A REPRINT

FROM

Volume 11
Lectures in Applied Mathematics

MATHEMATICS OF THE DECISION SCIENCES

PART 1

R. T. Rockafellar[1]

Duality in Nonlinear Programming

1. **Introduction.** An ordinary nonlinear program in n variables may be defined as a problem of minimizing a quantity $f_0(x)$ subject to constraints $f_1(x) \leqq 0, \cdots, f_m(x) \leqq 0$, where $f_0, \cdots, f_m$ are certain real-valued functions of the vector $x = (x_1, \cdots, x_n) \in R^n$. The problem may be interpreted broadly or narrowly, however.

In the narrower sense, one is only interested in the infimum of a certain function given on a subset S of R^n. The elements x of the subset S are the so-called feasible solutions to the problem. Typical questions are the following. Is the infimum finite? Do there exist optimal solutions, i.e. feasible solutions at which the infimum is attained? Is there only one optimal solution? One seeks conditions which guarantee "yes" answers to these questions, as well as algorithms for actually computing the infimum and optimal solutions.

In the broader sense of the probelm, one is also concerned with the sensitivity of the infimum and optimal solutions to slight changes in the constraints. This is where duality and Lagrange multipliers come in. Let $p(u_1, \cdots, u_m)$ denote the infimum of $f_0(x)$ subject to $f_1(x) \leqq u_1, \cdots, f_m(x) \leqq u_m$. Each $u = (u_1, \cdots, u_m) \in R^m$ corresponds to a certain perturbation of the given problem and p gives the infimum in the perturbed problem as a function of u.

[1] This work was supported in part by grant AF-AFOSR-1202-67 from the Air Force Office of Scientific Research.

One is interested in the properties of p near $u = 0$. For instance, is p continuous or differentiable at $u = 0$?

It is especially important to look for numbers $u_1^*, \cdots, u_m^*$ such that

$$p(u_1, \cdots, u_m) \geqq p(0, \cdots, 0) - u_1^* u_1 - \cdots - u_m^* u_m, \quad \forall (u_1, \cdots, u_m) \in R^m. \tag{1.1}$$

Such numbers can be interpreted as "equilibrium prices" if the objective function f_0 is interpreted as a cost function. Suppose that in trying to minimize cost we are allowed to perturb the given problem in the above sense by any amount $(u_1, \cdots, u_m)$, but that this perturbation must be paid for, the price being u_i^* per unit of variable u_i. The minimum cost attainable in the problem perturbed by $(u_1, \cdots, u_m)$, plus the cost of this perturbation, is

$$p(u_1, \cdots, u_m) + u_1^* u_1 + \cdots + u_m^* u_m.$$

If the prices satisfy (1.1), this is never less than the minimal cost $p(0, \cdots, 0)$ in the unperturbed problem, so all the incentive for perturbation is neutralized and there is an "equilibrium".

Observe that (1.1) is satisfied if and only if

$$f_0(x) + u_1^* u_1 + u_m^* u_m \geqq p(0, \cdots, 0) \tag{1.2}$$

for every choice of x and $(u_1, \cdots, u_m)$ such that $f_i(x) \leqq u_i$ for $i = 1, \cdots, m$. Assuming $p(0, \cdots, 0)$ is finite, (1.2) is equivalent to the condition that $u_i^* \geqq 0$ for $i = 1, \cdots, m$ and

$$f_0(x) + u_1^* f_1(x) + \cdots + u_m^* f_m(x) \geqq p(0, \cdots, 0), \quad \forall x \in R^n.$$

(If u_i^* were negative for some i, (1.2) would fail for high values of u_i.) In other words, the equilibrium prices are the same as the nonnegative Lagrange multipliers $u_1^*, \cdots, u_m^*$ such that the unconstrained infimum of $f_0 + u_1^* f_1 + \cdots + u_m^* f_m$ coincides with the infimum of f_0 subject to the constraints $f_1(x) \leqq 0, \cdots, f_m(x) \leqq 0$.

These reflections on the nature of an ordinary nonlinear program lead us to propose a concept of a generalized nonlinear program as not just an isolated problem of minimizing a given function over a given set, but such a problem *together with a particular class of perturbations*. In such a program, one is to study not only the infimum in the problem corresponding to zero perturbation, but also the sensitivity of the infimum with respect to perturbations to the "neighboring" problems. The Lagrange multipliers are to be the "equilibrium prices" for the perturbations.

The terminology of "bifunctions" is useful in describing the dependence of an abstract minimization problem on a perturbation. Suppose that for each vector $u \in R^m$ we are given a pair (Su, Fu), where Su is a subset of R^n (possibly empty) and Fu is a function on Su with values in $[-\infty, +\infty]$. The correspondence

$$F: u \to (Su, Fu)$$

will be called a *bifunction* from R^m to R^n. A bifunction is to be regarded as a generalization of "multivalued mapping": the image of u under F is not just a set, but a set with a distinguished function attached to it. One can interpret the function Fu as assigning a relative cost $(Fu)(x)$ to each element x of the set Su.

For any bifunction F from R^m to R^n, we define a *generalized program* (P): minimized the function $F0$ on the set $S0$. The problem is to include the local analysis of the properties of the function $p = \inf F$ at $u = 0$, where

$$(\inf F)(u) = \inf\{(Fu)(x) \mid x \in Su\}.$$

(By convention, an infimum is $+\infty$ if the set over which it is taken is empty.) A vector $x \in R^n$ will be called an *optimal solution* to (P) if $(\inf F)(0)$ is finite and attained at x. If $(\inf F)(0)$ is finite, we define a *Kuhn-Tucker vector* for (P) to be a vector $u^* \in R^m$ such that

$$(\inf F)(u) + \langle u^*, u\rangle \geqq (\inf F)(0)$$

for every perturbation $u \in R^m$. (Here $\langle \cdot, \cdot \rangle$ denotes the ordinary inner product of two real vectors.)

Observe that, if F is extended by setting $(Fu)(x) = +\infty$ for all $x \notin Su$, the corresponding (P) is essentially the same. No generality is lost, therefore, if one considers only cases where $Su = R^n$ for all u.

Under simple convexity assumptions on the bifunction F, a comprehensive duality theory is possible for generalized programs, as will be explained below. A dual program (P*) may be constructed which is of the same type, except that it involves maximization rather than minimization. The dual of the program (P*) is in turn (P). The extrema in (P) and (P*) are generally equal. The optimal solutions to (P*) are generally the Kuhn-Tucker vectors for (P), while the optimal solutions to (P) are the Kuhn-

Tucker vectors for (P*). The pairs of optimal solutions to (P) and (P*) are the saddle-points of a certain Lagrangian function.

