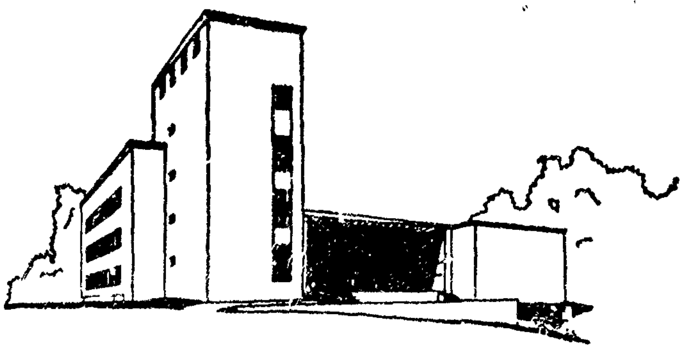
AD 697306

Carnegie-Mellon University

PITTSBURGH, PENNSYLVANIA 15213

GRADUATE SCHOOL OF INDUSTRIAL ADMINISTRATION

WILLIAM LARIMER MELLON, FOUNDER



WP - 9-69-7

Management Sciences Research Report No. 176

MINIMAX AND DUALITY FOR LINEAR AND NONLINEAR MIXED-INTEGER PROGRAMMING

by

Egon Balas

Lecture Notes
for the
NATO Advanced Study Institute
on
Integer and Nonlinear Programming
held in
Bendor (France), June 8-20, 1969

This report was prepared as part of the activities of the Management Sciences Research Group, Carnegie-Mellon University, under Contract NONR 760(24) NR 047-048 with the U.S. Office of Naval Research.

Management Sciences Research Group
Carnegie-Mellon University
Graduate School of Industrial Administration
Pittsburgh, Pennsylvania

CONTENTS

MINIMAX AND DUALITY FOR LINEAR AND NONLINEAR MIXED-INTEGER PROGRAMMING

by Egon Balas

This paper discusses duality for linear and nonlinear programs in which some of the variables are arbitrarily constrained. The most important class of such problems is that of mixed-integer (linear and nonlinear) programs. Part I introduces the duality constructions; part II discusses algorithms based on them.

PART I. SYMMETRIC DUAL MIXED-INTEGER PROBLEMS

1. The Linear Case

Consider the pair of dual linear programs

$$(\mathrm{LP})\quad \begin{array}{l} \max\ cx \\ Ax + y = b \\ x, y \geq 0 \end{array} \qquad\qquad (\mathrm{LD})\quad \begin{array}{l} \min\ ub \\ uA - v = c \\ u, v \geq 0 \end{array}$$

where A is an $m \times n$ matrix and $\{1,\ldots,m\} = M$, $\{1,\ldots,n\} = N$.

The main result of linear programming duality theory [1] is that the primal problem has an optimal solution if and only if the dual has one, in which case, denoting the two optimal solutions by $(\bar{x},\bar{y})$ and $(\bar{u},\bar{v})$ respectively, we have $c\bar{x} = \bar{u}b$, and $\bar{u}\bar{y} = \bar{v}\bar{x} = 0$. These relations play a central role in linear programming.

We wish to examine what happens to the above duality properties, if we constrain some of the primal and dual variables to belong to arbitrary sets--like, for instance, the set of integers. Suppose the first n_1 components of x and the first m_1 components of u $(0 \leq n_1 \leq n$, $0 \leq m_1 \leq m)$ are arbitrarily constrained, and the following notation is introduced: $(x_1,\ldots,x_{n_1}) = x^1$, $(u_1,\ldots,u_{m_1}) = u^1$, $x = (x^1,x^2)$, $u = (u^1,u^2)$, $\{1,\ldots,n_1\} = N_1$, $\{1,\ldots,m_1\} = M_1$. Then the above pair of problems becomes

$$(\mathrm{LPI})\quad \begin{array}{l} \max\ cx \\ Ax + y = b \\ x, y \geq 0 \\ x^1 \epsilon X^1 \end{array} \qquad\qquad (\mathrm{LDI})\quad \begin{array}{l} \min\ ub \\ uA - v = c \\ u, v \geq 0 \\ u^1 \epsilon U^1 \end{array}$$

where X^1 and U^1 are arbitrary sets of vectors in the n_1-dimensional and m_1-dimensional Euclidean space.

Let us partition A,b,c,y and v in accordance with the partitioning of x and u:

$$(1.1) \qquad A = \underbrace{\begin{array}{cc} \underbrace{}_{N_1} & \underbrace{}_{N_2} \end{array}}_{N} \left.\begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix}\begin{matrix} \}M_1 \\ \}M_2 \end{matrix}\right\} M \qquad \begin{array}{l} b = (b^1,b^2)\ ,\ c = (c^1,c^2) \\ y = (y^1,y^2)\ ,\ v = (v^1,v^2) \end{array}$$

Unless the constraints $x^1 \epsilon X^1$ and $u^1 \epsilon U^1$ happen to be redundant, it is clear that $cx < ub$ for any pair x,u satisfying (together with some y,v) the constraints of (LPI) and (LDI) respectively: a "gap" appears between the two optimal objective function values.

Suppose now that we attempt to dispose of this gap by "relaxing" each dual constraint associated with an arbitrarily constrained primal variable, and each primal constraint associated with an arbitrarily constrained dual variable; in other words, by dropping the nonnegativity requirement for each dual slack v_j, $j \epsilon N_1$, and for each primal slack y_i, $i \epsilon M_1$. Suppose, further, that while thus permitting the primal and dual constraints $j \epsilon N_1$, $i \epsilon M_1$ to be violated, we want the extent of this violation, as measured by the weighted sums $-v^1x^1$ and u^1y^1 respectively, to be as small as possible. This points towards replacing the initial primal and dual objective functions by

$$(1.2) \qquad \min_{u^1} \max_{x} cx + u^1y^1 = \min_{u^1} \max_{x} cx + u^1(b^1-A^{11}x^1-A^{12}x^2)$$

and

(1.3) $$\max_{x^1} \min_{u} ub - v^1x^1 = \max_{x^1} \min_{u} ub - (u^1A^{11}+u^2A^{21}-c^1)x^1$$

respectively.

However, it turns out that in order to obtain equality of the two objective functions, the term $-u^1A^{11}x^1$, occurring in both (1.2) and (1.3), has to be done away with. Thus, finally we are led to consider the following pair of problems:

$$\min_{u^1} \max_{x} cx + u^1y^1 + u^1A^{11}x^1$$

(P)
$$
\begin{aligned}
Ax + y &= b \\
x^1 \epsilon X^1, &u^1 \epsilon U^1 \\
x^2, y^2 &\geq 0 \\
y^1 &\text{ unconstrained}
\end{aligned}
$$

$$\max_{x^1} \min_{u} ub - v^1x^1 + u^1A^{11}x^1$$

(D)
$$
\begin{aligned}
uA - v &= c \\
u^1 \epsilon U^1, &x^1 \epsilon X^1 \\
u^2, v^2 &\geq 0 \\
v^1 &\text{ unconstrained}
\end{aligned}
$$

Here, as before, $X^1 \subset R^{n_1}$ and $U^1 \subset R^{m_1}$ are arbitrary sets of vectors in the respective spaces, with the only restriction that they are supposed to be independent of each other and of the other variables, i.e., none of them is supposed to be defined in terms of other problem variables.

Since in the above pair of problems y is uniquely defined by x and v is uniquely defined by u, a solution to P will be written as (x,u^1) and a solution to D as (u,v^1).

We define (D) to be the dual of (P). It is easy to see that the duality defined in this way is involutory (symmetric): the dual of the dual is the primal. Also, it is easy to see that the mixed-integer linear program is a special case (actually the most important special case) of (P), namely the one in which X^1 is the set of n_1-vectors with nonnegative integer components, and $m_1=0$, i.e., $M_1=\emptyset$.

The main feature of the above pair of dual problems is the special relationship between each primal variable x_j and the associated dual slack v_j, and between each dual variable u_i and the associated primal slack y_i, namely:

$$\begin{array}{llcl} & x_j \text{ arbitrarily constrained} & \longleftrightarrow & v_j \text{ unconstrained} \\ & x_j \geq 0 & \longleftrightarrow & v_j \geq 0 \\ (1.4) & y_i \text{ unconstrained} & \longleftrightarrow & u_i \text{ arbitrarily constrained} \\ & y_i \geq 0 & \longleftrightarrow & u_i \geq 0 \end{array}$$

We shall now state a lemma which will be used in the proof of the next theorem.

Let $s^1,s^2,\ldots,s^p$ be elements of arbitrary vector spaces. A vector function $G(s^1,s^2,\ldots,s^p)$ will be called <u>separable with respect to s^1</u> if there exist vector functions $H(s^1)$(independent of $s^2,\ldots,s^p$), and $K(s^2,\ldots,s^p)$(independent of s^1), such that

$$G(s^1,s^2,\ldots,s^p) \equiv H(s^1) + K(s^2,\ldots,s^p) \ .$$

$G(s^1,s^2,\ldots,s^p)$ will be called <u>componentwise separable with respect to s^1</u>, if each component g_i of G can be written either as $g_i(s^1)$, or as $g_i(s^2,\ldots,s^p)$.

Note that none of these definitions implies separability in each component of s^1. Obviously, the first of the above two definitions also applies to scalar functions (i.e., one-component vector functions).

Let r,s,t be elements of arbitrary vector spaces. Let f(r,s,t) be a scalar function and G(r,s,t) a vector function. We have.

Lemma 1.1. If f(r,s,t) is separable and G(r,s,t) is componentwise separable with respect to r or s, then

$$\inf_{s} \sup_{r,t} \left\{f(r,s,t) \mid G(r,s,t) \leq 0\right\} = \sup_{r} \inf_{s} \left\{\sup_{t}\{f(r,s,t) \mid G(r,s,t) \leq 0\}\right\}$$

Proof. Suppose $f(r,s,t) \equiv f_1(r) + f_2(s,t)$, and the constraint set can be written as $G_1(r) \leq 0$, $G_2(s,t) \leq 0$.

Then both sides of the equality in the Lemma become

$$\sup_{r} \left\{f_1(r) \mid G_1(r) \leq 0\right\} + \inf_{s} \sup_{t} \left\{f_2(s,t) \mid G_2(s,t) \leq 0\right\}$$

Similarly, if $f(r,s,t) \equiv f_1(r,t) + f_2(s)$ and the constraint set can be written as $G_1(r,t) \leq 0$, $G_2(s) \leq 0$, then both sides of the equality can be written as

$$\sup_{r,t} \left\{f_1(r,t) \mid G_1(r,t) \leq 0\right\} + \inf_{s} \left\{f_2(s) \mid G_2(s) \leq 0\right\} .$$

Q.e.d.

To state our next theorem, let us recall that y^2 and v^2 are vector functions of x^1,x^2 and u^1,u^2 respectively:

$$y^2 = b^2 - A^{21}x^1 - A^{22}x^2 , \qquad v^2 = u^1A^{12} + u^2A^{22} - c^2$$

<u>Theorem 1.1</u>. Assume v^2 (or y^2) to be componentwise separable with respect to u^1 (to x^1). Then, if (P) has an optimal solution $(\overline{x},\overline{u}^1)$, there exists $\overline{u}^2$ such that $(\overline{u},\overline{x}^1)$, where $\overline{u} = (\overline{u}^1,\overline{u}^2)$, is an optimal solution to (D), with

$$\text{(1.5)}\qquad \min_{u^1\in U^1}\max_{x\in X} cx + u^1y^1 + u^1A^{11}x^1 = \max_{x^1\in X}\min_{u\in U} ub - v^1x^1 + u^1A^{11}x^1,$$

$$\text{(1.6)}\qquad \overline{u}^2\overline{y}^2 = 0\ , \qquad\qquad \overline{v}^2\overline{x}^2 = 0$$

and

$$\text{(1.7)}\qquad (c^2 - \overline{u}^1A^{12})\overline{x}^2 - \overline{u}^2(b^2 - A^{21}\overline{x}^1) = 0$$

<u>Proof</u>. Suppose v^2 is componentwise separable with respect to u^1. (An analogous reasoning holds for the case when y^2 is componentwise separable with respect to x^1).

