NBSIR 85-3206

Global Solutions to Factorable Nonlinear Optimization Problems Using Separable Programming Techniques

Garth P. McCormick

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
National Engineering Laboratory
Center for Applied Mathematics
Gaithersburg, MD 20899

June 1985

Sponsored by
U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
r Applied Mathematics
ns Research Division
urg, MD 20899

NBSIR 85-3206

GLOBAL SOLUTIONS TO FACTORABLE NONLINEAR OPTIMIZATION PROBLEMS USING SEPARABLE PROGRAMMING TECHNIQUES

Garth P. McCormick

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
National Engineering Laboratory
Center for Applied Mathematics
Gaithersburg, MD 20899

June 1985

Sponsored by
U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Center for Applied Mathematics
Operations Research Division
Gaithersburg, MD 20899

U.S. DEPARTMENT OF COMMERCE, Malcolm Baldrige, *Secretary*
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Director*

ABSTRACT

Many algorithms for obtaining global solutions to nonconvex optimization problems have been proposed in recent years. The methods farthest along computationally are those for separable problems. These use linear programming codes to solve sequences of LP problems formed from piece-wise linear approximations to the nonlinear functional forms. For a large class of optimization problems, called factorable programming problems, it is possible to create equivalent separable problems. This is done at a cost: additional variables and constraints. In this paper the procedure for creating the equivalent separable problems is outlined and a brief description is given of a global solution algorithm due to Falk. A small example is given illustrating the above techniques. The example is also solved using a more direct method. Application to the solution of nonlinear least squares is illustrated with another example. Discussion of areas of research for improving the efficiency of this approach concludes the paper.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Global Solutions to Separable Problems

Solving nonconvex programming problems, i.e., optimization problems where local minimizers may not be global minimizers was thought not long ago to be relegated to heuristic algorithms. (See [McCormick 1972a] for a survey of some of these methods.) Recent investigators (e.g. [Falk and Soland 1969], [Soland 1971], [Falk 1972], [Beale and Tomlin 1970], [Hoffman 1981], [Mancini and McCormick 1976], [Mancini and McCormick 1979], [McCormick 1980]) have developed rigorous algorithms for which convergence to global minimizers can be proved. The theories are well-established and some computational results are available.

The algorithms which have the most computational development are those which solve separable optimization problems using linear programming codes. Separable programming problems have the following form:

$$\begin{aligned} &\underset{x \,\varepsilon\, R^n}{\text{minimize}}\ F_o(x) \\ &\text{subject to } F_i(x) \leqslant b_i,\ i = 1, \ldots, m, \\ &\qquad \ell_j \leqslant x_j \leqslant L_j,\ j = 1, \ldots, n, \qquad (Q) \end{aligned}$$

where each $F_i(x)$ is written

$$F_i(x) = \sum_{j=1}^{n} F_{ij}(x_j),\ i = 0, 1, \ldots, m.$$

(A more general formulation allows for equality constraints: i.e. for $F_i(x) = b_i$, $i = m + 1, \ldots, p$. However, for simplicity in presentation only, the inequality constrained problem is considered in this section.)

The basic idea behind most of these methods is to approximate each nonlinear functional by a piece-wise linear functional and solve the resulting problem by solving a finite sequence of linear programming problems. The computer programs used to solve the example of this paper implemented the algorithm of Falk and is briefly described below. For more details the reader is referred to [Falk 1972] or Appendix A in [Grotte 1978].

The approximating problem of the original problem Q is obtained by replacing each function F_{ij} by a piece-wise linear approximation over the interval $[\ell_j, L_j]$. This involves selecting $p_j + 1$ grid points $y_{jo}, \ldots, y_{jp_j}$ in $[\ell_j, L_j]$, where $y_{jo} = \ell_j$ and $y_{jp_j} = L_j$. Suppose x_{j*} is some point in $[\ell_j, L_j]$. Let $y_{jk*}, y_{jk*+1}, \Theta_{yk*}$ be the unique values where

$$x_{j*} = \Theta_{jk*}y_{jk*} + (1 - \Theta_{jk*})\, y_{jk*+1},\ 0 \leq \Theta_{jk*} \leq 1.$$

Then

$$F_{ij}\,(x_{j*}) = \Theta_{jk*}F_{ij}(y_{jk*}) + (1-\Theta_{jk*})\, F_{ij}(y_{jk*+1}).$$

Figure 1 represents this type of approximation.

There are obviously many ways these $p_j + 1$ points can be selected. The computer program implementing this algorithm allows the user to specify the value p_j. It then computes equally spaced points.

Let $K_j = \{0, 1, \ldots, p_j\}$. The approximating problem can be stated in terms of the weights on the grid points, the Θ_{jk}'s:

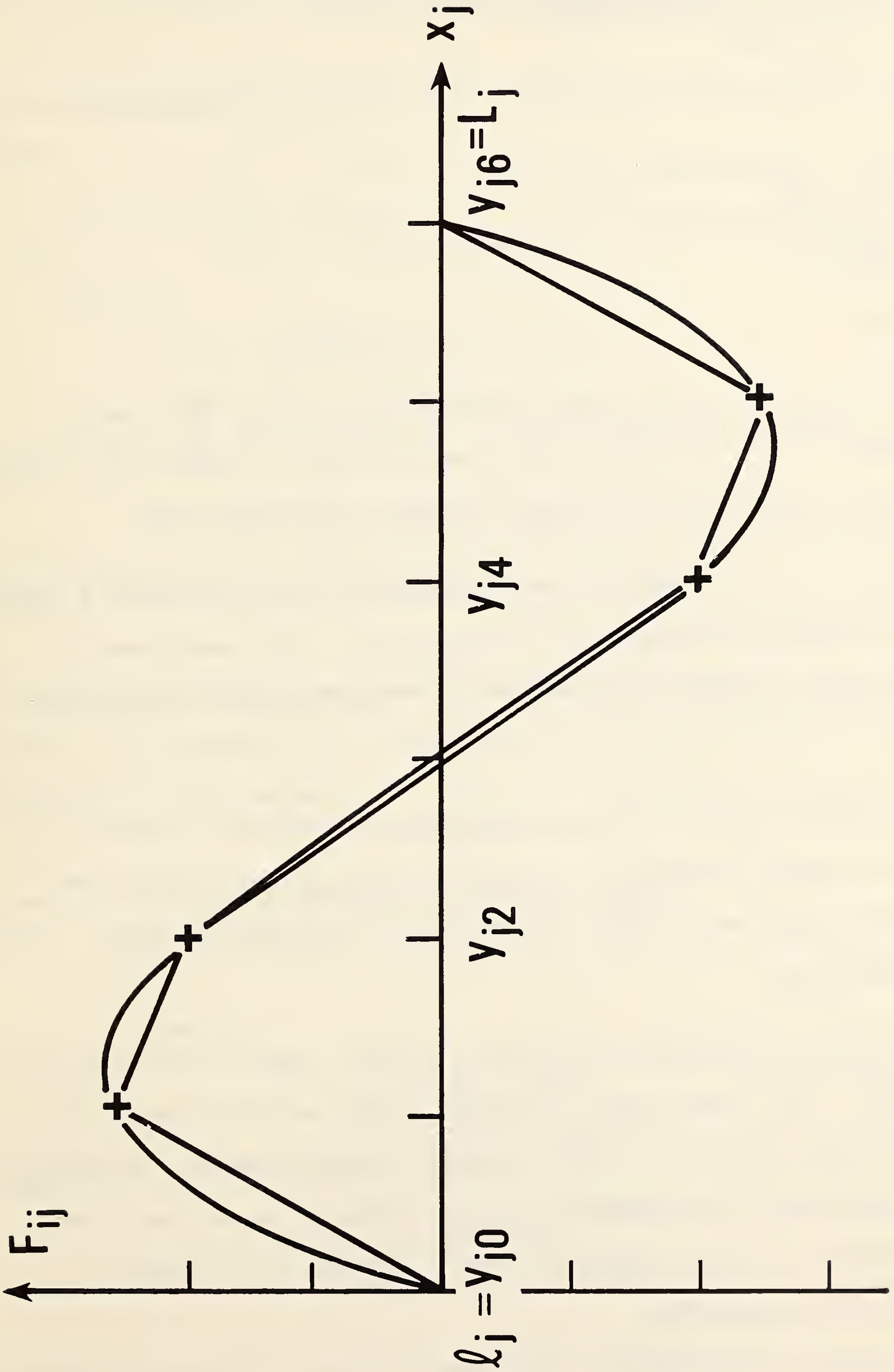

Figure 1. Piecewise Linear Approximations

$$\min_{\Theta} \sum_{j=1}^{n} \sum_{k\varepsilon K_j} \Theta_{jk} F_{oj} (y_{jk})$$