An intriguing mathematical feature of the theory to be explained is that it constitutes a new "convex algebra" closely parallel to linear algebra. The convex bifunction F plays a role analogous to that of a linear transformation. Duality is obtained by the construction of an adjoint bifunction F^* in terms of Fenchel's conjugacy correspondence. Whereas a linear transformation and its adjoint are related by a bilinear function, a convex bifunction and its adjoint are related by a concave-convex function. The formula $\langle Fu, x^* \rangle = \langle u, F^* x^* \rangle$ (in which $\langle \cdot, \cdot \rangle$ denotes a generalized inner product to be defined in §4) appears as an "inf = sup" theorem for a dual pair of programs. Minimax theory is associated with the "inverse" operation for bifunctions.

The results in this paper are based on the general theory of convex functions and especially on the very important notion of conjugacy due to Fenchel [**17**]. The elementary facts about convex functions are reviewed in §2. Further details can be found in the works of Fenchel, Brøndsted, Moreau and Rockafellar listed among the references.

The complete proofs of the new duality theorems and of the theorems about bifunctions are all contained in a forthcoming book [**44**]. Some of the main ideas have already appeared in previous papers, however. A perturbational approach to duality theory has been given by the author in [**43**] and [**38**]. The correspondence between concave-convex functions on $R^m \times R^n$ and convex functions on R^{m+n} (here the *graph functions* of convex bifunctions as defined in §3) has been established in [**38**]. A "convex algebra" for multivalued mappings has been developed in [**36**] and presented in [**37**].

Some applications of Fenchel's theory to general nonlinear programming have also been described by Ghouila-Houri [**2**], Dennis [**7**], Dieter [**8**], [**9**], Falk and Thrall [**15**], Karlin [**23**], and Whinston [**46**].

An excellent discussion of Lagrange multipliers as "equilibrium prices" has been given by Gale [**19**] in the case of concave maximization problems depending on parameters. Our idea of a "generalized program" is essentially derived from Gale's paper, but the dual problems we speak of are quite different.

2. **Convex functions and their conjugates.** The object of the finite-dimensional theory of convex functions is the study of

pairs (C,f), where C is a nonempty convex set in R^n and f is a real-valued convex function on C, i.e. a function from C to R satisfying

$$(2.1) \quad f((1-\lambda)x+\lambda y) \leqq (1-\lambda)f(x)+\lambda f(y), \quad 0<\lambda<1,$$

for any $x \in C$ and $y \in C$. There are technical advantages, however, in representing each such pair by a function which is defined on all of R^n but which may have infinity values, namely the function obtained by defining $f(x)$ to be $+\infty$ for $x \notin C$.

In general, let f be any function defined on all of R^n and having values which are real numbers or $\pm\infty$. The *epigraph* of f, denoted by $\operatorname{epi} f$, is the set of pairs (x,μ) in R^{n+1} such that $x \in R^n$, $\mu \in R$ and $\mu \geqq f(x)$. (Thus $\operatorname{epi} f$ can be regarded as the set of all "finite" points lying on or above the graph of f.) We define f to be a *convex function* on R^n if $\operatorname{epi} f$ is a convex subset of R^{n+1}. If there is no x such that $f(x) = -\infty$, this definition of convexity is equivalent to inequality (2.1) being satisfied throughout R^n with the obvious rules for manipulating $+\infty$. (If f takes on $-\infty$ as well as $+\infty$, (2.1) cannot be used, because it might involve the undefined espression $\infty - \infty$.)

If f is convex, the set

$$\operatorname{dom} f = \{x \mid f(x) < +\infty\},$$

which is the projection of $\operatorname{epi} f$ on R^n, is convex; it is called the *effective domain* of f. A convex function f on R^n is said to be *proper* if $\operatorname{dom} f$ is nonempty and f is finite on $\operatorname{dom} f$; in other words, if f is not the constant function $+\infty$ and there is no x such that $f(x) = -\infty$. The restriction of f to $C = \operatorname{dom} f$ is then a pair (C,f) of the type mentioned above, and every such pair arises in this way. Thus the study of the pairs (C,f) is replaced by the study of the proper convex functions f on R^n.

Convex functions which are improper can arise naturally as the result of certain operations, and they do have some technical uses. The fundamental fact about an improper convex function f on R^n is that f must be identically $-\infty$ on the interior of $\operatorname{dom} f$.

A useful example of a convex function is the *indicator function* $\delta(\cdot \mid C)$ of a convex set C in R^n, where $\delta(x \mid C) = 0$ for $x \in C$ and $\delta(x \mid C) = +\infty$ if $x \notin C$. If f_0 is a finite (i.e. real-valued, rather than extended-real-valued) convex function on R^n, the convex function $f = f_0 + \delta(\cdot \mid C)$ agrees with f_0 on C and is $+\infty$ elsewhere. Minimizing f_0 on C is equivalent to minimizing f over all of R^n.

We shall use this device to re-express all constrained extremum problems as formally unconstrained problems.

Let f be a convex function on R^n, and let D denote the collection of all pairs (x^*, μ^*) such that $x^* \in R^n$, $\mu^* \in R$ and

$$f(x) \geqq \langle x, x^* \rangle - \mu^*, \quad \forall x \in R^n.$$

The pointwise supremum of the corresponding collection of affine functions $h(x) = \langle x, x^* \rangle - \mu^*$ is called the *closure* of f and is denoted by $\mathrm{cl} f$. Thus by definition

$$(2.2) \qquad (\mathrm{cl} f)(x) = \sup \{ \langle x, x^* \rangle - \mu^* \mid (x^*, \mu^*) \in D \}.$$

When $\mathrm{cl} f = f$, one says that f is *closed*. If f is proper, it can be shown that the epigraph of $\mathrm{cl} f$ is simply the closure in R^{n+1} of the epigraph of f. Then $\mathrm{cl} f$ is a closed proper convex function on R^n, and

$$(2.3) \qquad (\mathrm{cl} f)(x) = \liminf_{y \to x} f(y), \quad \forall x \in R^n.$$

In particular, a proper convex function is closed if and only if it is *lower semicontinuous*, i.e. has the property that the convex level set $\{ x \mid f(x) \leqq \mu \}$ is closed in R^n for each real μ.

For a proper convex function f, $(\mathrm{cl} f)(x)$ must actually coincide with $f(x)$ for every x in the interior of $\mathrm{dom} f$ or outside the closure of $\mathrm{dom} f$. Thus $f \to \mathrm{cl} f$ may be regarded as a regularizing operation which simply redefines f at certain boundary points of its effective domain, so as to make f lower semicontinuous. For an improper convex function f, $\mathrm{cl} f$ is the constant function $-\infty$ or the constant function $+\infty$, depending on whether or not $\mathrm{dom} f$ is nonempty.

Fenchel's important notion of conjugacy is obtained by further consideration of the set D introduced above. Clearly D consists of the pairs (x^*, μ^*) in R^{n+1} such that $\mu^* \geqq f^*(x^*)$, where

$$(2.4) \qquad f^*(x^*) = \sup \{ \langle x, x^* \rangle - f(x) \mid x \in R^n \}.$$

Thus D is the epigraph of a certain function f^* on R^n. This f^* is called the *conjugate* of f.

