MANAGEMENT SCIENCE
Vol. 22, No. 4, December, 1975
Printed in U.S.A.

IMPROVED LINEAR INTEGER PROGRAMMING FORMULATIONS OF NONLINEAR INTEGER PROBLEMS*

FRED GLOVER
University of Colorado

A variety of combinatorial problems (e.g., in capital budgeting, scheduling, allocation) can be expressed as a linear integer programming problem. However, the standard devices for doing this often produce an inordinate number of variables and constraints, putting the problem beyond the practical reach of available integer programming methods.

This paper presents new formulation techniques for capturing the essential nonlinearities of the problem of interest, while producing a significantly smaller problem size than the standard techniques.

1. Introduction

Nonlinearities in integer programming are customarily handled by the use of techniques involving piecewise linear approximation [3], [14] or involving the transformation of a nonlinear function into a polynomial function of 0-1 variables [1], [11], and then transforming the polynomial function into a linear function of 0-1 variables [1], [25], [26]. Methods have also been devised for solving nonlinear integer programs directly, usually after conversion to 0-1 problems involving separable functions with certain monotonicity properties (such as polynomials—see, e.g., [4], [9], [10], [12], [13], [16], [17], [23]). The relative effectiveness of the transformed linear approach and the direct nonlinear approach depends on the problem to be solved—neither approach claims a uniform advantage over the other in all cases.

Nevertheless, the transformed linear approach often encounters a severe shortcoming. Standard procedures for "linearizing" nonlinear integer problems (including those of piecewise approximation) typically involve a radical increase in the number of problem variables and constraints. Consequently, the gains to be derived from dealing with linear functions (albeit of integer variables) are quite likely to be nullified by the increased problem size.

Methods to achieve more economical linear representations of 0-1 polynomial programming problems have been proposed in [7], [8] in an effort to expand the range of nonlinear problems for which the transformed linear approach may prove effective. A useful result in [8] demonstrated the possibility of linearizing 0-1 polynomial problems by the addition of variables that are *continuous* (or, more precisely, automatically 0-1 without explicitly imposing an integer restriction), thus giving rise to a linear integer problem containing the same number of integer variables as the polynomial problem. The advantage of this derives from the fact that the difficulty of integer programming problems is usually much more dependent on the number of integer variables than the number of continuous variables.

However, in the case of nonlinear integer problems that are not naturally 0-1, a great deal is lost to increased problem size in the initial conversion to a 0-1 problem before applying subsequent linearization. The purpose of this note is to give procedures for achieving improved linear representations of commonly encountered nonlinearities by giving special attention to this initial conversion as well as by examining situations in which the variables are already 0-1.

* Processed by Professor Arthur M. Geoffrion, Departmental Editor for Integer and Large-Scale Programming; received March 6, 1974, revised April 29, 1974 and October 22, 1974. This paper has been with the author 7 months for revision.

Copyright © 1975, The Institute of Management Sciences

2. A Simple Technique for Linearization by Discrete Variables

Many nonlinearities in integer programming appear in the form of polynomial functions (e.g., via the use of approximations), and of these a significant number involve terms no higher than the second order. Such "quadratic" integer programs arise commonly in capital budgeting [18], [19] and in scheduling [20], [21].

The standard approach to linearizing these polynomial functions is to express them first as functions of 0-1 variables and then to introduce new 0-1 variables to take the place of the cross product terms, simultaneously introducing auxiliary constraints to insure that the new variables will assume the appropriate values. Ways to accomplish these things economically are proposed in [7], [8]. However, it is possible to improve substantially on the previous proposals when dealing with certain types of nonlinearities.

A number of improvements can be realized by linking these nonlinearities to the following simple situation. Consider a variable w and a 0-1 variable x which are related to each other by the conditions $U_0 \geqslant w \geqslant L_0$ when $x = 0$, $U_1 \geqslant w \geqslant L_1$ when $x = 1$.

