CONJUGATE CONVEX FUNCTIONS IN NONLINEAR PROGRAMMING

R. T. Rockafellar

Let f_o, f_1, ..., f_m be real-valued functions which are convex, but not necessarily differentiable, and consider the convex program in which $f_o(x)$ is minimized subject to the constraints $f_1(x) \leq 0$, ..., $f_m(x) \leq 0$. Real numbers λ_1, ..., λ_m are called <u>Lagrange multipliers</u> for the program if they are nonnegative and the (unconstrained) infimum of the convex function $f_o + \lambda_1 f_1 + \ldots + \lambda_m f_m$ is finite and equal to the constrained infimum of f_o.

The meaning of such Lagrange multipliers is connected with a notion of perturbation, namely where the given program is perturbed by subtracting constants u_i from the constraint functions f_i. For each $u = (u_1, \ldots, u_m) \in R^m$, let $p(u) = p(u_1, \ldots, u_m)$ denote the infimum of $f_o(x)$ subject to the constraints

$$f_1(x) \leq u_1, \ \ldots, \ f_m(x) \leq u_m. \qquad (1)$$

Then $p(0)$ is the infimum in the given "unperturbed" program. Assuming that $p(0)$ is finite, one can show easily that $u^* = (\lambda_1, \ldots, \lambda_m)$ is a Lagrange multiplier vector if and only if

$$p(u) \geq p(0) - \langle u, u^* \rangle \text{ for every } u \in R^m, \qquad (2)$$

where $\langle .,. \rangle$ denotes the ordinary inner product of two vectors in R^m. In other words, Lagrange multipliers depend only on the function p, and they describe some sort of "linearization" of p around $u = 0$.

Now it happens that p is a <u>convex function</u> on R^m, and that condition (2) can be analyzed in terms of the directional derivatives of p at 0. According to the general definition of convexity for functions whose values may be not only real numbers but $+\infty$ and $-\infty$, the convexity of p means that the <u>epigraph</u> of p (which is the set of pairs (u, α) such that $u \in R^m$, $\alpha \in R$ and $\alpha \geq p(u)$) is a convex set in R^{m+1}. Condition (2) says that the graph of the affine function

$$h(u) = p(0) - \langle u, u^* \rangle$$

is a (nonvertical) supporting hyperplane to the epigraph of p at the point $(0, p(0))$. Theorems about the existence of Lagrange multipliers may thus be deduced from theorems about the existence of supporting hyperplanes to convex sets. The convexity of p also implies that the one-sided directional derivative

$$p'(0; u) = \lim_{\lambda \downarrow 0} [p(\lambda u) - p(0)]/\lambda$$

exists for every u (assuming that $p(0)$ is finite). It turns out that, in general, u^* is a Lagrange multiplier vector if and only if

$$p'(0; u) \geq - \langle u, u^* \rangle \quad \text{for every } u \in R^m.$$

A Lagrange multiplier vector fails to exist if and only if there is some u such that $p'(0; u) = -\infty$. A unique Lagrange multiplier vector u^* exists if and only if p is differentiable at 0, in which case $u^* = -\nabla p(0)$. Thus the Lagrange multipliers λ_i, when unique, give the rate of change of the infimum in the given program with respect to perturbing the constraints to those in (1):

$$\lambda_i = - \frac{\partial p}{\partial u_i} \bigg|_{u = 0} \qquad i = 1, \ldots, m.$$

For proofs of all these facts we refer to [9, section 29].