(D) can be stated as the problem of finding

$$\text{(1.8)}\qquad w = \max_{x^1\in X^1}\ \min_{u^1\in U^1}\ \left\{\min_{u^2\geq 0}\ \{c^1x^1+u^1b^1+u^2(b^2-A^{21}x^1)\ |u^2A^{22}\geq c^2-u^1A^{12}\}\right\}$$

In view of the separability assumption, lemma 1.1 can be applied to (1.8), i.e., $\max\limits_{x^1}$ and $\min\limits_{u^1}$ can be interchanged. Then we have

$$\text{(1.9)}\qquad w = \min_{u^1\in U^1}\ \max_{x^1\in X^1}\ \left\{c^1x^1+u^1b^1+\min_{u^2\geq 0}\{u^2(b^2-A^{21}x^1)\ |u^2A^{22}\geq c^2-u^1A^{12}\}\right\}$$

On the other hand, (P) can be written as the problem of finding

$$(1.10) \qquad z = \min_{u^1 \epsilon U^1} \max_{x^1 \epsilon X^1} \left\{c^1x^1+u^1b^1+\max_{x^2\geq 0} \{(c^2-u^1A^{12})x^2 \,|A^{22}x^2\leq b^2-A^{21}x^1\}\right\}$$

For any given u^1 and x^1 the linear programs in the inner brackets of (1.9) and (1.10) are dual to each other; and since for $x^1 = \bar{x}^1$ and $u^1 = \bar{u}^1$ the vector $\bar{x}^2$ is suppos to be an optimal solution of the linear program in (1.10)--or otherwise $(\bar{x},\bar{u}^1)$ could not be an optimal solution of (P)--it follows that the linear program in (1.9) also has an optimal solution $\bar{u}^2$, and that for $(u,x^1) = (\bar{u},\bar{x}^1)$, where $\bar{u} = (\bar{u}^1,\bar{u}^2)$, the objective function of (D) takes on the value of z. But then $(\bar{u},\bar{x}^1)$ must be an optimal solution to (D); for if it is not, i.e., if there exists some $\hat{x}^1 \epsilon X^1$ such that $\hat{w} > z$, where

$$(1.11) \qquad \hat{w} = \min_{u^1 \epsilon U^1} \left\{c^1\hat{x}^1+u^1b^1+\min_{u^2\geq 0} \{u^2(b^2-A^{21}\hat{x}^1 \,|u^2A^{22}\geq c^2-u^1A^{12}\}\right\}$$

then, following the above reasoning, there also exists a vector $\hat{x}^2$ such that $(\hat{x},\hat{u}^1)$, where $\hat{u}^1$ is the value taken on by u^1 in (1.11), is a feasible solution to (P) with an objective function value equal to $\hat{w}$--which contradicts the optimality of $(\bar{x},\bar{u}^1)$ for (P). This proves that (1.5) holds, while (1.6) follows from the fact that $\bar{x}^2$ and $\bar{u}^2$ are optimal solutions to the linear programs in the inner brackets of (1.10) and (1.9).

On the other hand, from

$$\bar{u}^2(b^2-A^{21}\bar{x}^1-A^{22}\bar{x}^2) = 0$$

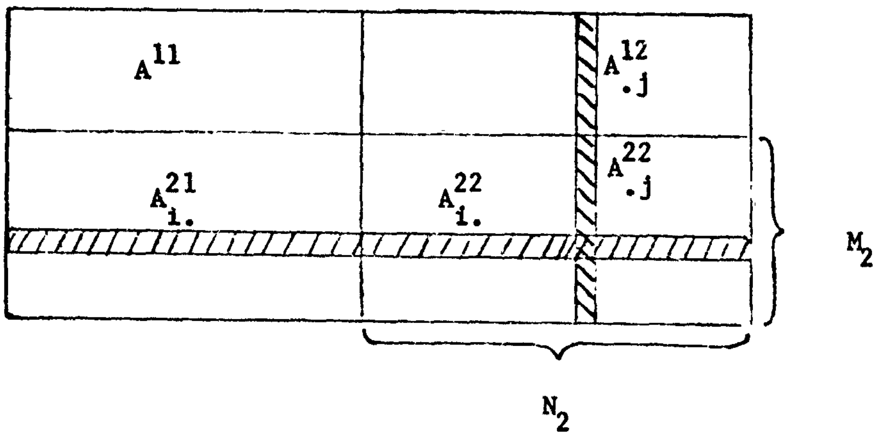
and

$$(c^2-\bar{u}^1A^{12}-\bar{u}^2A^{22})x^2 = 0$$

we have (1.7). <u>Q.e.d.</u>

According to the above theorem, the main results of linear programming duality theory carry over to the pair of dual problems (P) and (D), provided v^2 (or y^2) is componentwise separable with respect to u^1 (to x^1). Denoting by $|B_{i.}|$ and $|B_{.j}|$ respectively the norm of the i-th row and of the j-th column of a matrix B, the above assumption can also be expressed as a requirement that the matrix A satisfy the condition (see Figure 1):

$$(1.12) \qquad |A^{12}_{.j}|\cdot|A^{22}_{.j}| = 0\ ,\ j\epsilon N_2 \qquad \text{or} \qquad |A^{21}_{i.}|\cdot|A^{22}_{i.}| = 0\ ,\ i\epsilon M_2$$



<u>Figure 1</u>

This assumption is obviously a genuine restriction. However, it does not exclude from the class of problems to whic :he above results apply any of the special cases of known interest. In particular, it does not exclude the general all-integer and mixed-integer linear programs: since in these cases $M_1 = \emptyset$, A^{12} is a zero matrix and the separability requirement is satisfied.

The above duality construction is rooted in the ideas of Benders [2] and Stoer [3]. It also bears some relation to the general minimax theorem of Kakutani [4].

Additional properties of the pair of dual problems (P) and (D) are discussed in [5]. They include conditions for the existence of feasible and (finite) optimal solutions, uniqueness of the optimum, the relationship between (D) and the dual of the linear program over the convex hull of feasible points to a mixed-integer program. An economic interpretation is also given in [5] in terms of a generalized shadow price system, in which non-negative prices are associated with each constraint, and subsidies or penalties with each integer-constrained variable of a mixed-integer program. (For an alternative interpretation of pricing in integer programming see [6].)

2. The Nonlinear Case

We now discuss extensions of the above duality construction to the case of a nonlinear objective function and constraints [7],[8],[9]. This time our starting point is the pair of symmetric dual nonlinear programs studied by Dantzig, Eisenberg and Cottle [10]. Let $K(x,u)$ be a differentiable function of $x \epsilon R^n$ and $u \epsilon R^m$, and let $\nabla_x K(x,u)$ and $\nabla_u K(x,u)$ be the vectors of partial derivatives of K in the components of x and u respectively. The nonlinear programs of [10] can then be stated as

$$\text{(NP)} \qquad \begin{aligned} &\max K(x,u) - u\nabla_u K(x,u) \\ &\nabla_u K(x,u) \geq 0 \\ &x,u \geq 0 \end{aligned}$$

and

$$\text{(ND)}\qquad \begin{aligned} &\min K(x,u) - x\nabla_x K(x,u) \\ &\nabla K(x,u) \leq 0 \\ &x,u \geq 0 \end{aligned}$$

The generality of this formulation consists in the fact that K can be chosen so (see [10]) that the above pair of problems reduces to any of the dual programs studied by Dorn [11], or Cottle [12] or Wolfe [13], Mangasarian [14] and Huard [15].

The main result of [10] is that, assuming K to be twice differentiable in u, and concave in x for each u, convex in u for each x, if (NP) has an optimal solution $(\bar{x},\bar{u})$ such that the (Hessian) matrix $\nabla_u^2 K(\bar{x},\bar{u})$ of second partial derivatives of K in the components of u, evaluated at $(\bar{x},\bar{u})$, is positive definite, then $(\bar{x},\bar{u})$ is an optimal solution to (ND) and

$$\bar{u}\nabla_u K(\bar{x},\bar{u}) = \bar{x}\nabla_x K(\bar{x},\bar{u}) = 0$$

i.e., the two objective functions are equal.

As in the linear case, we now generalize the above pair of dual nonlinear programs by constraining some of the primal and dual variables to belong to arbitrary sets. Partitioning x and u in the same way as before and denoting again by X^1 and U^1 arbitrary sets of n_1-vectors and m_1-vectors respectively, we are led to consider the pair of problems

$$\text{(P)}\qquad \begin{aligned} &\min_{u^1} \max_{x,u^2} f = K(x,u) - u^2\nabla_{u^2} K(x,u) \\ &\nabla_{u^2} K(x,u) \geq 0 \\ &x^1 \epsilon X^1,\ u^1 \epsilon U^1 \\ &x^2, u^2 \geq 0 \end{aligned}$$

and

$$\max_{x^1} \min_{x^2,u} g = K(x,u) - x^2 \nabla_{x^2} K(x,u)$$

(D)
$$\nabla_{x^2} K(x,u) \leq 0$$
$$x^1 \epsilon X^1, \; u^1 \epsilon U^1$$
$$x^2, u^2 \geq 0$$

where $\nabla_{x^2} K(x,u)$ and $\nabla_{u^2} K(x,u)$ stand for the vectors of partial derivatives of K in the components of x^2 and u^2 respectively.

We define (D) to be the dual of (P). Obviously, the duality defined in this way is symmetric (involutory). It is easy to see that a mixed-integer nonlinear program is a special case of (P), in which X^1 is the set of n_1-vectors with nonnegative integer components, $m_1 = 0$, and

(2.1) $$K(x,u) \equiv f(x) - uF(x)$$

with $f(x) \epsilon R$ and $F(x) \epsilon R^m$.

In the following, we shall assume--as in the linear case--that the sets $X^1 \subset R^{n_1}$ and $U^1 \subset R^{m_1}$, while arbitrary, are independent of each other and of the other variables of the problem. Also, the concept of separability with respect to u^1 (or x^1) will again be used in the sense defined in section 1, i.e., it will <u>not</u> imply separability in each component of u^1 (or x^1).

When K(x,u) is twice differentiable in the components of x^2 and u^2, let $\nabla^2_{x^2} K(\bar{x},\bar{u})$ and $\nabla^2_{u^2} K(\bar{x},\bar{u})$ be the (Hessian) matrices of second partial derivatives of K in the components of x^2 and u^2 respectively, evaluated at $(\bar{x},\bar{u})$. We then define the following <u>regularity condition</u> for (P) and (D):

(a) If $(\bar{x},\bar{u})$ solves (P), $\nabla^2_{u^2} K(\bar{x},\bar{u})$ is positive definite;

(b) If $(\hat{x},\hat{u})$ solves (D), $\nabla^2_{x^2} K(\hat{x},\hat{u})$ is negative definite.