A natural way to model this situation is by the pair of inequalities

$$U_0 + (U_1 - U_0)x \geqslant w \geqslant L_0 + (L_1 - L_0)x. \tag{1}$$

Difficulties are encountered, however, if the U's and L's are not constants, but variables (e.g., expressions of other problem variables), since then $(U_1 - U_0)x$ and $(L_1 - L_0)x$ will usually be nonlinear, and we must find some way of dealing with these cross products in order to achieve a linear model. Assume for the moment that $U_1 \geqslant U_0$ and $L_1 \geqslant L_0$. Then we can linearize the cross product terms by means of an idea of Petersen [21]. Specifically, it is shown in [21] that cross products of the form xz, with z a nonnegative variable bounded above by a constant M, can be handled by replacing xz with a new variable y which is required to satisfy

$$Mx \geqslant y \geqslant z + Mx - M \quad \text{and} \quad z \geqslant y. \tag{2}$$

Thus, upon identifying appropriate upper bounds, each of the cross products of (1) can be accommodated by introducing a new variable and three new constraints by (2), or a total of two new variables and six new constraints to accommodate both of these cross products. A minor extension of Petersen's observation makes it possible to handle (1) similarly when $U_1 \geqslant U_0$ and $L_1 \geqslant L_0$ do not hold, but we can specify a more economical approach to dealing with (1) that also implies the extension of (2), and in which it is necessary only to introduce four new constraints and *no* new variables. To do this, we identify constants $\overline{U}_0$, $\underline{U}_0$, $\overline{L}_0$, $\underline{L}_0$, etc., such that

$$\overline{U}_0 \geqslant U_0 \geqslant \underline{U}_0, \quad \overline{L}_0 \geqslant L_0 \geqslant \underline{L}_0, \qquad \overline{U}_1 \geqslant U_1 \geqslant \underline{U}_1, \quad \overline{L}_1 \geqslant L_1 \geqslant \underline{L}_1,$$

where "upper bars" represent upper bounds and "lower bars" represent lower bounds. Then the appropriate set of inequalities is given by

$$U_0 + \left(\overline{U}_1 - \underline{U}_0\right)x \geqslant w \geqslant L_0 + \left(\underline{L}_1 - \overline{L}_0\right)x,$$

$$U_1 + \left(\overline{U}_0 - \underline{U}_1\right)(1 - x) \geqslant w \geqslant L_1 + \left(\underline{L}_0 - \overline{L}_1\right)(1 - x). \tag{3}$$

Note that the first pair of these inequalities is actually the same as (1) with the inclusion of bars in appropriate places, and the second pair is "equivalent" to (1) in a similar manner.

When $x = 0$, the first of these inequalities becomes $U_0 \geqslant w \geqslant L_0$, as desired, and (upon rearranging terms) the second becomes $\overline{U}_0 + (U_1 - \underline{U}_1) \geqslant w \geqslant \underline{L}_0 - (\overline{L}_1 - L_1)$.

Since $\bar{U}_0 \geqslant U_0$, $L_0 \geqslant \underline{L}_0$, and the quantities in parentheses are nonnegative, the second inequality is redundant relative to the first. In a similar manner, when $x = 1$ the second inequality of (3) becomes $U_1 \geqslant w \geqslant L_1$, as desired, and the first inequality becomes redundant.

The foregoing inequalities can help considerably to reduce the number of new variables and constraints standardly introduced to accommodate cross products. An example is the "quadratic" capital budgeting problem [15], [18], in which it is desired to minimize $\sum_{i,j \in N} x_i d_{ij} x_j$ subject to $x_i = 0$ or 1 for all $i \in N = \{1, \ldots, n\}$, with all remaining constraints linear. The standard approach introduces $n(n-1)/2$ new 0-1 variables (one for each cross product term). This can be improved in the manner of [8] which makes these variables continuous (automatically 0 or 1 when the original x_i variables are 0 or 1). By the use of (3) we can reduce the number of new variables to n *and* make them continuous. Specifically, define $w_i = x_i \sum_j d_{ij} x_j$. Then the desired restrictions translate into the following conditions involving w_i:

$$0 \geqslant w_i \geqslant 0 \qquad \text{if } x_i = 0,$$

$$\sum_j d_{ij} x_j \geqslant w_i \geqslant \sum_j d_{ij} x_j \qquad \text{if } x_i = 1.$$

Thus, appropriate constants for (3) (with $w = w_i$) are

$$\bar{U}_0 = \underline{U}_0 = \bar{L}_0 = \underline{L}_0 = 0,$$

$$\bar{U}_1 = \bar{L}_1 = D_i^+ = \text{the sum (over } j\text{) of the positive } d_{ij}\text{'s},$$

$$\underline{U}_1 = \underline{L}_1 = D_i^- = \text{the sum (over } j\text{) of the negative } d_{ij}\text{'s}.$$

Consequently, (3) becomes

$$D_i^+ x_i \geqslant w_i \geqslant D_i^- x_i, \quad \sum_j d_{ij} x_j - D_i^- (1 - x_i) \geqslant w_i \geqslant \textstyle\sum_j d_{ij} x_j - D_i^+ (1 - x_i)$$

for $w = w_i$ and $x = x_i$, and we have succeeded in modeling the quadratic objective function by introducing n new continuous variables (w_i, $i \in N$) and $4n$ constraints.