It can be seen that f^* is a closed convex function on R^n, proper if and only if f itself is proper. The conjugate f^{**} of f^* is in turn given by

$$f^{**}(x) = \sup \{ \langle x, x^* \rangle - f^*(x^*) \mid x^* \in R^n \}.$$

But this supremum is the same as the supremum in (2.2). Thus $f^{**} = \mathrm{cl} f$. In particular, if f is closed it is the conjugate of its conjugate f^*.

Conjugacy therefore defines a one-to-one symmetric correspondence in the class of all closed convex functions on R^n.

As an example, the conjugate of the indicator function $\delta(\cdot \mid C)$ of a convex set C in R^n is given by

$$\delta^*(x^* \mid C) = \sup_{x \in R^n} \{ \langle x, x^* \rangle - \delta(x \mid C) \} = \sup_{x \in C} \langle x, x^* \rangle.$$

The function $\delta^*(\cdot \mid C)$ is called the *support function* of C.

A convex function f on R^n is necessarily continuous on the interior of its effective domain. It is differentiable almost everywhere on any open set where it is finite.

Assume that x is any point where f is finite. The (one-sided) directional derivative

$$f'(x; y) = \lim_{\lambda \downarrow 0} \frac{f(x + \lambda y) - f(x)}{\lambda} \tag{2.5}$$

exists and is a convex function of y (possibly with the values $\pm \infty$). Of course, if f is actually differentiable at x, we have

$$f'(x; y) = \langle \nabla f(x), y \rangle, \tag{2.6}$$

where $\nabla f(x)$ is the gradient of f at x,

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \cdots, \frac{\partial f}{\partial x_n}(x) \right).$$

If f is not differentiable at x, the directional derivatives can still be described in terms of "subgradients". A *subgradient* of f at x is a vector $x^* \in R^n$ such that

$$f(z) \geqq f(x) + \langle z - x, x^* \rangle, \quad \forall z \in R^n. \tag{2.6}$$

This condition means that the graph of the function $h(z) = f(x) + \langle z - x, x^* \rangle$ is a nonvertical supporting hyperplane in R^{n+1} to the epigraph of f at the point $(x, f(x))$. The set of subgradients x^* at x is a certain closed convex (possibly empty) set denoted by $\partial f(x)$.

The case where $\partial f(x)$ consists of just one x^* is precisely the case where f is finite and differentiable at x, the unique subgradient then being $\nabla f(x)$. It can be shown that, if x is actually an interior point of $\mathrm{dom} f$, $\partial f(x)$ is nonempty and compact, and

$$f'(x; y) = \max \{ \langle x^*, y \rangle \mid x^* \in \partial f(x) \} = \delta^*(y \mid \partial f(x)) \tag{2.7}$$

for each $y \in R^n$. In general, $\partial f(x)$ is empty if and only if $f'(x: y) = -\infty$ for some y.

When $\partial f(x)$ is nonempty, one necessarily has $(\mathrm{cl} f)(x) = f(x)$. On the other hand, when $(\mathrm{cl} f)(x) = f(x)$ one has $x^* \in \partial f(x)$ if and only if $x \in \partial f^*(x^*)$. Thus the multivalued mapping $\partial f^*: x^* \to \partial f^*(x^*)$ is the inverse of the multivalued mapping $\partial f: x \to \partial f(x)$, when f is a closed proper convex function.

Note that $0 \in \partial f(x)$ if and only if f attains its minimum (over R^n) at x. We shall use this fact later in a slightly different form: when $(\mathrm{cl} f)(0) = f(0)$, the vectors x^* in $\partial f(0)$ are the same as those for which $0 \in \partial f^*(x^*)$, i.e. for which f^* attains its minimum.

The conjugate of a differentiable convex function f on R^n is closely related to the Legendre transform of f. Let C^* be the set of all gradients x^* of f, i.e. the image of R^n under the mapping $x \to \nabla f(x)$. Given any $x^* \in C^*$, the vectors x for which the supremum in (2.4) is attained are precisely those for which $x^* = \nabla f(x)$; thus

$$f^*(x^*) = \langle x, x^* \rangle - f(x) \text{ when } x^* = \nabla f(x). \tag{2.8}$$

If the mapping ∇f is one-to-one, we get

$$f^*(x^*) = \langle (\nabla f)^{-1}(x^*), x^* \rangle - f((\nabla f)^{-1}(x^*)), \quad x^* \in C^*. \tag{2.9}$$

The right side of (2.9) is the formula for the *Legendre transform* of f.

If ∇f is not one-to-one, we can still conceive of parameterizing C^* in terms of x by means of the nonlinear substitution $x^* = \nabla f(x)$; the substitution yields the formula

$$f^*(\nabla f(x)) = \langle x, \nabla f(x) \rangle - f(x). \tag{2.10}$$

This function of x is one which is commonly mentioned in the literature of nonlinear programming. It is generally not convex, of course, and it generally does not express f^* completely, since it only gives the values of f^* on C^*. The set C^* need not be convex in R^n, and there may be points outside of C^* where f^* is finite but the Legendre transform is undefined.

It will be convenient in what follows to place concave functions on an equal footing with convex functions. A function g from R^n to $[-\infty, +\infty]$ is said to be *concave*, of course, if $f = -g$ is convex. All the above facts and definitions for convex functions have obvious analogues for concave functions, in which the roles of $+\infty$, inf and $\leqq$ are interchanged with those of $-\infty$, sup and

$\geqq$. In particular, the conjugate of concave function g is defined by

$$g^*(x^*) = \inf\{\langle x, x^*\rangle - g(x) \mid x \in R^n\}.$$

It should be noted that g^* is *not* the same as $-f^*$, where $f = -g$. Instead one has $g^*(x^*) = -f^*(-x^*)$.

3. **Dual programs and adjoint bifunctions.** By a *convex bifunction* from R^m to R^n, we shall mean a correspondence F which assigns to each $u \in R^m$ a function Fu from R^n $(= Su)$ to $[-\infty, +\infty]$, such that $(Fu)(x)$ is a (jointly) convex function of (u, x) on R^{m+n}. This function on R^{m+n} is called the *graph function* of F. We shall say that F is closed, or proper, according to whether its graph function is closed or proper, respectively. The *effective domain* of F is defined to be the convex set which is the projection on R^m of the effective domain of the graph function of F, i.e.

$$\operatorname{dom} F = \{u \mid \exists x, (Fu)(x) < +\infty\}.$$

If F is closed, proper and convex, then in particular Fu is a closed convex function on R^n for every u, proper for $u \in \operatorname{dom} F$ but identically $+\infty$ for $u \notin \operatorname{dom} F$.

For example, let $f_0, f_1, \cdots, f_m$ be finite convex functions on R^n, and for each $u = (u_1, \cdots, u_m)$ define the function Fu by

$$\begin{aligned}(Fu)(x) &= f_0(x) \quad \text{if } f_1(x) \leqq u_1, \cdots, f_m(x) \leqq u_m, \\ &= +\infty \quad \text{otherwise.}\end{aligned} \tag{3.1}$$

It is easily demonstrated that F is a closed proper convex bifunction. Note that $\operatorname{dom} F$ consists of the vectors u such that the corresponding inequality system

$$f_1(x) \leqq u_1, \cdots, f_m(x) \leqq u_m$$

has at least one solution x.

