If the $1/\lambda_k$ were unbounded above, the last expression would be $-\infty$ whenever $f(x) < 0$ (which is possible since $\inf f < 0$), and this would contradict the assumed finiteness of μ. Therefore, taking subsequences if necessary, we can suppose that

$$\lim_{k \to \infty} (1/\lambda_k) = \lambda_0, \quad 0 \leq \lambda_0 < \infty.$$

Since f is finite everywhere, we can now re-express (B.12) as

$$\mu \geq [x, x^*] - \lambda_0 f(x) \text{ for all } x \in E.$$

This implies $\lambda_0 > 0$, because $x^* \neq 0$ and μ is finite. Therefore $\lambda_0 = (1/\lambda)$ for some positive real number λ, and

$$\mu \geq \sup_x \left\{ [x, x^*] - (1/\lambda) f(x) \right\} = f^*(\lambda x^*)/\lambda.$$

This proves assertion (B.11), and hence proves the theorem.

It is interesting to observe that if f is a closed proper convex function on E, the graph sets of the convex support functions of dom f^* and lev f^* are related in a simple geometric manner to gph f. The second graph set is the closed convex cone in $E \oplus R$ generated by gph f, while the first is the asymptotic cone of gph f. (The asymptotic cone of a closed convex set C consists of all vectors x such that $z + \lambda x \in C$ for all $z \in C$ and $\lambda > 0$. See [19, p.41].)

APPENDIX C

Some Differential Properties of Convex Functions

In §2 we generalized the concept of "differential" by defining $x^* = \partial f(x)$, where f is a proper convex function on E, to mean simply that

(C.1) $$f(z) \geq f(x) + [z-x, x^*] \text{ for all } x \in E.$$

It was pointed out that if f were actually differentiable in the ordinary sense at x, then there would be exactly one vector x^* satisfying (C.1), namely

(C.2) $$x^* = \langle \frac{\partial f}{\partial \xi_1}(x), \ldots, \frac{\partial f}{\partial \xi_n}(x) \rangle$$

Thus the new definition is compatible with the familiar one (in which the differential is often called the "gradient"). Nevertheless, one would like to know whether the new differentials are somehow related to limits of difference quotients, so that the terminology is more justified. Ordinary differentials are characterized by their relation to directional derivatives. Namely, if f is differentiable at x, then the directional derivative of f at x with respect to $z = \langle \zeta_1, \ldots, \zeta_n \rangle$ is given by

(C.3) $$\lim_{\lambda \to 0} (f(x+\lambda z)-f(x))/\lambda = \zeta_1 \frac{\partial f}{\partial \xi_1}(x) + \ldots + \zeta_n \frac{\partial f}{\partial \xi_n}(x) = [z, \partial f(x)]$$

One might wonder whether an analog of (C.3) is true for the new differentials as well. Actually, the answers to these questions are already contained in Fenchel's work, although Fenchel did not make use of the vectors x^* satisfying (C.1) in theoretical developments (as we do, for example in devising "equilibrium conditions" in §3).

We shall explain Fenchel's result below.

The conjugate operation is essentially an extension of the classical Legendre tranformation (see [10; vol. I, p.238-242, and vol. II, p.26-31] which was used by Dennis [11, § F] to construct duals for certain convex programs with linear constraints. We shall analyze Legendre transformation here from the standpoint of the general theory of convex functions and their differentials as outlined in § 2. In particular, we shall explain the extent to which conjugates can be determined by the methods of the calculus.

DEFINITION C-A

Let f be a proper convex function on E and let $x \in$ dom f. Let $z \in E$. Then the <u>directional derivative</u> $f'(x;z)$ <u>of f at x with respect to z</u> is given by

$$f'(x',z) = \inf_{\lambda > 0} \big(f(x+\lambda z) - f(x)\big)/\lambda = \lim_{\lambda \to 0^+} \big(f(x+\lambda z) - f(x)\big)/\lambda$$

By a well known argument [15, p.49], the difference quotient appearing in this definition is a non-decreasing function of $\lambda > 0$, whence the equality between the "inf" and the "lim". The classical theory of directional derivatives of convex functions on open convex sets was developed by Bonnesen and Fenchel [4, p.18-21], and was extended to other convex sets by Fenchel in [19].

