Reprinted from:
NONLINEAR PROGRAMMING
© 1970
ACADEMIC PRESS INC., NEW YORK

Some Convex Programs Whose Duals Are Linearly Constrained

R. TYRRELL ROCKAFELLAR

ABSTRACT

Let $f_0, f_1, \ldots, f_m$ be convex functions on R^n, and let (P) denote the problem of minimizing $f_0(x)$ subject to $f_1(x) \le 0, \ldots, f_m(x) \le 0$. According to the theory of conjugate functions, many different dual problems can be associated with (P), each one corresponding to a particular class of perturbations of (P). Thus, in developing dual methods of solutions of (P), one has considerable flexibility in the choice of the dual problem, and the choice can be made in view of its suitability for a given purpose.

This paper treats some simple possibilities in the important case where each of the functions f_i satisfies the following condition: f_i is not affine along any line segment, unless it is affine along the entire line extending the segment (The latter holds, for example, if f_i is analytic.) It is shown that the perturbations can be chosen in this case so that the corresponding dual problem (P^*) consists essentially of maximizing a differentiable concave function subject to linear constraints. The duality theorems applicable to (P) and (P^*) are then somewhat more refined than those in the general theory; e.g. the infimum in (P) and the supremum in (P^*) are necessarily equal, if (P) is consistent.

The duality theory for the geometric programs of Duffin, Peterson and Zener, and the quadratic and ℓ^p programs of Peterson and Ecker, is derived as an illustration.

1. Introduction

Let $f_0, f_1, \ldots, f_m$ be convex functions on R^n, and let (P) denote the problem of minimizing $f_0(x)$ subject to the constraints $f_1(x) \leq 0, \ldots, f_m(x) \leq 0$. According to the theory of conjugate functions, many different dual problems can be associated with (P), each one corresponding to a particular class of perturbations of the objective function and constraints. Thus, in developing dual methods of solution of (P), one has considerable flexibility in the choice of the dual problem, and the choice can be made in view of its suitability for a given purpose.

In this paper we describe cases where the dual can be regarded essentially as a problem of maximizing a differentiable concave function subject to linear constraints. Actually, it is possible to find many such cases, and the approaches to them are quite diverse. Rather than attempting a general survey, however, we concentrate here on presenting a few especially sharp results for problems (P) of a restricted type.

To this end, we assume that the functions f_i are everywhere differentiable and satisfy the following condition, called faithful convexity: f_i is not affine (simultaneously convex and concave) along any line segment, unless f_i is affine along the entire line extending the line segment. The class of faithfully convex functions obviously includes all strictly convex functions and all affine or quadratic convex functions. In fact, it includes all analytic convex functions. Thus the results in this paper are applicable in particular to convex programming problems with analytic objective and analytic constraints.

We begin with a discussion of the "ordinary" dual (D_0) of (P), where the constraint functions are perturbed by subtracting constants. Further perturbations are then introduced, leading to a dual problem (D_1) which involves more variables, but which can be handled more directly, provided that the Legendre transformation can be carried out. This extended dual, originally presented in a less elaborate form in 1964 [11], has hitherto not been exploited for its computational properties.

Our assumptions on the nature of the functions f_i allow certain refinements of the general theorems in [12]. These results encompass the duality theorems of Duffin, Peterson and Zener for geometric programming [1, 2, 3] and those of Peterson and Ecker for (quadratic and) ℓ^p-programming [7, 8, 9, 10].

We show that, after an optimal solution has been determined for (D_1), an optimal solution may be obtained for (P), if not immediately, then by solving the dual (D_1') of just one further problem (P') of the same type as (P). This may be contrasted with the procedure given by Peterson and Ecker in ℓ^p-programming, where a sequence of subsidiary dual problems, numbering perhaps as many $m+1$, might have to be solved in order to determine an optimal solution to (P). When the simplified procedure given here is applied to an ℓ^p-program (P), the problem (P') is another ℓ^p-program.

2. Dual problems

In the terminology of [12], the problem (P) is not actually called a <u>convex program</u> until a suitable class of perturbations has been singled out. Unless otherwise specified, however, it is assumed that the perturbations are the following, in which event one speaks of an <u>ordinary</u> convex program: for each vector $u = (u_1, \ldots, u_m)$ in R^m one considers the problem of minimizing $f_0(x)$ subject to the constraints

$$f_i(x) - u_i \leq 0, \quad i = 1, \ldots, m . \tag{2.1}$$

The dual problem corresponding to this class of perturbations is that of maximizing the concave function

$$g(y) = \inf\{f_0(x) + y_1 f_1(x) + \quad + y_m f_m(x) \mid x \in R^n\} \tag{2.2}$$

over the convex set

$$C = \{y = (y_1, \ldots, y_m) \in R^m \mid y \geq 0, \; g(y) > -\infty\} \tag{2.3}$$

We call this the ordinary dual of (P) and denote it by (D_0).

Problems closely related to (D_0) have, of course, been studied by many authors. Concerning computation, the papers of Falk [4] and Geoffrion [5] are especially noteworthy. Roughly speaking, (D_0) may be expected to be useful computationally in solving (P) if the minimization in (2.2) is relatively easy to carry out for any $y \geq 0$, as for example if the functions f_i are all separable, or all quadratic. In the separable case, solving (P) by way of (D_0) is an application of the decomposition principle (for a general discussion see [12, pp. 285-290]).

