DUALITY THEORY

Jørgen Tind, Department of Statistics and Operations Research, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen O, Denmark.

Keywords: Duality, Saddle point, Complementary slackness, Karush-Kuhn-Tucker conditions, Perturbation function, Supporting functions.

Contents

Summary

The purpose of this article is to present the duality theory in mathematical programming. The mathematical setup of duality depends on the actual problem under study, which for example may be an integer programming problem or a convex programming problem for example. The main intention is to introduce a unifying framework immediately exhibiting the basic involutory property of duality. It is then demonstrated how to derive a duality theory for some of the most important classes of mathematical programming problems. Emphasis is put on the description of the interrelationships among the various theories of duality. Detailed mathematical derivations, most of which are easily available in the textbooks mentioned, have not been included.

1 Introduction

Duality is an important concept in many areas of mathematics and its neighboring disciplines. Thus the general *Encyclopedia Britannica* has an article entirely devoted to duality and explains the concept as follows: "In mathematics, principle whereby one true statement can be obtained from another by merely interchanging two words." Here, we shall here consider duality in the context of optimization and the two words to be interchanged are going to be the terms, *maximum and minimum.*

First some notation. Introduce two arbitrary nonempty sets S and T and a function $K(x, y) : S \times T \rightarrow \mathbb{R}$. Consider the following *primal* problem.

$$z_0 = \max_{x \in S} \min_{y \in T} K(x, y). \tag{1}$$

Let $x_0 \in S$. If z_0 is finite and $z_0 = \min_{y \in T} K(x_0, y)$ then x_0 is said to be an *optimal solution* of the problem. (For simplicity of notation we assume that for fixed x the inner minimization problem of (1) is obtained for an argument of $y \in T$ unless it is unbounded below).

Thus, the primal problem is considered as an optimization problem with respect to (w.r.t.) the variable x. We shall next introduce the *dual* problem as an optimization problem w.r.t. y.

$$w_0 = \min_{y \in T} \max_{x \in S} K(x, y). \tag{2}$$

Let $y_0 \in T$. If w_0 is finite and $w_0 = \max_{x \in S} K(x, y_0)$ then y_0 is said to be an optimal solution of the problem. (Again, for simplicity of notation we assume that the inner maximum w.r.t. x is obtained unless the value is unbounded above).

Note that the dual problem is equivalent to

$$-\max_{y\in T}\min_{x\in S} -K(x,y),$$

which has the form of a primal problem. By the above definition its dual problem is

$$-\min_{x\in S}\max_{y\in T} -K(x,y),$$

which is equivalent to (1). This exhibits the nice property of *involution* which says that *the dual of a dual problem is equal to the primal problem.* We may thus speak of a pair of mutually dual problems, in accordance with the above quotation from *Encyclopedia Britannica.*

The entire construction may be interpreted in the framework of a so-called zero-sum game with two players, player 1 and player 2. Player 1 selects a strategy x among a possible set of strategies S. Similarly, player 2 selects a strategy $y \in T$. According to the choice of strategies an amount $K(x,y)$, the so-called payoff, is transferred from player 2 to player 1. In the primal problem (1) player 1 selects a (cautious) strategy for which this player is sure to receive at least the amount z_0. Player 2 selects in the dual problem (2) a strategy such that w_0 is a guaranteed maximum amount to be paid to player 1.

It appears that the interesting cases are obtained when optimal solutions exist for both problems such that $z_0 = w_0$. In this case we speak of *strong duality.* In general we have the following so-called weak duality.

Proposition 1 $z_0 \le w_0$.

Proof: For any $x_1 \in S$ we have $K(x_1,y) \le \max_{x\in S} K(x,y)$. Hence

$$\min_{y\in T} K(x_1,y) \le \min_{y\in T}\max_{x\in S} K(x,y).$$

Since $x_1 \in S$ is arbitrarily selected we get

$$\max_{x\in S}\min_{y\in T} K(x,y) \le \min_{y\in T}\max_{x\in S} K(x,y),$$

i.e. $z_0 \le w_0$.

When strong duality does not exist, i.e. when $z_0 < w_0$ we speak of a *duality gap* between the dual problems.