THEOREM C-B

Let f be a proper convex function on E and let $x \in$ dom f. Then $f(x;z)$ is a positively homogeneous (A.1) convex function of $z \in E$, which is always closed and proper when $x \in$ ri(dom f).

Proof: This can be proved by an elementary extension of the classical arguments (see also [19, p.79 ff]). When $x \in ri(dom\ f)$, the effective domain of $f'(x;z)$ as a function of z is a subspace of E, according to the characterization (b) of relative interiors in Theorem A-C. The fact that $f'(x;z)$ is closed in this case follows from (2.7), because a subspace has an empty relative boundary.

The directional derivatives in Definition C-A are related to the generalized differentials in Definition 2-G by the following result, due to Fenchel [19, p.103], which is based on the theory of support functions (see Appendix A).

THEOREM C-C

Let f be a proper convex function on E and let $x \in dom\ F$. Then $x^* = \partial f(x)$ if and only if

(C.4) $$f'(x;z) \geq [z, x^*] \text{ for all } z \in E.$$

Proof: If x^* is a differential of f at x, i.e. satisfies (C.1), then in particular

(C.5) $$f(x+\lambda z) \geq f(x) + [(x+\lambda z)-x, x^*] \text{ for all } z \in E,\ \lambda > 0,$$

from which (C.4) follows by Definition C-A. If x^* satisfies (C.4) then (C.5) holds by Definition C-A which in turn implies (C.1), i.e. that $x^* = \partial f(x)$.

COROLLARY C-D

Let f be a proper convex function on E and let $x \in E$. Then $x^* = \partial f(x)$ for at least one x^* if and only if $x \in dom\ f$ and $f'(x;z)$ is proper in z, and in this case

(C.6) $\quad cl_z f'(x;z) = \sup \left\{ [z,x^*] \mid x^* = \partial f(x) \right\}$

The closure operation in (C.6) is unnecessary when $x \in$ ri(dom f).

Remark: This is the analog of (C.3).

Proof: It is obvious from (C.1) that f has no differentials at x if $x \notin$ dom f, since then $f(x) = \infty$. The corollary therefore follows from Theorems C-B and A-A(b).

COROLLARY C-E

Let f be a proper convex function on E. Then $\left\{ x^* \mid x^* = \partial f(x) \right\}$ is a closed convex subset of E^* for each $x \in E$.

Proof: This results, along with the last corollary, from Theorems C-B and A-A(b).

COROLLARY C-F

Let f be a proper convex function on E and let $x \in E$. Then the following situations are equivalent to each other.

(a) x is in the interior of dom f and f is differentiable at x in the ordinary sense,

(b) $f'(x;z)$ is a linear function of z,

(c) f has exactly one differential x^* at x.

Moreover, in these situations the unique differential x^* of f at x is given by (C.2), and

$$f'(x;z) = [z,x^*] \text{ for all } z \in E.$$

Proof: The equivalence of (b) and (c) is evident from Corollary C-D. Also, if (b) holds x must be in the interior of dom f, for otherwise

$f'(x;z)$ would not be finite for all z. (See Theorem A-C.) This reduces the corollary to the classical case [4, p.20].

Suppose for a moment that f is a differentiable convex function defined on an open convex set C in E. Then the relation $x^* = \partial f(x)$ reduces to a system of partial differential equations

$$\xi_i^* = \frac{\partial f}{\partial \xi_i}(\xi_1, \ldots, \xi_n), \; i = 1, \ldots, n,$$

which conceivably can be solved for $\xi_1, \ldots, \xi_n$ in terms of $\xi_1^*, \ldots, \xi_n^*$. Denote the general solution abstractly by $x = (\partial f)^{-1}(x^*)$. The function $L(f)$ given by

$$L(f)(x^*) = [(\partial f)^{-1}(x^*), x^*] - f((\partial f)^{-1}(x^*))$$

is then called the <u>Legendre transform</u> of f. If $L(f)$ happens to be defined on an open set, and is differentiable there, the same procedure can be applied to $L(f)$. It turns out that $L(L(f)) = f$. Ordinarily this transformation is treated rather informally, and, in particular, questions about domains are neglected. Dennis [11, § E] proved rigorously, however, that, if f is strictly convex on C, the domain of $L(f)$ is open and $L(f)$ is also differentiable and strictly convex. He then derived a class of dual convex programs based on this Legendre correspondence. (He seems to assume in proving theorems about these programs, however, that f and $L(f)$ are both everywhere defined [11, § F].) Actually, the Legendre transformation is quite close to the conjugate operation. If we extend f to all of E, by assigning f the value ∞ outside of C in accordance with our custom, f^* may be calculated and compared with the similarly extended function $L(f)$. If $x^* \in \operatorname{dom} L(f)$ and $x = (\partial f)^{-1}(x^*)$, then $x^* = \partial f(x)$

