24)

Duplicate [22]

Reprint from

Lecture Notes in Operations Research and Mathematical Economics, Vol. 11

Mathematical Systems Theory and Economics I

Springer-Verlag Berlin Heidelberg New York 1969

Printed in Germany. Not for sale

R. T. Rockafellar

Convex Functions and Duality in Optimization Problems and Dynamics

Everyone is aware of the importance of convex sets in the study of optimization problems. Much of the modern theory of convex functions is less well known, however, and for this reason has not sufficiently been exploited. This is true especially of Fenchel's theory of conjugacy [11], which ought to be made the vehicle for all results involving duality. Fenchel's theory and some of its consequences will be described below.

Elementary facts about convex functions, their conjugates and their continuity and differentiability properties are set forth in §1, 2, and 3. Proofs of these facts in the finite-dimensional case may be found in [12] and the forthcoming book [39]. For the extensions to infinite dimensional cases, see the papers of Brøndsted, Moreau and Rockafellar listed in the bibliography.

The approach which we take to convex programs has been suggested by a paper of Gale [14]. It leads to a concept of "generalized convex program" in §4 for which an extensive duality theory is possible. This duality theory, explained in §5, is new and is being announced here for the first time.

---

*Preparation of this manuscript was supported in part by U.S. Air Force Grant AF-AFOSR-1202-67 at the Department of Mathematics, University of Washington, Seattle.

The details are contained in the author's book [39].

In §6, certain applications of conjugate convex functions to control theory, the calculus of variations and Hamiltonian dynamics are discussed. Publication of the proofs of the new results announced in this section has not yet been fixed.

1. Definition of a Convex Function.

Let E be an arbitrary vector space over the real numbers R. Let C be a convex set in E. According to the classical definition, a function f from C to R is convex if

$$f(1-\lambda)x+\lambda y) \leq (1-\lambda)f(x) + \lambda f(y) \; , \; 0 < \lambda < 1 \; ,$$

for every x and y in C. The geometric meaning of this definition is that the set of points in $E \oplus R$ lying "on or above" the graph of f is a convex set.

It is convenient to extend a given convex function f on C to all of E by defining $f(x) = +\infty$ for $x \notin C$. In general, a function f on all of E whose values are real numbers or $+\infty$ or $-\infty$ is said to be convex if the set

$$\text{epi } f = \{(x, \mu) \,|\, x \in E \, , \, \mu \in R \, , \, \mu \geq f(x)\},$$

which is called the epigraph of f, is convex as a subset of the vector space $E \oplus R$. The projection of epi f on E, which is the set

$$\text{dom } f = \{x \,|\, f(x) < \infty \},$$

is then convex too. It is called the effective domain of f. The convex functions f on E obtained by extending finite convex functions on non-empty convex subsets of E by $+\infty$ are precisely those such that $f(x) < +\infty$ for at least one x and $f(x) > -\infty$ for every x. Such an f is said to be a proper convex function. Improper convex functions are not really of interest in themselves, but they are technically useful in the general theory.

An important example of a convex function is the indicator function $\delta(\cdot|C)$ of a convex set C, which is defined by

$$\delta(x|C) = \begin{cases} 0 & \text{if } x \in C \, , \\ +\infty & \text{if } x \notin C \, . \end{cases}$$

If f_o is a finite convex function on E, the convex function f given by

$$f(x) = f_o(x) + \delta(x|C)$$

corresponds in a certain sense to the restriction of f_o to C. Observe that minimizing f_o over C is equivalent to minimizing f over E. By this device, constrained minimization problems can be represented formally as unconstrained problems. For example, in the case of minimizing $f_o(x)$ subject to $f_1(x) \leq 0, \ldots, f_m(x) \leq 0$, where $f_1, \ldots, f_m$ are finite convex functions on E (the type of problem we shall refer to as an ordinary convex program), one takes C to be the intersection of the convex sets

$$C_i = \{x \mid f_i(x) \leq 0\}, \quad i = 1, \ldots, m.$$

It should be kept in mind below that such cases are included when we speak simply of minimizing a convex function f over E.

There are many useful operations which can be performed in the collection of all convex functions on E. If f and g are proper convex functions, then $f + g$ is convex. Of course, $f + g$ might be identically $+\infty$ (and hence improper), because the set

$$\mathrm{dom}(f+g) = \mathrm{dom}\, f \cap \mathrm{dom}\, g$$

might be empty. The reason we ask that f and g be proper when forming $f + g$, is that this is a simple way to ensure that $f(x) + g(x)$ is not $\infty - \infty$. The combination $\infty - \infty$ is undefined, like division by zero, and is carefully avoided.

If f is a convex function and $\lambda > 0$, then λf is convex.

Given a collection of convex functions $\{f_i \mid i \in I\}$ on E, where I is an arbitrary index set, the pointwise supremum f of the collection, given by

$$f(x) = \sup \{f_i(x) \mid i \in I\},$$

is a convex function. The convexity of f is obvious from the fact that $\mathrm{epi}\, f$ is the intersection of the convex sets $\mathrm{epi}\, f_i$. The convex function which is the pointwise supremum of the collection of all convex functions g such that $g \leq f_i$ for every $i \in I$ is called the convex hull of $\{f_i \mid i \in I\}$. Its epigraph is essentially the convex hull of the collection of convex sets $\mathrm{epi}\, f_i$.

One of the most interesting operations is infimal convolution. The infimal convolute $f \square g$ of two proper convex functions f and g is defined by

$$(f \square g) = \inf_y \{f(x-y) + g(y)\}.$$

The epigraph of $f \square g$ is obtained essentially by vector addition of $\mathrm{epi}\, f$ and $\mathrm{epi}\, g$ in $E \oplus R$, whence its convexity. Infimal convolution is a commutative, associative operation.

As an example of infimal convolution, consider the case where $E = R^n$, $f(x) = |x|$ (the Euclideen norm) and $g(x) = \delta(x|C)$, where C is a convex set. Then

$$(f \square g)(x) = \inf_y \{|x-y| + \delta(y|C)\} = \inf \{|x-y| \,|\, y \in C\} .$$

This convex function gives the distance of x from C.

For another example, let L be any subspace of E and let $g(x) = \delta(x|L)$. For any convex function f on E, we have

$$(f \square g)(x) = \inf \{f(x+z) \mid z \in L\} .$$

Thus $f \square g$ gives the infimum of f over the affine set $x + L$ as a function of the translation x.

These examples illustrate the fact that, in certain minimization problems depending on parameters, the infimum is a convex function of the parameters. A very important case concerns perturbations of an ordinary convex program. For each vector

$$u = (\upsilon_1, \ldots, \upsilon_m) \in R^m , \text{ let } p(u) = p(\upsilon_1, \ldots, \upsilon_m)$$

denote the infimum of f_o subject to the constraints $f_i(x) \leq \upsilon_i$, $i = 1, \ldots, m$ (where f_i is a finite convex function on E for $i = 0, 1, \ldots, m$). The given problem corresponds to $u = 0$, and other values of u are conceived of as perturbations away from 0. We shall call p the <u>perturbation function</u> for the program. Lagrange multipliers, as we shall explain in §4, can be studied in terms of the differentiability properties of p at $u = 0$. The fundamental and easily proved fact about p is that p is a convex function on R^m. Note that dom p consists of the vectors u such that the inequality system

$$f_1(x) \leq \upsilon_1, \ldots, f_m(x) \leq \upsilon_m ,$$

is satisfied by at least one $x \in E$, whereas the interior of dom p consists of the vectors u such that the inequality system

$$f_1(x) < \upsilon_1, \ldots, f_m(x) < \upsilon_m ,$$

is satisfied by at least one $x \in E$. If $0 \in$ dom p, the program is said to be <u>consistent</u>, and if $0 \in$ int (dom p) it is said to be <u>strictly consistent</u>.