For another example, let A be a linear transformation from R^m to R^n and let

$$\begin{aligned}(Fu)(x) &= 0 \quad \text{if } x = Au, \\ &= +\infty \quad \text{if } x \neq Au.\end{aligned} \tag{3.2}$$

This F is a closed proper convex bifunction which we call the *indicator bifunction* of A. We shall see that the "convex algebra" below reduces to ordinary linear algebra when the bifunctions are taken to be such indicator bifunctions.

Henceforth we assume for simplicity that F is a certain closed proper convex bifunction from R^m to R^n.

The program (P) associated with F, as in the introduction, is that of minimizing $F0$ on R^n. Of course, minimizing $F0$ on R^n is equivalent to minimizing $F0$ over the convex set $\operatorname{dom}(F0)$, since $F0$ has only the value $+\infty$ outside this set. The elements of $\operatorname{dom}(F0)$ will be called the *feasible solutions* to (P). This is suggested by the case of (P) where F is given by (3.1), which we refer to as the case of an *ordinary convex program*. Feasible solutions to (P) exist if and only if $0 \in \operatorname{dom} F$, in which event we say (P) is *consistent*. If 0 is actually an interior point of $\operatorname{dom} F$, we say (P) is *strictly consistent*. In the case of an ordinary convex program, (P) is strictly consistent if and only if there exists an x such that $f_i(x) < 0$ for $i = 1, \cdots, m$.

The fundamental and easily proved fact on which our analysis of (P) depends is that the extended-real-valued function $\inf F$ on R^m defined by

$$(\inf F)(u) = \inf(Fu) = \inf\{(Fu)(x) \mid x \in R^n\}$$

is a *convex* function (not necessarily proper) whose effective domain is the same as $\operatorname{dom} F$. The theory of closures, conjugates, directional derivatives and subgradients of convex functions can therefore be applied to the study of $\inf F$ at $u = 0$.

For example, if (P) is strictly consistent, 0 is in the interior of the effective domain of $\inf F$, so we may conclude at once that $(\inf F)(u)$ depends continuously on u for sufficiently small perturbations u.

Assume that $(\inf F)(0)$ is finite. By definition, u^* is a *Kuhn-Tucker vector* for (P) if and only if

$$(\inf F)(u) \geqq (\inf F)(0) - \langle u, u^* \rangle, \quad \forall u \in R^m.$$

in other words if $-u^*$ is a subgradient of $\inf F$ at 0, i.e.

$$-u^* \in \partial(\inf F)(0). \tag{3.3}$$

If (P) is strictly consistent, so that 0 is an interior point of $\operatorname{dom}(\inf F)$, we know from the general theory that $\partial(\inf F)(0)$ is a nonempty compact convex set in R^m whose support function is the directional derivative function

$$(\inf F)'(0; u) = \lim_{\lambda \downarrow 0} \frac{(\inf F)(\lambda u) - (\inf F)(0)}{\lambda}. \tag{3.4}$$

In particular, a Kuhn-Tucker vector $u^* = (u_1^*, \cdots, u_m^*)$ does exist when (P) is strictly consistent. This u^* is unique if and only if $\inf F$ is actually differentiable at 0, in which case one has

$$u_i^* = \frac{-\partial}{\partial u_i}(\inf F)(0), \quad i = 1, \cdots, m. \tag{3.5}$$

(Thus, for example, in an ordinary convex program the "equilibrium" values of the Lagrange multipliers, if unique, give the rates of change of the infimum in the program with respect to changes of the constant terms in the corresponding constraint inequalities.)

By the general theory of subgradients, a Kuhn-Tucker vector fails to exist for (P) if and only if there exists a u such that $(\inf F)'(0; u) = -\infty$. The interpretation of this case is that there is some direction of perturbation in which the "minimal cost" drops off infinitely steeply, so that no finite "prices" for the perturbation variables can bring about a state of equilibrium.

To get the program which is dual to (P), we need to introduce the *adjoint* of the convex bifunction F. This is the bifunction F^* from R^n to R^m given by $x^* \to F^*x^*$, where

$$\begin{aligned}(F^*x^*)(u^*)& \\ &= \inf\{(Fu)(x) - \langle x, x^*\rangle + \langle u, u^*\rangle \mid u \in R^m, x \in R^n\}.\end{aligned} \tag{3.6}$$

Note that, in terms of the graph function f of F, one has

$$\begin{aligned}(F^*x^*)(u^*) &= -\sup_{u,x}\{\langle u, -u^*\rangle + \langle x, x^*\rangle - f(u,x)\} \\ &= -f^*(-u^*, x^*),\end{aligned}$$

where f^* is the conjugate of f on R^{m+n}. Thus F^* is a *closed proper concave* bifunction in the obvious sense.

The adjoint of a concave bifunction is defined as in (3.6), except of course that "sup" replaces "inf". Thus the adjoint F^{**} of F^* is defined in turn by

$$\begin{aligned}(F^{**}u)(x) &= \sup\{(F^*x^*)(u^*) - \langle u, u^*\rangle + \langle x, x^*\rangle \mid x^* \in R^n, u^* \in R^m\} \\ &= \sup_{u^*,x^*}\{\langle u, u^*\rangle + \langle x, x^*\rangle - f^*(u^*, x^*)\} = f^{**}(u, x).\end{aligned}$$

Since $f^{**} = f$ under the conjugacy correspondence, we have $F^{**} = F$.

It is easy to see that, when F is the convex indicator bifunction of a linear transformation A from R^m to R^n, F^* is the concave indicator bifunction of the adjoint linear transformation A^* from R^n back to R^m (corresponding to the transpose matrix), i.e. $(F^*x^*)(u^*)$ is 0 if $u^* = A^*x^*$ and $-\infty$ if $u^* \neq A^*x^*$. In this sense, the adjoint operation for bifunctions generalizes the one for linear transformations. Further justification of the "adjoint" terminology will be given in the next section.

We define the dual program (P*) to be that of maximizing the concave function F^*0 on R^m. In (P*) we are also interested in the properties of the function $\sup F^*$ at $x^* = 0$, where $\sup F^*$ is the function on R^n defined by

$$(\sup F^*)(x^*) = \sup(F^*x^*) = \sup\{(F^*x^*)(u^*) \mid u^* \in R^m\}.$$

Thus x^* is taken to be the perturbation variable in (P*), while u^* is the vector variable over which one maximizes. Of course, $\sup F^*$ turns out to be a concave function. Everything that has been said about $\inf F$ in (P) applies to $\sup F^*$ in (P*) with only the obvious changes. The dual of the generalized program (P^*) is in turn (P), inasmuch as $F^{**} = F$.