and, by definition of L(f),

$$L(f)(x) = [x^*,x] - f(x).$$

But this expression is also $f^*(x^*)$, by Theorems 2-F and 2-H. Thus f^* coincides with L(f) where L(f) is defined, at least in the case studies by Dennis. The reverse is not necessarily true; in fact f^* may be finite on a much larger set, as we shall see.

Our objective below is, first of all, to deduce an analog of the Legendre transformation formula for general conjugate pairs of closed convex functions. Secondly, we shall characterize the closed convex functions whose differential mappings are one-to-one where defined, and then compare this situation with the classical one.

For a proper convex function f on E we define

$$\text{dom}\ \partial f = \{x \mid x^* = \partial f(x) \text{ for some } x^*\},$$
$$\text{range}\ \partial f = \{x^* \mid x^* = \partial f(x) \text{ for some } x\},$$

and similarly for the differential ∂f^* of f^*.

LEMMA C-G

Suppose that f is a <u>closed</u> proper convex function on E. Then

(a) $\text{ri}(\text{dom}\ f) \subseteq \text{dom}\ \partial f = \text{range}\ \partial f^* \subseteq \text{dom}\ f$,

(b) $\text{ri}(\text{dom}\ f^*) \subseteq \text{range}\ \partial f = \text{dom}\ \partial f^* \subseteq \text{dom}\ f^*$.

<u>Proof</u>: This just combines Theorem 2-H with Corollary 2-I and its dual.

<u>Remark</u>: The hypothesis that f is closed is crucial in the Lemma, as can be seen from the following example in the one-dimensional case. Let

$$f(\xi) = 0 \text{ if } -1 < \xi < 1,\ f(\xi) = \infty \text{ if } |\xi| \geq 1.$$

Then f is a non-closed proper convex function and

$$f^*(\xi^*) = |\xi^*| \text{ for all } \xi^*.$$

Moreover, now

$$\text{ri}(\text{dom } f) = \text{dom } \partial f = \text{dom } f = \{\xi \mid -1 < \xi < 1\},$$

$$\text{ri}(\text{dom } f^*) = \text{dom } \partial f^* = \text{dom } f^* = \{\xi^* \mid -\infty < \xi^* < \infty\},$$

$$\text{range } \partial f = \{0\}, \quad \text{range } \partial f^* = \{\xi \mid -1 \leq \xi \leq 1\}.$$

Note especially the fact that range ∂f can be much smaller than ri(dom f^*) when f is not closed.

It is not necessarily true, incidentally, that dom ∂f and range ∂f are convex sets, even when f is closed. However, when f is closed these sets must be "almost convex" by the Lemma, because they differ then from the convex sets dom f and dom f^* by at most the lack of certain relative boundary points.

The next theorem demonstrates that the conjugate operation can be thought of as a generalized Legendre transformation.

THEOREM C-H

Suppose that f is a closed proper convex function on E. Then the following facts are true, and provide a means of calculating f^* whenever ∂f is completely known.

(a) If $x^* \in$ range ∂f, then $f^*(x^*) = [x, x^*] - f(x)$, where x is any vector such that $x^* = \partial f(x)$,

(b) If $x^* \notin$ cl(range ∂f), then $f^*(x^*) = \infty$,

(c) ri(range ∂f) is non-empty,

(d) If $x^* \in$ cl(range ∂f) and $x_0^* \in$ ri(range ∂f), then

$$\lambda x_0^* + (1-\lambda)x^* \in \text{ri}(\text{range } \partial f) \text{ for } 0 < \lambda < 1,$$

and $f^{*}(x^{*})$ is given by

$$f^{*}(x^{*}) = \lim_{\lambda \to 0^{+}} f^{*}(\lambda x_0^{*} + (1-\lambda)x^{*}).$$

Proof: If $x^{*} = \partial f(x)$ then $f(x) + f^{*}(x^{*}) = [x, x^{*}]$ by Theorems 2-F and 2-H. This proves (a). Next we observe from (2.1) and Lemma C-G that

(C.7) $\quad \mathrm{cl}(\text{range } \partial f) = \mathrm{cl}(\text{dom } f^{*})$ and $\mathrm{ri}(\text{range } \partial f) = \mathrm{ri}(\text{dom } f^{*})$.

