

EXTENSION OF FENCHEL'S DUALITY THEOREM FOR CONVEX FUNCTIONS

BY R. T. ROCKAFELLAR

1. Introduction. Let E be a locally convex Hausdorff topological vector space over the real numbers R with dual E^* . Let f be a proper convex function on E , i.e. an everywhere-defined function with values in $(-\infty, +\infty]$, not identically $+\infty$, such that

$$(1.1) \quad f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \text{ if } x \in E, y \in E, 0 < \lambda < 1.$$

Let g be a proper concave function on E (i.e. $-g$ is proper convex). This paper is concerned with characterizing the solutions and the extremum in the following problem:

$$(I) \quad \text{minimize } f(x) - g(x) \text{ on } E.$$

Many constrained, as well as unconstrained, extremum problems can be represented in the model form (I), because the functions are allowed to be infinite-valued. For example, if D is a convex set in E and $g(x) = 0$ for $x \in D$, $g(x) = -\infty$ for $x \notin D$, then (I) is essentially the same as minimizing f on D .

Closely associated with (I) is a "dual" problem of similar type,

$$(II) \quad \text{maximize } g^*(x^*) - f^*(x^*) \text{ on } E^*,$$

where the concave function g^* and the convex function f^* are the *conjugates* [2, 4, 7] of f and g defined by

$$(1.2) \quad f^*(x^*) = \sup \{(x, x^*) - f(x)\},$$

$$(1.3) \quad g^*(x^*) = \inf \{(x, x^*) - g(x)\}$$

for each $x^* \in E^*$. It is immediate from (1.2) and (1.3) that

$$(1.4) \quad f(x) - g(x) \geq g^*(x^*) - f^*(x^*) \text{ for all } x \in E \text{ and } x^* \in E^*.$$

Problem (II) was first introduced (in the finite-dimensional case, and in a slightly different formulation) by Fenchel [5], who showed that (1.4) could often be strengthened to

$$(A) \quad \inf \{f(x) - g(x)\} = \max \{g^*(x^*) - f^*(x^*)\}.$$

Fenchel's duality theorem [5; 108] asserts, namely, that (A) is true when $E = R^n$, if the relative interiors of the convex sets

Received September 2, 1964. This work was supported in part by the Air Force Grant Air Force Office of Scientific Research. The author wishes to thank the Mathematics Institute of the University of Copenhagen for its hospitality during the spring and summer of 1964 while this research was being conducted.

$$(1.5) \quad C = \{x \mid f(x) < \infty\}, \quad D = \{x \mid g(x) > -\infty\},$$

have a point in common.

Although Fenchel's duality theorem is not generally valid for $E \neq R^n$ under the original hypothesis, we shall be able to show that the following version of it is true for spaces of arbitrary dimension.

THEOREM 1. *If either f or g is continuous at some point where both functions are finite, then (A) is true.*

This result will enable us to characterize the solutions of (I) using the sub-differentiability notion recently studied by Moreau [9]. We shall also apply it to some particular problems in the theory of conjugate convex functions, and obtain from it a generalization of the Kuhn-Tucker theorem for Lagrange multipliers.

2. Proof of Theorem 1. Let C and D be the convex sets in (1.5), and let $x_0 \in C \cap D$ be a point at which f is continuous. Then $x_0 \in \text{int } C$ (the interior of C) and

$$(2.1) \quad \infty > f(x_0) - g(x_0) \geq \inf \{f(x) - g(x)\} = \alpha.$$

Since (A) is trivial when the infimum α is $-\infty$ because of (1.4), we can assume $\alpha \in R$. In $E \oplus R$, the sets

$$\begin{aligned} C' &= \{ \langle x, \mu \rangle \mid x \in \text{int } C, \mu > f(x) \}, \\ D' &= \{ \langle x, \mu \rangle \mid x \in D, \mu \leq g(x) + \alpha \}, \end{aligned}$$