2. Closures and Conjugates.

Let E^* be a real vector space in duality with E with respect to a certain bilinear form $\langle\cdot,\cdot\rangle$, and let E and E^* be provided with locally convex topologies compatible with this duality. Thus E and E^* are henceforth to be locally convex (Hausdorff) topological vector spaces, and $\langle x,x^*\rangle$ is to be a (separately) continuous function of $x \in E$ and $x^* \in E^*$, such that the continuous linear functions on E are the functions of the form $x \to \langle x,x^*\rangle$ and the continuous linear functions on E^* are the functions of the form $x^* \to \langle x,x^*\rangle$. In particular, of course, one may take $E = E^* = R^n$ with the ordinary topology and $\langle x,x^*\rangle$ as the ordinary inner product of two numerical vectors x and x^*.

Let f be any convex function on E. The pointwise supremum of the collection of all the continuous affine functions

$$h(x) = \langle x,x^*\rangle - \mu^* , \quad x^* \in E^* , \quad \mu^* \in R ,$$

such that $h \leq f$ is called the <u>closure</u> of f and is denoted by $\text{cl } f$.

The epigraph of $\text{cl } f$ is the intersection of the "non-vertical" closed half-spaces in $E \oplus R$ containing the convex set $\text{epi } f$, and this is the same as the closure of $\text{epi } f$ if f is proper and E is finite-dimensional (Fenchel [11]), or if f is bounded from below in a neighborhood of some point of $\text{dom } f$ (Brøndsted [2]).

A function f from E to $[-\infty, +\infty]$ is said to be <u>lower semi-continuous</u> if the sets $\{x \mid f(x) \leq \mu\}$, $\mu \in R$, are all closed. This condition is actually equivalent to the epigraph of f being a closed set in $E \oplus R$. Thus a proper convex function is closed if and only if it is lower semi-continuous.

The nature of the closure operation is seen clearly from the following example. Let C be a closed (Euclidean) disk in R^2. Let $f(x) = 0$ for interior points x of C, $f(x) = +\infty$ for points $x \notin C$, and assign arbitrary non-negative values to $f(x)$ at boundary points of C. Then f is a proper convex function on R^2. One obtains the epigraph of $\text{cl } f$ in this case by closing the epigraph of f, and this amounts to redefining $f(x)$ to be 0 on the boundary of C.

It can be shown that for any proper convex function f, if $\text{epi } f$ has a non-empty interior, or if E is finite-dimensional, $\text{cl } f$ agrees with f everywhere except perhaps on the boundary of $\text{dom } f$. The closure operation may thus be regarded as a harmless regularization procedure for making a given convex function lower semi-continuous.

Continuity is almost as easy to come by.

THEOREM 1. *Let f be a convex function on E. Suppose that epi f has a non-empty interior, or that E is finite-dimensional, or that epi f is closed and E is a Banach space (or a tonnelé space), or that there exists a point at which f is finite and continuous. Then f is necessarily continuous at all points other than boundary points of dom f, and in particular f is continuous on any open set on which it is finite.*

This theorem is classical and well-known, except for the case where f is closed and E is a Banach space, which is due to Brøndsted [2]. (The extension to tonnelé spaces, and hence to all non-Banach reflexive spaces, such as those in the theory of distributions, is due to the author [29].)

The importance of Theorem 1 in optimization problems stems from fact that such problems often give rise to convex functions in a manner which should not in general be expected to ensure continuity. For example, the perturbation function p of a classical convex program is a convex function on R^m, and if the program is strictly consistent the interior of the effective domain of p contains the origin. It follows then from Theorem 1 that the infimum $p(u)$ is a continuous function of the perturbation u in some neighborhood of $u = 0$.

Almost all of the rich duality which pervades the theory of convex functions flows from Fenchel's notion of conjugacy. Let f be a closed convex function on E. On the one hand, we can describe f in terms of the collection of all points $(x,\mu) \varepsilon E \oplus R$ such that $\mu \geq f(x)$, i.e. the epigraph of f. But there is also a dual description. Since f is the pointwise supremum of the collection of all continuous affine functions h such that $h \leq f$ (by definition of f being closed), we can describe f in terms of the points $(x^*,\mu^*) \varepsilon E^* \oplus R$ such that

$$\langle x,x^* \rangle - \mu^* \leq f(x) \ ,\forall\, x \,\varepsilon\, E .$$

What Fenchel noticed was that this collection of points (x^*,μ^*) forms the epigraph of a certain function f^* on E^*,

$$f^*(x^*) = \sup_x \{\langle x,x^* \rangle - f(x)\} .$$

This f^* is called the conjugate of f. The formula expresses f^* as the pointwise supremum of a collection of continuous affine functions of x^*, so f^* is a closed convex function on E^*. The conjugate f^{**} of f is in turn given by

$$f^{**}(x) = \sup_{x^*} \{\langle x,x^*\rangle - f^*(x^*)\} ,$$

i.e. f^{**} is the pointwise supremum of the continuous affine functions $\langle\cdot,x^*\rangle - \mu^*$ on E such that $\mu^* \geq f^*(x^*)$. But this supremum is f, by the definition of f^*. Thus $f^{**} = f$. *Conjugacy is therefore a symmetric one-to-one correspondence between the closed convex functions on* E *and the closed convex functions on* E^*. Of course, the conjugate of a convex function f which is not closed may be defined by the same formula. One then has $f^* = (\mathrm{cl}\, f)^*$, and consequently $f^{**} = \mathrm{cl}\, f$.

For example, if f is the indicator of a convex set C in E, the conjugate of f is given by

$$f^*(x^*) = \sup_x \{\langle x,x^*\rangle - \delta(x|C)\} = \sup \{\langle x,x^*\rangle | x \in C\} .$$

This function is known as the *support function* of C, since it may be used to describe all the (closed) supporting hyperplanes to C. Note that, in general, f^* is closely related to the support function of the convex set epi f in $E \oplus R$, in that $f^*(x^*)$ gives the supremum over epi f of the linear function

$$(x,\mu) \to \langle x,x^*\rangle + \alpha\mu$$

in the normalized case where $\alpha = -1$.

Another case of conjugacy, which we would like to mention for its beauty, occurs when $E = R^n$ and

$$f(x) = (1/p)\langle x,Qx\rangle^{p/2} , \quad 1 < p < \infty ,$$

Q being a symmetric $n \times n$ positive definite matrix. Then f is a closed proper convex function and

$$f^*(x^*) = (1/q)\langle Q^{-1}x^*,x^*\rangle^{q/2} , \quad 1 < q < \infty ,$$

where $(1/p) + (1/q) = 1$.

