PROCEEDINGS OF THE FOURTH CONFERENCE ON PROBABILITY THEORY

September 12—18, 1971

Braşov, Romania

REPRINT

EDITURA ACADEMIEI REPUBLICII SOCIALISTE ROMÂNIA

1973

R. TYRRELL ROCKAFELLAR

NEW APPLICATIONS OF DUALITY IN NONLINEAR PROGRAMMING

Duality has been studied in optimization problems since the origins of linear and nonlinear programming and even earlier. Many have been intrigued by it and a considerable theory of duality has been built up. Our aim here is not to review in detail what is already known, but to speculate on some of the current prospects for this theory. It is helpful first to say a few words about the general role of duality in nonlinear programming.

Duality is valuable fundamentally as a conceptual tool for understanding optimization problems and uncovering their deeper properties. In this way it is often helpful in devising and interpreting algorithms, or in analyzing mathematical models of reality from which the problems are derived. For example, characterizations of the globally optimal solutions to a problem almost always involve duality.

A common and direct use of duality in computation is the replacement of a given problem or subproblem by a dual problem (or conceivably a minimax problem) which is easier to solve. Sometimes the presence of a dual problem makes it possible to bracket the unknown maximum or minimum value between upper and lower bounds which improve in the course of computation and provide a criterion for termination. Duality theory also reveals the meaning of the special values of Lagrange multipliers and other auxilliary variables which many algorithms produce as a bonus, in addition to solving a given problem. These values can then be put to work, for instance in sensitivity analysis.

Several developments over the last few years suggest the possibility of new applications of duality. The first is the extension of the theory to problems of a more general or abstract character. The practical advantage of this is clear when one recalls that, in dealing with large-scale systems or nonconvex systems with substantial convex subsystems, one of the main hopes is to break the problem down into manageable pieces [1, 2]. The subproblems in such a structuring may not fit the classical mold, since the objective function and set of feasible solutions may be described only indirectly, in terms of other subproblems. In particular, classical assumptions concerning differentiability or domains of definition may not be reasonable in such a context. Of course, the extension of duality theory to new classes of problems should also open up applications of a more immediate sort. Optimal control problems are especially promising in this respect.

A second development that may turn out to be fruitful is the recognition that a single optimization problem can be dualized in many different ways. This idea has always been present, in the sense that a given problem could be formulated in different ways, depending on what variables were introduced and how the constraints were written down, and each formulation would correspond to a different dual problem. In this form the idea has been exploited very little outside of linear programming, but lately it has assumed importance in geometric programming. However, another general approach is now known, where the given problem (say, of minimizing a certain function over a certain set) is not "reformulated", but subjected to different classes of "perturbations". Each class of perturbations leads to a corresponding dual problem and Lagrangian function (with an associated minimax problem). This is described briefly below (further details may be found in [3, 4]). More recently there has also been work directly on replacing the classical Lagrangian function by some other function, thereby obtaining a different dual problem. Actually, as long as certain convexity assumptions are satisfied, this approach is equivalent to the preceding one, because it is known that every "generalized Lagrangian function" corresponds conversely to some class of perturbations of the original problem.

The main point here is that there is no reason to limit attention to a fixed dual problem, even in the case of a convex programming problem of classical type. It may be possible to construct a much better dual for a particular application by taking advantage of special properties of the objective function and constraints.

Another thing to note is the fact that many technical results, for example on the continuity and differentiability properties of convex functions and saddle-functions, have now been strengthened and dualized. Such results should prove useful in particular in the construction of algorithms based on duality, since they facilitate the analysis of the dual problem, making it usable even in cases where it can not be written down explicitly in a classical form.

DUALITY BASED ON CONJUGATE CONVEX FUNCTIONS

Before going on to describe a few potential applications of a specific sort, we sketch briefly the general theory of dual problems that has grown out of Fenchel's notion of conjugate convex functions.

Let us consider a convex programming problem of the type :

(1) minimize $f_0(x)$ subject to

$$f_i(x) \leqslant 0 \text{ for } i = 1, \dots, m,$$

where $f: R^n \to R$ is convex for $i = 0, 1, \ldots, m$. This problem may be represented "abstractly" as :

$$(2) \qquad \text{minimize } f(x) \text{ over all } x \in R^n,$$

where

$$(3) \qquad f(x) = f_0(x) \text{ if } x \text{ is feasible}$$

$$= +\infty \text{ if } x \text{ is not feasible.}$$

We may call f the "essential" objective function in the problem. It is another convex function, but extended-real-valued.