can have no common point. The convexity of f and the concavity of g imply that C' and D' are convex. Furthermore, C' is open in the product topology on $E \oplus R$, because f must be continuous at all points of the open set $\text{int } C$ when it is continuous at one such point x_0 (see [1; 92]). Hence C' and D' can be separated by some non-trivial closed hyperplane in $E \oplus R$ (see [1; 71]). The hyperplane cannot be "vertical", for otherwise its projection on E would separate the projections $\text{int } C$ and D of C' and D' , which is impossible because $x_0 \in D \cap (\text{int } C)$. The hyperplane must therefore be the graph of some continuous affine function on E . Thus there exists some $\bar{x}^* \in E^*$ and $\beta \in R$ such that

$$(2.2) \quad f(x) \geq \langle x, \bar{x}^* \rangle - \beta \quad \text{for all } x \in \text{int } C,$$

$$(2.3) \quad \langle x, \bar{x}^* \rangle - \beta \geq g(x) + \alpha \quad \text{for all } x \in D.$$

Since $g(x) = -\infty$ for $x \notin D$, (2.3) must be true for all x . Hence

$$(2.4) \quad \alpha + \beta \leq g^*(\bar{x}^*)$$

by (1.3). Inequality (2.2) is true similarly for $x \notin C$. If x belongs to C , but perhaps not to its interior, we have $\lambda x_0 + (1 - \lambda)x \in \text{int } C$ for $0 < \lambda < 1$ because $x_0 \in \text{int } C$, and hence

$$(2.5) \quad (\lambda x_0 + (1 - \lambda)x, \bar{x}^*) - \beta \leq f(\lambda x_0 + (1 - \lambda)x) \\ \leq \lambda f(x_0) + (1 - \lambda)f(x)$$

for $0 < \lambda < 1$ by (1.1). The outer inequality in (2.5) must therefore also hold for $\lambda = 0$. Thus inequality (2.2) holds for all x , so that

$$(2.6) \quad \beta \geq f^*(x^*)$$

by (1.2). But (2.4) and (2.6), combined with (1.4), say that

$$(2.7) \quad \alpha \leq g^*(\bar{x}^*) - f^*(\bar{x}^*) \leq \sup (g^* - f^*) \leq \inf (f - g).$$

This proves (A), because the outer terms in (2.7) are equal by definition.

3. Special cases. The hypothesis of Theorem 1 can be weakened somewhat when E is *tonnelé* (in particular, when E is a Banach space). For such spaces, namely, a convex function is continuous at a point x_0 if it is finite and lower semi-continuous (l.s.c.) on a neighborhood of x_0 (see [12, Corollary 7C]). Thus we have

COROLLARY 1. *If E is tonnelé, if f and g are l.s.c. and u.s.c. on E respectively, and if one of the convex sets in (1.5) contains an interior point of the other, then (A) is true.*

Especially interesting is the case where E is reflexive (and hence automatically *tonnelé*). The semi-continuity conditions then guarantee that the conjugates f^{**} and g^{**} of f^* and g^* on the bidual can be identified with f and g again (see [2] or [7]), so that problem (I) can in turn be viewed as the dual of problem (II).

COROLLARY 3. *Let E be a reflexive space, and let h be an l.s.c. convex function finite on all of E whose conjugate h^* is finite on all of E^* . Let K be any non-empty closed convex cone in E and let K^* be the negative of its polar in E^* , i.e.*

$$K^* = \{x^* \mid (x, x^*) \geq 0 \text{ for all } x \in K\}.$$

Then, for all $a \in E$ and $a^* \in E^*$, one has

$$(B) \quad \min_{x \in K} \{h(a+x) - (x, a^*)\} + \min_{x^* \in K^*} \{h^*(a^* + x^*) - (a, x^*)\} = (a, a^*).$$

Proof. Define f and g by

$$(3.2) \quad \begin{aligned} f(x) &= h(a+x) - (a+x, a^*), \\ g(x) &= 0 \quad \text{if } x \in K, \quad g(x) = -\infty \quad \text{if } x \notin K. \end{aligned}$$

Formulas (1.2) and (1.3) then yield

$$(3.3) \quad \begin{aligned} f^*(x^*) &= h^*(a^* + x^*) - (a, x^*), \\ g^*(x^*) &= 0 \quad \text{if } x^* \in K^*, \quad g^*(x^*) = -\infty \quad \text{if } x^* \notin K^*. \end{aligned}$$

Applying Corollary 2(c) to f and g proves (B).