Closely related to strong duality is the following notion. $(x_0,y_0) \in S\times T$ is a *saddle point* provided that

$$K(x,y_0) \le K(x_0,y_0) \le K(x_0,y) \text{ for all } (x,y) \in S\times T.$$

The next proposition states the relationship.

Proposition 2 $(x_0,y_0) \in S\times T$ *is a saddle point if and only if* i) x_0 *is an optimal solution of the primal problem,* ii) y_0 *is an optimal solution of the dual problem and* iii) $z_0 = w_0$.

Proof: By a reformulation of the definition we get that (x_0, y_0) is a saddle point if and only if

$$\min_{y\in T} K(x_0, y) = K(x_0, y_0) = \max_{x\in S} K(x, y_0).$$

This implies that

$$z_0 = \max_{x\in S}\min_{y\in T} K(x, y) \geq K(x_0, y_0) \geq \min_{y\in T}\max_{x\in S} K(x, y) = w_0.$$

In general, by the weak duality in proposition 1 $z_0 \leq w_0$. Hence (x_0, y_0) is a saddle point if and only if

$$\max_{x\in S}\min_{y\in T} K(x, y) = K(x_0, y_0) = \min_{y\in T}\max_{x\in S} K(x, y).$$

This is equivalent to i), ii) and iii) in the proposition.

In the above framework of a game a saddle point constitutes an equilibrium point among all strategies.

In the sequel we shall consider some specific types of optimization problems derived from the primal problem (1) and dual problem (2) by further specifications of the sets S and T and the function $K(x, y)$.

Let $x \in \mathbb{R}^n$ be a variable, $f(x) : \mathbb{R}^n \to \mathbb{R}$ a function, $g(x) : \mathbb{R}^n \to \mathbb{R}^m$ a multivalued function and $b \in \mathbb{R}^m$ a vector of constants. Let $K(x, y) = f(x) + y(b - g(x))$ with $y \in \mathbb{R}^m_+$, i.e. T is defined as the non-negative orthant of $\mathbb{R}^m$. The primal problem then has the form

$$\max_{x\in S}\min_{y\geq 0} f(x) - yg(x) + yb. \tag{3}$$

Observe that if for a given x an inequality of $g(x) \leq b$ is violated then the inner minimization in the primal problem (3) is unbounded. Hence the corresponding x can be neglected as a candidate in the outer maximization. Otherwise, if a given $x \in S$ satisfies $g(x) \leq b$ we say that x is a *feasible solution*. In this case the inner minimization over $y \in \mathbb{R}^m_+$ yields $f(x)$.

Hence the primal problem may be converted to

$$\begin{array}{rl} \max & f(x) \\ \text{s.t.} & g(x) \leq b \\ & x \in S. \end{array} \tag{4}$$

The notation "s.t." stands for "subject to". The above is a standard format for a general *mathematical programming problem* consisting of an *objective function* to be maximized such that the optimal solution found is feasible. In the sequel we shall study various types of mathematical programming problems, first by a discussion of the classical case of convex programming. For each type we shall create the specific dual problem to be derived from the general dual (2).

2 Convex programming

Here $f(x)$ is assumed to be concave and the components of $g(x)$ are assumed to be convex. Moreover $S = \mathbb{R}^n_+$ which gives us a standard form of a convex programming problem.

$$\begin{aligned} \max \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq b \\ & x \geq 0. \end{aligned} \tag{5}$$

Let $d \in \mathbb{R}^m$ and consider the so-called *perturbation function* $\phi(d)$ of (5) defined by

$$\begin{aligned} \phi(d) = \max \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq d \\ & x \geq 0. \end{aligned} \tag{6}$$

Let $D = \{d \mid \exists x \geq 0 \text{ such that } g(x) \leq d\}$. The above assumptions imply that the perturbation function is concave as shown by the following.

Proposition 3 *$\phi(d)$ is concave.*

Proof: Introduce $d_1 \in D$ and $d_2 \in D$. Let x_1 be feasible in (6) with $d = d_1$ and x_2 be feasible in (6) with $d = d_2$. Introduce the scalar $\lambda \in [0, 1]$. By convexity $g(\lambda x_1 + (1-\lambda)x_2) \leq \lambda g(x_1) + (1-\lambda)g(x_2) \leq \lambda d_1 + (1-\lambda)d_2$.