subject to

$$\begin{aligned} &\sum_{j=1}^{n} \sum_{k\varepsilon K_j} \Theta_{jk}F_{ij}(y_{jk}) \leq b_i, \quad i = 1, \ldots, m, \\ &\sum_{k\varepsilon K_j} \Theta_{jk} = 1 \quad j = 1, \ldots, n\ , \qquad \text{(P)} \\ &\Theta_{jk} \geq 0, \quad j = 1, \ldots, n;\ k\varepsilon k_j \end{aligned}$$

and (for j = 1, ..., n) at most two of the weights $\{\Theta_{jk}, k\varepsilon k_j\}$ can be nonzeros. If two are nonzero, they must correspond to adjacent grid points. This last constraint is the adjacency weighted restriction (AWR).

If it were not for the AWR, the approximating problem (P) would be a simple linear programming problem with variables $\{\Theta_{jk}\}$*. This restriction is necessary in order that the piecewise linear approximation indicated be valid.

In [Falk 1972] a branch and bound algorithm for solving (P) is presented. It involves solving a sequence of linear programming problems without the adjacency weight restrictions. The details of his algorithm however, will not be pursued here.

Most nonlinear programming problems are, in their simplest formulation, not separable. For a large class of problems, called factorable programming problems it is possible to create separable programs which are "equivalent" to the original ones in the sense that local minimizers are in a one-to-one correspondence. This is done at a cost; an increase in the number of variables and constraints.

The description of factorable optimization problems is in the next section.

2. Factorable Programming Problems

A factorable programming problem is one which can be written in the following form:

$$\min f_N(x) \equiv \sum_{p=1}^{N-1} T_{N,p}[f_p(x)] + \sum_{p=1}^{N-1} \sum_{q=1}^{p} U_{N,p,q}[f_q(x)]\cdot V_{N,q,p}[f_p(x)]$$

subject to

$$a_j \leq f_j(x) \equiv x_j \leq b_j,\ j = 1, \ldots, n, \tag{2.1}$$

$$a_j \leq f_j(x) \equiv \sum_{p=1}^{j-1} T_{j,p}[f_p(x)] + \sum_{p=1}^{j-1} \sum_{q=1}^{p} U_{j,p,q,}[f_q(x)]V_{j,q,p}[f_p(x)] \leq b_j,$$

$j = m+1, \ldots, N-1$ where some of the a_j's are possibly equal to $-\infty$, and the T's, U's and V's are scalar functions of one variable.

Most practical nonlinear programming problems are factorable. The reader is referred to [McCormick 1975] and [Ghaemi and McCormick 1979] for a fuller discussion of this subject.

Consider the following nonlinear programming problem:

$$\min_{(x_1,\, x_2)} \quad \text{ERF}\ (x_1+x_2) + \sin\ (x_1)\cdot \exp(-.5x_2)$$

subject to

$$-(x_1+x_2)^2 \leq -10 \tag{2.2}$$

$$0 \leq x_1 \leq 10,\ 0 \leq x_2 \leq 10,$$

where

$$k = 2/\sqrt{\pi},$$

$$\text{ERF}(z) = k\int_0^z \exp(-t^2)dt$$

Written as a factorable programming problem $f_j = f_j(x))$ this has the form:

$$\min_{(x_1,x_2)} \; f_5 = \mathrm{ERF}(f_3) + \sin(f_1) \cdot \exp(-.5f_2) \qquad (2.3)$$

subject to

$$0 \leq f_1 \equiv x_1 \leq 10$$

$$0 \leq f_2 \equiv x_2 \leq 10$$

$$-\infty \leq f_3 \equiv f_1 + f_2 \leq +\infty$$

$$-\infty \leq f_4 \equiv -(f_3)^2 \leq -10$$

The isovalue contours of the objective function are plotted in Figure 2. There are two local minimizers, one at (10,0) with objective function value .45599, and another at approximately ($3\pi/2$, 0) with objective function value approximately 0. If the inequality constraint $(x_1 + x_2)^2 \geq 10$ were removed from the problem, (0,0) would also be a local minimizer. The point ($3\pi/2$,0) is the global minimizer in both cases with a function value slightly less than zero.

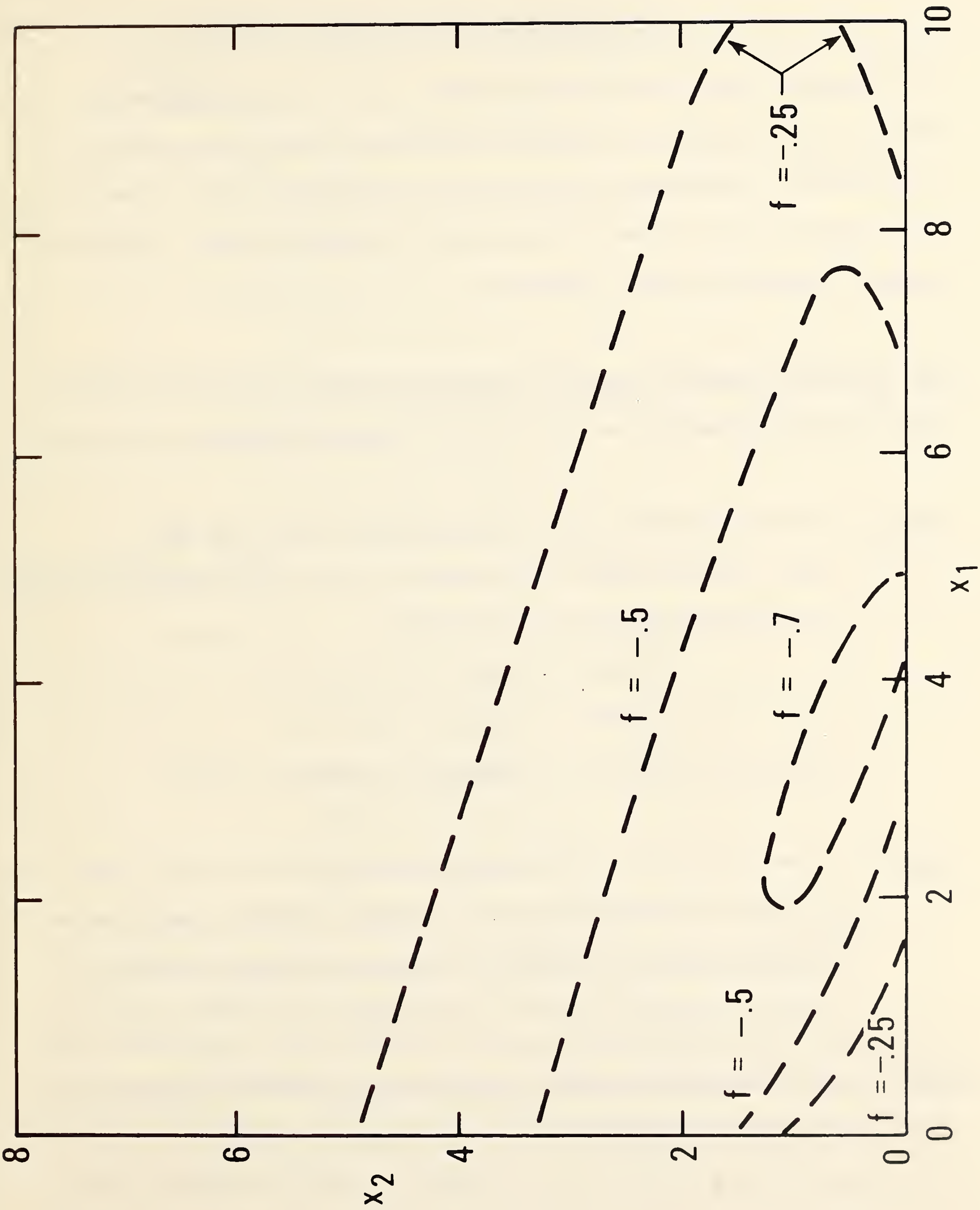

Figure 2. Isovalue Contours of Two Variable Problem

3. Creating Equivalent Separable Problems

The techniques which can be used to convert most nonlinear programming problems into "equivalent" separable programs are part of the tradition of optimization. The origins of these simple techniques are lost in the past. A description of two of these are contained in [McCormick 1972b]. Below is a summary statement of these techniques.