3. *Directional Derivatives and Subgradients.*

Let f be a convex function on E and let x be a point where f is finite. It can be seen that the one-sided directional derivative

$$f'(x;y) = \lim_{\lambda \downarrow 0} [f(x+\lambda y) - f(x)]/\lambda$$

exists for every $y \in E$ ($+\infty$ and $-\infty$ being allowed as limits), and $f'(x;y)$ is a convex function of y which is positively homogeneous (of degree one). If $f'(x;y)$ is actually a continuous linear function of y, i.e. if there exists a vector $x^* \varepsilon E^*$ such that

$$f'(x;y) = \langle y, x^* \rangle \ , \forall y \ \varepsilon \ E \ ,$$

then f is said to be differentiable at x in the sense of Gâteaux, and x^* is said to be the *gradient* of f at x and is denoted by $\nabla f(x)$. It is a classical theorem about convex functions that, if E is finite-dimensional, Gâteaux differentiability is equivalent to differentiability at x in the sense of Fréchet. Fréchet differentiability requires that

$$\lim_{z \to x} \frac{f(z) - f(x) - \langle x^*, z-x \rangle}{|z-x|} = 0 \ .$$

A more general concept of gradient can be exploited by convexity methods. A vector $x^* \varepsilon E^*$ is called a *subgradient* of f at x if

$$f(z) \geq f(x) + \langle z-x, x^* \rangle \ , \forall x \ \varepsilon \ E \ .$$

This inequality means that graph of the affine function

$$h(z) = f(x) + \langle z-x, x^* \rangle$$

is a supporting hyperplane to the convex set epi f at the point $(x, f(x))$. There may be more than one subgradient at x or none at all. At all events, the subgradients x^* at x form a closed convex set in E^* which is denoted by $\partial f(x)$. If $\partial f(x) \neq \emptyset$, f is said to be *subdifferentiable* at x. Theorems about subdifferentiability are easily deduced from well-known theorems about the existence of supporting hyperplanes.

THEOREM 2. *Let f be a convex function on E, and let x be a point where f is finite and continuous (cf. Theorem 1). Then f is subdifferentiable at x, and*

$$f'(x;y) = \sup \{ \langle y, x^* \rangle \mid x^* \ \varepsilon \ \partial f(x) \} < \infty \ , \forall y \ \varepsilon \ E \ .$$

The set $\partial f(x)$ consists of a unique x^* if and only if f is Gâteaux differentiable at x, in which case $x^* = \nabla f(x)$.

THEOREM 3. *Let f be a convex function on E, and let x be a point where f is finite. Then f fails to be subdifferentiable at x if and only if the directional derivative function $f'(x;\cdot)$ is unbounded from below in every neighborhood of the origin of E. If E is finite-dimensional, this condition actually implies the existence of a vector y such that $f'(x;y) = -\infty$.*

The value of these theorems can easily be appreciated in connection with Lagrange multipliers. Real numbers $\lambda_1, \dots, \lambda_m$ are said to be Lagrange multipliers for a given ordinary convex program if $\lambda_i \geq 0$ for $i = 1, \dots, m$ and the infimum of the convex function $f_o + \lambda_1 f_1 + \dots + \lambda_m f_m$ on E is the same as the infimum of f_o subject to $f_i(x) \leq 0$, $i = 1, \dots, m$. (We assume in what follows that the latter infimum, which is $p(0, \dots, 0)$, is finite.) It is easy to see that, in terms of the perturbation function p for the program, this condition is equivalent to having

$$p(0, \dots, 0) \leq p(\upsilon_1, \dots, \upsilon_m) + \lambda_1 \upsilon_1 + \dots + \lambda_m \upsilon_m , \forall \upsilon_1, \dots, \upsilon_m .$$

Setting $u^* = (\lambda_1, \dots, \lambda_m)$, $u = (\upsilon_1, \dots, \upsilon_m)$, we can write this as

$$p(u) \geq p(0) + \langle u, -u^* \rangle , \forall u \in R^m .$$

But this means that $-u^* \in \partial p(0)$. Thus $u^* \in R^m$ is a Lagrange multiplier vector if and only if $-u^*$ is a subgradient of p at 0 .

It follows that all questions about the existence or interpretation of Lagrange multipliers in an ordinary convex program correspond to questions about the subgradients of a convex function on R^m . According to Theorem 2, a Lagrange multiplier vector exists if the program is strictly consistent, since then $0 \in \text{int} (\text{dom } p)$ and p is continuous at 0 . Moreover, in this case Lagrange multiplier vectors completely describe the directional derivatives of p at 0 , i.e. the rates of change of the infimum in the progam with respect to different directions of perturbation. The Lagrange multipliers λ_i are unique if and only if p is actually differentiable at 0 , in which case one has

$$\lambda_i = - \frac{\partial p}{\partial \upsilon i} \Big|_{u = 0} \quad i = 1, \dots, m .$$

In general, by Theorem 3, Lagrange multipliers fail to exist if and only if there exists a vector with respect to which p has directional derivative $-\infty$ at the origin. In this circumstance, the program is highly unstable, in the sense that there is a direction of perturbation in which the infimum drops off infinitely steeply. Thus, apart from this unstable case, Lagrange multiplier vectors always exist.

Lagrange multipliers can also be interpreted as "equilibrium prices". Let us think of the given problem as one of minimum cost. If we could perturb the problem by altering the constraints from $f_i(x) \leq 0$ to $f_i(x) \leq \upsilon_i$ for a certain choice of $\upsilon_1, \dots, \upsilon_m$, we might be able to achieve a lower cost. Suppose perturbations can be bought at prices λ_i per unit of variable υ_i . The minimum cost in the perturbed problem, plus the cost of this perturbation, is then

$$p(\upsilon_1,\ldots,\upsilon_m) + \lambda_1\upsilon_1 +\ldots+ \lambda_m\upsilon_m .$$

The perturbation $u = (\upsilon_1,\ldots,\upsilon_m)$ is worth buying only if this total cost is less than the minimum cost in the unperturbed problem, which is the amount $p(0,\ldots,0)$. Thus, according to the analysis above, Lagrange multipliers are precisely the prices λ_i which have the property that no perturbation is worth buying. For such prices there is an equilibrium, in the sense that the incentive for perturbation is neutralized.

The subgradients of a closed convex function f and its conjugate f^* are related in a simple way: one has $x^* \in \partial f(x)$ if and only if $x \in \partial f^*(x^*)$. In other words, the multivalued mappings ∂f and ∂f^* are the inverses of each other.

Fenchel's conjugacy correspondence can be regarded as a generalization of the classical Legendre transformation, which is so important in Hamiltonian dynamics and the calculus of variations, and this is the idea behind the applications to be described in §6. Let f be a differentiable convex function on R^n . By definition, for each $x^* \in R^n$, $f^*(x^*)$ is the supremum of $\langle x,x^*\rangle - f(x)$ as a function of x . This is a differentiable function, and its supremum is attained (if at all) at the points where the gradient, which is $x^* - \nabla f(x)$, vanishes. Thus

$$f^*(x^*) = \langle x,x^*\rangle - f(x) \text{ if and only if } \nabla f(x) = x^* .$$

If the gradient mapping ∇f is actually one-to-one from R^n onto R^n , i.e. if the equation $\nabla f(x) = x^*$ has a unique solution x for each x^* , we have

$$f^*(x^*) = \langle (\nabla f)^{-1}(x^*),\, x^*\rangle - f((\nabla f)^{-1}(x^*)), \forall x^* .$$

This is the function which is called the <u>Legendre transform</u> of f . The general theory of conjugate convex functions implies that the Legendre transform, if it exists, is another differentiable convex function whose Legendre transform is turn f . The gradient mapping of the Legendre transform of f is $(\nabla f)^{-1}$.