3. Handling Other Common Nonlinearities

Another, more general, nonlinearity is frequently encountered in an objective function of the form $\sum_{i \in N; j \in M} x_i d_{ij} y_j$ where, as before, the x_i are 0-1 but the variables y_j need not be so constrained. The procedure for handling this is the same as in the preceding section redefining D_i^+ and D_i^- appropriately to provide upper and lower bounds for $\sum_j d_{ij} y_j$. A more interesting case is when the integer variables x_i are not constrained to be 0 or 1. Such a situation typically occurs in quadratic capital budgeting problems in which a project is not merely to be accepted or rejected, but can be accepted at various levels of investment. It is still possible to accommodate this within the framework of (3) by expressing each x_i as a linear combination of 0-1 variables; e.g.,

$$x_i = 1x_{i1} + 2x_{i2} + \cdots + rx_{ir}, \quad x_{i1} + x_{i2} + \cdots + x_{ir} \leqslant 1, \tag{4}$$

where $r \geqslant x_i \geqslant 0$. Then

$$x_i \sum_j d_{ij} y_j = \sum_{k=1}^{r} k x_{ik} \sum_j d_{ij} y_j$$

and (3) can be applied for $x = x_{ik}$ and $w = w_{ik}$, where $w_{ik} = kx_{ik}\sum_j d_{ij}y_j$. Consequently, if D_{ik}^+ and D_{ik}^- are upper and lower bounds on $k\sum_j d_{ij}y_j$ we obtain the constraints

$$D_{ik}^+ x_{ik} \geqslant w_{ik} \geqslant D_{ik}^- x_{ik},$$

$$k\textstyle\sum_j d_{ij}y_j - D_{ik}^-(1 - x_{ik}) \geqslant w_{ik} \geqslant k\sum_j d_{ij}y_j - D_{ik}^+(1 - x_{ik})$$

thus introducing a total of nr new variables (w_{ik}) and $4nr$ constraints (after the initial replacement of each x_i by the 0-1 variables x_{ik}). Alternatively, one may express each x_i in the familiar "binary expansion" of 0-1 variables $x_i = x_{i1} + 2x_{i2} + 4x_{i3} + \cdots$ giving roughly $\log_2 r$ integer 0-1 variables for each x_i. This results, correspondingly, in about $n \log_2 r$ new w_{ik} variables and $4n \log_2 r$ constraints. However, I would like to suggest that this is one instance in which reducing the number of variables may not be particularly advantageous. The actual number of 0-1 solutions is not reduced in the binary expansion, and the structure of the "direct expansion" (4), in which a sum of variables cannot exceed 1, is highly exploitable both in the continuous and in the integer settings. (This is true both in branch and bound [2], [5], [24] and in cutting [6].)

Moreover, the direct expansion actually permits a more substantial reduction in the number of remaining new variables and constraints than the binary expansion. This is accomplished by the following generalization of (3).

Consider the situation in which

$$U_k \geqslant w \geqslant L_k \quad \text{for } x = k,\ k = 0, 1, \ldots, r. \tag{5}$$

Via the direct expansion (4) (suppressing the i subscript) the inequalities of (5) can be modeled by

$$U_k + (U - \underline{U}_k)(1 - x_k) \geqslant w \geqslant L_k + (L - \overline{L}_k)(1 - x_k) \quad \text{for } k = 0, 1, \ldots, r, \tag{6}$$

where the constants $\underline{U}_k$, $\overline{L}_k$, U and L satisfy

$$\underline{U}_k \leqslant U_k, \quad \overline{L}_k \geqslant L_k, \quad U \geqslant \operatorname*{Max}_k \{U_k\}, \quad L \leqslant \operatorname*{Min}_k \{L_k\}$$

and where, definitionally, $x_0 \equiv 1 - \sum_{k=1}^r x_k$. That (6) accomplishes the intended effect is immediately apparent; its form is essentially that of the second inequality of (3), and becomes exactly that of this inequality by the minor refinement of replacing U and L with $\overline{U}_k$ and $\underline{L}_k$, where $\overline{U}_k \geqslant \mathrm{Max}_{h \neq k}\{U_h\}$, $\underline{L}_k \leqslant \mathrm{Min}_{h \neq k}\{L_h\}$. To use (6) to linearize the expression $\sum_{i \in N; j \in M} x_i d_{ij} y_j$ it suffices to introduce a single variable $w_i = x_i \sum_j d_{ij} y_j$ for each $i \in M$ and require

$$k\sum_j d_{ij}y_j \geqslant w_i \geqslant k\sum_j d_{ij}y_j \quad \text{if } x_i = k.$$