This means that $\lambda x_1 + (1-\lambda)x_2$ is feasible in (6) with $d = \lambda d_1 + (1-\lambda)d_2$. Moreover, by a similar argument using the concavity of the objective function $f(x)$ we obtain that $f(\lambda x_1 + (1-\lambda)x_2) \geq \lambda f(x_1) + (1-\lambda)f(x_2)$. These observations imply that $\phi(\lambda d_1 + (1-\lambda)d_2) \geq \lambda\phi(d_1) + (1-\lambda)\phi(d_2)$, showing the concavity desired. This also applies in the instance under which $\phi(d)$ is unbounded for some values of $d \in D$.

Now let us look at the dual problem. With the specifications as above it has the following form

$$\min_{y \geq 0} \max_{x \geq 0} f(x) - yg(x) + yb$$

or with $u \in \mathbb{R}$ alternatively the form

$$\begin{aligned} \min \quad & u + yb \\ \text{s.t.} \quad & u + yg(x) \geq f(x) \text{ for all } x \geq 0 \\ & u \text{ unrestricted}, \ y \geq 0. \end{aligned} \tag{7}$$

An additional third form is also available by

$$\begin{aligned} \min \quad & u + yb \\ \text{s.t.} \quad & u + yd \geq \phi(d) \text{ for all } d \in D \\ & u \text{ unrestricted}, \ y \geq 0. \end{aligned} \tag{8}$$

Proposition 4 *The two programs* (7) *and* (8) *are equivalent.*

Proof: Notice that the two programming problems have identical objective functions. Hence, for the proposition to hold it suffices to show that the two problems also have identical sets of feasible solutions. Assume therefore that (u, y) is feasible in (8). Let $x \geq 0$ and $d = g(x)$. This implies that $d \in D$ and that $u + yg(x) = u + yd \geq \phi(d) = \phi(g(x)) \geq f(x)$. Therefore, x is also feasible in (7). Conversely, assume next that (u, y) is feasible in (7). For any $d \in D$ there exists an $x \geq 0$ such that $g(x) \leq d$. We then get that $u + yd \geq u + yg(x) \geq f(x)$ for any $x \geq 0$ such that $g(x) \leq d$. This shows that $u + yd \geq \phi(d)$ implying that (u, y) is also feasible in (8).

The dual form (8) has a nice geometric interpretation. For fixed values u and y we deal with an affine function and the corresponding points $(d, u + yd)$ constitute a hyperplane in the $m+1$ dimensional space. The purpose of (8) is to push this hyperplane without passing points in $(d, \phi(d))$ for any $d \in D$ and such that the value $u + yb$ is as low as possible. If the hyperplane touches the perturbation function ϕ at b strong duality exists. If so, the affine function is said to support the perturbation function at the point b. This can be stated in mathematical terms as follows.

$$\phi(d) \leq \phi(b) + y(d - b) \text{ for all } d \in D. \tag{9}$$

If (9) is fulfilled we say that the perturbation function is *superdifferentiable* at b. (In the converse case of a convex function one usually speaks of a subdifferentiable function). Assume that $\phi(b)$ is finite. A general result about convex functions then states that ϕ is superdifferentiable in the relative interior of D. Thus, exceptions for superdifferentiability may only occur in the extreme case when b is located at the boundary of D. Here problems may occur if for example the perturbation function has a directional derivative of infinite value. In case of no exceptions the perturbation function is said to be *stable*. In other words, strong duality holds in convex programming except for very few peculiar instances.

If the perturbation function is additionally differentiable at b then by (9) we get that the vector y_0 obtained as an optimal solution of the dual problem is equal to the gradient $\bigtriangledown\phi(b)$. This shows the potential of y_0 for marginal analysis. In practical problems b may stand for given resources or demands, and if those change to new values d then the term $y_0(d - b) = \bigtriangledown\phi(b)(d - b)$ gives a linear approximation of the corresponding change of the optimal value for the mathematical programming problem (5). Due to this property the elements of an optimal solution y_0 of the dual problem are commonly termed the *shadow prices*.