Observe that, even if ∇f is not one-to-one, it is still true that

$$f^*(\nabla f(x)) = \langle x, \nabla f(x)\rangle - f(x) \ , \forall x .$$

The expression on the right is a common one in the convex programming literature. It is generally not a convex function of x , of course, and it only expresses the values of the conjugate function on the range of the mapping ∇f .

We refer the reader to [36] and [39] for more details about the relationship between conjugacy and the Legendre transformation.

4. Generalized Convex Programs.

As we have seen, in an ordinary convex program one is concerned with minimizing a certain convex function (possibly infinity-valued) over E. This function is embedded in a natural way in a whole class of functions depending on a parameter vector u, which may be regarded as a perturbation. Lagrange multipliers evaluate the effects of the given class of perturbations. We shall take these notions as the starting point for a much more general theory of convex programs.

Let D and D^* be two more topological real vector spaces paired together as described at the beginning of §2, just like E and E^*.

By a convex bifunction from D to E, we shall mean a correspondence F which assigns to each $u \in D$ a convex function Fu on E in such a way that the function

$$(u,x) \to (Fu)(x) ,$$

which we call the graph function of F, is convex on the space $D \oplus E$. Obviously each convex function on $D \oplus E$ is the graph function of one and only one convex bifunction F. The bifunction is the graph function broken down into two stages:

$$F{:}u \to Fu : x \to (Fu)(x) .$$

The reason why we introduce the concept of "bifunction", instead of dealing directly with functions on $D \oplus E$ is the same as the reason why one speaks of multivalued mappings from D to E instead of the corresponding subsets of $D \oplus E$ (their graphs). We want to stress certain analogies with ordinary mappings, particularly linear transformations.

Note that if A is a linear transformation from D to E and

$$(Fu)(x) = \delta(x\,|\,Au) = \begin{cases} 0 & \text{if } x = Au , \\ +\infty & \text{if } x \neq Au , \end{cases}$$

then F is a convex bifunction. We call this F the convex indicator bifunction of A.

By the convex program (P) associated with a convex bifunction F from D to E, we shall mean the problem of minimizing the convex function F0 in E, as viewed in the context of the whole class of problems in which, for different choices of u, Fu is minimized over E. The vector u is regarded as a perturbation. Thus a general convex program (P), in our conception, is not just an isolated minimization problem but a minimization problem with a specified class of perturbations. The perturbation function for (P) is the function $p = \inf F$ on D defined by

$$(\inf F)(u) = \inf(Fu) = \inf_x (Fu)(x) .$$

A vector $u^* \in D^*$ is called a Lagrange multiplier vector for (P) if $(\inf F)(O)$ (the infimum in (P)) is finite and

$$(\inf F)(O) \le (\inf F)(u) + \langle u,u^* \rangle , \forall u \in D .$$

(Such a u^* can be interpreted as an equilibrium price vector for perturbations.) A vector $x \in E$ is called an optimal solution to (P) , of course, if the infimum of FO is finite and attained at x .

An ordinary convex program corresponds to a bifunction F from $D = R^m$ to E of the form

$$(Fu)(x) = \begin{cases} f_o(x) & \text{if } f_1(x) \le u_1, \ldots, f_m(x) \le u_m , \\ +\infty & \text{if not,} \end{cases}$$

where $f_o, f_1, \ldots, f_m$ are finite convex functions. It is easily verified that such an F is a convex bifunction.

By virtue of the following theorem, all the results which have been described for Lagrange multipliers of ordinary convex programs can immediately be extended to general convex programs.

THEOREM 4. The perturbation function inf F of a general convex program (P) is a convex function on D . When $(\inf F)(O)$ is finite, the Lagrange multiplier vectors for (P) are precisely the vectors u^* such that $-u^*$ is a subgradient of inf F at O .

It follows from Theorem 3, for example, that a Lagrange multiplier vector exists for (P) as long as (P) is reasonably stable under the given perturbations (and the infimum in (P) is finite).

Theorem 2 can be applied to inf F to get results about Lagrange multiplier vectors for (P) when inf F is continuous at O . Here is one criterion for continuity.

THEOREM 5. Suppose there exists a vector $x_o \in E$ such that, for some real number α and some open neighborhood U of O in D , one has

$$(Fu)(x_o) \le \alpha , \forall u \in U .$$

Suppose also that $\inf F)(O)$ is finite. Then inf F is finite and continuous at every point $u \in U$, and in particular continuous at $u = O$.

Proof. The condition implies that the epigraph in $D \oplus R$ of the convex function $u \to (Fu)(x_o)$ contains an open set whose projection on D is U . This epigraph is contained in the epigraph of inf F . Apply Theorem 1. ||

If (P) has a unique Lagrange multiplier vector u^* , for instance, and the condition in Theorem 5 is satisfied, then $-u^*$ is the Gâteaux gradient of inf F at O by Theorem 2, and one has

$$\lim_{\lambda \to 0} [(\inf F)(\lambda u) - (\inf F)(0)]/\lambda = -\langle u,u^* \rangle \ , \forall u \ .$$

5. The Dual of a Convex Program.

The results above concerning Lagrange multipliers can be dualized to results about optimal solutions to convex programs. This is made possible by the fact that each general convex program (P) has a certain "dual", which is a "concave" program (P^*) . Broadly speaking, the optimal solutions to either program are the Lagrange multiplier vectors for the other program.

A function g on E is said to be concave, of course, if $-g$ is convex. We shall find it convenient to place concave functions on an equal plane with convex functions in what follows. The theory of concave functions is a mirror image of the theory of convex functions, of course. The symbols $\leq$, inf , $+\infty$, are everywhere interchanged with $\geq$, sup , $-\infty$. There is no need for us to write down the definitions of dom g , cl g , ∂g , etc., when g is concave. These definitions are all obvious. There is a possible source of confusion, however, in the case of conjugates. The conjugate g^* of g is defined by

$$g^*(x^*) = \inf_x \{ \langle x,x^* \rangle - g(x) \} ,$$

and what needs caution is the fact that $g^* \neq -(-g)^*$. (Instead one has

$$g^*(x^*) = -(-g)^*(-x^*) .)$$

A concave bifunction G from D to E is associated with a concave program (Q) , in which one maximizes rather than minimizes. The perturbation function for this program is denoted accordingly by sup G :

$$(\sup G)(u) = \sup(Gu) = \sup_x (Gu)(x) \ .$$

The dual of a convex program is defined in terms of a notion of the adjoint of a convex bifunction. Let F be a convex bifunction from D to E . We shall assume for simplicity that F is closed, i.e. that the graph function of F is closed. We define the adjoint of F to be the bifunction F^* from E^* to D^* given by

$$(F^* x^*)(u^*) = \inf_{u,x} \{ (Fu)(x) - \langle x,x^*\rangle + \langle u,u^*\rangle \} .$$

It is not difficult to show, from the basic facts about conjugacy, that F^* is a closed concave bifunction. The adjoint of a concave bifunction is defined in the same way, except with "sup" in place of "inf", and it is a convex bifunction. One has $F^{**} = F$.