Let C' denote the subset of C consisting of the vectors y such that the infimum in (2.2) is attained. Results of Falk [4] show that, if f_0 is strictly convex, then C' is convex and open relative to the orthant R^m_+, and g is continuously differentiable relative to C' with relative gradient $(f_1(x), \ldots, f_m(x))$, where the x corresponding to a given $y \in C'$ is the unique element of R^n for which the infimum in (2.2) is attained. Furthermore, if (P) is strictly consistent (i.e. satisfies the Slater condition) and has an optimal solution, or if C' is nonempty and the supremum of g over C' is attained, then

$$\sup_{y \in C'} g(y) = \sup(D_0) = \inf(P) .$$

The main restriction in applying Falk's results is the requirement that f_0 be strictly convex. Thus the case where (P) is a linear programming problem is not covered, even though this is the case on which the results are patterned. On the other hand, if f_0 is not strictly convex it can be made so by adding a strictly convex term. For instance one can add $\varepsilon|x - \overline{x}|^2$, where $\overline{x}$ is an estimate of an optimal solution to (P) and $|\cdot|$ denotes the Euclidean norm, and one then has $C = C' = R^n_+$ [14, p. 136]. This device has computational uses, but one disadvantage could be an increase in the dimensionality of the dual problem. In the altered (D_0) the convex set C is m-dimensional in R^m, whereas in the original (D_0) it might be of smaller dimension, corresponding to the fact that the dual variables had to satisfy certain linear relations as in linear programming.

Under the conditions we have imposed on the functions f_i, Falk's results can be generalized in a rather thorough way to the case where f_0 is not strictly convex, and the relationship between C and C' can also be described in greater detail. Instead of doing this here, however, we develop related results for a different dual problem in which linear relations among the dual variables appear explicitly.

Let each f_i be expressed in the form

$$f_i(x) = h_i(A_i x + a_i) + b_i x + c_i, \qquad (2.5)$$

where h_i is a differentiable, faithfully convex function on R^{n_i}, A_i is a matrix of dimension $n_i \times n$, $a_i \in R^{n_i}$, $b_i \in R^n$ and $c_i \in R^1$. Certainly such an expression (2.5) is possible, since one can always take $n_i = n$, $A_i = I$, $a_i = 0$, $b_i = 0$ and $c_i = 0$. It is easily seen from the theory of lineality vectors of convex functions [12, pp. 70-71] that the faithful convexity property of f_i is equivalent to the existence of a representation (2.5) with h_i strictly convex and n_i = rank f_i. (As an extreme case of (2.5), we allow $n_i = 0$; then the term $h_i(A_i x + a_i)$ is omitted.) In what follows we do not assume, however, that h_i is strictly convex, since that would make it awkward to treat certain examples such as geometric programs.

The dual problem we want to discuss is the one corresponding to the following class of perturbations of (P), as explained in [12, pp. 324-325]. With each vector

$$(u, v_0, \ldots, v_m) \in R^m \times R^{n_0} \times \ldots \times R^{n_m} \tag{2.6}$$

one associates the problem of minimizing

$$h_0(A_0 x + a_0 - v_0) + b_0 x + c_0 \tag{2.7}$$

subject to the constraints

$$h_i(A_i x + a_i - v_i) + b_i x + c_i - u_i \leq 0, \quad i = 1, \ldots, m. \tag{2.8}$$

The dual problem, denoted by (D_1), consists of maximizing

$$c_0 + a_0 z_0 - h_0^*(z_0) + \sum_{i=1}^{m} [c_i y_i + a_i z_i - y_i h_i^*(y_i^{-1} z_i)] \tag{2.9}$$

subject to the constraints

$$b_0 + \sum_{i=1}^{m} y_i b_i + \sum_{i=0}^{m} A_i^* z_i = 0 , \tag{2.10}$$

$$z_0 \in C_0, \quad z_i \in y_i C_i \text{ and } 0 \leq y_i \in R^1 \text{ for } i = 1, \ldots, m, \tag{2.11}$$

where A_i^* is the transpose of A_i, h_i^* is the convex function conjugate to h_i, that is,

$$h_i^*(z_i) = \sup\{z_i v_i - h_i(v_i) \mid v_i \in R^{n_i}\} , \tag{2.12}$$

and C_i is the (convex) effective domain of h_i^*,

$$C_i = \{z_i \in R^{n_i} \mid h_i^*(z_i) < +\infty\} . \qquad (2.13)$$

We use the convention in (2.9) that

$$y_i h_i^*(y_i^{-1} z_i) = 0 \text{ if } y_i = 0 \text{ and } z_i = 0 . \qquad (2.14)$$

The circumstances in which (D_1) is likely to be more useful computationally than (D_0) are those in which it is comparatively easy to determine a solution v_i (if it exists, not necessarily uniquely) to any equation of the form $\nabla h_i(v_i) = z_i$. Then, as we explain in the next section, the values of the objective function (2.9) and its directional derivatives are readily available, and the possibly nonlinear aspects of the constraints $z_i \in y_i C_i$ can, in a sense, be ignored.

3. The nature of problem (D_1)

We now state some general facts about the functions h_i^* and sets C_i and how they may be determined from h_i, particularly in light of our assumptions of differentiability and faithful convexity.

Let L_i denote the lineality space of h_i [12, p. 70]. Thus L_i is a subspace of R^{n_i}, and a vector s_i belongs to L_i if and only if the difference quotient

$$[h_i(v_i + \lambda s_i) - h_i(v_i)]/\lambda, \qquad \lambda \neq 0,$$

is a constant independent of λ and v_i. Faithful convexity implies that $v_i' - v_i \in L_i$ if h_i is affine on the line segment joining v_i and v_i'; cf. [12, Theorem 8.8].

The affine hull M_i of C_i can be obtained from L_i as follows [12, Cor. 13. 3. 4(d)]. Let the vectors $s_{ik}(k = 1, \ldots, \ell_i)$ generate L_i, and let r_{ik} be the constant (3.1) corresponding to s_{ik}. Then M_i is the set of vectors $z_i \in R^{n_i}$ satisfying the linear equation

$$S_i z_i = r_i , \qquad (3.2)$$

where S_i is the matrix of dimension $\ell_i \times n_i$ whose k^{th} row is s_{ik}, and r_i is the vector in R^{ℓ_i} with components r_{ik}. Therefore, the linear equation

$$y_i r_i - S_i z_i = 0 \qquad (3.3)$$

is a constraint implicit in the condition $z_i \in y_i C_i$ in (2.11).

The affine set M_i is all of R^{n_i} if and only if $L_i = \{0\}$, which means, because of faithful convexity, that h_i is strictly convex. Thus it could be arranged by appropriate choice of the representations (2.5) that $M_i = R^{n_i}$ for $i = 0, \ldots, m$, and then every C_i would have a nonempty interior in R^{n_i}.