Denoting the constraint sets of (P) and (D) by Z and W respectively, we have

Theorem 2.1. Assume that

1. $K(x,u)$ is concave in x^2 for each x^1,u, and convex in u^2 for each x,u^1.
2. $K(x,u)$ is twice differentiable in x^2 and u^2; (P) and (D) meet the regularity condition.
3. $K(x,u)$ is separable with respect to u^1 or x^1.

Given 1,2,3, if $(\bar{x},\bar{u})$ solves (P), then it also solves (D) and

$$\min_{u^1}\max_{x,u^2}\{f\,|\,(x,u)\in Z\} = \max_{x^1}\min_{x^2,u}\{g\,|\,(x,u)\in W\} \tag{2.2}$$

with

$$\bar{u}^2\cdot\nabla_{u^2}K(\bar{x},\bar{u}) = \bar{x}^2\cdot\nabla_{x^2}K(\bar{x},\bar{u}) = 0 \tag{2.3}$$

Proof. Denote

$$\begin{aligned} z &= \min_{u^1}\max_{x,u^2}\{f\,|\,(x,u)\in Z\} \\ w &= \max_{x^1}\min_{x^2,u}\{g\,|\,(x,u)\in W\} \end{aligned} \tag{2.4}$$

Assume that $K(x,u)$ is separable with respect to u^1, i.e.,

$$K(x,u) \equiv K^1(u^1) + K^2(x,u^2) \tag{2.5}$$

(An analogous reasoning holds if K is separable with respect to x^1.)

Then z can be written as

$$z = \min_{u^1\in U^1}\ \max_{\substack{x^1\in X^1 \\ x^2,u^2\geq 0}} \left\{K^1(u^1) + K^2(x,u^2) - u^2\nabla_{u^2}K^2(x,u^2)\,|\,\nabla_{u^2}K^2(x,u^2)\geq 0\right\}$$

or

$$(2.6) \qquad z = \max_{x^1 \epsilon X^1} \min_{u^1 \epsilon U^1} \left\{K^1(u^1) + f_2(x^1)\right\}$$

where

$$(2.7) \qquad f_2(x^1) = \max_{x^2, u^2 \geq 0} \left\{K^2(x,u^2) - u^2 \nabla_{u^2} K^2(x,u^2) \,\middle|\, \nabla_{u^2} K^2(x,u^2) \geq 0\right\}$$

and w can be written as

$$(2.8) \qquad w = \max_{x^1 \epsilon X^1} \min_{u^1 \epsilon U^1} \left\{K^1(u^1) + g_2(x^1)\right\}$$

where

$$(2.9) \qquad g_2(x^1) = \min_{x^2, u^2 \geq 0} \left\{K^2(x,u^2) - x^2 \nabla_{x^2} K^2(x,u^2) \,\middle|\, \nabla_{x^2} K(x,u^2) \leq 0\right\}$$

For any given x^1, (2.7) and (2.9) are a pair of symmetric dual nonlinear programs of the type discussed in [10]. Hence, using the above mentioned results of [10], in view of assumptions 1 and 2 we have, for $x^1 = \bar{x}^1$,

$$(2.10) \qquad \bar{u}^2 \nabla_{u^2} K^2(\bar{x},\bar{u}^2) = \bar{x}^2 \nabla_{x^2} K^2(\bar{x},\bar{u}^2) = 0$$

and

$$(2.11) \qquad f_2(\bar{x}^1) = g_2(\bar{x}^1)$$

It remains to be shown that $(\bar{x},\bar{u})$ is indeed optimal for (D). If this is not the case, there exists $\hat{x}^1 \epsilon X^1$ such that $g_2(\hat{x}^1) > g_2(\bar{x}^1)$. But then, in view of the regularity condition for (D), we have

$$(2.12) \qquad g_2(\hat{x}^1) = f_2(\hat{x}^1) > f_2(\bar{x}^1)$$

which contradicts the optimality of $(\bar{x},\bar{u})$ for (P).

This, together with (2.6),(2.8) and (2.11), proves (2.2), whereas (2.3) follows from (2.5) and (2.10).

Q.e.d.

Assumptions 1 and 2 are the same as the ones required by Dantzig, Eisenberg and Cottle [10] in the absence of arbitrary constraints, except that the regularity condition is required in [10] only for the primal. Assumption 3 is an additional requirement, which represents a genuine restriction. However, this restriction does not exclude from the class of problems for which Theorem 2.1 holds the most important special case, namely, mixed-integer nonlinear programs. Indeed, when $m_1 = 0$ then u^1 disappears from the problem, which means that the separability requirement is met.

The assumptions of Theorem 2.1 can be weakened for various specific functions K(x,u). Thus, for

$$K(x,u) \equiv cx + ub - uAx + u^1A^{11}x^1 \tag{2.13}$$

(P) and (D) become the pair of dual problems discussed in section 1. In this case assumptions 1 and 2 can be dropped (1 is satisfied by definition, 2 is simply not required), whereas assumption 3 can be replaced by the weaker separability requirement of Theorem 1.1 (weaker, since assumption 3 would require A^{12} or A^{21} to be a zero matrix).

Further, for

$$K(x,u) \equiv cx + ub - uAx + \frac{1}{2}(xCx - uEu) + u^1A^{11}x^1 \tag{2.14}$$

where

$$(2.15)\quad C = \begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{pmatrix} \quad \text{and} \quad E = \begin{pmatrix} E^{11} & E^{12} \\ E^{21} & E^{22} \end{pmatrix}$$

are symmetric matrices of order n and m respectively, with C^{22} and E^{22} negative semi-definite and of order n_1 and m_1 respectively, our pair of dual problems becomes

$$\min_{u^1}\ \max_{x,u^2}\ cx + \frac{1}{2}xCx + \frac{1}{2}uEu + u^1y^1 + u^1A^{11}x^1$$

(P1)

$$Ax + Eu + y = b$$
$$x^1 \epsilon X^1\ ,\ u^1 \epsilon U^1$$
$$x^2, u^2, y^2 \geq 0$$
$$y^1 \text{ unconstrained}$$

$$\max_{x^1}\ \min_{x^2,u}\ ub - \frac{1}{2}uEu - \frac{1}{2}xCx - v^1x^1 + u^1A^{11}x^1$$

(D1)

$$uA - xC - v = c$$
$$u^1 \epsilon U^1\ ,\ x^1 \epsilon X^1$$
$$u^2, x^2, v^2 \geq 0$$
$$v^1 \text{ unconstrained}$$

This generalizes the symmetric dual quadratic programs of Cottle [12] by letting some of the primal and dual variables to be arbitrarily constrained. In this case, the regularity condition is not required, and the separability assumption can be weakened, viz., replaced by the requirement that $E^{21} = 0$ and v^2 be componentwise separable with respect to u^1, or $C^{12} = 0$ and y^2 be componentwise separable with respect to x^1.

The mixed-integer quadratic programming problem is a special case of (P1), in which X^1 is the set of n_1-vectors with nonnegative integer components, $m_1 = 0$ and E is a null matrix. (For a detailed discussion of the quadratic case, see [7].)

Finally, let us consider the case when $K(x,u) \equiv f(x) - uF(x)$, where f(x) is a scalar function and F(x) an m-component vector function of $x \epsilon R^n$, and let $F(x) = [F^1(x), F^2(x)]$, where $F^1(x)$ and $F^2(x)$ have m_1 and $m-m_1$ components respectively. Then our pair of dual problems generalizes the dual nonlinear programs studied by Wolfe [13], Mangasarian [14], and Huard [15]:

$$\min_{u^1} \max_{x} f(x) - u^1F^1(x)$$

(P2)
$$F^2(x) \leq 0$$
$$x^1 \epsilon X^1, u^1 \epsilon U^1$$
$$x^2 \geq 0$$

$$\max_{x^1} \min_{x^2,u} f(x) - uF(x) - x^2 \nabla_{x^2}[f(x) \quad uF(x)]$$

(D2)
$$\nabla_{x^2}[f(x) - uF(x)] \leq 0$$
$$x^1 \epsilon X^1, u^1 \epsilon U^1$$
$$x^2, u^2 \geq 0$$

Assumptions 1,2, and 3 of Theorem 1 are now to be maintained, but the regularity condition for (P) and (D) can be weakened so as to read:

(a) If $(\bar{x},\bar{u})$ solves (P2), the inequality set $F^2(\bar{x}^1,x^2) \leq 0$ satisfies the Kuhn-Tucker constraint qualification [16] at $x^2 = \bar{x}^2$.

(b) If $(\hat{x},\hat{u})$ solves (D2), the matrix $\nabla^2_{x^2}[f(\hat{x}) - \hat{u}F(\hat{x})]$ is nonsingular.

Theorem 2.1 then becomes

Corollary 2.1. Given the assumptions 1,2,3 of Theorem 2.1, if (P2) has an optimal solution $(\bar{x},\bar{u}^1)$, there exists $\bar{u}^2$ such that $(\bar{x},\bar{u}) = (\bar{x},\bar{u}^1,\bar{u}^2)$ is an optimal solution to (D2). Conversely, if (D2) has an optimal solution $(\hat{x},\hat{u})$ then $(\hat{x},\hat{u}^1)$ is an optimal solution to (P2).

In both cases, (2.2) and (2.3) hold.

3. Linearization of the Dual

An undesirable characteristic of the dual problems (P) and (D) discussed in the previous section is the presence of the arbitrarily constrained primal variables x^1 in the dual inequality set. This was not the case for the linear problem discussed in section 1.

Now consider again the nonlinear problem (P) of section 2, and let $K(x,u)$ be also differentiable in x^1 on the set $\{x^1 \epsilon R^{n_1} | x^1 \geq 0\}$ for each x^2,u. Then consider the problem [9]:

$$\max_{s}\ \min_{x,u}\ g' = K(x,u) - x\nabla_x K(x,u) + s\nabla_{x^1} K(x,u)$$

$$\nabla_{x^2} K(x,u) \leq 0$$

$$(D') \qquad s \epsilon X^1, \quad u^1 \epsilon U^1$$

$$x,u^2 \geq 0$$

where s is an n_1-vector. Let W' be the constraint set of (D').

The inequality set of (D'), unlike that of (D), is independent of the arbitrarily constrained variables $s \epsilon X^1$; and the optimand of (D'), unlike that of (D), is linear in these same variables $s \epsilon X^1$. We shall show, however, that with two additional assumptions (D') is equivalent to (D). In view of its linearity in the arbitrarily constrained variables s, (D') will be called the linearized dual of (P).

Theorem 3.1. Assume 1,2,3 as in Theorem 2.1 (regularity also assumed for (D')), and

4. $K(x,u)$ is concave in x^1 on the set $\{x^1 \epsilon R^{n_1} | x^1 \geq 0\}$ for each x^2,u.
5. $X^1 \subset \{s \epsilon R^{n_1} | s \geq 0\}$.