Observe that, if we define

$$f(u, x) = \begin{cases} f_0(x) & \text{if } f_1(x) \leq u_1, \ldots, f_m(x) \leq u_m, \\ +\infty & \text{if not,} \end{cases} \tag{3}$$

then f is a convex function on R^{m+n}. Minimizing $f_0(x)$ subject to $f_i(x) \leq 0$, $i = 1, \ldots, m$, is equivalent to minimizing the function f(0, x) as x ranges over all of R^n. On the other hand, p can be expressed in terms of f too:

$$p(u) = \inf_x f(u, x). \tag{4}$$

This leads to the following concept of a generalized convex program, which is essentially contained in a recent paper of Gale [2]. Let f be any convex function on R^{m+n} (not necessarily finite everywhere). In the program (P) corresponding to f, one is to minimize f(0, x) in $x \in R^m$. This minimization problem is not taken in isolation, however, but rather in the context of a certain specified system of perturbations. The perturbations corresponding to (P) are to be those in which the objective function f(0, x) is replaced by f(u, x) for various choices of $u \in R^m$. (In the case where f is given by (3), such perturbations amount to changes of constraints as in (1).) The perturbation function for (P) is the function p defined by (4). When p(0) (the optimal value in (P)) is finite, the Lagrange multiplier vectors for (P) are defined to be the vectors $u^* \in R^m$ satisfying (2), while the optimal solutions to (P) are defined to be the vectors $x \in R^n$ at which the infimum of the convex function f(0, .) is attained.

It can be proved that the perturbation function p of a generalized convex program (P) is convex. Therefore all the results stated above for Lagrange multipliers in "classical" convex programs are also valid for Lagrange multipliers in generalized convex programs.

We shall now explain how Fenchel's theory of conjugate convex functions [3], [4], can be used to obtain a complete and symmetric duality theory for generalized convex programs. The details of this new duality theory will be published in [9].

For the moment, let p denote any convex function on R^m. The closure of p is defined to be the convex function cl p whose epigraph is the intersection of all the nonvertical closed half-spaces in R^{m+1} containing the epigraph of p. Thus cl p is the pointwise supremum of the collection of all affine (i. e. linear-plus-a-constant) functions h such that $h(u) \leq p(u)$ for every u. If cl p = p, then p is said to be closed. (It can be shown that, if p nowhere takes on the value $-\infty$, $(\mathrm{cl}\ p)(u) = p(u)$ for every u which is not a boundary point of the convex set $\{u | p(u) < \infty\}$.)

The conjugate of p is the function p^* on R^m defined by

$$p^*(u^*) = \sup_u \{\langle u, u^* \rangle - p(u)\}.$$

Fenchel proved that p^* is a closed convex function, and that $p^{**} = \mathrm{cl}\ p$. In particular, if p is closed then p is in turn the conjugate of p^* and the relationship between p and p^* is a symmetric one. The epigraph of p^* consists of the pairs $(u^*, \alpha^*) \in R^{m+1}$ such that the affine function $h(u) = \langle u, u^* \rangle - \alpha^*$ satisfies $h(u) \leq p(u)$ for every u, and this set of pairs gives a dual description of p when p is closed.

We would like to point out that, in the case where p is everywhere finite and differentiable, one has

$$p^*(u^*) = \langle u, \nabla p(u) \rangle - p(u) \quad \text{when } \nabla p(u) = u^*.$$

This fact can be used to explain the connection between duality based on Fenchel's conjugacy and the duality theories for nonlinear programming in which expressions like $\langle u, \nabla p(u) \rangle - p(u)$ are important, for instance the theory of Wolfe [10] and the theory of Dantzig, Eisenberg and Cottle [11]. We must forego explaining the connection here, however, and refer instead to the exposition in [8] and [9].

Given a generalized convex program (P) corresponding as above to a closed convex function f on R^{m+n}, we define the dual program (P^*) as follows. Let f^* be the conjugate of f and let

$$g(x^*, u^*) = -f^*(-u^*, x^*).$$

This g is a concave function, and (P^*) is the generalized concave program corresponding to g. Thus in (P^*) one is to maximize the function $g(0, u^*)$ in $u^* \in R^m$. The perturbations considered in (P^*) are those which replace the objective function $g(0, u^*)$ by $g(x^*, u^*)$ for various choices of $x^* \in R^n$. The perturbation function in (P^*) is the concave function q on R^n defined by

$$q(x^*) = \sup_{u^*} g(x^*, u^*).$$

When the optimal value $q(0)$ in (P^*) is finite, the Lagrange multiplier vectors for (P^*) are defined to be the vectors $x \in R^n$ such that

$$q(x^*) \leq q(0) - \langle x, x^* \rangle \quad \text{for every } x^* \varepsilon R^n,$$

while the optimal solutions to (P*) are defined to be the vectors $u^* \varepsilon R^m$ at which the supremum of the concave function g(0, .) is attained.