This definition of " adjoint" may be regarded as a generalization of the adjoint of a linear tranformation. Convex bifunctions from D to E and their adjoints correspond to concave-convex functions on $D \times E^*$ much as linear transformations from D to E correspond to bilinear functions on $D \times E^*$. See [39].

Let F be a closed convex bifunction from D to E and let (P) be the associated convex program. The concave program (P*) associated with the adjoint bifunction F^* is defined to be the dual of (P) . The dual of (P*) is (P) again, inasmuch as $F^{**} = F$. In (P) , one minimizes the convex function $F0$ on E in the context of minimizing "neighboring" functions Fu corresponding to various perturbation vectors $u \in D$. In (P*) , one maximizes the concave function F^*0 on D^* in the context of maximizing "neighboring" functions F^*x^* corresponding to various perturbation vectors $x^* \in E^*$. The "optimal value" in (P) is $(\inf F)(0)$, while the "optimal value" in (P*) is $(\sup F^*)(0)$.

This concept of "dual program" reduces to the familiar one in the case of linear programs. Suppose $D = D^* = R^m$ and $E = E^* = R^n$. Let A be a linear transformation from R^n to R^m , let $a \in R^m$ and $a^* \in R^n$. Define

$$(Fu)(x) = \begin{cases} \langle x,a^*\rangle & \text{if } x \geq 0 , Ax \geq a - u, \\ +\infty & \text{if not.} \end{cases}$$

Then F is a convex bifunction from R^m to R^n , and in the program (P) associated with F one minimizes $\langle x,a^*\rangle$ subject to $x \geq 0$, $Ax \geq a$. By a straightforward calculation, it is seen that the adjoint of F is given by

$$(F^*x^*)(u^*) = \begin{cases} \langle a,u^*\rangle & \text{if } u^* \geq 0 , A^*u^* \leq a^* - x^* , \\ -\infty & \text{if not,} \end{cases}$$

where A^* is the adjoint of A (corresponding to the transpose matrix). In the dual program (P*) , therefore, one maximizes $\langle a,u^*\rangle$ subject to $u^* \geq 0$, $A^*u^* \leq a^*$.

In general, we shall say that (P) is normal if the function $\inf F$ coincides with its closure at 0 . Assuming that $(\inf F)(0) < +\infty$, this is equivalent to the natural stability condition that

$$\liminf_{u \to 0} \lfloor (\inf F)(u) \rfloor \geq (\inf F)(0) \ .$$

Similarly, (P^*) is said to be normal if the concave function $\sup F^*$ agrees with its closure at 0. We then can state a basic general fact about the relationship between (P) and (P^*).

THEOREM 6. *The following three conditions are equivalent:*

(a) (P) *is normal;*

(b) (P^*) *is normal;*

(c) $(\inf F)(0) = (\sup F^*)(0)$ i.e. *the extrema in* (P) *and* (P^*) *are equal.*

Let us say simply that *normality holds* when these equivalent properties are present. We note in particular that normality holds when F satisfies the condition in Theorem 5, since then $\inf F$ is actually continuous at 0. Likewise, normality holds if F^* satisfies the dual version of Theorem 5. It can also be seen easily that normality holds if a Lagrange multiplier vector exists for (P) or for (P^*).

Another duality theorem is the following.

THEOREM 7. *Suppose that normality holds and that the common value for the infimum in* (P) *and the supremum in* (P^*) *is finite. Then* u^* *is a Lagrange multiplier vector for* (P) *if and only if* u^* *is an optimal solution to* (P^*). *Dually,* x *is a Lagrange multiplier vector for* (P^*) *if and only if* x *is an optimal solution to* (P).

This theorem implies, for example, that if F^* satisfies the hypothesis of the dual version of Theorem 5, then (P) has an optimal solution. Moreover, the optimal solution x is unique if and only if $-x$ is the Gâteaux gradient of the perturbation function $\sup F^*$ at 0.

6. Hamiltonian Dynamics and Control Theory.

Some new applications of convex function theory to the calculus of variations will now be sketched briefly. Specifically, we shall indicate how convexity methods make possible an extension of Hamiltonian dynamics broad enough to include many probelms of control theory.

In classical Hamiltonian dynamics, of course, the state of a given "system" is represented by a moving point in R^n. The possible trajectories $t \to x(t)$ are all differentiable, and they are characterized in terms of extremal properties of

$$\int_{\alpha}^{\beta} L(x(t), \dot{x}(t), t)dt ,$$

where L is a real-valued sufficiently differentiable function on $R^n \times R^n \times R$ called the Lagrangian of the system. The trajectories are thus the solutions to the Euler-Lagrange equations:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{x}} (x(t), \dot{x}(t), t) = \frac{\partial L}{\partial x} (x(t), \dot{x}(t), t)$$

(where $\partial L/\partial x$ is the vector in R^n consisting of the first partial derivatives of L with respect to the components of x, and similarly $\partial L/\partial \dot{x}$). The trajectories are also characterized as the solutions to the Hamiltonian equations

$$\dot{x}(t) = \frac{\partial H}{\partial p} (x(t), p(t), t) , \dot{p}(t) = - \frac{\partial H}{\partial x} (x(t), p(t), t).$$

The function H, called the Hamiltonian of the system, is obtained from L by taking the Legendre transform in $\dot{x}$ for each x and t. (It is assumed that this transformation is well-defined.) The vector variable p in the Hamiltonian equations is connected with duality. (In the case of the motion of planets, for instance, x is position, p is momentum and H is total energy.) The system is completely described by either its Langrangian or its Hamiltonian.

We propose to extend the classical theory by using Fenchel's conjugacy correspondence instead of the Legendre transformation. Differentiability assumptions are to be replaced by convexity assumptions. The Lagrangian function is to be convex, so that its extremals correspond to minima, but it need not be finite everywhere, and this is what will make the theory applicable to certain problems in control theory. The Pontryagin Maximum Principle for such problems appears in the form of generalized Hamiltonian equations involving subgradients instead of gradients. The paths $p(t)$ turn out to be the trajectories for a certain "dual system". This duality is an extension of the theory developed by Friedrichs [13] (see also [5] and [26]) in the case where the Legendre transformation is well-defined.

In the following discussion, we can only scratch the surface of a very large subject. There are many new results besides the ones we mention here. For simplicity, we shall keep to autonomous systems of the "nicest" kind.

Let E_τ denote the space of all absolutely continuous functions x from the interval $[0, \tau]$ to R^n (paths in R^n). Of course, E_τ is a Banach space under the norm

$$||x|| = |x(0)| + \int_0^\tau |\dot{x}(t)|dt .$$

Let L be any closed proper convex function on R^{2n}. Given a path x in E_τ, we can consider the integral

$$I_\tau[x] = \int_0^\tau L(x(t), \dot{x}(t))dt .$$

It turns out that this integral is always well-defined and is never $-\infty$. In fact I_τ is a closed convex function on E_τ [38]. The general theory of convex programs can therefore be applied to problems involving the minimization of I_τ.