Then the following statements hold:

a) If $(\bar{x},\bar{u})$ solves (P), then $(\bar{s},\bar{x},\bar{u})$, where $\bar{s} = \bar{x}^1$, solves (D').

b) If $(\tilde{x},\tilde{u})$ solves (D), then $(\tilde{x},\tilde{u})$ solves (P) and $(\tilde{s},\tilde{x},\tilde{u})$, where $\tilde{s} = \tilde{x}^1$, solves (D').

c) If $(\hat{s},\hat{x},\hat{u})$ solves (D'), then $\hat{s} = \hat{x}^1$ and $(\hat{x},\hat{u})$ solves (P) and (D).

d) In each of the cases a), b), c),

$$\min_{u^1} \max_{x,u^2} \{f | (x,u) \epsilon Z\} = \max_{s} \min_{x,u} \{g' | (s,x,u) \epsilon W'\} = \max_{x^1} \min_{x^2,u} \{g | (x,u) \epsilon W\}. \tag{3.1}$$

Proof. Consider the problem (P'), which clearly is just another way of writing (P) under assumption 5 above (here s is an n_1-vector):

$$\min_{u^1} \max_{s,x,u^2} K(x,u) - u^2 \nabla_{u^2} K(x,u)$$

$$\text{(P')} \qquad \begin{aligned} \nabla_{u^2} K(x,u) &\geq 0 \\ x^1 - s &\geq 0 \\ -x^1 + s &\geq 0 \\ s \epsilon X^1 \ , \ & u^1 \epsilon U^1 \\ x,u^2 &\geq 0 \end{aligned}$$

We now restate (P') in the form (P). Let

$$\xi^1 \epsilon R^{n_1}, \xi^2 \epsilon R^n, \eta^1 \epsilon R^{m_1}, \eta^2 \epsilon R^{m-m_1+2n_1}, t^1 \epsilon R^{n_1}, t^2 \epsilon R^{n_1}$$

$$\xi = (\xi^1, \xi^2) = (s,x) \ , \quad \text{where } \xi^1 = s \ , \ \xi^2 = x$$

$$\eta = (\eta^1, \eta^2) = (u, t^1, t^2) \ , \quad \text{where } \eta^1 = u^1 \ , \ \eta^2 = (u^2, t^1, t^2)$$

$$H(\xi, \eta) = K(x,u) + (t^1 - t^2)(x^1 - s)$$

Then (P') can be stated as the problem (P"):

$$\min_{\eta^1} \max_{\xi, \eta^2} \ H(\xi,\eta) - \eta^2 \nabla_{\eta^2} H(\xi,\eta)$$

(P")
$$\nabla_{\eta^2} H(\xi,\eta) \geq 0$$
$$\xi^1 \epsilon X^1 \ , \ \eta^1 \epsilon U^1$$
$$\xi^2 \ , \ \eta^2 \geq 0$$

We now write the dual (D") of (P"):

$$\max_{\xi^1} \min_{\xi^2, \eta} \ H(\xi,\eta) - \xi^2 \nabla_{\xi^2} H(\xi,\eta)$$

(D")
$$\nabla_{\xi^2} H(\xi,\eta) \leq 0$$
$$\xi^1 \epsilon X^1 \ , \ \eta^1 \epsilon U^1$$
$$\xi^2 \ , \ \eta^2 \geq 0$$

which upon substitution becomes

$$\max_{s}\ \min_{x,u,t^1,t^2} K(x,u) - x\nabla_x K(x,u) - (t^1-t^2)s$$

(3.2)

$$\begin{aligned} \nabla_{x^1} K(x,u) + t^1 - t^2 &\leq 0 \\ \nabla_{x^2} K(x,u) &\leq 0 \\ s\epsilon X^1 \ ,\ u^1 \epsilon U^1 & \\ x,u^2,t^1,t^2 &\geq 0 \end{aligned}$$

Introducing the slack vector $p \geq 0$ in the first inequality set of (3.2) and substituting in the objective function for t^1-t^2, we obtain

$$\max_{s}\ \min_{x,u,p}\ K(x,u) - x\nabla_x K(x,u) + s\nabla_{x^1} K(x,u) + sp$$

(3.3)

$$\begin{aligned} \nabla_{x^2} K((x,u) &\leq 0 \\ s\epsilon X^1 \ ,\ u^1 \epsilon U^1 & \\ x,u^2,p &\geq 0 \end{aligned}$$

Since p is nonnegative, (3.3) is equivalent to (D') in the sense that

α) if $(\overline{s},\overline{x},\overline{u},\overline{p})$ solves (3.3), then $\overline{sp} = 0$ and $(\overline{s},\overline{x},\overline{u})$ solves (D');

β) if $(\hat{s},\hat{x},\hat{u})$ solves (D'), then $(\hat{s},\hat{x},\hat{u},\hat{p})$, where $\hat{p} = 0$, solves (3.3).

Then statement a) of Theorem 3.1 follows from the application of Theorem 2.1 to (P'). Here we need assumption 4, since x plays in (P') the role of x^2 in (P).

To obtain statement b), note that $(\tilde{s},\tilde{x},\tilde{u})\epsilon W'$, where $\tilde{s} = \tilde{x}^1$. Also, from Theorem 2.1 applied to (D), $(\tilde{x},\tilde{u})$ solves (P), hence $(\tilde{s},\tilde{x},\tilde{u})$ solves (P').

Since (P) is assumed to meet the regularity condition, so is (P'), which implies that $(\tilde{s},\tilde{x},\tilde{u})$ solves (D').

Statement c) follows from the application of Theorem 2.1 to (D'). The fact that $(\hat{s},\hat{x},\hat{u})$ solves (P') implies that $\hat{s} = \hat{x}^1$ and $(\hat{x},\hat{u})$ solves (P). Applying again Theorem 2.1 to (P), one sees that $(\hat{x},\hat{u})$ solves (D).

In each of the cases a),b),c), statement d) follows directly from the proofs given above.

<u>Q.e.d.</u>

Theorem 3.1 on the linearization of the dual constitutes the basis of the method for solving mixed-integer nonlinear programs presented in section 5.

PART II. ALGORITHMS

The theory presented in Part I can be used for computational purposes. In the linear case, it leads to the same class of algorithms to which Benders' partitioning procedure [2] belongs. We shall describe a variant which differs from Benders' procedure in that it requires the solution of a single pure integer program instead of a sequence of such programs, and which is essentially the same as the one described by Lemke and Spielberg [17] (The differences will be mentioned later).

In the nonlinear case, the above theory leads to a new algorithm for solving pure or mixed-integer nonlinear programs, which can be regarded as a generalization of Benders' partitioning procedure (and its variations) to the nonlinear case.

4. Implicit Enumeration for Mixed-Integer Linear Programs

We shall consider the mixed-integer programming problem in the special form where the integer variables are zero-one variables ([18] and [7]) describe techniques for bringing any integer or mixed-integer linear program to this form):

$$
\text{(P)} \qquad \begin{aligned}
\min \; & c^1 y + c^2 x \\
& A^1 y + A^2 x \geq b \\
& y_j = 0 \text{ or } 1, \quad j \in N \\
& x_h \geq 0, \; h \in H
\end{aligned}
$$

where $c^1 \in R^n$, $c^2 \in R^p$, $b \in R^m$, $A^1_{(m \times n)}$, $A^2_{(m \times p)}$ are given, and $\{1,\dots,m\} = M$, $\{1,\dots,n\} = N$, $\{1,\dots,p\} = H$.

The dual of (P) is then the problem (see Section 1)

$$\text{(D)}\qquad \begin{aligned} \min_{y}\max_{u}\ & ub + v^1 y \\ & uA^1 + v^1 = c^1 \\ & uA^2 + v^2 = c^2 \\ & y_j = 0 \text{ or } 1, \quad j \in N \\ & v_h^2 \geq 0, \quad h \in H \\ & u_i \geq 0, \quad i \in M \end{aligned}$$

or, after substitution of v^1 (which is unconstrained)

$$\text{(D)}\qquad \begin{aligned} \min_{y}\max_{u}\ g = {} & ub + (c^1 - uA^1)y \\ & uA^2 \leq c^2 \\ & y_j = 0 \text{ or } 1, \quad j \in N \\ & u_i \geq 0, \quad i \in M \end{aligned}$$

Let

$$(4.1)\qquad Y = \{y \in R^n | y_j = 0 \text{ or } 1,\ j \in N\}$$

For each $y \in Y$, (D) becomes a linear program $L(y)$ in u. One could therefore solve (D) by solving $L(y)$ for each element y of the finite set Y, and by choosing that $\bar{y} \in Y$ which minimizes the optimal (maximal) solution of $L(y)$. On the other hand, one could use an implicit enumeration technique [19] if one could generate constraints to be satisfied by any $y \in Y$ which is a candidate for optimality. The reason why this can indeed be done, is that the inequalities of (D) are <u>independent of y</u>.

Assume we have solved $L(y)$ for a sequence $y^1, y^2, .., y^q$ of vectors $y \in Y$. We shall ignore the trivial case when $L(y)$, and hence (D), has no feasible solution (then P has no finite optimum).

Let

$$\{1,\ldots,q\} = Q = Q_1 \cup Q_2 \tag{4.2}$$

where

$$Q_1 = \{k \in Q | L(y^k) \text{ has a finite optimum}\} \tag{4.3}$$

$$Q_2 = \{k \in Q | L(y^k) \text{ has no finite optimum}\}$$

For $k \in Q_1$, let u^k be an optimal solution of $L(y^k)$, and let g^k be the optimal value of the objective function of $L(y^k)$. Further, let

$$g^* = \min_{k \in Q_1} g^k \tag{4.4}$$

For $k \in Q_2$, $L(y^k)$ has a feasible solution of the form

$$u^k + \lambda t^k, \qquad \lambda \geq 0 \tag{4.5}$$

where u^k is an extreme point and t^k a direction vector for an extreme ray of the convex polytope of feasible solutions to $L(y^k)$, t^k being a solution of the homogeneous system $tA^2 \leq 0$.

Since the constraints of $L(y)$ are independent of y, any optimal solution u^k to a linear program $L(y^k)$, as well as any feasible solution $u^k + \lambda t^k$ of the type described above, is a feasible solution to all other linear programs $L(y)$. Hence, we have

<u>Theorem 4.1</u>. Any $y \in Y$ (if one exists) such that

$$\max_{u \geq 0} \{u(b - A^1 y) + c^1 y | uA^2 \leq c^2\} < g^* \tag{4.6}$$

satisfies the constraints

$$(c^1 - u^k A^1)y < g^* - u^k b, \quad k \in Q \tag{4.7}$$

and

$$(4.8) \qquad -t^k A^1 y \leq - t^k b, \qquad k \in Q_2$$

<u>Proof</u>. Suppose y violates (4.7) for $p \in Q$, i.e.

$$u^p b + (c^1 - u^p A^1) y \geq g^*$$

Then, since u^p is a feasible solution to $L(y)$,

$$\max_{u \geq 0} \{u(b - A^1 y) + c^1 y | uA^2 \leq c^2\} \geq u^p(b - A^1 y) + c^1 y \geq g^*$$

which contradicts (4.6).

On the other hand, if y violates (4.8) for $p \in Q_2$, i.e., if $t^p(b - A^1 y) > 0$, then the objective function of $L(y)$ can be increased indefinitely by setting $u = u^p + \lambda t^p$, $\lambda > 0$, and by increasing λ, which again contradicts (4.6). <u>Q.e.d.</u>

We can now systematically search the set Y by applying the exclusion tests of implicit enumeration [18], [19] to the constraints (4.7), (4.8). Whenever a $y \in Y$ is found that satisfies the current constraints, it is introduced into the objective function of the linear program $L(y)$ which is then post-optimized. This in turn yields a new constraint (4.7), and possibly (4.8), which is not satisfied by the current y. It may also yield an improved value of g^*. A typical iteration of the algorithm consists then of the following two phases:

I. (Steps 1-4 below). Using implicit enumeration techniques, find a vector $y^s \in Y$ satisfying the current constraints (4.7) and (4.8). Then go to II.

II. (Steps 5-6 below). Solve (post-optimize) $L(y^s)$, add a new constraint to (4.7) and possibly to (4.8), and (possibly) update g^*. Then go to I.