As an example, let A be a linear transformation from R^n to R^m, fix $a \in R^m$ and $a^* \in R^n$, and set

$$(3.7)\qquad \begin{aligned}(Fu)(x) &= \langle x, a^*\rangle, \quad \text{if } x \geqq 0 \text{ and } Ax \geqq a - u,\\ &= +\infty, \quad \text{otherwise.}\end{aligned}$$

(This is the case of (3.1) where the functions f_i are all affine.) Minimizing $F0$ in (P) is then the same as minimizing $\langle x, a^*\rangle$ subject to $x \geqq 0$ and $Ax \geqq a$, so (P) is a typical linear program. By a straightforward calculation from the definition of F^*,

$$(3.8)\qquad \begin{aligned}(F^*x^*)(u^*) &= \langle a, u^*\rangle, \quad \text{if } u^* \geqq 0 \text{ and } A^*u^* \leqq a^* - x^*,\\ &= -\infty, \quad \text{if not.}\end{aligned}$$

Maximizing F^*0 in (P*) is the same as maximizing $\langle a, u^*\rangle$ subject to $u^* \geqq 0$ and $A^*u^* \leqq a^*$. Thus (P*) is the dual linear program.

The dual programs of Fenchel, extended by the author in [**43**], may also be represented as a special case of the above. Again let A be a linear transformation from R^n to R^m, let F be a closed proper convex function on R^n and let g be a closed proper concave function on R^m. Define F by

$$(3.9)\qquad (Fu)(x) = f(x) - g(Ax + u).$$

Then F is a closed proper convex bifunction, and (P) consists of minimizing $f(x) - g(Ax)$ in $x \in R^n$. Note that the perturbation u here corresponds to a translation of the function g on R^m. By elementary calculation,

$$(3.10) \qquad (F^*x^*)(u^*) = g^*(u^*) - f^*(A^*u^* + x^*),$$

so that (P*) consists of maximizing $g^*(u^*) - f^*(A^*u^*)$ in $u^* \in R^m$. Fenchel's original programs are obtained by taking A to be the identity transformation.

For an ordinary convex program, the adjoint bifunction is given by

$$\begin{aligned}(F^*x^*)(u^*) &= -(f_0 + u_1^*f_1 + \cdots + u_m^*f_m)^*(x^*), \\ &\qquad\qquad \text{if } u^* = (u_1^*, \cdots, u_m^*) \geqq 0, \\ &= -\infty, \quad \text{if } u^* \ngeqq 0.\end{aligned}$$

Thus the dual program (P*) is to maximize $-(f_0 + u_1^*f_1 + \cdots + u_m^*f_m)^*(0)$ subject to $u_i^* \geqq 0$, $i = 1, \cdots, m$. To calculate the conjugate of $f = f_0 + u_1^*f_1 + \cdots + u_m^*f_m$ explicitly, one would have to know more about the given functions f_i. However, if every f_i is differentiable one can apply the Legendre transformation in the weakened form of (2.10) to f to get a problem which is "almost" equivalent to (P*). Since $-f^*(\nabla f(x)) = f(x)$ by (2.10) when $\nabla f(x) = 0$, and

$$\nabla f = \nabla f_0 + u_1^* \nabla f_1 + \cdots + u_m^* \nabla f_m,$$

the problem is essentially that of maximizing

$$f_0(x) + u_1^*f_1(x) + \cdots + u_m^*f_m(x)$$

in $u^* \in R^m$ and $x \in R^n$ subject to the constraints

$$u^* \geqq 0, \quad \nabla f_0(x) + u_1^* \nabla f_1(x) + \cdots + u_m^* \nabla f_m(x) = 0.$$

This is the well-known dual problem which was discovered by Wolfe [47].

It should be pointed out that an ordinary convex program can be modified in many ways by introducing additional perturbations. For instance, one can perturb the constraint $f_i(x) \leqq u_i$ by a translation y_i to the constraint $f_i(x - y_i) \leqq u_i$. The dual problem would then turn out to involve an additional Lagrange multiplier vector $y_i^* \in R^n$ dual to the perturbation vector y_i. This would essentially be the dual problem for the ordinary convex program given by the author in [40]. The possibilities for perturbation are endless. The perturbations can be chosen to suit the situation, according to what "equilibrium prices" one is interested in. To apply the duality theory described here, it is only necessary that the perturbations be "convex", in the sense that the dependence of the problem on the perturbations be representable in terms of a convex bifunction F.

All the results relating the general dual pair of programs (P) and (P*) are based on one elementary fact, which follows directly from the definitions: the convex minimand $F0$ in (P) is the conjugate of the convex function $-\sup F^*$ on R^n, while the concave maximand F^*0 in (P*) is the conjugate of the concave function $-\inf F$ on R^m. This implies that

$$(F0)^* = (-\sup F^*)^{**} = -\operatorname{cl}(\sup F^*),$$
$$(F^*0)^* = (-\inf F)^{**} = -\operatorname{cl}(\inf F),$$

and hence that

$$\begin{aligned} \operatorname{cl}(\sup F^*)(0) &= -\sup_{x\in R^n}\{\langle x,0\rangle - (F0)(x)\} = (\inf F)(0), \\ \operatorname{cl}(\inf F)(0) &= -\inf_{u^*\in R^m}\{\langle 0,u^*\rangle - (F^*0)(u^*)\} = (\sup F^*)(0). \end{aligned} \tag{3.12}$$

The infimum $(\inf F)(0)$ in (P) is thus always greater than or equal to the supremum $(\sup F^*)(0)$ in (P*), and any possible discrepancy between these extrema is completely explained in terms of the closure operations for convex and concave functions.

Let us call (P) *normal* if $\operatorname{cl}(\inf F)(0) = (\inf F)(0)$. If (P) is consistent, this is equivalent to the semicontinuity condition that

$$\liminf_{u\to 0}(\inf F)(u) = (\inf F)(0).$$

Similarly, let us call (P*) normal if $\operatorname{cl}(\sup F^*)(0) = (\sup F^*)(0)$ in the sense of the closure operation for concave functions. Formulas (3.12) then yield a duality theorem: (P) *is normal if and only if* (P*) *is normal. Moreover, the normal case is precisely the one where the extrema in* (P) *and* (P*) *are equal,* i.e.

$$(\inf F)(0) = (\sup F^*)(0). \tag{3.13}$$

For brevity, we shall say that *normality holds* when both programs are normal and the "inf" and "sup" are equal. Normality holds in particular, then, when (P) is strictly consistent (since then $\inf F$ is continuous at 0), or when a Kuhn-Tucker vector exists for (P) (since then $\partial(\inf F)(0) \neq \emptyset$, implying that $\operatorname{cl}(\inf F)$ agrees with $\inf F$ at 0). Likewise, normality holds when (P*) is strictly consistent, or when a Kuhn-Tucker vector exists for (P*).

Suppose that normality holds, and that the common extremum value in (3.13) is finite. As we have already pointed out, u^* is a Kuhn-Tucker vector for (P) if and only if $u^* \in \partial(-\inf F)(0)$. Since $(-\inf F)^* = F^*0$, this is equivalent to the condition that $0 \in \partial(F^*0)(u^*)$, i.e. that the concave function F^*0 attain its maximum at u^*. Similarly, the Kuhn-Tucker vectors x for (P*) are the vectors where the convex function $F0$ attains its minimum. This gives us another duality theorem: *assuming that normality holds, the Kuhn-Tucker vectors u^* for* (P) *are precisely the optimal solutions (if any) to* (P*), *while the optimal solutions x to* (P) *are precisely the Kuhn-Tucker vectors for* (P*). This type of duality has previously been known only in the linear programming case.