It follows that (b) and (c) are true and that range ∂f can be replaced by dom f^{*} where it occurs in (d). The first assertion in (d) now reduces to a familiar fact about convex sets [15, p.9], while the second is a consequence of (2.8).

Needless to say, Theorem C-H is easy to apply only when ∂f is easy to calculate. This will often be the case, however, when the methods of the calculus can be applied, for example when f is finite and differentiable everywhere. (See also §8).

DEFINITION C-I

A proper convex function f on E will be called regular if f is closed and ∂f is a one-to-one mapping from dom ∂f onto range ∂f. If f is regular and dom $\partial f = E$, range $\partial f = E^{*}$, f will be called completely regular.

THEOREM C-J

If f is a regular (completely regular) convex function on E, then f^{*} is a regular (completely regular) convex function on E^{*}.

Proof: Immediate from Theorem 2-H.

THEOREM C-K

Let f be a proper convex function on E. Then cl f is regular if and only if f has the following properties.

(a) ri(dom f) is actually open, and f is differentiable at every point of ri(dom f).

(b) f is <u>strictly convex</u> on ri(dom f), i.e.

$$f(\lambda x_1+(1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda)f(x_2) \text{ for } 0 < \lambda < 1,$$

for all x_1 and x_2 in ri(dom f), $x_1 \neq x_2$;

(c) there exists a point $x_0 \in$ ri(dom f) such that, for each $x \in$ rb(dom f),

(C.8) $$\lim_{\lambda \to 0^+} f'(x+\lambda(x_0-x);\ x_0-x) = -\infty.$$

<u>Remark</u>: If (c) holds for one such $x_0 \in$ ri(dom f), then it will hold for every point of ri(dom f), as the argument below shows. With ri(dom f) open, as in (a), (c) says that f becomes infinitely steep along each ray proceeding from the interior point x_0 of dom f to a boundary point x of dom f.

<u>Proof</u>: The proof is divided into several parts.

<u>Part 1</u>: It is enough to prove the theorem in the case where f is already closed. We shall show, namely, that f has these three properties if and only if cl f has them. Properties (a) and (b) depend only on the values of f on ri(dom f). The same is true of (c) because of the general fact (see [15, p.9]): if C is convex, $x_0 \in$ ri C and $x \in$ cl C, then $\lambda x_0 + (1-\lambda)x \in$ ri C for $0 < \lambda < 1$. Moreover cl f(x) = f(x) for all $x \in$ ri(dom f) by (2.7). Finally,

$$\text{ri}(\text{dom}(\text{cl } f)) = \text{ri}(\text{dom } f),$$

$$\text{rb}(\text{dom}(\text{cl } f)) = \text{rb}(\text{dom } f),$$

as can easily be proved from (2.1) by applying the operations "ri" and "cl" to (2.6).

Part 2: ∂f is single-valued at all points of dom ∂f if and only if (a) holds and dom ∂f = ri(dom f). This is a direct consequence of C-F and C-G.

Part 3: Suppose f is closed and let $x_0 \in$ ri(dom f). Then dom ∂f = ri(dom f) if and only if, for each $x \in$ rb(dom f), either $f(x) = \infty$, or $f(x) < \infty$, but $f'(x;x_0-x) = -\infty$. Indeed, by Lemma C-G, dom ∂f = ri(dom f) if and only if, for each $x \in$ rb(dom f), either $x \notin$ dom f, i.e. $f(x) = \infty$, or $x \in$ dom f but $x \notin$ dom ∂f. The latter means that $f'(x;z)$ is improper in z, according to Corollary C-D, which is certainly the case if $f'(x;x_0-x) = -\infty$. Moreover $f'(x;0) = 0$ trivially, so if $f'(x;z)$ is improper it must take on the value $-\infty$. In this event $f'(x;z) = -\infty$ for all $z \in$ ri K, where K is the effective domain of $f'(x;z)$ in z, by (2.7) and (2.9). But by Definition C-A

$$K = \{\lambda(z-x) \mid z \in \text{dom } f,\ \lambda \geq 0\} \supseteq \{z-x \mid z \in \text{dom } f\},$$

and this implies

$$\text{ri } K \supseteq \{z-x \mid z \in \text{ri}(\text{dom } f)\}.$$

Thus $x_0-x \in$ ri K, so that $f'(x;x_0-x) = -\infty$ if $f'(x;z)$ is improper in z.