Whenever a new phase I is started, the implicit enumeration over the set Y is continued from where it had been interrupted at the end of the previous phase I: those elements of Y that had been excluded as infeasible for the current constraint set, do certainly not become feasible by the addition of new constraints. The procedure ends when there is no $y \in Y$ satisfying the current constraints (4.7) and (4.8). Then, if $Q_1 \neq \emptyset$, the vector y associated with the current g^* yields an optimal solution, or, if $Q_1 = \emptyset$, (P) has no feasible solution at all.

To discuss the algorithm in detail, we shall change the notation. Q and Q_2 will now be considered disjoint ordered sets (i.e., each inequality (4.8) will have a different index from each inequality (4.7)), denoted by Q and T respectively, and the two sets of inequalities (4.7), (4.8) will be written as a single set

$$\sum_{j\in N} \alpha_{ij} y_j \geq \beta_i , \quad i \in V = Q \cup T \tag{4.9}$$

with

$$\alpha_{ij} = \begin{cases} u^i A^l - c^l & \text{for } i \in Q \\ t^i A^l & \text{for } i \in T \end{cases}$$
$$\beta_i = \begin{cases} u^i b - g^* + \varepsilon^i & \text{for } i \in Q \\ t^i b & \text{for } i \in T \end{cases} \tag{4.10}$$

where ε^i is a positive number sufficiently small to enable us to replace the strict inequalities of (4.7) by ordinary inequalities, without unduly excluding from consideration any $y \in Y$. In other words, ε^i can be any number satisfying

$$0 < \varepsilon^i < |\alpha_{ij} - \alpha_{ih}| \tag{4.11}$$

for all pairs of indices j,h such that $\alpha_{ij} \neq \alpha_{ih}$.

We are interested in generating vectors $y \in Y$ satisfying (4.9). Any $y \in Y$ will be called a <u>solution</u>, and a solution satisfying (4.9) will be called <u>feasible</u>. In the process, we shall generate a sequence of <u>pseudo-solutions</u> $\psi_1,\ldots,\psi_s$, a pseudo-solution (or partial solution) ψ_k being defined as a set of 0-1 value-assignments to some components of y:

$$\psi_k = \{y_j^k = \delta_j^k,\ j = j_1,\ldots,j_q\}\ ,\ 1 \leq q \leq n \tag{4.12}$$

where each δ_j^k represents one of the values 0 and 1.

Let J_k^1 (and J_k^o respectively) be the set of those $j \in N$ such that the j^{th} component of y is assigned by ψ_k the value 1 (the value 0), i.e.,

$$J_k^1 = \{j \in N | \delta_j^k = 1\}\ ,\ J_k^o = \{j \in N | \delta_j^k = 0\} \tag{4.13}$$

and let

$$N_k = N - J_k^1 \cup J_k^o \tag{4.14}$$

We shall say that, at the stage characterized by the pseudo-solution ψ_k, y_j is <u>fixed at 1</u> if $j \in J_k^1$, <u>fixed at 0</u> if $j \in J_k^o$, and <u>free</u> if $j \in N_k$.

The solution y^k defined by

$$y_j^k = \begin{cases} \delta_j^k & \text{for } j \in J_k^1 \cup J_k^o \\ 0 & \text{for } j \in N_k \end{cases} \tag{4.15}$$

will be called the solution associated with ψ_k.

In order to keep track of the sequence of pseudo-solutions that will be generated, we shall associate with this sequence an arborescence (rooted tree) $\mathfrak{A}$. Each node h of $\mathfrak{A}$ corresponds to a pseudo-solution ψ_h, to a solution y^h associated with ψ_h via (4.15), and to a linear program $L(y_h)$. Each arc (h,k) of $\mathfrak{A}$ corresponds to a pair of pseudo-solutions ψ_h, ψ_k such that ψ_k has been generated from ψ_h. Since the generating procedure is such that

$$J_h^1 \subset J_k^1,\ J_h^o \subset J_k^o,\ |J_k^1|-|J_h^1| = 1 \tag{4.16}$$

i.e., ψ_k is generated from ψ_h by fixing at 1 a free component of y, an arc (h,k) will also be associated with the (unique) variable y_j which is free at node h and fixed at 1 at node k. For the same reason, any pseudo-solution ψ_t such that $J_h^1 \subset J_t^1$ and $J_h^o \subset J_t^o$, will be called a descendant of ψ_h, if actually generated, and a potential descendant otherwise.

The implicit enumeration procedure that we are going to apply to the elements of Y is based on the use of tests of the type introduced in [19]. We shall assume that $c^1 \geq 0$, which is not a restriction, since c_j^1, if negative, can always be made positive by a substitution of the form $y_j' = 1 - y_j$. Further, in order to be able to use in this context tests which place bounds on the value of the objective function, we compute a lower bound γ on c^2x (the existence of which follows from that of a finite optimum for (P)):

$$\gamma = \min_{x\geq 0} \{c^2x \mid A^1y + A^2x \geq b,\ 0 \leq y_j \leq 1,\ j \in N\} \tag{4.17}$$

We start with $V = \emptyset$ which admits an arbitrary $y \in Y$. We choose as a starting solution (root of $\mathfrak{A}$) $y = 0$, and set $g^* = +\infty$.

In order to describe a typical iteration, let us suppose that the last pseudo-solution generated was ψ_k, with the associated solution y^k satisfying (4.9), and that by solving $L(y^k)$ the system (4.9) has been augmented and updated so that it is not satisfied any more by y^k (we shall see that this is the situation at the beginning of each new iteration).

Let J_k^1, J_k^o and N_k be the index sets defined by (4.13), (4.14) associated with ψ_k, and let

(4.18) $$N_k^{i+} = \{j \in N_k | \alpha_{ij} > 0\} \ , \ N_k^{i-} = \{j \in N_k | \alpha_{ij} \leq 0\}, \ i \in V$$

(4.19) $$\hat{\beta}_i = \beta_i - \sum_{j \in J_k^1} \alpha_{ij}, \qquad i \in V$$

(4.20) $$V^+ = \{i \in V | \hat{\beta}_i > 0\}$$

We then proceed as follows:

<u>Step 1</u>. Compute

(4.21) $$\bar{\beta}_i = \hat{\beta}_i - \sum_{j \in N_k^{i+}} \alpha_{ij}, \qquad i \in V^+$$

If $\bar{\beta}_i > 0$ for some $i \in V^+$, backtrack (go to Step 4).

If $\bar{\beta}_i \leq 0$, $\forall i \in V^+$, go to Step 2.

<u>Step 2</u>. Let

(4.22) $$\hat{\beta}_{i_o} = \max_{i \in V^+} \hat{\beta}_i$$

Order the indices $j \in N_k^{i_o+}$ so that

(4.23) $$c^1_{j_1}/\alpha_{i_o j_1} \leq c^1_{j_2}/\alpha_{i_o j_2} \leq \dots \leq c^1_{j_t}/\alpha_{i_o j_t}$$

and find an index $j_r \in \{j_1,\dots,j_t\}$ such that

$$(4.24) \qquad \sum_{h=1}^{r-1} \alpha_{i_o j_h} \le \hat{\beta}_{i_o} < \sum_{h=1}^{r} \alpha_{i_o j_h}$$

Compute

$$(4.25) \qquad \Delta g = g^* - \gamma - \sum_{j \in J_k^1} c_j^1 - \sum_{h=1}^{r-1} c_{j_h}^1 - \frac{c_{j_r}^1}{\alpha_{i_o j_r}} \left(\hat{\beta}_{i_o} - \sum_{h=1}^{r-1} \alpha_{i_o j_h}\right)$$

where γ is defined by (4.17).

If $\Delta g \le 0$, backtrack (go to Step 4).

If $\Delta g > 0$, go to Step 3.

Steps 1 and 2 are exclusion tests meant to identify such nodes of $\mathcal{A}$ that cannot have among their potential descendants nodes associated with feasible solutions $y \in Y$ "better" than the currently best one. Thus, in the first test, if $\bar{\beta}_i > 0$ for some $i \in V^+$, then the i^{th} constraint cannot be satisfied by assigning whatever values (0 or 1) to the free variables. Hence, one can backtrack, i.e., abandon the current node of $\mathcal{A}$ (i.e., the current y) with all its potential descendants.

The second test consists in choosing the "most violated" constraint, and computing a lower bound on the "cost" of satisfying it by assigning values 1 to some of the free variables. Δg is the difference between g^* and this lower bound, the latter being expressed as a sum of γ (a lower bound on c^2x) and the rest of the expression on the right-hand side of (4.25) (a lower bound on c^1y). Hence, if $\Delta g \le 0$, no descendant of the current node can yield a lower value of g than g^*, and again we can backtrack, i.e., abandon the current node with all its potential descendants.

Other tests used in [18, 19] or suggested elsewhere in a similar context can also be introduced at this point.

Step 3. Generate the pseudo-solution ψ_{k+1} (and the associated node of $\mathcal{a}$), defined by

$$(4.26) \qquad J^1_{k+1} = J^1_k \cup \{j_1\}, \; J^o_{k+1} = J^o_k$$

where j_1 is given by (4.23), and update $\hat{\beta}_i$, $i \in V$, i.e., set

$$(4.27) \qquad \hat{\beta}_i = \beta_i - \sum_{j \in J^1_{k+1}} \alpha_{ij} , \qquad i \in V$$

If $\hat{\beta}_i > 0$ for some $i \in V$, set $k + 1 = k$ and go to Step 1.

If $\hat{\beta}_i \leq 0$, $i \in V$, introduce $y^s = y^{k+1}$, the solution associated with ψ_{k+1}, into the objective function of $L(y)$, and go to Step 5.

Step 4. Backtrack to the predecessor h of the current node k in $\mathcal{a}$. Let y_j be the variable associated with the backtracking arc. Update the sets N_h and J^o_h by replacing them through $N_h - \{j\}$ and $J^o_h \cup \{j\}$, respectively, i.e., remove j from the set of free indices by fixing y_j at 0. Go to Step 1. If backtracking is not possible (if we are at the root of $\mathcal{a}$ and instructed to backtrack), terminate: if $g^* < \infty$, the solution associated with g^* is optimal; if $g^* = \infty$, P has no feasible solution.

Step 3 generates a new solution by fixing a hitherto free variable at 1. If the solution associated with the new pseudo-solution obtained in this way is not feasible, the tests are repeated. If it is, one introduces the new vector y^s into the objective function of $L(y)$ and one goes to the step dealing with $L(y^s)$.

In Step 4 we backtrack to the predecessor of the current node, and by fixing at 0 the variable associated with the backtracking arc we make sure that the abandoned node and its potential descendants will never be visited in any future step.

Step 5. Solve (post-optimize) $L(y^s)$.

If $L(y^s)$ has an optimal solution u^s, add to (4.9) the constraint

$$(u^sA^1 - c^1)y \geq u^sb - g^* + \epsilon^s \quad . \tag{4.28}$$

Then, if $g^s < g^*$, update g^* in all constraints of type (4.7) by setting $g^* = g^s$. If $L(y^s)$ has no finite optimum, let $u^s + \lambda t^s$ be a feasible solution for any $\lambda \geq 0$. Add to (4.9) the constraint (4.28), and the constraint

$$t^sA^1y \geq t^sb \tag{4.29}$$

In all cases, if $|V| \leq 2n$, where $|V|$ stands for the current number of constraints (4.9), go to Step 1. Otherwise go to Step 6.