4. **Lagrangian functions and minimax theory.** We shall now describe the correspondence between convex bifunctions from R^m to R^n and concave-convex functions on $R^m \times R^n$ which is analogous to the correspondence between linear transformations from R^m to R^n and bilinear functions $R^m \times R^n$. This correspondence gives further insight into the nature of the adjoint bifunction. It enables us to construct for each dual pair of programs (P) and (P*) as in the last section a certain convex-concave function whose saddle-points correspond to optimal solutions to the programs, much as in the classical Kuhn-Tucker theory [**24**].

Let K be a concave-convex function on $R^m \times R^n$, i.e. a function with values in $[-\infty, +\infty]$ such that $K(u,v)$ is concave in u for each v and convex in v for each u. Closure operations may be applied to K for the sake of regularization. Let $\mathrm{cl}_v K$ be the function on $R^m \times R^n$ obtained by closing $K(u,v)$ as a convex function of v for each u. Similarly let $\mathrm{cl}_u K$ denote the function obtained by closing K as a concave function of u for each v. Then $\mathrm{cl}_u K$ and $\mathrm{cl}_v K$ are concave-convex functions on $R^m \times R^n$ [**35**].

We can proceed now to form the concave-convex functions $\mathrm{cl}_v(\mathrm{cl}_u K)$ and $\mathrm{cl}_u(\mathrm{cl}_v K)$. The first of these is called the *lower closure* of K (since the final regularization involves lower semicontinuity), and the second is called the *upper closure* of K. If K coincides with its lower closure, it is said to be *lower closed*, and so forth. It turns out that $\mathrm{cl}_v(\mathrm{cl}_u K)$ is itself always lower closed, and $\mathrm{cl}_u(\mathrm{cl}_v K)$ is upper closed, but these two functions may disagree at certain points of $R^m \times R^n$.

Since the operations $\mathrm{cl}_v \mathrm{cl}_u$ and $\mathrm{cl}_u \mathrm{cl}_v$ do not quite produce the same result, there is not a *unique* natural closure operation for concave-convex functions. Nevertheless, there is an important phenomenon of pairing of closures. It may be shown that, if $\underline{K}$ is any lower closed concave-convex function on $R^m \times R^n$, then $\overline{K} = \mathrm{cl}_u \underline{K}$ is an upper closed concave-convex function such that $\mathrm{cl}_v \overline{K} = \underline{K}$. Thus there is a simple one-to-one correspondence between the lower closed functions and the upper closed functions. Corresponding functions $\underline{K}$ and $\overline{K}$ cannot differ very greatly from each other, since the closure operations for convex and concave functions only redefine functions at special points.

For example, let C and D be closed convex sets in R^m and R^n, respectively, and let K be any continuous real-valued concave-convex function defined on $C \times D$. Set

$$\underline{K}(u,v) = \begin{cases} K(u,v) & \text{if } u \in C \text{ and } v \in D, \\ +\infty & \text{if } u \in C \text{ and } v \notin D, \\ -\infty & \text{if } u \notin C, \end{cases}$$

$$\overline{K}(u,v) = \begin{cases} K(u,v) & \text{if } u \notin C \text{ and } v \in D, \\ +\infty & \text{if } v \notin D, \\ -\infty & \text{if } u \notin C \text{ and } v \in D. \end{cases} \tag{4.1}$$

Then $\underline{K}$ and $\overline{K}$ are lower closed and upper closed concave-convex functions, respectively, which are paired together in the manner just described. Observe, incidentally, that

$$\sup_{u \in R^m} \inf_{v \in R^n} \underline{K}(u,v) = \sup_{u \in R^m} \inf_{v \in R^n} \overline{K}(u,v) = \sup_{u \in C} \inf_{v \in D} K(u,v),$$

$$\inf_{v \in R^n} \sup_{u \in R^m} \underline{K}(u,v) = \inf_{v \in R^n} \sup_{u \in R^m} \overline{K}(u,v) = \inf_{v \in D} \sup_{u \in C} K(u,v).$$

Thus the minimax analysis of K with respect to $C \times D$ can be represented by the formally unconstrained minimax analysis of $\underline{K}$ or of $\overline{K}$ (or of any extension of K to all of $R^m \times R^n$ such that $\underline{K} \leqq K \leqq \overline{K}$).

In order to apply these facts to the study of bifunctions in a manner suggestive of linear algebra, we introduce an inner product notation for the conjugate of a convex (or concave) function f:

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

This inner product is a true generalization of the ordinary one, in the following sense: if f is the indicator of a vector $x \in R^n$, i.e. if

$$f(y) = \delta(y\,|\,x) = 0 \qquad \text{if } y = x,$$
$$= +\infty \qquad \text{if } y \neq x,$$

then $\langle f, x^*\rangle = \langle x, x^*\rangle$. Incidentally, by means of Fenchel's Duality Theorem it is possible [44] to generalize further to inner products of the form $\langle f, g\rangle$, where f is a convex function on R^n and g is a concave function on R^n. We shall not need this degree of generalization here, however.

For any convex bifunction from R^m to R^n, we can form

$$\langle Fu, x^*\rangle = \langle x^*, Fu\rangle = (Fu)^*(x^*) \tag{4.2}$$

as a function of $u \in R^m$ and $x^* \in R^n$. Note that, if F is the indicator bifunction of a linear transformation $A\colon R^m \to R^n$ as in (3.1), then $\langle Fu, x^*\rangle$ is simply the bilinear function $\langle Au, x^*\rangle$ associated with A.

The basic theorem is the following. If F is any closed convex bifunction from R^m to R^n, then $\langle Fu, x^*\rangle$ is a lower closed concave-convex function on $R^m \times R^n$. Conversely, given any function $\underline{K}$ of the latter type, there exists a unique closed convex bifunction F from R^m to R^n such that $\underline{K}(u, x^*) = \langle Fu, x^*\rangle$, namely the F given by

$$(Fu)(x) = \sup\{\langle x, x^*\rangle - \underline{K}(u, x^*)\,|\, x^* \in R^n\}.$$

The upper closed $\overline{K}$ on $R^m \times R^n$ paired with $\underline{K}$ is precisely the concave-convex function associated with the adjoint bifunction F^*, i.e.

$$\overline{K}(u, x^*) = \langle u, F^*x^*\rangle = (F^*x^*)^*(u).$$

Thus the formulas

$$\begin{aligned} \mathrm{cl}_u \langle Fu, x^*\rangle &= \langle u, F^*x^*\rangle, \\ \langle Fu, x^*\rangle &= \mathrm{cl}_{x^*} \langle u, F^*x^*\rangle, \end{aligned} \tag{4.3}$$

hold for any closed convex bifunction and its adjoint.