Thus, letting D_{ik}^+ and D_{ik}^- represent upper and lower bounds on $k\sum_j d_{ij}y_j$, as before, and letting $D_i^+ = \mathrm{Max}\{D_{ik}^+\}$ and $D_i^- = \mathrm{Min}\{D_{ik}^-\}$, the inequalities of (6) become

$$k\textstyle\sum_j d_{ij}x_j + (D_i^+ - D_{ik}^-)(1 - x_{ik}) \geqslant w_i \geqslant k\sum_j d_{ij}x_j + (D_i^- - D_{ik}^+)(1 - x_{ik})$$

$$\text{for } i = 1, \ldots, n \text{ and } k = 0, 1 \ldots r,$$

introducing a total of only n new continuous variables and n (r + 1) constraints (as contrasted with nr new variables and $4nr$ constraints for the preceding use of the expansion (4)).

It should be noted that the direct expansion immediately accommodates the generalization of the foregoing to the case in which x_i is replaced by the nonlinear function $f(x_i)$, requiring only that $k\sum d_{ij}y_j$ be replaced by $f(k)\sum d_{ij}y_j$. The binary expansion, on the other hand, can be used in this situation only if one carries out a secondary linearization of the function $f(x_i)$, upon replacing x_i by its "bit terms." This secondary linearization consists in identifying the 0-1 polynomial function corresponding to $f(x_i)$ under such a replacement of variables, and then linearizing the cross product terms of this polynomial—altogether an arduous process requiring the creation of new 0-1 variables for the linearization of the polynomials.

Moreover, one can use the direct expansion (4) and still get by with only $\log_2 r$ integer variables, using an observation of David Sommer [22]. Suppose one has an equation of the form $x_0 + x_1 + \cdots + x_r = 1$ in 0-1 variables (x_0 may be viewed as a slack variable in the context of (4)). Then these variables can be treated as continuous upon introducing the new 0-1 variables y_i, $i = 0, \ldots, p$, where p is the smallest integer such that $2^p \geqslant r$, together with the associated equations

$$y_i = \sum_{h=0}^{p+i-1} \sum_{k \in S_{ih}} x_k, \qquad i = 0, \ldots, p,$$

where S_{ih} is the set of integers k satisfying $0 \leqslant k \leqslant r$ and $2^i + h \cdot 2^{i+1} \leqslant k \leqslant (h+1) \cdot 2^{i+1} - 1$. This is just a formal way of saying that y_p is the sum of all the x_k whose indexes fall in the right half of the interval $(0, 1, \ldots, 2^p)$, y_{p-1} is the sum of all x_k whose indexes fall in the right halves of the two intervals that result when the initial interval is split in two, and in general y_{p-q} is the sum of the x_k whose indexes fall in the right halves of the 2^{q-1} intervals that result when the initial interval is divided into 2^{q-1} equal parts. (The equation requiring the x_k to sum to one must be retained.) This essentially accomplishes the transformation $\sum_{k=0}^{r} k x_k = \sum_{i=0}^{p} 2^i y_i$ in such a manner that each 0-1 assignment of values to the y_i variables will automatically give appropriate 0-1 values to the x_k variables. (If $2^{p+1} - 1$ exceeds r, then the foregoing equation must be introduced explicitly, or its right-hand side constrained not to exceed r.)

In any event, as in the case of the ordinary binary expansion, it seems likely that this sort of reduction of the number of integer variables will not lead to real computational gains over the direct expansion, particularly if one uses approaches such as [6], [24] to exploit the structure in (4).

4. Concluding Remarks

The simple linearizations proposed in this note are designed to alleviate the difficulties frequently encountered in representing a nonlinear (or combinatorial) relationship by means of integer variables. Principally, the proposed procedures seek to reduce the traditional "blow up" in problem size that is often a disaster both for data manipulation and, more importantly, for the algorithms used to solve these problems. The suggested procedures are straightforward and easy to apply, thus providing a practical alternative to some of the more cumbersome formulations currently in use.