Step 6. If at Step 5 we have generated one constraint, drop from (4.9) the constraint i_* defined by

$$\hat{\beta}_{i_*} = \min_{i \in V} \hat{\beta}_i \tag{4.30}$$

If at Step 5 two constraints have been generated, drop from (4.9) the constraints i_*, defined by (4.30), and i_{**}, defined by

$$\hat{\beta}_{i_{**}} = \min_{i \in V - \{i_*\}} \hat{\beta}_i \tag{4.31}$$

Go to Step 1.

In Step 5 the solution (post-optimization) of $L(y^s)$ is used to generate one or two new constraints for (4.9). If the objective function of $L(y^s)$ at the optimum is smaller than g^*, the latter is replaced by the new value in all constraints of type (4.7).

Step 6 is meant to keep the number of constraints constant after a certain level has been reached, by eliminating the "loosest" constraint (or pair of constraints). The level chosen here, $2n$, is arbitrary, and can of course be changed (the more constraints are retained, the more efficient the tests tend to be, but the more time it takes to apply them).

From the above comments it should be clear that the algorithm ends in a finite number of iterations. The solution associated with the last value g^* is optimal; if $g^* = +\infty$, P has no feasible solution.

Indeed, Y is a finite set, and in the process of enumerating its elements we abandon a subset of elements (associated with a node of $\mathcal{a}$ and its potential descendants) only when we can conclude from the tests that there is no element of the subset which satisfies the current constraints and is "better" than the currently "best" element. On the other hand, Theorem 4.1 shows that a vector $y \in Y$ can possibly be "better" than the current "best" one only if it satisfies the current constraints (4.7),(4.8). Finally, the implicit enumeration technique is such that no abandoned node can ever be visited again - nor can any of the potential descendants of such a node be generated.

The above algorithm is closely related to the partitioning procedure of Benders [2]. The Benders procedure, however, prescribes for phase I the finding of an "optimal" $y \in Y$, i.e., one that maximizes $\hat{\beta}_{i_0}$, which implies the solution of an integer programming problem each time we get into phase I. Our procedure avoids this, and requires only the finding of a feasible $y \in Y$ in each phase I, so that the complete sequence of phases I amounts to solving one simple integer program. This procedure is essentially identical with the one described by Lemke and Spielberg [17], with the following minor differences:

(a) we work with $L(y)$ rather than its dual, which permits the use of a primal algorithm for the post-optimization required in each phase II; (b) we generate the lower bounds (4.17) and (4.25) and use them in what seems to be a strong exclusion test; (c) we work with a fixed number of constraints (4.9).

5. An Algorithm for Integer and Mixed-Integer Nonlinear Programming

We shall now discuss a generalization of the procedure described in section 4 to the nonlinear case [7,9].

Consider the mixed-integer nonlinear program

$$
\text{(P)} \qquad \begin{array}{l} \max f(y,x) \\ \quad F(y,x) \le 0 \\ \quad y \in Y \, , \; x \ge 0 \end{array}
$$

where $f(y,x)$ is a scalar function and $F(y,x)$ an m-component vector function of $y \in R^n$, $x \in R^p$, and $Y \subset R^n$ is the set of n-vectors with nonnegative integer components. This is a special case of problem (P2) of section 2, in which $m_1 = 0$.

Let $u \in R^m$ and let the function

$$
(5.1) \qquad K(y,x,u) \equiv f(y,x) - uF(y,x)
$$

be differentiable in y and twice differentiable in x.

The dual of (P), as defined in section 2, is then

$$
\text{(D)} \qquad \begin{array}{l} \max\limits_{y} \min\limits_{x,u} g = K(y,x,u) - x\nabla_x K(y,x,u) \\ \qquad \nabla_x K(y,x,u) \le 0 \\ \qquad y \in Y \, ; \; x,u \ge 0 \end{array}
$$

Problem (D) does not seem to be of any use in solving (P), since its inequality set contains the integer-constrained primal variables y, and its objective function is nonlinear in the latter. However, in section 3 we have introduced a linearized (in $y \in Y$) dual (D') of (P). We shall use a slightly different notation here, in that we shall continue to denote by y the integer-constrained variable of the dual, and shall let the newly introduced variable $s \in R^n$ to be continuous:

$$\max_{y} \min_{s,x,u} g' = K(s,x,u) - (s,x)\nabla_{s,x}K\,(s,x,u) + y\nabla_s K(s,x,u)$$

$$\text{(D')} \qquad \nabla_x K(s,x,u) \leq 0$$

$$y \in Y \;;\; s,x,u \geq 0$$

Here $\nabla_{s,x}K = (\nabla_s K, \nabla_x K)$, $\nabla_s K$ being the vector of partial derivatives of K in the components of s.

The inequality set of (D') is independent of the integer-constrained variables y; moreover, the objective function g' is linear in y. In view of the results of section 3, this opens the way to the approach of solving (P) by solving (D'). To restate those results relating (D') to (P) for the special case under consideration, we recall from section 2 that the regularity condition for the above problems (P), (D') is as follows:

(a) If (P) has an optimal solution $(\bar{y},\bar{x})$, the inequality set $F(\bar{y},x) \leq 0$ satisfies the Kuhn-Tucker [16] constraint qualification at $x = \bar{x}$.

(b) If (D') has an optimal solution $(\hat{y},\hat{s},\hat{x},\hat{u})$, the matrix $\nabla_x^2 K(\hat{s},\hat{x},\hat{u})$ is nonsingular.

Denoting by Z and W' the constraint sets of (P) and (D') respectively, the relevant parts of Theorem 3.1 become for this case

Theorem 5.1. Let f(y,x) and each component of -F(y,x) be differentiable and concave in y,x on the set $\{(y,x) \in R^n \times R^p \,|\, y,x \geq 0\}$, and assume that (P) and (D') meet the regularity condition. Then

a) If $(\bar{y},\bar{x})$ solves (P), there exists $\bar{u}\epsilon R^m$ such that $(\bar{y},\bar{s},\bar{x},\bar{u})$, where $\bar{s} = \bar{y}$, solves (D').

b) If $(\hat{y},\hat{s},\hat{x},\hat{u})$ solves (D'), then $\hat{y} = \hat{s}$ and $(\hat{y},\hat{x})$ solves (P).

c) In both cases a) and b),

$$(5.2) \qquad \max \{f(y,x) | (y,x)\epsilon Z\} = \max_{y} \min_{s,x,u} \{g' | (y,s,x,u)\epsilon W'\}$$

The proof of this theorem is along the same lines as that of Theorem 3.1, with the following observations:

α) The linearity of K(y,x,u) in u, along with the assumptions on f(y,x) and F(y,x) and the regularity condition, make up for assumptions 1,2, and 4 of Theorem 3.1. As to assumption 3 of that theorem, it is taken care of by the fact that $m_1 = 0$. Assumption 5 holds by the definition of Y.

β) The regularity condition required for Theorem 3.1 can be replaced by the weaker regularity condition stated above, because the duality theorems of Wolfe [13] and Huard [15] can now replace the one by Dantzig, Eisenberg and Cottle [10] in the proof of the above theorem.

Remark 1. In the regularity condition stated above, the Kuhn-Tucker constraint qualification can of course be replaced by that of Slater [20] or Arrow-Hurwicz-Uzawa [21], or any other constraint qualification under which the duality theorem of [13] holds. On the other hand, if the regularity condition for (D') is replaced by the weaker "low-value property" requirement introduced by Mangasarian and Ponstein [22], then the "strict" converse duality statement b), based on [15], has to be replaced by a weaker converse duality statement of the type [22]. In all these cases, the theorem can still serve as a basis for the algorithm to be described below.

Remark 2. If $R^n \times R^p$ reduces to R^n, i.e., (P) is a pure integer nonlinear program in y, its linearized dual (D') becomes a mixed-integer max-min problem (D^o) in nonnegative variables, otherwise unconstrained, and linear in the integer-constrained variables:

$$(D^o) \qquad \max_{y \in Y} \min_{s,u \geq 0} K(s,u) + (y-s)\nabla_s K(s,u)$$

Before discussing the algorithm, let us consider the case when

$$(5.3) \qquad K(y,x,u) = c^1y + c^2x + ub - u(A^1y + A^2x) + \frac{1}{2}(y,x)C\begin{pmatrix} y \\ x \end{pmatrix}$$

where $b, c = (c^1, c^2)$, $A = (A^1, A^2)$ and C are of appropriate dimensions, C being symmetric. (P) is then the mixed-integer quadratic program

$$(P) \qquad \begin{aligned} &\max c^1y + c^2x + \frac{1}{2}(y,x)C\begin{pmatrix} y \\ x \end{pmatrix} \\ &A^1y + A^2x \leq b \\ &y \in Y \ ; \ x \geq 0 \end{aligned}$$

whose dual is

$$(D) \qquad \begin{aligned} &\max_{y} \min_{x,u} ub - \frac{1}{2}(y,x)C\begin{pmatrix} y \\ x \end{pmatrix} - v^1y \\ &uA - (y,x)C - v = c \\ &y \in Y \ ; \ x,u,v^2 \geq 0 \ ; \ v^1 \text{ unconstrained} \end{aligned}$$

and whose linearized dual (D') is

$$\max_{y}\ \min_{t,x,u}\ ub - \frac{1}{2}(s,x)C\begin{pmatrix} s \\ x \end{pmatrix} - v^1 y$$

(D') $$uA - (s,x)C - v = c$$

$$y \epsilon Y\ ;\ s,x,u,v^2 \geq 0\ ;\ v^1 \text{ unconstrained}$$

No regularity condition is required for this case, and Theorem 5.1 becomes

Theorem 5.2. Let C be negative semi-definite. Then

a) If $(\bar{y},\bar{x})$ solves (P), there exists $\bar{u}\epsilon R^m$ such that $(\bar{y},\bar{s},\bar{x},\bar{u})$, where $\bar{s} = \bar{y}$, solves (D').

b) If $(\hat{y},\hat{s},\hat{x},\hat{u})$ solves (D'), there exists $\tilde{x}\epsilon R^p$ such that $(\hat{y},\tilde{s},\tilde{x},\hat{u})$, where $\tilde{s} = \hat{y}$, also solves (D'), while $(\hat{y},\tilde{x})$ solves (P).

The proof is along the same lines as for Theorem 3.1, with the use of the quadratic duality theorem of Cottle [12] in place of the strict non-linear duality theorem of [10].

We shall now discuss a method for solving integer or mixed-integer nonlinear programs, based on the above results. The basic idea of the method is to solve (D') instead of (P).

We shall consider the mixed-integer nonlinear program (P) introduced at the beginning of this section, and assume that f(y,x) and each component of -F(y,x) is concave and differentiable in y and x on $\{(y,x)\epsilon R^n X R^p | y,x \geq 0\}$. Further, we shall assume that the integer-constrained variables are bounded, i.e., Y is finite.

Now consider the linearized dual (D') of (P), which is a mixed-integer nonlinear problem in (y,s,x,u), with an objective function linear in y, and a constraint set independent of y. For any given $y \in Y$, (D') becomes a (continuous) nonlinear program in (s,x,u), which we shall denote by D'(y).

Let g'(y) be the objective function and W" the constraint set of D'(y), i.e.,

$$(5.3) \qquad W'' = \{(s,x,u) \mid \nabla_x K(s,x,u) \leq 0 \ , \ (s,x,u) \geq 0\}$$

We assume that $W'' \neq \emptyset$ (this is always the case when (P) has an optimal solution and meets the regularity condition).

The method we are going to discuss involves, as in the linear case, the solution of a sequence of problems D'(y) defined by a sequence of vectors $y \in Y$.