If K is a subspace in (B), then K^* is of course its annihilator. If E is compatibly partially-ordered, one can take K to be the non-negative orthant; then K^* is the non-negative orthant of E^* in the dual ordering. An interesting pair of conjugate convex functions for which (B) is valid, when E is a reflexive Banach space, is

$$(3.4) \quad h(x) = (1/p) \|x\|^p, \quad h^*(x^*) = (1/q) \|x^*\|^q, \quad (1/p) + (1/q) = 1.$$

(See [9; 16-17].) If $E = R^n = E^*$, one can also take

$$(3.5) \quad h(x) = (1/2)x^T A x, \quad h^*(x^*) = (1/2)x^{*T} A^{-1} x^*,$$

where A is any positive definite symmetric real matrix.

4. Characterization of solutions. A subgradient of f at a point $x \in E$ is an $x^* \in E^*$ such that

$$(4.1) \quad f(y) \geq f(x) + (y-x, x^*) \quad \text{for all } y \in E.$$

The set of subgradients of f at x is denoted by $\partial f(x)$. If f has a gradient $\nabla f(x)$ at x in the sense of Gateaux (or Frechet), then $\partial f(x)$ is the singleton $\{\nabla f(x)\}$ (see [9]). If f is the indicator function of a convex set C in E , i.e. $f(x)$ is 0 on C and $+\infty$ outside of C , then $x^* \in \partial f(x)$ if and only if the linear function (\cdot, x^*) attains its maximum on C at x , and hence defines a (possibly trivial) supporting hyperplane to C at x . It follows from the definitions of f^* and $\partial f(x)$ that, in general,

$$(4.2) \quad x^* \in \partial f(x) \quad \text{if and only if} \quad f(x) + f^*(x^*) \leq (x, x^*).$$

Subgradients of the concave function g have analogous properties (with the defining inequality in (4.1) reversed).

THEOREM 2. *Assume that f and g satisfy some hypothesis, like that of Theorem 1, guaranteeing that (A) is true. Then \bar{x} is a point where $f - g$ achieves its minimum in (I), if and only if $\partial f(\bar{x})$ and $\partial g(\bar{x})$ have some \bar{x}^* in common. Moreover, such vectors \bar{x}^* are then precisely the points where $g^* - f^*$ achieves its maximum in (II).*

Proof. By (4.2) and its concave analog, the conditions $\bar{x}^* \in \partial f(\bar{x})$ and $\bar{x}^* \in \partial g(\bar{x})$ are equivalent to

$$(4.3) \quad f(\bar{x}) + f^*(\bar{x}^*) \leq (\bar{x}, \bar{x}^*) \leq g(\bar{x}) + g^*(\bar{x}^*).$$

But the opposite inequalities are always true by definition of f^* and g^* , so (4.3) is equivalent to

$$f(\bar{x}) - g(\bar{x}) = g^*(\bar{x}^*) - f^*(\bar{x}^*).$$

The conclusion of the theorem is now obvious from (A).

5. Application to sums of convex functions. Our next result is a contribution to the study of the conjugate and subgradients of a sum (see [2, 8, 9]).

THEOREM 3. *Let f_1 and f_2 be proper convex functions on E . Suppose there exists a point at which both functions are finite and at least one is continuous. Then, for all $x \in E$ and $x^* \in E^*$,*

$$(a) \quad (f_1 + f_2)^*(x^*) = \min \{f_1^*(x^* - z^*) + f_2^*(z^*) \mid z^* \in E^*\},$$

$$(b) \quad \partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x).$$

Proof. For any fixed $x^* \in E^*$, let $f(x) = f_2(x)$ and $g(x) = (x, x^*) - f_1(x)$. Then $f^*(z^*) = f_2^*(z^*)$, and $g^*(z^*) = -f_1^*(x^* - z^*)$ by (1.3). Our hypothesis guarantees via Theorem 1 that (A) is true for f and g , in other words

$$(5.1) \quad \inf \{f_1(x) + f_2(x) - (x, x^*) \mid x \in E\} = \max \{-f_1^*(x^* - z^*) - f_2^*(z^*) \mid z^* \in E^*\}.$$