The method for creating equivalent separable problems has two basic steps which are used repeatedly until they can no longer be applied.

Step 1. If the optimization poblem has a product term of the form $q_1(x) \cdot q_2(x)$, replace that product by $(z_1)^2-(z_2)^2$ wherever it occurs, and introduce the equality constraints

$$q_1(x) = z_1 + z_2$$
$$q_2(x) = z_1 - z_2.$$

The problem variables are augmented to include z_1 and z_2.

Step 2. If the optimization problem has a term of the form $T[t(x)]$ where $T(t)$ is a scalar function and $t(x)$ is a scalar function of more than one variable, replace $T[t(X)]$ by $T(y)$ wherever it occurs, add the equality constraint $t(x) = y$, and augment the variable list with y.

Consider the example given earlier. Applying step 2 for the expression $t(x) = x_1+x_2$ yields the program

$$\min_{(x_1,x_2,y_1)} \mathrm{ERF}(y1) + \sin(x_1) \cdot \exp(-.5x_2) \tag{3.1}$$

subject to

$$-(y_1)^2 \le -10$$

$$x_1 + x_2 = y_1,$$

$$0 \le x_1 \le 10,$$

$$0 \le x_2 \le 10.$$

Application of Step 1 to the product term in the objective function yields the equivalent separable program

$$\min_{(x_1,x_2,y_1,z_1,z_2)} \mathrm{ERF}(y_1) + (z_1)^2 - (z_2)^2 \tag{3.2}$$

subject to

$$-(y_1)^2 \le -10$$

$$x_1 + x_2 = y_1,$$

$$\sin(x_1) = z_1 + z_2,$$

$$\exp(-.5x_2) = z_1 - z_2,$$

$$0 \le x_1 \le 10,\ 0 \le x_2 \le 10.$$

For this simple problem, separation occurred after just two applications of the separating techniques. Often the situation is more complicated.

The "equivalence" of the problems is shown in [McCormick 1972c], where it is shown that local minimizers are in one-to-one correspondence. For the purpose here it is sufficient to emphasize that if (x_1,x_2,y_1,z_1,z_2) is a local (global) minimizer for (3.2), then (x_1,x_2) is a local (global) minimizer for (3.1).

Not all nonlinear programming problems can be separated by repeated application of steps 1 and 2. In particular, if a program cannot be written as a factorable optimization problem it probably cannot be separated. Certainly, if it can be written in factorable form, separation is possible.

Make the correspondence $y_j = x_j$, $j = 1, \ldots, m$. Then using steps 1 and 2 on the factorable programming problem (2.1) yields directly

$$\min_{(y,z)} \sum_{p=1}^{N-1} T_{N,p}(y_n) + \sum_{p=1}^{N-1} \sum_{q=1}^{p} [(Z_{q,p}^{N,1})^2 - (Z_{q,p}^{N,2})^2] \qquad (3.3)$$

subject to the inequality constraints

$$a_j < y_j < b_j, \; j = 1, \ldots, N-1 \qquad (3.4)$$

and the equality constraints

$$y_j = \sum_{p=1}^{j-1} T_{j,p}(y_p) + \sum_{p=1}^{j-1} \sum_{q=1}^{p} [(Z_{q,p}^{j,1})^2 - (Z_{q,p}^{j,2})^2], \; j=m+1, \ldots, N-1; \qquad (3.5)$$

$$\begin{aligned} U_{j,p,q}(y_q) &= Z_{q,p}^{j,1} + Z_{q,p}^{j,2} \qquad & j &= n+1, \ldots, N; \\ V_{j,p,q}(y_p) &= Z_{q,p}^{j,1} - Z_{q,p}^{j,2} & p &= 1, \ldots, j-1, \\ & & q &= 1, \ldots, p. \end{aligned} \qquad (3.6)$$

Here $y = (x_1, \ldots, x_m, y_{m+1}, \ldots, y_{N-1})^T$, and z is the vector containing the $Z_{q,p}^{j,1}$, $Z_{q,p}^{j,2}$'s appearing above.

Application of these equations to the example problem given in form (2.3) yields the separable problem

$$\min \mathrm{ERF}(y_3) + (Z_{1,2}^{5,1})^2 - (Z_{1,2}^{5,2})^2 \qquad (3.8)$$

subject to

$$0 \leqslant y_1 \leqslant 10$$

$$0 \leqslant y_2 \leqslant 10$$

$$-\infty \leqslant y_3 \leqslant +\infty$$

$$-\infty \leqslant y_4 \leqslant +\infty \qquad (3.9)$$

$$y_3 = y_1 + y_2,$$

$$y_4 = (y_3)^2 \qquad (3.10)$$

$$\sin(y_1) = Z_{1,2}^{5,1} + Z_{1,2}^{5,2}$$

$$\exp(-.5y_2) = Z_{1,2}^{5,1} - Z_{1,2}^{5,2}\ ,$$

where the problem variables are

$$(y_1,\ y_2,\ y_3,\ y_4,\ Z_{1,2}^{5,1},\ Z_{1,2}^{5,2}\ )$$

Problem (3.8) is similar to the separable problem (3.2) obtained directly by applying steps 1 and 2 except that it has one variable more and one more equality constraint. This is an example of a general modification which should be made to the separation rules based to create (3.3). If some f_j ($j = m+1, \ldots, N-1$) does not enter into the computation of any future f_i, the introduction of the new variable y_j should be omitted, and no equality constraint of the form (3.5) should be included. The simple bound in (3.4) in this instance should remain

$$a_j \leqslant \sum_{p=1}^{j-1} T_{j,p}(y_p) + \sum_{p=1}^{j-1} \sum_{q=1}^{p} [(Z_{q,p}^{j,1})^2 - (Z_{q,p}^{j,2})^2] \leqslant b_j.$$

Using this modification, problem (3.8) above would be changed in that y_4 would no longer be a variable, (3.10) would be omitted, and (3.9) would become $-\infty \leqslant -(y_3)^2 \leqslant -10$. It is then equivalent, except for the variable names, to problem (3.2).

4. Obtaining Bounds on the New Variables

Algorithms which obtain global solutions to separable optimization problems invariably require upper and lower bounds on the variables. (An inductive process to do this when all original variables $\{x_j\}$ are bounded is given below).

The basic assumption here is that it is possible to compute the suprema and infima of T's, U's, and V's used in defining the factorable programming problem (2.1).

Define

$$\begin{aligned}
\ell(T,p,j) &= \inf T_{j,p}(y) \quad \text{s.t. } a_p \leqslant y \leqslant b_p,\\
u(T,p,j) &= \sup T_{j,p}(y) \quad \text{s.t. } a_p \leqslant y \leqslant b_p,\\
\ell(U,p,q,j) &= \inf U_{j,p,q}(y) \quad \text{s.t. } a_q \leqslant y \leqslant b_q,\\
u(U,p,q,j) &= \sup U_{j,p,q}(y) \quad \text{s.t. } a_q \leqslant y \leqslant b_q.
\end{aligned}$$

Analagous notation is used for the other quantities.

Then the lower bound on y_j is given by

$$y_j \geqslant \max\,(a_j, c_j)$$

where

$$c_j = \sum_{p=1}^{j-1} \ell(T,p,j) + \sum_{p=1}^{j-1} \sum_{q=1}^{p} \max\,[\ell(U,p,q,j)\cdot\ell(V,p,q,j), \ell(U,p,q,j)\cdot u(V,p,q,j),$$

$$u(U,p,q,j)\cdot\ell(V,p,q,j),\ u(U,p,q,j)\cdot u(V,p,q,j)]$$

Analogous bounds are computed for the other quantities.