Since each problem $D'(\bar{y})$ is the dual of the concave program $P(\bar{y})$ obtained from (P) by setting $y = \bar{y}$ (see the proof of Theorem 3.1), one can solve D(y) by solving P(y) whenever the latter satisfies (or can be perturbed so as to satisfy) the required constraint qualification. By "solving" a problem D'(y) we mean finding an optimal solution or an ϵ-solution (in the sense defined, for instance, in [23]), or establishing the fact that D'(y) has no finite optimum. Further, we shall have to assume that at the end of the whole procedure, when an optimal solution (or ϵ-solution) to (D') has been found, the regularity condition required in Theorem 5.1 holds (or can be made to hold by some perturbation). However, this assumption is not needed in the case of a mixed-integer quadratic program, as it was mentioned above.

Now suppose we solve $D'(y)$ for $y = y^1,\ldots,y^q$, $\{1,\ldots,q\} = Q$, $(y^k \epsilon Y, k \epsilon Q)$. For each $k \epsilon Q$, exactly one of the following two situations holds:

a) $D'(y^k)$ has an optimal solution (or an ϵ-solution) (s^k,x^k,u^k).

b) $g'(y^k)$ is unbounded from below on W''.

For case b) we have

Theorem 5.3. If $g'(y^k)$ is unbounded from below on W'', there exist vectors $s^k \epsilon R^n$, $x^k \epsilon R^p$, $u^k \epsilon R^m$ and $t^k \epsilon R^m$, such that

(5.4) $$(s^k,x^k,u^k) \epsilon W'' \ , \ t^k \geq 0$$

(5.5) $$\nabla_x t^k F(s^k,x^k) \geq 0$$

and

(5.6) $$-t^k F(s^k,x^k) + (s^k,x^k)\nabla_{s,x} t^k F(s^k,x^k) - y^k \nabla_s t^k F(s^k,x^k) < 0$$

Proof. Let $e = (1,\ldots,1) \epsilon R^m$ and let $\xi \epsilon R^p$, $\xi \geq 0$ be such that $K(y^k,\xi,e) \equiv f(y^k,\xi) - eF(y^k,\xi)$ is finite. The existence of such a vector ξ follows from the assumption that $f(y,x)$ and $F(y,x)$ are differentiable (hence continuous). Then for any $(s,x,u) \epsilon W''$

$$g'(y^k) \geq K(s,x,u) + [(y^k,\xi) - (s,x)]\nabla_{s,x}K(s,x,u)$$

$$[\text{since } \xi\nabla_x K(s,x,u) \leq 0]$$

$$\geq K(y^k,\xi,u) \qquad [\text{by the concavity of } K(s,x,u)].$$

Since $K(y^k,\xi,e)$ is finite, it follows that for any finite $u \epsilon R^m$, $K(y^k,\xi,u)$ is also finite, and $g'(y^k)$ is bounded from below. Hence a necessary condition for $g'(y^k)$ to have no lower bound on W'' is the existence of s^k, x^k, u^k and t^k such that, if λ is a scalar,

a) $(s^k, x^k, u^k + \lambda t^k) \epsilon W''$ for arbitrary $\lambda \geq 0$, which implies (5.4) and (5.5)

b) for $(s,x,u) = (s^k, x^k, u^k + \lambda t^k)$ and $\lambda \geq 0$, $g'(y^k)$ is a decreasing function of λ, which implies (5.6) <u>Q.e.d.</u>

Having solved $D'(y)$ for $y = y^k \epsilon Y$, $k \epsilon Q = \{1,\dots,q\}$, let $Q = Q_1 \cup Q_2$, with

$$Q_1 = \{k \epsilon Q \,|\, D'(y^k) \text{ has an optimal solution } (s^k, x^k, u^k)\}$$

(5.7) $$Q_2 = \left\{ k \epsilon Q \,\middle|\, \begin{array}{l} g'(y^k) \text{ is unbounded from below on } W'' \text{ and} \\ s^k, x^k, u^k, t^k \text{ satisfy (5.4), (5.5) and (5.6)} \end{array} \right\}$$

For each $k \epsilon Q$, let g^k stand for the value of $g'(y^k)$ for $(s,x,u) = (s^k, x^k, u^k)$, i.e., let

(5.8) $$g^k = K(s^k, x^k, u^k) - (s^k, x^k) \nabla_{s,x} K(s^k, x^k, u^k) + y^k \nabla_s K(s^k, x^k, u^k)$$

Further, let

(5.9) $$g^* = \begin{cases} g^{k_o} = \max\limits_{k \epsilon Q_1} g^k & \text{if } Q_1 \neq \emptyset \\ -\infty & \text{if } Q_1 \neq \emptyset. \end{cases}$$

<u>Theorem 5.4.</u> Any $y \epsilon Y$ (if one exists) such that

(5.10) $$\min_{s,x,u} \{g'(y) \,|\, (s,x,u) \epsilon W''\} > g^*$$

satisfies the constraints

(5.11) $\quad y\nabla_s K(s^k,x^k,u^k) > g^* - g^k + y^k\nabla_s K(s^k,x^k,u^k)\,, \qquad k\epsilon Q$

a.

(5.12) $\quad -y\nabla_s t^k F(s^k,x^k) \geq t^k F(s^k,x^k) - (s^k,x^k)\nabla_{s,x} t^k F(s^k,x^k), \quad k\epsilon Q_2$

where s^k, x^k, u^k, t^k and g^k are defined by (5.7) and (5.8).

Proof. Suppose $y\epsilon Y$ does not satisfy (5.11) for $p\epsilon Q$. Since $(s^P,x^P,u^P)\epsilon W''$, this implies

$$\inf_{s,x,u} \{g'(y)\,|\,(s,x,u)\epsilon W''\} \leq K(s^P,x^P,u^P) - (s^P,x^P)\nabla_{s,x}K(s^P,x^P,u^P) + y\nabla_s K(s^P,x^P,u^P)$$

$$\leq g^*$$

which contradicts (5.10).

Now suppose $y\epsilon Y$ violates (5.12) for $p\epsilon Q_2$. Then, since $(s^P,x^P,u^P + \lambda t^P)\epsilon W''$ for any $\lambda \geq 0$, we have

$$\inf_{s,x,u} \{g'(y)\,|\,(s,x,u)\epsilon W''\} < K(s^P,x^P,u^P+\lambda t^P)-(s^P,x^P)\nabla_{s,x}K(s^P,x^P,u^P+\lambda t^P)+y\nabla_s K(s^P,x^P,u^P+\lambda t^P)$$

$$= K(s^P,x^P,u^P)-(s^P,x^P)\nabla_{s,x}K(s^P,x^P,u^P)+y\nabla_s K(s^P,x^P,u^P)$$

$$+ \lambda[-t^P F(s^P,x^P)+(s^P,x^P)\nabla_{s,x}t^P F(s^P,x^P)-y\nabla_s t^P F(s^P,x^P)]$$

But then in view of (5.7) and Theorem 5.3 the right-hand side, and hence also the left-hand side of the above expression can be decreased arbitrarily by increasing λ, which contradicts (5.10). <u>Q.e.d.</u>

Corollary 5.4. If there is no $y \epsilon Y$ satisfying the system (5.11),(5.12), then either

a) $Q_1 = \emptyset$ and (P) has no feasible solution, or

b) $Q_1 \neq \emptyset$ and the vector $y^* \epsilon Y$ associated with the last g^* defines an optimal solution to (P).

Proof. If $Q_1 = \emptyset$, $g'(y)$ has no lower bound on W" for any $y \epsilon Y$. Hence (Theorem 1, [13]) the dual of the convex program $D'(y)$ has no feasible solution for any $y \epsilon Y$, and so (P) itself has no feasible solution.

If $Q_1 \neq \emptyset$, denote by (s^*,x^*,u^*) the optimal solution to $D'(y^*)$. Then, if (D') meets the regularity condition, (y^*,x^*) is an optimal solution to (P) (Theorem 5.1). If not, and if the regularity condition is not required (like in the quadratic case), then the optimal solution to the concave (quadratic) program $P(y^*)$ obtained from (P) by setting $y = y^*$ is also an optimal solution to (P)(Theorem 5.2). Q.e.d.

Based on the above results, we can now formulate a procedure for solving integer or mixed-integer nonlinear programs with the required properties (shown in Theorems 5.1 and 5.2), which generalizes to these cases the algorithm discussed in section 4.

Phase I. Find $y^s \epsilon Y$ satisfying the linear inequalities (5.11),(5.12). (At the start this constraint set is vacuous; thus $y^1 \epsilon Y$ is arbitrary.) Go to Phase II.

Phase II. Solve $D'(y^s)$. If it has an optimal solution (ϵ-solution), generate a constraint (5.11) and, if $g^s > g^*$, update g^* (i.e., set $g^* = g^s$). If $g'(y^s)$ has no lower bound on W", generate a constraint (5.11) and a constraint (5.12). Then go to phase I.

Theorem 5.5. In a finite number of iterations, the algorithm consisting of phases I and II ends with the set (5.12),(5.13) having no feasible solution $y \epsilon Y$.

Proof. When a new constraint (5.12) or (5.13) is generated in phase II, it is violated by the last $y \epsilon Y$ found in phase I. Hence no constraint is generated twice (a new constraint, violated by y, cannot be identical with any of the old ones, satisfied by y); and no $y \epsilon Y$ is generated twice (a new $y \epsilon Y$, satisfying all current constraints, cannot be identical with any of the old ones, each of which violates at least one of the current constraints). Since Y is assumed to be finite, the theorem follows.

Remark. This proof is valid as long as all the constraints generated under the procedure are kept and used in each phase I. If they are not, convergence will depend on the non-redundancy (convergence) of the procedure for generating the elements of the finite set Y, as in the case of the algorithm of section 4. On the other hand, it is easy to see that the above convergence proof is not affected if in phase II, whenever $g'(y)$ has no lower bound on W", we generate only a constraint (5.12), instead of also generating a constraint (5.11). This may sometimes be preferable [7], as a direction vector t^s may be easier to obtain than the associated feasible solution (s^s,x^s,u^s) to $D'(y^s)$.

The procedure outlined above can be implemented in several ways.

Phase I is a search for a solution y to the constraints (5.11),(5.12) over the set Y. As shown in section 4, this search is not to be restarted from the beginning for each phase I; rather the successive applications of phase I should constitute successive stages of a single search process over Y. If $Y = \{0,1\}^n$, the implicit enumeration techniques known for linear programs in 0-1 variables, with their various exclusion tests, can be used here as in section 4. If Y is the set of nonnegative integers, then a technique of the type discussed in [16], p. 942-943, or in [7], can be used to transform the problem in integer variables into one in 0-1 variables at a relatively modest price in terms of problem size, and the implicit enumeration techniques are again applicable.

As to phase II, from a computational standpoint it seems preferable, whenever it is possible (see Theorems 5.1,5.2), to find an optimal solution to $D'(y^s)$ by solving the problem $P(y^s)$ obtained from (P) by setting $y = y^s$. If, for some $s \in Q_1$, $P(y^s)$ does not satisfy the constraint qualification at the optimum, the optimal solution of $P(y^s)$ may still yield an ε-solution to $D'(y^s)$. Should this not be the case, the current y^s can simply be dropped and another $y \in Y$ generated. This will not affect the convergence of the procedure, provided one makes sure that y^s is not repeated.