Formulas (4.3) generalize the familiar formula

$$\langle Au, x^*\rangle = \langle u, A^*x^*\rangle$$

relating a linear transformation and its adjoint. Since the closure operations in (4.3) merely redefine the functions at special points, one will actually have

$$\langle Fu, x^*\rangle = \langle u, F^*x^*\rangle \tag{4.4}$$

for "most" values of u and x^*.

Observe that (4.4) expresses a duality between two different extremum problems, because by definition any convex bifunction F and its (concave) adjoint F^* satisfy

$$(4.5)\qquad \begin{aligned} \langle Fu, x^* \rangle &= \sup_x \{ \langle x, x^* \rangle - (Fu)(x) \}, \\ \langle u, F^*x^* \rangle &= \inf_{u^*} \{ \langle u, u^* \rangle - (F^*x^*)(u^*) \}. \end{aligned}$$

In particular, we have

$$(4.6)\qquad \begin{aligned} -\langle Fu, 0 \rangle &= \inf_x (Fu)(x) = (\inf F)(u), \\ -\langle 0, F^*x^* \rangle &= \sup_{u^*} (F^*x^*)(u^*) = (\sup F^*)(x^*). \end{aligned}$$

The equality of the extrema in the programs (P) and (P*) in the last section is therefore expressed simply by

$$\langle F0, 0 \rangle = \langle 0, F^*0 \rangle .$$

Minimax characterizations of duality are obtained through the introduction of inverse bifunctions. The *inverse* of a convex bifunction F from R^m to R^n is the concave bifunction F_* from R^n to R^m defined by

$$(4.7)\qquad (F_* x)(u) = -(Fu)(x).$$

If F is closed, F_* is closed. The inverse of a concave bifunction is also defined by (4.7). It is easily seen that $F_{**} = F$ and $(F^*)_* = (F_*)^*$. The latter bifunction from R^m to R^n will be denoted simply by F^*_*.

As an example, if $m = n$ and F is the convex indicator bifunction of a one-to-one linear transformation A from R^n onto R^n as in (3.2), then F_* is the concave indicator bifunction of A^{-1}, i.e. $(F_* x)(u)$ is 0 if $u = A^{-1}x$ and $-\infty$ if $u \neq A^{-1}x$. Likewise, F^*_* is the convex indicator bifunction of A^{*-1}.

Given any closed proper convex bifunction F from R^m to R^n, we define the *Lagrangian function* of the corresponding program (P) to be $\langle u^*, F_* x \rangle$ as a function of u^* and x. Since $F_* x$ is concave, we have by definition

$$(4.8)\qquad \begin{aligned} \langle u^*, F_* x \rangle &= \inf_u \{ \langle u, u^* \rangle - (F_* x)(u) \} \\ &= \inf_u \{ \langle u, u^* \rangle + (Fu)(x) \}. \end{aligned}$$

This is, of course, an upper closed concave-convex function on $R^m \times R^n$ by the correspondence theory already outlined.

In the case of an ordinary convex program, where F is given by (3.1), the Lagrangian is evidently given by

$$(4.9)\qquad \begin{aligned}\langle u^*, F_*x\rangle &= f_0(x) + u_1^* f_1(x) + \cdots + u_m^* f_m(x) \\ &\qquad\qquad \text{if } u^* = (u_1^*, \cdots, u_m^*) \geqq 0, \\ &= -\infty \quad \text{if } u^* \ngeqq 0.\end{aligned}$$

Except for the convenient concave extension by means of $-\infty$, this is the Lagrangian associated with (P) by the familiar Kuhn-Tucker theory.

In the case where F is given by (3.9), the Lagrangian function is given by

$$(4.10)\qquad \begin{aligned}\langle u^*, F_*x\rangle &= f(x) + g^*(u^*) - \langle Ax, u^*\rangle \quad \text{if } f(x) < +\infty, \\ &= +\infty \quad \text{if } f(x) = +\infty.\end{aligned}$$

A *saddle-point* of the Lagrangian function is, of course, defined to be a vector pair (u^*, x) such that

$$(4.11)\qquad \langle u^{*\prime}, F_*x\rangle \leqq \langle u^*, F_*x\rangle \leqq \langle u^*, F_*x'\rangle, \ \forall\, u^{*\prime}, \ \forall\, x'.$$

The main result is this: *a vector pair (u^*, x) is a saddle-point of the Lagrangian of* (P) *if and only if u^* is a Kuhn-Tucker vector for* (P) *and x is an optimal solution to* (P). In this event normality holds, and the minimax value $\langle u^*, F_*x\rangle$ coincides with the infimum in (P) and the supremum in (P*). Moreover, as explained in the last section, u^* is then dually an optimal solution to (P*), and x is a Kuhn-Tucker vector for (P*).

Given any upper closed concave-convex function $\overline{K}$ on $R^m \times R^n$ (for instance a $\overline{K}$ of the type in (4.1)), there is, as we know, a unique closed concave bifunction G from R^n to R^m such that $\overline{K}(u^*, x) = \langle u^*, Gx\rangle$. Hence there is a unique program (P) having $\overline{K}$ as its Lagrangian, namely the (P) corresponding to $F = G_*$. The inverse operation for bifunctions therefore corresponds to a general minimax theory for concave-convex functions in the same way that the adjoint operation for bifunctions corresponds to a general duality theory for convex programs. It is clear from the definitions that F and F^* are expressible here in terms of $\overline{K}$ by

$$(4.12)\qquad \begin{aligned}(Fu)(x) &= \sup_{u^*} \{\overline{K}(u^*, x) - \langle u, u^*\rangle\}, \\ (F^*x^*)(u^*) &= \inf_{x} \{\overline{K}(u^*, x) - \langle x, x^*\rangle\}.\end{aligned}$$

In particular, the minimand in (P) is given by

$$(F0)(x) = \sup_{u^*} \overline{K}(u^*, x),$$

and the maximand in (P*) is given by

$$(F^*0)(u^*) = \inf_x \overline{K}(u^*, x).$$

The dual programs of Dantzig, Eisenberg and Cottle [**6**], Stoer [**45**], Mangasarian and Ponstein [**26**], may be obtained in this way, for instance by applying the Legendre transformation formula (2.10) to (4.12) and similar devices; see [**38**].