Part 4: Suppose f is closed and let $x_0 \in$ ri(dom f), $x \in$ rb(dom f). Then (C.8) holds if and only if either $f(x) = \infty$, or $f(x) < \infty$ but $f'(x;x_0-x) = -\infty$. To prove this, we first reduce the situation to the one-dimensional case. Let $h(\lambda) = f(x+\lambda(x_0-x))$ for all λ. Then h is a proper convex function on the real line, $1 \in \mathrm{ri}(\mathrm{dom}\ h)$, $0 \in \mathrm{rb}(\mathrm{dom}\ h)$, and h is closed by (2.8). Moreover $f(x) = h(0)$ and

$$f'(x+\lambda(x_0-x);x_0-x) = h'(\lambda;1)$$

for all $\lambda \in \mathrm{dom}\ h$. We must prove that

(C.9) $$\lim_{\lambda \to 0^+} h'(\lambda;1) = -\infty$$

if and only if

(C.10) either $h(0) = \infty$, or $h(0) < \infty$ but $h'(0;1) = -\infty$.

First we show that $h'(\lambda;1)$ is a non-decreasing function of λ on the interval dom h. Let $\lambda \in \mathrm{dom}\ h$ and $\mu \in \mathrm{dom}\ h$, and suppose $\mu < \lambda$. By the definition of directional derivatives,

$$h(\lambda) - h(\mu) = h(\mu+1(\lambda-\mu))-h(\mu) \geq h'(\mu;\lambda-\mu).$$

Interchanging λ and μ are adding the second inequality to the first, we get

$$0 \geq h'(\mu;\lambda-\mu) + h'(\lambda;\mu-\lambda) = (\lambda-\mu)(h'(\mu;1)+h'(\lambda;-1))$$

by positive homogeneity (see Theorem C-B). Hence

(C.11) $$h'(\mu;1) \leq -h'(\lambda;-1)$$

On the other hand,

(C.12) $$-h'(\lambda;-1) \leq h'(\lambda;1) \text{ for } \lambda \in \mathrm{dom}\ h.$$

This is trivial if λ is an end-point of the interval dom h. Otherwise $\lambda \in \mathrm{ri}(\mathrm{dom}\ h)$ so $h'(\lambda;\lambda')$ is a proper convex function of λ' by Theorem C-B. Then

$$0 = h'(\lambda;0) = h'(\lambda;\ \tfrac{1}{2}(1)+\tfrac{1}{2}(-1)) \leq \tfrac{1}{2}(h'(\lambda;,1)+h'(\lambda;-1)),$$

from which (C.12) follows. Thus $h'(\lambda;1)$ is non-decreasing on dom h as asserted. Therefore

$$\infty > \inf_{0<\lambda<1} h'(\lambda;1) = \lim_{\lambda \to 0^+} h'(\lambda;1) = \lambda_0^* \geq -\infty,$$

and $\lambda_0^* > h'(0;1)$ whenever $0 \in \text{dom } h$. It is obvious from this that (C.9) implies (C.10). Conversely, suppose λ_0^* is finite. We shall show that then $h(0) < \infty$ and $h'(0;1) > -\infty$; this will prove that (C.10) implies (C.9). Let $1 > \lambda_1, \lambda_2, \ldots,$ be a sequence of positive real numbers decreasing to 0, and let

$$\lambda_k^* = h'(\lambda_k;1), \quad k = 1,2,\ldots .$$

(Each $\lambda_k \in \text{ri}(\text{dom } h)$, so the λ_k^* are all finite). Then

$$\text{(C.13)} \qquad \lim_{k\to\infty} \lambda_k = 0 \text{ and } \lim_{k\to\infty} \lambda_k^* = \inf_k \lambda_k^* = \lambda_0^* > -\infty.$$