Consider the separable problem created in section 3. Since $y3 = y_1 + y_2$ and the bounds on both y_1 and y_2 are 0 and 10,

$$0 < y_3 < 20.$$

(Recall y_4 was eliminated from the problem as unnecessary). Now

$$Z_{1,2}^{5,1} = [\sin(y_1) + \exp(-.5y_2)]/2$$

and

$$Z_{1,2}^{5,1} = [\sin(y_1) - \exp(-.5y_2)]/2.$$

Since

$$-1 = \inf \sin(y_1) < \sin(y_1) < \sup \sin(y_1) = 1,$$

where

$0 < y_1 < 10$

and

$0 = \inf \exp(-.5y_2) < \exp(-.5y_2) < \sup \exp(-.5y_2) < 1$ where $0 < y_2 < 10$,

it follows that the bounds on these new variables are

$$-.5 < Z_{1,2}^{5,1} < 1,$$

$$-1. < Z_{1,2}^{5,2} < .5.$$

5. Computer Solution To Simple Problem

The computer program which implements Falk's branch and bound procedure (see section 1) is called NUGLOBAL and is available from the Operations Research Division of the National Bureau of Standards. It uses a linear programming package called SEXOP written by Dr. Roy E. Marsten.

The user is required to write a FORTRAN subroutine (GETPHI) which supplies the values of the nonlinear functions defining the separable optimization problem. Input to this routine is the constraint index i, variable index j, and variable value x_j. The output is $F_{ij}(x_j)$. For the simple two variable example the SUBROUTINE GETPHI is listed in Appendix A.

The user is also required to supply a data file. The one used for this example is given in Appendix A. (This corresponds to the Run 3 to be discussed next.)

Computer instructions on how to use the NUGLOBAL system are in [Hoffman 1975].

For notational convenience and to conform to the separable format, problem (3.2) is rewritten

$$\min_{(x_1,\ \dots,\ x_5)} \mathrm{ERF}(x_3) + (x_4)^2 - (x_5)^2$$

Subject to:

$$\begin{aligned} x_1 + x_2 - x_3 &= 0 \\ -\sin(x_1) + x_4 + x_5 &= 0 \\ -\exp(-.5x_2) + x_4 - x_5 &= 0 \\ - x_3 &\leq -10 \\ 0 \leq x_1 &\leq 10, \\ 0 \leq x_2 &\leq 10, \\ 0 \leq x_3 &\leq 20, \\ -.5 \leq x_4 &\leq 1, \\ 1 \leq x_5 &\leq .5. \end{aligned}$$

For "Run 1" the first variable was "divided into nine variables", i.e. p_1=8 (see Section 10). That is, in the linear programming problem $\Theta_{1,0}, \ldots, \Theta_{1,8}$ "represented" x_1. Variables $x_2.x_3.x_4.x_5$ were replaced by 6 variables each. The resulting linear programming problems had 33 Θ variables.

To get the global minimizer to problem (P) required the solution of fifteen linear programming problems. The solution obtained was

$$x^* = (5.0,\ 0.0,\ 5.0,\ 0.020538,\ -0.97946).$$

The value of the piece-wise linear approximation to the objective function was (.052861). The value of the true objective function at this point is (0.04107572).

The second run ("Run 2") used $p_j = 10$, j=1, ..., 5. This created linear programming problems with 55 Θ variables. This required the solution of 12 linear programming problems to obtain the global solutions to Problem (P).

The solution obtained was the same as run 1, x*=(5.0, 0.0, 5.0, 0.020538, -0.97946). The value of the piece-wise linear objective function was (0.04402). The true objective function value there is, as before, (0.04107).

These two values are closer than in run 1 because of the finer grid approximation in Run 2.

In Run 3, $p_j = 25$, j = 1, ..., 5. Each x_j was represented by 26 Θ variables. Each linear programming problem had 130 variables. Nine linear programs were required to find the solution x* = (4.8, 0.0, 4.8, 0.0019177, -0.99808). The true objective function value there is (0.00383) with piece-wise linear approximate value of (0.00455). These results are summarized in Table 1.

# of LP Vars.	LPS Solved	x_1	x_2	x_3	x_4	x_5
Run 1 33	15	5.0	0.0	5.0	.0205	-0.9795
Run 2 55	12	5.0	0.0	5.0	|0205	-0.9795
Run 3 130	9	4.8	0.0	4.8	.0019	-0.9981
Theoretical Solution		4.71 $=3\pi/2$	0.0	4.71	0.0	-1.0

Table 1 - Computer Effort Required to Solve LP Approximating Problems

6. Direct Solution of Simple Problem

Another approach for finding global minimizers to nonconvex programming problems (see [McCormick 1976] for more details) is to solve a sequence of "convex underestimating problems." Convex programming problems are problems which can be written:

$$\text{minimize } f(x)$$

$$\text{subject to}$$

$$g_i(x) \geqslant 0,\ i = 1, \ldots, m$$

where the functions

$$f,\ \{-g_i\}_{i=1}^{m}$$

are convex functions.

The nice property of convex programming problems is that local minimizers are global minimizers. The isovalue contours resemble ellipses, parabolas, or lines. It is clear from Figure 2 that the objective function of problem (2.2) is not a convex function.

The constraints of a convex programming problem define a convex set for the feasible region. For this simple problem, the constraint $(x_1 + x_2)^2 \geqslant 10$ does not define a convex set. However, the intersection of the set of points satisfying this with the feasible region is a convex set.

A convex underestimating problem for a general problem is one where the objective function is convex and underestimates the given objective function in the feasible region and where the constraint region is a convex set and contains the original feasible region. Presented in [McCormick 1976] is an algorithm for finding a convex underestimating problem when the original optimization problem is factorable with bounded variables.

The key to implementing this approach is the ability to compute, for a function of a single variable in an interval, its convex envelope. This is the highest convex function which underestimates the function in the interval. Thus, for example, the convex envelope of $\sin(x_1)$ where $0 \leq x_1 \leq 10$ is

$$\text{vexsin}(x_1) = \begin{array}{ll} -0.21723\, x_1, & 0 \leq x_1 \leq 4.49342 \\ \sin(x_1), & 4.49342 \leq x_1 \leq 4.79946 \\ (-1.41353 + 0.08695 x_1), & 4.79945 \leq x_1 \leq 10 \end{array}$$

This is shown in Figure 3, where the notation vex $f(\cdot)$ is used. Proper notation would indicate the interval. Thus vexsin $[x_1, 0, 10]$ should be used, but for notational simplicity this is not done.

$$\begin{array}{ll} -.21723x & 0 \leq x \leq 4.493719 \\ \text{Sin}(x) & 4.493419 \leq x \leq 4.7994496 \\ -1.4135281 & 4.7994496 \leq x \leq 10. \\ +\ .0869507x & \end{array}$$