In general C_i need not have a nonempty interior in R^{n_i}, but it has a nonempty interior relative to M_i, which we denote here by C_i'. Of course, C_i' is convex and has the same closure as C_i. The following facts are elementary generalizations to faithful convexity of facts derived in [12, §26] for differentiable, strictly convex functions.

(a) The set C_i' is the range of the gradient mapping ∇h_i. Thus z_i belongs to C_i' if and only if the supremum in (2.12) is attained by some v_i. Moreover, if z_i belongs to C_i' and v_i is any vector such that $\nabla h_i(v_i) = z_i$, one has

$$h_i^*(z_i) = z_i v_i - h_i(v_i) . \tag{3.4}$$

(b) If z_i and z_i' are elements of R^{n_i} such that

$$z_i + \lambda z_i' \in C_i', \quad 0 < \lambda < \lambda_0 , \tag{3.5}$$

then

$$h_i^*(z_i) = \lim_{\lambda \downarrow 0} h_i^*(z_i + \lambda z_1') . \tag{3.6}$$

Thus the values of h_i^* on the closure of C_i can be obtained as simple limits of the values of h_i^* on C_i'.

(c) h_i^* is a continuously differentiable, strictly convex function relative to C_1'. Indeed, suppose that z_i and z_i' satisfy (3.6), and let

$$\varphi(\lambda) = h_i^*(z_i + \lambda z_i'), \quad \lambda \geq 0 . \tag{3.7}$$

If z_i belongs to C_i', and v_i is any vector such that $\nabla h_i(v_i) = z_i$, then $\varphi'(\lambda)$ decreases to $\varphi'(0)$ as λ tends to 0, and one has $\varphi'(0) = z_i' v_i$. (Thus, in particular, v_i gives the directional derivatives of h_i^* at z_i; in fact, $v_i = \nabla h_i^*(z_i)$ if C' is full-dimensional.) On the other hand, if z_i does not belong to C_i', then the derivative $\varphi'(\lambda)$ decreases to $-\infty$ as λ tends to 0. (In other words, h_i^* becomes "infinitely steep" as one approaches the relative boundary of C_i'.)

(d) One has $C_i = C_i' = M_i$ if and only if

$$\lim_{\lambda \to +\infty} h_i(\lambda s_i)/\lambda = +\infty \text{ for every } s_i \notin L_i . \tag{3.8}$$

These facts yield much information about the nature of (D). For notational convenience, let us set $y = (y_1, \ldots, y_m)$ in R^m, $z = (z_0, \ldots, z_m)$ in $R^N (N = n_0 + + \ldots + n_m)$,

$$G(y, z) = k_0(z_0) + k_1(y_1, z_1) + \ldots + k_m(y_m, z_m), \quad (3.9)$$

$$k_0(z_0) = c_0 + a_0 z_0 - h_0^*(z_0), \quad (3.10)$$

$$\begin{aligned} k_i(y_i, z_i) &= c_i y_i + a_i z_i - y_i h_i^*(y_i^{-1} z_i) \text{ if } y_i > 0, \\ &= 0 \text{ if } y_i = 0 \text{ and } z_i = 0, \quad (3.11) \\ &= -\infty \text{ otherwise } (i = 1, \ldots, m). \end{aligned}$$

Note that for $i \neq 0$ one has

$$k_i(\lambda y_i, \lambda z_i) = \lambda k_i(y_i, z_i), \quad \lambda \geq 0. \quad (3.12)$$

In (D_1), G is to be maximized subject to (2.10). The functions k_i are concave and upper semicontinuous [12, p. 67 and Theorem 13.3], and therefore G is concave and upper semicontinuous. Thus for every real number α, the set of feasible solutions (y, z) to (D_1) giving a value $\geq \alpha$ to the objective function in (D_1) is a closed convex set.

The differential properties of G can be derived from those of the functions k_i, which are apparent from (c) above. In particular, for $i = 1, \ldots, m$ let

$$F_i = \{(y_i, z_i) | k_i(y_i, z_i) > -\infty\} = \{(y_i, z_i) | y_i \geq 0, z_i \in y_i C_i\}. \quad (3.13)$$

Then F_i is a convex cone whose relative interior is

$$\{(y_i,\ z_i) \mid y_i > 0,\ \ z_i \in y_i C_i'\} \tag{3.14}$$

[12, Theorem 6.8], and k_i is continuously differentiable relative to this relative interior. Furthermore, k_i becomes "infinitely steep" as one approaches a relative boundary point of F_i, unless the boundary point is the origin. At the origin, k_i is linear on every ray in F_i by (3.12), and the directional derivatives of k_i are therefore trivial to calculate.

It is possible, in view of all this, to regard (D_1) essentially as a problem of maximizing a differentiable concave function subject to only linear constraints. The exact sense is explained by the theorem which follows.

Let F denote the set of all feasible solutions $(y,\ z)$ to (D_1), and let F' be the modification of F obtained by substituting C_i' for C_i in (2.11). Let F'' be the modification obtained by not only substituting C_i' for C_i, but also strengthening the constraint $y_i \geq 0$ to $y_i > 0$, except for indices i such that $n_i = 0$ (f_i affine). Of course, the sets F, F' and F'' are convex, but they need not be closed. Their closures coincide, however, if $F'' \neq \emptyset$.

<u>Theorem 1.</u> <u>Suppose that</u> $F'' \neq \emptyset$. <u>Then the objective function in</u> (D_1) <u>has the same supremum over</u> F' <u>as it has over</u> F, <u>and the optimal solutions to</u> (D_1), <u>if any, all belong to</u> F'.

<u>Furthermore, let</u> $(y,\ z)$ <u>and</u> $(y',\ z')$ <u>be such that</u>

$$(y + \lambda y',\ z + \lambda z') \in F',\quad 0 < \lambda < \lambda_0 . \tag{3.15}$$

<u>Then the concave function</u>

$$\varphi(\lambda) = G(y + \lambda y',\ z + \lambda z') \tag{3.16}$$

<u>is continuous for</u> $0 < \lambda < \lambda_0$ <u>and continuously differentiable for</u> $0 < \lambda < \lambda_0$. <u>The derivative</u> $\varphi'(\lambda)$ <u>increases to</u> $+\infty$ <u>as</u> λ <u>tends to</u> 0, <u>unless</u> $(y, z) \in F'$, <u>in which event</u> $\varphi'(\lambda)$ <u>increases to the (finite) right derivative of</u> φ <u>at</u> $\lambda = 0$.