The pair of functions $\langle F^*_* u^*, x\rangle$, $\langle u^*, F_* x\rangle$, is *conjugate* to the pair of functions $\langle Fu, x^*\rangle$, $\langle u, F^* x^*\rangle$, in the following sense. If K is any one of the concave-convex functions such that

$$\langle Fu, x^*\rangle \leqq K(u, x^*) \leqq \langle u, F^* x^*\rangle \tag{4.13}$$

(such functions all being essentially the same), one has, according to the definitions,

$$\begin{aligned} &\inf_{u \in R^m} \sup_{x^* \in R^n} \{ \langle u, u^*\rangle + \langle x, x^*\rangle - K(u, x^*) \} = \langle u^*, F_* x\rangle, \\ &\sup_{x^* \in R^n} \inf_{u \in R^m} \{ \langle u, u^*\rangle + \langle x, x^*\rangle - K(u, x^*) \} = \langle F^*_* u^*, x\rangle. \end{aligned} \tag{4.14}$$

On the other hand, if K^* is any one of the functions satisfying

$$\langle F^*_* u^*, x\rangle \leqq K^*(u^*, x) \leqq \langle u^*, F_* x\rangle, \tag{4.15}$$

one has in turn

$$\begin{aligned} &\inf_{u^* \in R^m} \sup_{x \in R^n} \{ \langle u, u^*\rangle + \langle x, x^*\rangle - K^*(u^*, x) \} = \langle u, F^* x^*\rangle, \\ &\sup_{x \in R^n} \inf_{u^* \in R^m} \{ \langle u, u^*\rangle + \langle x, x^*\rangle - K^*(u^*, x) \} = \langle Fu, x^*\rangle. \end{aligned} \tag{4.16}$$

Applying (4.3) to the convex bifunction F^*_* in place of F, we have

$$\begin{aligned} \mathrm{cl}_{u^*} \langle F^*_* u^*, x\rangle &= \langle u^*, F_* x\rangle, \\ \langle F^*_* u^*, x\rangle &= \mathrm{cl}_x \langle u^*, F_* x\rangle, \end{aligned} \tag{4.17}$$

and this makes possible a detailed comparison of the "inf sup" and "sup inf" in (4.14). In particular we see that these two extrema are "usually" equal; the fact that they can be different in some cases is exactly dual to the fact that the upper and lower closure operations for concave-convex functions do not always coincide. A minimax theory from this point of view has been developed by the author in [**35**].

References

1. K. J. Arrow, L. Hurwicz and H. Uzawa, *Studies in linear and nonlinear programming*, Stanford Univ. Press, Stanford, Calif., 1958.

2. C. Berge and A. Ghouila-Houri, *Programmes, jeux et réseaux de transport*, Dunod, Paris, 1962.

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

4. 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.

5. R. W. Cottle, *Symmetric dual quadratic programs*, Quart. Appl. Math. **21** (1963), 237.

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

7. J. B. Dennis, *Mathematical programming and electrical networks*, Technology Press, Cambridge, Mass., 1959.

8. U. Dieter, *Dualität bei konvexen Optimierungs-(Programmierungs-) Aufgaben*, Unternehmensforschung **9** (1965), 91-111.

9. ———, *Optimierungsaufgaben in topologischen Vektorräumen* I: *Dualitätstheorie*, Z. Wahrscheinlichkeitstheorie Verw. Geb. **5** (1966), 89-117.

10. R. J. Duffin, *Infinite programs*, Ann. Math. Study 38 (1956), 157-170.

11. W. S. Dorn, *Duality in quadratic programming*, Quart. App. Math. **18** (1960), 155-162.

12. ———, *A duality theorem for convex programs*, IBM J. Res. Develop. **4** (1960), 407-413.

13. R. J. Duffin, *Dual programs and minimum cost*, J. Soc. Indust. Appl. Math. **10** (1962), 119-124.

14. E. Eisenberg, *Duality in homogeneous programming*, Proc. Amer. Math. Soc. **12** (1961), 783-787.

15. J. E. Falk and R. M. Thrall, *A constrained Lagrangian approach to nonlinear programming*, Technical report, Dept. of Math., Univ. of Michigan, Ann Arbor, 1965.

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

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

18. ———, *Convex cones, sets and functions*, Lecture notes, Princeton Univ., 1951.

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

20. M. A. Hanson, *A duality theorem in nonlinear programming with nonlinear constraints*, Austral. J. Statist. **3** (1961), 64-72.

21. P. Huard, *Dual programs*, IBM J. Res. Develop. **6** (1962), 137-139.

22. F. John, "Extremum problems with inequalities as subsidiary conditions" in *Studies and essays, Courant Anniversary Volume*, Interscience, New York, 1948.

23. S. Karlin, *Mathematical methods and theory in games, programming and economics*, McGraw-Hill, New York, 1960.

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

25. O. L. Mangasarian, *Duality in nonlinear programming*, Quart. Appl. Math. **20** (1962), 300-302.

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

27. J.-J. Moreau, *Fonctions convexes en dualité*, multigraph, Seminaires de Math., Faculté des Sciences, Université de Montpellier, 1962.

28. ———, *Etude locale d'une fonctionelle convexe,* multigraph, Seminaires de Math., Faculté des Sciences, Université de Montpellier, 1962.

29. ———, *Fonctionelles sous-differentiables,* C. R. Acad. Sci. **257** (1963), 4117-4119.

30. ———, *Théorèmes 'inf-sup',* C. R. Acad. Sci. **258** (1964), 2720-2722.

31. ———, *Proximité et dualité dans une espace hilbertien,* Bull. Soc. Math. France **93** (1965), 273-299.

32. ———, *Fonctionelles convexes,* Lecture notes, Seminaire "Equations aux dérivées partielles", College de France, 1966.

33. R. T. Rockafellar, *Convex functions and dual extremum problems,* Thesis, Harvard, 1963.

34. ———, *Duality theorems for convex functions,* Bull. Amer. Math. Soc. **70** (1964), 189-192.

35. ———, *Minimax theorems and conjugate saddle-functions,* Math. Scand. **14** (1964), 151-173.

36. ———, *Monotone processes of concave and convex type,* Mem. Amer. Math. Soc. No. 77 (1968).

37. ———, *A monotone convex analog of linear algebra,* Proceedings of the Colloquium on Convexity, Copenhagen, 1965, edited by W. Fenchel, Matematisk Institut, Copenhagen, 1967, 261-276.

38. ———, *A general correspondence between dual minimax problems and convex programs,* Pacific J. Math. **25** (1968), no. 3.

39. ———, *Helly's theorem and minima of convex functions,* Duke Math. J. **32** (1965), 381-398.

40. ———, *An extension of Fenchel's duality theorem for convex functions,* Duke Math. J. **33** (1966), 81-90.

41. ———, *Convex programming and systems of elementary monotonic relations,* J. Math. Anal. Appl. **19** (1967), 543-564.

42. ———, *Conjugates and Legendre transforms of convex functions,* Canad. J. Math. **19** (1967), 200-205.

43. ———, *Duality and stability in extremum problems involving convex functions,* Pacific J. Math. **21** (1967), 167-187.

44. ———, *Convex analysis,* in preparation for the Princeton University Press.

45. J. Stoer, *Über einen Dualitätsatz der nichtlinearen Programmierung,* Num. Math. **6** (1964), 55-58.

46. A. Whinston, "Some applications of the conjugate function theory to duality" in *Nonlinear programming,* edited by J. Abadie, North-Holland, Amsterdam, 1967, pp. 75-96.

47. P. Wolfe, *A duality theorem for nonlinear programming,* Quart. Appl. Math. **19** (1961), 239-244.

UNIVERSITY OF WASHINGTON