Let $h^*(\lambda^*)$ be the conjugate of the convex function h. Then

$$\text{(C.14)} \qquad h(\lambda_k) + h^*(\lambda_k^*) \leq \lambda_k \lambda_k^* \text{ for } k = 1,2,\ldots .$$

This can be proved in the following way. By Theorem 2-H, (C.14) says that $\lambda_k^* = \partial h(\lambda_k)$, while the latter is true if and only if

$$\text{(C.15)} \qquad h(\lambda_k;\mu) \geq \mu\lambda_k^* \text{ for all } \mu \in R$$

by Theorem C-C. By the positive homogeneity of directional derivatives, (C.15) is equivalent to

$$-h'(\lambda_k;-1) \leq \lambda_k^* \leq h'(\lambda_k;1).$$

This is certainly true for the λ_k^* we have chosen, inasmuch as the outer inequality holds for all λ_k by (C.12). The proper convex functions h and h^* are both closed, so (C.13) and (C.14) now gives us

$$h(0) + h^*(\lambda_0^*) \leq 0 \cdot \lambda_0^*.$$

This implies that $h(0) < \infty$, and that $\lambda_0^* = \partial h(0)$ by Theorem 2-H. Moreover $h'(0;1) > -\infty$ in this case by Corollary C-D.

Part 5: Suppose ∂f is single-valued on ri(dom f). Then ∂f is also one-to-one on ri(dom f) if and only if f has property (b). Namely, let $x_1 \in \text{ri}(\text{dom } f)$, $x_2 \in \text{ri}(\text{dom } f)$, $x_1^* = \partial f(x_1)$, $x_2^* = \partial f(x_2)$, $0 < \lambda < 1$. If f has property (b) and $x_1^* = x_2^* = x^*$, Theorems 2-F and

2-H imply that

$$\begin{aligned} f(\lambda x_1+(1-\lambda)x_2) &\geq [\lambda x+(1-\lambda)x_2, x^*] - f^*(x^*) \\ &= \lambda([x_1,x^*] - f^*(x^*)) + (1-\lambda)([x_2,x^*]-f^*(x^*)) \\ &\geq \lambda f(x_1) + (1-\lambda)f(x_2), \end{aligned}$$

and hence that $x_1 = x_2$. Thus ∂f must be one-to-one in this case. Conversely suppose ∂f is one-to-one and that $x_1 \neq x_2$. Then x_1^*, x_2^* and $x^* = \partial f(\lambda x_1+(1-\lambda)x_2)$ are all different. Again applying Theorems 2-F and 2-H, we have

$$\begin{aligned} f(\lambda x_1+(1-\lambda)x_2) &\leq [\lambda x_1+(1-\lambda)x_2, x^*] - f^*(x^*) \\ &= \lambda([x_1,x^*]-f^*(x^*)) + (1-\lambda)([x_2,x^*]-f^*(x^*)) \\ &< \lambda f(x_1) + (1-\lambda)f(x_2). \end{aligned}$$

By this argument, f must have property (b).

<u>Part</u> 6: As shown in Part 1, f can be assumed closed. From Parts 2, 3 and 4, we know that ∂f is then single-valued at all points of dom ∂f if and only if (a) and (c) hold. Finally, ∂f is in addition one-to-one if and only if f has property (b), according to Part 5. This proves the theorem, and at the same time proves:

COROLLARY C-L

Let f be a proper convex function on E. Then $\partial(\text{cl } f)$ is single-valued on dom $\partial(\text{cl } f)$ if and only if f has properties (a) and (c) in Theorem C-H. In this case dom $\partial(\text{cl } f)$ = ri(dom f).

Theorem C-K precisely characterizes those convex functions for which the conjugate operation amounts to the Legendre transformation considered by Dennis. Indeed, the convex functions f whose closures are regular are those whose conjugates can be determined as follows.

Calculate ∂f on the open convex set ri(dom f) by ordinary differentiation. Then range ∂f is an open convex set and the equation $x^* = \partial f(x)$ has a unique solution $x = (\partial f)^{-1}(x^*)$ at each point x^* of range ∂f. Moreover

$$(C.16) \qquad f^*(x^*) = [(\partial f)^{-1}(x^*), x^*] - f((\partial f)^{-1}(x^*))$$

on range ∂f, and this set is actually ri(dom f). Finally, calculate the boundary values of f^* by the closure operation (2.8) and give f^* the value ∞ elsewhere. (Calculate the values of cl f similarly, so as to get a conjugate pair of closed convex functions.)