<u>Proof</u>. The hypothesis implies that F lies between F' and the closure of F'. Therefore G, being concave, has the same supremum over F' as over F [12, Cor. 7.3.1]. The differential properties of G described in the theorem are immediate from the properties of the functions k_i noted above, and they imply in particular that the supremum cannot occur at a point of F not in F'.

Theorem 1 asserts that, if $F'' \neq \emptyset$, the constraints $z_0 \in C_0$ and $z_i \in y_i C_i$ $(i = 1, \ldots, m)$ can be replaced in (D_1) by $z_i \in y_i C_i'$. Moreover, latter constraints are automatically taken care of, in the sense that, as one approaches a point (y, z) which is excluded by these constraints but not by the other constraints, G becomes "infinitely steep." Thus the only constraints in (D_1) which have a practical effect, in terms of gradient projections and related computational techniques, are the linear constraints (2.10) and

$$r_0 - S_0 z_0 = 0, \quad y_i r_i - S_i z_i = 0 \text{ and } y_i \geq 0 \quad \text{for } i = 1, \ldots, m. \tag{3.17}$$

Of course, if the closures of the convex sets C_i (and C_i') are all polyhedral, then the closure of F (and F') is polyhedral and hence describable entirely by a system of linear equations and inequalities. In particular, suppose that every h_i satisfies condition (3.8). Then one has $F' = F$, and the closure of F is described simply by (2.10) and (3.17). As one approaches boundary points of F not in F itself, the objective function tends to $-\infty$.

4. Examples

The following pairs of problems (P) and (D_l) illustrate the facts described in §3, as well as indicate areas of application of the duality results to be developed in §5.

Example 1. (Geometric programming [1, 2, 3]). Let

$$h_i(v_i) = \log\left(\sum_{k=1}^{n_i} e^{v_{ik}}\right) \qquad i = 0, \ldots, m, \tag{4.1}$$

where v_{ik} is the k^{th} component of $v_i \in R^{n_i}$. Then h_i is an analytic convex function (hence a faithfully convex function) such that

$$C_i = \{z_i \in R^{n_i} \mid z_{ik} \geq 0, \sum_{k=1}^{n_i} z_{ik} = 1\}, \tag{4.2}$$

$$h_i^*(z_i) = \sum_{k=1}^{n_i} z_{ik} \log z_{ik}, \quad z_i \in C_i , \tag{4.3}$$

(with $0 \log 0 = 0$). Setting $x_i = \log t_i$, one sees that (P) is equivalent to a typical geometric program. Let $b_i = 0$ and $c_i = 0$, and let the components of A_i and a_i be denoted by a^i_{kj} and a^i_{k0}, respectively. The dual problem (D_l) consists of maximizing the concave function

$$\sum_{i=0}^{m} \sum_{k=1}^{n_i} z_{ik}(a^i_{k0} - \log z_{ik}) + \sum_{i=1}^{m} y_i \log y_i \tag{4.4}$$

subject to the linear constraints

$$z_{ik} \geq 0 \text{ for } i = 0, \ldots, m \text{ and } k = 1, \ldots, n_i ,$$

$$\sum_{k=1}^{n_0} z_{0k} = 1, \text{ and } \sum_{k=1}^{n_i} z_{ik} = y_i \text{ for } i = 1, \ldots, m, \tag{4.5}$$

$$\sum_{i=0}^{m} \sum_{k=1}^{n_i} a_{kj}^i z_{ik} = 0 \text{ for } j = 1, \ldots, m .$$

Here the feasible set F is actually polyhedral, because the sets C_i are polyhedral and bounded.

Example 2. (Quadratically constrained quadratic programming; cf. [7, 8, 9, 10]). If each of the convex functions f_i is quadratic, it is simple to write down representations of the form (2.5) with

$$h_i(v_i) = \frac{1}{2} |v_i|^2 = \frac{1}{2} \sum_{k=1}^{n_i} v_{ik}^2 . \tag{4.6}$$

One then has $C_i = R^{n_i}$ and

$$h_i^*(z_i) = \frac{1}{2} |z_i|^2 = \frac{1}{2} \sum_{k=1}^{n_i} z_{ik}^2 , \tag{4.7}$$

so that in (2.9) one has

$$\begin{aligned} y_i h_i^*(y_i^{-1} z_i) &= |z_i|^2/2y_i \quad \text{if } y_i > 0 \ (z_i \text{ arb.}), \\ &= 0 \text{ if } y_i = 0 \text{ and } z_i = 0 , \\ &= +\infty \text{ otherwise.} \end{aligned} \tag{4.8}$$

Observe that the last remarks of §3 are applicable to this example; h_i satisfies (3.8).

The more general ℓ^p-programs of Peterson and Ecker may be obtained by letting h_i be of the form

$$h_i(v_i) = \sum_{k=1}^{n_i} (1/p_{ik})|v_{ik}|^{p_{ik}}, \quad 1 < p_{ik} < +\infty, \tag{4.9}$$

in which event one has

$$h_i^*(z_i) = \sum_{k=1}^{n_i} (1/q_{ik})|z_{ik}|^{q_{ik}}, \quad 1 < q_{ik} < +\infty, \tag{4.10}$$

where $(1/p_{ik}) + (1/q_{ik}) = 1$. The next example shows that a much broader class of problems can actually be handled just as easily.