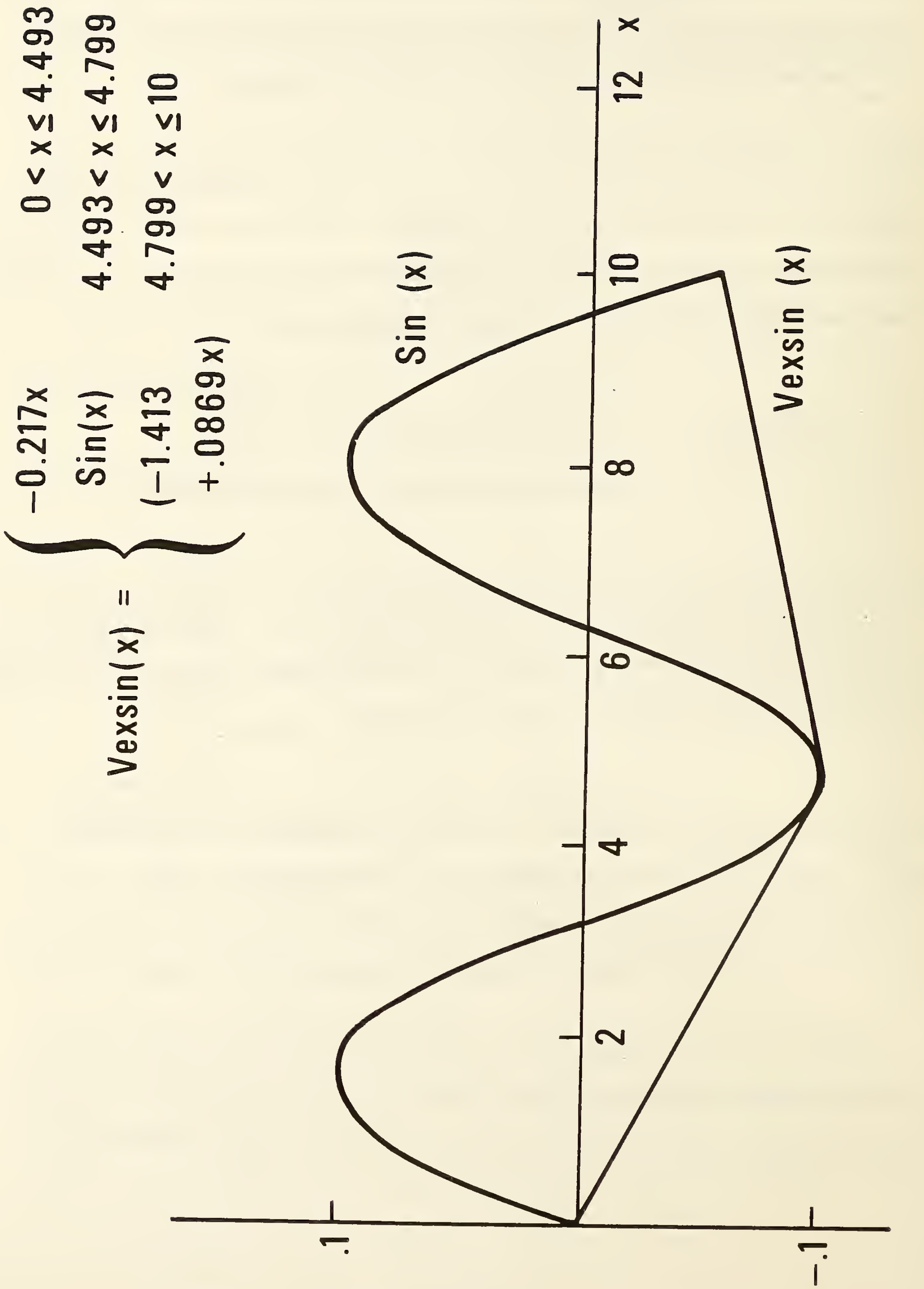

Figure 3. Convex Envelope of Sine Function [0, 10]

Using the general techniques in the above referenced article, a convex underestimating function for the objective function of example problem (2.2) will be constructed.

Since $0 \leq x_1 \leq 10$ and $0 \leq x_2 \leq 10$, then $0 \leq x_1 + x_2 \leq 20$. Consider

$$\phi(x_1 + x_2) = k \int_{o}^{x_1+x_2} \exp(-t^2)dt$$

in this interval. It is clear from Figure 4, that $.05\ (x_1 + x_2)$ is a "high" convex underestimating function in this interval to a close approximation.

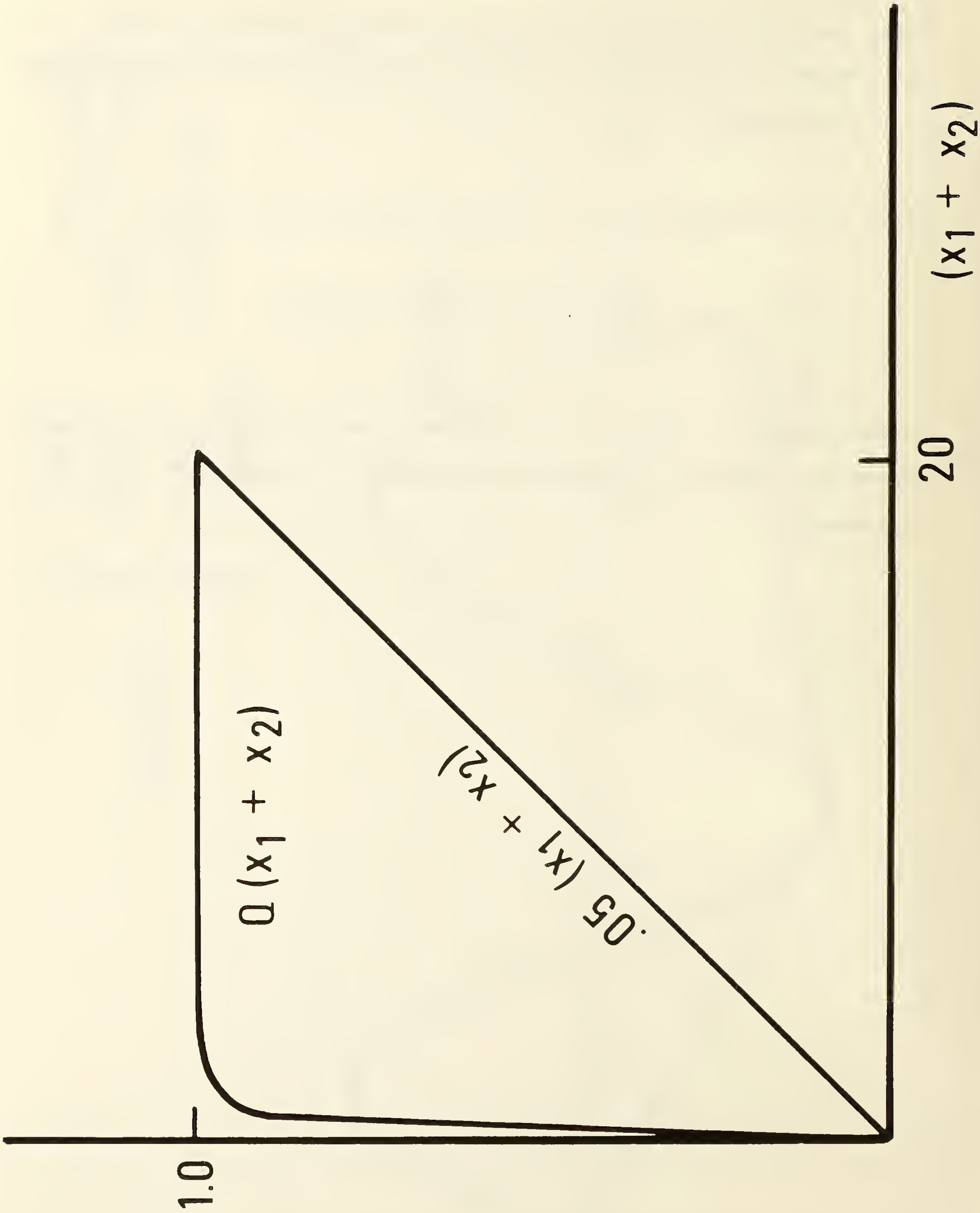

Figure 4. A Convex Underestimating Function

Since $0 \leq x1 \leq 10$, then $-1 \leq \sin (x_1) \leq 1$. Since $0 \leq x_2 \leq 10$, then

$$c_1 \equiv \exp(-5) \leq (-.5x_2) \leq 1.$$

Thus,

$$[\sin(x_1) + 1] \; [\exp(-.5x_2) - c_1] \leq 0,$$

or

$$\begin{aligned} \sin(x_1) \cdot \exp(-.5x_2) &\geq -\exp(-.5x_2) + c_1 \sin(x_1) + c_1 \\ &\geq -1 + c_2x_2 + c_1 \; [\text{vexsin} \; (x_1)] + c_1 \qquad (6.2) \end{aligned}$$

where

$$c_2 = (-c_1 + 1)/10 = .0993262.$$

Note that $-1 + .099326x_2$ is the convex envelope of $-\exp(-.5x_2)$ in the interval $[0, 10]$.