In the case that $f(x)$ and the components of $g(x)$ are also differentiable then the conditions for a saddlepoint or strong duality can be expressed in an alternative way. We here get that $(x_0, y_0) \geq 0$ is a saddle point when

$$\begin{aligned} \max_{x \geq 0} f(x) - y_0(g(x) - b) &= f(x_0) - y_0(g(x_0) - b) \\ &= \min_{y \geq 0} f(x_0) - y(g(x_0) - b). \end{aligned} \tag{10}$$

By differentiability and convexity we see that the first equation of (10) is obtained if and only if the following conditions i) and ii) are satisfied.

i) $\bigtriangledown_{x_0} (f(x_0) - y_0 g(x_0)) \leq 0$ and

ii) $x_0 \bigtriangledown_{x_0} (f(x_0) - y_0 g(x_0)) = 0$.

Similarly, the second equation of (10) is true if and only if the following conditions iii) and iv) are satisfied.

iii) $g(x_0) \leq b$ and

iv) $y_0(g(x_0) - b) = 0$.

Together with $(x_0, y_0) \geq 0$ the conditions i) - iv) constititute the famous Karush-Kuhn-Tucker conditions (see Non-Linear Programming). Subject to the assumption about convexity and differentiability they are as seen here necessary and sufficient conditions for the existence of a saddle point (and hence for strong duality as well).

3 Linear programming

Linear programming is a particular case of convex programming where $f(x) = cx$ and $g(x) = Ax$ with $c \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m\times n}$ (see Linear Programming). The corresponding linear programming problem may thus be formulated as

$$\begin{aligned} \max \quad & cx \\ \text{s.t.} \quad & Ax \leq b \\ & x \geq 0. \end{aligned} \tag{11}$$

The perturbation function becomes

$$\begin{aligned} \phi(d) = \max \quad & cx \\ \text{s.t.} \quad & Ax \leq d \\ & x \geq 0. \end{aligned} \tag{12}$$

Since $\phi(d)$ is positively homogeneous, i. e. $\phi(\lambda d) = \lambda\phi(d)$ for any scalar $\lambda > 0$ and any $d \in D$, we conclude that all affine supports of ϕ contain the origin and hence become linear. This implies that $u = 0$ in (7), and we get

$$\begin{aligned} \min \quad & yb \\ \text{s.t.} \quad & yAx \geq cx \text{ for all } x \geq 0 \\ & y \geq 0. \end{aligned} \tag{13}$$

We shall see that this is equivalent to the ordinary dual linear programming problem stated as

$$\begin{aligned} \min \quad & yb \\ \text{s.t.} \quad & yA \geq c \\ & y \geq 0. \end{aligned} \tag{14}$$

The constraints of (13) are all possible non-negative combinations of the constraints in (14). Hence any feasible solution of (14) is also feasible in (13). On the other hand, by putting the weigths x equal to the unit vectors, the constraints of (14) are included in (13). Hence the sets of feasible solutions are the same for both problems showing their equivalence.

The perturbation function is known to be polyhedral with finitely many “slopes” (mathematically: the epigraph of ϕ over D has finitely many facets). It is stable at all points in D. So strong duality holds for all points in D for which the objective function is bounded. Strong duality in linear programming is a fundamental result and all educational textbooks on optimization contain a section on this topic.

The Karush-Kuhn-Tucker conditions ii) and iv) have a nice symmetric setup in linear programming given by

$$x_0(y_0A - c) = 0 \tag{15}$$

and

$$y_0(Ax_0 - b) = 0. \tag{16}$$

They are called the *complementary slackness conditions* for the following reason. The left hand sides of (15) and (16) are sums of product terms, which are non-negative due to the feasibility of x_0 and y_0 in their respective programs (or equivalently due to i) and iii) of the Karush-Kuhn-Tucker conditions). Hence each product must be zero. Let a_j denote the jth column of A and x_{0j} the jth element of the vector x_0. Also let c_j denote the jth element of c. For each j in (15) the elements x_{0j} and $y_0a_j - c_j$ may be considered as complementary elements of a pair to be mutually multiplied. If an element is positive it is said to have a slack. The condition (15) expresses that at most one element may have a slack. Similar comments apply to condition (16).