This is just the negative of (a), because the left side of (5.1) is $-(f_1 + f_2)^*(x^*)$ by definition. Next we observe from (a) and (4.2) that $x^* \in \partial(f_1 + f_2)$ if and only if

$$(5.2) \quad f_1(x) + f_2(x) + f_1^*(x^* - z^*) + f_2^*(z^*) \leq (x, x^*) = (x, x^* - z^*) + (x, z^*)$$

for some $z^* \in E^*$. By definition of f_1^* and f_2^* , (5.2) is equivalent to

$$f_1(x) + f_1^*(x^* - z^*) \leq (x, x^* - z^*) \quad \text{and} \quad f_2(x) + f_2^*(z^*) \leq (x, z^*).$$

Using (4.2) again, we conclude that $x^* \in \partial(f_1 + f_2)(x)$ if and only if there exists some $z^* \in E^*$ such that $x^* - z^* \in \partial f_1(x)$ and $z^* \in \partial f_2(x)$. Thus (b) is true.

Remark 1. If $E = R^n$, Theorem 3 is true if one assumes, instead of continuity, that the relative interiors of the convex sets $C_1 = \{x \mid f_1(x) < \infty\}$ and

$C_2 = \{x \mid f_2(x) < \infty\}$ have a point in common. The proof in this case is the same as the one above, except that one applies the original version of Fenchel's duality theorem in place of Theorem 1.

Remark 2. By reversing the above proof, one can derive Theorem 1 from Theorem 3a and Theorem 2 from Theorem 3b. Indeed, as a general rule one can prove a version of Fenchel's duality theorem by first proving a theorem about conjugates of sums of convex functions. This approach could have been used here, because Theorem 3 can be deduced from the sum theorems of Brøndsted [2] or Moreau [8], if one uses the new theorem of Moreau [10] about the duality between inf-compactness and continuity in the Mackey topology and then shows that the result can be extended to the case where f_1 and f_2 are not everywhere lower semi-continuous. (In the lower semi-continuous case, Theorem 3b has already been deduced this way by Moreau in [10].) The approach we have taken, however, has the advantage that the arguments are self-contained and more elementary.

6. Application to Lagrange Multipliers. Let g_1, \dots, g_m be concave functions on E which are everywhere finite, continuous and Gateaux differentiable, and let

$$D = \{x \mid g_i(x) \geq 0, \quad i = 1, \dots, m\}.$$

Let f be a proper convex function on E . The following theorem characterizes the solutions to the problem of minimizing f on the (closed convex) set D .

THEOREM 4. *Suppose that f is finite at some point x satisfying*

$$(6.1) \quad g_i(x) > 0 \quad \text{for } i = 1, \dots, m.$$

Then \bar{x} is a point where f achieves its minimum on D , if and only if there exist real numbers $\bar{\lambda}_1, \dots, \bar{\lambda}_m$ (Lagrange multipliers) which along with \bar{x} satisfy

$$(C) \quad \begin{aligned} \bar{\lambda}_i &\geq 0, & g_i(\bar{x}) &\geq 0, & \bar{\lambda}_i g_i(\bar{x}) &= 0 \quad \text{for } i = 1, \dots, m, \\ \bar{\lambda}_1 \nabla g_1(\bar{x}) &+ \dots + \bar{\lambda}_m \nabla g_m(\bar{x}) &\in &\partial f(\bar{x}). \end{aligned}$$

Proof. Let $g = h_1 + \dots + h_m$, where the concave functions h_i are defined by

$$(6.2) \quad h_i(x) = 0 \quad \text{if } g_i(x) \geq 0, \quad h_i(x) = -\infty \quad \text{if } g_i(x) < 0.$$

Minimizing f on D is the same as minimizing $f - g$ on E . Any solution of (6.1) where f is finite is a point where the functions g, h_1, \dots, h_m are all finite and continuous. Hence, by Theorem 2 and by induction on the concave analog of Theorem 3(b), f achieves its minimum on D at \bar{x} if and only if $\partial f(\bar{x})$ contains an element of $\partial h_1(\bar{x}) + \dots + \partial h_m(\bar{x})$. To complete the proof, one need only show that

$$(6.3) \quad \begin{aligned} \partial h_i(\bar{x}) &= \{0\} \quad \text{if } g_i(\bar{x}) > 0, & \partial h_i(\bar{x}) &= \emptyset \quad \text{if } g_i(\bar{x}) < 0, \\ \partial h_i(\bar{x}) &= \{\lambda_i \nabla g_i(\bar{x}) \mid \lambda_i \geq 0\} \quad \text{if } g_i(\bar{x}) = 0. \end{aligned}$$