The ordinary duality associated with problem (1) is derived from the Kuhn-Tucker theory of Lagrange multipliers. It corresponds to the (extended) Lagrangian function L on $R^n \times R^m$ defined by :

$$(4) \qquad L(x, y) = f_0(x) + y_1 f_1(x) + \ldots + y_m f_m(x) \text{ if } y \geqslant 0,$$

$$= -\infty \text{ if } y \geqslant 0.$$

Note that $L(x, y)$ is convex in x and concave in y, and that

$$(5) \qquad f(x) = \sup_{y \in R^m} L(x, y).$$

Thus the problem (1), represented as (2), is the "$\inf_x \sup_y$" problem for L, and we are led naturally to the dual problem :

$$(6) \qquad \text{maximize } g(y) \text{ over all } y \in R^m,$$

where

$$(7) \qquad g(y) = \inf_{x \in R^n} L(x, y).$$

This is called the "ordinary" dual of (1). The function g is concave.

The ordinary dual has been studied by many authors and has turned out to extremely useful. However, other duals can also be constructed.

To do this, a space R^s is introduced, and on $R^n \times R^s$ an arbitrary (extended-real-valued) convex function F is chosen with the property that

$$(8) \qquad F(x, u) = f(x) \text{ for } u = 0.$$

The problem of minimizing $F(x, u)$ over all $x \in R^n$ for a fixed non-zero $u \in R^s$ is regarded as the "perturbation" of (2) corresponding to u.

The existence of many convex functions F satisfying (8) is obvious geometrically : the epigraph of F is simply a convex set having the epigraph of the given function f as a particular cross-section. We assume below that F is lower semicontinuous, i.e. that the epigraph of F is a closed set.

For each choice of F, there is a corresponding Lagrangian function L defined by

$$L(x, y) = \inf_{u \in R^s} \{F(x, u) + u \cdot y\}. \tag{9}$$

(Thus L is formed essentially by taking the conjugate of $F(x, u)$ as a function of u for each x and then changing some signs). It is a theorem that $L(x, y)$ is convex in x and concave in y, and that

$$F(x, u) = \sup_{y \in R^s} \{L(x, y) - u \cdot y\}. \tag{10}$$

In particular (5) holds, in view of (8) and (10), so that problem (2) corresponds as before to the "$\inf_x \sup_y$" problem for L. The dual problem corresponding to F is therefore defined to be (6), where g is again defined by (7) but with the new L.

The ordinary duality above arises when the choice of F is

$$F(x, u) = f_0(x) \text{ if } f_i(x) \leqslant u_i \text{ for } i = 1, \ldots, m, \tag{11}$$

$$= +\infty \text{ otherwise (where } s = m).$$

However, the general construction works even if f is not of the form (3). It can be applied to problem (2) in the case of an arbitrary, extended-real-valued, convex function f. Furthermore, a similar construction leads from the general concave dual problem (6) back to (2), and in this sense the duality theory is symmetric. The perturbation of (6) corresponding to a vector $v \in R^n$ is taken to be the problem of maximizing $G(y, v)$ over all $y \in R^n$, where

$$G(y, v) = \inf_{x \in R^n} \{L(x, y) - x \cdot v\}. \tag{12}$$

The function G is concave on $R^s \times R^n$, and

$$G(y, v) \equiv g(y) \text{ for } v = 0. \tag{13}$$

We shall not discuss this further here, except to remark that the duality theorems that can be proved relating problems (2) and (6) are dependent on these notions of perturbation [3, 4].

We have mentioned that the general Lagrangian functions L in this theory are convex-concave, but a converse fact has also been demonstrated. Every (extended real-valued, "lower closed" [3]) convex-concave function L is in turn the Lagrangian corresponding to some choice of perturbations of an abstract convex programming problem (2) (where f is not necessarily of form (3)). Indeed, f must be given by (5), and the F from which L arises must be given by (10). Thus convex-concave minimax theory and duality theory are equivalent: every convex-concave minimax problem is essentially a Lagrangian problem. (A minimax problem over a product of subsets of R^n and R^s can always be represented as a minimax problem over all of $R^n \times R^s$ for a certain extended-real-valued function [3, § 36]).