It is important to note that the above procedure is not enough to determine f^* when f satisfies (a) and (b) but not (c), even though the Legendre transform is well-defined in this case. For example, consider the one dimensional situation where

$$f(\xi) = \begin{cases} \frac{1}{2}\xi^2 & \text{if } 0 < \xi < 1, \\ \infty & \text{otherwise.} \end{cases}$$

Then $L(f) = f$, but

$$f^*(\xi^*) = \begin{cases} \frac{1}{2}(\xi^*)^2 & \text{if } 0 \le \xi^* \le 1, \\ \frac{1}{2}\xi^* & \text{if } 1 \le \xi^* < \infty, \\ 0 & \text{if } -\infty < \xi^* \le 0. \end{cases}$$

The final theorem of this appendix characterizes the completely regular convex functions.

THEOREM C-M

A convex function f on E is completely regular if and only if

(a) f is finite, differentiable and strictly convex everywhere on E, and

(b) $\lim_{\lambda \to \infty} (f(x_0+\lambda x)-f(x_0))/\lambda$ for all non-zero $x \in E$,

for some $x_0 \in E$.

<u>Proof</u>: Property (a) guarantees that f is regular and that dom $\partial f = E$, by Theorem C-K. But (b) is equivalent the condition that range $\partial f = E^*$ by Corollary B-C and Lemma C-G.

If f is completely regular, then, of course, formula (C.16) holds for every $x^* \in E^*$. An example of a completely regular convex function is

(C.17) $f(x) = \frac{1}{2} [x, Bx]$,

where B is a positive definite matrix. In this case $\partial f(x) = Bx$ for all x, so by (C.16)

(C.18) $f^*(x^*) = \frac{1}{2} [B^{-1}x^*, x^*]$,

where B^{-1} is the inverse of B.

REFERENCES

1. R. Bellman and W. Karush, "Mathematical programming and the maximum transform," J. Soc. Ind. Appl. Math. 10 (1962), 550-566.

2. C. Berge and A. Ghouila-Houri, Programmes Jeux et Réseaux de Transport (Dunod, Paris, 1962).

3. G. Birkhoff and J. B. Diaz, "Non-linear network problems," Q. Appl. Math. XIII (1956), 431-443.

4. T. Bonnesen and W. Fenchel, Theorie der Konvexen Korper (Springer, Berlin, 1934).

5. N. Bourbaki, Espaces Vectoriels Topologiques, Chap. I & II (Hermann & Cie., Paris, 1953).

6. N. Bourbaki, Espaces Vectoriels Topologiques, Chap. III, IV, & V (Hermann & Cie., Paris, 1955).

7. A. Charnes, "Constrained games and linear programming," Proc. Nat. Acad. Sci. U.S.A. 39 (1953), 639-641.

8. A. Charnes, W. W. Cooper and K. Kortanek, "A duality theory for convex programs with convex constraints," Proc. Amer. Math. Soc. 68 (1962), 605-608.

9. A. Charnes and W. W. Cooper, "Non-linear network flows and convex programming over incidence matrices," Naval Res. Log. Q. 5 (1958), 231-236.

10. R. Courant and D. Hilbert, Methods of Mathematical Physics, Vols. I & II (Interscience, New York, 1943).

11. J. B. Dennis, Mathematical Programming and Electrical Networks (Technology Press, Cambridge, Mass., 1959).

12. W. S. Dorn, "Duality in quadratic programming," Q. Appl. Math. XVIII (1960), 155-162.

13. W. S. Dorn, "A duality theorem for convex programs," IBM J. of Res. Devel. 4 (1960), 407-413.

14. R. J. Duffin, "Dual programs and minimum cost," J. Soc. Ind. Appl. Math. 10 (1962), 119-124.

15. H. G. Eggleston, *Convexity* (Cambridge Univ. Press, 1958).

16. E. Eisenberg, "Duality in homogeneous programming," Proc. Amer. Math. Soc. 12 (1961), 783-787.

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

18. W. Fenchel, "On conjgate convex functions," Canad. J. Math. 1 (1949), 73-77.

18. W. Fenchel, *Convex Cones, Sets and Functions*, lecture notes (multilith, Princeton, 1953).

20. D. Gale, *The Theory of Linear Economic Models* (McGraw-Hill, New York, 1960).

21. D. Gale, H. W. Kuhn and A. W. Tucker, "Linear programming and the theory of games," in *Activity Analysis of Production and Allocation* (T. C. Koopmans, ed.) (Wiley, New York, 1951), 317-329.