For example, suppose one wants to minimize I_τ subject the boundary conditions $x(0) \in C$, $x(\tau) \in D$, where C and D are closed convex sets in R^n. The set of curves satisfying these boundary conditions is a closed convex subset of E_τ, so the problem is one of minimizing a convex function over a convex set. Existence theorems, transversality conditions, ect. can be deduced by convexity methods alone. "Convex" perturbations may be introduced, and these lead to various dualities including the duality between the x and p systems below.

Note that $I_\tau[x] = +\infty$ unless the point $(x(t), \dot{x}(t))$ belongs to the convex set dom L for almost every t. The latter condition is thus an implicit constraint in any problem where I_τ is minimized. For instance, if

$$L(x,\dot{x}) = \begin{cases} L_0(x,\dot{x}) & \text{if } L_i(x,\dot{x}) \leq 0,\ i = 1,\ldots, m, \\ +\infty & \text{if not,} \end{cases}$$

where $L_0, L_1,\ldots, L_m$ are finite convex functions, then minimizing I_τ is equivalent to minimizing

$$\int_0^\tau L_0(x(t), \dot{x}(t))dt$$

subject to the constraints that

$$L_1(x,\dot{x}) \leq 0,\ldots, L_m(x,\dot{x}) \leq 0 ,$$

for almost every t.

For each pair of points a and b in R^n, let $f_\tau(a,b)$ denote the infimum of $I_\tau[x]$ over all the paths $x \in E_\tau$ such that $x(0) = a$ and $x(\tau) = b$. Then f_τ is a convex function on R^{2n}. Paths along which the infimum defining f_τ is attained and finite are called extremals of L. More exactly, an extremal of L is any absolutely continuous function x from an interval $[\alpha,\beta]$ to R^n such that

$$\int_{\alpha}^{\beta} L(x(t),\ \dot{x}(t))dt = f_{\beta-\alpha}(x(\alpha),\ x(\beta)) < \infty \ .$$

Some limitations on L are needed in order to obtain a good existence theorem for extremals. We shall denote by W the class of all closed convex functions f on R^{2n} such that, for each $x \in R^n$, the convex function $f(x,.)$ on R^n is proper and has no non-vertical half-lines in its epigraph.

THEOREM 8. _If_ $L \in W$, _then_ $f_\tau \in W$ _for every_ $\tau \geq 0$, _and the infimum defining_ $f_\tau(a,b)$ _is always attained by at least one path_ x. _Moreover, if the point_ (a,b) _belongs to the relative interior of the convex set_ $\mathrm{dom}\, f_\tau$ _(the interior of_ $\mathrm{dom}\, f_\tau$ _with respect to the smallest affine set -- translate of a subspace -- containing_ $\mathrm{dom}\, f_\tau$), _then the path_ x _for which the infimum is attained has a bounded derivative (i.e. the function_ $\dot{x}$ _belongs to_ L^∞ _rather than to_ L^1).

We may think of each $L \in W$ as the Lagrangian of a certain "convex dynamical system". The system always behaves in such a way as to minimize the integrals

$$\int_{\alpha}^{\beta} L(x(t),\ \dot{x}(t))dt \ .$$

The trajectories in R^n representing changing states of the system are the extremals of the Lagrangian L. According to Theorem 8, if there exists a path from a at $t = 0$ to b at $t = \tau$, such that the integral of L along the path is not $+\infty$, then there actually exists a trajectory of the system passing from state a to state b in τ time units.

For example, suppose the given system is an "economy" whose states are expressed by vectors $x \in R^n$ (the components of x being the amounts of various goods that are present). Let $L(x,\dot{x})$ be the cost of producing goods at a rate $\dot{x} \in R^n$ when the economy is in the state $x \in R^n$. (Certain "production schedules" $\dot{x}$ may be forbidden by setting their cost $= +\infty$.) The integral of L along any given path from a to b is the total cost of passing from state a to state b along this path. If we assume that the economy always behaves so as to minimize total cost, it follows that the state trajectories of the economy will be the extremals of its Lagrangian L. (In connection with this assumption about ideal behavior, one can think of the system as not just the economy itself, but the economy combined with a perfect controller. Thus the extremals of L show how an _optimally controlled economy_ would behave, if optimality is taken to mean minimum "cost".)

If L is the (convex) indicator function of the graph of a linear transformation A from R^n to R^n, then $L \in W$ and the extremals of L are just the solutions to the differential equation $\dot{x} = Ax$. In this case f_τ is of course the indicator function of the linear transformation $e^{\tau A}$ from R^n to R^n, where

$$e^{\tau A} = I + \tau A + (\tau^2/2)A^2 + \ldots$$

The transformations $e^{\tau A}$ form a one-parameter group which is a representation of the additive group of real numbers, i.e. one has

$$e^{(\tau+\sigma)A} = e^{\tau A}e^{\sigma A}, \forall \tau, \sigma \in R .$$

This is the so-called <u>dynamical group</u> of the system with Lagrangian L.

We would like to mention that "convex dynamical systems" given by Lagrangians $L \in W$ correspond similarly to <u>dynamical semigroups</u> consisting of <u>bifunctions</u> rather than linear transformations. It can be shown that the class of all convex bifunctions from R^n to R^n whose graph functions belong to W is a (non-commutative) semigroup if multiplication of bifunctions is defined by

$$((GF)u)(y) = \inf_x \{ (Fu)(x) + (Gx)(y) \} .$$

According to Theorem 8, the bifunctions F given by $(F^\tau a)(b) = f_\tau(a,b)$ are elements of this semigroup, and it happens that

$$F^{\tau+\sigma} = F^\tau F^\sigma, \forall \tau \geq 0, \forall \sigma \geq 0 .$$

Thus one has a one-parameter semigroup whose "infinitesimal generator" is the bifunction $x \to L(x,\cdot)$.

Given a "convex dynamical system" with Lagrangian $L \in W$, we define the <u>dual</u> "convex dynamical system" to be the one with Lagrangian M, where

$$M(p,\dot{p}) = L^*(\dot{p},p) = \sup_{\substack{x \in R^n \\ \dot{x} \in R^n}} \{ \langle x,\dot{p}\rangle + \langle \dot{x},p\rangle - L(x,\dot{x}) \} .$$

It can be proved that M likewise belongs to W, so that Theorem 8 is applicable to the dual system. The dual of the dual system is the original system, as follows from the symmetry of Fenchel's conjugacy correspondence.

The integral of M along a path $p \in E_\tau$ will be denoted by $J_\tau(p)$. The infimum of $J_\tau(p)$ with respect to all paths p such that $p(0) = c$ and $p(\tau) = d$ will be denoted by $g_\tau(c,d)$.

Observe that, for any two paths x and p in E_τ, one has

$$I_\tau[x] + J_\tau[p] = \int_0^\tau [L(x(t), \dot{x}(t)) + M(p(t), \dot{p}(t))]dt$$
$$\geq \int_0^\tau [\langle x(t), \dot{p}(t)\rangle + \langle \dot{x}(t), p(t)\rangle]dt$$
$$= \int_0^\tau \frac{d}{dt} \langle x(t), p(t)\rangle dt = \langle x(\tau), p(\tau)\rangle - \langle x(0), p(0)\rangle .$$

Therefore, for every a, b, c and d

$$f_\tau(a,b) + g_\tau(c,d) \geq \langle b,d\rangle - \langle a,c\rangle ,$$

and one has

$$g_\tau(c,d) \geq \sup_{a,b} \{ \langle a,-c\rangle + \langle b,d\rangle - f_\tau(a,b)\} = f^*_\tau(-c,d) .$$

It can be proved that actually $g_\tau(c,d) = f^*_\tau(-c,d)$ and $f_\tau(a,b) = g^*_\tau(-a,b)$ (at least when $L \in W$) .