Note that this complete correspondence between minimax problems and dual convex and concave optimization problems would not hold if we adopted the slightly different point of view that, instead of introducing "perturbations", we were simply "reformulating" problem (2) in terms of some further variables u_i as

$$\text{minimize } F(x, u) \text{ subject to} \tag{14}$$

$$u_i = 0 \text{ for } i = 1, \ldots, m.$$

In this event, since we apparently have a convex programming problem with linear constraints, we would be obliged to define the Lagrangian function by

$$L(x, u, y) = F(x, u) + u \cdot y. \tag{15}$$

The Lagrangian would thus be of a special type (always linear in y), and it would not be possible to say that "every" convex-concave function L was the Lagrangian for some convex minimization problem (2) under some class of perturbations. Also, of course, the linearity of the constraints in (14) is only superficial, since F is extended-real-valued. In particular, F may be given by (11), and it would be unwieldy if a theory of duality required that all ordinary convex programming problems (1) be reformulated in this manner as (14). In fact, then we would not even get back as (15) the Lagrangian function (4) assigned to (1) by the Kuhn-Tucker theory !

Some examples related to penalty functions and saddle-point algorithms

A simple, but illuminating class of examples of Lagrangians associated with the convex programming problem (1) is obtained by taking F to be of the form

$$F(x, u) = f_0(x) + \Sigma_{i=1}^{m} \varphi(f_i(x), u_i). \tag{16}$$

It can be shown that the assumptions of F will then be satisfied if φ is any lower semicontinuous, convex function on R^2 with the property that

$$\text{(17)}\qquad \varphi(\alpha,\ 0) = 0 \text{ if } \alpha \leqslant 0, \quad = +\infty \text{ if } \alpha > 0.$$

The corresponding Lagrangian for (1) is

$$\text{(18)}\qquad L(x,\ y) = f_0(x) + \Sigma_{i=1}^m \psi(f_i(x),\ y_i),$$

where

$$\text{(19)}\qquad \psi(\alpha,\ \gamma) = \inf_\beta \{\varphi(\alpha,\ \beta) + \beta\lambda\}.$$

"Ordinary duality", with L given by (4), is the case where

$$\text{(20)}\qquad \varphi(\alpha,\ \beta) = 0 \text{ if } \alpha \leqslant \beta, \quad = +\infty \text{ if } \alpha > \beta.$$

If we take, say,

$$\text{(21)}\qquad \varphi(\alpha,\ \beta) = \beta^2 / |\alpha| \text{ if } \alpha < 0, \quad = 0 \text{ if } \alpha = 0 \text{ and } \beta = 0, \quad = +\infty \text{ otherwise},$$

the perturbation of (1) corresponding to a vector u is a "penalty function approximation" of (1) of a kind commonly used in interior point algorithms (see [5]). Specifically, if $u_i \neq 0$ for $i = 1, \ldots, m$ the perturbed problem is that of minimizing

$$\text{(22)}\qquad f_0(x) + \Sigma_{i=1}^m u_i^2 / |f_i(x)|$$

over the set of all x satisfying $f_i(x) < 0$ for $i = 1, \ldots, m$. In this case the Lagrangian is :

$$\text{(23)}\qquad L(x,\ y) = f_0(x) + (1/4)\Sigma_{i=1}^m y_i^2 f_i(x) \text{ if } x \text{ is feasible}, \quad = +\infty \text{ if } x \text{ is not feasible},$$

and the solution to the dual problem is trivially $y = 0$. Here it would not be sensible to think of solving the primal problem by way the dual problem or the Lagrangian minimax problem. Nevertheless the duality theory is not without implications. The dual objective function g is by a general theorem the conjugate of the concave function $-p$, where $p(u)$ is the infimum in the perturbed problem corresponding to u. Properties of the dual problem, which can be analyzed in various ways, thus correspond to properties of p, and the latter are obviously crucial computationally. Properties of the Lagrangian similarly have a bearing on the convergence of the interior point method associated with the given perturbations. Other interior point and exterior point methods can also be viewed in this light.

A rather surprising case is (for any $r > 0$)

$$\varphi(\alpha, \beta) = \beta^{2'2} \text{ if } \alpha \leqslant \beta, \qquad = +\infty \text{ if } \alpha > \beta. \tag{24}$$

The perturbation of (1) corresponding to a vector u is then the same as the perturbation in the case of ordinary duality, except that the constant $r|u|^2$ is added to the objective function of the perturbed problem. Thus it seems that the dual problem and Lagrangian should be essentially the same in ordinary duality, but in fact there is quite a difference. In the Lagrangian L given by (18), one has

$$\psi(f_i(x), y_i) = y_i f_i(x) + r\, f_i(x)^2 \text{ if } f_i(x) \geqslant -y_i/2r, \qquad = -y_i^2/4r \text{ if } f_i(x) \leqslant -y_i/2r. \tag{25}$$

Note that L is finite everywhere. Thus in the corresponding minimax problem, there are *no implicit constraints*. (In ordinary duality, the multipliers y_i are constrained by $y_i \geqslant 0$). Furthermore, L is *continuously differentiable*, assuming the functions f_i are.