References

1. Balas, Egon, "Extension de l'algorithme additif à la programmation en nombres entiers et à la programmation non linéaire," *C.R. Acad. Sc. Paris* (May 1964).
2. Beale, E. M. L., "Selecting an Optimum Subset," in J. Abadie, ed., *Integer and Nonlinear Programming*, North-Holland, The Netherlands, 1970.
3. Dantzig, G. B., *Linear Programming and Extensions,* Princeton University Press, Princeton, 1962.
4. Ginsburgh, V. and Van Peetersen, A., "Un algorithme de programmation quadratique en variables binares," *Révue Françoise d'Informatique et de Recherche Opérationnelle*, Vol. 3 (1969), pp. 57–64.
5. Glover, Fred, "Flows in Arborescences," *Management Science*, Vol. 17, No. 9 (May 1971).
6. ———, "Convexity Cuts for Multiple Choice Problems," *Discrete Mathematics*, Vol. 6 (1973), pp. 221–234.
7. ——— and Woolsey, Eugene, "Further Reduction of Zero-One Polynomial Programming Problems to Zero-One Linear Programming Problems," *Operations Research*, Vol. 22, No. 1 (January-February 1974).

8. ——— AND WOOLSEY, R. E., "Converting the 0-1 Polynomial Programming Problem to a 0-1 Linear Program," *Operations Research*, Vol. 22, No. 1 (January-Feburary 1974).
9. HAMMER, PETER L., "A B-B-B Method for Linear and Nonlinear Bivalent Programming," *Operations Research*, Statistics and Economics Mimeograph Series No. 48, Technion, May 1969.
10. ——— AND HANSEN, P. "Quadratic 0-1 Programming," C.O.R.E. Division Paper, 72–19 (1972).
11. ——— AND RUDEAU, S., *Boolean Methods in Operations Research and Related Areas*, Springer, New York, 1968.
12. HANSEN, P., "Nonlinear 0-1 Programming by Implicit Enumeration," paper presented at the 7th Mathematical Programming Symposium, The Hague, September 1971.
13. ———, "Quadratic 0-1 Programming by Implicit Enumeration," in *Numerical Methods in Nonlinear Optimization*, ed., F. A. Lootsma, Academic Press, 1972, pp. 265–278.
14. HU, T. C., *Integer Programming and Network Flows*, Addison-Wesley, 1969.
15. LAUGHHUNN, D. J., "Quadratic Binary Programming with Applications to Capital Budgeting Problems," *Operations Research*, Vol. 18, No. 3 (May-June 1970).
16. LAWLER, E. AND BELL, M., "A Method for Solving Discrete Optimization Problems," *Operations Research*, Vol. 15 (1967), pp. 1098–1112.
17. MAO, J. C. AND WALLINGFORD, B. A., "An Extension of Lawler and Bell's Method of Discrete Optimization with Examples from Capital Budgeting," *Management Science*, Vol. 15 (1968), pp. 51–60. Corrections and Comments in *Management Science*, Vol. 15 (1969), p. 481.
18. MARKOWITZ, HARRY M., *Portfolio Selection*, Monograph 16, Cowles Foundation, Wiley, New York, 1959.
19. MERVILLE, LARRY J., "An Investment Decision Model for the Multinational Firm: A Chance-Constrained Programming Approach," Ph.D. Dissertation, University of Texas, May 1971.
20. MODER, J. J. AND PHILLIPS, C. R., *Project Management with CPM and PERT*, Reinhold, New York, 1964.
21. PETERSEN, CLIFFORD C., "A Note on Transforming the Product of Variables to Linear Form in Linear Programs," Working Paper, Purdue University, 1971.
22. SOMMER, DAVID C., "Computational Experience with the Ophelie-Mixed I.P. Code," talk presented at the International TIMS Conference, Houston, April 1972.
23. TAHA, HAMDY A., "A Balasian-Based Algorithm for 0-1 Polynomial Programming," Research Report No. 70-2, University of Arkansas, May 1970.
24. TOMLIN, J. A., "Branch and Bound Methods for Integer and Nonconvex Programming," in A. Abadie, ed., *Integer and Nonlinear Programming*, North-Holland, The Netherlands, 1970.
25. WATTERS, LAWRENCE J., "Reduction of Integer Polynomial Programming Problems to Zero-One Linear Programming Problems," *Operations Research*, Vol. 15 (1967).
26. ZANGWILL, WILLARD I., "Media Selection by Decision Programming," *Journal of Advertising Research*, Vol. 5, No. 3 (1965).