Example 3. (Quasiseparable Programming). The convex function f_i is said to be quasiseparable if it can be expressed as a sum of functions, each of which is a linear function on R^n composed with a convex function (possibly infinite) on R^1. Thus f_i is quasiseparable if and only if f_i can be represented as in (2.5) with h_i separable:

$$h_i(v_i) = \sum_{k=1}^{n_i} h_{ik}(v_{ik}). \tag{4.11}$$

Assuming that every f_i has this property, in addition to the properties already specified, we can actually get representations in which the functions h_{ik} on R^1 are all differentiable and strictly convex. The conjugate functions h_{ik}^* are then, of course, relatively simple to determine, and one has

$$h_i^*(z_i) = \sum_{k=1}^{n_i} h_{ik}^*(z_{ik}). \tag{4.12}$$

Furthermore, C_i is the product of the (nondegenerate) intervals

$$C_{ik} = \{z_{ik} \in R^1 \mid h_{ik}^*(z_{ik}) < +\infty\}. \tag{4.13}$$

Thus C_i has a polyhedral closure and a nonempty interior ($M_i = R^{n_i}$). It follows that the feasible set F has a polyhedral closure.

<u>Example 4</u>. (Convex programming with linear constraints). Suppose that f_i is affine for $i = 1, \ldots, m$, so that the term $h_i(A_i x + a_i)$ can be omitted in (2.5). Then in (D_1) one maximizes

$$c_0 + \sum_{i=1}^{m} c_i y_i + a_0 z_0 - h_0^*(z_0) \tag{4.14}$$

subject to the constraints

$$b_0 + \sum_{i=1}^{m} y_i b_i + A_0^* z_0 = 0, \quad y_i \geq 0, \quad z_0 \in C_0 . \tag{4.15}$$

If these constraints can be satisfied with $z_0 \in C_0'$, then, as explained in §3, the condition $z_0 \in C_0$ reduces for practical purposes to the linear constraint $S_0 z_0 = r_0$ (which is vacuous if C_0 has a nonempty interior). If h_0 is strictly convex and satisfies the growth condition (3.8), then $C_0 = R^{n_0}$.

If f_0 itself is affine, so that (P) is a linear programming problem, everything concerning z_0 can be omitted from (4.14) and (4.15), and (D_1) is the usual dual linear programming problem, coinciding with (D_0).

<u>Example 5</u>. This miscellaneous, but specific example illustrates some useful tricks, as well as a particular computation of the conjugate functions h_i^*. We consider the problem of minimizing

$$(1/4)(1 + 3x_1)^4 - 7x_2 + \exp|x_1 - x_2| \tag{4.16}$$

over all $(x_1, x_2) \in R^2$ satisfying the constraints

$$(x_1^5 + x_2^5)^{1/3} + 2x_2^2 \leq 9 \, , \tag{4.17}$$

$$x_1 \geq 0, \;\; x_2 \geq 0, \;\; x_1 x_2 \geq 1 \, . \tag{4.18}$$

In the given formulation, this problem does not satisfy our assumptions, because the absolute value term in (4.16) spoils the differentiability of the objective function, and the quantity $1 - x_1 x_2$ in (4.18) is not convex as a function of x_1 and x_2. Also, the first term in (4.17) is not globally convex as a function of x_1 and x_2, although it is convex for $x_1 \geq 0$ and $x_2 \geq 0$. Note, however, that the objective function is the pointwise maximum of two analytic convex functions, and that the set of points satisfying (4.18) is indeed convex.

We may transform this problem into the desired form as follows. First, where the expression $|x_1 - x_2|$ occurs, we replace it by a new variable x_3, which is required to satisfy

$$x_3 \geq x_1 - x_2 \quad \text{and} \quad x_3 \geq x_2 - x_1 \, .$$

(A similar device can be used whenever the given objective function is the maximum of several functions which are differentiable and faithfully convex.) We next replace x_1 and x_2 by $|x_1|$ and $|x_2|$ in (4.17) to get a globally convex function (see below); this involves no loss of generality, because of the nonnegativity in (4.18). Finally, we replace (4.18) by an equivalent convex constraint:

$$(4 + (x_1 - x_2)^2)^{1/2} - x_1 - x_2 \leq 0 \, . \tag{4.20}$$

(The trick used to obtain this constraint is the following one, which is applicable under quite general circumstances. Let the set of all (x_1, x_2) satisfying (4.18) be denoted by H.

We observe that for each $(x_1, x_2) \in R^2$ there exists a unique smallest real number λ such that $(x_1 + \lambda, x_2 + \lambda)$ belongs to H. Denoting this λ by $r(x_1, x_2)$, we have

$$H = \{(x_1, x_2) \mid r(x_1, x_2) \leq 0\} ,$$

and r is convex. The function r is easily computed in this case, and the left side of (4.20) is $2r(x_1, x_2)$.)

The given problem is thus equivalent to minimizing

$$f_0(x) = (1/4)(1 + 3x_1)^4 - 7x_2 + \exp x_3$$

subject to the constraints

$$f_1(x) = (|x_1|^5 + |x_2|^5)^{1/3} + 2x_2^2 - 9 \leq 0 ,$$

$$f_2(x) = (4 + (x_1 - x_2)^2)^{1/2} - x_1 - x_2 \leq 0 ,$$

$$f_3(x) = x_1 - x_2 - x_3 \leq 0 ,$$

$$f_4(x) = -x_1 + x_2 - x_3 \leq 0 ,$$

where $x = (x_1, x_2, x_3) \in R^3$. The functions f_i are all differentiable and faithfully convex, so this is a problem (P) of the desired type.

The next step is to choose suitable representations of the form (2.5). Here convenience in computing the conjugate functions is the chief guide, and this is dependent on one's knowledge of general rules and examples, such as those in [12] (cf. the situation in computing indefinite integrals). We take

$$h_0(v_{01}, v_{02}) = (1/4)v_{01}^4 + \exp v_{02} ,$$

$$h_1(v_{11}, v_{12}, v_{13}) = (|v_{11}|^5 + |v_{12}|^5)^{1/3} + 2v_{13}^2 ,$$

$$h_2(v_{21}) = (4 + v_{21}^2)^{1/2} ,$$

$$A_0 = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad A_2 = [1, -1, 0]$$

$$a_0 = (1, 0), \quad a_1 = (0, 0, 0), \quad a_2 = 0$$

$$b_0 = (0, -7, 0), \quad b_1 = (0, 0, 0), \quad b_2 = (-1, -1, 0) ,$$

$$b_3 = (1, -1, -1), \quad b_4 = (-1, 1, -1)$$

$$c_0 = c_2 = c_3 = c_4 = 0, \quad c_1 = -9 .$$

We do not define h_i, A_i or a_i for $i = 3, 4$, because f_3 and f_4 are linear. The only trick here that deserves special mention is the introduction of the variable v_{13}, where v_{12} could apparently have been used just as well. This makes it easier to compute h_1^*. (The same trick can be used whenever a function h_i is a sum of convex functions, each of which has a known conjugate.)