By definition, $x^* \in \partial h_i(\bar{x})$ if and only if

$$h_i(x) \leq h_i(\bar{x}) + (x - \bar{x}, x^*) \quad \text{for all } x \in E.$$

This condition says that the linear function (\cdot, x^*) achieves its minimum on $D_i = \{x \mid g_i(x) \geq 0\}$ at \bar{x} . Since $\sup g_i > 0$ by the hypothesis, (6.3) follows from this by an elementary argument using the continuity and differentiability of g_i . This proves Theorem 4.

Theorem 4 is "open-ended," in the sense that one may incorporate further constraints into f . For example, suppose $f = f_0 + h$ where h is proper convex, and f_0 is the convex function which is 0 at all solutions of a certain system of linear inequalities

$$(6.4) \quad (x, a_j^*) - \alpha_j \geq 0, \quad j = 1, \dots, k,$$

and is $+\infty$ elsewhere. If h is finite and continuous at some solution of (6.4), we can substitute $\partial f_0(\bar{x}) + \partial h(\bar{x})$ for $\partial f(\bar{x})$ in (C) by Theorem 3(b). In other words, h then attains its minimum on D at \bar{x} subject to the additional constraints (6.4), if and only if

$$\bar{\lambda}_1 \nabla g_1(\bar{x}) + \dots + \bar{\lambda}_m \nabla g_m(\bar{x}) - x_0^* \in \partial h(\bar{x}),$$

where \bar{x} and $\bar{\lambda}_i$ satisfy the first half of (C) and x_0^* is some element of $\partial f_0(\bar{x})$. Moreover, $x_0^* \in \partial f_0(\bar{x})$ if and only if the linear function (\cdot, x_0^*) attains its maximum subject to (6.4) at \bar{x} . This is equivalent to the existence of real numbers μ_1, \dots, μ_k which along with \bar{x} satisfy

$$\begin{aligned} \mu_i \geq 0, \quad (\bar{x}, a_i^*) - \alpha_i \geq 0, \quad \mu_i [(\bar{x}, a_i^*) - \alpha_i] &= 0, \\ x_0^* &= -(\mu_1 a_1^* + \dots + \mu_k a_k^*). \end{aligned}$$

(See [3; 108]). Combining this condition with the earlier one, and changing notation slightly, we get the following version of Theorem 4.

THEOREM 4'. *Suppose that g_i is affine for $i = 1, \dots, k$, and that f is finite and continuous at some point x satisfying*

$$(6.1') \quad g_i(x) \geq 0 \quad \text{for } i = 1, \dots, k, \quad g_i(x) > 0 \quad \text{for } i = k+1, \dots, m.$$

Then the conclusion of Theorem 4 is valid.

Theorem 4' can be applied to systems of constraints containing finitely many linear equations, using the standard trick of representing each equation by a pair of inequalities.

When f is finite and Gateaux differentiable, condition (C) says that $\langle \bar{x}, \bar{\lambda}_1, \dots, \bar{\lambda}_m \rangle$ is a saddle-point of the Lagrangian function

$$L(x; \lambda_1, \dots, \lambda_m) = f(x) - \lambda_1 g_1(x) - \dots - \lambda_m g_m(x),$$

minimizing for $x \in E$ and maximizing for $\langle \lambda_1, \dots, \lambda_m \geq 0$ in R^m . Condition (C) was developed in this case for $E = R^n$ by Kuhn and Tucker [6].