Also,

$$[\sin(c_1) - 1] \; [\exp(-.5x_2) - 1] \geq 0,$$

or

$$\begin{aligned} \sin(x_1) \cdot \exp(-.5x_2) &\geq \exp(-.5x_2) + \sin(x_1) - 1 \\ &\geq \exp(-.5x_2) + \text{vexsin} \; (x_1) -1. \qquad (6.3) \end{aligned}$$

Since the maximum of two convex functions is convex, a convex underestimating function for the product takes the form (combining (6.2) and (6.3))

$$\sin(x_1) \cdot \exp(-.5x_2) \geq \max[c_1 - 1 + c_2x_2 + c_1\text{vexsin}(x_1), \; \exp(-5x_2) + \text{vexsin}(x_1) - 1]. \qquad (6.4)$$

Using this with (6.1) gives the full convex underestimating function:

$$k \int_{o}^{x_1+x_2} \exp(-t^2)dt + \sin(x_1) \cdot \exp(-.5x_2) \qquad (6.5)$$

$$\geq .05(x_1 + x_2) + \max[c_1 - 1 + c_2x_2 + c_1\text{vexsin}(x_1), \; \exp(-.5x_2) + \text{vexsin}(x_1) - 1].$$

Isovalue contours of this function are plotted in Figure 5.

The global minimizer for function (6.5) in the rectangle $0 \leq x1 \leq 10$, $0 \leq x_2 \leq 10$ is (4.6623681, 0.). The first component is a solution of the equation $\arccos(x_1) = -.05$. This is to be compared with the true solution of the original problem (4.712389, 0.).

Development of the convex underestimating function above took no advantage of the restriction that

$$-(x_1 + x_2)^2 \leq -10$$

must hold. A convex underestimating function of $-(x_1 + x_2)^2$ in the rectangle

$$0 \leq x_i \leq 10, \; i = 1, 2$$

is

$$-20\,(x_1 + x_2) \leq -(x_1 + x_2)^2.$$

Thus the constraint can be imposed

$$-20\,(x_1 + x_2) \leq -10,$$

or

$$x_1 + x_2 \geq .5.$$

The convex underestimating function of the objective function of (2.2) can now be restricted to the region where $.5 \leq (x_1 + x_2) \leq 20$. This gives the tighter convex underestimating function

$$.024593(x_1 + x_2) + .5091315.$$

Using this with (6.4) gives a tighter convex underestimating function and the resulting solution of this problem is (4.687793, 0.). The first component is a solution of $\arccos(x_1) = -.024593$.

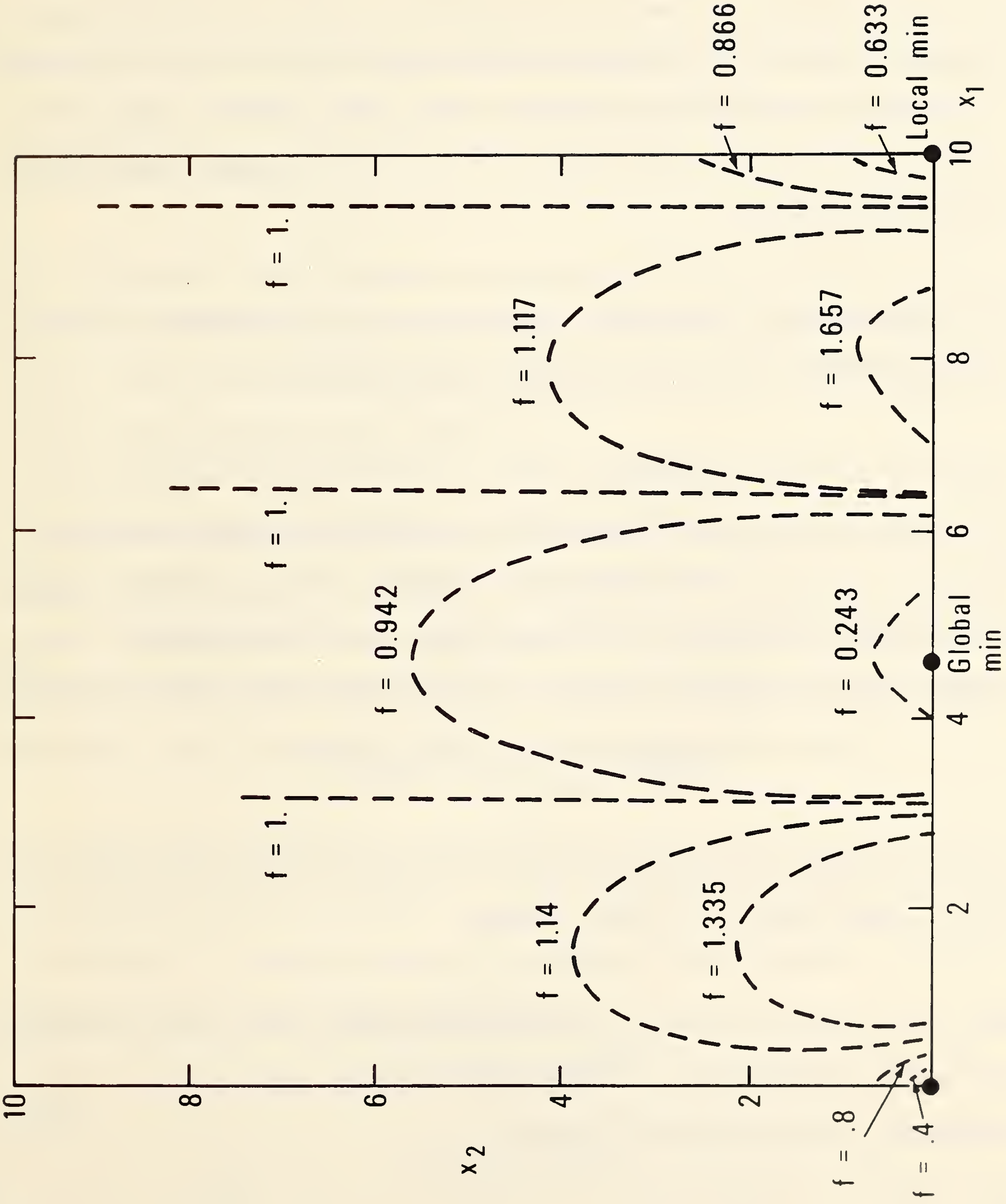

Figure 5. Isovalue Contours of Convex Underestimating Function

7. Application To Nonlinear Least Squares

An important source of nonlinear programming problems is the area of parameter estimation. In particular the finding of the "best" parameter values which define a functional form often gives rise to an unconstrained, nonconvex optimization problem.

Let x be a vector of "independent" variables, and y a "dependent" variable which is thought to be related to x via the equation

$$y = f\ (x,E),$$

where E is a vector of parameters and f is a predetermined functional form. Let $\{y_i\}$, i = 1, ..., r be observations of y simultaneous with observations $\{x_i\}$, i = 1, ..., r of the independent variables. There is a body of theory (See [Goldfeld and Quandt 1972]) which states, under the appropriate assumptions, that the "best" values of $\{E_j\}$ are those which solve the problem

$$\underset{E}{\text{minimize}} \quad \sum_{i=1}^{r} [y_i - f(x_i,E)]^2.$$

In the reference above is also a theory of the probability distribution of the point estimate given by the above problem.

When f is linear in E, this is the well-known linear least squares problem and the function minimized is a positive (semi) definite quadratic form. Local minimizers of this are global minimizers.

When f is nonlinear in the parameters $\{Ej\}$ the problem is much more complicated and the possibility that local, but not truly global solutions of that problem exist must be considered. When the E_j's are physical constants to be estimated, as in many scientific applications, this difficulty has serious implications.

In this section a small parameter estimation problem is considered. The general separation method given in Section 3 plus some ad hoc techniques are used to create an equivalent separable problem. This is in turn solved by the solution algorithm of Section 1.

It should be noted that the number of variables in the separated problem is greater than or equal to the number of parameters to be estimated plus the number of observations (r). If the function f is complicated, more separation is required and this can give rise to a separable problem in a large number of variables with an accompanying large linear approximation problem.