22. A. J. Goldman, "Resolution and separation theorems for polyhedral convex sets," in [37], 41-52.

23. A. J. Goldman and A. W. Tucker, "Theory for linear programming," in [37], 53-98.

24. A. J. Goldman and A. W. Tucker, "Polyhedral convex cones," in [37], 19-39.

25. M. A. Hanson, "A duality theorem in non-linear programming with non-linear constraints," Austral. J. Statist. 3 (1961), 64-72.

26. L. Hörmander, "Sur la fonction d'appui des esembles convexes dan une espace localement convexe," Arkiv för Mat. 3 (1954), 181-186.

27. S. Karlin, Mathematical Methods and Theory in Games, Programming and Economics, Vol. I (Addison-Wesley, Reading, Mass., 1960).

28. H. W. Kuhn and A. W. Tucker, "Non-linear programming," Proc. 2nd Berkeley Symp. Prob. and Statist. (Univ. of Calif. Press, Berkeley, 1951), 481-492.

29. R. D. Luce and H. Raiffa, Games and Decisions (Wiley, New York, 1957).

30. J. von Neumann and O. Morgenstern, Theory of Games and Economic Behavior (Princeton, 1953, 3rd edition).

31. R. T. Rockafellar, "A composite duality theorem for games and programs," submitted to Pac. J. Math.

32. F. W. Sinden, "A geometric representation for pairs of dual quadratic or linear programs," J. Math. Anal. Appl. 5 (1962), 378-302.

33. A. W. Tucker, "Dual systems of homogeneous linear relations, in [37], 3-18.

34. A. W. Tucker, "Linear and non-linear programming," J. Operations Res. Soc. Amer. 2 (1957), 244-257.

35. P. Wolfe, "Determinateness of polyhedral games," in [37], 195-198.

36. P. Wolfe, "A duality theorem for non-linear programming," Q. Appl. Math. XIX (1961), 239-244.

37. Linear Inequalities and Related Systems (H. W. Kuhn and A. W. Tucker, eds.), Ann. of Math. Study 38 (Princeton, 1956).

38. R. W. Cottle, "Symmetric dual quadratic programs," Operations Research Center, Research Report 19 (Univ. of Calif., Berkeley; May, 1962).

39. G. B. Dantzig, E. Eisenberg and R. W. Cottle, "Symmetric dual non-linear programs," Operations Research Center, Research Report 30 (Univ. of Calif., Berkeley; December, 1962).

40. P. Huard, "Dual programs," IBM J. Research & Devel. 6 (1962), 137-139.

41. O. L. Mangasarian, "Duality in non-linear programming," Q. Appl. Math. XX (1962), 300-302.

42. J.-J. Moreau, "Fonctions convexes en dualite" (multigraph Faculte des Sciences de Montpellier, Seminaires de Mathematiques, 1962).

43. J.-J. Moreau, "Applications 'prox'" (multigraph, Faculte des Sciences de Montpellier, Seminaires de Mathematiques, 1963).

44. J.-J. Moreau, "Decomposition orthogonale d'un espace hilbertien selon deux cônes mutuellement polaires," C. R. Acad. Sci. 255 (July, 1962), 238-240.

45. J.-J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," C. R. Acad. Sci. 255 (November, 1962), 2897-2899.

46. J.-J. Moreau, "Properties des applications 'prox'," C. R. Acad. Sci. 256 (January, 1963), 1069-1071.

47. J.-J. Moreau, "Les liaisions unilaterales et le principe de Gauss," C. R. Acad. Sci. (January, 1963), 871-874.

48. T. A. Dwyer, "Non-linear networks and boundary value problems," Q. Appl. Math. XIX (1961), 285-300.

49. C. Berge, "Sur l'equivalence du problème du transport généralisé et du problème des réseaux," C. R. Acad. Sci. 251 (1960), 324-326.

50. W. Millar, "On some general theorems for non-linear systems possessing resistance," Philosophical Mag. 42 (1951), 1150-1177.

51. G. J. Minty, "Monotone networks," Proc. Roy. Soc. London A, 257 (1960), 194-212.