We should like finally to describe a dual to the problem of minimizing f on D . We have already indicated, in the proof of Theorem 4, that this problem can be represented in the form (I) with g as the sum of the functions h_i in (6.2). To see the nature of the corresponding problem (II), we must calculate the conjugate of g . Under the hypothesis of Theorem 4, there exists a point at which all the h_i are continuous. In this event

$$(6.5) \quad \begin{aligned} g^*(x^*) &= (h_1 + \cdots + h_m)^*(x^*) \\ &= \max \left\{ \sum_{i=1}^m h_i^*(Z_i^*) \mid Z_i^* \in E^*, \sum_{i=1}^m Z_i^* = x^* \right\} \end{aligned}$$

by induction on the concave analog of Theorem 3(a). Furthermore, according to formulas developed by the author elsewhere (see [12, Corollary 4B and Corollary 3C(d)]), we have

$$(6.6) \quad h_i^*(Z_i^*) = \max \{ \lambda_i g_i^*((1/\lambda_i)Z_i^*) \mid \lambda_i > 0 \} \quad \text{if } Z_i^* \neq 0, \quad h_i^*(0) = 0.$$

(The fact that g_i is finite everywhere, and the assumption in Theorem 4 that $\sup g_i > 0$, are both required for this formula to be true.) Setting $Z_i^* = \lambda_i x_i^*$ and

$$D_i^* = \{ x_i^* \in E^* \mid g_i^*(x_i^*) > -\infty \},$$

we can combine (6.5) and (6.6) as

$$(6.7) \quad g^*(x^*) = \max \left\{ \sum \lambda_i g_i^*(x_i^*) \mid x_i^* \in D_i^*, \lambda_i \geq 0, \sum \lambda_i x_i^* = x^* \right\}$$

with the maximum taken to be $-\infty$ when the constraints cannot be satisfied. Since, under the hypothesis of Theorem 4, f is finite at a point where g is finite and continuous, we can now apply Theorem 1 to get the following result.

THEOREM 5. *If the hypothesis of Theorem 4 is satisfied, then*

$$(A') \quad \begin{aligned} \inf \{ f(x) \mid x \in E, g_i(x) \geq 0 \text{ for } i = 1, \dots, m \} \\ = \max \left\{ \sum_{i=1}^m \lambda_i g_i^*(x_i^*) - f^* \left(\sum_{i=1}^m \lambda_i x_i^* \right) \mid \lambda_i \geq 0, x_i^* \in D_i^* \right\}. \end{aligned}$$

It could be shown similarly that (A') is true under the hypothesis of Theorem 4', but the argument will not be given. If f achieves its minimum on D at some point \bar{x} , it turns out that the maximum in (A') occurs when the λ_i are the multipliers in (C) and $x_i^* = \nabla g_i(\bar{x})$. Theorem 5 is valid, however, even if the g_i are not differentiable, since this assumption was not used in its proof.

REFERENCES

1. N. BOURBAKI, *Espaces Vectoriels Topologiques*, Livre V, Chapters I and II, Paris, 1953.
2. A. BRØNSTED, *Conjugate convex functions in topological vector spaces*, Matematiskfysiske Meddelelser udgivet af Det Kongelige Danske Videnskabernes Selskab, Copenhagen, vol. 34(1964).
3. K. FAN, *On systems of linear inequalities*, Annals of Mathematics Studies, vol. 38(1956), pp. 99-155.

4. W. FENCHEL, *On conjugate convex functions*, Canadian Journal of Mathematics, vol. 1(1949), pp. 73-77.
5. W. FENCHEL, *Convex cones, sets and functions*, lecture notes, Princeton, 1953.
6. H. KUHN AND A. W. TUCKER, *Non-linear programming*, Proceedings of the second Berkeley Symposium on Mathematical Statistical Problems, University of California Press, 1951, pp. 481-492.
7. J. J. MOREAU, *Fonctions convexes en dualité*, Faculté des Sciences de Montpellier, Séminaires de Mathématiques, 1962, (Multigraph).
8. J. J. MOREAU, *Inf-convolution*, Faculté des Sciences de Montpellier, Séminaires de Mathématiques, 1963, (Multigraph).
9. J. J. MOREAU, *Etude locale d'une fonctionnelle convexe*, Faculté des Sciences de Montpellier, Séminaires de Mathématiques, 1963, (Multigraph).
10. J. J. MOREAU, *Sur la polaire d'une fonction semi-continue supérieurement*, Comptes Rendus Academie des Sciences, vol. 258(1964).
11. R. T. ROCKAFELLAR, *Duality theorems for convex functions*, Bulletin of the American Mathematical Society, vol. 70(1964), pp. 189-192.
12. R. T. ROCKAFELLAR, *Level sets and continuity of conjugate convex functions*, to appear in Transactions of the American Mathematical Society, 1966.

PRINCETON UNIVERSITY