In Figure 6 is plotted the fraction of college men married versus age. The data for the problem is in Table 2. It is thought that the dependent variable y (fraction of college men married) is related to the independent variable x (age in years) by the functional equation

$$y = p\Phi[(x-\mu)/\sigma]$$

where

$$\Phi(z) = \int_{-\infty}^{y} \phi(t)dt,$$

and

$$\phi(t) = (2\pi)^{-.5} \exp(-t^2/2).$$

(The normal density function.)

The parameters, whose best values are to be estimated from the data of Table 2 are p (the ultimate fraction of these married), μ (the mean of the distribution) and σ (the standard deviation).

i	1	2	3	4	5	6	7
x_i	14.6	16.8	18.7	20.6	23.1	27.1	32.0
y_i	0.000	0.004	0.015	0.075	0.315	0.571	0.737

Table 2
Data for Example

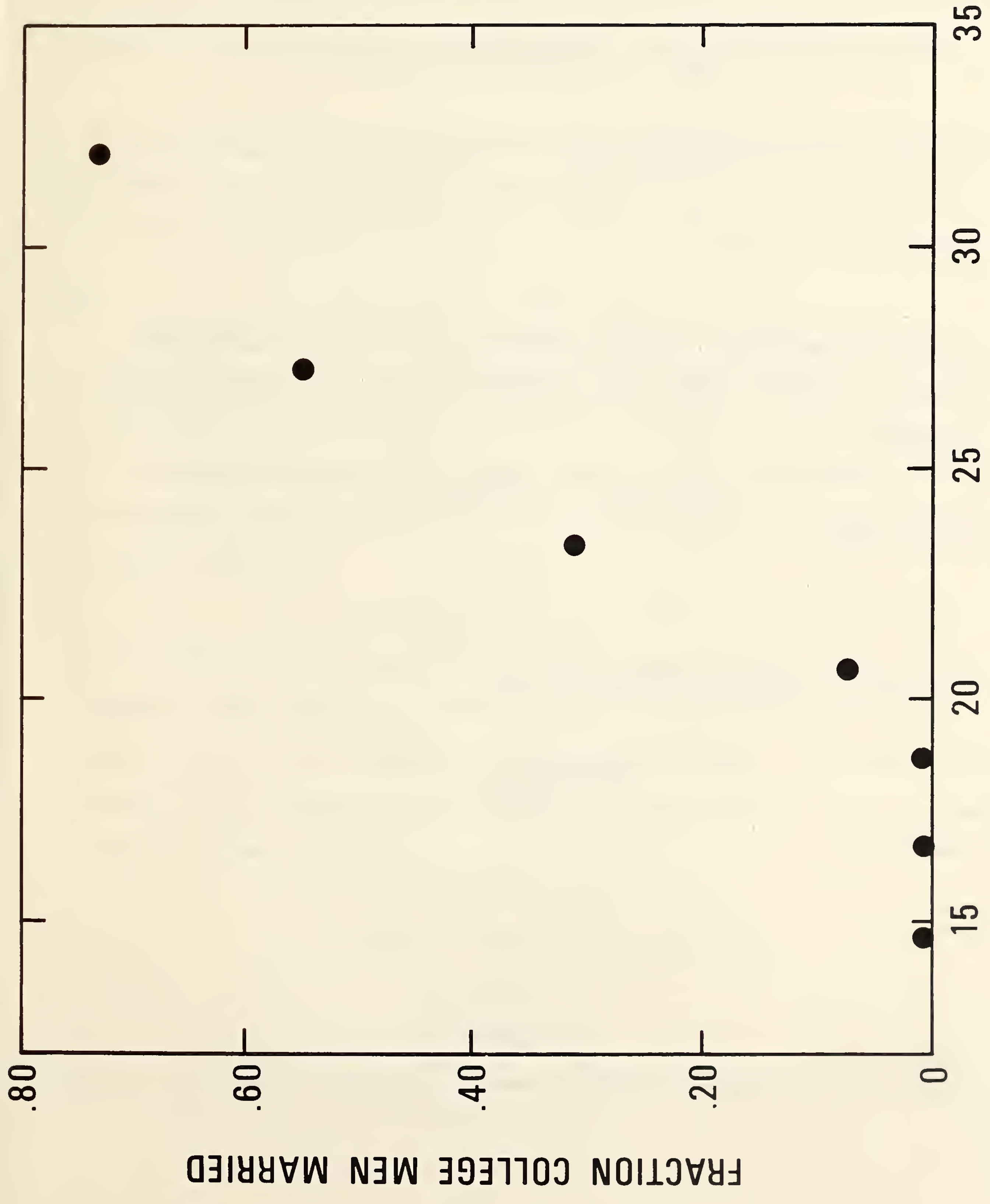

Figure 6. Plotted Data for Example

The optimization problem to be solved is

$$\min_{(\mu,\delta,p)} \sum_{i=1}^{7} \{y_i - p\Phi([x_i - \mu]/\sigma)\}^2. \tag{7.1}$$

As a first step in converting this to a separable problem set

$$v_i = p\Phi\ [(x_i - \mu)/\sigma],\ i=1,\ \ldots,\ 7, \tag{7.2}$$

and

$$z_i = (x_i - \mu)/\sigma \quad,\ i=1,\ \ldots,\ 7. \tag{7.3}$$

A trick not mentioned in section 3 but one which is often useful is to transform a product into a sum by taking logarithms. Doing this to (7.2) yields

$$\ln v_i = \ln p + \ln \Phi(zi),\ i=1,\ \ldots,\ 7.$$

By setting

$$\mu = \omega_1 + \omega_2,$$

$$1/\sigma = \omega_1 - \omega_2,$$

the original nonseparable problem becomes

$$\text{minimize} \sum_{i=1}^{7} (y_i - v_i)^2$$

subject to

$$1_n\ p - \ln v_i + \ln \Phi(z_i) = 0,\ i=1,\ \ldots,\ 7$$

$$x_i(\sigma^{-1}) - z_i - (\omega_1)^2 + (\omega_2)^2 = 0,\ i=1,\ \ldots,\ 7$$

$$\mu - \omega_1 - \omega_2 = 0$$

$$\sigma^{-1} - \omega_1 + \omega_2 = 0$$

where the minimization is performed over (μ, $1/\sigma$, ln p, $\{v_i\}$, $\{z_i\}$, ω_1, ω_2).

Note that the separated problem contains $1/\sigma$ and $\ln p$ instead of σ and p. This was done because the new formulation is equivalent and is linear in the transformed parameters (which are variables of the optimization problem).

From the physical model it is obvious that the parameters are constrained as $\mu>0$, $\sigma>0$, $0<p<1$. If the data is consistent with the model it is expected that these constraints are implicit and will not be binding at the solution.

One requirement of the global minimization technique outlined in Section 1 is that lower and upper bounds are required on the problem variables. It should be pointed out that this can be an advantage since often, from knowledge of the problem, good bounds are available. This can reduce the region in which optimization takes place and thus cut down on the amount of computer effort required to solve it.

From the data, obvious bounds on the original three parameters are: $18<\mu<30$, $1<\sigma<10$, and $.5<p<1$. The techniques of Section 4 are then used to place bounds on the new variables of the separated problem. The names of the variables of the separated problem, their x_j indices and their upper and lower bounds are given in Table 3.

Variable Index j	Variable Name	Lower Bound	Upper Bound	Cuts Run 1	Cuts Run 2
1	μ	18.	30.	1	1
2	$1/\sigma$	.1	1.	1	1
3	$\ln p$	-.693	0.	1	1
4	v_1	.0006	.37	5	5
5	v_2	.0006	.45	5	5
6	v_3	.0006	.76	5	5
7	v_4	.0006	1.	5	5
8	v_5	.0006	1.	5	5
9	v_6	.0009	1.	5	5
10	v_7	.280	1.	5	5
11	z_1	(-3.)	-.34	3	5
12	z_2	-3.	-.12	3	5
13	z_3	-3.	0.7	3	5
14	z_4	-3.	2.6	3	5
15	z_5	-3.	3.	3	5
16	z_6	-2.9	3.	3	5
17	z_7	.2	3.	3	5
18	ω	8.0	16.	5	5
19	ω	8.0	16.	5	5

Table 3: Data for Least Squares Problem in Separable Form

A modified Newton method was applied directly to problem (7.1) and gave the solution

$$\mu = 24.4, \sigma = 3.28, p = 0.73.$$

Two computer runs were made on the separated problem. The data files and subroutine are in Appendix B.