The conjugate of a concave function is defined by the same formula as the conjugate of a convex function, except that "inf" replaces "sup". It can be seen from the symmetry of the conjugacy correspondence that

$$f(u, x) = - g^*(- x, u),$$

so that (P) is in turn the dual of (P*). The duality between (P) and (P*) is thus symmetric.

The dual pairs of minimization and maximization problems which one obtains by this scheme include not only the familiar linear programs, but in essence all the other pairs of (convex or concave) extremum problems for which substantial duality theories are presently known. Details and examples may be found in [7], [8] and [9].

Theorems about the duality between (P) and (P*) are based on the following fact: the objective function g(0, .) in (P*) is the conjugate of the concave function - p, whereas the objective function f(0, .) in (P) is the conjugate of the convex function - q.

We shall say that (P) is normal if cl p agrees with p at the origin. Normality of (P*) is defined similarly. It can be demonstrate that generalized convex or concave programs are always normal, unle they are quite freakish. (In other words, heuristically speaking, any natural program can in practice be expected to be normal.)

Two main duality results can be stated.

Theorem 1. The following conditions are equivalent:

(a) program (P) is normal;

(b) program (P*) is normal;

(c) the infimum in (P) and the supremum in (P*) are equal

Theorem 2. Suppose that the equivalent conditions in Theorem 1 hold. 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).

Existence and uniqueness theorems for optimal solutions can be obtained via Theorem 2 from existence and uniqueness theorems for Lagrange multipliers, such as those cited earlier.

Optimal solutions and Lagrange multipliers can be given a saddle-point characterization. In fact, there is a one-to-one correspondence between dual pairs of generalized convex and concave programs and "regularized" concave-convex minimax problems of the most general type. This correspondence, also based on Fenchel's conjugacy, is explained in [5], [8] and [9].

References

1. Dantzig, G. B, E. Eisenberg, R. W. Cottle, "Symmetric Dueal Nonlinear Programs", Pacific J. Math., 15 (1965), pp. 809-812.

2. Gale, D., "A Geometric Duality Theorem with Economic Application", Review of Econ. Studies, 34 (1967), pp. 19-24.

3. Fenchel, W., "On Conjugate Convex Functions", Canad. J. Math., 1 (1949), pp. 73-77.

4. Fenchel, W., Convex Cones, Sets and Functions, Lecture Notes, Princeton University, 1951 (copies available in limited number from W. Fenchel, Mathematisk Institut, University of Copenhagen, Denmark).

5. Rockafellar, R. T., "A General Correspondence Between Dual Minimax Problems and Convex Programs", Pacific J. Math.

6. Rockafellar, R. T., "Duality and Stability in Dual Extremum Problems Involving Convex Functions", Pacific J. Math., 21 (1967), pp. 167-187.

7. Rockafellar, R. T., "Convex Functions and Duality in Optimization Problems and Dynamics", Mathematical Systems Theory and Economics (Lecture Notes in Operations Research and Math. Economics 11), Springer-Verlag, 1969, pp. 117-141.

8. Rockafellar, R. T., "Duality in Nonlinear Programming", Mathematics of the Decision Sciences, Part, 1, Lectures in Appl. Math., Vol. 11, Amer. Math. Soc., 1968, pp. 401-422.

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

10. Wolfe, P., "A Duality Theorem for Nonlinear Programming", Q. Appl. Math., 19 (1961), pp. 239-244.