The conjugate of h_0 is easy to determine directly (cf. [12, p. 106]):

$$h_0^*(z_{01}, z_{02}) = (3/4)z_{01}^{4/3} + z_{02}(\log z_{02} - 1) ,$$

$$C_0 = \{(z_{01}, z_{02}) \in R^2 \mid z_{02} \geq 0\} .$$

The conjugate of h_1 is obtained by a fancier method, although in principle the Legendre transformation would give the global answer [12, Theorem 26.6]. We write

$$h_1(v_{11}, v_{12}, v_{13}) = (5/3)w_1(v_{11}, v_{12}) + 4w_2(v_{13}) ,$$

where

$$w_1(v_{11}, v_{12}) = (3/5)[(|v_{11}|^5 + |v_{12}|^5)^{1/5}]^{5/3}$$

$$w_2(v_{13}) = (1/2)v_{13}^2 .$$

(The convexity of w_1 follows from [12, Theorem 5.1] and the convexity of ℓ^p norms; this provides in particular one way of verifying the convexity of h_1.) We then have

$$h_1^*(z_{11}, z_{12}, z_{13}) = (5/3)w_1^*((3/5)z_{11}, (3/5)z_{12}) + \\ + 4w_2^*((1/4)z_{13})$$

with $C_1 = R^3$, where, using [12, Theorem 15.3],

$$w_1^*(z_{11}, z_{12}) = (2/5)[(|z_{11}|^{5/4} + |z_{12}|^{5/4})^{4/5}]^{5/2} ,$$

$$w_2^*(z_{13}) = (1/2)z_{13}^2 .$$

Finally, we compute from the definition that

$$h_2^*(z_{21}) = -2(1-z_{21}^2)^{1/2}, \quad C_2 = \{z_{21} \mid -1 \leq z_{21} \leq 1\} .$$

Substituting these expressions in (2.9), (2.10) and (2.11), we obtain the following dual problem (D_1) in the variables y_i $(i = 1, \ldots, 4)$, z_{0k} $(k = 1, 2)$, z_{1k} $(k = 1, 2, 3)$ and z_{21}: maximize

$$z_{01} + z_{02} - (3/4)z_{01}^{4/3} - z_{02}\log z_{02} - 9y_1$$
$$- (2/3)(3/5)^{5/2} y_1^{-3/2}(|z_{11}|^{5/4} + |z_{12}|^{5/4})^2 \qquad (4.21)$$
$$-(1/8)y_1^{-1} z_{13}^2 + 2(y_2^2 - z_{21}^2)^{1/2}$$

subject to the linear constraints

$$-y_2 + y_3 - y_4 + 3z_{01} + z_{11} + z_{21} = 0,$$
$$-y_2 - y_3 + y_4 + z_{12} + z_{13} - z_{21} = 7\ , \qquad (4.22)$$
$$-y_3 - y_4 + z_{02} = 0\ ,$$
$$y_1 \geq 0,\ y_2 \geq 0,\ y_3 \geq 0,\ y_4 \geq 0,\ z_{02} \geq 0,\ -y_2 \leq z_{21} \leq y_2\ .$$

5. Relationships between (P), (D_0) and (D_1)

The general duality results in [12, §30] are, of course, applicable to (P), (D_0) and (D_1). It would be repetitious to state these results here, so we only present certain refinements which follow from the assumption that the functions f_i are faithfully convex. The main goal is to indicate the extent to which (P) can be solved by solving (D_1).

We denote the optimal value in (P) by inf (P); this is $+\infty$ by convention if (P) has no feasible solutions. The expressions sup (D_0) and sup (D_1) have an analogous meaning.

Before giving a duality theorem concerning these optimal values, we describe the basic relationship between problems (D_0) and (D_1). Let g_0 and g_1 be the upper semicontinuous, concave functions defined by

$$\begin{aligned} g_0(y) &= g(y) \text{ if } y \geq 0, \\ &= -\infty \text{ if } y \not\geq 0, \end{aligned} \tag{5.1}$$

$$\begin{aligned} g_1(y, z) &= G(y, z) \text{ if (2.10) holds}, \\ &= -\infty \text{ if (2.10) does not hold}, \end{aligned} \tag{5.2}$$

where g is given by (2.2) and G by (3.9). Then

$$\sup(D_0) = \sup\{g_0(y) \mid y \in R^m\}, \tag{5.3}$$

$$\sup(D_1) = \sup\{g_1(y, z) \mid y \in R^m, \ z \in R^N\}. \tag{5.4}$$

Theorem 2. For each $y \in R^m$ one has

$$g_0(y) = \sup\{g_1(y, z) \mid z \in R^N\}, \tag{5.5}$$

where the supremum is attained by some z. Thus

$$\sup(D_0) = \sup(D_1), \tag{5.6}$$

and (D_0) has an optimal solution if and only if (D_1) has an optimal solution.