The number of cuts for each of the nineteen variables are given in Table 3. The number of Θ-variables in the linear programming problems associated with each of the variables in the separated problem is equal to the number of cuts plus one. Thus in Run 1 each linear programming problem had 88 variables, and in Run 2 each had 102 variables. The lower bound for Z_1 in Run 1 was -3 and it was -5 in Run 2.

For Run 1 the approximate solution was obtained after the solution of 41 linear programming poblems and was

$$\mu = 22.19, \sigma = 4.28, \text{ and } p = .57.$$

For Run 2 the approximate solution was obtained after the solution of 62 linear programming problems with values

$$\mu = 22.54, \sigma = 3.31, \text{ and } p = .72.$$

These results are summarized in Table 4.

Method	μ	σ	p	# of Variables	MP Problems Solved
Newton	24.14	3.28	0.73	3	1(NLP)
Run 1	22.19	4.28	0.57	88	41(LP)
Run 2	22.54	3.31	0.72	102	62 (LP)

Table 4: Data Relating to Solution of
of Least Squares Problem

8. Discussion of Results and Areas of Future Research

In this paper it was shown how general nonlinear programming problems can automatically be converted to equivalent separable problems when they consist of factorable functions. Global solutions to linear approximations of these separable programs are then obtained using a branch and bound technique with linear programming subproblems. Two small examples were solved this way illustrating the feasibility of the approach.

First, experiments should be made on the placing of the points which determine the piece-wise linear approximations to the nonlinear functions of a single variable. In solving example (3.1), refinement of the grid from Run 1 to Run 2 did not result in any improvement in the solution approximation (see Table 1). The final answer obtained (Run 3) only agreed with the correct solution to one significant place even though the number of 0 variables approximating the nonlinear functions was increased significantly. The reason for this was that the grid refinement strategy used allowed only for equally spaced intervals. Judicious placement should have as a result better accuracy and fewer 0-variables.

Recent efforts in this area are encouraging. In [Grotzinger 1981] problem (1.2) was solved to the accuracy (4.20,0.) requiring 11 linear programming problems none of which had more than 25 0-variables.

The second area of improvement is to lower the effort required to solve the linear programming problems by using the inverse matrix (basis) information available from similar LP's solved by the branch and bound procedure. This should considerably reduce the computation time. No advantage of available partial basis information was used in solving the problems.

Advantage can also be taken of this information when new variables are added to improve the linear approximation.

Until efficient nonlinear programming codes are available, using linear programming codes as subroutines to solve general nonlinear programming problems seems a reasonable approach and one which could be implemented on a production basis without much additional effort.

9. References

1. Beale, E.M.L., and John Tomlin, Special Facilities in a General Mathematical Programming System for Nonconvex Problems Using Ordered Sets of Variables in (J. Lawrence ed.) Proc. of the Fifth International Conference on Operations Research (1970) pp. 447-454, Tavistock Publication, London.

2. Falk, James E. and Richard M. Soland, An Algorithm for Separable Nonconvex Programming Problems, Man. Sc. 15 (1969) pp. 550-569.

3. Falk, J.E., An Algorithm for Locating Approximate Global Solutions of Nonconvex, Separable Problems, Technical Paper Serial T-262, Institute for Management Science and Engineering, the George Washington University, Washington, DC, April 1972.

4. Ghaemi, Abofazl, and Garth P. McCormick, Symbolic factorable SUMT: What is it? How is it used?, Technical Paper Serial T-402, Institute for Management Science and Engineering, The George Washington University, Washington, DC, May 1979.

5. Goldfeld, Stephen M., and Richard E. Quandt, Nonlinear Methods in Econometrics North-Holland, Amsterdam, 1972.

6. Grotte, Jeffrey H., A Computer Program for Solving Separable Nonconvex Optimization Problems, IDA Paper P-1318, Institute for Defense Analysis, Arlington, Virginia, January 1978.

7. Grotzinger, Stephen J., Personal Communication.

8. Hoffman, Karla L., NUGLOBAL--USER'S GUIDE, Technical Memorandum Serial TM-64866, Institute for Management Science and Engineering, The George Washington University, March 1975.

9. Hoffman, Karla L., A Method for Globally Minimizing Concave Functions Over Convex Sets, Math. Prog. 20 (1981), pp. 22-32.

10. Mancini, Louis J. and Garth P. McCormick, Bounding Global Minima, Math. Oper. Res. 1 (1976) pp. 50-53.

11. Mancini, Louis J. and Garth P. McCormick, Bounding Global Minima with Interval Arithmetic, Operations Research 27 (1979) pp. 743-754.

12. McCormick, Garth P., Attempts to Calculate Global Solutions of Problems that May Have Local Minima, in (F.A. Lootsma, ed.) Numerical Methods for Nonlinear Optimization, 1972, Academic Press, New York, pp. 209-221.

13. McCormick, Garth P., Converting General Nonlinear Programming Problems to Separable Nonlinear Programming Problems, Technical Paper Serial T-267, Institute for Management Science and Engineering, The George Washington University, Washington, DC, June 1972.

14. McCormick, Garth P., Optimal Design of a Corrugated Transverse Bulkhead; An Example of the Use of the Factorable Programming Language, Technical Paper Serial T-313, Institute for Managements Science and Engineering, The George Wahington University, Washington, DC, 1975.

15. McCormick, Garth P., Computability of Global Solutions to Factorable Nonconvex Programs: Part I--Convex Underestimating Problems, Math. Prog. 10 (1976) pp. 147-175.

16. McCormick, Garth P., Locating An Isolated Global Minimizer of a Constrained Nonconvex Program, Math. of Oper. Res., 5(1980) pp. 435-443.

17. Soland, Richard M., An Algorithm for Separable Nonconvex Programming Problems II: Nonconvex Constraints, Man. Sc. 17 (1971) pp. 759-773.

NBS-114A (REV. 2-80)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET *(See instructions)*	1. PUBLICATION OR REPORT NO. NBSIR 85-3206	2. Performing Organ. Report No.	3. Publication Date JULY 1985

4. TITLE AND SUBTITLE

Global Solutions to Factorable Nonlinear Optimization Problems Using Separable Programming Techniques

5. AUTHOR(S)

Garth P. McCormick

6. PERFORMING ORGANIZATION *(If joint or other than NBS, see instructions)*

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

7. Contract ~~Grant No.~~

N00014-75-C-0611

8. Type of Report & Period Covered

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS *(Street, City, State, ZIP)*

National Bureau of Standards
Center for Applied Mathematics
Operations Research Division
Gaithersburg, MD 20899

10. SUPPLEMENTARY NOTES

□ Document describes a computer program; SF-185, FIPS Software Summary, is attached.

11. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here)*

Many algorithms for obtaining global solutions to nonconvex optimization problems have been proposed in recent years. The methods farthest along computationally are those for separable problems. These use linear programming codes to solve sequences of LP problems formed from piece-wise linear approximations to the nonlinear functional forms. For a large class of optimization problems, called factorable programming problems, it is possible to create equivalent separable problems. This is done at a cost: additonal variables and constraints. In this paper the procedure for creating the equivalent separable problems is outlined and a brief description is given of a global solution algorithm due to Falk. A small example is given illustrating the above techniques. The example is also solved using a more direct method. Application to the solution of nonlinear least squares is illustrated with another example. Discussion of areas of research for improving the efficiency of this approach concludes the paper.

12. KEY WORDS *(Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons)*

branch and bound; factorable functions; mathematical programming; nonconvex programming; nonlinear least squares; nonlinear optimization; separable optimization.

13. AVAILABILITY

☒ Unlimited

□ For Official Distribution. Do Not Release to NTIS

□ Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.

☒ Order From National Technical Information Service (NTIS), Springfield, VA. 22161

14. NO. OF PRINTED PAGES

45

15. Price

$8.50