THEOREM 9. <u>Assume that</u> $L \in W$. <u>In order that a given path</u> $x \in E_\tau$ <u>be an extremal of</u> L , <u>it is sufficient that there exist a path</u> $p \in E_\tau$ <u>such that</u>

$$(\dot{p}(t), p(t)) \in \partial L(x(t), \dot{x}(t))$$

<u>for almost every</u> t , <u>in which case</u> p <u>is an extremal of</u> M . <u>This condition is necessary, as well as sufficient, if the point</u> $(x(0), x(\tau))$ <u>is in the relative interior of the convex set</u> $\mathrm{dom}\, f_\tau$.

The subgradient condition in Theorem 9 may be called the <u>Euler-Lagrange</u> condition for L , since if L is differentiable it reduces to the classical Euler-Lagrange equations:

$$\frac{dp(t)}{dt} = \frac{\partial L}{\partial x} (x(t), \dot{x}(t)), \; p(t) = \frac{\partial L}{\partial \dot{x}} (x(t), \dot{x}(t)) .$$

The Euler-Lagrange conditions for L and M are equivalent: one has $(\dot{p}, p) \in \partial L(x, \dot{x})$ if and only if $(\dot{x}, x) \in \partial M(p, \dot{p})$. Paths x and p satisfying these conditions will be called extremals <u>dual</u> to each other.

It can be shown that extremals dual to each other can be expressed as optimal solutions to certain dual programs. In the case of an "economy" as above, the dual system may be interpreted in terms of "market states", with $p(t)$ as a price vector.

Given a "convex dynamical system" with Lagrangian $L \in W$, we define the <u>Hamiltonian</u> H of the system to be the function on $R^n \times R^n$ obtained by taking the conjugate of $L(x, \cdot)$ for each $x \in R^n$. Thus

$$H(x,p) = \sup_{\dot{x} \in R^n} \{\langle \dot{x}, p\rangle - L(x, \dot{x})\}$$
$$L(x, \dot{x}) = \sup_{p \in R^n} \{\langle \dot{x}, p\rangle - H(x,p)\} .$$

Since this correspondence between Lagrangians $L \in E$ and Hamiltonians H is one-to-one, the system can be described completely by specifying H instead of by specifying L. We may ask just what class of functions H is involved here. The answer, proved in [30] and [39] is perhaps surprising: the Hamiltonians corresponding to Lagrangians $L \in W$ are precisely the real-valued (everywhere finite) functions H on $R^n \times R^n$ such that $H(x,p)$ is a concave function of x for each $p \in R^n$ and a convex function of p for each $x \in R^n$.

Given such a function H, we denote by $\partial_p H(x,p)$ the set of subgradients of the convex function $H(x,\cdot)$ at the point p, and by $-\partial_x H(x,p)$ the set of subgradients of the convex function $-H(\cdot,p)$ at the point x. Thus, when H is actually differentiable we have

$$\partial_p H(x,p) = \nabla_p H(x,p) = \frac{\partial H}{\partial p}(x,p) ,$$

$$\partial_x H(x,p) = \nabla_x H(x,p) = \frac{\partial H}{\partial x}(x,p) .$$

THEOREM 10. Assume $L \in W$. In order that paths x and p be extremals dual to each other, it is necessary and sufficient that they satisfy the Hamiltonian conditions

$$\dot{x}(t) \in \partial_p (H(x(t),p(t)), \; -\dot{p}(t) \in \partial_x H(x(t),p(t)) ,$$

for almost every t.

It can be proved from the existence theory for solutions to contingent differential equations that the generalized Hamiltonian equations in Theorem 10 have at least one solution proceeding from each choice of the initial point $(x(0),p(0))$.

We cannot furnish an adequate explanation here of the relationship between the above results and results in control theory, but the following special example will serve to indicate some of the connections.

Let f be a finite convex function on R^n, and let g be a closed proper convex function on R^m whose epigraph contains no non-vertical half-lines. We are interested in minimizing

$$\int_0^\tau [f(x(t)) + g(u(t))]\,dt$$

subject to endpoint conditions $x(0) = a$, $x(\tau) = b$, and the differential equation

$$\dot{x} = Ax + Bu ,$$

where A is $n \times n$, B is $n \times m$, $x \in E_\tau$ and u is a function from $[0,\tau]$ to R^n belonging to L^1. Here u may be interpreted as a control for a system whose states or positions are given by x , and f and g may be interpreted as cost functions. For each $x \in R^n$ and $\dot{x} \in R^n$, let

$$L(x,\dot{x}) = f(x) + \inf \{g(u) \mid u \in R^m , Bu = \dot{x} - Ax\} .$$

It can be shown that the infimum is always attained by some $u \in R^m$ (when the infimum is not merely $+\infty$). Moreover, $L \in W$. The "convex dynamical system" with Lagrangian L can be interpreted as the original system combined with a system which, for any position x and velocity $\dot{x}$, fixes the control u so as to yield the given velocity as cheaply as possible.

What are the trajectories for this system, assuming it behaves so as to minimize total cost? Every trajectory which has a dual (and this includes "most" trajectories by Theorem 9) occurs as a solution to the generalized Hamiltonian equations. We calculate the Hamiltonian H corresponding to L as:

$$\begin{aligned} H(x,p) &= \sup_x \{\langle x,p\rangle - f(x) - \inf \{g(u) \mid Bu = \dot{x} - Ax\}\} \\ &= \sup_u \{\langle Bu + Ax,p\rangle - f(x) - g(u)\} \\ &= \langle Ax,p\rangle - f(x) + \sup_u \{\langle u,B^*p\rangle - g(u)\} \\ &= \langle Ax,p\rangle - f(x) + g^*(B^*p) , \end{aligned}$$

where B^* is the transpose of B . The assumptions on g actually imply that g^* is finite everywhere. The generalized Hamiltonian equations reduce to

$$\dot{x} \in (Ax + B\, \partial g^*(B^*p)) ,$$

$$-\dot{p} \in (A^*p - \partial f(x)) ,$$

where the first condition can also be written as

$$\dot{x} = Ax + Bu , \quad u \in \partial g^*(B^*p) .$$

To make things specific, let us suppose that $m = n$, $A = 0$, $B = I$, $f(x) = \langle x,e\rangle$ where $e \neq 0$, and

$$g(u) = \begin{cases} |u| & \text{if } |u| \leq 1 , \\ +\infty & \text{if } |u| > 1 . \end{cases}$$