4 Integer programming

A similar duality theory also exists in integer programming. However, due to the absence of convexity the perturbation function cannot in general be supported by an affine function. The idea is here to extend the supporting functions to a wider class of functions.

Again, let $c = (c_1, \dots c_n)$ and $A = (a_1, \dots a_n)$ such that a_j denotes the jth column of the $m \times n$ dimensional matrix A. Assume also that all elements are integer. Also let $x = (x_1, \dots x_n)$. With this notation the integer programming problem may be formulated as

$$\begin{aligned} \max \quad & \sum_{j=1}^{n} c_j x_j \\ \text{s.t.} \quad & \sum_{j=1}^{n} a_j x_j \le b \\ & x_j \ge 0 \text{ and integer for } j = 1, \dots, n. \end{aligned} \tag{17}$$

This looks similar to the linear programming problem (11) except for the quite complicating additional requirement that the components of an optimal solution must be integer (see: Combinatorial Optimization and Integer Programming).

The supporting functions to be introduced are the so-called Chvatal-Gomory functions. This class, to be denoted by $\mathcal{C}$, consists of functions $F : \mathbb{R}^m \to \mathbb{R}$ which are recursively generated by using the operations

i) multiplication by a non-negative scalar,
ii) addition and
iii) round down operation to the nearest integer number, denoted by $\lfloor \cdot \rfloor$.

Application of only i) and ii) will just generate the linear functions used in linear programming duality. It is the introduction of iii) that makes the difference. An example of a Chvatal-Gomory function where $m = 2$ and $d = (d_1, d_2) \in \mathbb{R}^2$ could be $F(d_1, d_2) = d_1 + \lfloor 2d_1 + 3\lfloor d_1 + 7d_2 \rfloor \rfloor$.

With the introduction of the function class $\mathcal{C}$ we use the following specifications in the general framework. Let $S = \{x = (x_1, \dots, x_n) \in \mathbb{R}^n_+ \mid x \text{ integer}\}$, $T = \mathcal{C}$, $y = F$ and $K(x, F) = \sum_{j=1}^n c_j x_j - \sum_{j=1}^n F(a_j) x_j + F(b)$. Since the linear functions are included in the class $\mathcal{C}$ we may utilize the same arguments as in the case of convex programming showing that the integer programming problem (17) is equivalent to (1) with the specifications given above. Similarly, the dual problem (2) takes the form

$$\begin{array}{ll} \min & F(b) \\ \text{s.t.} & F(a_j) \geq c_j \text{ for all } j = 1, \dots n \\ & F \in \mathcal{C}. \end{array} \tag{18}$$

In this context let $D = \{d \in \mathbb{R}^m \mid \exists x \geq 0 \text{ and integer, such that } Ax \leq d\}$. With c and A integer a supporting function exists for all points in D of the perturbation function for (17) if the objective is bounded. This result proves strong duality for the dual programs (17) and (18).

Corresponding to the Karush-Kuhn-Tucker conditions for convex and linear programming we get the conditions

i) $F_0(a_j) - c_j \geq 0$ for all $j = 1, \dots, n$

ii) $(F_0(a_j) - c_j) x_{0j} = 0$ for all $j = 1, \dots, n$

iii) $Ax_o \leq b$

iv) $F_0(Ax_0) - F_0(b) = 0$.

With the additional requirement that $F_0 \in \mathcal{C}$, $x_0 \geq 0$ and integer we get here that satisfaction of i) - iv) is necessary and sufficient for strong duality to hold.

5 General mathematical programming

In general, as can be extracted from the above discussion, the fundamental idea is to create a function class that is appropriate to support the perturbation function of the problem in question.