Proof. The proof of this result does not depend on faithful convexity, but it uses the finiteness of the functions h_i in applying Fenchel's Duality Theorem [12, Theorem 31.1]. Let p_0 and p_1 be the perturbation functions associated with (D_0) and (D_1), respectively. Thus $p_0(u)$ is for each $u \in R^m$ the infimum of $f_0(x)$ subject to (2.1), while $p_1(u, v)$ is for each $u \in R^m$ and $v \in R^N$ the infimum of (2.7) subject to (2.8). The functions p_0 and p_1 are convex, and we have the conjugacy relations

$$g_0(y) = \inf\{yu + p_0(u) \mid u \in R^m\} = (-p_0) * (y), \quad (5.7)$$

$$g_1(y, z) = \inf\{yu + zv + p_1(u, v) \mid u \in R^m, v \in R^N\} = \\ = (-p_1) * (y, z) \quad (5.8)$$

(see [12, Theorem 30.2]). Also

$$p_0(u) = p_1(u, 0) . \quad (5.9)$$

Therefore, assuming for the moment that $p_1(u, v)$ is never $-\infty$, we have for every $\overline{y} \in R^m$

$$g_0(\overline{y}) = \inf\{q(u, v) + p_1(u, v) \mid u \in R^m, v \in R^N\}, \quad (5.10)$$

where q is the convex function on $R^m \times R^N$ whose value at (u, v) is $\overline{y}u$ if $v = 0$ and $+\infty$ if $v \neq 0$. From Fenchel's Duality Theorem, we then have

$$g_0(\overline{y}) = \sup\{(-p_1)*(y, z) - q^*(y, z) | y \in R^m, z \in R^N\}, \quad (5.11)$$

where the supremum is attained, because the interior of the convex set $\{(u, v) | p_1(u, v) < +\infty\}$ meets the subspace $\{(u, v) | q(u, v) < +\infty\}$. Of course, the conjugate function q^* vanishes at points of the form $(\overline{y}, z)$ and has the value $+\infty$ at all other points. Thus (5.11) is equivalent to (5.5) by (5.8). If p_1 takes on $-\infty$, the expression $\infty - \infty$ could occur in (5.10), and a different argument must be given. In this degenerate case, p_1 takes on $-\infty$ throughout the interior of its effective domain $\{(u, v) \mid p_1(u, v) < +\infty\}$ [12, Theorem 7.2], so that p_0 also takes on $-\infty$ by (5.9). But then g_0 and g_1 are identically $+\infty$ by (5.7) and (5.8), and (5.5) holds trivially.

Theorem 3. (Duality) *If* (P) *has a feasible solution, then*

$$\inf(P) = \sup(D_1) . \tag{5.12}$$

Moreover, if (P) *has a feasible solution* x *such that* $f_i(x) < 0$ *for every index* i *such that* f_i *is not affine, and if the common extremum in* (5.12) *is not* $-\infty$, *then* (D_1) *has an optimal solution.*

On the other hand, suppose (D_1) *has a feasible solution* (y, z) *such that* $y_i > 0$ *and* $z_i \in C_i^{!}$ *for every index* i *such that* $n_i \neq 0$ (*that is, the set* F'' *is nonempty, as in Theorem* 1). *Then again* (5.12) *necessarily holds, and if the common extremum is not* $+\infty$, (P) *has an optimal solution.*

Proof. The assertions in the first paragraph are immediate, in view of Theorem 2, from the corresponding assertions for (D_0) (see [13] and [12, Theorem 28.2]). To prove the assertions in the second paragraph, we express g_1 as $-(K_0 + \ldots + K_{m+1})$, where the convex functions K_i on $R^m \times R^N$ are defined from the functions k_i in (3.10) and (3.11) by

$$K_0(y, z) = -k_0(z_0) , \tag{5.13}$$

$$K_i(y, z) = -k_i(y_i, z_i) \text{ for } i = 1, \ldots, m, \tag{5.14}$$

$$\begin{aligned} K_{m+1}(y, z) &= 0 \text{ if (2.10) is satisfied,} \\ &= +\infty \text{ if (2.10) is not satisfied .} \end{aligned} \tag{5.15}$$

We then have

$$\sup(D_1) = -\inf(K_0 + \ldots + K_{m+1}) = (K_0 + \ldots + K_{m+1})^*(0,0)$$

$$= (K_0^* \,\Box\, \ldots \,\Box\, K_{m+1}^*)(0, 0) \qquad (5.16)$$

by Theorem 20.1 of [12], where $\Box$ denotes infimal convolution, and the infimum in the definition of $\Box$ is attained. (The hypothesis of the cited theorem, for the functions K_i, is the condition that $F'' \neq \emptyset$.) The conjugates K_i^* are easily computed, and one sees thereby that the infimum symbolized by the final expression in (5.16) is inf(P) . We leave the straightforward details of this to the reader.

Remark 1. If inf(P) is finite and the functions h_i all satisfy (3.8), as in Example 2, then (P) has an optimal solution. This follows from [12, Cor. 27.3.3].

Remark 2. Theorem 2, applied to Examples 1 and 2, yields the duality theorems of Duffin, Peterson and Zener in geometric programming [1, 2, 3] and Peterson and Ecker in ℓ^p-programming [7, 8, 9, 10], except for those results involving the subinfimum in (P). The latter results are covered by the general theorems stated in [12, §30], as has already been pointed out in the case of geometric programming by Hamala [6].

It is clear from Theorem 3 that the optimal value in (P) can usually be obtained by solving (D_1), but further analysis is needed to see how optimal solutions to (P) can likewise be obtained by a dual approach. In this analysis, it is convenient to represent each of the lineality spaces L_i introduced in §3 as

$$L_i = \{w \in R^{n_i} \mid B_i w = 0\}, \qquad (5.17)$$

where B_i is some matrix. Such a representation is of course, easy to obtain from the matrix S_i by elementary linear algebra. In particular, if h_i is strictly convex, one can take B_i to be the $n_i \times n_i$ identity matrix. Note that the subspace

$$L_i' = \{w \in R^n \mid B_i A_i w = 0\} \tag{5.18}$$

is the lineality space of f_i. Thus $f_i(x + \lambda w)$ is an affine function of λ for all x, if w satisfies $B_i A_i w = 0$.

Theorem 4. Let (y, z) be an optimal solution to (D_1), and let I be the index set consisting of 0 and all the indices $i \in \{1, \ldots, m\}$ such that $y_i \neq 0$. Then (P) has an optimal solution if and only $y_i^{-1} z_i \in C_i'$ for every $i \in I$.