(For this choice of g , the cost of a control function $t \to u(t)$ is $+\infty$ unless $u(t)$ belongs to the Euclidean unit ball for almost every t . This ball is the control region, in other words.) The Lagrangian in this case is given by

$$L(x,\dot{x}) = \begin{cases} \langle x,e\rangle + |\dot{x}| & \text{if } |\dot{x}| \leq 1 , \\ +\infty & \text{if } |\dot{x}| > 1 , \end{cases}$$

and the Hamiltonian equations are

$$\dot{x} \in \partial g^*(p) , \quad \dot{p} = e .$$

One calculates easily that

$$g^*(p) = \max\{\,0,\ |p| - 1\,\},$$

$$\partial g^*(p) = \begin{cases} 0 & \text{if } |p| < 1, \\ \{\lambda p \mid 0 \le \lambda \le 1\} & \text{if } |p| = 1, \\ |p|^{-1}p & \text{if } |p| > 1. \end{cases}$$

Thus, given any points a and c in R^n, there is a unique dual pair of extremals x and p such that $x(0) = a$ and $p(0) = c$, namely

$$p(t) = c + te,$$

$$x(t) = a + \int_0^t s(c + te)\,dt,$$

where

$$s(c + te) = \begin{cases} 0 & \text{if } |c + te| \le 1, \\ |c + te|^{-1}(c + te) & \text{if } |c + te| > 1. \end{cases}$$

The choice of c determines whether a given terminal condition $x(\tau) = b$ will be satisfied.

BIBLIOGRAPHY

1. E. Asplund and R.T. Rockafellar, "Gradients of convex functions", in preparation.

2. A. Brøndsted, "Conjugate convex functions in topological vector spaces", Mat.-Fys Medd. Dansk. Vid. Selsk. 34 (1964), 1-26.

3. A. Brøndsted and R.T. Rockafellar, "On the subdifferentiability of convex functions", Proc. Amer. Math. Soc. 16 (1965), 605-611.

4. L. Cesari, "Existence theorems for weak and usual optimal solutions in Lagrange problems with unilateral constraints I", Trans. Amer. Math. Soc. 124 (1966), 369-412.

5. R. Courant and D. Hilbert, Methods of Mathematical Physics, Vol. 1 (Springer, Berlin, 1937), Chap. 4, §9.

6. G. Dantzig, E. Eisenberg and R.W. Cottle, "Symmetric dual nonlinear programs", Pacific J. Math. 15 (1965), 809-812.

7. U. Dieter, "Dualität bei konvexen Optimierungs - (Programmierungs -) Aufgaben", Unternehmungsforschung 9 (1965), 91-111.

8. U. Dieter, "Optimierungsaufgaben in topologischen Vektorräumen I: Dualitätstheorie", Z. Wahrscheinlichkeitstheorie verw. Geb. 5 (1966), 89-117.

9. R.J. Duffin, "Infinite programs", Annals of Math. Study 38 (1956), 157-170.

10. K. Fan, I. Glicksberg and A.J. Hoffman, "Systems of inequalities involving convex functions", Proc. Amer. Math. Soc. 57 (617-622).

11. W. Fenchel, "On conjugate convex functions", Canad. J. Math. 1 (1949), 73-77.

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

13. K.O. Friedrichs "Ein Verfahren der Variationsrechnung das Minimum eines Integral als das maximum eines anderen Ausdruckes darzustellen", Nachr. der Gesellschaft der Wiss., Göttingen (1929), 13-20.

14. D. Gale, "A geometric duality theorem with economic application", Review of Econ. Studies 34 (1967).

15. F. John, "Extremum problems with inequalities as subsidiary conditions", Studies and Essays, Courant Anniversary Volume (Interscience, New York, 1948).

16. S. Karlin, Mathematical Methods and Theory in Games, Programming and Economics (McGraw-Hill, New York, 1960), Vol. 1.

17. H.W. Kuhn and A.W. Tucker, "Nonlinear programming", Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (Univ. of California Press, Berkeley, 1951), 481-492.

18. O.L. Mangasarian and J. Ponstein, "Minimax and duality in nonlinear programming", J. Math. Anal. Appl. 11 (1965), 504-518.

19. J.-J. Moreau, "Fonctions convexes en dualité", multigraph, Séminaires de Math., Faculté des Sciences, Université de Montpellier (1962).

20. J.-J. Moreau, "Étude locale d'une fonctionelle convexe", multigraph, Séminaires de Math., Faculté des Sciences, Université de Montpellier (1963).

21. J.-J. Moreau, "Fonctionelles sous-differentiables", C.R. Acad. Sci 257 (1963), 4117-4119.

22. J.-J. Moreau, "Théorémes 'inf-sup' ", C.R. Acad. Sci. 258 (1964), 2720-2722.

23. J.-J. Moreau, "Proximité et dualité dans une espace hilbertien", Bull. Soc. Math. France 93 (1965), 273-299.

24. J.-J. Moreau, "Sous-differentiabilité", *Proceedings of the Colloquium on Convexity, Copenhagen, 1965*, W. Fenchel, ed. (Copenhagen Matematisk Institut, 1967), 185-201.

25. J.-J. Moreau, *Fonctionelles Convexes*, lecture notes, Séminaire "Equations aux dérivées partielles", Collége de France, 1966.

26. J.D. Pearson, "Reciprocity and duality in control programming problems", J. Math. Anal. Appl. 10 (1965), 338-408.

27. R.T. Rockafellar, *Convex Functions and Dual Extremum Problems*, thesis, Harvard, 1963.

28. R.T. Rockafellar, "Minimax theorems and conjugate saddlefunctions", Math. Scand. 14 (1964), 151-173.

29. R.T. Rockafellar, "Level sets and continuity of conjugate convex functions", Trans. Amer. Math. Soc. 123 (1966), 46-63.

30. R.T. Rockafellar, "A general correspondence between dual minimax problems and convex programs", (1965), to appear in Pacific J. Math., 25 (June 1968).

31. R.T. Rockafellar, "An extension of Fenchel's duality theorem for convex functions", Duke Math. J. 33 (1966), 81-90.

32. R.T. Rockafellar, "Characterization of the subdifferentials of convex functions", Pacific J. Math. 17 (1966), 81-90.

33. R.T. Rockafellar, "A monotone convex analogue of linear algebra", *Proceedings of the Colloquium on Convexity, Copenhagen, 1965*, W. Fenchel, ed. (Mathematisk Institut, Copenhagen, 1967), 261-276.

34. R.T. Rockafellar, *Monotone Processes of Concave and Convex Type*, Memoirs of the Amer. Math. Soc., no. 77 (1967), 74 pp.

35. R.T. Rockafellar, "Duality theorems for convex functions", Bull. Amer. Math. Soc. 70 (1964), 189-192.

36. R.T. Rockafellar, "Conjugates and Legendre transforms of convex functions", Canad. J. Math. 19 (1967), 200-205.

37. R.T. Rockafellar, "Duality and stability in extremum problems involving convex functions", Pacific J. Math. 21 (1967), 167-187.

38. R.T. Rockafellar, "Integrals which are convex functionals", Pacific J. Math. 24 (1968), 525-539.

39. R.T. Rockafellar, *Convex Analysis*, 650 page manuscript, to be published in 1969 by the Princeton University Press.

40. A. Whinston, "Some applications of the conjugate function theory to duality", *Nonlinear Programming*, J. Abadie, ed. (Nordt-Holland, Amsterdam, 1967), 75-96.

41. P. Wolfe, "A duality theorem for nonlinear programming", Q. Appl. Math 19 (1961), 239-244.