We shall here return to the general mathematical programming problem (4). When this was derived from (1) we specified T as the set of all non-negative coefficients y defining a linear function. We shall here expand T to define a larger set of non-decreasing functions F to be denoted by $\mathcal{F}$, which by its greater size somewhat limits the usefulness of a strong duality result. We still include the set of non-decreasing linear functions enabling us to use the formulation (4) for the primal problem. With this notation the dual problem (2) becomes

$$\min_{F \in \mathcal{F}} \max_{x \in S} f(x) - F(g(x)) + F(b). \tag{19}$$

In Section 2 on convex programming we included affine functions by the addition of a constant u which allows for translations. We keep this possibility and assume that translations are possible within the class $\mathcal{F}$. Repeating the arguments, on which (7) was derived, the dual program (19) turns into

$$\begin{array}{ll} \min & F(b) \\ \text{s.t.} & F(g(x)) \geq f(x) \text{ for all } x \in S \\ & F \in \mathcal{F}. \end{array} \tag{20}$$

Again the essential question is how to support the perturbation function of (4), this time without any assumptions about convexity etc. Naturally one may use the perturbation function itself, which supports anywhere. But we then ask for too much in an actual situation. However, a useful result is at hand. If the perturbation function is upper semicontinuous, it can be supported by a quadratic type non-decreasing function. So even in this quite general case not too complicated supporting functions are available to ensure strong duality between (4) and its dual (20).

6 Conclusion

Above we have discussed the duality theory for some major disciplines of mathematical programming without going into mathematical and algorithmic details. These are carefully treated in the basic textbooks listed in the bibliography.

However, at this stage it should be noted that the development of a duality theory for a selected class of problems is intimately connected to a construction method or algorithm for the generation of an optimal solution. Consider for example linear programming where most problems are going to be solved by the simplex method or one of its variants. This method calculates simultaneously an optimal solution of the dual problem. It is also the situation for the other cases discussed. In fact, verification of the optimality of a solution derived by an algorithm is intimately connected to the verification of strong duality by finding an appropriate supporting function. In this context it should be noted that the duality theory which has been selected and described here for the integer programming case is connected to the so-called cutting plane algorithms, which can be used to construct the Chvatal-Gomory functions. The so-called branch and bound methods, which generate supporting functions of a polyhedral type, belong to the other major class of algorithms in integer programming.

Glossary

Duality: A concept operating on two mathematical programming problems and the possible coincidence of their values.
Karush-Kuhn-Tucker conditions: A set of linear constraints aimed to check the optimality of a solution for a mathematical programming problem.
Perturbation function: The value of a mathematical programming problem as a function of the right hand sides of the constraints.
Saddle point: A generalization of the geometric shape of a horse saddle into higher dimensions. Intimately connected to the concept of duality.
Strong duality: The coincidence of the values of two mutually dual mathematical programming problems.
Supporting function: A function that coincides with the perturbation function at a selected point and lies above the perturbation function otherwise.
Weak duality: When the values of two mutually dual mathematical programming problems differ.

Bibliography

Encycyclopedia Britannica Online, 2000. [A general and widely used encyclopedia available at all major libraries and on the internet at http:// www.eb.com/]

Minoux, M. (1986) *Mathematical programming. Theory and algorithms*, Wiley. [This is a classic monograph with a detailed description of all major developments in convex and general mathematical programming]

Nemhauser, G. L. and Wolsey L. A. (1999) *Integer and combinatorial optimization*, paperback edition, Wiley. [This is a reprint of a classic monograph which appeared in 1988. It gives a detailed presentation of all major aspects of integer programming]

Owen, G. (1995) *Game theory*, 3rd edition, Academic Press. [A classic book on game theory for the mathematically oriented reader]

Rockafellar, R. T. (1996) *Convex analysis*, reprint edition, Princeton Univ Press. [This is a reprint of a main monograph about convexity. Appeared in 1970. It gives a complete axiomatic development of all major issues in the area]

Schrijver, A. (1986) *Theory of linear and integer programming*, Wiley. [A classic book with a complete presentation and discussion of all major optimization problems of linear and combinatorial nature including graph and network problems]

Wolsey, L. A. (1998) *Integer programming*, Wiley. [A fundamental textbook in the field giving a theoretical and algorithmic analysis of all major areas within integer programming]