Furthermore, suppose that the latter condition is satisfied, and for each $i \in I$ let v_i be an element of R^{n_i} such that $\nabla h_i(v_i) = y_i^{-1} z_i$ (with the factor y_i^{-1} omitted if $i = 0$). Let M be the affine subset of R^n consisting of the vectors x such that

$$B_i A_i x = B_i(v_i - a_i) \text{ for every } i \in I . \tag{5.19}$$

The functions f_i for $i \in I$ are then affine on M, and the optimal solutions to (P) are the vectors x such that

$$x \in M \text{ and } f_i(x) = 0 \text{ for every } i \in I, \; i \neq 0, \tag{5.20}$$

$$f_i(x) \leq 0 \text{ for every } i \notin I . \tag{5.21}$$

Proof. It follows from the first assertion of Theorem 3 that x is an optimal solution to (P) if and only if x is a feasible solution to (P) such that

$$f_0(x) = g_1(y, z) . \tag{5.22}$$

Using the defintion of the conjugate functions h_i^*, one can easily verify that (5.22) is equivalent to (5.20) and (5.21), if M is taken to be the set of all $x \in R^n$ such that

$$\nabla h_i(A_i x_i + a_i) = y_i^{-1} z_i \quad \text{for every } i \in I \tag{5.23}$$

(with y_i^{-1} omitted if $i = 0$). The equivalence of this description of M with the one in the theorem follows from the fact that h_i is faithfully convex. Indeed, assuming that v_i satisfies $\nabla h_i(v_i) = y_i^{-1} z_i$, one has $\nabla h_i(w) = y_i^{-1} z_i$, if and only if $w - v_i \in L_i$. Thus (5.23) holds if and only if

$$B_i(A_i x_i + a_i - v_i) = 0 \quad \text{for every } i \in I,$$

or in other words (5.19). This completes the proof of Theorem 4.

The affine set M in Theorem 4, if nonempty, is a translate of the subspace L equal to the intersection of the L_i' for $i \in I$, and hence one has

$$\dim M = \dim L \leq \min_{i \in I} \dim L_i' . \tag{5.24}$$

Therefore in particular, if one of the functions f_i for $i \in I$ (e.g. the objective function f_0) is strictly convex, so that $\dim L_i' = 0$, there is only one element x in M. This x must automatically satisfy the equations and inequalities in (5.20) and (5.21) and thus be the unique optimal solution in (P).

More generally, if M is not zero-dimensional, we may look at the set M' consisting of the vectors x which satisfy (5.20). Since the functions f_i, $i \in I$, are affine on M, this involves solving a further system of linear equations, and M' is an affine set. If M' is zero-dimensional, its unique element x is the unique optimal solution to (P), as in the case just considered. Otherwise, the problem is reduced to the following: find an $x \in M'$ satisfying (5.21). This can be solved by linear programming if the functions f_i for $i \notin I$, like those for $i \in I$, are affine on M', as is true certainly if $-L_i' \supset L$ for every $i \notin I$. (Observe that this holds in Example 4. It also holds if $L_i' \supset L_0'$ for $i = 1, \ldots, m$.)

At all events, if none of these shortcuts can be used, one can pass to the following convex programming problem, in order to obtain a solution to (5.20) and (5.21) and thereby an optimal solution to (P):

$$(P') \quad \text{minimize } \frac{1}{2} |x|^2 \text{ over all } x \in M' \text{ satisfying (5.21).}$$

This problem is in fact of the same type as (P). (The linear equations expressing the condition $x \in M'$ could be represented as linear inequalities, but they could also be used to eliminate some of the variables x_j and thus transform (P') into a similar problem of reduced dimensionality.) Thus (P') can be attacked via its dual (D_1'), just like (P). Since the objective function in (P') is strictly convex, an optimal solution to (D_1') immediately yields an optimal solution to (P'), as just explained.

Thus, having computed an optimal solution to (D_1), one can determine an optimal solution to (P) by solving at most one more problem (D_1'), which is similar to (D_1) but probably of a vastly lower dimension.

REFERENCES

1. R. J. Duffin, "Linearizing geometric programs," SIAM Review, 1970.

2. R. J. Duffin and E. L. Peterson, "Duality theory for geometric programming," SIAM J. App. Math. 14 (1966), 1307-1349.

3. R. J. Duffin, E. L. Peterson and C. Zener, Geometric Programming, Wiley, New York, 1967.

4. J. E. Falk, "Lagrange multipliers and nonlinear programming," J. Math. Anal. Appl. 19 (1967), 141-159.

5. A. Geoffrion, "Duality in nonlinear programming: a simplified applications-oriented development," Working Paper no. 150, Western Management Science Institute, U. of California, Los Angeles, 1969.

6. M. Hamala, "Geometric programming in terms of conjugate functions," discussion paper no. 6811, Center for Operations Research and Econometrics, Louvain, 1968.

7. E. L. Peterson and J. G. Ecker, "A unified duality theory for quadratically constrained quadratic programs and ℓ_p-constrained ℓ_p-approximation problems," Bull. Amer. Math. Soc. 74 (1968), 316-321.

8. E. L. Peterson and J. G. Ecker, "Geometric programming: duality in quadratic programming and ℓ_p-approximation I," Proc. Internat. Symp. on Math. Programming, Princeton, 1967, H. W. Kuhn and A. W. Tucker, eds. (to appear).

9. E. L. Peterson and J. G. Ecker, "Geometric programming: duality in quadratic programming and ℓ_p-approximation II (canonical programs)," SIAM J. Appl. Math. 17 (1969), 317-340.

10. E. L. Peterson and J. G. Ecker, "Geometric programming: duality in quadratic programming and ℓ_p-approximation III (degenerate programs)," J. Math. Anal. App. 29 (1970), 365-383.

11. R. T. Rockafellar, "An extension of Fenchel's duality theorem," Duke Math. J. (1966), 81-90.

12. R. T. Rockafellar, <u>Convex Analysis</u>, Princeton Univ. Press, 1969.

13. R. T. Rockafellar, "Ordinary convex programs without a duality gap," to appear.

14. A. V. Fiacco and G. P. McCormick, Nonlinear Programming: Sequential Unconstrained Minimization Techniques, Wiley, 1968.