This procedure is perfectly valid (in fact, considerably simplified) in the special case when all the variables of (P) are integer-constrained. The inequality set of (D') is then vacuous, and (D') becomes the problem (D^o) shown in Remark 2 to Theorem 5.1. Since the concavity of $K(s,u)$ in s implies the relation

$$(5.13) \qquad K(s,u) + (y-s)\nabla_s K(s,u) \geq K(y,u)$$

which holds as an equality for s = y, phase II reduces to solving the problem $D^o(y^k)$ in u:

$$D^o(y^k) \qquad \min_{u\geq 0} K(y^k,u) \equiv \min_u \{f(y^k) - uF(y^k) | u \geq 0\}$$

Whenever $F(y^k) \leq 0$, $u^k = 0$ solves $D^o(y^k)$, and a constraint (5.11) which now becomes

$$(5.14) \qquad y\nabla_s f(y^k) > g^* - f(y^k) + y^k \cdot \nabla_s f(y^k)$$

is generated for phase I. Whenever $F_i(y^k) > 0$ for $i \in M^+$, $K(y^k,u)$ has no lower bound on $\{u \in R^m | u \geq 0\}$. Then the vector t^k such that $t_i^k = 1$ for $i \in M^+$ and $t_i^k = 0$ for $i \notin M^+$ defines a constraint of type (5.12) for phase I.

A detailed discussion of the above algorithm as specialized to integer and mixed-integer quadratic programming, along with numerical examples, is given in [7].

*

We shall now briefly explore the relationship of the procedure described in this section to some other methods.

As mentioned above, our method can be viewed as a generalization for the nonlinear case of the ideas underlying the partitioning procedure of Benders [2] or the closely related technique of Lemke and Spielberg [17].

While Benders' partitioning procedure is generally used for solving mixed-integer linear programs, it is in fact slightly more general than that. Benders partitions a mixed-variables program into two subproblems: a linear program (say, P^1) and a more general problem (say, P^2, which may be, for instance, an integer program--whether linear or not); then he solves the original problem by solving a sequence of subproblems P^1, P^2. But this partitioning method is subject to the following limitations (also valid for the Lemke-Spielberg algorithm):

1. The objective function and each constraint has to be separable with respect to the continuous variables, i.e., no term containing both integer and continuous variables is allowed.

2. The objective function and the constraints have to be linear in the continuous variables.

3. If the objective function and/or the constraints are not linear in the integer variables, then the subproblem P^2 will be a pure integer nonlinear program for which a solution method has yet to be found.

The algorithm described in the present paper does not have any of these limitations: 1 and 2 are not required, and 3 does not apply: our correspondent of Benders' subproblem P^2 is a pure integer linear program.

Furthermore, while Benders' partitioning method becomes meaningless when applied to a pure integer linear program (it replaces the integer program with itself), the algorithm discussed in the previous section replaces an integer nonlinear program by an integer linear program.

We shall now discuss the relationship between our method and the cutting plane method of Kelley [24] for nonlinear programming, which, as Kelley has shown, can be combined with Gomory's [25] cutting plane method for integer programming. The constraints (5.11),(5.12) generated in our procedure are hyperplanes that cut off portions of the set Y containing the current $y \in Y$, hence they can also be regarded as "cutting planes". But there are some basic differences:

1. Kelley's method generates a sequence of points outside the feasible set, which converges on a feasible point. The first point which is feasible, is also optimal, but no feasible point is available before the end of the procedure. In this sense it is a "dual" method. The same

is true when Kelley's method is combined with Gomory's one to solve an integer nonlinear program (in this case of course "feasible" means a solution which is also integer in the required components).

On the other hand, the method described in this paper generates a finite sequence of feasible and (occasionally) infeasible (but integer in the required components) points, with a subsequence of feasible points such that each point in the subsequence is strictly "better" than the previous one. At each stage, a currently "best" feasible solution is available. In this sense this is a "primal" method.

2. Kelley's cutting hyperplanes define a convex set S' containing the original constraint set S. The role of each newly generated hyperplane is to cut off a portion of the set S'-S containing the current (infeasible) solution. Similarly, Gomory's hyperplanes are meant to cut off a portion of the set S'-S", where S" is the convex hull of the feasible integer points. Thus, both types of hyperplanes cut off sets of points lying outside the feasible (integer-feasible) set.

In our procedure, two types of hyperplanes are generated. Both of them are hyperplanes in n-space, rather than (n+p)-space, i.e., in the space of the integer-constrained variables rather than the space of all variables, and they are used as constraints on the (and only on the) integer-constrained variables $y \in Y$. The main role belongs to the hyperplanes of type (5.11), which are meant to cut off as large a portion of Y (whether feasible or not) as a hyperplane containing the current point $\bar{y}$ can possibly cut off without cutting off any $y \in Y$ which could yield, in conjunction with an appropriate x, a "better" integer-feasible solution than the current "best" one. When hyperplanes of the type (5.12) are generated, they are meant to cut off portions of Y containing points which cannot yield, in conjunction with any x, a feasible solution.

3. In Kelley's procedure, a cutting plane is generated by replacing a constraint function by its first order Taylor series approximation in the neighborhood of the current solution. In the notation of this section, this would be

$$F_i(\overline{y},\overline{x}) + [(y,x) - (\overline{y},\overline{x})]\nabla F_i(\overline{y},\overline{x}) \leq 0$$

The dual problem does not play any role in the derivation of this constraint.

To give a comparable interpretation to the cutting planes generated in our procedure, consider the Lagrangian expression associated with the primal problem

$$K(y,x,u) \equiv f(y,x) - uF(y,x) \ .$$

If the current integer point $\overline{y}$ (in n-space) is such that the function $K(\overline{y},x,u)$ in (x,u) has a saddle-point at $(\overline{x},\overline{u})$, we generate a cutting plane by requiring the first order Taylor series approximation of $K(y,\overline{x},\overline{u})$ (considered as a function in y defined on $\{y|y \geq 0\}$) in the neighborhood of $y = \overline{y} = \overline{t}$ to satisfy

$$(5.15) \qquad K(\overline{s},\overline{x},\overline{u}) + (y-\overline{y})\nabla_s K(\overline{s},\overline{x},\overline{u}) > g^*$$

where g^* is defined by (5.9). It is easy to see that (5.15) is the same as (5.11).

If $K(\overline{y},x,u)$ has no saddle-point and $\overline{x},\overline{u}$ and $\overline{t}$ are such that $K(\overline{y},\overline{x},\overline{u} + \lambda\overline{t}) \to -\infty$ when $\lambda \to +\infty$, then two cutting planes are generated, one of the type (5.11) and a second one of the type (5.12). In each case the dual vector $\bar{u}$ (or $\bar{t}$) plays a key role in generating the constraints.

Hence, while our method also generates a certain type of cutting planes, it differs substantially from Kelley's.

References

[1] GALE, D., KUHN, H.W., AND TUCKER, A.W.: Linear Programming and the Theory of Games. Activity Analysis of Production and Allocation, T.C. Koopmans (ed.), Chapter 19.

[2] BENDERS, J.F.: Partitioning Procedures for Solving Mixed-Variables Programming Problems. Numerische Mathematik, 4, 1962, 238-252.

[3] STOER, J.: Duality in Nonlinear Programming and the Minimax Theorem. Numerische Mathematik, 5, 1963, 371-379.

[4] KAKUTANI, S.: A Generalization of Brouwer's Fixed Point Theorem. Duke Math. J., 8, 1941, 457-458.

[5] BALAS, E., Duality in Discrete Programming. Technical Report No. 67-5, July 1967, Department of Operations Research, Stanford University. Revised December 1967.

[6] GOMORY, R.E., AND BAUMOL, W.J.: Integer Programming and Pricing. Econometrica, 28, 1960, 521-550.

[7] BALAS, E.,, Duality in Discrete Programming II. The Quadratic Case. Management Science (Theory), 16, 1969, No. 1.

[8] __________, Duality in Discrete Programming III. Nonlinear Objective Function and Constraints. M.S.R.R. No. 121, February 1958, Carnegie-Mellon University.

[9] __________, Duality in Discrete Programming IV. Applications. M.S.R.R. No. 145, October 1968, Carnegie-Mellon University.

[10] DANTZIG, G.B., EISENBERG, E., AND COTTLE, R.W., Symmetric Dual Nonlinear Programs. Pacific Journal of Mathematics, 15, 1965, pp. 809-812.

[11] DORN, W.S., Duality in Quadratic Programming. Quarterly of Applied Mathematics, 18, 1960, pp. 155-162.

[12] COTTLE, R.W., Symmetric Dual Quadratic Programs. Quarterly of Applied Mathematics, 21, 1963, pp. 237-243.

[13] WOLFE, P., A Duality Theorem for Nonlinear Programming. Quarterly of Applied Mathematics, 19, 1961, pp. 239-244.

[14] MANGASARIAN, O.L., Duality in Nonlinear Programming. Quarterly of Applied Mathematics, 20, 1962, pp. 300-302.

[15] HUARD, P., Dual Programs. Recent Advances in Mathematical Programming (R. Graves and P. Wolfe, eds.), Wiley, 1963, pp. 55-62.

[16] KUHN, H.W. AND TUCKER, A.W., Nonlinear Programming. Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, 1951, pp. 481-492.

[17] LEMKE, C.F. AND SPIELBERG, K., Direct Search Algorithms for Zero-One and Mixed-Integer Programming. Operations Research, 15, pp. 892-914.

[18] BALAS, E., Discrete Programming by the Filter Method. Operations Research, 15, 1967, p. 915-957.

[19] BALAS, E., An Additive Algorithm for Solving Linear Programs with Zero-One Variables. Operations Research, 13, 1965, p. 517-546.

[20] SLATER, R., Lagrange Multipliers Revisited: A Contribution to Nonlinear Programming. Cowles Commission Discussion Paper, Mathematics 403, 1950.

[21] ARROW, K. J., HURWICZ, L. and UZAWA, S., Constraint Qualifications in Maximization Problems, Naval Research Logistics Quarterly, 8, 1961, p. 175-191.

[22] MANGASARIAN, O. L. and PONSTEIN, J., Minmax and Duality in Nonlinear Programming. Journal of Mathematical Analysis and Applications, 11, 1965, p. 504-518.

[23] WOLFE, P., Methods of Nonlinear Programming. Nonlinear Programming (J. Abadie, editor), North-Holland, 1967, p. 99-134.

[24] KELLEY, J. E., The Cutting Plane Method for Solving Convex Programs. Journal Soc. Ind. Appl. Math., 8, 1960, p. 703-712.

[25] GOMORY, R. E., Outline of an Algorithm for Integer Solutions to Linear Programs. Bull. of Am. Math. Soc., 64, 1958, No. 5.

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Graduate School of Industrial Administration Carnegie-Mellon University	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP Not applicable

3. REPORT TITLE

MINIMAX AND DUALITY FOR LINEAR AND NONLINEAR MIXED-INTEGER PROGRAMMING

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical report

5. AUTHOR(S) (First name, middle initial, last name)

Balas, Egon

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June, 1969	53	25

8a. CONTRACT OR GRANT NO. NONR 760(24) b. PROJECT NO. NR 047-048 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Management Sciences Research Report No.176 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES None	12. SPONSORING MILITARY ACTIVITY Logistics and Mathematical Statistics Br. Office of Naval Research Washington, D.C. 20360

13. ABSTRACT

This paper discusses duality for linear and nonlinear programs in which some of the variables are arbitrarily constrained. The most important class of such problems is that of mixed-integer (linear and nonlinear) programs. ~~Part I~~ introduces the duality constructions; ~~part~~ II discusses algorithms based on them.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

Unclassified

Security Classification

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Integer Programming Duality Non-linear Programming Optimization under Arbitrary Constraints Paritioning Methods						

DD FORM 1 NOV 65 1473 (BACK)

Security Classification