In this case, therefore, in place of the Kuhn-Tucker Theorem, we get the result that problem (1) can be solved, at least in principle, by finding the saddle-point of a certain continuously differentiable, convex-concave function on $R^n \times R^m$ (no constraints). Whether or not this is a reasonable approach to computation in some cases is not clear, but the question is certainly open. Past attempts at solving convex programming problems by direct calculation of saddle-points have been hampered in particular by the constraints $y_i \geqslant 0$. In any event, it is known that for a differentiable concave-convex function L the differential equation

$$(-\dot{x}, \dot{y}) = \nabla L(x, y) \tag{26}$$

has a unique solution starting from any initial point $(x(0), y(0))$. Under mild assumptions (e.g. a sort of strict convexity or concavity condition on L at a saddle-point) every such solution $(x(t), y(t))$ converges to a saddle-point as $t \to +\infty$. "Small step" algorithms for finding saddle-points, based on (26), can therefore be constructed. The real question computationally, however, is whether "large step" algorithms of some good kind are possible.

Like the Lagrangian problem, the dual problem corresponding to (24) has the advantage that there are no implicit constraints, and the objective function is continuously differentiable. This makes it possible to simplify or improve certain dual methods of solving (1).

CONVERSION FROM NONLINEAR TO LINEAR CONSTRAINTS

The constraints in the convex programming problem (1) are in general nonlinear, but it may be possible to construct a dual problem in which the implicit constraints are linear. Solving the primal by way of the dual might then be easier than solving the primal directly. In the example just described, the dual is in fact unconstrained. However, there is another approach to this matter which exploits special properties of the functions f_i. The duality obtained in this way generalizes that in geometric programming,where the standard dual problem is well-known to be "essentially" linearly constrained. We shall not go into the details, since they are given in [6]. It is enough to point out here that this is a new kind of application of general duality theory, and that other possibilities may lie in this direction.

NONCONVEX PROGRAMMING AND DUALITY GAPS

If the functions f_i are not convex, problem (1) is much more difficult, but duality theory might still be applied and could lead to progress. The construction of Lagrangians and dual problems outlined above can easily be modified to fit this case : $F(x, u)$ is required to be convex in u for each x, but not necessarily convex in (x, u). As before, L is defined by (9) and the dual problem by (6) and (7). The dual problem is still *concave*, and formulas (5) and (10) still hold (assuming $F(x, u)$ is lower semicontinuous in u). The real trouble, of course, is that typically the infimum in (1) and the supremum in the dual problem (6) will not be equal, the difference being termed the duality gap. Thus while the dual can still be solved essentially as a concave programming problem, its solution does not yield a solution any more to the primal.

In many cases, however, we can obtain by this procedure an "estimate" of a solution to the primal, where the "estimate" is relatively good if the duality gap is relatively small. A potential application of duality theory now is the following. Since different choices of the dual problem are available, it may be possible to reduce the duality gap and thereby come closer to solving the given primal. As a matter of fact, it can be shown that this is always possible in principle "to within any ε", although whether this can be effected in a computationally sound manner is another issue; the case of (16), (24), is especially promising.

Observe that here, since no convexity is assumed with respect to x, the linear structure of R^n is not needed. Thus R^n could be replaced by a discrete set X, if desired. Naturally, almost any dual method of solution would rest on the availability of a good algorithm for unconstrained minimization of a real-valued function on X.

REFERENCES

1. A. M. Geoffrion, *Elements of large-scale programming*, Management Science, **16**, 652—691 (1970).
2. L. S. Lasdon, *Optimization Theory for Large Systems*, MacMillan, 1970.
3. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.
4. — *Duality in nonlinear programming*, in Mathematics of the Decision Sciences, Part I (G. B. Dantzig and A. F. Veinott, eds.), Lectures in Applied Math. 11, Amer. Math. Soc., 401—422 (1968).
5. A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, 1968.
6. R. T. Rockafellar, *Some convex programs whose duals are linearly constrained*, in Nonlinear Programming, J. B. Rosen O. L. Mangasarian and K. Ritter (eds.), Academic Press, 1970, p. 293—322.